DISSERTATION

TRANSFORMED-LINEAR MODELS FOR TIME SERIES EXTREMES

Submitted by

Nehali Mhatre

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: Daniel Cooley

Piotr Kokoszka
Benjamin Shaby
Tianyang Wang

ABSTRACT

TRANSFORMED-LINEAR MODELS FOR TIME SERIES EXTREMES

In order to capture the dependence in the upper tail of a time series, we develop nonnegative regularly-varying time series models that are constructed similarly to classical non-extreme ARMA models. Rather than fully characterizing tail dependence of the time series, we define the concept of weak tail stationarity which allows us to describe a regularly-varying time series through the tail pairwise dependence function (TPDF) which is a measure of pairwise extremal dependencies. We state consistency requirements among the finite-dimensional collections of the elements of a regularly-varying time series and show that the TPDF's value does not depend on the dimension being considered. So that our models take nonnegative values, we use transformed-linear operations. We show existence and stationarity of these models, and develop their properties such as the model TPDF's. Motivated by investigating conditions conducive to the spread of wildfires, we fit models to hourly windspeed data using a preliminary estimation method and find that the fitted transformed-linear models produce better estimates of upper tail quantities than traditional ARMA models or than classical linear regularly-varying models.

The innovations algorithm is a classical recursive algorithm used in time series analysis. We develop an analogous transformed-linear innovations algorithm for our time series models that allows us to perform prediction which is fundamental to any time series analysis. The transformed-linear innovations algorithm also enables us to estimate parameters of the transformed-linear regularly-varying moving average models, thus providing a tool for modeling.

We construct an inner product space of transformed-linear combinations of nonnegative regularly-varying random variables and prove its link to a Hilbert space which allows us to employ the projection theorem. We develop the transformed-linear innovations algorithm using the properties of the projection theorem. Turning our attention to the class of MA(∞) models, we talk about esti-

mation and also show that this class of models is dense in the class of possible TPDFs. We also develop an extremes analogue of the classical Wold decomposition. Simulation study shows that our class of models provides adequate models for the GARCH and another model outside our class of models.

The transformed-linear innovations algorithm gives us the best prediction and we also develop prediction intervals based on the geometry of regular variation. Simulation study shows that we obtain good coverage rates for prediction errors. We perform modeling and prediction for the hourly windspeed data by applying the innovations algorithm to the estimated TPDF.

ACKNOWLEDGEMENTS

Words cannot express my gratitude to my advisor, Dan Cooley. I feel fortunate to have Dan as a mentor. He has been a completely supportive advisor who never made me feel stressed out about my research. His dedication towards his students is commendable and inspiring. I thank him for the vast knowledge of time series and extremes that I have acquired from him. He has been an incredible listener and has always been patient with all my questions. He is instrumental in preparing me for great things in the future.

I would like to thank my committee members, Piotr Kokoszka, Ben Shaby, and Tianyang Wang for their valuable discussions that helped me towards my dissertation. Special thanks to Jay Breidt for taking the time to chat about time series. I am thankful to Julia Sharp for all that I learned from her about collaborative research and technical writing. A special thanks to Aaron Nielson for always encouraging me and for appreciating my hard work as a TA. A very special thanks to Jeongjin Lee. Grad school would not have been the same without a friend and colleague like him.

I am thankful to the CSU Department of Statistics for their financial support. I also acknowledge financial support from National Science Foundation grant DMS-1811657.

I would not have been able to achieve any of this without the love and support of my family. My heartiest thanks to Jaishree Mhatre, who has been the most wonderful mother and the greatest friend I could ever expect. My father, Late Ashok Mhatre, who loved me and cared for me dearly. My uncle, Bhushan Mhatre, who is responsible for laying the foundations for my academic excellence and who has always inspired me to keep learning. My aunt, Piroz Shanbhag, who always inspired me to do better in life. My cousin brother, Aditya Shanbhag, who has been a constant support throughout my life, always looking out for me. I would be remiss in not mentioning my sister-in-law, Rucha Shanbhag, for always encouraging me, and my cutest nephew, Ryansh Shanbhag, for not forgetting me even though I am not around. I would like to thank my family-in-law, Sanjay Kirve and Swapnal Kirve, for their unconditional love and support. I would also like to thank all my friends for always being there for me. Especially my RamFam friends, thank

you for all the fun that kept me sane. Finally, I am grateful to my husband, Ameya Kirve, for his incessant love and support that helped me through this PhD journey. You constantly encouraged me and believed in me more than I ever believed in myself. I could not have done this without you. Thanks for everything.

DEDICATION

This dissertation is dedicated to

my father, Late Ashok Mhatre,

and

my grandmother, Late Sumati Mhatre.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

Measuring risk associated with rare events that are extreme in magnitude requires extreme value methods. Univariate extremes methods are well-developed, but there is a need to develop easily implementable statistical methods that can describe and model extremal dependence in the time series context. This dissertation will use transformed-linear operations to construct straightforward and flexible models for nonnegative regularly-varying time series.

To motivate our models, consider the time series of hourly windspeeds (m/s) observed at the March AFB station in Southern California (HadISD (refer Dunn (2019)); revisited in Section 3.9.1). High windspeeds are one of the factors that contribute to wildfire risk. Prolonged exposure to high winds could be a quantity of interest for firefighters. For instance, we can consider the average run-length in hours over a high windspeed threshold or the mean windspeed over a 12-hour period. Figure 1.1 (upper panel) shows the hourly windspeeds for December 1-10, 2017. The windspeeds increase for a period of time beginning on December 4, the ignition day of the Thomas Fire, a wildfire attributed to the weather phenomenon known as the Santa Ana winds.

Windspeed data are not stationary, as there is a diurnal cycle. To remove the diurnal behavior, we subtract off the mean from each hour of a 24-hour cycle creating a time series of windspeed anomalies that can reasonably be assumed stationary (Figure 1.1, lower panel, data available *here*). Our goal is to model the upper tail behavior of this time series. Exploratory analysis indicates the data at short lags exhibit asymptotic dependence. Two realizations X_t and X_{t+h} of a stationary time series with marginal distribution F are said to be asymptotically independent if $\chi(h) = \lim_{u \to 1} P\{F(X_{t+h}) > u | F(X_t) > u\} = 0$, and asymptotically dependent if $\chi(h) \neq 0$ with the value of $\chi(h)$ summarizing the magnitude of pairwise tail dependence at lag h. A chiplot (not shown) for the upper tail shows that $\hat{\chi}(1) \approx 0.55$ which is relatively stable at the higher

quantiles. We do not use $\chi(\cdot)$ to describe asymptotic dependence in our time series. Instead we define a different measure of pairwise tail dependence and show evidence for apparent asymptotic dependence at lags greater than 1. Classical time series models that do not focus on extremes tend to not capture asymptotic dependence. Extremes specific models such as the model based on Markov assumptions (Smith et al. (1997)) capture asymptotic dependence and also fit into the regular variation framework that we use to model asymptotic dependence in our time series.

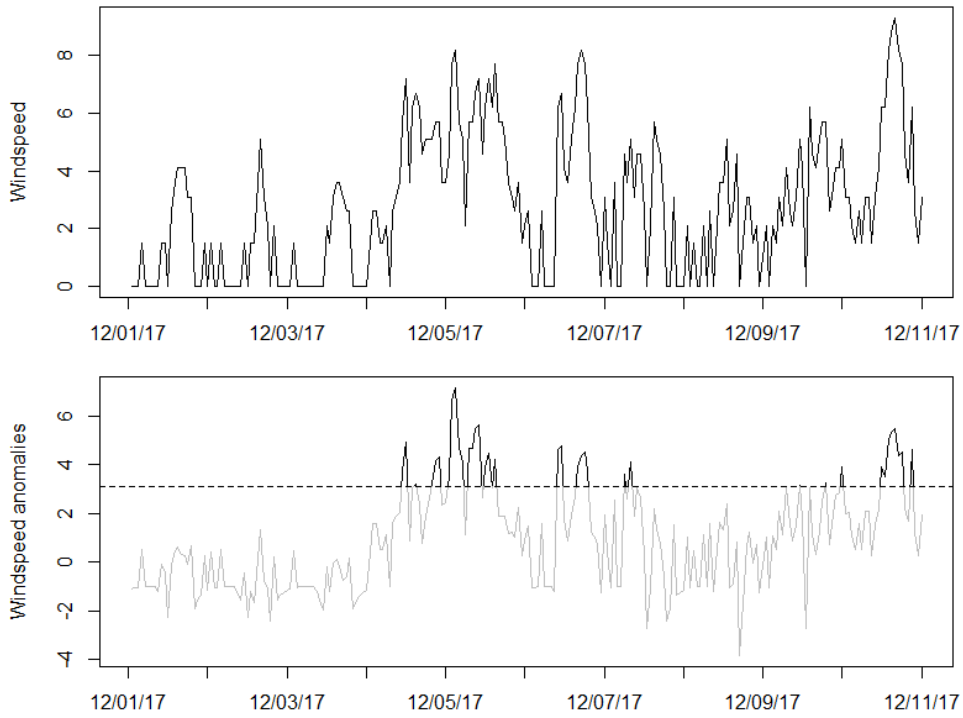

Figure 1.1: Time series of hourly windspeed (m/s) for 12/01/2017 through 12/11/2017. The Thomas fire started on 12/04/2017. Original windspeeds (upper panel) and windspeed anomalies (lower panel): bold above a threshold of 3.09 (95^{th} percentile).

Linear time series models, which include the familiar ARMA models, have many attractive qualities such as simplicity, interpretability, and widespread familiarity. Much of classical time series analysis can be done with only an assumption of weak or second-order stationarity, which allows one to focus on the autocovariance function (ACVF). Mean and covariance are poorly-suited to describing tail behavior as they describe the behavior near the center of the distribution. If the

time series is assumed Gaussian, then elements of the time series are asymptotically independent. For these reasons, this classical approach is poorly suited for describing extremal behavior, and classical ARMA models are not designed with tail behavior in mind.

1.2 Outline

In this dissertation, we build time series models via transformed-linear combinations of regularly-varying terms. The framework of regular variation is useful as it is defined only in terms of the tail, it is naturally linked to extreme value theory, and it is well-suited for modeling data that exhibit asymptotic dependence. Models take only nonnegative values in order to focus attention on the upper tail. To construct models that feel similar to linear models, we use transformed-linear operations which define addition and multiplication in a manner such that nonnegative values are obtained.

We develop an extremes analogue to weak stationarity in Chapter 2. We construct transformed linear regularly-varying analogues to ARMA models and study their properties in Chapter 3. We will show that fitting a transformed-linear model is done easily, and that our fitted model captures the tail dependence in the windspeed data better than a Gaussian or two classical linear regularly-varying models. In Chapter 4 we discuss prediction for our transformed-linear regularly-varying time series models.

1.3 Regular Variation

One of many formal definitions says X is a regularly-varying random variable, denoted $X \in RV(\alpha)$, if

$$\Pr\left(|X| > x\right) = L(x)x^{-\alpha},$$

for $x > 0$ and a slowly varying function L, that is, $\lim_{x\to\infty} L(ax)/L(x) = 1$ for all $a > 0$. The tail index α describes the power-law-like behavior of the tail. The upper and lower tails can be

described individually as

$$\Pr(X > x) = uL(x)x^{-\alpha} \text{ and } \Pr(X < -x) = vL(x)x^{-\alpha}, \tag{1.1}$$

where $u, v \geq 0, u + v = 1$.

X is a p-dimensional multivariate regularly-varying random vector ($X \in RV^p(\alpha)$) if there exists a non-trivial limit measure $\nu_X(\cdot)$ and function $b(s) \to \infty$ as $s \to \infty$, such that

$$s\Pr\left(\frac{X}{b(s)} \in \cdot\right) \xrightarrow{v} \nu_X(\cdot) \text{ as } s \to \infty, \tag{1.2}$$

where $\xrightarrow{v}$ denotes vague convergence in $M_+(\mathbb{R}^p \setminus \{0\})$, the space of nonnegative Radon measures on $\mathbb{R}^p \setminus \{0\}$ (refer Resnick (2007, Section 6)). The normalizing function is of the form $b(s) = U(s)s^{1/\alpha}$ where $U(s)$ is slowly varying. The limiting measure ν_X has scaling property

$$\nu_X(aC) = a^{-\alpha}\nu_X(C),$$

for any $a > 0$ and any set $C \subset \mathbb{R}^p \setminus \{0\}$. When described in polar coordinates, ν_X decomposes into independent radial and angular components. Given any norm, define the unit ball $\mathbb{S}_{p-1} = \{x \in \mathbb{R}^p : \|x\| = 1\}$. Let $C(r, B) = \{x \in \mathbb{R}^p : \|x\| > r, \|x\|^{-1}x \in B\}$ for some $r > 0$, and some Borel set $B \subset \mathbb{S}_{p-1}$. Then $\nu_X\{C(r, B)\} = r^{-\alpha}H_X(B)$ where H_X is the angular measure taking values on $\mathbb{S}_{p-1}$. Equivalently,

$$\nu(dr \times dw) = \alpha r^{-\alpha-1}drdH_X(w).$$

The normalizing function $b(s)$ and measures ν_X or H_X are not uniquely defined by (1.2), as $b(s)$ can be scaled by any positive constant which can be absorbed into the limiting measure.

Modeling with the regular variation framework usually involves thresholding in terms of the norm or the marginals to extract a subset of extreme data, and using this subset of threshold exceedances for inference. Modeling is primarily concerned with the angular measure H, which gets difficult to model at high dimensions. The regular variation framework is often used to model

heavy-tailed data. However, the framework has been used to model extremal dependence for data which is not heavy-tailed by viewing it as a copula-like dependence framework for extreme behavior. This approach is defended by Proposition 5.10 of Resnick (1987) which says the fundamental nature of tail dependence is unchanged by monotone transformation of the marginals.

1.4 Transformed-Linear Operations for Regularly-Varying Random Vectors

Reconciling linear operations with positive random variables or vectors is not simple. For example, regular variation restricted to the positive orthant cannot accommodate multiplication by a negative number. Max-linear approaches (e.g., Davis & Resnick (1989), Strokorb & Schlather (2015)) have been used to build max-stable models, but linear-algebra operations do not have direct analogues due to the maximum operation.

In the finite dimensional setting, Cooley & Thibaud (2019) propose a way to link traditional linear algebra operations to regular variation on the positive orthant. For two vectors X_1 and $X_2 \in \mathbb{R}_+^p$ and $f : \mathbb{R} \mapsto \mathbb{R}_+$ applied componentwise to vectors, transformed-linear summation is

$$X_1 \oplus X_2 := f\{f^{-1}(X_1) + f^{-1}(X_2)\}$$

and scalar multiplication is

$$a \circ X_1 := f\{af^{-1}(X_1)\},$$

for $a \in \mathbb{R}$. Using the transform $f(y) = \log\{1 + \exp(y)\}$, Cooley & Thibaud (2019) apply transformed-linear operations to $X \in RV_+^p(\alpha)$ and show that regular variation is "preserved", so long as the lower tail condition

$$s\Pr\left[X_i \leq \exp\{-kb(s)\}\right] \to 0,\ k > 0,\ i = 1, ..., p;\ s \to \infty, \tag{1.3}$$

is met. The lower tail condition prevents any of the marginals from having nonzero mass at 0. Standard regularly-varying distributions such as the Frèchet meet this condition. If $X_1, X_2 \in RV_+^p(\alpha)$ are independent, both with normalizing function $b(s)$ and respective limiting measures ν_{X_1} and ν_{X_2}, then

$$s\Pr\left(\frac{X_1 \oplus X_2}{b(s)} \in \cdot\right) \xrightarrow{v} \nu_{X_1}(\cdot) + \nu_{X_2}(\cdot), \text{ and} \tag{1.4}$$

$$s\Pr\left(\frac{a \circ X_1}{b(s)} \in \cdot\right) \xrightarrow{v} \begin{cases} a^\alpha \nu_{X_1}(\cdot) & \text{if } a > 0, \\ 0 & \text{if } a \leq 0, \end{cases} \text{ as } s \to \infty. \tag{1.5}$$

Any transform f that has a negligible effect on large values, that is,

$$\lim_{y\to\infty} \frac{f(y)}{y} = \lim_{x\to\infty} \frac{f^{-1}(x)}{x} = 1$$

and preserves regular variation is a possible candidate function with a corresponding different lower tail condition.

Chapter 2

Regularly-Varying Time Series and Weak Tail Stationarity

2.1 Regularly-Varying Time Series

The recent volume by Kulik & Soulier (2020) provides a comprehensive treatment of regularly-varying time series. We restrict attention to univariate time series. Following Kulik & Soulier (2020, Definition 5.1.1), define a regularly-varying time series as a sequence $\{X_t\}$, $t \in \mathbb{Z}$, of real-valued random variables whose finite-dimensional distributions are regularly-varying. Letting $X_{t,p} = (X_t, X_{t+1}, \ldots, X_{t+p-1})^T$ for any t and $p > 0$, $X_{t,p}$ is multivariate regularly-varying and there exists a normalizing sequence $b(s) \to \infty$ as $s \to \infty$ yielding limit measure $\nu_{X_{t,p}}(\cdot)$ and angular measure $H_{X_{t,p}}$ on $\mathbb{S}_{p-1}$.

Having defined finite-dimensional limiting measures, Kulik & Soulier (2020, Chapter 5) go on to define and prove existence of the tail measure ν, which is the infinite-dimensional limiting measure of the time series. Kulik & Soulier (2020) then characterize the stochastic properties of the time series by focusing on its behavior respective to $t = 0$. They define the tail process as a random element $\{Y_t\}$, $t \in \mathbb{Z}$, with values in $E_0 = \{y \in (\mathbb{R})^{\mathbb{Z}} : |y_0| > 1\}$ and distribution η defined by $\eta = \nu(\cdot \cap E_0)$. Kulik & Soulier (2020) characterize the time series focusing on the spectral tail process $\{\Theta_t = Y_t/|Y_0|\}$, $t \in \mathbb{Z}$, which is independent of $|Y_0|$.

Rather than fully characterizing the tail dependence via the tail measure ν or the spectral tail process Θ_t, we will characterize a time series only via pairwise tail dependency summary measures. While such a specification does not contain the full information of ν or Θ_t, we believe that a time series model which matches these summary measures of dependence is useful for answering many questions and that it can be difficult to infer the full tail dependence structure from data. Furthermore, our approach nicely ties to traditional linear time series modeling, which often focuses

only on second-order properties. To do so, we require a notion of weak tail stationarity (Section 2.3), contrary to strict stationarity discussed in Kulik & Soulier (2020).

2.2 Consistency between finite-dimensional measures

We present consistency results required of regularly-varying time series. These consistency results show that our tail dependence summary measure is sensible regardless of the dimension of the random vector under consideration. Our presentation only requires consideration of the finite-dimensional distributions.

To focus attention on the upper tail, consider nonnegative time series $\{X_t\}$, $t \in \mathbb{Z}$. Then $X_{t,p}$, taking values in $\mathbb{R}_+^p := [0,\infty)^p$, is multivariate regularly-varying with tail index α, denoted $X_{t,p} \in RV_+^p(\alpha)$. $X_{t,p}$ has limiting measure $\nu_{X_{t,p}}$ for sets in $\mathbb{R}_+^p \setminus \{0\}$ and angular measure $H_{X_{t,p}}$ on $\mathbb{S}_{p-1}^+ = \{x \in \mathbb{R}_+^p : \|x\| = 1\}$.

Let $X_{t,p}^{(-i)}$ be the $(p-1)$-dimensional vector obtained by excluding the i^{th} component of $X_{t,p}$. Then for $A_1 \in \mathbb{R}_+^{i-1}$, $A_2 \in \mathbb{R}_+^{p-i}$, and $\{0\} \notin A_1 \times A_2$, consistency across dimensions requires

$$\nu_{X_{t,p}}(A_1 \times [0,\infty] \times A_2) = \nu_{X_{t,p}^{(-i)}}(A_1 \times A_2),$$

or equivalently,

$$\int_{(r,w)\in A_1\times[0,\infty]\times A_2} \alpha r^{-\alpha-1} \mathrm{d}r \mathrm{d}H_{X_{t,p}}(w) = \int_{(r,v)\in A_1\times A_2} \alpha r^{-\alpha-1} \mathrm{d}r \mathrm{d}H_{X_{t,p}^{(-i)}}(v),$$

where $w \in \mathbb{S}_{p-1}^+$ and $v \in \mathbb{S}_{p-2}^+$. In Proposition 1, the representation of the lower dimensional marginal angular measure is derived. For ease and generality, we temporarily drop time series notation, letting X be a p-dimensional random vector.

Proposition 1. *Let $X = (X_1, \cdots, X_p)^T \in RV_+^p(\alpha)$ with angular measure H_X on $\mathbb{S}_{p-1}^+$. Let $X_{(l)} \in RV_+^l(\alpha)$ be the marginal l-dimensional random vector, $l < p$, with angular measure $H_{X_{(l)}}$ on $\mathbb{S}_{l-1}^+$. Let $A_{(l)}(r, B_{l-1}) = \{X_{(l)} \in \mathbb{R}_+^l : \|X_{(l)}\| > r, \|X_{(l)}\|^{-1} X_{(l)} \in B_{l-1}\}$ where $B_{l-1} \subset \mathbb{S}_{l-1}^+$.*

Let $A^*_{(l)}(r, B_{l-1}) = \{X \in \mathbb{R}^p_+ : \|X_{(l)}\| > r, \|X_{(l)}\|^{-1} X_{(l)} \in B_{l-1}\}$. *Then,*

$$H_{X_{(l)}}(B_{l-1}) = \int_{w \in \mathbb{S}^+_{p-1} : \|w_{(l)}\|^{-1} w_{(l)} \in B_{l-1}} \|w_{(l)}\|^\alpha dH_X(w). \tag{2.1}$$

Proof: By the relation between limiting measure and angular measure, we have that

$$\nu_{\mathbf{X}_{(l)}}\{A_{(l)}(r_0, B_{l-1})\} = r_0^{-\alpha} H_{\mathbf{X}_{(l)}}(B_{l-1}).$$

Consider the limiting measure $\nu_{\mathbf{X}}\{A^*_{(l)}(r_0, B_{l-1})\}$.

$$\begin{aligned}
\nu_{\mathbf{X}}\{A^*_{(l)}(r_0, B_{l-1})\} &= \int_{\mathbf{w} \in \mathbb{S}^+_{p-1} : \|\mathbf{w}_{(l)}\|^{-1} \mathbf{w}_{(l)} \in B_{l-1}} \int_{r = r_0/\|\mathbf{w}_{(l)}\|}^{\infty} \alpha r^{-(\alpha+1)} \mathrm{d}r \mathrm{d}H_{\mathbf{X}}(\mathbf{w}) \\
&= r_0^{-\alpha} \int_{\mathbf{w} \in \mathbb{S}^+_{p-1} : \|\mathbf{w}_{(l)}\|^{-1} \mathbf{w}_{(l)} \in B_{l-1}} \|\mathbf{w}_{(l)}\|^\alpha \mathrm{d}H_{\mathbf{X}}(\mathbf{w}).
\end{aligned}$$

By consistency, $\nu_{\mathbf{X}}\{A^*_{(l)}(r_0, B_{l-1})\} = \nu_{\mathbf{X}_{(l)}}\{A_{(l)}(r_0, B_{l-1})\}$, and (2.1) follows. □

2.3 Tail Stationarity defined via the Tail Pairwise Dependence Function

We define a notion of weak tail stationarity for regularly-varying time series, analogous to weak or second-order stationarity of non-extreme time series discussed in Brockwell & Davis (1991) and elsewhere. Assuming a constant mean, classical time series analysis focuses on the ACVF which summarizes pairwise dependence, and weak stationarity implies the ACVF is a function only of lag.

For tail stationarity, we likewise require a measure of pairwise dependence. Many pairwise extremal dependence measures have been suggested. For example, $\chi(h)$ is a specific example of the more general extremogram of Davis & Mikosch (2009). However, unlike the ACVF, $\chi(h)$ is not nonnegative-definite. So that our dependence measure has properties similar to covariance, we will henceforth assume $\{X_t\}$ is a regularly-varying time series with $\alpha = 2$ and we will use the L_2

norm to define the angular measures $H_{X_{t,p}}$ which are defined on $\Theta_{p-1}^{+} = \{x \in \mathbb{R}_{+}^{p} : \|x\|_2 = 1\}$. Considering the vector $(X_t, X_{t+h})^T$, define the tail pairwise dependence function (TPDF) as

$$\sigma(X_t, X_{t+h}) = \int_{\Theta_1^+} w_t w_{t+h} \mathrm{d}H_{X_t, X_{t+h}}(w). \tag{2.2}$$

A nonnegative regularly-varying time series is weakly tail stationary if $\sigma(X_t, X_{t+h}) = \sigma(h)$ for all t.

The TPDF is essentially the extremal dependence measure defined by Larsson & Resnick (2012), but for the special case of $\alpha = 2$ and using L_2-norm to define the angular measure. As we will show in Section 3.4, these restrictions will be important to link the TPDF to an inner product, which connects our models to Hilbert space ideas in time series (see Brockwell & Davis (1991, Chapter 2)). We also do not impose that the angular measure be a probability measure on Θ_1^+, as this would invalidate the relationship between angular measures as given in (2.1). The assumption that $\alpha = 2$ may seem restrictive; however, we will show in Section 3.9 that our models are widely applicable.

In the finite dimensional setting, Cooley & Thibaud (2019) constructed the tail pairwise dependence matrix (TPDM) defined in terms of the p-dimensional joint angular measure of random vector $X \in RV_{+}^{p}(2)$, and showed that this matrix was nonnegative definite. In the time series setting, there is no "correct" higher dimension to consider and therefore we consider the bivariate angular measures as in (2.2). Recall that a real-valued function κ defined on the integers is nonnegative definite if the matrix $K = \{\kappa(i-j)\}_{i,j=1}^{p}$ is nonnegative definite for any p; i.e., $a^T K a \geq 0$ for all $a \in \mathbb{R}^p \setminus \{0\}$. Proposition 2 will first show that the TPDF defined in terms of the bivariate angular measure is equivalent to that defined in terms of the higher dimensional angular measure. We again speak of a general p-dimensional vector X.

Proposition 2. *Let $X = (X_1, \cdots, X_p)^T \in RV_{+}^{p}(2)$, $p > 2$ with limiting measure ν_X on $[0,\infty]^p \setminus \{0\}$ and angular measure H_X on Θ_{p-1}^{+}. Let $(X_i, X_j)^T \in RV_{+}^{2}(2)$ be a marginal random vector*

with $\nu_{(X_i,X_j)}$ *on* $[0,\infty]^2 \setminus \{0\}$ *and* $H_{(X_i,X_j)}$ *on* Θ_1^+. *Then,*

$$\sigma(X_i, X_j) = \int_{v\in\Theta_1^+} v_i v_j dH_{(X_i,X_j)}(v) = \int_{w\in\Theta_{p-1}^+} w_i w_j dH_X(w). \tag{2.3}$$

Proof: By construction, $v_i = x_i/\|(x_i, x_j)\|$ and $w_i = x_i/\|\mathbf{x}\|$. Therefore,

$$\begin{aligned} v_i &= w_i \frac{\|\mathbf{x}\|}{\|(x_i,x_j)\|} = w_i \frac{\sqrt{x_1^2+\cdots+x_p^2}}{\sqrt{x_i^2+x_j^2}} = w_i \frac{\sqrt{\frac{x_1^2}{\|\mathbf{x}\|^2}+\cdots+\frac{x_p^2}{\|\mathbf{x}\|^2}}}{\sqrt{\frac{x_i^2}{\|\mathbf{x}\|^2}+\frac{x_j^2}{\|\mathbf{x}\|^2}}} \\ &= w_i \frac{\sqrt{w_1^2+\cdots+w_p^2}}{\sqrt{w_i^2+w_j^2}} = \frac{w_i}{\|(w_i,w_j)\|}. \end{aligned}$$

Similarly, $v_j = w_j/\|(w_i, w_j)\|$. Therefore, by Proposition 1, for $l = 2$ and $\alpha = 2$, we have,

$$\begin{aligned} \sigma(X_i, X_j) &= \int_{\mathbf{v}\in\Theta_1^+} v_i v_j \mathrm{d}H_{(X_i,X_j)}(\mathbf{v}) = \int_{\mathbf{w}\in\Theta_{p-1}^+} \frac{w_i}{\|(w_i,w_j)\|}\frac{w_j}{\|(w_i,w_j)\|}\|(w_i,w_j)\|^2 \mathrm{d}H_{\mathbf{X}}(\mathbf{w}) \\ &= \int_{\mathbf{w}\in\Theta_{p-1}^+} w_i w_j \mathrm{d}H_{\mathbf{X}}(\mathbf{w}). \end{aligned}$$

□

Using Proposition 2, we can show the TPDF $\sigma(h)$ of a weakly tail stationary time series $\{X_t\}$ is nonnegative definite. Let $p > 0$ and $a = (a_1, \ldots, a_p)^T \in \mathbb{R}^p \setminus \{0\}$ be given. Then

$$\sum_{i,j=1}^{p} a_i \sigma(i-j) a_j = a^T \Sigma_{X_{1,p}} a,$$

where $X_{1,p} = (X_1, \ldots, X_p)^T$, and $\Sigma_{X_{1,p}}$ is the matrix whose i, jth element is $\int_{w\in\Theta_{p-1}^+} w_i w_j \mathrm{d}H_{X_{1,p}}(w)$. Let $m = H_X(\Theta_{p-1}^+)$ and let w be a random vector such that

$$\Pr(w \in B_{p-1}) = m^{-1} H_{X_{1,p}}(B_{p-1}),$$

for any set $B_{p-1} \in \Theta^+_{p-1}$. Then by definition, $\Sigma_{X_{1,p}} = m\mathrm{E}(ww^T)$, and thus for any vector $a \in \mathbb{R}^p \setminus \{0\}$,

$$a^T \Sigma_{X_{1,p}} a = m\mathrm{E}(a^T w w^T a) \geq 0.$$

The inequality becomes strict if no element of w is a linear combination of the others.

In the finite dimensional case, Cooley & Thibaud (2019, Proposition 5) additionally show that the TPDM is a completely positive matrix, that is, there exists a $p \times q_*$, $q_* < \infty$, nonnegative matrix A_* such that $\Sigma_{X_{1,p}} = A_* A_*^T$. This notion can be extended to the idea of a completely positive function. Analogous to nonnegative definiteness, we say that a real-valued function κ defined on the integers is completely positive if the matrix $K = \{\kappa(i-j)\}_{i,j=1}^p$ is completely positive for any p. Thus, the TPDF $\sigma(\cdot)$ is a completely positive function since $\Sigma_{X_{1,p}}$ is a completetly positive matrix by Proposition 5 of Cooley & Thibaud (2019).

Proposition 2 also gives a way of defining the TPDF at lag 0. By (1.2),

$$\begin{aligned}\lim_{s\to\infty} s\mathrm{Pr}\left(\frac{X_1}{b(s)} > c\right) &= \int_{\Theta^+_{p-1}} \int_{c/w_1}^{\infty} 2r^{-3}\mathrm{d}r\mathrm{d}H_{X_{1,p}}(w) \\ &= c^{-2}\int_{\Theta^+_{p-1}} w_1^2 \mathrm{d}H_{X_{1,p}}(w) = c^{-2}\sigma(0). \end{aligned} \tag{2.4}$$

It is perhaps useful to think of the TPDF similarly to the ACVF in that its value at lag h can most easily be interpreted with knowledge of the relative scale provided by $\sigma(0)$, as $\sigma(h)/\sigma(0) \in [0,1]$. For a general (perhaps non-stationary) regularly-varying time series, we will find it useful to rewrite the left expression in (2.4) by recalling that $b(s) = U(s)s^{1/2}$ and letting $x = cU(s)s^{1/2}$ to obtain

$$\lim_{x\to\infty} \frac{\mathrm{Pr}(X_t > x)}{x^{-2}L(x)} = k_t, \tag{2.5}$$

where $L(x)$ is a slowly varying function given by $L(x) = (U(s))^2 = (b(s))^2 s^{-1}$. We refer to the left hand side of (2.5) as the tail ratio of X_t. Just as scale information can be passed between $b(s)$ and ν_X in (1.2), there is ambiguity in (2.5) as k_t and $L(x)$ can be scaled by any positive number. This scale ambiguity has been handled in various ways. Larsson & Resnick (2012) (see also Kim

& Kokoszka (2020)) impose that the angular measure be a probability measure, a restriction we find inconvenient for consistency across dimensions. Cooley & Thibaud (2019) assumed a Pareto tail letting $b(s) = s^{1/2}$ pushing all scale information into the angular measure. In Section 3, we will handle this ambiguity as the normalizing sequence $x^{-2}L(x)$ can be defined by the generating noise sequence.

Chapter 3

Transformed-Linear Regularly-Varying Models

3.1 Previous Linear Constructions of Regularly-Varying Time Series

Researchers have been studying linear regularly-varying time series models for decades. Generally, $\{X_t\}$ is defined as

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \tag{3.1}$$

where $\{Z_t\}, t \in \mathbb{Z}$ is an iid sequence of regularly-varying random variables with tail index α. If $\exists$ $\delta \in (0, \alpha) \cap [0, 1]$ such that $\sum_{j=0}^{\infty} |\psi_j|^\delta < \infty$, then $\{X_t\}$ can be shown to converge with probability 1 (refer Embrechts et al. (1997, Section 7.2)). Because $\{Z_t\}$ is iid, $\{X_t\}$ is strictly stationary.

If $\{Z_t\}$ has tail index $0 < \alpha < 2$, characterization of dependence via the ACVF is not possible. Authors still frequently summarize the dependence structure of $\{X_t\}$ by pairwise summary metrics. For a linear time series with infinite variance, Brockwell & Davis (1991, Section 13.3) define the dependence metric

$$\rho(h) := \frac{(\sum_{j=0}^{\infty} \psi_j \psi_{j+h})}{(\sum_{j=0}^{\infty} \psi_j^2)}, \quad h = 1, 2, \cdots .$$

Much of the previous work considers regularly-varying time series $\{X_t\}$ which take both positive and negative values. One could use (3.1) to construct nonnegative time series by setting $v = 0$ in the tail balance condition (1.1) and restricting $\psi_j \geq 0$ for all j. We will show in Section 3.9 that transformed-linear constructions which allow for negative coefficients but which still result in nonnegative regularly-varying time series allow for more flexibility than standard linear constructions restricted to have positive coefficients.

3.2 Transformed-Linear Regularly-Varying Time Series

For application to extremes, we extend familiar linear time series models by considering transformed-linear combinations of regularly-varying terms. We say the time series $\{X_t\}$ is a transformed-linear process of regularly-varying terms if for all t, it can be represented as

$$X_t = \bigoplus_{j=-\infty}^{\infty} \psi_j \circ Z_{t-j}, \tag{3.2}$$

where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, and $\{Z_t\}$ is a noise sequence of independent and tail stationary $RV_+(2)$ random variables. Henceforth, we assume normalizing functions $b(s)$ and $x^{-2}L(x)$ are such that

$$\lim_{s\to\infty} s\text{Pr}\left\{\frac{Z_t}{b(s)} > c\right\} = c^{-2}$$

and

$$\lim_{x\to\infty} \frac{\text{Pr}(Z_t > x)}{x^{-2}L(x)} = 1.$$

We also assume that lower-tail condition (1.3) holds for Z_t's. The TPDF of $\{Z_t\}$ is $\sigma_{\{Z_t\}}(0) = 1$ and $\sigma_{\{Z_t\}}(h) = 0$ for all $h \neq 0$. As $\alpha = 2$, the condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ guarantees that the infinite transformed-sum in (3.2) converges with probability 1, which we show in Section 3.4. As defined, $\{X_t\}$ is not strictly stationary because the noise sequence $\{Z_t\}$ is not required to be identically distributed. In fact so long as $\sigma_{\{Z_t\}}(h) = 0$ for all $h \neq 0$, $\{Z_t\}$ is not really required to be independent for the following results to hold; however, there seems little practical value in imagining time series constructed from such noise sequences.

3.3 Transformed Regularly-Varying MA(q) Process

X_t is a transformed regularly-varying moving average (MA) process of order q if for $\theta_j \in \mathbb{R}$, $\theta_q > 0$,

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \theta_2 \circ Z_{t-2} \oplus \cdots \oplus \theta_q \circ Z_{t-q}. \tag{3.3}$$

We require that $\theta_q > 0$ because if $\theta_q \leq 0$, the TPDF of the MA(q) will be the same as that of a lower order MA process with the same θ_j coefficients, $j < q$. To show that the transformed sum in (3.3) exists, we need to show that it takes a finite value with probability 1, i.e., $\lim_{x\to\infty} \text{Pr}(X_t > x) = 0$. We can easily extend (1.4) and (1.5) to the finite linear combination of $q+1$ $RV_+(2)$ independent variables to give that $X_t \in RV_+(2)$ and

$$\lim_{s\to\infty} s\text{Pr}\left(\frac{X_t}{b(s)} > c\right) = \lim_{s\to\infty} s\text{Pr}\left(\frac{(\bigoplus_{j=0}^{q} \theta_j \circ Z_{t-j})}{b(s)} > c\right) = c^{-2}\sum_{j=0}^{q}(\theta_j^{(0)})^2, \tag{3.4}$$

where $a^{(0)} = \max(a, 0)$ for $a \in \mathbb{R}$. Existence follows as (3.4) implies

$$\text{Pr}(X_t > x) \sim x^{-2}L(x)\sum_{j=0}^{q}(\theta_j^{(0)})^2 \to 0 \text{ as } x \to \infty. \tag{3.5}$$

Rearranging (3.5), the tail ratio is

$$\lim_{x\to\infty} \frac{\text{Pr}(X_t > x)}{x^{-2}L(x)} = \lim_{x\to\infty} \frac{\text{Pr}(X_t > x)}{\text{Pr}(Z_1 > x)} = \sum_{j=0}^{q}(\theta_j^{(0)})^2.$$

To get the TPDF, let $h > 0$ and let X_t and X_{t+h} be two elements of the series in (3.3), i.e.,

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \theta_2 \circ Z_{t-2} \oplus \cdots \oplus \theta_q \circ Z_{t-q},$$
$$X_{t+h} = Z_{t+h} \oplus \theta_1 \circ Z_{t+h-1} \oplus \theta_2 \circ Z_{t+h-2} \oplus \cdots \oplus \theta_q \circ Z_{t+h-q}.$$

Let $Z = (Z_{t+h}, Z_{t+h-1}, \cdots, Z_{t-q})^T$ be the $h+q$ dimensional vector of the regularly-varying noise terms. Let $\theta_t = (0, \cdots, 0, 1, \theta_1, \cdots, \theta_q)^T$ and $\theta_{t+h} = (1, \theta_1, \cdots, \theta_q, 0, \cdots, 0)^T$ be the coefficient vectors for X_t and X_{t+h}, respectively, such that, $X_t = \theta_t^T \circ Z$, and $X_{t+h} = \theta_{t+h}^T \circ Z$. By Corollary A2 and Proposition 2 of Cooley & Thibaud (2019), $(X_t, X_{t+h})^T$ has the joint angular measure

$$H_{(X_t,X_{t+h})}(\cdot) = \sum_{j=0}^{h+q} \|\theta_{\cdot j}^{(0)}\|_2^2 \delta_{\theta_{\cdot j}^{(0)}/\|\theta_{\cdot j}^{(0)}\|}(\cdot),$$

where $\theta_{\cdot j} = (\theta_{tj}, \theta_{(t+h)j})^T = (\theta_j, \theta_{j+h})^T$, $j = 0, \cdots, h+q$; $\theta_{\cdot j}^{(0)} = \max(\theta_{\cdot j}, 0)$ applied component-wise and δ is the Dirac mass function. By (2.2), the TPDF is then given by

$$\sigma(h) = \sum_{j=0}^{h+q} \left(\frac{\theta_{tj}^{(0)}}{\|\theta_{\cdot j}^{(0)}\|_2} \right) \left(\frac{\theta_{(t+h)j}^{(0)}}{\|\theta_{\cdot j}^{(0)}\|_2} \right) \|\theta_{\cdot j}^{(0)}\|_2^2 = \sum_{j=0}^{h+q} \theta_j^{(0)} \theta_{j+h}^{(0)} = \sum_{j=0}^{q} \theta_j^{(0)} \theta_{j+h}^{(0)}, \tag{3.6}$$

where $\theta_0 = 1$ and we drop the final h terms in the last summation since $\theta_j = 0$ for all $j > q$. As the TPDF depends only on lag, $\{X_t\}$ is stationary.

3.4 Transformed Regularly-Varying MA(∞) Process

Consider the transformed regularly-varying MA(∞) time series model

$$X_t = \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j}, \tag{3.7}$$

where $\psi_j \in \mathbb{R}$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. In order to draw on previous results for (standard) linear regularly-varying time series, we consider the time series of preimages $\{Y_t\}$ where

$$X_t = f\left\{ \sum_{j=0}^{\infty} \psi_j f^{-1}(Z_{t-j}) \right\} := f(Y_t).$$

Let $Y_t^{(q)}$ be the truncated series of the first q terms of Y_t. This truncated series $Y_t^{(q)}$ is $RV(2)$.

Consider the time series which is the difference in the infinite time series in (3.7) and the truncated MA(q) time series in (3.3), as

$$X_t^{(q)'} = \bigoplus_{j=q+1}^{\infty} \psi_j \circ Z_{t-j} = f\left\{ \sum_{j=q+1}^{\infty} \psi_j f^{-1}(Z_{t-j}) \right\} := f(Y_t^{(q)'}). \tag{3.8}$$

Lemma A3.26 in Embrechts et al. (1997) implies that, as $x \to \infty$,

$$\Pr(Y_t^{(q)'} > x) \sim x^{-2} L(x) \sum_{j=q+1}^{\infty} |\psi_j|^2 \implies \lim_{q\to\infty} \Pr(Y_t^{(q)'} > x) \sim \lim_{q\to\infty} x^{-2} L(x) \sum_{j=q+1}^{\infty} |\psi_j|^2 = 0.$$

By Lemma A2 of Cooley & Thibaud (2019), the limiting measure of $X_t^{(q)'}$ also tends to 0 as $q \to \infty$ and $X_t^{(q)'} \in RV_+(2)$. Rearranging, we get that the tail ratio of $X_t^{(q)'}$ tends to 0 as $q \to \infty$. We say that the MA(q) time series converges to the MA(∞) time series in tail ratio, as $q \to \infty$. Convergence in tail ratio is analogous to mean square convergence in the classical non-heavy-tail case. Consequently, the infinite series in (3.7) converges since,

$$\Pr(X_t > x) \sim x^{-2} L(x) \sum_{j=0}^{\infty} (\psi_j^{(0)})^2 \to 0, \text{ as } x \to \infty.$$

Also, taking limit as $q \to \infty$ in (3.6), the TPDF is

$$\sigma(h) = \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_{j+h}^{(0)} < \infty, \tag{3.9}$$

where $\psi_j^{(0)} = \max(\psi_j, 0)$ and $\psi_0 = 1$. Thus the infinite series in (3.7) is stationary.

3.5 Inner Product Space $\mathbb{V}$

Before we discuss AR and ARMA time series models, we broaden the discussion to better understand why we consider $\alpha = 2$ and use the L_2 norm. Consider the space

$$\mathbb{V} = \left\{ X_t : X_t = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j, \sum_{j=0}^{\infty} |\psi_j| < \infty \right\},$$

where Z_j's are independent and tail stationary $RV_+(2)$ random variables with $\lim_{x\to\infty} \Pr(Z_j > x)/\{x^{-2}L(x)\} = 1$ and $\psi_{t,j} \in \mathbb{R}$. It can be easily shown that $\mathbb{V}$ is a vector space (Section 4.1.1). For any $X_t \in \mathbb{V}$ we can define a mapping $T : \mathbb{V} \to \ell^1 = \left\{ \{a_j\}_{j=0}^{\infty}, a_j \in \mathbb{R} : \sum_{j=0}^{\infty} |a_j| < \infty \right\}$ such that $T(X_t) = \{\psi_{t,j}\}_{j=0}^{\infty} \in \ell^1$. As T is a linear map and an isomorphism (Section 4.1.3), $\mathbb{V}$ is isomorphic to ℓ^1.

Let X_t and X_s be two elements of vector space $\mathbb{V}$. We define the inner product between X_t and X_s as

$$\langle X_t, X_s \rangle := \sum_{j=0}^{\infty} \psi_{t,j}\psi_{s,j}. \tag{3.10}$$

We show (3.10) satisfies the properties of an inner product (Section 4.1.2). For $X_t \in \mathbb{V}$, the norm of X_t is defined as

$$\|X_t\| = \sqrt{\langle X_t, X_t \rangle} = \sqrt{\sum_{j=0}^{\infty} \psi_{t,j}^2},$$

which is finite as $\sum_{j=0}^{\infty} |\psi_{t,j}| < \infty$. Elements $X_t, X_s \in \mathbb{V}$ are said to be orthogonal if

$$\langle X_t, X_s \rangle = \sum_{j=0}^{\infty} \psi_{t,j}\psi_{s,j} = 0.$$

If $\{X_t\}$ is an MA(∞) time series, $X_t \in \mathbb{V}$ for all t. As $\{X_t\}$ is stationary, it is natural to think of the inner product as a function of lag:

$$\gamma(h) = \langle X_t, X_{t+h} \rangle = \sum_{j=0}^{\infty} \psi_j\psi_{j+h}.$$

Because we assume $\alpha = 2$ and use the L_2 norm, the TPDF $\sigma(h)$ is closely related to $\gamma(h)$. Clearly, $\gamma(h)$ is equivalent to $\sigma(h)$ if $\psi_j \geq 0$ for all j. In Chapter 4, we discuss prediction for stationary time series. Although $\mathbb{V}$ itself is not a Hilbert space since ℓ^1 is not complete in the metric induced by the ℓ^2 inner product, the set of predictors based on previous n observations is isomorphic to a closed linear subspace of ℓ^2 and we can employ the projection theorem.

3.6 Transformed Regularly-Varying Auto-Regressive Processes

Consider the stationary transformed regularly-varying autoregressive (AR) model of order 1 where X_t is defined as the stationary solution to the equation

$$X_t = \phi \circ X_{t-1} \oplus Z_t, \tag{3.11}$$

where $|\phi| < 1$. Iterating, we obtain

$$\begin{aligned} X_t = \phi \circ X_{t-1} \oplus Z_t &= \phi \circ (\phi \circ X_{t-2} \oplus Z_{t-1}) \oplus Z_t \\ &= \phi^2 \circ X_{t-2} \oplus \phi \circ Z_{t-1} \oplus Z_t \\ &= \cdots = \phi^k \circ X_{t-k} \bigoplus_{j=0}^{k-1} \phi^j \circ Z_{t-j}. \end{aligned}$$

If $|\phi| < 1$, the first term becomes small as k increases.

To show $X_t = \bigoplus_{j=0}^{\infty} \phi^j \circ Z_{t-j}$ is a solution, rearrange (3.11) and rewrite in terms of transform f to get

$$Z_t = X_t \oplus (-\phi) \circ X_{t-1} = f\left\{f^{-1}(X_t) - \phi f^{-1}(X_{t-1})\right\}. \tag{3.12}$$

Brockwell & Davis (1991) define the backward shift operator as $BX_t = X_{t-1}$. For our transformed-linear operations, we use the same notation to define the transformed backward shift operator as,

$$Bf^{-1}(X_t) = f^{-1}(X_{t-1}),$$

or equivalently,

$$f\{Bf^{-1}(X_t)\} = X_{t-1}.$$

Powers of the operator B are defined as $B^j f^{-1}(X_t) = f^{-1}(X_{t-j})$. Transformed-polynomials in B are isomorphic to polynomial functions of real variables and can be manipulated the same way.

Rewriting (3.12) in terms of the operator B, we get,

$$Z_t = f\left\{(1-\phi B)f^{-1}(X_t)\right\} = (1-\phi B)\circ X_t.$$

Defining function $\phi(B) = 1-\phi B$, $|\phi| < 1$, we get, $Z_t = \phi(B)\circ X_t$. We want to find the inverse $\pi(B)$ of $\phi(B)$ such that,

$$\pi(B)\circ Z_t = \{\pi(B)\phi(B)\}\circ X_t. \tag{3.13}$$

To find the inverse $\pi(B)$, consider $\phi(z) = 1-\phi z$. It can be shown through a Taylor series expansion that

$$\pi(z) = \sum_{j=0}^{\infty}\frac{z^j\pi^{(j)}(0)}{j!} = \sum_{j=0}^{\infty}\phi^j z^j,$$

the series being convergent only if $|\phi z| < 1$. In terms of the transformed backward shift operator B, $\pi(B) = \sum_{j=0}^{\infty}\phi^j B^j$. Applying $\pi(B) = \sum_{j=0}^{\infty}\phi^j B^j$ to $\{\pi(B)\phi(B)\}\circ X_t = \pi(B)\circ Z_t$, we get,

$$\left\{\sum_{j=0}^{\infty}\phi^j B^j(1-\phi B)\right\}\circ X_t = \left(\sum_{j=0}^{\infty}\phi^j B^j\right)\circ Z_t. \tag{3.14}$$

Let us first consider the left hand side of (3.14):

$$\begin{aligned}\left\{\sum_{j=0}^{\infty}\phi^j B^j(1-\phi B)\right\}\circ X_t &= f\left\{\sum_{j=0}^{\infty}\phi^j B^j(1-\phi B)f^{-1}(X_t)\right\}\\ &= f\left\{\sum_{j=0}^{\infty}\phi^j B^j f^{-1}(X_t) - \sum_{j=0}^{\infty}\phi^{j+1}B^{j+1}f^{-1}(X_t)\right\}\\ &= f\{f^{-1}(X_t)\} = X_t. \end{aligned}\tag{3.15}$$

Let us now consider the right hand side of (3.14):

$$\left(\sum_{j=0}^{\infty}\phi^j B^j\right)\circ Z_t = f\left\{\sum_{j=0}^{\infty}\phi^j B^j f^{-1}(Z_t)\right\} = f\left\{\sum_{j=0}^{\infty}\phi^j f^{-1}(Z_{t-j})\right\} = \bigoplus_{j=0}^{\infty}\phi^j\circ Z_{t-j}. \tag{3.16}$$

Putting together (3.15) and (3.16),

$$X_t = \bigoplus_{j=0}^{\infty} \phi^j \circ Z_{t-j}. \tag{3.17}$$

To show uniqueness, given any stationary solution Y_t of (3.11),

$$Y_t = \phi \circ Y_{t-1} \oplus Z_t = \cdots \text{iterating} \cdots = \phi^{k+1} \circ Y_{t-(k+1)} \oplus \bigoplus_{j=0}^{k} \phi^j \circ Z_{t-j}.$$

Rearranging, we get,

$$Y_t \oplus \bigoplus_{j=0}^{k} (-\phi^j) \circ Z_{t-j} = \phi^{k+1} \circ Y_{t-(k+1)}. \tag{3.18}$$

The right hand side of (3.18) is a regularly-varying random variable. The stationary time series $\{Y_t\}$ has a finite tail ratio. The tail ratio of $\phi^{k+1} \circ Y_{t-(k+1)} = (\phi^{2k+2}) \times$ tail ratio of Y_t which is finite. Taking limit as $k \to \infty$,

$$\lim_{k\to\infty} \left[\text{tail ratio of } \{Y_t \oplus \bigoplus_{j=0}^{k} (-\phi^j) \circ Z_{t-j}\} \right] = \lim_{k\to\infty} \left[\phi^{2k+2} \times \text{ tail ratio of } \{Y_t\}\right] = 0, \tag{3.19}$$

as $|\phi| < 1$. The limit in (3.19) is 0, indicating that Y_t is equal to the tail ratio limit $\bigoplus_{j=0}^{\infty} \phi^j \circ Z_{t-j}$ and that the process defined by (3.17) is the unique stationary solution of (3.11).

The TPDF for the AR(1) is then given by,

$$\sigma_{X_t, X_{t+h}} = \sigma(h) = \sum_{j=0}^{\infty} \phi^{j^{(0)}} \phi^{j+h^{(0)}},$$

and the tail ratio for the AR(1) is,

$$\sigma_{X_t, X_t} = \sigma(0) = \sum_{j=0}^{\infty} (\phi^{j^{(0)}})^2.$$

In the case $|\phi| > 1$, the series in (3.17) does not converge as $\sum_{j=0}^{\infty} |\phi^j|$ does not converge. However, we can write (3.11) for X_{t+1} as,

$$\begin{aligned}
&\phi \circ X_t = X_{t+1} \oplus (-Z_{t+1}) \\
&\Rightarrow X_t = \phi^{-1} \circ \{X_{t+1} \oplus (-Z_{t+1})\} = \phi^{-1} \circ X_{t+1} \oplus (-\phi^{-1}) \circ Z_{t+1} \\
&\qquad = \cdots = \phi^{-k-1} \circ X_{t+k+1} \oplus (-\phi^{-k-1}) \circ Z_{t+k+1} \oplus \cdots \oplus (-\phi^{-1}) \circ Z_{t+1},
\end{aligned}$$

which shows, by the same arguments as earlier that, $X_t = \bigoplus_{j=1}^{\infty} (-\phi^{-j}) \circ Z_{t+j}$, is the unique stationary solution of (3.11) as $\sum_{j=1}^{\infty} |\phi|^{-j} < \infty$. Similar to non-extreme time series notion of causality, we say that $\{X_t\}$ is causal if X_t can be expressed in terms of the current and past values, Z_s, $s \leq t$. Thus for $\{X_t\}$ defined as a solution to (3.11), $\{X_t\}$ is causal if $|\phi| < 1$, $\{X_t\}$ is non-causal if $|\phi| > 1$ and there is no stationary solution if $|\phi| = 1$.

3.7 Transformed Regularly-Varying ARMA(1,1) Process

Consider the transformed regularly-varying ARMA(1,1) model where X_t is defined as the stationary solution to the equation

$$X_t \oplus (-\phi) \circ X_{t-1} = Z_t \oplus \theta \circ Z_{t-1}, \tag{3.20}$$

where $\theta + \phi \neq 0$. If $\phi = 0$ or $\theta = 0$ then (3.20) reduces to a MA(1) and AR(1), respectively. Using the transformed backward shift operator B as defined in Section 3.6, (3.20) can be rewritten as,

$$\phi(B) \circ X_t = \theta(B) \circ Z_t, \tag{3.21}$$

where $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$. Following Brockwell & Davis (2002), we investigate the range of values of ϕ and θ for which a stationary solution of the ARMA(1,1) exists. If $|\phi| < 1$,

we can rewrite (3.21) as,

$$X_t = \left[\{\phi(B)^{-1}\}\theta(B)\right] \circ Z_t = \{\pi(B)\theta(B)\} \circ Z_t = \psi(B) \circ Z_t, \tag{3.22}$$

where $\pi(B) = \sum_{j=0}^{\infty} \phi^j B^j$. Expanding $\psi(B)$, we get,

$$\begin{aligned}\psi(B) = \pi(B)\theta(B) &= \left(\sum_{j=0}^{\infty} \phi^j B^j\right)(1+\theta B) = (1+\phi B+\phi^2 B^2+\phi^3 B^3+\cdots)(1+\theta B) \\ &= 1+(\phi+\theta)B+(\phi+\theta)\phi B^2+(\phi+\theta)\phi^2 B^3+\cdots = \sum_{j=0}^{\infty}\psi_j B^j, \end{aligned} \tag{3.23}$$

where $\psi_0 = 1$ and $\psi_j = (\phi+\theta)\phi^{j-1}$ for $j \geq 1$. Substituting (3.23) in equation (3.22), we get,

$$\begin{aligned} X_t &= \left(\sum_{j=0}^{\infty}\psi_j B^j\right) \circ Z_t \\ &= \{1+(\phi+\theta)B+(\phi+\theta)\phi B^2+(\phi+\theta)\phi^2 B^3+\cdots\} \circ Z_t \\ &= f[\{1+(\phi+\theta)B+(\phi+\theta)\phi B^2+(\phi+\theta)\phi^2 B^3+\cdots\} f^{-1}(Z_t)] \\ &= f\{f^{-1}(Z_t)+(\phi+\theta)Bf^{-1}(Z_t)+(\phi+\theta)\phi B^2 f^{-1}(Z_t)+(\phi+\theta)\phi^2 B^3 f^{-1}(Z_t)+\cdots\} \\ &= f\{f^{-1}(Z_t)+(\phi+\theta)f^{-1}(Z_{t-1})+(\phi+\theta)\phi f^{-1}(Z_{t-2})+(\phi+\theta)\phi^2 f^{-1}(Z_{t-3})+\cdots\} \\ &= Z_t \oplus (\phi+\theta) \circ \left(\bigoplus_{j=1}^{\infty}\phi^{j-1} \circ Z_{t-j}\right) \end{aligned} \tag{3.24}$$

We conclude that the MA(∞) process (3.24) is the unique stationary solution of (3.20), and is causal.

The TPDF for the ARMA(1,1) is complicated, and is developed from (3.24) in the Section 3.11. In contrast to the MA and AR(1) models where negative parameters θ_j or ϕ do not affect the TPDF because of the zero operation, negative parameter values do influence the TPDF of the transformed-linear ARMA(1,1) which does not match the ACVF of a non-extreme ARMA(1,1).

If $|\phi| > 1$, we can express $1/\phi(z)$ as a power series of z with absolutely summable coefficients by expanding in powers of z^{-1}, giving $1/\phi(z) = -\sum_{j=1}^{\infty}\phi^{-j}z^{-j}$. Applying the same argument

as in the case where $|\phi| < 1$, we can obtain the unique stationary solution of (3.20). Letting $\zeta(B) = -\sum_{j=1}^{\infty} \phi^{-j} B^{-j}$, and applying $\zeta(B)$ to both sides of (3.21), we get,

$$\begin{aligned}
X_t &= \left\{ \left(-\sum_{j=1}^{\infty} \phi^{-j} B^{-j} \right) (1 + \theta B) \right\} \circ Z_t \\
&= \left\{ (-\phi^{-1} B^{-1} - \phi^{-2} B^{-2} - \phi^{-3} B^{-3} - \cdots)(1 + \theta B) \right\} \circ Z_t \\
&= \left\{ -\phi^{-1}\theta - (\phi + \theta)\phi^{-2} B^{-1} - (\phi + \theta)\phi^{-3} B^{-2} - \cdots \right\} \circ Z_t \\
&= f \left[\left\{ -\phi^{-1}\theta - (\phi + \theta)\phi^{-2} B^{-1} - (\phi + \theta)\phi^{-3} B^{-2} - \cdots \right\} f^{-1}(Z_t) \right] \\
&= f \left\{ -\phi^{-1}\theta f^{-1}(Z_t) - (\phi + \theta)\phi^{-2} B^{-1} f^{-1}(Z_t) - (\phi + \theta)\phi^{-3} B^{-2} f^{-1}(Z_t) - \cdots \right\} \\
&= f \left\{ -\phi^{-1}\theta f^{-1}(Z_t) - (\phi + \theta)\phi^{-2} f^{-1}(Z_{t+1}) - (\phi + \theta)\phi^{-3} f^{-1}(Z_{t+2}) - \cdots \right\} \\
&= -\phi^{-1}\theta Z_t \oplus -(\phi + \theta) \circ \left(\bigoplus_{j=1}^{\infty} \phi^{-j-1} \circ Z_{t+j} \right). \qquad (3.25)
\end{aligned}$$

In this case the solution is noncausal. If $\phi = \pm 1$, there is no stationary solution of (3.20).

We can show that the ARMA(1,1) process in (3.20) is invertible, i.e., Z_t is expressible in terms of current and past values, X_s, $s \leq t$. Let $\xi(z) = 1/\theta(z) = \sum_{j=0}^{\infty} (-\theta)^j z^j$ with absolutely summable coefficients. Applying $\xi(B)$ to both sides of (3.21),

$$Z_t = \{\xi(B)\phi(B)\} \circ X_t = \pi(B) \circ X_t,$$

where

$$\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j = (1 - \theta B + (-\theta)^2 B^2 + \cdots)(1 - \phi B).$$

As earlier,

$$Z_t = X_t \oplus -(\phi + \theta) \circ \{ \bigoplus_{j=1}^{\infty} (-\theta)^{j-1} \circ X_{t-j} \}.$$

Following an argument like the one used to show noncausality when $|\phi| > 1$, the ARMA(1,1) process is noninvertible when $|\theta| > 1$, since then Z_t is expressed in terms of current and future

values, X_s, $s \geq t$, by

$$Z_t = (-\phi\theta^{-1}) \circ X_t \oplus (\phi + \theta) \circ \{\bigoplus_{j=1}^{\infty} (-\theta)^{-j-1} \circ X_{t+j}\}.$$

Analogous to the non-extreme time series case, if $\theta = \pm 1$, the ARMA(1,1) process is invertible in the more general sense that Z_t is a tail ratio limit of finite transformed-linear combinations of X_s, $s \leq t$, although it cannot be expressed explicitly as an infinite transformed-linear combination of X_s, $s \leq t$. A noncausal or noninvertible ARMA(1,1) process $\{X_t\}$ can be re-expressed as a causal and invertible ARMA(1,1) process relative to a new regularly-varying noise sequence $\{Z_t^*\}$. Thus, we can restrict attention to causal and invertible ARMA(1,1) models with $|\phi| < 1$ and $|\theta| < 1$. This is also valid for higher-order ARMA models.

3.8 Transformed Regularly-Varying ARMA(p, q) Process

Consider the stationary transformed regularly-varying ARMA(p, q) model where X_t is defined as the stationary solution to the equation

$$X_t \oplus (-\phi_1) \circ X_{t-1} \oplus \cdots \oplus (-\phi_p) \circ X_{t-p} = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \cdots \oplus \theta_q \circ Z_{t-q}, \tag{3.26}$$

where polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors. We can express (3.26) in terms of the transformed backward shift operator as

$$\phi(B) \circ X_t = \theta(B) \circ Z_t, \tag{3.27}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z$ and $\theta(z) = 1 + \theta_1 z - \cdots - \theta_q z$. $\{X_t\}$ is said to be a transformed regularly-varying autoregressive process of order p (or AR(p)) if $\theta(z) = 1$, and a transformed regularly-varying moving average process of order q (or MA(q)) if $\phi(z) = 1$.

In Section 3.7 we showed that a unique stationary solution exists for the ARMA$(1, 1)$ if and only if $\phi_1 \neq \pm 1$. Equivalently, the AR polynomial $\phi(z) = 1 - \phi(z) \neq 0$ for $z = \pm 1$. The

analogous condition for the general ARMA(p, q) process is $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z \neq 0$ for all complex z with $|z| = 1$. As discussed in Brockwell & Davis (2002, Section 3.1), z could be complex, since the zeros of a polynomial of degree $p > 1$ may be either real or complex.

If $\phi(z) \neq 0$ for $z = \pm 1$, then there exists $\delta > 0$ such that $\zeta(z) = \phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \zeta_j z^j$ for $1 - \delta < |z| < 1 + \delta$, and $\sum_{j=-\infty}^{\infty} |\zeta_j| < \infty$. We can then define $1/\phi(B)$ as the linear filter with absolutely summable coefficients, i.e., $1/\phi(B) = \sum_{j=-\infty}^{\infty} \zeta_j B^j$. Applying $\zeta(B)$ to both sides of (3.27), we get,

$$X_t = \{\zeta(B)\phi(B)\} \circ X_t = \{\zeta(B)\theta(B)\} \circ Z_t = \psi(B) \circ Z_t = \bigoplus_{j=-\infty}^{\infty} \psi_j \circ Z_{t-j}, \tag{3.28}$$

where $\psi(z) = \zeta(z)\theta(z) = \sum_{-\infty}^{\infty} \psi_j z^j$. By similar argument as in Section 3.7, we can show that $\psi(B) \circ Z_t$ is the unique stationary solution of (3.26).

In Section 3.7 we saw that the ARMA(1,1) process is causal if and only if $|\phi_1| < 1$. Equivalently, the ARMA(p, q) process is causal, i.e., X_t can be represented as $\bigoplus_{j=0}^{\infty} \psi_j Z_{t-j}$, with $\sum_{j=0}^{\infty} |\psi_j| < \infty$, if $\phi(z) \neq 0$ for $|z| \leq 1$, i.e., the roots of the AR polynomial should lie outside the unit circle. Similarly, we saw that the ARMA(1,1) process is invertible, i.e., Z_t can be represented as $\bigoplus_{j=0}^{\infty} \pi_j X_{t-j}$, with $\sum_{j=0}^{\infty} |\pi_j| < \infty$, if and only if $|\theta_1| < 1$. Equivalently, the ARMA(p, q) process is invertible if $\theta(z) \neq 0$ for $|z| \leq 1$, i.e., the roots of the MA polynomial should lie outside the unit circle.

As shown in Brockwell & Davis (2002, Section 3.1), $\{\psi_j\}$ can be determined by,

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = \theta_j, \quad j = 0, 1, ...,$$

where $\theta_0 = 1$, $\theta_j = 0$ for all $j > q$ and $\psi_j = 0$ for all $j < 0$. Similarly, $\{\pi_j\}$ can be determined by,

$$\pi_j + \sum_{k=1}^{q} \theta_k \pi_{j-k} = -\phi_j, \quad j = 0, 1, ...,$$

where $\phi_0 = -1$, $\phi_j = 0$ for all $j > p$ and $\pi_j = 0$ for all $j < 0$. By the causal representation of the ARMA(p, q), its TPDF is given by (3.9).

3.9 Application to Santa Ana Winds

3.9.1 Data and Preprocessing

Most regularly-varying time series models have been constructed with the aim of modelling heavy-tailed data. If $\{X_t^{(orig)}\}$ is a nonnegative regularly-varying time series with tail index α, letting $X_t = (X_t^{(orig)})^{\alpha/2}$ yields a time series with tail index 2, and a transformed linear model could be suitable for capturing tail dependence after this simple power transformation. Because regular variation provides a dependence model which focuses on tail behavior and allows for asymptotic dependence, we think our models can be widely employed to capture dependence in the upper tail. As mentioned in Chapter 1, the windspeed data appears to exhibit very strong tail dependence at short lags, and we feel that an asymptotically dependent model is well-suited to capture this dependence. However, the windspeed data are not heavy-tailed. To model this windspeed data, we will use our transformed-linear models to capture the tail dependence of this data after it has been marginally transformed to have regularly-varying tails with $\alpha = 2$. Separating the marginal distribution from the dependence structure is justified by Sklar's theorem (Sklar (1959), see also Resnick (1987, Proposition 5.15)), and such marginal transformations are relatively common in extremes (refer Smith et al. (1997) for an example in the time series case).

We return to the March AFB hourly windspeed data for the years 1973 - 2019, first introduced in Chapter 1. We choose to focus only on the autumn season (September 22 - December 22) as this is the period when fire risk due to the Santa Ana winds phenomenon is greatest. Our data set consists of 103,630 hourly observations.

Let $\{x_t^{(\text{orig})}\}$ be the hourly windspeed anomalies after removal of the diurnal cycle. Based on mean residual life plots (refer Coles (2001, Section 4.3)), we fit a generalized Pareto distribution (GPD) to the upper 2.5% of $x_t^{(\text{orig})}$. The windspeed anomalies appear to have a bounded tail (shape parameter estimate: $\hat{\xi}$ = -0.1 (se = 0.02)). Letting $\hat{\mu}$, $\hat{\psi}$ and $\hat{\xi}$ denote the empirical $.975$ quantile,

GPD scale and shape estimates, respectively, our estimated marginal distribution is

$$\hat{F}(x) = \begin{cases} (n+1)^{-1} \sum_{t=1}^{n} \mathbb{I}(x_t^{(\text{orig})} \leq x) & \text{for } x \leq \hat{\mu}, \\ 1 - 0.025\{1 + \hat{\xi}(x - \hat{\mu})/\hat{\psi}\}^{-1/\hat{\xi}} & \text{for } x > \hat{\mu}. \end{cases}$$

To obtain data which fits into our regular-variation modeling framework, we define $x_t = G^{-1}\{\hat{F}(x_t^{(\text{orig})})\}$, where $G(x) = \exp(-x^{-2})$. Thus, our transformed time series $\{x_t\}$ will have a marginal distribution which is approximately Fréchet with $\alpha = 2$ and $\sigma(0) = 1$.

3.9.2 Determination of Model TPDF's

Our proposed model fitting method is to find the parameters which minimize the squared difference between the empirical and model TPDF's. Due to preprocessing, $\sigma(0)$ is known to be 1. For lag $h > 0$, the model TPDFs with tail ratio 1 for the three models that we fit to the application data are as follows:

$$\text{AR(1): } \sigma(h) = \max(0, \phi^h),$$

$$\text{MA(1): } \sigma(h) = \begin{cases} \frac{\theta}{1+\theta^2} & \text{if } h = 1, \theta > 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{ARMA(1, 1): } \sigma(h) = \begin{cases} \frac{(\phi+\theta)\phi^h(1+\phi\theta)}{1+2\phi\theta+\theta^2} & \text{if } \phi > 0, \phi+\theta > 0 \\ 0 & \text{if } \phi > 0, \phi+\theta < 0 \\ \frac{(\phi+\theta)^2\phi^h}{1-\phi^4+(\phi+\theta)^2} & \text{if } \phi < 0, \phi+\theta > 0, h \text{ is even} \\ \frac{(\phi+\theta)\phi^{h-1}(1-\phi^4)}{1-\phi^4+(\phi+\theta)^2} & \text{if } \phi < 0, \phi+\theta > 0, h \text{ is odd} \\ \frac{(\phi+\theta)\phi^{h-1}(1+\theta\phi^3)}{1+\phi^2\theta^2+2\phi^3\theta} & \text{if } \phi < 0, \phi+\theta < 0, h \text{ is even} \\ 0 & \text{if } \phi < 0, \phi+\theta < 0, h \text{ is odd.} \end{cases}$$

3.9.3 Estimation of TPDF, Model Fitting, and Comparison

To estimate the TPDF, we use the estimator defined in Cooley & Thibaud (2019) in which the true angular measure is replaced by an empirical estimate. Assuming $(x_t, x_{t+h})^T, (t = 1, \cdots, n-h)$ be lag-h pairs of observations from a tail stationary time series $\{X_t\}$ and letting $r_t = \|(x_t, x_{t+h})^T\|_2$, and $w = (w_t, w_{t+h})^T = (x_t, x_{t+h})^T/r_t$, the TPDF estimator is defined as

$$\hat{\sigma}(h) = 2\int_{\Theta_1^+} w_t w_{t+h} d\hat{N}_{X_t,X_{t+h}}(w) = \frac{2}{\sum_{t=1}^{n} \mathbb{I}(r_t > r_0)} \sum_{t=1}^{n-h} w_t w_{t+h} \mathbb{I}(r_t > r_0), \tag{3.29}$$

where r_0 is some high threshold for the radial components, $H_{X_t,X_{t+h}}(\Theta_1^+) = 2$ because $\sigma(0) = 1$, and $N_{X_t,X_{t+h}}(\cdot) = 2^{-1} H_{X_t,X_{t+h}}(\cdot)$.

It is known that tail dependence estimates tend to have positive bias in the case of weak dependence (refer Huser et al. (2016)). Simulation study (refer Appendix A.1) shows that preprocessing the data to have $\sigma(0) = 1$, thus allowing to consider $H_{X_t,X_{t+h}}(\Theta_1^+) = 2$ instead of estimating $H_{X_t,X_{t+h}}(\Theta_1^+)$ gives better TPDF estimates. Simulation study also shows that subtracting off the mean of the time series considerably reduces bias in TPDF estimation. We subtract off the mean of the transformed time series $\{x_t\}$ and replace the negative observations by 0. We estimate the TPDF for the first 30 lags using (3.29). The upper left panel of Figure 3.1 gives the empirical TPDF for the data after bias correction.

Parameter estimates for the three transformed regularly-varying time series models are obtained using a preliminary method of numerical least squares optimization. Estimation will be further examined in Chapter 4. The parameter estimates and sum of squared differences (SS) between the empirical and model TPDF's are given in Table 3.1, and the ARMA(1,1) has the lowest SS. The fitted MA(1) model is not flexible enough to capture the dependence and $\hat{\theta} = 1$, rendering the fitted MA(1) non-invertible. We can also see from Figure 3.1 that compared to the other two models, the theoretical TPDF of the ARMA(1,1) model (lower right panel) seems to be a closer fit to the empirical TPDF (upper left panel).

Table 3.1: Fitted transformed-linear regularly-varying models

Model	Parameter estimates	SS
MA(1)	$\hat{\theta} = 1$	2.71
AR(1)	$\hat{\phi} = 0.9$	0.19
ARMA(1,1)	$\hat{\phi} = 0.93, \hat{\theta} = -0.51$	0.01

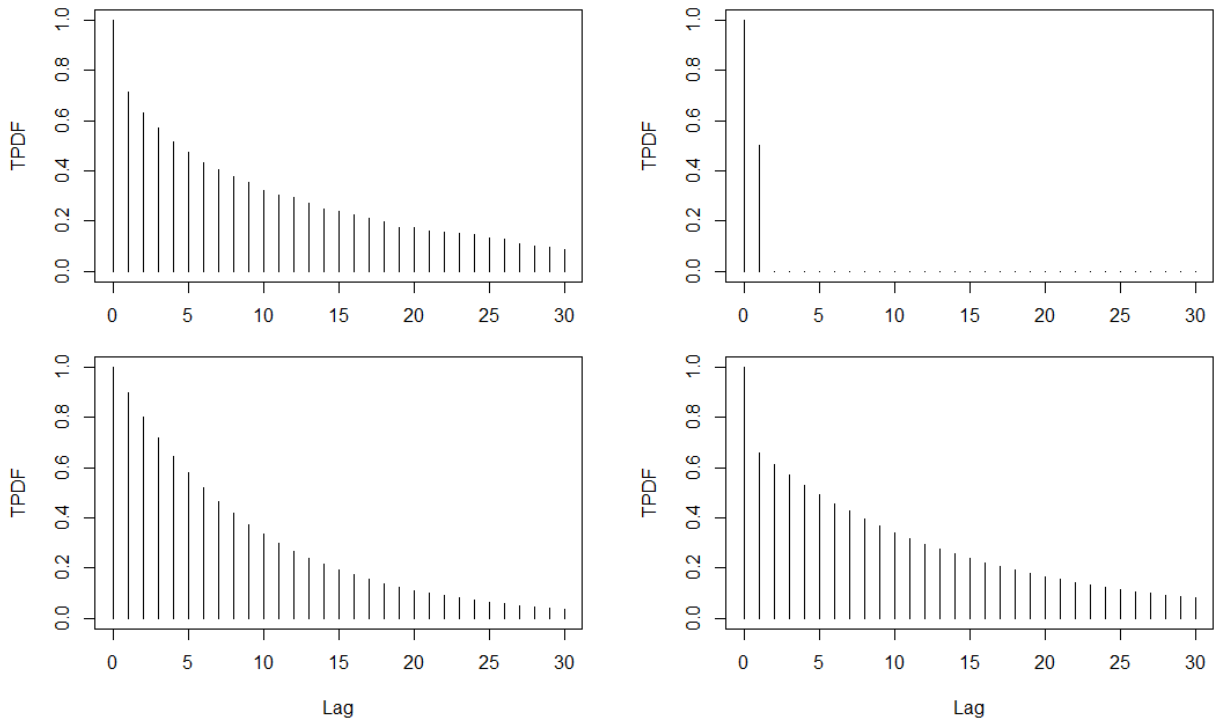

Figure 3.1: Comparison of estimated TPDF from data (upper left panel) and theoretical TPDF of fitted models: MA(1) (upper right panel), AR(1) (lower left panel) and ARMA(1,1) (lower right panel).

We generate a realization of the fitted transformed regularly-varying ARMA(1,1) time series model and transform it to the marginal of the observed hourly windspeed anomalies time series. Figure 3.2 compares the actual time series (upper panel) with the generated synthetic time series (lower panel) above a threshold of 1m/s. The two time series look quite similar to each other above this threshold.

For comparison, we fit three alternative ARMA(1,1) models, each with a two-step procedure. The first is a linear Gaussian time series model. We transform the marginal to be standard normal, estimate the ACVF, and then estimate the ARMA(1,1) parameters. The second is a linear regularly-varying time series model which follows the spirit of the models reviewed in Section 3.1 that take

values in $\mathbb{R}$. To be comparable, we perform a marginal transformation so that the data is unit Fréchet with $\alpha = 2$ in both directions. We then estimate the TPDF for this transformed data, which differs because the bivariate angular measure is not restricted to the positive orthant. For this model, the TPDF's summarized tail dependence is symmetric because it assesses both the upper and lower tail. Dependence in the lower tail of the windspeed anomalies is not as strong as in the upper tail ($\hat{\chi}(1) \approx 0.4$ for the negated anomaly time series). Consequently, the estimated TPDM ($\hat{\sigma}(1) = 0.49$) is not as strong as when focused only on the upper tail ($\hat{\sigma}(1) = 0.71$). The last model for comparison is a (standard) linear regularly-varying time series restricted to take values in $\mathbb{R}^+$ by imposing that the parameter estimates be positive. This model was fitted to the same transformed anomalies as our model, and shares the same estimated TPDF. As our transformed-linear model resulted in a negative $\hat{\theta}$, restricting the ARMA coefficients to be positive results in $\hat{\theta} = 0$ and we essentially get an AR(1) fit.

To explore model performance in capturing tail dependence, we calculate some tail summary statistics, and compare them across the above discussed models. Table 3.2 gives average length (and standard errors) of run above higher quantiles for the actual time series data, the synthetic ARMA(1,1) time series generated from our transformed-linear regularly-varying model, and synthetic ARMA(1,1) time series generated from the three alternative models. Table 3.3 gives the higher quantiles for sum of twelve consecutive time series terms (and bootstrapped standard errors). In both tables, the transformed linear model performs best, and seems to produce reasonable estimates of these tail quantities. Likely due to its asymptotic independence, or perhaps due to the fact that it was fit to the entire data set rather than focusing on extreme behavior, the non-extreme Gaussian model underestimates both quantities. The linear regularly-varying model on $\mathbb{R}$, with regular variation in both directions, also exhibits a lower dependence in the upper tail because of the symmetry in the definition of the TPDF. This provides evidence for restricting the model to the positive orthant when interest is only in the upper tail. The linear regularly-varying model on the positive orthant, with coefficients restricted to be positive, overestimates the dependence in the

upper tail as it does not allow for a negative MA coefficient. Hence, the transformed-linear model provides more flexibility than this restricted model.

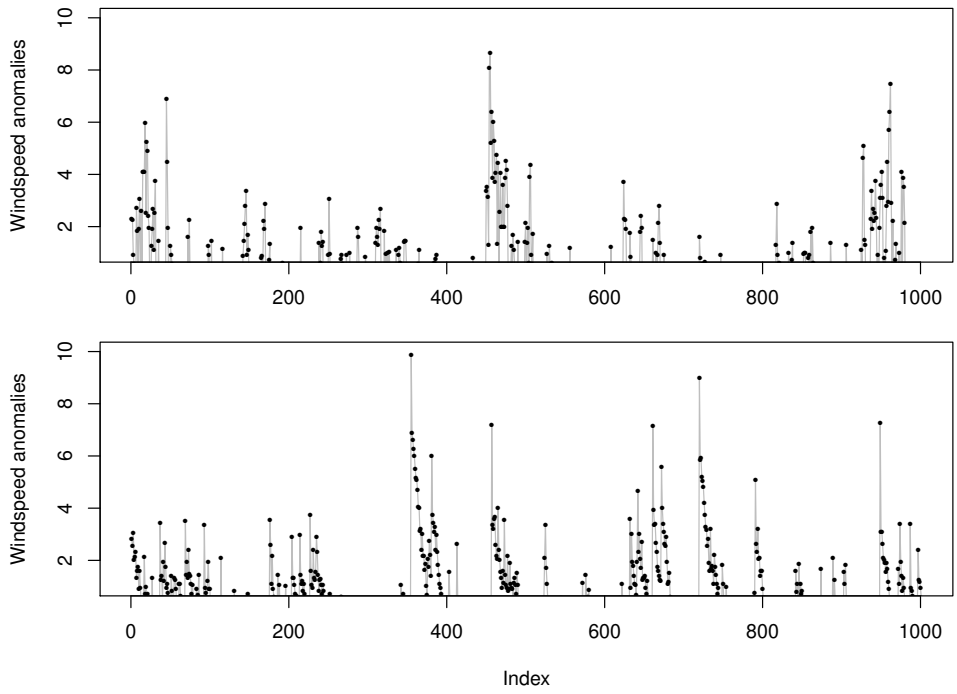

Figure 3.2: Comparison, above a threshold of 1 m/s, of actual windspeed anomalies time series (upper panel) and realization of synthetic time series generated from fitted transformed regularly-varying ARMA(1,1) model (lower panel), transformed to marginal of original time series.

Table 3.2: Average length (standard error) of run above a threshold

Threshold quantile	Actual	Trans-Lin Reg-Var in $\mathbb{R}^+$	Gaussian	Lin Reg-Var in $\mathbb{R}$	Lin Reg-Var in $\mathbb{R}^+$
0.95	2.43 (0.06)	2.60 (0.10)	1.48 (0.02)	1.56 (0.04)	5.41 (0.18)
0.98	2.35 (0.09)	2.60 (0.16)	1.34 (0.02)	1.52 (0.07)	5.74 (0.29)
0.99	2.10 (0.10)	2.66 (0.22)	1.27 (0.02)	1.59 (0.11)	5.93 (0.41)
0.995	1.77 (0.11)	2.61 (0.30)	1.21 (0.03)	1.75 (0.18)	5.83 (0.53)
0.999	1.40 (0.10)	2.21 (0.47)	1.08 (0.03)	2.04 (0.40)	5.78 (1.06)

Table 3.3: Quantiles for sum (standard error) of twelve consecutive terms

Quantile	Actual	Trans-Lin Reg-Var in $\mathbb{R}^+$	Gaussian	Lin Reg-Var in $\mathbb{R}$	Lin Reg-Var in $\mathbb{R}^+$
0.95	27.70 (0.72)	30.00 (0.62)	24.15 (0.40)	23.80 (0.41)	31.58 (0.65)
0.98	43.74 (1.19)	45.11 (1.10)	33.63 (0.60)	34.75 (0.89)	47.67 (1.22)
0.99	56.65 (1.66)	58.33 (1.80)	41.19 (0.91)	44.82 (1.60)	61.18 (1.73)
0.995	69.67 (2.11)	72.06 (2.63)	48.64 (0.91)	59.77 (3.17)	74.52 (2.47)
0.999	91.89 (3.60)	101.46 (5.15)	63.58 (2.50)	97.51 (5.63)	103.93 (5.04)

3.10 Discussion

This chapter constructs straightforward, flexible, and interpretable time series models by applying transformed-linear operations to regularly-varying random variables. Unlike other linear time series models, our time series models only take positive values, allowing one to focus modeling and inference entirely on the upper tail. As is common in time series, we characterize dependence between pairs of elements, and the TPDF has properties analogous to those of the ACVF. Our notion of weak tail stationarity helps relax the assumption of iid noise terms and more importantly allows characterization of the time series' upper tail via the TPDF. Our transformed-linear time series models have similar interpretations to and share some properties of their non-extreme ARMA analogues. Application to the Santa Ana hourly windspeed time series shows that our models appropriately capture extremal dependence. Fitting the MA(1), AR(1), and ARMA(1,1) transformed-linear time series models requires only minutes.

Our fitted model's run length estimate (Table 3.2) seems to exhibit "threshold stability", common to asymptotically dependent models. Recent work, mainly in spatial extremes, has aimed to develop models with more nuanced handling of tail dependence (e.g., Huser et al. (2018), Wadsworth & Tawn (2019), Bopp et al. (2020)). There are likely time series analyses where similar models would prove beneficial; but these models and their estimation procedures are quite complex. We believe that there is value in simple models.

A current challenge in extremes is tail dependence estimation, which was evident when estimating the TPDF, particularly in cases of weak tail dependence. We are confident that tail dependence estimation methods will continue to improve, aiding in TPDF estimation.

The fact that the transformed linear MA(∞) class of models can be linked to a Hilbert space opens avenues for further exploration. In future work we aim to investigate method-of-moments estimation procedures analogous to traditional time series methods such as Yule-Walker or the innovations algorithm. We would also like to investigate forecasting methods for transformed-ARMA models.

3.11 Derivation of the TPDF expression for a transformed-linear regularly-varying ARMA(1,1) time series model

The general form for the TPDF of a transformed-linear regularly-varying ARMA(1,1) time series is given as,

$$\sigma(h) = \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_{j+h}^{(0)},$$

where $\psi_0 = 1$, $\psi_j = (\phi + \theta)\phi^{j-1}$ for $j > 0$, $\psi_j^{(0)} = \max(\psi_j, 0)$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. We consider different cases of ϕ and θ below.

Case 1: $\phi > 0$ and $\phi + \theta > 0$, so that $\psi_0 = 1$ and $\psi_j > 0$ for all j.

For $h > 0$,

$$\begin{aligned}
\sigma(0) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} &= \psi_0 \psi_h + \sum_{j=1}^{\infty} (\phi+\theta)^2 \phi^{j-1} \phi^{j+h-1} = (\phi+\theta)\phi^{h-1} + (\phi+\theta)^2 \sum_{k=0}^{\infty} \phi^{2k+h} \\
&= (\phi+\theta)\phi^{h-1} + (\phi+\theta)^2 \phi^h \sum_{k=0}^{\infty} \phi^{2k} = (\phi+\theta)\phi^{h-1} + \frac{(\phi+\theta)^2 \phi^h}{1-\phi^2} \\
&= (\phi+\theta)\phi^h \left\{ \frac{1}{\theta} + \frac{(\phi+\theta)}{1-\phi^2} \right\} = \frac{(\phi+\theta)\phi^h(1+\phi\theta)}{1-\phi^2}. \qquad (3.30)
\end{aligned}$$

For $h = 0$,

$$\begin{aligned}
\sigma(h) = \sum_{j=0}^{\infty} \psi_j \psi_j &= \psi_0 \psi_0 + \sum_{j=1}^{\infty} (\phi+\theta)^2 \phi^{j-1} \phi^{j-1} = 1 + (\phi+\theta)^2 \sum_{k=0}^{\infty} \phi^{2k} \\
&= 1 + \frac{(\phi+\theta)^2}{1-\phi^2} = \frac{1 + 2\phi\theta + \theta^2}{1-\phi^2}. \qquad (3.31)
\end{aligned}$$

Case 2: $\phi > 0$ and $\phi + \theta < 0$, so that $\psi_0 = 1$ and $\psi_j < 0$ for all $j > 0$.

$$\sigma(h) = \begin{cases} 1, & \text{if } h = 0, \\ 0, & \text{if } h > 0. \end{cases} \tag{3.32}$$

Case 3: $\phi < 0$ and $\phi + \theta > 0$, so that $\psi_0 = 1$, $\psi_j < 0$ if j is even, and $\psi_j > 0$ if j is odd.

For $h > 0$,

$$\begin{aligned}
\sigma(h) &= \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_{j+h}^{(0)} = \psi_h^{(0)} + \psi_1 \psi_{h+1}^{(0)} + \psi_3 \psi_{h+3}^{(0)} + \psi_5 \psi_{h+5}^{(0)} + \cdots, \text{ as } \psi_j < 0 \text{ if } j \text{ is even,} \\
&= \begin{cases} 0 + \psi_1 \psi_{h+1} + \psi_3 \psi_{h+3} + \psi_5 \psi_{h+5} + \cdots, & \text{if } h \text{ is even,} \\ \psi_h, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} (\phi+\theta)\phi^0(\phi+\theta)\phi^h + (\phi+\theta)\phi^2(\phi+\theta)\phi^{h+2} + (\phi+\theta)\phi^4(\phi+\theta)\phi^{h+4} + \cdots, & \text{if } h \text{ is even,} \\ \psi_h, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} (\phi+\theta)^2\phi^h + (\phi+\theta)^2\phi^{h+4} + (\phi+\theta)^2\phi^{h+8} + \cdots, & \text{if } h \text{ is even,} \\ (\phi+\theta)\phi^{h-1}, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} (\phi+\theta)^2\phi^h \left(1 + \phi^4 + \phi^8 + \cdots\right), & \text{if } h \text{ is even,} \\ (\phi+\theta)\phi^{h-1}, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} \frac{(\phi+\theta)^2\phi^h}{1-\phi^4}, & \text{if } h \text{ is even,} \\ (\phi+\theta)\phi^{h-1}, & \text{if } h \text{ is odd.} \end{cases}
\end{aligned} \tag{3.33}$$

For $h = 0$,

$$\begin{aligned}
\sigma(h) &= \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_j^{(0)} = 1 + \psi_1^2 + \psi_3^2 + \psi_5^2 + \cdots \\
&= 1 + (\phi+\theta)^2 \left(1 + \phi^4 + \phi^8 + \cdots\right) = 1 + \frac{(\phi+\theta)^2}{1-\phi^4} = \frac{1 - \phi^4 + (\phi+\theta)^2}{1-\phi^4}
\end{aligned} \tag{3.34}$$

Case 4: $\phi < 0$ and $\phi + \theta < 0$, so that $\psi_0 = 1$, $\psi_j > 0$ if j is even, and $\psi_j < 0$ if j is odd.

For $h > 0$,

$$\begin{aligned}
\sigma(h) &= \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_{j+h}^{(0)} = \psi_h^{(0)} + \psi_2 \psi_{h+2}^{(0)} + \psi_4 \psi_{h+4}^{(0)} + \psi_6 \psi_{h+6}^{(0)} + \cdots, \text{ as } \psi_j < 0 \text{ if } j \text{ is odd,} \\
&= \begin{cases} \psi_h + \psi_2 \psi_{h+2} + \psi_4 \psi_{h+4} + \psi_6 \psi_{h+6} + \cdots, & \text{if } h \text{ is even,} \\ 0, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} (\phi+\theta)\phi^{h-1} + (\phi+\theta)\phi(\phi+\theta)\phi^{h+1} + (\phi+\theta)\phi^3(\phi+\theta)\phi^{h+3} + \cdots, & \text{if } h \text{ is even,} \\ 0, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} (\phi+\theta)\phi^{h-1} + (\phi+\theta)^2\phi^{h+2}\left(1 + \phi^4 + \phi^8 + \cdots\right), & \text{if } h \text{ is even,} \\ 0, & \text{if } h \text{ is odd,} \end{cases} \\
&= \begin{cases} (\phi+\theta)\phi^{h-1} + \frac{(\phi+\theta)^2\phi^{h+2}}{1-\phi^4}, & \text{if } h \text{ is even,} \\ 0, & \text{if } h \text{ is odd.} \end{cases}
\end{aligned} \tag{3.35}$$

For $h = 0$,

$$\begin{aligned}
\sigma(h) &= \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_j^{(0)} = 1 + \psi_2^2 + \psi_4^2 + \psi_6^2 + \cdots \\
&= 1 + (\phi+\theta)^2\phi^2\left(1 + \phi^4 + \phi^8 + \cdots\right) = 1 + \frac{(\phi+\theta)^2\phi^2}{1-\phi^4} = \frac{1 + \phi^2\theta^2 + 2\phi^3\theta}{1-\phi^4}
\end{aligned} \tag{3.36}$$

Putting together (3.30) to (3.36), for ARMA(1,1) with tail ratio 1, i.e. $\sigma(0) = 1$, the TPDF for lag $h > 0$ is given by,

$$\sigma(h) = \begin{cases} \frac{(\phi+\theta)\phi^h(1+\phi\theta)}{1+2\phi\theta+\theta^2} & \text{if } \phi > 0, \phi+\theta > 0 \\ 0 & \text{if } \phi > 0, \phi+\theta < 0 \\ \frac{(\phi+\theta)^2\phi^h}{1-\phi^4+(\phi+\theta)^2} & \text{if } \phi < 0, \phi+\theta > 0, h \text{ is even} \\ \frac{(\phi+\theta)\phi^{h-1}(1-\phi^4)}{1-\phi^4+(\phi+\theta)^2} & \text{if } \phi < 0, \phi+\theta > 0, h \text{ is odd} \\ \frac{(\phi+\theta)\phi^{h-1}(1+\theta\phi^3)}{1+\phi^2\theta^2+2\phi^3\theta} & \text{if } \phi < 0, \phi+\theta < 0, h \text{ is even} \\ 0 & \text{if } \phi < 0, \phi+\theta < 0, h \text{ is odd,} \end{cases}$$

Chapter 4

Transformed-Linear Innovations Algorithm for Modeling and Forecasting of Time Series Extremes

A primary aim of time series analysis is forecasting. In Chapter 2 and 3 we developed transformed-linear time series models, a class of time series which are nonnegative and regularly-varying but which are similar to familiar ARMA models in the non-extreme setting. We showed that these models can capture dependence in the upper tail of the time series. Now let us address the problem of forecasting.

Our approach for forecasting is to develop innovations algorithm for transformed-linear time series. The innovations algorithm, which relies on the autocovariance function, is a well known approach in classical time series for forecasting. In traditional time series analysis and elsewhere, the best linear predictor $\hat{X}_t$ is the one which minimizes mean squared prediction error (MSPE) given by $E[(X_{n+1} - \hat{X}_{n+1})^2]$ and Gaussian assumptions are usually used to create prediction intervals. However, expected-squared error is not a natural or intuitive measure of loss for extremes. Despite these differences, we show that the form of the transformed-linear predictor is of the same form as in the non-extreme setting.

The innovations algorithm gives us more than just a method for prediction. The innovations algorithm also has implications for modeling. Using the innovations algorithm we show that if the true model is in our transformed-linear space then applying the innovations algorithm iteratively, our parameter estimates converge to the true parameters. Furthermore we show that even if the underlying model is not a transformed-linear model, applying the innovations algorithm will yield a transformed-linear model whose TPDF matches closely the estimated TPDF of the underlying model.

We construct a vector space $\mathbb{V}$ of a series of absolutely summable transformed-linear combinations of nonnegative regularly-varying random variables. We show that $\mathbb{V}$ is an inner product space

and is isomorphic to ℓ^1, the space of absolutely summable sequences. Although $\mathbb{V}$ itself is not a Hilbert space, we show that the set of predictors based on previous n observations is isomorphic to a closed linear subspace of ℓ^2, the space of square summable sequences, and we can employ the projection theorem. Using the properties of the projection theorem we develop a transformed-linear analogue of the classical innovations algorithm that allows us to do modeling and prediction iteratively. We show that the class of transformed-linear regularly-varying MA(∞) time series is dense in the class of possible TPDFs. We also develop a transformed-linear analogue of the Wold decomposition. To demonstrate the richness of the class of transformed-linear regularly-varying MA(∞) models we run the innovations algorithm on data simulated from two different models. The first model is a GARCH(1,1) process and the second model is a first order Markov chain where each pair of consecutive observations has a bivariate logistic distribution. Neither of these models is in the family of transformed-linear time series. We show for both these models that by running the innovations algorithm on the estimated TPDF we can get estimates for coefficients of a transformed-linear regularly-varying MA time series whose TPDF closely matches the estimated TPDF of the simulated data.

Because the regular variation geometry differs from the elliptical geometry underlying standard linear prediction settings, uncertainty quantification is significantly different from the non-extreme setting. We extend the method proposed by Lee & Cooley (2021+) to develop prediction intervals to the time series setting. We perform modeling and prediction for the windspeed anomalies data discussed in Section 3.9 by applying the innovations algorithm to the estimated TPDF.

4.1 Inner Product Space $\mathbb{V}$

4.1.1 Vector Space $\mathbb{V}$

Consider the space $\mathbb{V} = \{X_t : X_t = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j, \sum_{j=0}^{\infty} |\psi_{t,j}| < \infty\}$ where Z_j's are independent and tail stationary $RV_+(2)$ random variables with $\lim_{x\to\infty} \Pr(Z_j > x)/\{x^{-2}L(x)\} = 1$ for some slowly-varying function $L(x)$, $\psi_{t,j} \in \mathbb{R}$, and $t \in \mathbb{Z}$. We first show that $\mathbb{V}$ is a vector space. Consider X_t, X_u, $X_v \in \mathbb{V}$, $a, b \in \mathbb{R}$ and $0_{\mathbb{V}} := \bigoplus_{j=0}^{\infty} 0 \circ Z_j$, where the subscript $\mathbb{V}$ in $0_{\mathbb{V}}$

is used to distinguish this element of our vector space from the number 0. Let $f : \mathbb{R} \to (0, \infty)$ defined as $f(y) = \log\{1 + \exp(y)\}$ be the transform defined in Cooley & Thibaud (2019).

Vector addition is closed: $X_t \oplus X_u \in \mathbb{V}$.

Proof:

$$\begin{aligned} X_t \oplus X_u &= \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j \oplus \bigoplus_{j=0}^{\infty} \psi_{u,j} \circ Z_j = f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\} \oplus f\left\{\sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\} \\ &= f\left(f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\}\right] + f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\}\right]\right) \\ &= f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\}, \qquad (4.1) \\ &= f\left\{\sum_{j=0}^{\infty} (\psi_{t,j} + \psi_{u,j}) f^{-1}(Z_j)\right\} = \bigoplus_{j=0}^{\infty} (\psi_{t,j} + \psi_{u,j}) \circ (Z_j) \qquad (4.2) \end{aligned}$$

which is in $\mathbb{V}$, since $\sum_{j=0}^{\infty} |\psi_{t,j} + \psi_{u,j}| \leq \sum_{j=0}^{\infty} |\psi_{t,j}| + \sum_{j=0}^{\infty} |\psi_{u,j}| < \infty$.

Vector addition is commutative: $X_t \oplus X_u = X_u \oplus X_t$.

Proof: By (4.1),

$$\begin{aligned} X_t \oplus X_u &= f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\} \\ &= f\left\{\sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\} = X_u \oplus X_t. \end{aligned}$$

Vector addition is associative: $X_t \oplus (X_u \oplus X_v) = (X_t \oplus X_u) \oplus X_v$.

Proof: By (4.2),

$$\begin{aligned}
X_t \oplus (X_u \oplus X_v) &= f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\} \oplus f\left\{\sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{v,j} f^{-1}(Z_j)\right\} \\
&= f\left(f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\}\right] + f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{v,j} f^{-1}(Z_j)\right\}\right]\right) \\
&= f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{v,j} f^{-1}(Z_j)\right\} \\
&= f\left(f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\}\right] + f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{v,j} f^{-1}(Z_j)\right\}\right]\right) \\
&= f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\} \oplus f\left\{\sum_{j=0}^{\infty} \psi_{v,j} f^{-1}(Z_j)\right\} \\
&= (X_t \oplus X_u) \oplus X_v.
\end{aligned}$$

Additive identity: $X_t \oplus 0_{\mathbb{V}} = X_t$.

Proof: By (4.2),

$$X_t \oplus 0_{\mathbb{V}} = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j \oplus \bigoplus_{j=0}^{\infty} 0 \circ Z_j = \bigoplus_{j=0}^{\infty} (\psi_{t,j} + 0) \circ Z_j = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j = X_t.$$

Additive inverse: $X_t \oplus -X_t = 0_{\mathbb{V}}$.

Proof:

$$X_t \oplus -X_t = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j \oplus \bigoplus_{j=0}^{\infty} -\psi_{t,j} \circ Z_j = \bigoplus_{j=0}^{\infty} (\psi_{t,j} + (-\psi_{t,j})) \circ Z_j = \bigoplus_{j=0}^{\infty} 0 \circ Z_j = 0_{\mathbb{V}}.$$

Scalar multiplication is closed: $a \circ X_t \in \mathbb{V}$.

Proof:

$$
\begin{aligned}
a \circ X_t = a \circ \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j &= f\left(a \cdot f^{-1} \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j\right) \\
&= f\left(a \cdot f^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\}\right]\right) = f\left\{a \sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\} \\
&= f\left\{\sum_{j=0}^{\infty} a\psi_{t,j} f^{-1}(Z_j)\right\} = \bigoplus_{j=0}^{\infty} (a\psi_{t,j}) \circ Z_j \in \mathbb{V}. \qquad (4.3)
\end{aligned}
$$

Scalar multiplication is distributive: $a \circ (X_t \oplus X_u) = a \circ X_t \oplus a \circ X_u$.

Proof: By (4.1),

$$
\begin{aligned}
a \circ (X_t \oplus X_u) &= a \circ f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\} \\
&= f\left(af^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\}\right]\right) \\
&= f\left\{a \sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j) + a \sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\} \\
&= f\left(af^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{t,j} f^{-1}(Z_j)\right\}\right] + af^{-1}\left[f\left\{\sum_{j=0}^{\infty} \psi_{u,j} f^{-1}(Z_j)\right\}\right]\right) \\
&= f\left\{af^{-1}\left(\bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j\right) + af^{-1}\left(\bigoplus_{j=0}^{\infty} \psi_{u,j} \circ Z_j\right)\right\} \\
&= a \circ X_t \oplus a \circ X_u.
\end{aligned}
$$

Scalar multiplication is associative: $a \circ (b \circ X_t) = (ab) \circ X_t$.

Proof: By (4.3)

$$
\begin{aligned}
a \circ (b \circ X_t) = a \circ \left\{\bigoplus_{j=0}^{\infty} (b\psi_{t,j}) \circ Z_j\right\} &= \left\{\bigoplus_{j=0}^{\infty} (ab\psi_{t,j}) \circ Z_j\right\} \\
&= (ab) \circ \left(\bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j\right) = (ab) \circ X_t. \qquad (4.4)
\end{aligned}
$$

Multiplicative identity: $1 \circ X_t = X_t$.

Proof: By (4.3),

$$1 \circ X_t = \bigoplus_{j=0}^{\infty} (1 \cdot \psi_{t,j}) \circ Z_j = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j = X_t.$$

4.1.2 Inner Product in $\mathbb{V}$

Let X_t and X_s be two elements of vector space $\mathbb{V}$ such that $X_t = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j$ and $X_s = \bigoplus_{j=0}^{\infty} \psi_{s,j} \circ Z_j$ where $\sum_{j=0}^{\infty} |\psi_{t,j}| < \infty$ and $\sum_{j=0}^{\infty} |\psi_{s,j}| < \infty$. We define the inner product between X_t and X_s as,

$$\langle X_t, X_s \rangle := \sum_{j=0}^{\infty} \psi_{t,j} \psi_{s,j}. \tag{4.5}$$

We can show that (4.5) is indeed an inner product by showing that it satisfies the properties of an inner product. Consider X_t, X_u, $X_v \in \mathbb{V}$ and $a \in \mathbb{R}$.

Linearity: $\langle X_t \oplus X_u, X_v \rangle = \langle X_t, X_v \rangle + \langle X_u, X_v \rangle$ and $\langle a \circ X_t, X_v \rangle = a \langle X_t, X_v \rangle$.

Proof: By (4.2),

$$\begin{aligned} \langle X_t \oplus X_u, X_v \rangle &= \left\langle \bigoplus_{j=0}^{\infty} (\psi_{t,j} + \psi_{u,j}) \circ Z_j, \bigoplus_{j=0}^{\infty} \psi_{v,j} \circ Z_j \right\rangle = \sum_{j=0}^{\infty} (\psi_{t,j} + \psi_{u,j}) \psi_{v,j} \\ &= \sum_{j=0}^{\infty} \psi_{t,j} \psi_{v,j} + \sum_{j=0}^{\infty} \psi_{u,j} \psi_{v,j} = \langle X_t, X_v \rangle + \langle X_u, X_v \rangle. \end{aligned}$$

Also,

$$\begin{aligned} \langle a \circ X_t, X_v \rangle &= \left\langle a \circ \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j, \bigoplus_{j=0}^{\infty} \psi_{v,j} \circ Z_j \right\rangle = \left\langle \bigoplus_{j=0}^{\infty} (a \psi_{t,j}) \circ Z_j, \bigoplus_{j=0}^{\infty} \psi_{v,j} \circ Z_j \right\rangle \\ &= \sum_{j=0}^{\infty} a \psi_{t,j} \psi_{v,j} = a \sum_{j=0}^{\infty} \psi_{t,j} \psi_{v,j} = a \langle X_t, X_v \rangle \end{aligned}$$

Symmetric property: $\langle X_t, X_u \rangle = \langle X_u, X_t \rangle$.

Proof:

$$\langle X_t, X_u \rangle = \sum_{j=0}^{\infty} \psi_{t,j}\psi_{u,j} = \sum_{j=0}^{\infty} \psi_{u,j}\psi_{t,j} = \langle X_u, X_t \rangle.$$

Positive definite property: $\langle X_t, X_t \rangle \geq 0$ and $\langle X_t, X_t \rangle = 0$ if and only if $X_t = 0_{\mathbb{V}} := \bigoplus_{j=0}^{\infty} 0 \circ Z_j$.

Proof:

$$\langle X_t, X_t \rangle = \sum_{j=0}^{\infty} \psi_{t,j}\psi_{t,j} = \sum_{j=0}^{\infty} \psi_{t,j}^2 \geq 0.$$

Let $\langle X_t, X_t \rangle = 0$, which means that $\sum_{j=0}^{\infty} \psi_{t,j}^2 = 0$. For $\psi_{t,j} \in \mathbb{R}$, this is possible only if $\psi_{t,j} = 0$ for all j, equivalently, when $X_t = 0_{\mathbb{V}}$. Conversely, if $X_t = 0_{\mathbb{V}}$, $\langle X_t, X_t \rangle = \sum_{j=0}^{\infty} 0^2 = 0$.

Thus, by satisfying all the above properties, the vector space $\mathbb{V}$ is an inner product space. For $X_t \in \mathbb{V}$, the norm of X_t is defined as $\|X_t\| = \sqrt{\langle X_t, X_t \rangle} = \sqrt{\sum_{j=0}^{\infty} \psi_{t,j}^2}$, which is finite as $\sum_{j=0}^{\infty} |\psi_{t,j}| < \infty$. Also, $X_t, X_s \in \mathbb{V}$ are said to be orthogonal if $\langle X_t, X_s \rangle = \sum_{j=0}^{\infty} \psi_{t,j}\psi_{s,j} = 0$.

As vector space $\mathbb{V}$ is an inner product space, it follows that $\mathbb{V}$ is a metric space. We can define a metric given by the squared norm of $X_t \ominus X_s$ as,

$$\langle X_t \ominus X_s,\ X_t \ominus X_s \rangle = \sum_{j=0}^{\infty} (\psi_{t,j} - \psi_{s,j})^2.$$

4.1.3 Isomorphism of $\mathbb{V}$ to ℓ^1

Consider the infinite dimensional space of absolutely summable sequences,

$$\ell^1 = \left\{ \{a_j\}_{j=0}^{\infty}, a_j \in \mathbb{R} : \sum_{j=0}^{\infty} |a_j| < \infty \right\}.$$

Vector addition and scalar multiplication in ℓ^1 are defined component-wise: $(a_0, a_1, a_2, ...) + (b_0, b_1, b_2, ...) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, ...)$ for $\{a_j\}, \{b_j\} \in \ell^1$ and the product of $(a_0, a_1, a_2, ...)$ with a scalar $\alpha \in \mathbb{R}$ is the sequence $(\alpha a_0, \alpha a_1, \alpha a_2, ...)$.

For any $X_t \in \mathbb{V}$ we can define a mapping $T : \mathbb{V} \to \ell^1$ such that $T(X_t) = \{\psi_{t,j}\}_{j=0}^{\infty} \in \ell^1$. We will first show that the mapping T is a linear map. Let X_t and X_s be two elements of vector space $\mathbb{V}$ such that $X_t = \bigoplus_{j=0}^{\infty} \psi_{t,j} \circ Z_j$ and $X_s = \bigoplus_{j=0}^{\infty} \psi_{s,j} \circ Z_j$ where $\sum_{j=0}^{\infty} |\psi_{t,j}| < \infty$ and $\sum_{j=0}^{\infty} |\psi_{s,j}| < \infty$. By (4.2) and (4.3), for any scalars $a, b \in \mathbb{R}$,

$$\begin{aligned} T(a \circ X_t \oplus b \circ X_s) &= T\left\{\bigoplus_{j=0}^{\infty}(a\psi_{t,j}) \circ Z_j \oplus \bigoplus_{j=0}^{\infty}(b\psi_{s,j}) \circ Z_j\right\} = T\left\{\bigoplus_{j=0}^{\infty}(a\psi_{t,j} + b\psi_{s,j}) \circ Z_j\right\} \\ &= \{a\psi_{t,j} + b\psi_{s,j}\}_{j=0}^{\infty} = \{a\psi_{t,j}\}_{j=0}^{\infty} + \{b\psi_{s,j}\}_{j=0}^{\infty} \\ &= a\{\psi_{t,j}\}_{j=0}^{\infty} + b\{\psi_{s,j}\}_{j=0}^{\infty} = aT(X_t) + bT(X_s). \end{aligned} \tag{4.6}$$

Recall that a linear map T is called an isomorphism if T is one-to-one and onto. We will show that the linear map T defined above is an isomorphism. Let X_t and X_s be as defined above. Then,

$$T(X_t) = T(X_s) \implies \{\psi_{t,j}\}_{j=0}^{\infty} = \{\psi_{s,j}\}_{j=0}^{\infty} \implies t = s \implies X_t = X_s. \tag{4.7}$$

Thus T is a one-to-one map. T is also onto since, for any sequence $\{a_j\}_{j=0}^{\infty}$ in ℓ^1 we can define $X = \bigoplus_{j=0}^{\infty} a_j \circ Z_j$ which is in $\mathbb{V}$ as $\sum_{j=0}^{\infty} |a_j| < \infty$. Thus, $\mathbb{V}$ is isomorphic to ℓ^1.

We know that vector space $\ell^1 \subset \ell^2$, where $\ell^2 = \{\{d_j\} : \sum_{j=0}^{\infty} d_j^2 < \infty\}$, the space of square-summable sequences. The inner product defined in (4.5) is isomorphic to the usual inner product on ℓ^2.

4.2 Transformed-Linear Innovations

4.2.1 Best Transformed-Linear Time Series Prediction

We want to use the projection theorem to perform prediction in our time series setting. To use the projection theorem we need to work with Hilbert spaces. However, $\mathbb{V}$ is isomorphic to ℓ^1, and hence is not a Hilbert space since ℓ^1 is not complete in the metric induced by the ℓ^2 inner product.

Consider performing prediction in terms of previous n steps. Our predictor will be a transformed-linear combination of previous n observations, that is,

$$\hat{X}_{n+1} = \bigoplus_{j=1}^{n} b_{nj} \circ X_{n+1-j}. \tag{4.8}$$

Let $\mathbb{V}_n$ be the set of all such predictors $\hat{X}_{n+1}$. Let us consider the analogous problem in ℓ^1. Consider sequences $\{a_1\}, \{a_2\}, \cdots, \{a_n\}$ in ℓ^1. The sequence corresponding to our predictors will be linear combinations of these sequences. Let $\mathbb{C}_n$ be the set of sequences c, corresponding to the predictors, where

$$c = b_{n1}\underline{a_1} + \cdots + b_{nn}\underline{a_n}.$$

$\mathbb{C}_n$ is the space spanned by $\underline{a_1}, \cdots, \underline{a_n}$, and dim($\mathbb{C}_n$) $\leq n$. Since any n-dimensional subspace of a complex topological vector space is closed (Rudin (1991, Theorem 1.21)), $\mathbb{C}_n$ is closed. Thus, $\mathbb{C}_n$ is a closed subspace of $\ell^1 \subset \ell^2$. By the projection theorem, there is a unique $\hat{c} \in \mathbb{C}_n$ such that $\|x - \hat{c}\| = \delta := \inf_{c \in \mathbb{C}_n} \|x - c\|$, for every x in ℓ^2. Thus, the set of predictors $\mathbb{V}_n$ based on previous n observations is isomorphic to a closed linear subspace of ℓ^2 and we can employ the projection theorem since ℓ^2 is known to be a Hilbert space.

Following Brockwell & Davis (1991, Chapter 5), we investigate the problem of predicting the value X_{n+1} in terms of $\{X_1, ..., X_n\}$. Again, let $\mathbb{V}_n$ be the closed transformed-linear subspace $\bar{\text{sp}}\{X_1, \cdots, X_n\}$, $n \geq 1$. By the projection theorem,

$$\hat{X}_{n+1} = \begin{cases} 0, & \text{if } n = 0, \\ P_{\mathbb{V}_n} X_{n+1} & \text{if } n \geq 1, \end{cases} \tag{4.9}$$

where $\hat{X}_{n+1}$ denotes the one-step predictor and $P_{\mathbb{V}_n}$ denotes the projection mapping onto $\mathbb{V}_n$. Thus, $\hat{X}_{n+1}$ is a transformed-linear combination of $\{X_1, ..., X_n\}$ as given in (4.8). The prediction

equations given by the projection theorem are

$$\left\langle X_{n+1} \ominus \bigoplus_{j=1}^{n} b_{nj} \circ X_{n+1-j},\ \ X_{n+1-k} \right\rangle = 0,\ \ k = 1, \cdots, n.$$

Equivalently,

$$\left\langle \bigoplus_{j=1}^{n} b_{nj} \circ X_{n+1-j},\ \ X_{n+1-k} \right\rangle = \langle X_{n+1},\ X_{n+1-k} \rangle,\ \ k = 1, \cdots, n. \tag{4.10}$$

By linearity of the inner product, the prediction equations can be rewritten in matrix form as

$$\Gamma_n \boldsymbol{b}_n = \boldsymbol{\gamma}_n \tag{4.11}$$

where $\Gamma_n = [\langle X_{n+1-j}, X_{n+1-k} \rangle]_{j,k=1}^{n}$, $\boldsymbol{b}_n = (b_{n1}, \cdots, b_{nn})^T$, and $\boldsymbol{\gamma}_n = [\langle X_{n+1}, X_{n+1-k} \rangle]_{k=1}^{n}$. If Γ_n is non-singular, then the solution is given as

$$\hat{\boldsymbol{b}}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n. \tag{4.12}$$

It can be shown that the above is equivalent to minimizing the squared norm $\|X_{n+1} \ominus \hat{X}_{n+1}\|^2$ by setting the appropriate derivative to zero. We see that (4.12) is of the familiar form for linear prediction in the non-extreme setting where the inner product terms are autocovariances.

4.2.2 Transformed-Linear Innovations

Following Brockwell & Davis (1991), we develop a transformed-linear analogue of the recursive innovations algorithm to obtain the one-step predictors $\hat{X}_{n+1}$, $n \geq 1$, defined in (4.9), without having to perform matrix inversion of Γ_n, defined in (4.11). Also, the innovations algorithm is a classical approach which allows one to iteratively perform better prediction in time series analysis.

Consider the transformed-linear innovation, $(X_{n+1} \ominus \hat{X}_{n+1})$, $n \geq 1$. Since $\mathbb{V}_n = \bar{\text{sp}}\{X_1, \cdots, X_n\}$, letting $\hat{X}_1 := 0$, $\mathbb{V}_n = \bar{\text{sp}}\{X_1 \ominus \hat{X}_1, \cdots, X_n \ominus \hat{X}_n\}$. We can rewrite the predictor in (4.8) in terms

of the innovations as,

$$\hat{X}_{n+1} = \bigoplus_{j=1}^{n} \theta_{nj} \circ \left(X_{n+1-j} \ominus \hat{X}_{n+1-j} \right).$$

By properties of projection mappings, $\hat{X}_{n+1} \in \mathbb{V}_n$ and by (4.10),

$$\langle X_{n+1} \ominus \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = 0.$$

That is, the transformed-linear innovation $(X_{n+1} \ominus \hat{X}_{n+1})$ is orthogonal to a transformed-linear combination of $X_1, ..., X_n$. Thus, the innovation is orthogonal to each of $X_1, ..., X_n$.

Consider the set of transformed-linear innovations, $\{X_{n+1-j} \ominus \hat{X}_{n+1-j}\}_{j=1,...,n}$. The innovation $(X_i \ominus \hat{X}_i) \in \mathbb{V}_{j-1}$ for $i < j$, as $(X_i \ominus \hat{X}_i)$ is a transformed-linear combination of $X_1, ..., X_i$. Also, by (4.10), $(X_j \ominus \hat{X}_j) \perp \mathbb{V}_{j-1}$. Thus, the set $\{X_1 \ominus \hat{X}_1, X_2 \ominus \hat{X}_2, ..., X_n \ominus \hat{X}_n\}$ is orthogonal. In fact, $\{X_{n+1-j} \ominus \hat{X}_{n+1-j}\}_{j=1,...,n}$ is an orthogonal basis of $\mathbb{V}_n$.

Let the squared distance of prediction be denoted by ν_n, that is, $\nu_n = \|X_{n+1} \ominus \hat{X}_{n+1}\|^2$. Following Brockwell & Davis (1991, Proposition 5.2.2), the innovations algorithm for a transformed-linear time series in $\mathbb{V}$ is given as follows:

Proposition 3 (The Transformed-Linear Innovations Algorithm)**.** *If $\{X_t\}$ is a transformed-linear time series in $\mathbb{V}$, where the matrix $\Gamma_n = [\langle X_i, X_j \rangle]_{i,j=1}^{n}$ is non-singular for each $n \geq 1$, then the one-step predictors $\hat{X}_{n+1}$, $n \geq 0$, and their squared distances of prediction ν_n, $n \geq 1$, are given by*

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{if } n = 0 \\ \bigoplus_{j=1}^{n} \theta_{nj} \circ (X_{n+1-j} \ominus \hat{X}_{n+1-j}) & \text{if } n \geq 1, \end{cases} \tag{4.13}$$

and

$$\begin{cases} \nu_0 & = \langle X_1, X_1 \rangle \\ \theta_{n,n-k} & = \nu_k^{-1}\Big(\langle X_{n+1}, X_{k+1} \rangle - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j\Big), \quad k = 0, 1, ..., n-1, \cdot \\ \nu_n & = \langle X_{n+1}, X_{n+1} \rangle - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \end{cases} \tag{4.14}$$

Proof: As stated above, the set $\{X_1 \ominus \hat{X}_1, X_2 \ominus \hat{X}_2, ..., X_n \ominus \hat{X}_n\}$ is orthogonal. Taking the inner product on both sides of (4.13) with $(X_{k+1} \ominus \hat{X}_{k+1})$, $0 \leq k < n$, we get

$$\begin{aligned} \Big\langle \hat{X}_{n+1}, (X_{k+1} \ominus \hat{X}_{k+1}) \Big\rangle &= \left\langle \left\{ \bigoplus_{j=1}^{n} \theta_{nj} \circ (X_{n+1-j} \ominus \hat{X}_{n+1-j}) \right\}, (X_{k+1} \ominus \hat{X}_{k+1}) \right\rangle \\ &= \sum_{j=1}^{n} \theta_{nj} \Big\langle (X_{n+1-j} \ominus \hat{X}_{n+1-j}), (X_{k+1} \ominus \hat{X}_{k+1}) \Big\rangle \\ &= \theta_{n,n-k}\nu_k, \end{aligned}$$

since $(X_{n+1-j} \ominus \hat{X}_{n+1-j}) \perp (X_{k+1} \ominus \hat{X}_{k+1})$ for all $j \neq n-k$.

Also, since $(X_{n+1} \ominus \hat{X}_{n+1}) \perp (X_{k+1} \ominus \hat{X}_{k+1})$ for $k = 0, \cdots, n-1$, we get,

$$\langle \hat{X}_{n+1}, (X_{k+1} \ominus \hat{X}_{k+1}) \rangle = \langle X_{n+1}, (X_{k+1} \ominus \hat{X}_{k+1}) \rangle.$$

Hence, the coefficients $\theta_{n,n-k}$, $k = 0, ..., n-1$ are given by

$$\theta_{n,n-k} = \nu_k^{-1} \langle X_{n+1}, (X_{k+1} \ominus \hat{X}_{k+1}) \rangle. \tag{4.15}$$

Using the representation in (4.13) with n replaced by k, we get

$$\theta_{n,n-k} = \nu_k^{-1}\Big(\langle X_{n+1}, X_{k+1} \rangle - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, (X_{j+1} \ominus \hat{X}_{j+1}) \rangle\Big). \tag{4.16}$$

Since by (4.15), $\langle X_{n+1}, (X_{j+1} \ominus \hat{X}_{j+1})\rangle = \nu_j \theta_{n,n-j}$, $0 \leq j < n$, we can rewrite (4.16) as

$$\theta_{n,n-k} = \nu_k^{-1}\Big(\langle X_{n+1}, X_{k+1}\rangle - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j\Big).$$

By properties of projection mapping, we have

$$\nu_n = \|X_{n+1} \ominus \hat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2 = \langle X_{n+1}, X_{n+1}\rangle - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 \nu_k.$$

□

4.3 Implications for Modeling of Stationary Time Series

We discussed prediction via the innovations algorithm in Section 4.2. However, the innovations algorithm also provides us with a method for understanding modeling and the richness of our class of models.

Naturally, we want to restrict attention to prediction of stationary time series. If $\{X_t\}$ is an MA(∞) time series, $X_t \in \mathbb{V}$ for all t. As $\{X_t\}$ is stationary, it is natural to think of the inner product as a function of lag:

$$\gamma(h) = \langle X_t, X_{t+h}\rangle = \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.$$

Being an inner product, $\gamma(.)$ is nonnegative definite and by the Cauchy-Schwarz inequality, $|\gamma(h)| \leq \gamma(0)$.

Because we assume $\alpha = 2$ and use the L_2 norm, the TPDF $\sigma(h)$ is closely related to $\gamma(h)$. Clearly, $\gamma(h)$ is equivalent to $\sigma(h)$ if $\psi_j \geq 0$ for all j.

The following corollary shows that given an invertible transformed-linear regularly-varying MA time series, running the transformed-linear innovations algorithm long enough gives us coefficient estimates that converge to the true coefficients of the MA.

Corollary 1. *If $\{X_t\}$ is an invertible MA process in $\mathbb{V}$, that is,*

$$Z_t = X_t \oplus \bigoplus_{j=1}^{\infty} \pi_j \circ X_{t-j},$$

with $\lim_{x\to\infty} Pr(Z_j > x)/\{x^{-2}L(x)\} = 1$*, then as* $n \to \infty$*,*

(i) $\nu_n \to 1$*,*

(ii) $\|X_n \ominus \hat{X}_n \ominus Z_n\|^2 \to 0$*, and*

(iii) $\theta_{nj} \to \psi_j, j = 1, 2, \cdots$*.*

Proof:

Let $\mathbb{M}_n = \bar{\text{sp}}\{X_s, -\infty < s \le n\}$ and $\mathbb{V}_n = \bar{\text{sp}}\{X_1, \cdots, X_n\}$. Because $\{X_t\}$ is invertible,

$$Z_{n+1} \ominus X_{n+1} = \bigoplus_{j=1}^{\infty} \pi_j \circ X_{n+1-j} = P_{\mathbb{M}_n}(Z_{n+1} \ominus X_{n+1}) = \ominus P_{\mathbb{M}_n} X_{n+1},$$

since $Z_{n+1} \perp \mathbb{M}_n$. Also, we can think of Z_k as $Z_k = \bigoplus_{j=0}^{\infty} \psi_j \circ Z_j$, where $\psi_j = 1$ for $j = k$ and $\psi_j = 0$ for all $j \neq k$. Thus, $Z_k \in \mathbb{V}$ and subsequently, $\|Z_k\|^2 = 1$ for all k. Then,

$$\begin{aligned}
1 = \|Z_{n+1}\|^2 = \|X_{n+1} \oplus \bigoplus_{j=1}^{\infty} \pi_j \circ X_{n+1-j}\|^2 &= \|X_{n+1} \ominus P_{\mathbb{M}_n} X_{n+1}\|^2 \\
&\le \|X_{n+1} \ominus P_{\mathbb{V}_n} X_{n+1}\|^2 = \nu_n \\
&\le \|X_{n+1} \oplus \bigoplus_{j=1}^{n} \pi_j \circ X_{n+1-j}\|^2 = \|Z_{n+1} \ominus \bigoplus_{j=n+1}^{\infty} \pi_j \circ X_{n+1-j}\|^2 \\
&= \|Z_{n+1}\|^2 + \|\bigoplus_{j=n+1}^{\infty} \pi_j \circ X_{n+1-j}\|^2 = 1 + \sum_{i,j=n}^{\infty} \pi_i \pi_j \langle X_{n+1-i}, X_{n+1-j}\rangle \\
&\le 1 + \left(\sum_{j=n+1}^{\infty} \pi_j\right)^2 \gamma(0).
\end{aligned}$$

Thus (i) is established since,

$$1 \le \nu_n \le 1 + \left(\sum_{j=n+1}^{\infty} \pi_j\right)^2 \gamma(0) \implies \nu_n \to 1 \text{ as } n \to \infty.$$

Consider,

$$\begin{aligned}
\|X_n \ominus \hat{X}_n \ominus Z_n\|^2 &= \|X_n \ominus \hat{X}_n\|^2 - 2\langle Z_n, X_n \ominus \hat{X}_n\rangle + \|Z_n\|^2 \\
&= v_{n-1} - 2\left[\langle Z_n, X_n\rangle - \langle Z_n, \hat{X}_n\rangle\right] + 1 \\
&= v_{n-1} - 2\left[\langle Z_n, \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{n-j}\rangle - \langle Z_n, \bigoplus_{j=1}^{n-1} b_{nj} \circ X_{n-j}\rangle\right] + 1 \\
&= v_{n-1} + 2[\|Z_n\|^2 - 0] + 1 \\
&= v_{n-1} - 1, \qquad (4.17)
\end{aligned}$$

where (4.17) converges to 0 as $n \to \infty$ by (i), thus proving (ii).

Since $X_{n+1} = \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{n+1-j}$, we have that

$$\psi_j = \langle X_{n+1}, Z_{n+1-j}\rangle.$$

Also, by (4.15),

$$\theta_{nj} = \nu_{n-j}^{-1}\langle X_{n+1}, (X_{n+1-j} \ominus \hat{X}_{n+1-j})\rangle.$$

Then,

$$\begin{aligned}
|\theta_{nj} - \psi_j| &= \left|\theta_{nj} - \langle X_{n+1}, (X_{n+1-j} \ominus \hat{X}_{n+1-j})\rangle + \langle X_{n+1}, (X_{n+1-j} \ominus \hat{X}_{n+1-j})\rangle - \psi_j\right| \\
&\leq \left|\theta_{nj} - \langle X_{n+1}, (X_{n+1-j} \ominus \hat{X}_{n+1-j})\rangle\right| + \left|\langle X_{n+1}, (X_{n+1-j} \ominus \hat{X}_{n+1-j})\rangle - \langle X_{n+1}, Z_{n+1-j}\rangle\right| \qquad (4.18) \\
&= |\theta_{nj} - \theta_{nj} v_{n-j}| + \left|\langle X_{n+1}, (X_{n+1-j} \ominus \hat{X}_{n+1-j} \ominus Z_{n+1-j})\rangle\right| \\
&\leq |\theta_{nj} - \theta_{nj} v_{n-j}| + \sqrt{\gamma(0)}\left\|(X_{n+1-j} \ominus \hat{X}_{n+1-j} \ominus Z_{n+1-j})\right\|, \qquad (4.19)
\end{aligned}$$

where the inequalities in (4.18) and (4.19) hold by the triangle inequality and the Cauchy-Schwarz inequality, respectively. Since θ_{nj} and $\gamma(0)$ are bounded, as $n \to \infty$, the first term on the right-hand

side of (4.19) converges to 0 by (i) and the second term on the right-hand side of (4.19) converges to 0 by (ii). Thus, $\theta_{nj} \to \psi_j$ as $n \to \infty$, proving (iii). □

We show later in Section 4.5 that the class of MA time series is a rich class. For now, one way to see that the MA class is rich is through Proposition 4. Before that, we need to prove the following Lemma.

Lemma 1. *If X_t is a tail stationary process in $\mathbb{V}$, then*

$$P_{\bar{sp}\{X_j, t-n \le j \le t-1\}} X_t \overset{\text{tail ratio}}{\to} P_{\bar{sp}\{X_j, -\infty < j \le t-1\}} X_t, \quad \text{as } n \to \infty.$$

Proof: Consider the transformed-linear combination

$$\bigoplus_{j=n+1}^{\infty} a_j \circ X_{t-j}.$$

As X_t is a tail stationary process,

$$P(X_t > x) \sim x^{-2} L(x) \sigma(0),$$

where $\sigma(0) = \sigma(X_t, X_t)$. As shown in Section 3.4,

$$P\left(\bigoplus_{j=n+1}^{\infty} a_j \circ X_{t-j} > x \right) \sim x^{-2} L(x) \sigma(0) \sum_{j=n+1}^{\infty} |a_j|^2, \quad \text{as } x \to \infty. \tag{4.20}$$

Taking limit on both sides of (4.20) we get, as $x \to \infty$,

$$\lim_{n \to \infty} P\left(\bigoplus_{j=n+1}^{\infty} a_j \circ X_{t-j} > x \right) \sim \lim_{n \to \infty} x^{-2} L(x) \sigma(0) \sum_{j=n+1}^{\infty} |a_j|^2 = 0,$$

$$\implies \lim_{x \to \infty} \frac{P\left(\bigoplus_{j=n+1}^{\infty} a_j \circ X_{t-j} > x \right)}{x^{-2} L(x)} \to 0, \quad \text{as } n \to \infty.$$

That is, tail ratio of $\bigoplus_{j=n+1}^{\infty} a_j \circ X_{t-j}$ converges to 0 as $n \to \infty$. Thus, by tail ratio convergence described in Section 3.4,

$$\bigoplus_{j=1}^{n} a_j \circ X_{t-j} \overset{\text{tail ratio}}{\to} \bigoplus_{j=1}^{\infty} a_j \circ X_{t-j},$$

$$P_{\bar{\text{sp}}\{X_j, t-n \le j \le t-1\}} X_t \overset{\text{tail ratio}}{\to} P_{\bar{\text{sp}}\{X_j, -\infty < j \le t-1\}} X_t, \quad \text{as } n \to \infty.$$

□

Analogous to Proposition 3.2.1 in Brockwell & Davis (1991), the following proposition shows that a q-tail-dependent tail stationary regularly-varying time series can be represented as a transformed-linear regularly-varying MA(q) process.

Proposition 4. *If $\{X_t\}$ is a regularly-varying tail stationary process in $\mathbb{V}$ with inner product function $\gamma(.)$ such that $\gamma(h) = 0$ for $|h| > q$ and $\gamma(q) \neq 0$, then $\{X_t\}$ is a transformed-linear regularly-varying MA(q) process, i.e. there exists a regularly-varying noise sequence $\{Z_t\}$ of independent and tail stationary Z_t's such that*

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \cdots \oplus \theta_q \circ Z_{t-q}.$$

Proof: For each t, let $\mathbb{M}_t$ be the closed transformed-linear subspace $\bar{\text{sp}}\{X_s, s \le t\}$ of $\mathbb{V}$ and set

$$Z_t = X_t \ominus P_{\mathbb{M}_{t-1}} X_t. \tag{4.21}$$

That is,

$$Z_t = X_t \ominus \bigoplus_{j=1}^{\infty} a_j \circ X_{t-j}, \quad a_j \in \mathbb{R}.$$

Thus, $Z_t \in \mathbb{M}_t$. By definition of $P_{\mathbb{M}_{t-1}}$, $P_{\mathbb{M}_{t-1}} X_t \in \mathbb{M}_{t-1}$ and $Z_t = X_t \ominus P_{\mathbb{M}_{t-1}} X_t \in \mathbb{M}_{t-1}^{\perp}$. Thus $Z_s \in \mathbb{M}_s \subset \mathbb{M}_{t-1}$ and hence $\langle Z_s, Z_t \rangle = 0$ for $s < t$. Also, by Lemma 1

$$P_{\bar{\text{sp}}\{X_s, t-n \le s \le t-1\}} X_t \overset{\text{tail ratio}}{\to} P_{\mathbb{M}_{t-1}} X_t, \quad \text{as } n \to \infty.$$

By stationarity and continuity of norm,

$$\begin{aligned}
\|Z_{t+1}\| &= \|X_{t+1} \ominus P_{\mathbb{M}_t} X_{t+1}\| \\
&= \lim_{n\to\infty} \|X_{t+1} \ominus P_{\bar{\text{sp}}\{X_s, s=t+1-n,\cdots,t\}} X_{t+1}\| \\
&= \lim_{n\to\infty} \|X_t \ominus P_{\bar{\text{sp}}\{X_s, s=t-n,\cdots,t-1\}} X_t\| \\
&= \|X_t \ominus P_{\mathbb{M}_{t-1}} X_t\| = \|Z_t\|.
\end{aligned}$$

Letting $c^2 = \|Z_t\|^2$, $\{Z_t\}$ is a sequence of independent and tail stationary regularly-varying random variables with scale c, that is, $\Pr(Z_t > x)/\{x^{-2}L(x)\} = c^2$.

By (4.21),

$$X_{t-1} = Z_{t-1} \oplus P_{\mathbb{M}_{t-2}} X_{t-1}.$$

Consequently,

$$\begin{aligned}
\mathbb{M}_{t-1} &= \bar{\text{sp}}\{X_s, s \le t-1\} \\
&= \bar{\text{sp}}\{X_s, s < t-1, Z_{t-1}\} \\
&= \bar{\text{sp}}\{X_s, s < t-q, Z_{t-q}, \cdots, Z_{t-1}\}.
\end{aligned}$$

Therefore, $\mathbb{M}_{t-1}$ can be decomposed into two orthogonal subspaces, $\mathbb{M}_{t-q-1}$ and $\bar{\text{sp}}\{Z_{t-q}, \cdots, Z_{t-1}\}$. Since $\gamma(h) = 0$ for $|h| > q$, it follows that $X_t \perp \mathbb{M}_{t-q-1}$ and since $\bar{\text{sp}}\{c^{-2} \circ Z_{t-q}, \cdots, c^{-2} \circ Z_{t-1}\}$

is an orthonormal set, by properties of projection mappings,

$$\begin{aligned} P_{\mathbb{M}_{t-1}} X_t &= P_{\mathbb{M}_{t-q-1}} X_t \oplus P_{\overline{\text{sp}}\{Z_{t-q},\cdots Z_{t-1}\}} X_t \\ &= 0 \ \oplus \ \left(c^{-2}\langle X_t, Z_{t-1}\rangle\right) \circ Z_{t-1} \ \oplus \cdots \oplus \ \left(c^{-2}\langle X_t, Z_{t-q}\rangle\right) \circ Z_{t-q} \\ &= \theta_1 \circ Z_{t-1} \oplus \cdots \oplus \theta_q \circ Z_{t-q}, \end{aligned} \tag{4.22}$$

where $\theta_j := c^{-2}\langle X_t, Z_{t-j}\rangle$, which by stationarity is independent of t for $j = 1, \cdots, q$. By (4.21) and (4.22),

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \cdots \oplus \theta_q \circ Z_{t-q}.$$

□

4.4 Modeling in Subset $\mathbb{V}_+$

So far we defined $\mathbb{V}$ allowing for negative ψ_j's since the negative coefficients give us the needed flexibility to define an inner product and employ the projection theorem. However, in our setting, a time series which has negative coefficients is indistinguishable, in terms of tail behavior, from a time series which has zeroes in place of those coefficients. Hence we restrict our attention to a subset $\mathbb{V}_+$ defined in the following section.

Consider a subset of $\mathbb{V}$ defined as

$$\mathbb{V}_+ = \{X_t : X_t = \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j}, \ \psi_j \geq 0, \sum_{j=0}^{\infty} \psi_j < \infty\}.$$

Proposition 5 below follows from the definition of TPDF.

Proposition 5. *If a transformed-linear MA(∞) time series in $\mathbb{V}$ has TPDF $\sigma(h)$, then there exists a transformed-linear MA(∞) time series in subset $\mathbb{V}_+$ which has the same TPDF $\sigma(h)$, for all lag h.*

Proof: Let $X_t = \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j} \in \mathbb{V}$. The TPDF of X_t is given by $\sigma(h) = \sum_{j=0}^{\infty} \psi_j^{(0)} \psi_{j+h}^{(0)}$, which is equal to the TPDF of $X_t^* = \bigoplus_{j=0}^{\infty} \psi_j^{(0)} \circ Z_{t-j} \in \mathbb{V}_+$. □

In other words, X_t and X_t^* are indistinguishable in terms of tail dependence. X_t and X_t^* are also indistinguishable in terms of tail ratio since recall that tail ratio is equal to $\sigma(0)$. Furthermore, the TPDF gives full information for a time series in $\mathbb{V}_+$ and unlike the inner product function, the TPDF is estimable. Also, it can be clearly seen that $\gamma(h) = \sigma(h)$, for $X_t^*, X_{t+h}^* \in \mathbb{V}_+$, for all lag h. Hence it is reasonable to restrict our attention to $\mathbb{V}_+$.

As the inner product is equivalent to the TPDF in $\mathbb{V}_+$, equation 4.12 can be rewritten as

$$\hat{\boldsymbol{b}}_n = \Sigma_n^{-1} \boldsymbol{\sigma}_n. \tag{4.23}$$

where $\Sigma_n = [\sigma(i-j)]_{i,j=1}^n$ and $\boldsymbol{\sigma}_n = [\sigma(i)]_{i=1}^n$.

Also, if our time series is in $\mathbb{V}_+$, we can rewrite the equations (4.14) of the innovations algorithm in terms of the TPDF $\sigma(\cdot)$ instead of the inner product as

$$\begin{cases} \nu_0 & = \sigma(0) \\ \theta_{n,n-k} & = \nu_k^{-1} \left(\sigma(n-k) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j \right), \quad k = 0, 1, ..., n-1, \cdot \\ \nu_n & = \sigma(0) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \end{cases} \tag{4.24}$$

Rewriting Corollary 1 for $\mathbb{V}_+$, we get the following corollary.

Corollary 2. *If $\{X_t\}$ is an invertible MA process in $\mathbb{V}_+$ with $\lim_{x\to\infty} Pr(Z_j > x)/\{x^{-2}L(x)\} = 1$, then as $n \to \infty$,*

(i) $\nu_n \to 1$,

(ii) $\|X_n \ominus \hat{X}_n \ominus Z_n\|^2 \to 0$, and

(iii) $\theta_{nj} \to \psi_j, j = 1, 2, \cdots; \quad \psi_j \geq 0$.

Also, rewriting Proposition 4 for $\mathbb{V}_+$,

Corollary 3. *If $\{X_t\}$ is a regularly-varying tail stationary process in $\mathbb{V}_+$ with tail pairwise dependence function $\sigma(.)$ such that $\sigma(h) = 0$ for $|h| > q$ and $\sigma(q) \neq 0$, then $\{X_t\}$ is an transformed-linear regularly-varying MA(q) process, i.e. there exists a regularly-varying noise sequence $\{Z_t\}$ of independent and tail stationary Z_t's such that*

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \cdots \oplus \theta_q \circ Z_{t-q}.$$

The relation between the TPDF and the inner product gives an important result described in the following remark:

Remark 1. *If X_t is an MA(∞) time series in $\mathbb{V}$ and $X_t \notin \mathbb{V}_+$, then by Proposition 5 there exists $X_t^* \in \mathbb{V}_+$, obtained by applying the zero-operator on the coefficients of X_t, which has the same TPDF as X_t. Thus, the innovations algorithm applied to the TPDF of $X_t \in \mathbb{V}$ will give us the one-step predictors for the corresponding $X_t^* \in \mathbb{V}_+$.*

4.5 Flexibility of the MA(∞) Class for Modeling

In this section we show that the class of MA(∞) models is a rich class for modeling.

4.5.1 Richness of the MA(∞) Class in terms of the TPDF

Given a valid TPDF (that is, a completely positive function) that converges to 0 as lag increases, we can run the innovations algorithm to get the θ_{nj}'s and ν_n as defined in (4.14). If we apply the TPDF formula to these θ_{nj}'s we will get a TPDF that gets arbitrarily close to the given TPDF. In other words, if we consider random noise terms $\{Z_j\}$ that are Frèchet with $\alpha = 2$ and scale $\sqrt{\nu_n}$, and generate a process applying the coefficients θ_{nj} to the Z's, the TPDF of this generated process will be arbitrarily close to the given TPDF. Thus, given any valid TPDF, we can run the transformed-linear innovations algorithm long enough to find a transformed-linear regularly-varying MA(∞) time series whose TPDF will get arbitrarily close to the given TPDF. As such, the class of MA(∞) time series is rich in the class of possible TPDFs that converge to 0.

To show this, first we prove the result for a q-tail-dependent TPDF in the following corollary.

Corollary 4. *If $\{X_t\}$ is any regularly-varying tail stationary process with TPDF $\sigma(.)$ such that $\sigma(h) = 0$ for $|h| > q$ and $\sigma(q) \neq 0$, then as $n \to \infty$, the θ_{nj}s generated from the transformed-linear innovations algorithm approach $\theta_1, \cdots, \theta_q$ of an MA(q) whose TPDF matches the given TPDF.*

Proof: Let us consider the form for the θ_{nj}s given by the transformed-linear innovations algorithm in 4.24:

$$\theta_{n,n-k} = \nu_k^{-1} \left(\sigma(n-k) - \sum_{l=0}^{k-1} \theta_{k,k-l}\theta_{n,n-l}\nu_l \right), \quad k = 0, 1, ..., n-1. \tag{4.25}$$

Rewriting (4.25) by letting $h = n - k$,

$$\begin{aligned} \theta_{n,h} &= \nu_{n-h}^{-1} \left(\sigma(h) - \sum_{l=0}^{n-h-1} \theta_{n-h,n-h-l}\theta_{n,n-l}\nu_l \right), \\ &= \nu_{n-h}^{-1} \left(\sigma(h) - \sum_{l=n-h-q}^{n-h-1} \theta_{n-h,n-h-l}\theta_{n,n-l}\nu_l \right), \quad h = 0, 1, ..., q, \end{aligned} \tag{4.26}$$

since $\theta_{n,n-l} = 0$ for all $l = 0, 1, \cdots, n-h-q-1$.

Rewriting (4.26) by letting $j = n - h - l$,

$$\theta_{n,h} = \nu_{n-h}^{-1} \left(\sigma(h) - \sum_{j=1}^{q} \theta_{n-h,j}\theta_{n,j+h}\nu_{n-h-j} \right). \tag{4.27}$$

As $n \to \infty$, let $\theta_{n,h} \to \theta_h$ and $\nu_n \to c^2$. Then (4.27) becomes,

$$\theta_h = c^{-2} \left(\sigma(h) - \sum_{j=1}^{q} \theta_j\theta_{j+h}c^2 \right). \tag{4.28}$$

Rearranging 4.28,

$$\begin{aligned}\sigma(h) &= \theta_h c^2 + \sum_{j=1}^{q} \theta_j \theta_{j+h} c^2 \\ &= c^2 \sum_{j=0}^{q} \theta_j \theta_{j+h}, \quad h = 0, 1, ..., q, \end{aligned} \tag{4.29}$$

which is the form for the TPDF at lag h for a regularly-varying tail stationary MA(q) with $\theta_j \geq 0$ for $j = 0, 1, \cdots, q$ and $\lim_{x\to\infty} \Pr(Z_j > x)/\{x^{-2}L(x)\} = c^2$. Thus, the TPDF of this MA(q) matches the given TPDF $\sigma(h)$. □

We are extending the above result to the MA(∞) case.

4.5.2 Transformed-Linear Wold Decomposition

If the TPDF of a time series does not converge to 0, analogous to the Wold decomposition discussed in Brockwell & Davis (1991) we can decompose the time series into an MA(∞) process and a deterministic process. Following Brockwell & Davis (1991) and Sargent (1979) we prove our Transformed-Linear Wold Decomposition as follows:

Theorem 1 (The Transformed-Linear Wold Decomposition)**.** *If* $c^2 = \|X_{n+1} \ominus \hat{X}_{n+1}\|^2 > 0$*, then* X_t *can be expressed as*

$$X_t = \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j} \ \oplus \ U_t, \tag{4.30}$$

where

(i) $\psi_0 = 1$ *and* $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,

(ii) $\{Z_t\}$ *is a sequence of independent and tail stationary regularly-varying random variables with scale* c,

(iii) $Z_t \in \mathbb{V}_t$ *for each* $t \in \mathbb{Z}$,

(iv) $\langle Z_t, U_s \rangle = 0$ *for all* $s, t \in \mathbb{Z}$,

(v) $\{U_t\}$ *is deterministic.*

The sequences $\{\psi_j\}$, $\{Z_j\}$, *and* $\{U_j\}$ *are uniquely determined by (4.30) and the conditions (i) - (v).*

Proof: Consider the sequences

$$Z_t = X_t \ominus P_{\mathbb{V}_{t-1}} X_t,$$

$$\psi_j = c^{-2} \langle X_t, Z_{t-j} \rangle, \tag{4.31}$$

$$U_t = X_t \ominus \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j}. \tag{4.32}$$

That is,

$$Z_t = X_t \ominus \bigoplus_{j=1}^{\infty} a_j \circ X_{t-j}, \quad a_j \in \mathbb{R},\ j = 1, \cdots, t-1.$$

Thus, $Z_t \in \mathbb{V}_t$, establishing (iii). By definition of $P_{\mathbb{V}_{t-1}}$, $P_{\mathbb{V}_{t-1}} X_t \in \mathbb{V}_{t-1}$ and $Z_t = X_t \ominus P_{\mathbb{V}_{t-1}} X_t \in \mathbb{V}_{t-1}^{\perp}$. Thus,

$$Z_t \in \mathbb{V}_{t-1}^{\perp} \subset \mathbb{V}_{t-2}^{\perp} \subset \cdots$$

Hence for $s < t$, $\langle Z_s, Z_t \rangle = 0$. By Lemma 1

$$P_{\bar{\text{sp}}\{X_s, t-n \le s \le t-1\}} X_t \overset{\text{tail ratio}}{\to} P_{\mathbb{V}_{t-1}} X_t, \quad \text{as } n \to \infty.$$

By stationarity and continuity of norm,

$$\begin{aligned}
\|Z_{t+1}\| &= \|X_{t+1} \ominus P_{\mathbb{V}_t} X_{t+1}\| \\
&= \lim_{n\to\infty} \|X_{t+1} \ominus P_{\bar{\text{sp}}\{X_s, s=t+1-n,\cdots,t\}} X_{t+1}\| \\
&= \lim_{n\to\infty} \|X_t \ominus P_{\bar{\text{sp}}\{X_s, s=t-n,\cdots,t-1\}} X_t\| \\
&= \|X_t \ominus P_{\mathbb{V}_{t-1}} X_t\| = \|Z_t\|.
\end{aligned}$$

Letting $c^2 = \|Z_t\|^2$, $\{Z_t\}$ is a sequence of independent and tail stationary regularly-varying random variables with scale c, thus establishing (ii).

By equation (4.22) in the proof of Proposition 4,

$$P_{\bar{\text{sp}}\{Z_j, j \le t\}} X_t = \sum_{j=0}^{\infty} \psi_j \circ Z_{t-j},$$

where $\psi_j = c^{-2}\langle X_t, Z_{t-j} \rangle$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. By stationarity, the coefficients ψ_j are independent of t. Also,

$$\psi_0 = c^{-2}\langle X_t, X_t \ominus P_{\mathbb{V}_{t-1}} X_t \rangle = c^{-2}\|X_t \ominus P_{\mathbb{V}_{t-1}} X_t\|^2 = c^{-2}\|Z_t\|^2 = 1,$$

thus proving (i). From equation (4.31) and (4.32), for $s \leq t$,

$$\begin{aligned}
\langle U_t, Z_s \rangle &= \left\langle X_t \ominus \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j},\ Z_s \right\rangle \\
&= \langle X_t, Z_s \rangle - \left\langle \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j},\ Z_s \right\rangle \\
&= \langle X_t, Z_s \rangle - \psi_{t-s} \langle Z_s, Z_s \rangle \\
&= \langle X_t, Z_s \rangle - \|Z_s\|^{-2} \langle X_t, Z_s \rangle \|Z_s\|^2 \\
&= 0.
\end{aligned}$$

In addition, if $s > t$, $Z_s \in \mathbb{V}_{s-1}^{\perp} \subset \mathbb{V}_t^{\perp}$. But $U_t \in \mathbb{V}_t$. Hence $\langle U_t, Z_s \rangle = 0$ for $s > t$. Thus (iv) is proved.

Since U_t is orthogonal to Z_t, $U_t \in \mathbb{V}_{t-1}$, that is U_t can be predicted perfectly from previous X's. To see this clearly, consider the projection of U_t on $\mathbb{V}_{t-1}$ to get

$$\begin{aligned}
P_{\mathbb{V}_{t-1}} U_t &= P_{\mathbb{V}_{t-1}} X_t \ominus P_{\mathbb{V}_{t-1}} \bigoplus_{j=0}^{\infty} \psi_j \circ Z_{t-j} \\
&= P_{\mathbb{V}_{t-1}} X_t \ominus \bigoplus_{j=1}^{\infty} \psi_j \circ Z_{t-j},
\end{aligned}$$

since $P_{\mathbb{V}_{t-1}} Z_t = 0$ and $P_{\mathbb{V}_{t-1}} Z_{t-k} = Z_{t-k}$ for $k \geq 1$. Transformed-linearly subtracting above equation from (4.32) gives

$$U_t \ominus P_{\mathbb{V}_{t-1}} U_t = \left(X_t \ominus P_{\mathbb{V}_{t-1}} X_t\right) \ \ominus \ \psi_0 \circ Z_t = 0_{\mathbb{V}},$$

since the one-step ahead prediction error for X_t is $\psi_0 \circ Z_t$. Hence, $U_t = P_{\mathbb{V}_{t-1}} U_t$. In general,

$$P_{\mathbb{V}_{t-k}} U_t = P_{\mathbb{V}_{t-k}} X_t \ominus \bigoplus_{j=k}^{\infty} \psi_j \circ Z_{t-j}.$$

Transformed-linearly subtracting above equation from (4.32) gives

$$U_t \ominus P_{\mathbb{V}_{t-k}} U_t = \left(X_t \ominus P_{\mathbb{V}_{t-k}} X_t\right) \ominus \bigoplus_{j=0}^{k-1} \psi_j \circ Z_{t-j} = 0_{\mathbb{V}},$$

since the k-step ahead prediction error for X_t is $\bigoplus_{j=0}^{k-1} \psi_j \circ Z_{t-j}$. Thus $\{U_t\}$ is deterministic as it can be predicted from past X's. □

4.5.3 Simulation Study

We conduct a simulation study that corroborates the richness of the class of transformed-linear regularly-varying MA(∞) models. We simulate data from two models. The first model is a GARCH(1,1) process (Bollerslev (1986)) with Gaussian noise terms and parameters $\alpha_0 = 0.2$, $\alpha_1 = 0.5$, and $\beta_1 = 0.3$. We consider the time series of absolute values of this GARCH process. A chi-plot (not shown) for the upper tail shows asymptotic dependence with $\hat{\chi}(1) \approx 0.34$. We apply our transform f to bound the data away from 0. Let this transformed data be denoted by x_{trans}. The Hill estimator (Hill (1975)) at the empirical 0.99 quantile of this transformed data gives an estimate $\hat{\alpha}_{trans} = 3.27$ of the tail index. The scale is estimated to be $\hat{c}_{trans} = 0.47$. We further transform the data into $x_t = \hat{c}_{trans}^{-1/2}(x_{trans})^{\hat{\alpha}_{trans}/2}$ so that our marginal now has $\alpha = 2$ and $\sigma(0) = 1$. As discussed in Section 3.9.3, preprocessing the data to have $\sigma(0) = 1$ allows us to reduce bias in TPDF estimation. Note that by doing this the noise terms Z_j are no longer such that $\sigma_{Z_j}(0) = 1$. As in Section 3.9.3, to reduce bias in TPDF estimation, we subtract off the mean of the transformed data and replace the negative observations with 0. We estimate the TPDF up to 500 lags using data whose radial components exceeds the 0.99 quantile. The squared distance of prediction ν_n converges to 0.65. Running the innovations algorithm on the estimated TPDF gives us converged θ estimates of an MA model. We consider the first 25 $\hat{\theta}$'s since the θ estimates are negligible beyond that. We then generate Fréchet noise terms with $\alpha = 2$ and scale $\sqrt{\nu_n} = \sqrt{0.65}$, and simulate a transformed-linear regularly-varying MA(25) time series using the estimated θ's from the innovations algorithm. We then back transform the simulated time series to the original

marginals. The average difference between the estimated TPDF from the original GARCH data and the estimated TPDF from our fitted model is -0.02 (se = 0.01).

The second model is a first order Markov chain such that each pair of consecutive observations has a bivariate logistic distribution with dependence parameter of 0.4 and common unit-Fréchet marginals (refer Smith et al. (1997)). A chi-plot (not shown) for the upper tail shows asymptotic dependence with $\hat{\chi}(1) \approx 0.7$. We perform the square-root transformation on the data so that $\alpha = 2$. Following the same process as for the first model, we simulate a transformed-linear regularly-varying MA(30) time series using the estimated θ's from the innovations algorithm and back transform the simulated time series to the original marginals. The average difference between the estimated TPDF from the original logistic data and the estimated TPDF from our fitted model is -0.0007 (se = 0.01).

Table 4.1 gives average length of run above higher quantiles for the original time series data and the fitted time series data (using coefficient estimates from the innovations algorithm) for the GARCH model and the logistic model. Table 4.2 gives the higher quantiles for sum of three consecutive time series terms. The fitted models seem to produce reasonable estimates of these tail quantities. Interestingly, for the fitted MA model in the logistic case (last column in Table 4.1), there is an increasing trend in the estimates and we do not see the "threshold stability" as exhibited by the fitted GARCH model and the earlier fitted models in Table 3.2.

Figure 4.1 gives the lag 1 plots for the true and fitted models for both the logistic and GARCH case. Clearly there is a difference in the dependence structure of the true model and our fitted model. This is not surprising since in the general time series case, we do not know the true model and we try to fit an infinite dimensional distribution to the time series.

Table 4.1: Average length (standard error) of run above a threshold for the simulation study

Threshold	GARCH		Logistic	
quantile	Original	Fitted	Original	Fitted
0.95	1.57 (0.02)	1.71 (0.03)	3.02 (0.09)	3.88 (0.11)
0.98	1.57 (0.03)	1.60 (0.04)	3.21 (0.15)	3.87 (0.16)
0.99	1.56 (0.05)	1.62 (0.06)	3.44 (0.22)	4.46 (0.26)
0.995	1.56 (0.06)	1.66 (0.09)	3.45 (0.29)	4.63 (0.35)
0.999	1.47 (0.11)	1.52 (0.15)	2.86 (0.48)	4.76 (0.76)

Table 4.2: Quantiles for sum (standard error) of twelve consecutive terms for the simulation study

Quantile	GARCH		Logistic	
	Original	Fitted	Original	Fitted
0.95	16.23 (0.20)	17.49 (0.16)	223.19 (8.29)	233.21 (8.26)
0.98	21.36 (0.31)	22.05 (0.34)	565.32 (37.79)	589.23 (37.77)
0.99	25.72 (0.68)	26.72 (0.58)	1252.96 (184.83)	1250.13 (118.99)
0.995	32.08 (1.24)	32.09 (1.03)	2789.53 (420.92)	2672.24 (338.49)
0.999	50.67 (3.67)	48.90 (2.98)	12833.18 (3736.45)	12732.06 (3705.63)

4.6 Prediction Error

We now return our attention to prediction and investigate the problem of assessing prediction uncertainty. Because the geometry of regular variation is very different from the elliptical geometry assumed in many non-extreme settings, we need to deal with uncertainty in prediction differently.

4.6.1 Completely Positive Decomposition of the Prediction TPDM

The squared distance of prediction, ν_n, is the analogue to mean square prediction error. In the finite-dimensional multivariate case Lee & Cooley (2021+) have shown that ν_n (K in Lee & Cooley (2021+)) is not useful to construct a prediction interval in the polar geometry of regular variation because the magnitude of error is dependent on the magnitude of the predicted value. We follow Lee & Cooley (2021+) and apply their method to construct prediction intervals when $\hat{X}_{n+1}$ is large.

The tail dependence between $\hat{X}_{n+1}$ and X_{n+1} can be characterized by the bivariate angular measure $H_{\hat{X}_{n+1},X_{n+1}}$. As shown in Lee & Cooley (2021+), the dependence of $H_{\hat{X}_{n+1},X_{n+1}}$ is summarized by the prediction TPDM

$$\Sigma_{\hat{X}_{n+1},X_{n+1}} = \begin{bmatrix} \boldsymbol{\sigma}_n^T \Sigma_n^{-1} \boldsymbol{\sigma}_n & \boldsymbol{\sigma}_n^T \Sigma_n^{-1} \boldsymbol{\sigma}_n \\ \boldsymbol{\sigma}_n^T \Sigma_n^{-1} \boldsymbol{\sigma}_n & \sigma(0) \end{bmatrix}, \tag{4.33}$$

where $\Sigma_n = [\sigma(i-j)]_{i,j=1}^n$ and $\boldsymbol{\sigma}_n = [\sigma(i)]_{i=1}^n$. Since $\Sigma_{\hat{X}_{n+1},X_{n+1}}$ is a 2×2 completely positive matrix, given any $q_* \geq 2$, there exist nonnegative matrices $B \in \mathbb{R}^{2\times q_*}$ such that $BB^T = \Sigma_{\hat{X}_{n+1},X_{n+1}}$.

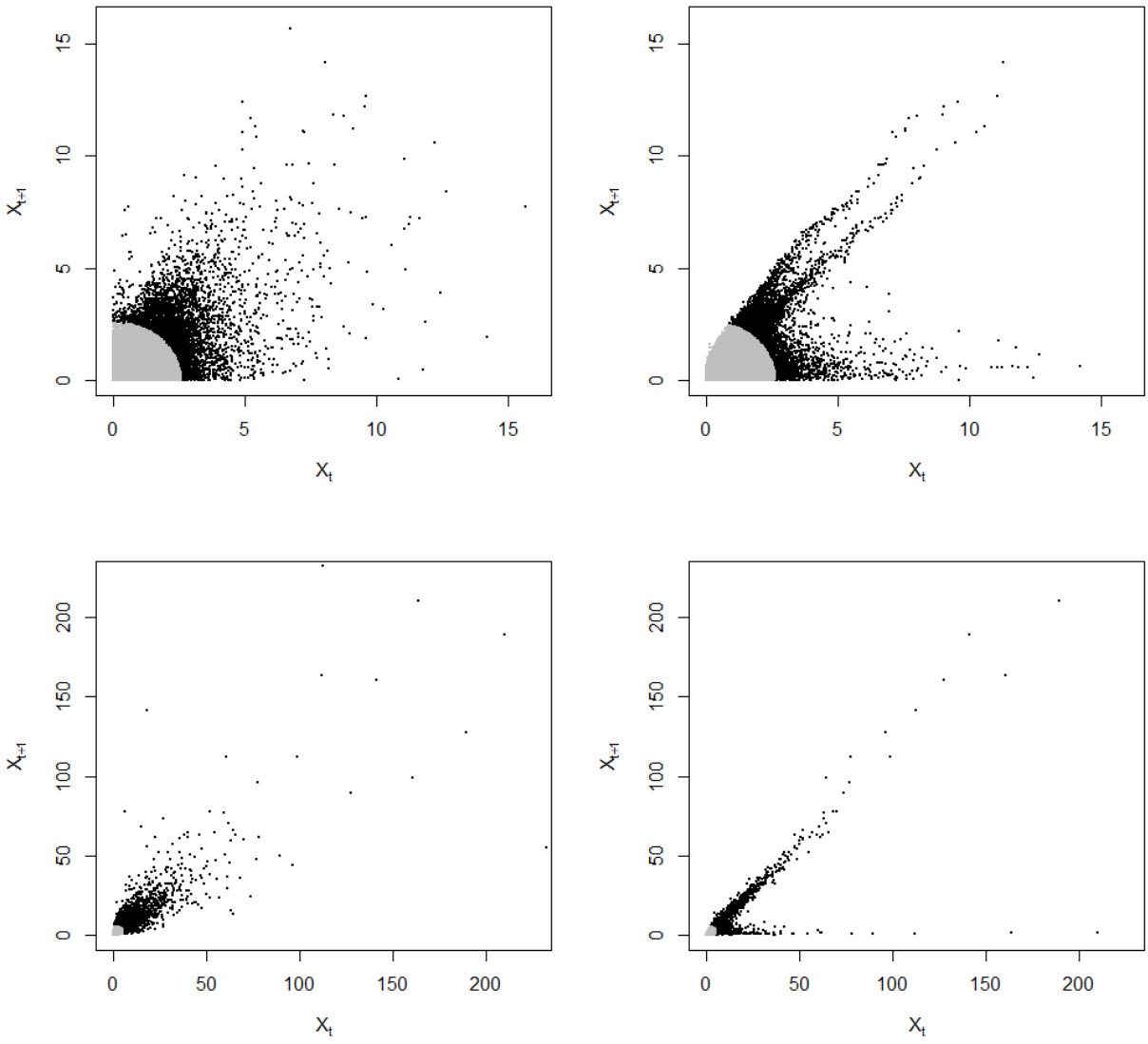

Figure 4.1: Lag-1 plots. Original data from the true GARCH model (top left panel), Simulated data from the fitted MA (back-transformed to the original marginals) on the GARCH data (top right panel), Original data from the true logistic model after square-root transformation (bottom left panel), Simulated data after square-root transformation from the fitted MA (back-transformed to the original marginals) on the logistic data (bottom right panel).

For feasible computation, Lee & Cooley (2021+) choose a moderate q_* and apply the algorithm in Groetzner & Dür (2020) repeatedly to get n_{decomp} nonnegative $B^{(k)}$ matrices, $k = 1, \cdots, n_{decomp}$, such that $B^{(k)}B^{(k)^T} = \Sigma_{\hat{X}_{n+1}, X_{n+1}}$ for all k. Then,

$$\hat{H}_{\hat{X}_{n+1}, X_{n+1}} = n_{decomp}^{-1} \sum_{k=1}^{n_{decomp}} \sum_{j=1}^{q_*} \|b_{kj}^{(0)}\|_2^2 \, \delta_{b_{kj}^{(0)}/\|b_{kj}^{(0)}\|_2}(\cdot),$$

where b_{kj} is the j^{th} column of the k^{th} matrix $B^{(k)}$ and δ is the Dirac mass function, and

$$\Sigma_{\hat{H}} = n_{decomp}^{-1} \sum_{k=1}^{n_{decomp}} B^{(k)}B^{(k)^T} = \Sigma_{\hat{X}_{n+1}, X_{n+1}}.$$

Defined this way, $\hat{H}_{\hat{X}_{n+1}, X_{n+1}}$ consists of $n_{decomp} \times q_*$ point masses.

As in Section 4.5.3, we simulate 100,000 random observations of a first order Markov chain model such that each pair of consecutive observations has a bivariate logistic distribution with dependence parameter of 0.4 and common unit-Fréchet marginals. We perform the square-root transformation on the data so that $\alpha = 2$. We consider the first 70,000 observations as training data and the remaining as test data. In Section 4.5.3 we fitted a transformed-linear regularly-varying MA(30) time series to data simulated from the logistic model because the converged innovations algorithm gave negligible θ estimates beyond θ_{30}. Hence we consider the problem of predicting any observation X_{n+1}, $n \geq 30$, based on the previous 30 observations. Let us denote this predicted observation as $\hat{X}_{n+1|n:n-29}$. Using equation (4.23) we obtain $\hat{\boldsymbol{b}}$ and the prediction TPDM $\Sigma_{\hat{X}_{n+1|n:n-29}, X_{n+1}}$ is obtained from equation (4.33). We apply the algorithm given in Groetzner & Dür (2020) repeatedly to compute 2×5 matrices $B^{(k)}$, $k = 1, \cdots, 100$, each of which is a completely positive decomposition of $\Sigma_{\hat{X}_{n+1|n:n-29}, X_{n+1}}$. Thus our estimated angular measure $\hat{H}_{\hat{X}_{n+1|n:n-29}, X_{n+1}}$ has 500 point masses. The 0.025 and 0.975 quantiles of $\hat{H}_{\hat{X}_{n+1|n:n-29}, X_{n+1}}$ give us a 95% joint region. The left panel of Figure 4.2 gives 95% joint region on the 30,000 test data. Thresholding at the 0.95 quantile of $\|\hat{X}_{n+1|n:n-29}, X_{n+1}\|$, 99.6% of the large data points fall within this joint region.

4.6.2 Conditional Prediction Intervals

The conditional density of $X_2|X_1 = x_1$ if x_1 is large is given in Lee & Cooley (2021+) as approximately

$$f_{X_2|X_1}(x_2|x_1) = 2c^{-1}\|(x_1, x_2)\|_2^{-5} x_2 h\left(\frac{(x_1, x_2)}{\|(x_1, x_2)\|_2}\right), \tag{4.34}$$

where $c = \int_0^\infty 2\|(x_1, x_2)\|_2^{-5} x_2 h\left(\frac{(x_1,x_2)}{\|(x_1,x_2)\|_2}\right) \mathrm{d}x_2$. We obtain an estimate of the conditional density of X_{n+1} given a large value of $\hat{X}_{n+1}$ using equation (4.34). The angular density h is estimated through a kernel density estimate of $\hat{H}_{\hat{X}_{n+1}, X_{n+1}}$. The 0.025 and 0.975 quantiles of the estimated conditional density in equation (4.34) give us a 95% conditional prediction interval. The right panel of Figure 4.2 gives the scatterplot after thresholding the test data at the 0.95 quantile of $\hat{X}_{n+1|n:n-29}$, along with the 95% conditional prediction intervals. These prediction intervals have a coverage rate of 0.975.

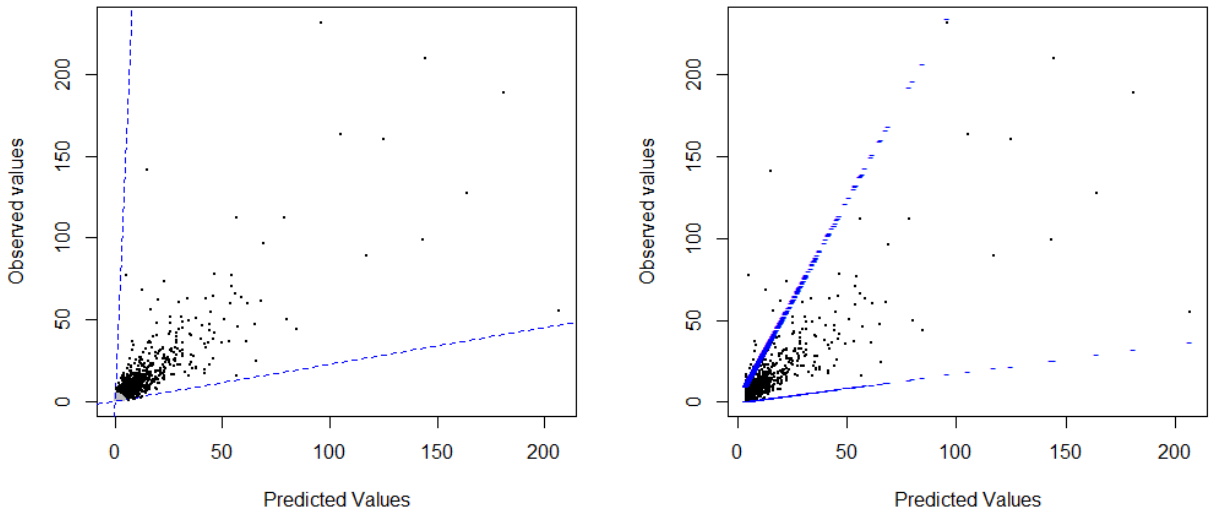

Figure 4.2: Scatterplot of the Logistic model test data, on the transformed Fréchet scale with $\alpha = 2$, with the estimated 95% joint prediction region (left panel). 95% conditional prediction intervals given each large value of $\hat{X}_{n+1|n:n-29}$ of the Logistic model test data, on the transformed Fréchet scale with $\alpha = 2$ (right panel).

4.7 Application to Santa Ana Winds

We return to the March AFB hourly windspeed data that we fitted a transformed-linear regularly-varying ARMA(1,1) time series to in Section 3.9. Recall that we marginally transformed the data to have regularly-varying tails with $\alpha = 2$ and $\sigma(0) = 1$. After subtracting off the mean of the transformed data and replacing the negative observations with 0, we estimate the TPDF up to 500 lags using data whose radial components exceeds the 0.99 quantile. Running the innovations algorithm on the estimated TPDF gives us converged θ estimates of an MA model. The squared distance of prediction, ν_n, converges to We consider the first 40 $\hat{\theta}$'s since the θ estimates are negligible beyond that. We then generate Fréchet noise terms, with $\alpha = 2$ and scale $\sqrt{\nu_n} = \sqrt{0.65}$, and simulate a transformed-linear regularly-varying MA(40) time series using the estimated θ's from the innovations algorithm and back transform the simulated time series to the original marginals. The average difference between the estimated TPDF from the original windspeed anomalies data and the estimated TPDF from our fitted model is -0.02 (se = 0.01). Table 4.3 gives average length of run above higher quantiles for the original windspeed anomalies time series data and the fitted time series data (using coefficient estimates from the innovations algorithm). Table 4.4 gives the higher quantiles for sum of three consecutive time series terms. The fitted models seem to produce reasonable estimates of these tail quantities.

Table 4.3: Average length (standard error) of run above a threshold for the windspeed data

Threshold quantile	Original	Fitted
0.95	2.43 (0.06)	2.32 (0.07)
0.98	2.35 (0.09)	2.27 (0.11)
0.99	2.10 (0.10)	2.46 (0.18)
0.995	1.77 (0.11)	2.35 (0.23)
0.999	1.40 (0.10)	2.04 (0.33)

We now perform prediction of the windspeed anomalies time series. Out of the 103,630 observations, we consider the first 70,000 observations as training data and the remaining as test data. We consider the problem of predicting an observation X_{n+1}, $n \geq 40$, based on the previous 40 observations. Let us denote this predicted observation as $\hat{X}_{n+1|n:n-39}$. Using equation (4.23)

Table 4.4: Quantiles for sum (standard error) of twelve consecutive terms for the windspeed data

Quantile	Original	Fitted
0.95	27.70 (0.72)	28.44 (0.52)
0.98	43.74 (1.19)	42.93 (1.07)
0.99	56.65 (1.66)	54.82 (1.62)
0.995	69.67 (2.11)	68.91 (2.50)
0.999	91.89 (3.60)	97.93 (4.49)

we obtain $\hat{\boldsymbol{b}}$ and the prediction TPDM $\Sigma_{\hat{X}_{n+1|n:n-39},X_{n+1}}$ is obtained from equation (4.33). We apply the algorithm given in Groetzner & Dür (2020) repeatedly to compute 2×5 matrices $B^{(k)}$, $k = 1, \cdots, 100$, each of which is a completely positive decomposition of $\Sigma_{\hat{X}_{n+1|n:n-39},X_{n+1}}$. Thus our estimated angular measure $\hat{H}_{\hat{X}_{n+1|n:n-39},X_{n+1}}$ has 500 point masses. The 0.025 and 0.975 quantiles of $\hat{H}_{\hat{X}_{n+1|n:n-39},X_{n+1}}$ give us a 95% joint region. The left panel of Figure 4.3 gives 95% joint region on the test data. Thresholding at the 0.95 quantile of $\|\hat{X}_{n+1|n:n-39}, X_{n+1}\|$, 98.99% of the large data points fall within this joint region.

We obtain an estimate of the conditional density of X_{n+1} given a large value of $\hat{X}_{n+1}$ using equation (4.34). The 0.025 and 0.975 quantiles of the estimated conditional density in equation (4.34) give us a 95% conditional prediction interval. The center panel of Figure 4.3 gives the scatterplot after thresholding at the 0.95 quantile of $\hat{X}_{n+1|n:n-39}$ of the test data along with the 95% conditional prediction bounds. These prediction intervals have a coverage rate of 0.96. The right panel of Figure 4.3 gives the prediction intervals on the original scale of the anomalies obtained by taking the inverse of the marginal transformation. We compare our prediction intervals to the standard Gaussian method. We transform the marginal of the original windspeed anomalies data to be standard normal and estimate the ACVF. We use the estimated covariance matrix to find the best linear unbiased predictor and to estimate the MSPE. We then create 95% Gaussian prediction intervals from the estimated MSPE and get a coverage rate of 0.94. For the windspeed anomalies data, our prediction intervals do not show significant advantage over the standard Gaussian based prediction intervals because our data is not too heavy-tailed, resulting into a negligible difference between the corresponding predicted weight vectors $\hat{\boldsymbol{b}}$. We are investigating a heavy-tailed pre-

cipitation data set where we suspect the difference between the corresponding predicted weight vectors will be more significant.

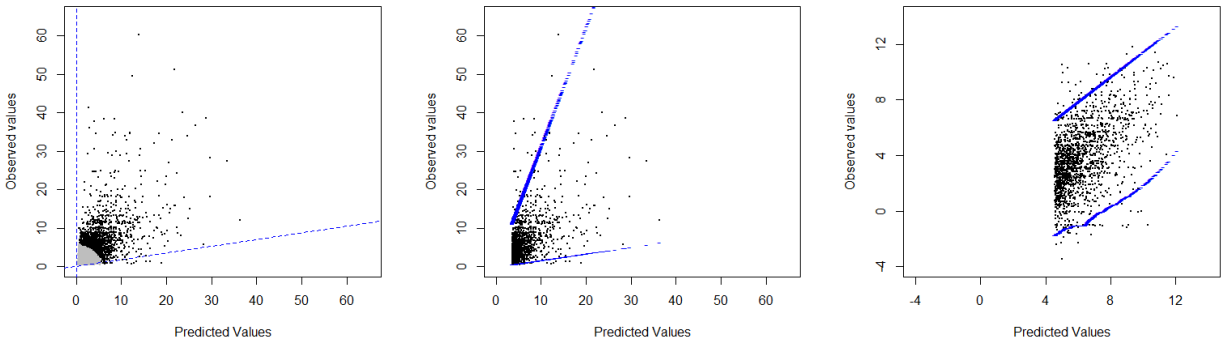

Figure 4.3: Scatterplot of the windspeed anomalies test data on the Fréchet scale with the estimated 95% joint prediction region (left panel). 95% conditional prediction intervals given each large value of $\hat{X}_{n+1|n:n-39}$ of the windspeed anomalies test data on the Fréchet scale (center panel). 95% conditional prediction intervals given each large value of $\hat{X}_{n+1|n:n-39}$ of the windspeed anomalies test data on the original scale (right panel).

4.8 Summary

In this chapter we address the goal of performing prediction. Unlike the classical setting, we are specifically interested in prediction for time series extremes, that is, when the previous time series terms indicate that the following observation will be large. We work in the transformed-linear regular variation framework presented in Chapter 2 and Chapter 3.

We show that the vector space $\mathbb{V}$ of a series of absolutely summable transformed-linear combinations of nonnegative regularly-varying random variables is an inner product space. We show that the set of predictors based on previous n observations of our time series is isomorphic to a closed linear subspace of a Hilbert space which enables us to use the projection theorem to construct methods for prediction. Using the properties of the projection theorem, we develop the transformed-linear analogue of the classical innovations algorithm that allows us to do prediction iteratively. Furthermore, the transformed-linear innovations algorithm also provides us a tool for modeling our time series iteratively. Restricting our attention to a subset $\mathbb{V}_+$ with nonnegative

coefficients links the inner product to the TPDF and allows us to use the innovations algorithm on the estimable TPDF. We show that the class of transformed-linear regularly-varying MA(∞) models is rich in the sense that, through the innovations algorithm, we can fit a transformed-linear regularly-varying MA(∞) to any valid TPDF. Using the polar geometry of regular variation, we develop prediction intervals when predicted values are large. We discuss some future work in the next chapter.

Chapter 5

Conclusions and Future Work

In this dissertation we extend familiar linear time series models to extremes. Our models are motivated by time series data found in atmospheric sciences, which exhibit dependence in the upper tail. To model extremal dependence we work in the regular variation framework which is a probabilistic framework that is linked to classical extreme value methods and helps model dependence in the tail. To model dependence in the upper tail we restrict our model to take only nonnegative values by considering transformed-linear operations on regularly-varying noise terms.

In Chapter 2, we develop a notion of tail stationarity through defining the tail pairwise dependence function (TPDF), a dependence measure that focuses on the dependence in the tail and has properties analogous to the autocovariance function in classical non-extreme time series. Our notion of tail stationarity is analogous to second order stationarity in classical non-extreme time series models and helps us to focus our entire time series analysis on the TPDF.

In Chapter 3, we develop the transformed-linear regularly-varying models which are analogous to the classical ARMA models. The transformed-linear operations give flexibility to our models by allowing for negative coefficients while still restricting the time series to the positives. Investigating some summary tail statistics that could be of interest, we demonstrate that our transformed-linear regularly-varying time series model outperforms linear regularly-varying models and a non-extreme Gaussian model.

In Chapter 4 we perform prediction for our models. We link the transformed-linear regularly-varying MA(∞) class of models to a Hilbert space which enables us to use the projection theorem to construct methods for prediction via the transformed-linear innovations algorithm. The transformed-linear innovations algorithm also has implications for modeling of our time series. This class of MA(∞) models is rich since using the innovations algorithm, we can fit a transformed-linear regularly-varying MA(∞) to any valid TPDF. We develop prediction intervals when predicted values are large using the polar geometry of regular variation.

There are many avenues for future work. A prevalent method for identifying the orders of AR and MA in classical time series is through looking at autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. This necessitates the development of a PACF analogue for our models. Some initial results on partial tail correlation in the finite dimensions have been worked on by Lee & Cooley in their upcoming paper.

There might also be motivation in extremes to think about a non-causal time series. In the simulated time series in our causal setting, we saw that there is a spike in value and then the time series values reduce in time, that is, we do not get anything ahead of the spike. The fact that our time series models drop off after a high value is somewhat non-physical. There is a reason to think of non-causality because the windspeed this hour has some information about the windspeed in the next hour. Thus there is a need to explore non-causal time series in our setting.

There are some remaining challenges that might be of interest for future work. As discussed at the end of Chapter 3, there is a need to develop improved methods for estimation of tail dependence. Also, in classical time series, model selection is done via likelihood based methods. However, a likelihood based method is challenging in our setting and needs to be investigated further.

Appendix A

Supplementary material for Chapter 3

A.1 Bias in TPDF Estimation of Transformed Regularly-Varying Time Series

A.1.1 Background

Let $\{X_t\}$ be a regularly-varying tail stationary time series with $\alpha = 2$. Let $(X_t, X_{t+h})^T$ be a two-dimensional random vector of elements at lag h of $\{X_t\}$. Thus, there exists a function $b(s)$ and Radon measure $H_{X_t,X_{t+h}}$ on $\Theta_1^+ = \{\mathbf{w} = (w_t, w_{t+h}) \in \bar{\mathbb{X}}^2 : \|\mathbf{w}\|_2 = 1\}$ such that as $s \to \infty$,

$$s\text{Pr}\left(\frac{(X_t, X_{t+h})}{b(s)} \in \cdot\right) \xrightarrow{v} \nu_{X_t,X_{t+h}}(\cdot), \text{ where } \nu_{X_t,X_{t+h}}(dr \times d\mathbf{w}) = 2r^{-3}drdH_{X_t,X_{t+h}}(\mathbf{w}). \quad \text{(A.1)}$$

We also assume that the lower tail condition

$$s\text{Pr}\{X_i \leq \exp(-kb(s))\} \to 0, \quad k > 0, \quad i = 1, ..., p, \quad s \to \infty, \quad \text{(A.2)}$$

is met. We define the tail pairwise dependence function (TPDF) as

$$\sigma(X_t, X_{t+h}) = \int_{\Theta_1^+} w_t w_{t+h} dH_{X_t,X_{t+h}}(\mathbf{w}).$$

To estimate the tail pairwise dependence function (TPDF), we use the estimator defined in Cooley & Thibaud (2019) in which the true angular measure is replaced by an empirical estimate. Let $\{x_t\}$, $(t = 1, \cdots, n)$, be the time series observations. Let $(x_t, x_{t+h})^T, (t = 1, \cdots, n-h)$ be lag-h pairs of observations from $\{x_t\}$. Let $r_t = \|(x_t, x_{t+h})^T\|_2$, and $\mathbf{w}_t = (w_t, w_{t+h})^T =$

$(x_t, x_{t+h})^T/r_t$. The TPDF estimator is defined as

$$\hat{\sigma}(h) = \hat{m} \int_{\Theta_1^+} w_t w_{t+h} d\hat{N}_{X_t, X_{t+h}}(w) = \frac{\hat{m}}{\sum_{t=1}^n \mathbb{I}(r_t > r_0)} \sum_{t=1}^n w_t w_{t+h} \mathbb{I}(r_t > r_0), \tag{A.3}$$

where r_0 is some high threshold for the radial components, $N_{X_t, X_{t+h}}(\cdot) = m^{-1} H_{X_t, X_{t+h}}(\cdot)$, and $\hat{m}$ is an estimate of $H_{X_t, X_{t+h}}(\Theta_1^+)$. Because we preprocessed the time series to have a unit scale, $m = 2$ and does not need to be estimated. When the data are not preprocessed to have a unit scale, an empirical estimator is $\hat{m} = \frac{r_0^2}{n} \sum_{t=1}^n \mathbb{I}(r_t > r_0)$.

It is known that tail dependence estimates tend to have positive bias in the case of weak dependence (refer Huser et al. (2016)). In this supplementary material, we explore the bias associated with estimation of the TPDF. Simulation study shows that subtracting off the mean of the series, considerably reduces bias in TPDF estimation.

A.1.2 Simulated data

First we consider a transformed regularly-varying MA(2) time series

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \theta_2 \circ Z_{t-2},$$

where $\{Z_t\}$ is a sequence of independent positive Fréchet random numbers with unit scale and shape $\alpha = 2$, for four different pairs of values of θ_1 and θ_2. Similarly, we simulate a transformed regularly-varying MA(5) time series

$$X_t = Z_t \oplus \theta_1 \circ Z_{t-1} \oplus \cdots \oplus \theta_5 \circ Z_{t-5},$$

with two different sets of values for $\theta_1, \cdots, \theta_5$, and a transformed regularly-varying AR(1) time series

$$X_t = \phi \circ X_{t-1} \oplus Z_t,$$

with four different values of ϕ. We simulate 50,000 observations x_t, of each of the above time series. For each of the time series, we then obtain the transformed observations $x_t^{(\text{trans})} = g(x_t)$, where g is the inverse cdf of the desired Fréchet. That is, $g(x) = \{-\log \hat{F}(x)\}^{-1/2}$, so that $\Pr(X_t^{(\text{trans})} \leq x) \approx \exp(-x^{-2})$, and $X_t^{(\text{trans})} \in RV_+(2)$ with scale 1.

A.1.3 TPDF estimation

We estimate the TPDF for the first 10 lags for the MA time series, for the first 50 lags for the AR time series with positive ϕ, and for the first 20 lags for the AR time series with negative ϕ, using the estimator in (A.3). TPDF estimation is done in four different ways. Method 1: $\hat{m}$ in (A.3) is estimated as $\frac{r_0^2}{n}\sum_{t=1}^{n}\mathbb{I}(r_t > r_0)$. Method 2: m is considered to be 2, as it theoretically should be after preprocessing the time series to have a unit scale. Method 3: Mean of the time series $X_t^{(\text{trans})}$ is subtracted off before estimation and m is estimated as in Method 1. Method 4: Mean of the time series $X_t^{(\text{trans})}$ is subtracted off before estimation and m is considered to be 2. In method 3 and 4, after we subtract off the mean of the time series, we set the negative values to zero.

The theoretical TPDF for the unit scale MA time series is given by $\frac{\sum_{j=0}^{q}\theta_j^{(0)}\theta_{j+h}^{(0)}}{\sum_{j=0}^{q}(\theta_j^{(0)})^2}$, where $q = 2$ or 5 for the MA(2) and MA(5) time series respectively, and that for the unit scale AR(1) time series is $\left(\phi^h\right)^{(0)}$, where $a^{(0)} = \max(a, 0)$.

A.1.4 Results

Figure A.1 shows the theoretical and estimated TPDF using the four methods described in Section A.1.3 for different transformed MA time series. It can be seen that there is substantial positive bias in TPDF estimation by method 1 and 2. When we subtract off the mean of the time series in method 3 and 4, the bias reduces considerably. However, there is still some bias at higher lags where the theoretical TPDF is zero. Also, estimating m in method 3 underestimates the TPDF substantially. We see similar results in Figure A.2 for the transformed AR time series.

When we simulate data using any regularly-varying noise terms, there always exists a bias problem. Our aim is to minimize the bias problem. A lag-1 bivariate scatter plot of the time series

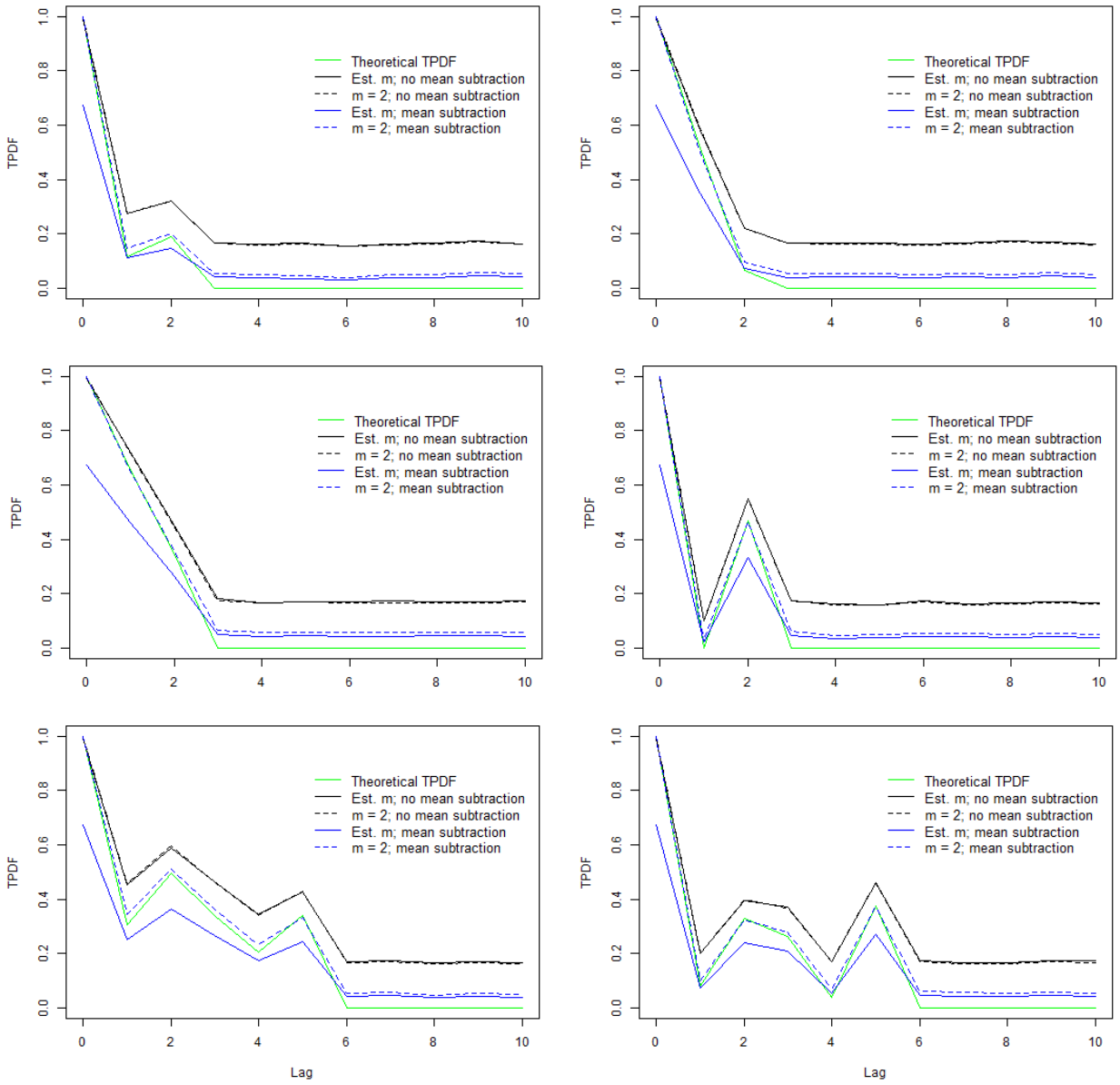

Figure A.1: Comparison of theoretical TPDF and estimated TPDF for transformed regularly-varying MA time series. MA(2): $\theta_1 = 0.1, \theta_2 = 0.2$ (upper left panel), MA(2): $\theta_1 = 0.7, \theta_2 = 0.1$ (upper right panel), MA(2): $\theta_1 = 0.9, \theta_2 = 0.9$ (middle left panel), MA(2): $\theta_1 = -0.5, \theta_2 = 0.7$ (middle right panel), MA(5): $\theta_1 = 0.1, \theta_2 = 0.7, \theta_3 = 0.2, \theta_4 = 0.4, \theta_5 = 0.8$ (lower left panel), and MA(5): $\theta_1 = 0.1, \theta_2 = 0.7, \theta_3 = -0.2, \theta_4 = -0.4, \theta_5 = 0.8$ (lower right panel).

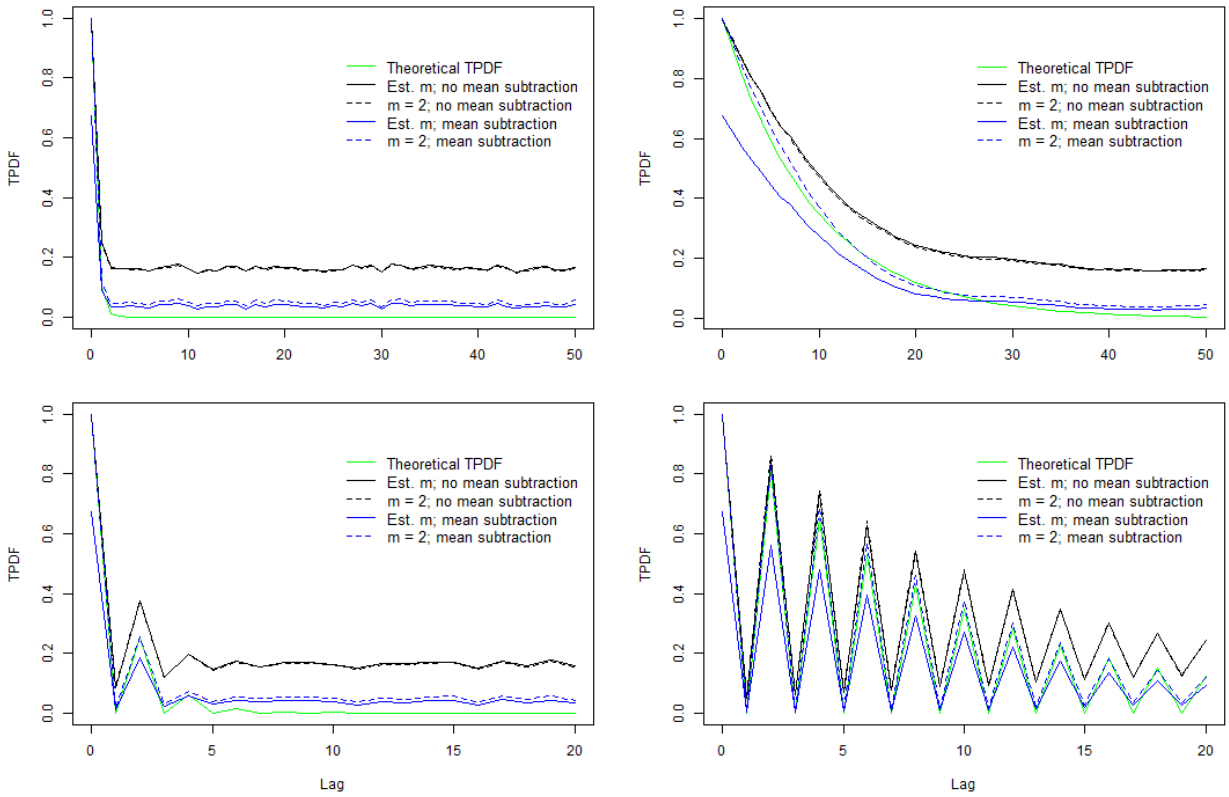

Figure A.2: Comparison of theoretical TPDF and estimated TPDF for transformed regularly-varying AR(1) time series. $\phi = 0.1$ (upper left panel), $\phi = 0.9$ (upper right panel), $\phi = -0.5$ (lower left panel), and $\phi = -0.9$ (lower right panel).

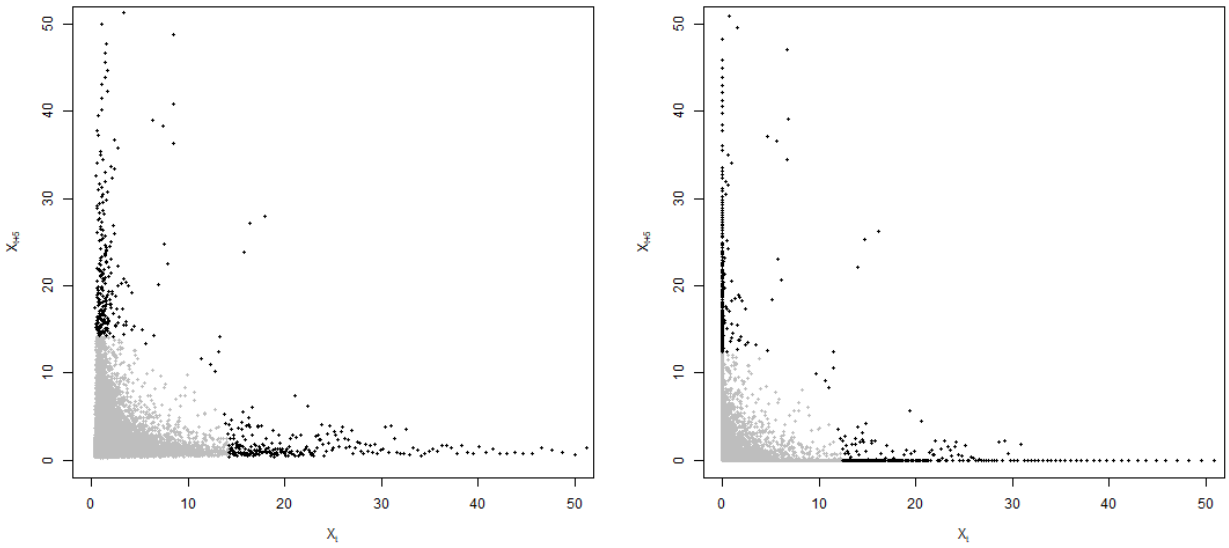

Figure A.3: Lag-5 plots of the MA(2) time series with $\theta_1 = 0.9$ and $\theta_2 = 0.9$. No mean subtracted from the time series (left panel), mean subtracted from the time series (right panel).

observations will have a bunch of large points away from the axes. As the lag increases, the points on the scatter plot spread out and move towards the axes. If there is no long range dependence, those points should eventually align along the axes, but that does not happen. However, the closer we can make the irrelevant points to zero, the better is the behavior of our estimator. Subtracting off the mean of the time series, reduces the bias in TPDF estimation. This can be seen in Figure A.3 which gives lag-5 plots for the MA(2) time series without subtracting the mean off (left panel) and after subtracting the mean off (right panel). We can clearly identify the few points towards the middle of the graph in both the plots, that have a higher value for both of the axes. Subtracting off the mean does not affect these points in the middle of the graph. However, it drives the points near the axes, closer to the axes, thus reducing bias in estimation.

A.1.5 Conclusion

Simulation results indicate that subtracting off the mean of the marginally transformed Fréchet time series considerably reduces bias in TPDF estimation. Also, preprocessing the data to have $\sigma(0) = 1$, thus allowing to consider $m = 2$ instead of estimating m gives better TPDF estimates. There is still bias in the tail which can be further reduced by subtracting the bias off of the TPDF

at higher lags. But it is tricky to address the bias this way since selecting the lag, beyond which to subtract off the bias, will be subjective. However, after subtracting off the mean of the time series, the TPDF estimates at lower lags where dependence exists, are not too different from their theoretical counterparts.

Bibliography

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics.* **31**(3), 307–327.

Bopp, G. P., Shaby, B. A. & Huser, R. (2020). A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association.* 1–14.

Brockwell, P. J. & Davis, R. A. (1991). *Time series: theory and methods.* Springer, New York.

Brockwell, P. J. & Davis, R. A. (2002). *Introduction to time series and forecasting.* Second edition. Springer, New York.

Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer-Verlag, London.

Cooley, D. & Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika.* **106**(3), 587–604.

Davis, R. A. & Mikosch, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli.* **15**(4), 977–1009.

Davis, R. A. & Resnick, S. I. (1989). Basic properties and prediction of max-arma processes. *Advances in Applied Probability.* **21**(4), 781–803.

Dunn, R. J. H. (2019). Hadisd version 3: Monthly updates. hadley centre tech, Technical report, Note 103, 10 pp., www.metoffice.gov.uk/research/library-and-archive

Embrechts, P., Mikosch, T. & Klüppelberg, C. (1997). *Modelling Extremal Events: For Insurance and Finance.* Springer-Verlag, Berlin, Heidelberg.

Groetzner, P. & Dür, M. (2020). A factorization method for completely positive matrices. *Linear Algebra and its Applications.* **591**, 1–24.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics.* **3**(5), 1163–1174.

Huser, R., Davison, A. C. & Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes.* **19**(1), 79–103.

Huser, R., Opitz, T. & Thibaud, E. (2018). Max-infinitely divisible models and inference for spatial extremes. *arXiv preprint.* arXiv:1801.02946.

Kim, M. & Kokoszka, P. (2020). Extremal dependence measure for functional data. *Preprint.*

Kulik, R. & Soulier, P. (2020). *Heavy-tailed time series.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York.

Larsson, M. & Resnick, S. I. (2012). Extremal dependence measure and extremogram: the regularly-varying case. *Extremes.* **15**(2), 231–256.

LEE, J. & COOLEY, D. (2021+). Transformed-linear prediction for extremes. *arXiv preprint arXiv:2111.03754v2.* 2021+.

MHATRE, J. & COOLEY, D. (2020+). Transformed-Linear models for time series extremes. *arXiv preprint arXiv:2012.06705v3.* 2020+.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer-Verlag, New York.

Rudin, W. (1991). *Functional analysis, 2nd ed.* McGraw-hill Inc, New York.

Sargent, T. J. (1979). *Macroeconomic theory.* Economic theory, econometrics, and mathematical economics. Academic Press.

Sklar, M. (1959). Fonctions de Répartition À N Dimensions Et Leurs Marges. *Publications de l'Institut Statistique de l'Université de Paris.* 8, 229-231.

Smith, R. L., Tawn, J. & Coles, S. (1997). Markov chain models for threshold exceedances. *Biometrika.*, 84, 249-268.

Strokorb, K. & Schlather, M. (2015). An exceptional max-stable process fully parameterized by its extremal coefficients. *Bernoulli.* **21**(1), 276–302.

Wadsworth, J. L. & Tawn, J. (2019). Higher-dimensional spatial extremes via single-site conditioning. *arXiv preprint.* arXiv:1912.06560.