8-2013

Extreme Value Theory in Periodic Time Series

Zhiyun Gong
Clemson University, zgong@g.clemson.edu

Extreme Value Theory in Periodic Time Series

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

---

by
Zhiyun Gong
August 2013

Accepted by:
Dr. Robert Lund, Committee Chair
Dr. Peter Kiessler (Co-Chair)
Dr. Colin Gallagher
Dr. Brian Fralix

Abstract

Extreme data points are important in environmental, financial, and insurance settings. In this work, we consider two topics on extremes from environmental data. Many environmental time series have a seasonal structure. The first part presents an approach to identify the rare events of such series based on time series residuals. Here, periodic autoregressive moving-average models are applied to describe the series. The methods justify the application of classical peaks over threshold methods to estimated versions of the one-step-ahead prediction errors of the series. Such methods enable the seasonal means, variances, and autocorrelations of the series to be taken into account. Even in stationary settings, the proposed strategy is useful as it bypasses the need for blocking runs of extremes. The mathematics are justified via a limit theorem for a periodic autoregressive moving-average time series. A detailed application to a daily temperature series from Griffin, Georgia, is pursued.

In the second part, the asymptotic independence between sample means and maxima from a periodic time series is derived. The setup entails a causal periodic autoregressive moving-average model driven by IID periodic noise having a finite $(2 + \delta)$th moment. Our approach takes a regenerative process framework, truncating to reduce the analysis to a periodic moving-average model. The regenerative process is allowed to be partitioned into IID cycles. Here, point process techniques are used to quantify the distribution of the maximums and the Découpage de Lévy Theorem gives the asymptotic independence between maximums and partial sums for the periodic time series.

Acknowledgments

First I would like to thank my advisors, Dr. Robert Lund and Dr. Peter Kiessler. I am deeply indebted to them. They are role models in my heart. Dr. Lund showed to me that excellence is a habit; he works very hard all the time. Dr. Kiessler is smart and always thinks so hard. They are such great advisors. They always encourage me and train me to do research independently and gradually. Some things come to my mind as I write this down. I remember that Dr. Lund is always more excited and happier than me when I made any progresses in my studies. He taught me how to give a talk when I had some trouble in presenting. He is very patient with writing, no matter how many times he revises, no matter where he was, he would mail prompt updates to me. Dr. Kiessler encouraged me that everything will be fine, even when I am upset. He tells me that having fun is the most important issue even when I feel nervous. I can't express my full appreciation for their support and help in words.

I would also like to express my gratitude to my committee members, Dr. Colin Gallagher and Dr. Brian Fralix. I have benefitted from their great comments and suggestions in my research and teaching. I would also like to thank all my teachers at Clemson for helping me and developing my career. I highly appreciate my dear friends, I won't name anyone, for fear of exclusion. They provide a good environment for my studies and great help in life. Lastly I would like to thank my former advisor Dr. Zuoxiang Peng and Dr. Yongcheng Qi for their concern and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Introduction

Extreme value theory is a mathematically fascinating field. Applications abound in an enormous variety of contexts. Consider the following circumstances:

A researcher measures temperatures at each weather observation station at various sites about a city and checks whether they are above or below historical records.

An actuary examines rare but influential events, say quantified by the upper 5% quantile of the portfolio profit-loss distribution.

Dams or dikes at locations along a river or sea must be built high enough to exceed the maximum water height encountered.

A skyscraper is to be built near California and will be subjected to earthquakes. The design structure must be sufficient to withstand extreme shaking.

The above situations involve the largest or smallest data values, i.e., the extremes. Hence, the study of extreme data has attracted widespread interest from scholars in probability and statistics. Extreme value theory provides a theoretical foundation on which meaningful inferences on extreme events can be made.

The most basic result in extreme value theory is the extreme types theorem. The extreme types theorem tells us that for an independent and identically distributed (IID)

random sequence, the limiting distribution of normalizing maximums must have one of three types. Fisher and Tippett (1928) [19] first proposed and proved the theorem; Gnedenko (1943) [21] rigorized the proof. Galambos (1987) [20], Leadbetter, Lindgren and Rootzén (1983) [29], and Resnick (1987) [36] are classical references on extreme value theory. Embrechts, Klüppelberg, and Mikosch (1997) [18] is a good applied probabilistic extreme reference. Leadbetter et al. (1983) [29] extended classical theorems for IID sequences to dependent sequences including stationary series, mixing series, non-stationary series, and stationary continuous time processes.

In statistics, the focus of extreme value theory lies with modeling extreme events, parameter estimation, and hypotheses testing. Smith (1990) [37] and Embrechts, Klüppelberg, and Mikosch (1997) [18] present applications involving extremal events in environmental, insurance, and financial problems. Usually, there are two significant approaches for identifying extremes in real data. One fits a generalized extreme value distribution to block maximums. The second fits a generalized Pareto distribution (GPD) to the exceedances over some fixed threshold (Pickands (1975) [34], Davison and Smith (1990) [15]).

Multivariate extreme value theory is also very important in environmental problems, finance, and insurance. For example, hurricane losses may occur in several states, returns from multiple stocks may need to be studied, or loss amounts under multiple insurance companies may be of interest. Multivariate extreme value methods study modeling and inference issues to understand and manage these phenomena. The multivariate maxima is the vector of componentwise maxima. The limit distributions for multivariate extremes is given in Resnick (1987) [36]. Coles and Tawn (1991 [13], 1994 [14]) show how to apply statistical models for multivariate extremes to real data problems.

Another topic in extremes is the joint behavior of the maximums and partial sums. This problem has been investigated from many angles. This topic was motivated by environmental data, where people tried to determine how extremes influence the means. Historical work started from a sequence of IID random variables (Chow and Teugels (1978) [11]). Anderson and Turkman (1991) [2] extended some results to stationary sequences; they also

analyzed the limiting distribution of sums and maxima for stationary sequences having a heavy-tailed distribution (Anderson and Turkman (1995) [1]). Hsing (1995) [25] derived the asymptotic independence of the sum and maximum for strongly mixing stationary random processes. For Gaussian sequences, Ho and Hsing (1996) [22], Ho and McCormick (1999) [23], McCormick and Qi (2000) [32], and Peng and Nadarajah (2002) [33] derive the joint limiting distribution of the maxima and sums from stationary sequences. James, James, and Qi (2007) [26] also consider this problem for multivariate Gaussian sequences.

In the second chapter of this dissertation, the focus is on identifying rare events from a periodic time series by the peaks over threshold method. A prediction-residual approach for identifying rare events in a periodic time series will be proposed. Mathematical justification and a detailed application to periodic data are given. In the third chapter, we apply the multivariate extreme value theory to show the asymptotic independence between extremes and sample means for periodic time series data. We prove the asymptotic independence between maxima and sample means from a periodic time series. Comments and discussion are presented in section 4.

In the rest of this chapter, background knowledge and results will be stated. Classical extreme value theory is discussed in the Section 1.2. Periodic time series models for time series data will be introduced in Section 1.3. In Section 1.4, we review basic stochastic process knowledge.

1.2 Extreme Value Theory

Let $X_1, \cdots, X_n$ be a sequence of IID random variables with distribution function F and maximum $M_n = \max(X_1, \cdots, X_n)$. Then the distribution function of M_n is

$$P(M_n \leq x) = P(X_1 \leq x, \cdots, X_n \leq x) = F^n(x).$$

Fisher and Tippett (1928) [19] and Gnedenko (1943) [21] stated the extreme types theorem as follows: suppose there exist $a_n > 0, b_n \in R, n \geq 1$, such that

$$P\Big(\frac{M_n - b_n}{a_n} \leq x\Big) = F^n(a_n x + b_n) \to G(x)$$

weakly as $n \to \infty$, where G is a nondegenerate cumulative distribution function. Then G must be one of the following three distribution functions:

i) Fréchet:

$$\Phi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\}, & x \geq 0 \\ 0, & x < 0 \end{cases},$$

for some $\alpha > 0$;

ii) Weibull:

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{-\alpha}\}, & x < 0 \\ 1, & x \geq 0 \end{cases},$$

for some $\alpha > 0$;

iii) Gumbel:

$$\Lambda(x) = \exp\{-e^{-x}\}, \quad x \in R.$$

Here, Φ_α, Ψ_α, and Λ are used to denote the three classical extreme distribution types. By $'$types$'$, we mean that some linear transformation of the scaled maximum must have this distribution.

There are two classical ways to identify extremes in real data. The first approach models the maximum that the variable takes in successive periods. These extreme events are sometimes called block maxima. Consider a random variable representing daily temperature as depicted in Figure 1.1. The block maxima of each block is modeled by the distributions in the classical extreme type theorems. This says that the limiting distribution of the normalized maximum of the sequence must converge to either a Gumbel, Frechet, or Weibull distribution.

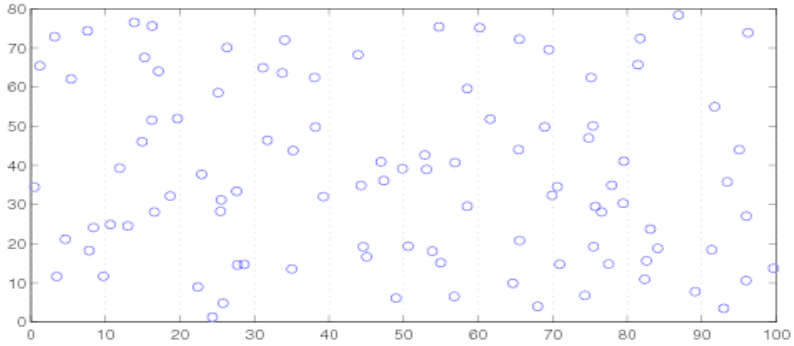

Figure 1.1: Block Maxima

Jenkinson (1955) [27] and von Mises (1936) [40] combined all three extreme distributions into one parametric form, called the generalized extreme value distribution (GEV):

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-\frac{1}{\xi}}\}, & \xi \neq 0 \\ \exp\{-\exp\{-x\}\}, & \xi = 0 \end{cases},$$

where the support set is all x such that $1+\xi x > 0$. The Fréchet distribution can be obtained by setting $\xi = \alpha^{-1}$; $\xi = -\alpha^{-1}$ gives the Weibull distributio; and $\xi = 0$ (take limits as $\xi \to 0$) yields the Gumbel distribution.

An alternative approach considers the process when it exceeds a given threshold u. All process exceedences of the threshold u constitute extreme events, so u is typically chosen to be large. Figure 1.2 below illustrates the idea, called the peaks over threshold (POT) paradigm.

For a random variable X with distribution function F, we are interested in estimating the distribution function F_u of excesses above the threshold u. The distribution function F_u is called the conditional excess distribution function and obeys

$$F_u(y) = P(X - u \leq y | X > u), \quad 0 \leq y \leq x_F - u,$$

where $x_F \leq \infty$ is the right endpoint of F: $x_F = \inf\{x : F(x) = 1\}$. The limiting distribution

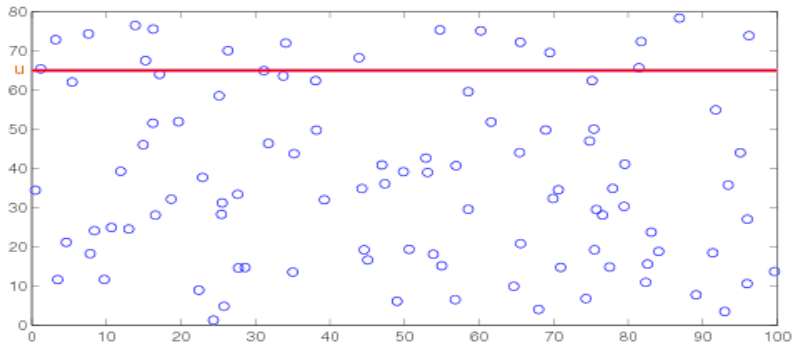

Figure 1.2: Peaks Over Threshold

of the excess as $u \to \infty$ is mathematically justified as the generalized Pareto distribution (GPD). Specifically, for a large class of underlying distribution functions F, the conditional excess distribution function F_u, for large u, is well approximated by

$$F_u(y) = P(X - u \leq y | X > u) \approx G_{\xi,\sigma}(y),$$

where the Pareto form is achieved:

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma} y)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}}, & \xi = 0 \end{cases},$$

for $y \in [0, (x_F - u)]$ if $\xi \geq 0$, and $y \in [0, -\frac{\sigma}{\xi}]$ if $\xi < 0$. (Pickands (1975), Balkema and de Haan (1974) [5]).

1.3 Periodic Stationary Time Series

Time series analysis is a popular branch of statistics and is an important tool in climate, hydrology, water resources, economics, etc. Time series analysis and modeling can be used to build mathematical models to generate synthetic data, to determine the likelihood of extreme events, to forecast future events, to detect trends and change-points,

and to interpolate missing data and extend records.

Definition 1 A time series is a set of time-ordered observations $\{X_t\}$, X_t being the value recorded at a specified time t.

The first step in the analysis of a time series is the selection of a suitable mathematical model for the data. To proceed, we need the notion of a stationary time series.

Definition 2 (Stationarity) The time series $\{X_t, t \in Z\}$ on the index set $Z = \{0, \pm 1, \pm 2, \cdots\}$ is said to be stationary if

$$(i) \quad E[X_t^2] < \infty \quad \text{for all } t \in Z,$$

$$(ii) \quad E[X_t] = m \quad \text{for all } t \in Z,$$

and

$$(iii) \quad E[(X_r - E[X_r])(X_s - E[X_s])] = \text{Cov}(X_r, X_s) = \gamma_X(r,s) = \gamma_X(r+t, s+t) \quad \text{for all } r, s, t \in Z.$$

There is an extremely important class of time series $\{X_t\}$ defined in terms of linear difference equations with constant coefficients: autoregressive moving-average (ARMA) processes. These series play a key role in the modelling of time series data.

Definition 3 (The ARMA(p,q) Process) The process $\{X_t\}$ is said to be an ARMA(p,q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim$ White Noise$(0, \sigma^2)$. We say that $\{X_t\}$ is an ARMA(p,q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p,q) process with zero mean.

Many environmental time series have periodic structures and are not stationary in raw form. Monthly river flows (Vecchia 1985 [39]; Anderson et al. 2007 [3]), monthly stratospheric ozone concentrations (Bloomfield et al. 1994 [8]), daily temperatures (Lund et al. 2006 [31]), and daily ground ozone concentrations (Horowitz 1980; [24] Ballerini

and McCormick 1989; [6] Eastoe and Tawn 2008 [16]) for example, are known to have periodicities in both their means and autocovariances. In these settings, periodic models can be used to describe any periodicities in the correlation structure. However, series with autocovariance periodicities should not be modeled with the seasonal autoregressive moving-average (SARMA) class discussed in Box et al. (1994) [9] and Brockwell and Davis (1991) [10]. An important class of periodic models is the periodic autoregressive moving average (PARMA) class, which are extensions of the commonly used ARMA models that allow for periodic parameters.

Definition 4 (The PARMA(p, q) Process) The process $\{X_t\}$ is said to be a PARMA(p, q) process if $\{X_t\}$ satisfies the periodic linear difference equation

$$(X_{nT+\nu} - \mu_\nu) - \sum_{k=1}^{p(\nu)} \phi_k(\nu)(X_{nT+\nu-k} - \mu_{\nu-k}) = E_{nT+\nu} + \sum_{k=1}^{q(\nu)} \theta_k(\nu) E_{nT+\nu-k}. \tag{3.1}$$

The quantities in (3.1) are as follows: $X_{nT+\nu}$ is the data point from season ν of the nth data cycle, T is the period of the data, which is assumed known throughout, and ν denotes the season and satisfies $1 \leq \nu \leq T$. There is no loss of generality in taking $p(\nu)$ and $q(\nu)$ to be constant in ν; for if $p(\nu)$ and $q(\nu)$ change with ν, one can set $p = \max_{1\leq\nu\leq T} p(\nu)$, and $q = \max_{1\leq\nu\leq T} q(\nu)$, taking $\phi_k(\nu) = 0$ for $p(\nu) < k \leq p$ and $\theta_k(\nu) = 0$ for $q(\nu) < k \leq q$. The autoregressive and moving-average model coefficients during season ν are $\phi_1(\nu), \ldots, \phi_p(\nu)$ and $\theta_1(\nu), \ldots, \theta_q(\nu)$, respectively, and $\{E_t\}$ is zero mean periodic IID noise. All PARMA coefficients are interpreted periodically with period T; for example, $\phi_1(0) = \phi_1(T)$. Seasonal white noise variances are denoted by $\mathrm{Var}(E_{nT+\nu}) = \sigma^2(\nu)$; we assume that $\sigma^2(\nu) > 0$ for all seasons ν so that the series is not perfectly predictable during some season.

PARMA models can be written in a T-variate ARMA form (see Vecchia, 1985):

$$\Phi_0 \mathbf{X}_n - \sum_{k=1}^{p^*} \Phi_k \mathbf{X}_{n-k} = \Theta_0 \mathbf{Z}_0 + \sum_{k=1}^{q^*} \Theta_k \mathbf{Z}_{n-k},$$

where $\mathbf{X}_n = (X_{nT+1}, \cdots, X_{nT+T})'$ and $\mathbf{Z}_n = (Z_{nT+1}, \cdots, Z_{nT+T})'$. The T-variate ARMA orders p^* and q^* are $p^* = \lceil p/T \rceil$ and $q^* = \lceil q/T \rceil$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The $T \times T$ autoregressive coefficients matrices $\{\Phi_k\}, 0 \leq k \leq p^*$ have (i,j)th entries

$$(\Phi_0)_{i,j} = \begin{cases} 1, & i = j \\ 0, & i < j \\ \phi_{i-j}(i), & i > j, \end{cases},$$

$$(\Phi_k)_{i,j} = \phi_{kT+i-j}(i), \quad 1 \leq k \leq p^*,$$

and the convention $\phi_k(\nu) = 0$ for $k > p$ is made. The $T \times T$ moving average coefficients $\{\Theta_k\}, 0 \leq k \leq q^*$ are obtained in a similar manner to the autoregressive coefficients with $\theta_k(\nu)$ replacing each occurrence of $\theta_k(\nu)$. The advantage of this representation is that it links multivariate stationary series to periodic series. In what follows, PARMA models will be the key building block underlying our extremes work for series with periodic properties.

1.4 Regenerative Processes

1.4.1 Markov Chains

A Markov chain is a stochastic process satisfying the Markov property on a state space. Consider a process $\{X_n\}_{n=0}^{\infty}$ with a discrete state space denoted by S. Usually, we take S to be a subset of integers such as $\{0, 1, \cdots, m\}$ (finite state space) or $\{0, 1, \cdots\}$ (infinite state space). The process is assumed to have the following properties:

$$P(X_0 = k) = a_k, \quad k \geq 0,$$

and for any $n \geq 0$,

$$P(X_{n+1} = j | X_n = i) = p_{ij}.$$

The so-called memoryless Markov property is imposed:

$$P(X_{n+1} = j | X_0 = i_0, \cdots, X_{n-1} = i_{n-1}, X_n = i) = p_{ij}.$$

Hence, the conditional probability of the future state X_{n+1} given the past states $X_0, X_1, \cdots, X_{n-1}$ and the present state X_n is independent of the past states, depending only on the present state X_n.

A crucial property of many Markov chains is the existence of a stationary distribution that quantifies the long-run behavior of the chain. Let $\pi = \{\pi_j, j \in S\}$ be a probability distribution on S. Then π is called a stationary distribution for $\{X_n\}_{n=0}^{\infty}$ with transition matrix P if

$$\pi = \pi P$$

i.e., $\pi_j = \sum_{k \in S} \pi_k p_{kj}, j \in S$, where the matrix of one-step transition probabilities p_{ij} is

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

In what follows, the well established theory of Markov chains will help us prove our results.

1.4.2 Regenerative Processes

Renewal processes are a flexible and powerful tool in stochastic processes and are fundamental building blocks in applied probability. Renewal processes model occurrences of events happening at random times, where the times between the events can be approximated by IID random variables. A complex stochastic model often has one or more embedded renewal processes. The idea of regeneration, which allows a process to be decomposed into IID blocks of random lengths, will be useful in this work.

Suppose $\{Y_n, n \geq 0\}$ is a sequence of independent random variables that take on

non-negative values. Furthermore, suppose the sequence $\{Y_n\}_{n=1}^{\infty}$ is identically distributed with common distribution F. Assume that

$$P(Y_n < 0) = 0 \quad \text{and} \quad P(Y_n = 0) < 1.$$

For $n \geq 0$, define

$$S_n = Y_0 + \cdots + Y_n.$$

The sequence $\{S_n\}_{n=0}^{\infty}$ is called a renewal sequence and S_n is called the nth renewal time or epoch. The process is called delayed if $P(Y_0 > 0) > 0$; otherwise it is called pure, and $S_0 = 0 = Y_0$. For a pure renewal process, time zero is considered a renewal epoch. Consider a simple example of returning to a fixed state in a Markov chain $\{X_n\}_{n=0}^{\infty}$; specifically, fix a state i and define the successive return times to state i by

$$\tau_0(i) = \inf\{n : X_n = i\},$$

and, for $n \geq 0$,

$$\tau_{n+1}(i) = \inf\{n > \tau_n(i) : X_n = i\}.$$

Then $\{\tau_n(i)\}_{n=0}^{\infty}$ is a renewal sequence.

Consider a stochastic process $\{X(t), t \in T\}$ where the index set T is either $[0, \infty)$ or $\{0, 1, \cdots\}$. Suppose there is a renewal process $\{S_n\}_{n=1}^{\infty}$ of times where the process regenerates. After one of these regeneration times, say S_n, the process looks probabilistically (exactly) as it did in the last cycle. In this dissertation, a positive recurrent Markov chain with irreducible state space $\{0, 1, \cdots\}$ is often assumed to have started from state zero. Renewal epochs then correspond to the return times to state zero.

Definition 5 (Regenerative Process) A process $\{X(t), t \in T\}$ is called regenerative if there are random times $\{S_n\}_{n=1}^{\infty}$ such that

1. $\{S_n\}_{n=1}^{\infty}$ is a renewal process.

2. The process after any S_n has the same distribution as the whole process. Specifically, for any $0 < t_1 < \cdots < t_k, t_i \in T, 1 \le i \le k$, we have the distributional equality

$$(X(S_n + t_i), 1 \le i \le k) \stackrel{d}{=} (X(t_i), 1 \le i \le k).$$

3. For any S_n, the process beyond time S_n is independent of $\{S_0, S_1, \cdots, S_n\}$; i.e., $\{X(t + S_n), t \in T\}$ is independent of $\{S_0, \cdots, S_n\}$. The nth cycle of the process is the sample path $(X(t), t \in T \cap [S_n, S_{n+1}))$.

There is an important limit theorem about regenerative processes:

Smith's Theorem (Resnick 1992 [35]) Suppose $\{X(t), t \in T\}$ is a regenerative process with the renewal times $\{S_n = \sum_{j=1}^{n} Y_j, n \ge 0\}$ which constitute a pure renewal process. Suppose $E[Y_1] < \infty$. If the distribution of the cycle length Y_1 satisfies a mild regularity condition (satisfied for instance if the distribution is absolutely continuous or concentrates on the integers), then the following limit exists:

$$\begin{aligned}\lim_{t\to\infty} P[X(t) \in j] &= \frac{1}{E[Y_1]} E\left[\int_0^{Y_1} 1_j(X(t))dt\right] \\ &= \frac{E\{\text{occupation time in state } j \text{ in a cycle}\}}{E\{\text{cycle length}\}}.\end{aligned}$$

1.4.3 Coupling

The coupling method is a great tool in modern stochastic processes. To deduce a property of one variable (or process) or to get distributional relations between two or more random variables (or processes), joint structures having the specified marginal properties are often constructed. In this work, coupling methods (Lindvall 1992 [30] and Thorisson 2000 [38]) will be used to connect two Markov chains.

Let $X = (X_n)_{n=0}^{\infty}$ be a Markov chain on a countable state space S with initial distribution $\lambda = (\lambda_i)_{i\in S}$ and transition matrix $P = (p_{ij})_{i,j\in S}$. Assume X is irreducible, aperiodic, and positive recurrent. Then it has a unique stationary distribution π satisfying

the equation $\pi = \pi P$ and

$$\lim_{n\to\infty} \lambda P^n = \pi.$$

Let $X^{'} = \{X_n^{'}\}_{n=0}^{\infty}$ be an independent copy of the same Markov chain, but starting from distribution π at time zero. Since $\pi P^n = \pi$ for all n, $X^{'}$ is stationary; that is, $X_n^{'}$ has distribution π for every $n \geq 1$. Run $\{X_n\}_{n=0}^{\infty}$ and $\{X_n^{'}\}_{n=0}^{\infty}$ together and let

$$T = \inf\{k \in N : X_k = X_k^{'}\}$$

be their first meeting time. Define

$$X^{''} = \begin{cases} X_n & \text{if } n < T \\ \\ X_n^{'} & \text{if } n \geq T \end{cases}.$$

Then, by the Markov strong property, $X^{''} = (X_n^{''})_{n=0}^{\infty}$ is a copy of X. Many things follow from this construction. For example, the bound

$$\sup_A |P_\lambda[X_n \in A] - \pi(A)| \leq P(T > n)$$

can be proven. If $P(T < \infty) = 1$, then $\|\lambda P^n - \pi P^n\| \to 0$, as $n \to \infty$ follows from the Markov inequality.

1.5 Poisson Processes

1.5.1 Point Processes

Point processes are components of solutions of many problems. Point process methods can be used to model the positions and times of earthquakes in the next 50 years, or the breakdown times of a machine or a group of machines.

Suppose $\{X_n\}_{n=1}^{\infty}$ are random elements in the state space E. Define the discrete

measure

$$\epsilon_{X_n}(A) = \begin{cases} 1, & \text{if } X_n \in A \\ \\ 0, & \text{if } X_n \notin A \end{cases}.$$

By summing over n, the total number of points that fall in the set A is obtained. Define the counting measure N by

$$N(\cdot) := \sum_{n=1}^{\infty} \epsilon_{X_n}(\cdot).$$

Then

$$N(A) = \sum_{n=1}^{\infty} \epsilon_{X_n}(A)$$

is the random number of points that fall in the set A. N is called a point process. The mean measure, usually called the intensity of the point process, assigns the measure

$$\mu(A) = E[N(A)]$$

to the set A. Thus, $\mu(A)$ is the expected number of points in the set A.

1.5.2 Poisson Processes

The most common point process model is the Poisson process. It arises as a limiting process of records (above a large threshold) in this work. Suppose N is a point process on the state space E. Let $\mathcal{B}$ be a class of measurable subsets of E. Then N is called a Poisson process with mean measure μ or Poisson random measure (PRM(μ)) if N satisfies

(1) For $A \in \mathcal{B}$,

$$P(N(A) = k) = \begin{cases} \frac{e^{-\mu(A)}(\mu(A))^k}{k!}, & \text{if } \mu(A) < \infty \\ \\ 0, & \text{if } \mu(A) = \infty. \end{cases}$$

(2) If $A_1, \cdots, A_k$ are disjoint subsets of E, then $N(A_1), \cdots, N(A_k)$ are independent random variables.

Hence, N is Poisson if the random number of points in the set A is Poisson dis-

tributed with parameter $\mu(A)$ and the number of points in disjoint regions are independent random variables.

Chapter 2

Identifying Rare Events from A Periodic Time Series

2.1 Introduction

Many environmental time series have apparent periodicities in both their means and autocovariance structures. Figure 2.1 plots 67 years of daily temperatures (24,455 points) from Griffin, Georgia from January 1, 1931 — December 31, 1997. Here, the periodic mean cycle is evident, with summer temperatures being hotter than winter temperatures. Comparing the relative homogeneity of the annual peaks (the July maximums) to the jaggedness of the corresponding troughs (the January minimums), one also sees a seasonal variance cycle with winter temperatures being more variable than summer temperatures. Figure 2.2 plots the Griffin sample means, sample variances, and sample lag-one autocorrelations by day of year from top to bottom. There, the seasonal mean and variance cycles are visually apparent. It turns out that the sample lag-one autocorrelations are also periodic (see Lund *et al.* 2006 [31] for justification), with higher correlations occurring during the Fall when long runs of clear sunny days are common; lower correlations occur during late winter when cold front passages are more frequent and serve to break runs of similar weather.

Before proceeding, we remark that the Griffin series was parsimoniously modeled in

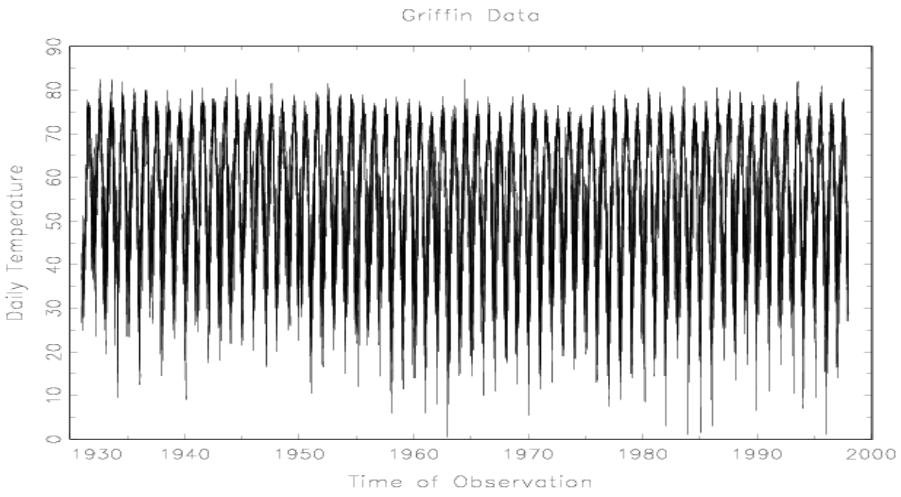

Figure 2.1: Daily Griffin Temperatures

Lund *et al.* (2006) [31]. There, a periodic autoregressive moving-average (PARMA) model was developed using only 12 model parameters. Even a first-order periodic autoregression would use $2 \times 365 = 730$ parameters, so there is considerable savings in this parsimonious parametrization. Leap year (February 29th) data was ignored to make the period 365 days. There is no missing data in this series, nor is there strong evidence of trends or changepoints (changepoints, or discontinuities in the mean of the series, frequently arise when temperature gauges are changed or the station location is moved). This data was recorded at The University of Georgia's Agricultural Experimental Station in Griffin, GA, which is approximately 38 miles south of Atlanta, GA. A daily temperature is the average of daily high and low temperatures. This definition is made for spring-loaded gauges (where marking needles are pressed out to daily lows and highs) as it makes data monitoring a once daily task.

While extreme problems frequently arise in environmetrics, methods to handle this

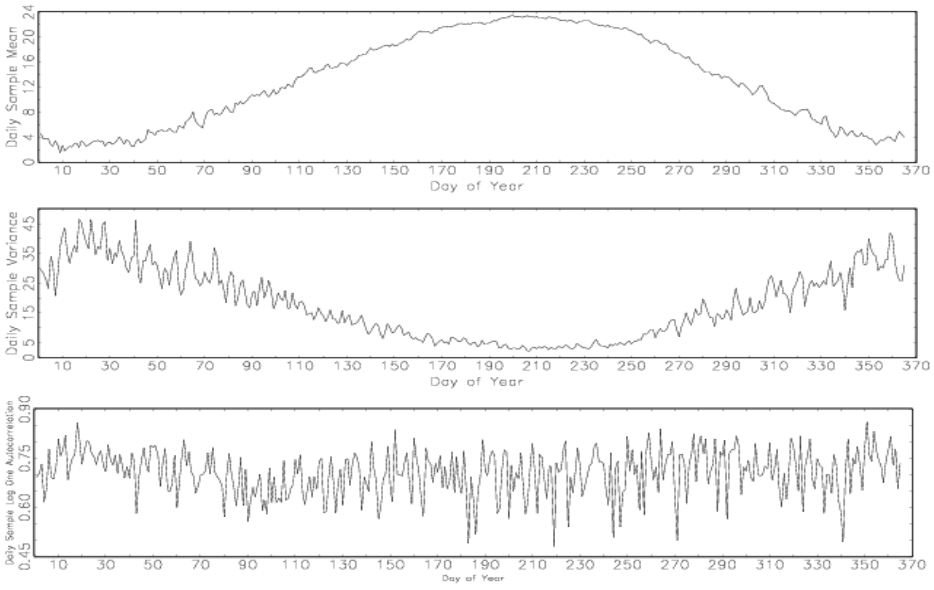

Figure 2.2: Daily Moments of The Griffin Series

issue are somewhat ad hoc. This said, two different approaches are apparent. The first approach simply examines the largest/smallest values from the raw series, irrespective of the season at which they occur. Threshold modeling techniques here adopt a time-constant threshold and examine the times and magnitudes of the threshold exceedences. There may be physical justification behind a time-constant threshold. For example, an insurance policy might specify the temperature threshold, or, events that breach a dam height that does not vary with the season might be of interest. Seasonality makes such an analysis with the Griffin series seem shortsighted. Indeed, Griffin temperatures below 25 degrees Fahrenheit would be common in January and virtually impossible in July. Hence, this approach would not account for rarity of the event as measured by frequency.

A second approach, and one that does account for seasonality, essentially sets a time-varying threshold. This allows data from all seasons to enter into the study. Because

such an approach does not consider true extremes of the process, we term it a "rare event" analysis. Our concern in this article lies with this type of approach. As the Griffin series is well-described by a PARMA model, even in the extremes, we use the PARMA model to construct a set of one-step-ahead prediction residuals that gauge how extreme each new observation is relative to the series' past. In computing these one-step-ahead prediction residuals, the periodic means, variances, and autocorrelations of the series are accounted for.

A drawback with the residual approach is that the residuals may not flag times where the series creeps up/down to an extreme; for example, a situation where the temperature becomes one degree warmer for 20 consecutive days may never produce a large enough residual (among the 20 days) to signal a rare event. Nonetheless, as we will see, residual methods produce useful information and will capably identify the rare events in the Griffin series. The residual approach here meshes well with the peaks over threshold (POT) paradigm of Pickands (1975) [34] (see Embrechts *et al.* (1997) [18] and Coles (2001) [12] for textbook discussions): one simply analyzes the threshold exceedences of the one-step-ahead prediction residuals. Intrinsically, this puts all seasons on the same footing and provides a better assessment of which observations are the rarest. Since autocorrelations are incorporated in the one-step-ahead predictions and the estimated prediction residuals are approximately uncorrelated, the task of blocking exceedence runs is bypassed. Essentially, the PARMA residuals serve to set a seasonal threshold and decluster runs of extremes in the daily temperatures. As such, the methods also have uses in stationary (non-periodic) settings. Daily temperatures, for example, are usually positively correlated, with runs of cold and warm days being common. Due to correlation, consecutive threshold exceedences are not independent events.

Methodologically, our paper involves both periodic time series methods and extreme value theory. If the series is uncorrelated, our methods would merely set a periodic threshold as in Koch *et al.* (2005) [28] or Anton *et al.* (2008) [4]. Eastoe and Tawn (2009) [16] is probably the most similar work to this paper as we both preprocess the original series

and then use POT techniques. Eastoe and Tawn (2009) [16], however, do not consider autocorrelation issues in detail — their focus was more on covariates, which we do not have with the Griffin series.

The rest of this chapter proceeds as follows. In Section 2, PARMA models and their properties are reviewed. Section 3 states our main mathematical result. There, we state that seasonally scaled estimates of the one-step-ahead prediction residuals have the same tail behavior as the scaled theoretical errors. Proofs are delegated to an Appendix. Section 4 returns to examine the Griffin series in detail. Comments conclude the chapter.

2.2 PARMA Series

Consider a periodic series $\{X_t\}$ satisfying the PARMA difference equation

$$X_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} + E_{nT+\nu} + \sum_{k=1}^{q} \theta_k(\nu) E_{nT+\nu-k}. \tag{2.1}$$

The quantities in (1.1) are as follows: T is the period of the data, which is assumed known throughout; p and q are the autoregressive and moving-average model orders, respectively, which are assumed constant in season ν, $1 \leq \nu \leq T$; the autoregressive and moving-average model coefficients during season ν are $\phi_1(\nu), \ldots, \phi_p(\nu)$ and $\theta_1(\nu), \ldots, \theta_q(\nu)$, respectively, and $\{E_t\}$ is zero mean periodic white noise. Jones and Brelsford (1967), Vecchia (1985), Anderson *et al.* (1999, 2007, and 2008), Lund and Basawa (2000), Basawa and Lund (2001), and Adams and Goodwin (2008) is a sample of articles in the PARMA-model literature. Hurd and Gerr (1991) and Gardner *et al.* (2006) are other useful periodic-process references.

We interchangeably use notations such as $\{X_t\}$ and $\{X_{nT+\nu}\}$, the latter being preferred when emphasis is on seasonality. The index ν denotes the season throughout and satisfies $1 \leq \nu \leq T$. Hence, $X_{nT+\nu}$ is the data point from season ν of the nth data cycle. All PARMA-model coefficients are interpreted periodically with period T; for example, $\phi_1(0) = \phi_1(T)$. To allow for a non-zero seasonal mean $\mu_\nu = E[X_{nT+\nu}]$, we merely examine $\{X_{nT+\nu} - \mu_\nu\}$. By examining $\{X_{nT+\nu} - \mu_\nu\}$, one need only consider the case where $\mu_\nu \equiv 0$,

sometimes, trends and seasonal means are first removed from the series to render the data as a zero mean. Seasonal white-noise variances are denoted by $\mathrm{Var}(E_{nT+\nu}) = \sigma^2(\nu)$, and we take $\sigma^2(\nu) > 0$ for all seasons ν so that the series is not perfectly predictable during some season. Equation (1.1) can be written as

$$X_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} + \sigma(\nu) Z_{nT+\nu} + \sum_{k=1}^{q} \theta_k(\nu)\sigma(\nu-k) Z_{nT+\nu-k}, \tag{2.2}$$

where $\{Z_t\}$ is mean-zero unit-variance white noise. For the moment, we assume that $\{Z_t\}$ is independent and identically distributed (IID). As is usual, we assume that the PARMA model is causal, invertible, and that all model parameters are identifiable. Causality implies that solutions to (2.1) are unique in mean square and have the form

$$X_{nT+\nu} = \sum_{k=0}^{\infty} \psi_k(\nu) Z_{nT+\nu-k}, \tag{2.3}$$

where the weights satisfy $\sum_{k=0}^{\infty} |\psi_k(\nu)| < \infty$ for each season ν.

Solutions to (2.1) and (2.2) are periodic in that $\mathrm{Cov}(X_{n+T}, X_{m+T}) = \mathrm{Cov}(X_n, X_m)$, for each integer n and m. In fact, with IID $\{Z_t\}$, $\{X_t\}$ is a strictly stationary sequence in a periodic sense (this means that the joint distributions of $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+T}, \ldots, X_{t_n+T})$ are identical for each $n \geq 1$ and choices of $t_1 < t_2 < \cdots < t_n$). Invertibility implies that

$$Z_{nT+\nu} = \sum_{k=0}^{\infty} \pi_k(\nu) X_{nT+\nu-k}, \tag{2.4}$$

for some sequence of weights $\{\pi_k(\nu)\}$ satisfying $\sum_{k=0}^{\infty} |\pi_k(\nu)| < \infty$. Invertibility is important in later proofs. The parameter-identifiability assumption allows us to estimate PARMA-model coefficients from d complete cycles of data; specifically, one can construct $\sqrt{d}$-consistent and asymptotically normal estimates of the PARMA parameters $\phi_1(\nu), \ldots, \phi_p(\nu)$; $\theta_1(\nu), \ldots, \theta_q(\nu); \sigma^2(\nu)$ for each season ν (Anderson and Vecchia 1993 and Basawa and Lund 2001 provide details).

Let $\gamma_\nu(h) = \mathrm{Cov}(X_{nT+\nu}, X_{nT+\nu-h})$ be the covariance of $\{X_t\}$ at lag $h \geq 0$ during season ν. A numerical algorithm for calculating the $\gamma_\nu(h)$'s for a causal PARMA series is presented in Shao and Lund (2004). Explicit expressions for $\gamma_\nu(h)$ in terms of the PARMA coefficients are known only for periodic moving-averages (Lund and Basawa 2000) and PARMA(1,1) series (Obeysekera and Salas 1986). From the PARMA parameters, it is easy to compute the one-step-ahead linear predictions,

$$\hat{X}_t = P(X_t | X_1, \ldots, X_{t-1}, 1), \tag{2.5}$$

and their unconditional mean squared errors, $v_t = E[(X_t - \hat{X}_t)^2]$. Anderson *et al.* (1999) and Lund and Basawa (2000) give Innovation Algorithm equations for this task in terms of the PARMA-model coefficients. The quantity 1 is included in the set of predictands in (2.5) to accommodate non-zero-mean series.

A prediction residual at time t is $(X_t - \hat{X}_t)/v_t^{1/2}$. For each fixed season ν, $v_{nT+\nu}$ monotonically decreases to $\sigma^2(\nu)$ as n increases (Lund and Basawa 2000); moreover, this convergence takes place at a geometric rate that is uniform over the season. Hence, we define the prediction residual R_t at time t as

$$R_{nT+\nu} = \frac{X_{nT+\nu} - \hat{X}_{nT+\nu}}{\sigma(\nu)}.$$

Suppose that $\hat{\phi}_1(\nu), \ldots, \hat{\phi}_p(\nu); \hat{\theta}_1(\nu), \ldots, \hat{\theta}_q(\nu); \hat{\sigma}^2(\nu)$ are $\sqrt{d}$-consistent estimators of the PARMA parameters at season ν. By plugging these estimates into the prediction equation for $\hat{X}_t$, one can construct an estimated version of $\hat{X}_t$, which we denote by $\hat{X}_t^*$. The estimated residuals are hence defined as

$$\hat{R}_{nT+\nu} = \frac{X_{nT+\nu} - \hat{X}^*_{nT+\nu}}{\hat{\sigma}(\nu)}. \tag{2.6}$$

An example with periodic autoregressions (PAR) may illuminate issues. Later, we will see that the Griffin data is well described by a third-order periodic autoregression.

Consider a causal PAR(p) series with mean μ_ν at season ν. This series obeys

$$X_{nT+\nu} = \mu_\nu + \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} + \sigma(\nu) Z_{nT+\nu}.$$

One can allow the PAR order p to depend on the season ν if desired, but we will not do this here for two reasons. First, such a model can be reduced to one with the constant PAR order $p^* = \max_{1 \le \nu \le T} p_\nu$, when one sets $\phi_k(\nu) = 0$ for $k > p_\nu$. Second, the changes in temperature data are smooth from season to season (one does not go from summer's peak to winter's low in a couple of days), making Fourier expansions more attractive than say wavelet expansions.

PAR models are frequently preferred in practice because prediction and estimation are easy tasks. Specifically, the one-step-ahead PAR(p) prediction is simply

$$\hat{X}_{nT+\nu} = \mu_\nu + \sum_{k=1}^{p} \phi_k(\nu)(X_{nT+\nu-k} - \mu_{\nu-k}), \qquad nT + \nu > p,$$

and the prediction errors are $v_{nT+\nu} = \sigma^2(\nu)$ when $nT + \nu > p$. The "start-up" residuals for t with $t \le p$ are usually neglected with little loss of precision. PAR coefficients can be estimated from periodic versions of the Yule-Walker equations (see Basawa and Lund 2001). Asymptotically, these moment estimators are as efficient as likelihood estimators. The seasonal means and variances are estimated via

$$\hat{\mu}_\nu = \frac{1}{d} \sum_{n=0}^{d-1} X_{nT+\nu}, \qquad \hat{\sigma}^2(\nu) = \frac{1}{d} \sum_{n=0}^{d-1} (X_{nT+\nu} - \hat{X}^*_{nT+\nu})^2.$$

From these, we obtain the estimated residuals,

$$\hat{R}_{nT+\nu} = \frac{X_{nT+\nu} - \hat{\mu}_\nu - \sum_{k=1}^{p} \hat{\phi}_k(\nu)(X_{nT+\nu-k} - \hat{\mu}_{\nu-k})}{\hat{\sigma}(\nu)}.$$

We now move our discussion back to extremes and rare events. Classical extreme value theory establishes limit laws for linearly scaled versions of $M_t = \max_{1 \le k \le t} X_k$ or examines process exceedences above a time-homogeneous threshold. Suppose that $\{X_t\}$ is

a strictly stationary (non-periodic) zero-mean series satisfying

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j},$$

where $\{Z_t\}$ is an IID zero-mean sequence. Typical extreme value assumptions are that the tail cumulative distribution function of Z_1 has a regularly varying tail with parameter ξ and that $\sum_{j=0}^{\infty} |c_j| < \infty$. As shown in Leadbetter *et al.* (1983), Resnick (1987), Chernick *et al.* (1991), and Embrechts *et al.* (1997, Section 8.1.1), if there exist extreme value norming sequences $\{a_n\}$ and $\{b_n\}$ such that $nP(Z_1 > a_n x + b_n) \to x^{-\xi}$ as $n \to \infty$ and $\xi > 0$, then

$$P\{M_n \leq a_n x + b_n\} \to (\exp^{-x^{-\xi}})^{\eta}, \tag{2.7}$$

where the parameter η is called the extremal index. Hsing *et al.* (1988) (see also section 8.1.2 of Embrechts *et al.* 1997) show how to interpret η as an inverse of the "expected cluster size of exceedence runs". In many settings, especially when series autocovariances are absolutely summable (this is the case for causal PARMA models), $\eta \leq 1$ (Section 8.1 of Embrechts *et al.* 1997 gives specifics and Theorem 4.4.7 there quantifies the Gaussian case where η is typically unity). While our work has non-zero seasonal means to contend with, ξ is still generally viewed as the most critical parameter. We suspect that the tail parameter of X_ν is also ξ for each season ν in the zero-mean case when $\xi > 0$, but do not consider this issue further because of the non-zero seasonal means and the desire to identify rare events over all seasons (and not simply the absolute maximum or minimum). We further comment about links to classical extreme value methods in our comments section.

For a threshold viewpoint, suppose Z_t follows the conditional tail Pareto law

$$\bar{F}(x) = P(Z_t > x + u | Z_t > u) = \left(1 + \frac{\xi x}{\kappa}\right)_+^{-1/\xi}, \tag{2.8}$$

where $u > 0$ is some pre-set large threshold, $x_+ = \max(x, 0)$, ξ is the shape parameter, and $\kappa > 0$ is a scale parameter. This is the classical POT paradigm of Pickands (1975) and has

vast applications. The results above suggest that the tail parameter of Z_t is related to that for X_t (at least in the zero-mean case). Hence, to quantify ξ, a sample of Z_t's would be useful. Unfortunately, we cannot recover the Z_t's directly from the X_t's. However, we do have the $\hat{Z}_t$'s. What remains to be seen is that no loss of asymptotic precision is incurred in working with $\{\hat{Z}_t\}$ in lieu of $\{Z_t\}$ — that is the subject of a theorem in the next section.

2.3 A Technical Result

The technical crux of this work establishes that the estimated residuals can be used in place of the residuals without asymptotic loss of precision.

Theorem 3.1: Suppose that $\{X_t\}$ is a causal, invertible, and identifiable PARMA series. Then for a PAR(p) model, $\max_{\ell(d)T<t\leq dT} |R_t - \hat{R}_t| \xrightarrow{\mathcal{P}} 0$. For a general PARMA series with a moving-average component, $\max_{\ell(d)T<t\leq dT} |R_t - \hat{R}_t| \xrightarrow{\mathcal{P}} 0$, for any sequence $\ell(d)$ such that $\ell(d) \to \infty$ and $\ell(d)/d \to 0$, as $d \to \infty$. Here, d is the number of complete cycles of observed data.

The proof of this result is presented in the Appendix A. One must crop the first p residuals in the PAR(p) series; this is the usual edge-effect issue with one-step-ahead predictions. Models with moving-average components are harder to analyze as the edge effects do not die out in any finite time horizon. This is why the sequence $\ell(d)$ tending to infinity is needed.

2.4 The Griffin Series

We now attempt to quantify the rarest cold events in the Griffin record. The coldest absolute daily temperature on record is 0.5 degrees Fahrenheit, recorded on December 13, 1962. For comparative purposes, the lowest low temperature recorded at the National Weather Service's Griffin location (this differs from the Agricultural Experiment Station) is -8 degrees for January 21, 1985 (the Agricultural Experiment Station lists 2.5 degrees as the

daily average for this day). Lund *et al.* (2005) fitted a parsimonious PARMA model to the Griffin series: a PAR(3) model was judged optimal. The PAR coefficients were consolidated with Fourier fits of the form

$$\phi_k(\nu) = A_k + B_k \cos\left(\frac{2\pi(\nu - \tau_k)}{T}\right), \quad k = 1, 2, 3.$$

Table 1 lists estimated coefficients. The parsimonious white-noise variances are

$$\hat{\sigma}^2(\nu) = 8.227 + 6.609 \cos\left(\frac{2\pi(\nu - 27.446)}{T}\right).$$

Table 2.1: Griffin Parsimonious PAR(3) Coefficients

k	A_k	B_k	τ_k
1	0.814	0.105	-1.834
2	-0.177	-0.149	15.986
3	0.0470	0.0689	24.832

This model cannot be further reduced at the 95% confidence level. While the seasonal means μ_ν were not consolidated with Fourier fits, we comment that it would take at least four harmonics to accurately tune them. Indeed, the descent from summer to winter takes place more rapidly than the ascent from winter to summer (see the top panel in Figure 2).

Figure 2.3 plots estimated residuals from the PAR(3) model fit, multiplied by minus one so that positive residuals correspond to cold conditions. All seasons appear to be on the same scale; specifically, the residuals have zero mean and unit variance (approximately) for each season ν. Figure 2.4 is the diagnostic sample mean threshold exceedence plot of Davison and Smith (1990): for each threshold, the average of all data points exceeding this threshold is plotted. If POT methods are to work well, this plot should be roughly linear with slope $1/(1-\xi)$. Indeed, Figure 2.4 seems "very linear"; in comparison to this plot for other data sets, this is indeed well-behaved.

We select the threshold $u = 3.0$ and fit the tail Pareto distribution in (2.8) to the

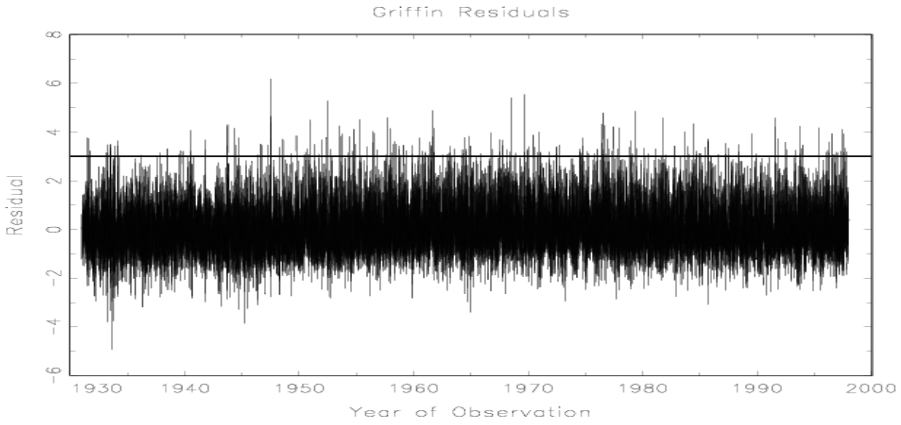

Figure 2.3: Daily Griffin Estimated Residuals

amount that the estimated residuals exceed 3.0. While one could push the threshold below 3.0, there are 232 residuals that exceed 3.0, which is plenty for our purposes. Maximum likelihood estimates of the Pareto parameters are $\hat{\xi} = 0.00640 \pm 0.00661$ and $\hat{\sigma} = 0.496 \pm 0.0465$. The error margins represent one standard error and were obtained by inverting the observed information matrix; a negative log likelihood of 70.894 was achieved. From this, one could infer that ξ is zero, which reduces the Pareto POT distribution to an exponential distribution. However, we work with a general ξ in what follows as the tail distribution in (2.8) has a closed form for any ξ and no computational simplicity is gained by taking $\xi = 0$.

Table 2 lists the times of the 10 largest residuals. The largest residual was 6.16 and occurred on July 23, 1947. Here, a daily temperature of 59.5 degrees was observed. Indeed, a mid-summer daily temperature this low is noteworthy (summer temperatures peak in mid July in the northern hemisphere), especially since July temperatures have small variabilities. Figure 2.5 plots the 1947 daily temperatures along with their residuals. There, one sees that large positive residuals generally occur on the colder days. The large temperature drop from the three days preceding July 23 (July 20 averaged 73 degrees, July 21 averaged 69.5

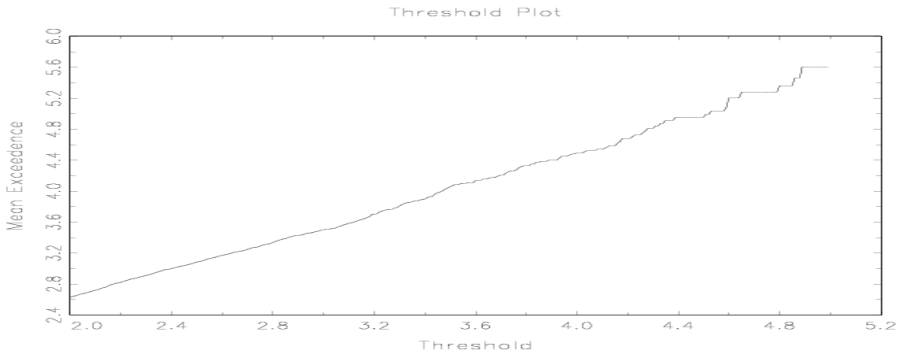

Figure 2.4: Sample Mean of Residuals Exceeding Threshold

degrees, and July 22 averaged 72.5 degrees) is also evident.

Table 2.2: The Ten Largest Residuals

Rank	Residual	Date	Rank	Residual	Date
1	6.160	Jul 23, 1947	6	4.859	May 26, 1979
2	5.548	Aug 23, 1969	7	4.795	Jul 19, 1976
3	5.403	Jun 29, 1968	8	4.643	Jun 23, 1947
4	5.286	Jul 11, 1952	9	4.599	Sep 25, 1957
5	4.889	Aug 16, 1961	10	4.594	Oct 3, 1981

There is an irregular aspect to Table 2: large residuals seem more likely during summer. This structure is investigated further in Figure 2.6, which shows an estimated daily probability of a residual exceeding $u = 3.0$. Elaborating, an empirical probability that a residual exceeds 3.0 is computed for each day of the year and then smoothed via a uniform moving average with a 21-day span. To assess whether or not summer threshold exceedences are more likely, a uniformity test was constructed. Specifically, the 92 day summer period from June 21 – September 21 was examined and seen to have 81 exceedances. If the exceedance pattern is uniform, the number of exceedances that fall into this period should be binomially distributed with $n = 232$ trials and success probability $p = 92/365$ (58.48

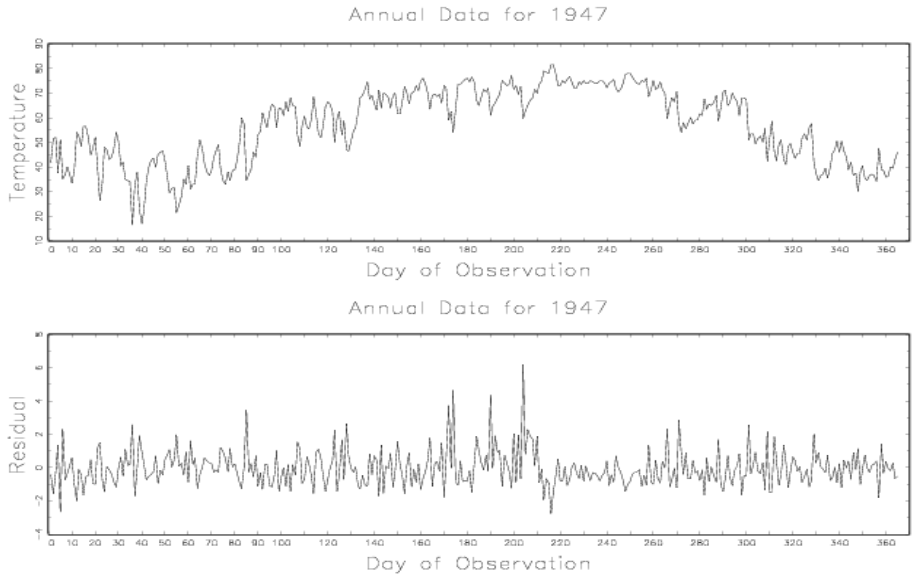

Figure 2.5: Data and Residuals for 1947

were expected). A large-sample standard normal statistic for this test is $z = 3.4056$ and has a one-sided p-value of 0.000325. Indeed, summer threshold exceedences appear more common.

One may also doubt the periodic IID noise assumption. Specifically, in (1.2), it was assumed that the season-ν innovation had the form $\sigma(\nu)Z_{nT+\nu}$ — a periodic multiplier times an IID sequence. This would not allow individual seasons to have differing tail distribution parameters. To investigate this aspect further, we refitted the 232 residuals above with a seasonal version of (2.8) that accounts for the day of year at which the exceedance occurred. In other words, we fitted the seasonal POT model

$$P(R_{nT+\nu} > x + u | R_{nT+\nu} > u) = \left(1 + \frac{\xi x}{\kappa_\nu}\right)_+^{-1/\xi} \tag{4.1}$$

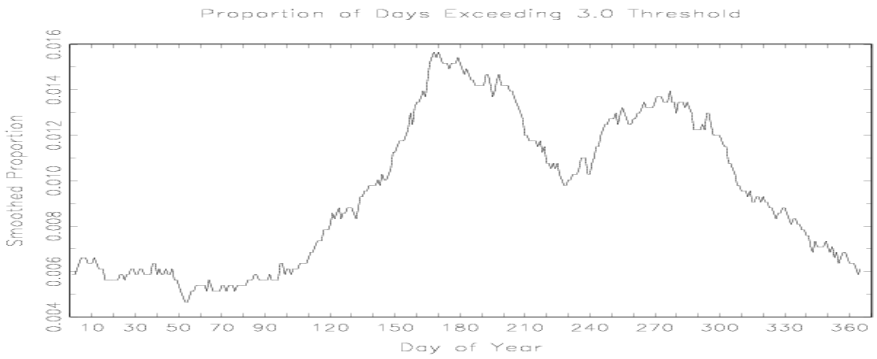

Figure 2.6: Smoothed Probabilities of Threshold Exceedence

under the first-order Fourier constraint

$$\kappa_\nu = D + G\cos\left(\frac{2\pi(\nu-\eta)}{T}\right). \tag{4.2}$$

Maximum likelihood estimators are $\hat{\xi} = -0.0977 \pm 0.0610$, $\hat{D} = 0.498 \pm 0.0477$, $\hat{G} = 0.223 \pm 0.0430$, and $\hat{\eta} = 214.264$, with a negative log likelihood of 59.999. As the negative log likelihood dropped by 10.895 points (also, the standard error of $\hat{G}$ makes a z-score of 5.186), one infers that G is nonzero and that a seasonal κ is preferable. Attempts to fit seasonally varying structures to ξ produced insignificant results. As Coles (2001, page 106) notes, it is usually difficult to estimate the shape parameter ξ with precision and modeling ξ as a smooth function of time is usually unwarranted.

A threshold stability property holds for the model in (4.1) with constraints as in (4.2). Indeed, if one increases the threshold parameter from u to $u^* > u$, then

$$
\begin{aligned}
P(R_{nT+\nu} > x | R_{nT+\nu} > u^*) &= \frac{P(R_{nT+\nu} > x)}{P(R_{nT+\nu} > u^*)} \\
&= \frac{P(R_{nT+\nu} > x | R_{nT+\nu} > u)}{P(R_{nT+\nu} > u^* | R_{nT+\nu} > u)} \\
&= \frac{\left(1 + \frac{\xi(x-u)}{\kappa_\nu}\right)_+^{-1/\xi}}{\left(1 + \frac{\xi(u^*-u)}{\kappa_\nu}\right)_+^{-1/\xi}}.
\end{aligned}
$$

Simplifying this expression gives

$$
P(R_{nT+\nu} > x | R_{nT+\nu} > u^*) = \left[1 + \frac{\xi(x-u^*)}{\kappa_\nu + \xi(u^*-u)}\right]_+^{-1/\xi}.
$$

Hence, if a larger threshold is selected, the ξ parameter stays constant and the seasonal κ_ν's all shift by $\xi(u^* - u)$. This is the seasonal analogy of the time-homogeneous threshold stability discussed by Coles (2001). Moreover, since the κ_ν's all shift by the same amount for all seasons ν, only the location parameter of the Fourier parametrization in the expansion for the κ_ν's changes with a higher threshold — the shape parameters are unchanged.

At this point, we have a probability model for the tail of the residuals at each season and a seasonal estimated probability of exceeding the threshold, which we call $e(\nu)$ at season ν. This information can be used to assess how rare each residual is. Table 3 lists the ten rarest cold events, rare being judged by overall probability of exceedance. For a residual at season ν, the overall probability of exceedance is

$$
P(R_{nT+\nu} > r + u) = P(R_{nT+\nu} > r + u | R_{nT+\nu} > u) e(\nu).
$$

In Table 3, $e(\nu) = P(R_{nT+\nu} > u)$ is estimated via Figure 6, and $P(R_{nT+\nu} > r + u | R_{nT+\nu} > u)$ is estimated with the seasonal Pareto model fitted in (4.1).

While Tables 2 and 3 contain some common events, March 27, 1955, which barely missed entry into Table 2, is now estimated as the rarest overall event. March 27, 1955 saw an average temperature of 21.5 degrees, very cold for this time of year. The three days

Table 2.3: The Ten Rarest Cold Temperatures

Rank	Date	Residual	Probability of Exceedence
1	Mar 27, 1955	4.520	0.0000272
2	Jul 23, 1947	6.160	0.0000416
3	Aug 23, 1969	5.548	0.0000837
4	Nov 26, 1950	4.506	0.000105
5	Dec 12, 1957	4.152	0.000129
6	Dec 7, 1977	4.177	0.000189
7	Jun 29, 1968	5.403	0.000229
8	May 26, 1979	4.859	0.000283
9	Feb 25, 1934	3.650	0.000294
10	Aug 16, 1961	4.889	0.000368

preceding March 27 had temperatures (residuals in parentheses) on March 24, 1977: 41.0 degrees (-0.116); on March 25, 1977: 45.0 degrees (-0.428); and on March 26, 1977: 50.0 degrees (-0.421). The July 23, 1947 event now is gauged as the second rarest.

It is instructive to flesh out the effects of autocorrelation in the analyses. A method that ignores autocorrelation but accounts for periodic means and autocovariances is easily devised by seasonally standardizing $\{X_t\}$. That is, we examine $\{[X_{nT+\nu} - \hat{\mu}_\nu]/\hat{\gamma}_\nu(0)^{1/2}\}$, where $\hat{\gamma}_\nu(0)$ is the sample day-ν variance. Such a method merely sets a periodic threshold for the data as in Koch *et al.* (2005) and Anton *et al.* (2008). To compare to previous results, we select a threshold of 2.5522 to get the same number (232) of threshold exceedences as before. Here, correlation is seen in the exceedences. Figure 2.7 plots the 1977 data against the seasonal threshold $\hat{\mu}_\nu - 2.5522\widehat{\mathrm{Var}}(X_{nT+\nu})$. There are only two exceedences for this year, which occurred on June 8 and June 10 (several near misses may change this count should the thresholds be smoothed). Table 4 lists specifics for the five-day period, June 7-11, 1977.

A combination of seasonal and non-seasonal methods may be useful. For example, one could set a time-constant threshold of say 20 degrees and use the residuals to decluster (in a manner that we will not elaborate upon here) runs of cold weather. The driving property here is that $\{R_t\}$ is uncorrelated in time but $\{X_t\}$ is not.

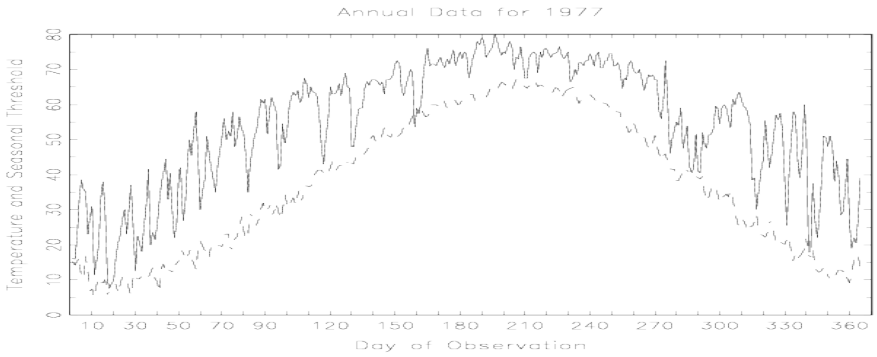

Figure 2.7: Data for 1977 with a Seasonal Threshold

Table 2.4: The Cold Snap of June 7-11, 1977

Date	Temperature	Seasonally Adjusted Residual	PAR(3) Residual
June 7, 1977	69.0	-0.167	0.258
June 8, 1977	53.5	3.308	4.219
June 9, 1977	59.0	2.382	-0.101
June 10, 1977	57.5	2.804	2.316
June 11, 1977	60.5	2.162	0.265

2.5 Remarks

The prediction-residual approach plausibly identified the rare events in the Griffin series. There are climatological implications of the results. For example, the cold day of Jan 21, 1985 (average temperature of 2.5 degrees; one-step-ahead residual of 3.132) is not found in Tables 2 or 3, implying that this mid-winter temperature record is more likely to be broken sooner than daily records set by some of the other listed events.

It also seems worthwhile to further connect this work to extreme value theory with a constant threshold — especially in a way that accounts for periodic means, variances, and a possibly negative ξ. Some results in this direction are known from Ballerini and McCormick (1989), Konstant and Piterbarg (1993), Ferreira and Martins (2003), and Scotto (2007).

Chapter 3

Asymptotic Independence Between Extremes and Sample Means

3.1 Introduction

In this chapter, we want to establish the asymptotic independence between maximums and partial sums in a periodic time series. The topic is interesting from both theoretical and practical standpoints. For example, while average temperatures are known to be warming at most locations on the planet, asymptotic independence between partial sums and maximums implies that changes in extremes need not follow the same pattern as mean changes.

Consider a PARMA series $\{X_t\}$ satisfying the PARMA(p, q) difference equation

$$X_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} + E_{nT+\nu} + \sum_{k=1}^{q} \theta_k(\nu) E_{nT+\nu-k}. \tag{1.1}$$

The quantities in (1.1) are as follows: $X_{nT+\nu}$ is the data point from season ν of the nth data cycle, T is the period of the data, which is assumed known throughout, and ν denotes the season and satisfies $1 \leq \nu \leq T$. The autoregressive and moving-average model orders are p and q, and are assumed constant in season ν for simplicity. The autoregressive and

moving-average model coefficients during season ν are $\phi_1(\nu), \ldots, \phi_p(\nu)$ and $\theta_1(\nu), \ldots, \theta_q(\nu)$, respectively, and $\{E_t\}$ is zero mean independent noise, that may have a marginal distribution at time t that depends on the season of time t.

All PARMA coefficients are interpreted periodically with period T; for example, $\phi_1(0) = \phi_1(T)$. Seasonal white noise variances are denoted by $\mathrm{Var}(E_{nT+\nu}) = \sigma^2(\nu)$; we assume that $\sigma^2(\nu) > 0$ for all seasons ν so that the series is not perfectly predictable during some season. Equation (1.1) can be written as

$$X_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} + \sigma(\nu) Z_{nT+\nu} + \sum_{k=1}^{q} \theta_k(\nu)\sigma(\nu-k) Z_{nT+\nu-k}, \tag{1.2}$$

where $\{Z_t\}$ is zero-mean unit-variance independent noise, (again, the distribution of Z_t at time t may depend on the season for time t). As usual, we assume that the PARMA model is causal, invertible, and that all model parameters are identifiable. Causality implies that solutions to (1.1) are unique in mean square and have the form

$$X_{nT+\nu} = \sum_{k=0}^{\infty} \psi_k(\nu) Z_{nT+\nu-k}, \tag{1.3}$$

where the weights satisfy $\sum_{k=0}^{\infty} |\psi_k(\nu)| < \infty$ for each season $\nu \in \{1, 2, \cdots, T\}$.

The PARMA model can be written in the vector form

$$\mathbf{X}_n = \sum_{k=0}^{\infty} \Psi_k \mathbf{Z}_{n-k}, \tag{1.4}$$

where

$$\mathbf{X}_n = (X_{nT+1}, \cdots, X_{nT+T})', \quad \text{and} \quad \mathbf{Z}_n = (Z_{nT+1}, \cdots, Z_{nT+T})'.$$

Here, Ψ_0 is the lower triangular matrix

$$\Psi_0 = \begin{bmatrix} \psi_0(1) & 0 & \cdots & 0 \\ \psi_1(2) & \psi_0(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{T-1}(T) & \psi_{T-2}(T) & \cdots & \psi_0(T) \end{bmatrix}_{T\times T},$$

and for $k \geq 1$,

$$\Psi_k = \begin{bmatrix} \psi_{kT}(1) & \psi_{kT-1}(1) & \cdots & \psi_{kT-T+1}(1) \\ \psi_{kT+1}(2) & \psi_{kT}(2) & \cdot & \psi_{kT-T+2}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{kT+T-1}(T) & \psi_{kT+T-2}(T) & \cdots & \psi_{kT}(T) \end{bmatrix}_{T\times T}.$$

Define the partial sums and normalized extremes via

$$\mathbf{S}_n = \sum_{k=0}^{n} \mathbf{X}_k$$

and

$$\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n) = \Big(\frac{\max_{k=0}^{n}\{X_{kT+1}\} - b_{n,1}}{a_{n,1}}, \cdots, \frac{\max_{k=0}^{n}\{X_{kT+T}\} - b_{n,T}}{a_{n,T}}\Big).$$

Here, $a_{n,\nu}$ and $b_{n,\nu}$ are the normalizing extreme value constants for $\mathbf{X}_t$ at season ν. We assume that the maginal distribution of $X_{nT+\nu}$ is in the domain of attraction of some extreme value distribution $G(\cdot)$ for each season ν, although the extreme distribution might depend on the season, and the partial sums of $\{\mathbf{X}_n\}$ are asymptotically normal–say $N(0,\Sigma)$. Our objective is to show the asymptotic independence between $n^{-\frac{1}{2}}\mathbf{S}_n$ and $\mathbf{a}_n{}^{-1}(\mathbf{M}_n - \mathbf{b}_n)$ for this periodic time series.

The approach taken here uses a regenerative stochastic process to analyze the PARMA series. This is accomplished by showing that our process can be approximated by an mth order Markov chain. This Markov approximation permits the time series to be partitioned into IID cycles. The maximum of the time series is the maximum of cycle maxima and the sum is the sum of all cycle sums. Hence, technical details essentially reduce

to showing that the maximum and the sum are asymptotically independent for IID random variables. Our work will use the Découpage de Lévy Theorem tol establish the asymptotic independence between maximums and partial sums.

The idea of the Découpage de Lévy Theorem is as follows (Resnick 2007) [36]: Suppose $\{X_n, n \geq 1\}$ are IID random elements of a metric space S with Borel sets $\mathscr{S}$. Fix a set $B \in \mathscr{S}$ such that $P[X_1 \in B] > 0$. Let $K^+(i)$ be those indices j for which $X_j \in B$; *i.e.*, let $K^+(0) = 0$ and $K^+(i) = \inf\{j > K^+(i-1) : X_j \in B\}$ for $i \geq 1$. Similarly define $\{K^-(i)\}$ by $K^-(0) = 0$ and $K^-(i) = \inf\{j > K^-(i-1) : X_j \in B^c\}$ for $i \geq 1$. Also, define $N(n) = \sup\{i : K^+(i) \leq n\}$ as the number of times X_t lies in B from time 1 to n. The Découpage de Lévy Theorem states that $\{X_{K^+(i)}\}, \{X_{K^-(i)}\}, \{N(i), i \geq 1\}$ are independent and $\{X_{K^\pm(i)}\}$ is IID with

$$P[X_{K^+(i)} \in A] = P[X_1 \in A | X_1 \in B], \quad A \subset B, \quad A \in \mathscr{S};$$

$$P[X_{K^-(i)} \in A] = P[X_1 \in A | X_1 \notin B], \quad A \subset B^c, \quad A \in \mathscr{S}.$$

Furthermore, $\{N(n), n \geq 1\}$ is a renewal counting function with $E[N(n)] = nP[X_1 \in B]$.

Our approach is completely different than previous analyses of similar problems, but seems to have merit beyond the PARMA paradigm considered here. Section 2 in this chapter constructs our regenerative process representation for the causal PARMA model. Section 3 constructs three independent sequences from maximums, partial sums, and point processes inspired by the Découpage de Lévy Theorem. The asymptotic independence between sample means and maximums is obtained from this setup, and the joint limit distribution is also derived in this section.

3.2 Regenerative Representation

The mth order truncated PARMA version of (1.4) is

$$\mathbf{X}_n^m = \sum_{k=0}^{m-1} \Psi_k \mathbf{Z}_{n-k}.$$

Our first result shows that the truncation remainder is negligible in probability as $m \to \infty$.

Lemma 1 If $\{\mathbf{X}_t\}$ is a causal PARMA series, then for any season ν, as $m \to \infty$,

$$\max_{1 \le i \le n} |X_{iT+\nu} - X_{iT+\nu}^m| \xrightarrow{P} 0.$$

Proof. For any season ν and $1 \le i \le n$,

$$X_{iT+\nu} - X_{iT+\nu}^m = \sum_{k=m}^{\infty} \sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j}.$$

By strict periodic stationarity, $P(A_1 \cup A_2 \cup \cdots \cup A_n) \le \sum_{i=1}^n P(A_i)$, and the Markov inequality, we have

$$\begin{aligned}
&P\Big(\max_{1 \le i \le n} \big| \sum_{k=m}^{\infty} \sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j} \big| > \epsilon \Big) \\
&\le nP\Big(\big| \sum_{k=m}^{\infty} \sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j} \big| > \epsilon \Big) \\
&\le \frac{n}{\epsilon^2} E\Big[\Big(\sum_{k=m}^{\infty} \sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j} \Big)^2 \Big].
\end{aligned}$$

Minkowski's integral inequality gives

$$\Big[E\Big(\sum_{k=m}^{\infty} \sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j} \Big)^2 \Big]^{\frac{1}{2}} \le \sum_{k=m}^{\infty} \Big[E\Big(\sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j} \Big)^2 \Big]^{\frac{1}{2}}$$

and the Cauchy-Schwarz inequality provides $|E[Z_t Z_s]| \leq E[Z_t^2]^{\frac{1}{2}} E[Z_s^2]^{\frac{1}{2}} = 1$. Thus,

$$\begin{aligned}
&E\Big[\Big(\sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j}\Big)^2\Big] \\
&= \sum_{j=1}^{T}\sum_{l=1}^{T} \psi_{kT+\nu-j}(\nu)\psi_{kT+\nu-l}(\nu) E(Z_{(i-k)T+j} Z_{(i-k)T+l}) \\
&\leq \sum_{j=1}^{T}\sum_{l=1}^{T} \psi_{kT+\nu-j}(\nu)\psi_{kT+\nu-l}(\nu).
\end{aligned}$$

Causality of the PARMA model implies that $|\psi_k| < \kappa r^k$ for some $r < 1$ and finite κ. Using this in the bound above gives

$$E\Big[\Big(\sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j}\Big)^2\Big] \leq T^2 \kappa^2 r^{2(kT+\nu-T)}.$$

Combining the above produces

$$\begin{aligned}
P\Big(\max_{1\leq i\leq n} |\sum_{k=m}^{\infty}(\sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) Z_{(i-k)T+j})| > \epsilon\Big) &\leq \frac{n}{\epsilon^2}\Big(\sum_{k=m}^{\infty} T\kappa r^{(kT+\nu-1)}\Big)^2 \\
&\leq \frac{nT^2\kappa^2 r^{2(mT+\nu-1)}}{(1-r^T)^2\epsilon^2} \to 0,
\end{aligned}$$

as $m \to \infty$. This finishes our work.

□

Next, we focus on the Markov structure of the truncated PARMA process. Let $P_{\mathbf{Z}}$ be the distribution of $\mathbf{Z}_n$, which does not depend on n. For a fixed $m \geq 1$, set $\mathbf{\Phi}_n = (\mathbf{Z}'_{n-m+1}, \cdots, \mathbf{Z}'_n)'$ and observe that $\{\mathbf{\Phi}_n\}$ forms a Markov chain on the state space R^{mT}. If the current state is $\phi_n = (\mathbf{z}'_{n-m+1}, \cdots, \mathbf{z}'_n)$, the chain next moves to state $(\mathbf{z}'_{n-m+2}, \cdots, \mathbf{z}'_n, \mathbf{Z}'_{n+1})$, according to the transition kernel

$$P(\mathbf{\Phi}_{n+1} \in B | \mathbf{\Phi}_n = \phi_n) = 1_{\{B_1 \times \cdots \times B_{m-1}\}}(\mathbf{z}_{n-m+2}, \cdots, \mathbf{z}_n) P_{\mathbf{Z}}(B_m),$$

where $B = B_1 \times \cdots \times B_m$ and each $B_i \in \mathcal{B}(R^T)$. When $n \geq m$, $\mathbf{\Phi}_n = (\mathbf{Z}'_{n-m+1}, \cdots, \mathbf{Z}'_n)'$ and $\mathbf{\Phi}_0 = (\mathbf{Z}'_{-m+1}, \cdots, \mathbf{Z}'_0)'$ are independent. Thus, for $n \geq m$,

$$P^{(n)}(\phi_0, B) = P_{\mathbf{Z}}(B_1) \times \cdots \times P_{\mathbf{Z}}(B_m) := \Pi_{\mathbf{\Phi}}(B),$$

where $\Pi_{\mathbf{\Phi}} = P_{\mathbf{Z}} \times \cdots \times P_{\mathbf{Z}}$ is a product measure on $\mathcal{B}(R^{mT})$.

It now follows that

$$\begin{aligned}
&\int_{R^T} \cdots \int_{R^T} \Pi_{\mathbf{\Phi}}(d\phi_0) P(\phi_0, B) \\
&= \int_{R^T} \cdots \int_{R^T} P_{\mathbf{Z}}(d\mathbf{z}_{-m+1}) \cdots P_{\mathbf{Z}}(d\mathbf{z}_0) 1_{\{B_1 \times \cdots \times B_{m-1}\}}(\mathbf{z}_{-m+2}, \cdots, \mathbf{z}_0) P_{\mathbf{Z}}(B_m) \\
&= \int_{R^T} P_{\mathbf{Z}}(d\mathbf{z}_{-m+1}) \int_{R^T} 1_{B_1}(\mathbf{z}_{-m+2}) P_{\mathbf{Z}}(d\mathbf{z}_{-m+2}) \cdots \int_{R^T} 1_{B_{m-1}}(\mathbf{z}_0) P_{\mathbf{Z}}(d\mathbf{z}_0) P_{\mathbf{Z}}(B_m) \\
&= 1 \cdot P_{\mathbf{Z}}(B_1) \cdots P_{\mathbf{Z}}(B_{m-1}) P_{\mathbf{Z}}(B_m) = \Pi_{\mathbf{\Phi}}(B);
\end{aligned}$$

that is, $\Pi_{\mathbf{\Phi}}(\cdot)$ is both a stationary and limiting distribution of the chain.

Since $\{\mathbf{\Phi}_n\}$ is an m-dependent Markov chain, the random vectors $\mathbf{\Phi}_m, \mathbf{\Phi}_{2m}, \cdots$ are IID with distribution $\Pi_{\mathbf{\Phi}}$ regardless of the initial state ϕ_0. For a set B, define $\tau_B = \inf\{n > 0; \mathbf{\Phi}_{nm} \in B\}$. Then τ_B is geometrically distributed with parameter $\Pi_{\mathbf{\Phi}}(B)$ (specifically, τ_B has a finite geometric moment of some order). Hence,

$$P(\tau_B = 1 | \mathbf{\Phi}_0 = \phi_0) = P(\mathbf{\Phi}_m \in B | \mathbf{\Phi}_0 = \phi_0) = P(\mathbf{\Phi}_m \in B) = \Pi_{\mathbf{\Phi}}(B).$$

and

$$P(\tau_{\mathcal{B}} = 2 | \mathbf{\Phi}_0 = \phi_0) = P(\mathbf{\Phi}_m \notin B, \mathbf{\Phi}_{2m} \in B | \mathbf{\Phi}_0 = \phi_0) = [1 - \Pi_{\mathbf{\Phi}}(B)]\Pi_{\mathbf{\Phi}}(B).$$

Induction now gives $P(\tau_B = n) = [1 - \Pi_{\mathbf{\Phi}}(B)]^{n-1}\Pi_{\mathbf{\Phi}}(B)$;

Let $\mathbf{\Phi}_0 = (\mathbf{Z}'_{-m+1}, \cdots, \mathbf{Z}'_0)'$ be a point in the support of $\Pi_{\mathbf{\Phi}}$. To avoid trivial work, we assume that this point is $(\mathbf{0}', \cdots, \mathbf{0}')'$. For every $\epsilon > 0$, there exists an open set $\mathbf{U}_\epsilon$ such

that $\mathbf{0} \in \mathbf{U}_\epsilon \subset \mathcal{B}(\mathbf{0}, \epsilon) = \{\mathbf{x} \in R^T, \|\mathbf{x}\| \leq \epsilon\}$. Here, $\|\cdot\|$ denotes the usual Euclidean norm.

Set $B_\epsilon = \mathbf{U}_\epsilon \times \cdots \times \mathbf{U}_\epsilon$ and suppose for simplicity that $\int_{\mathbf{U}_\epsilon} x dP_{\mathbf{Z}}(x) = 0$. Since $\Pi_{\mathbf{\Phi}}(B_\epsilon) > 0$, $\{\mathbf{\Phi}_n\}$ returns to B_ϵ infinitely often with probability 1. In fact, the above argument shows that return times to B_ϵ have a geometric tail.

The transition probabilities for $\{\hat{\mathbf{\Phi}}_t\}$ have the following property. If $\hat{\mathbf{\Phi}}_t = (\hat{\mathbf{z}}'_{t-m+1}, \cdots, \hat{\mathbf{z}}'_t)' \in B_\epsilon$, then $P(\hat{\mathbf{\Phi}}_{t+1} \in \mathcal{A} | \hat{\mathbf{\Phi}}_t) = P(\mathbf{\Phi}_{t+1} \in \mathcal{A} | \mathbf{\Phi}_t = \mathbf{0})$; if $\hat{\mathbf{\Phi}}_t = (\hat{\mathbf{z}}'_{t-m+1}, \cdots, \hat{\mathbf{z}}'_t)' \notin B_\epsilon$, then $P(\hat{\mathbf{\Phi}}_{t+1} \in \mathcal{A} | \hat{\mathbf{\Phi}}_t) = P(\mathbf{\Phi}_{t+1} \in \mathcal{A} | \mathbf{\Phi}_t)$. Observe that

$$\|\hat{\mathbf{\Phi}}_t, \mathbf{\Phi}_t\| = \left[\sum_{i=0}^{m-1} (\hat{\mathbf{Z}}_{t-m+i} - \mathbf{Z}_{t-m+i})^2\right]^{\frac{1}{2}} \leq (m4\epsilon^2)^{\frac{1}{2}}.$$

Now define a new series $\{\hat{\mathbf{X}}_t\}$ by ϵ-coupling as follows: If $\mathbf{\Phi}_t = (\mathbf{Z}'_{t-m+1}, \cdots, \mathbf{Z}'_t)' \in B_\epsilon$ and $|\mathbf{X}_t - \mathbf{X}_t^m| = |\sum_{k=m}^\infty \Psi_k \mathbf{Z}_{t-k}| < \epsilon$, then set $\hat{\mathbf{X}}_t = \mathbf{0}$; that is, $\hat{\mathbf{\Phi}}_t = (\hat{\mathbf{Z}}'_{t-m+1}, \cdots, \hat{\mathbf{Z}}'_t)' = (\mathbf{0}', \cdots, \mathbf{0}')'$, and $(\hat{\mathbf{Z}}'_{t-m}, \cdots)' = (\mathbf{0}', \cdots)'$. If $\mathbf{\Phi}_t = (\mathbf{Z}'_{t-m+1}, \cdots, \mathbf{Z}'_t)' \notin B_\epsilon$ or $|\mathbf{X}_t - \mathbf{X}_t^m| = |\sum_{k=m}^\infty \Psi_k \mathbf{Z}_{t-k}| > \epsilon$, then simply set $\hat{\mathbf{X}}_t = \mathbf{X}_t$.

The above construction shows that $\{\hat{\mathbf{X}}_t\}$ is a regenerative process with regeneration point $\mathbf{0}$ and that $\|\hat{\mathbf{X}}_t - \mathbf{X}_t\| < C\epsilon$ for every $t \geq 0$, and some constant C. To see this, observe that when t is not a regeneration point, $\|\mathbf{X}_t - \hat{\mathbf{X}}_t\| = 0$. When t is a regeneration point,

$$\begin{aligned}
\|\hat{\mathbf{X}}_t - \mathbf{X}_t\| &= \left\| \sum_{k=0}^{m-1} \Psi_k (\hat{\mathbf{Z}}_{t-k} - \mathbf{Z}_{t-k}) + \sum_{k=m}^{\infty} \Psi_k (\hat{\mathbf{Z}}_{t-k} - \mathbf{Z}_{t-k}) \right\| \\
&\leq \sum_{k=0}^{m-1} \|\Psi_k\| \|(\hat{\mathbf{Z}}_{t-k} - \mathbf{Z}_{t-k})\| + \left\| \sum_{k=m}^{\infty} \Psi_k \mathbf{Z}_{t-k} \right\| \\
&\leq m\epsilon \max_{0 \leq k \leq m-1} \|\Psi_k\| + \epsilon = C\epsilon.
\end{aligned}$$

Now embed cycles in the chain as follows. Set $C_0 = 0$ and

$$\sigma_0 = \inf \left\{ t \geq 0, \mathbf{\Phi}_t \notin B_\epsilon \text{ or } \left| \sum_{k=m}^{\infty} \Psi_k \mathbf{Z}_{n-k} \right| > \epsilon \right\}.$$

For $k \geq 1$, set

$$C_k = \inf \left\{ t > \sigma_{k-1}, \mathbf{\Phi}_t \in B_\epsilon \text{ and } \left| \sum_{k=m}^{\infty} \Psi_k \mathbf{Z}_{n-k} \right| < \epsilon \right\}$$

and

$$\sigma_k = \inf \left\{ t > C_k, \mathbf{\Phi}_t \notin B_\epsilon \text{ or } \left| \sum_{k=m}^{\infty} \Psi_k \mathbf{Z}_{n-k} \right| > \epsilon \right\}.$$

The above construction decomposes $\{\hat{\mathbf{X}}_t\}$ into independent and identically distributed cycles. In short, there exist nonnegative integer valued random variables $C_1, C_2, \cdots, C_{N(n)}, C_{N(n)+1}$, where $N(n)$ is the number of cycles up to time n, such that $\{\hat{\mathbf{X}}_t\}_{t=1}^n$ can be divided into IID cycles: $(\hat{\mathbf{X}}_{C_0}, \cdots, \hat{\mathbf{X}}_{C_1}); (\hat{\mathbf{X}}_{C_1+1}, \cdots, \hat{\mathbf{X}}_{C_2}); \cdots; (\hat{\mathbf{X}}_{C_{N(n)}+1}, \cdots, \hat{\mathbf{X}}_{C_{N(n)+1}})$; $(\hat{\mathbf{X}}_{C_{N(n)+1}+1}, \cdots, \hat{\mathbf{X}}_n)$. Note that the first and last cycle may be incomplete. For notation, let the index set of the kth cycle be $\mathcal{C}_k = \{C_k + 1, \cdots, C_{k+1}\}$, $k = 1, \cdots, N(n)$, where $\mathcal{C}_0 = \{0, \cdots, C_1\}$, and $\mathcal{C}_{N(n)+1} = \{C_{N(n)+1} + 1, \cdots, n\}$.

For the kth cycle, define approximating cycle sums and maxima via

$$\hat{\mathbf{S}}_k = (\hat{S}_{k,1}, \cdots, \hat{S}_{k,T}) = \left(\sum_{i \in \mathcal{C}_k} \hat{X}_{iT+1}, \cdots, \sum_{i \in \mathcal{C}_k} \hat{X}_{iT+T} \right);$$

$$\hat{\mathbf{M}}_k = (\hat{M}_{k,1}, \cdots, \hat{M}_{k,T}) = \left(\max_{i \in \mathcal{C}_k} \hat{X}_{iT+1}, \cdots, \max_{i \in \mathcal{C}_k} \hat{X}_{iT+T} \right).$$

The finite $(2+\eta)$th moment for partial sums in every cycle is proved as follows.

Lemma 2 Suppose that $\{\hat{\mathbf{X}}_t\}$ is the above regenative process constructed from the PARMA series $\{\mathbf{X}_t\}$ with $E[|Z_t|^{2+\delta}] < \infty$ for some $\delta > 0$ and all $\nu \in \{1, 2, \cdots, T\}$. Then for each season ν, $E[\hat{S}_{1,\nu}] = 0$ and $E[|\hat{S}_{1,\nu}|^{2+\eta}] < \infty$ for some $\eta > 0$.

Proof. The regenerative process decomposition gives, for each complete cycle length C,

$$\frac{E\left[\sum_{i=0}^{C-1} \hat{X}_{iT+\nu}\right]}{E(C)} = E(\hat{X}_\nu) = E\left[\sum_{k=0}^{m-1} \sum_{j=1}^{T} \psi_{kT+\nu-j}(\nu) \hat{Z}_{(1-k)T+j} \right] = 0.$$

Hence, $E[\hat{S}_{1,\nu}] = 0$ for any season ν.

For the finite $(2+\delta)$th moment claim of $\hat{\mathbf{X}}_k$, observe that

$$\left(E|\hat{X}_{iT+\nu}|^p\right)^{\frac{1}{p}} = \left(E\left|\sum_{k=0}^{m-1}\sum_{j=1}^{T}\psi_{kT+\nu-j}(\nu)\hat{Z}_{(i-k)T+j}\right|^p\right)^{\frac{1}{p}}.$$

By Minkowski's integal inequality with $p = 2+\delta$,

$$\left(E\left|\sum_{k=0}^{m-1}\sum_{j=1}^{T}\psi_{kT+\nu-j}(\nu)\hat{Z}_{(i-k)T+j}\right|^p\right)^{\frac{1}{p}} \leq \left(E\left[\sum_{k=0}^{m-1}\left|\sum_{j=1}^{T}\psi_{kT+\nu-j}(\nu)\hat{Z}_{(i-k)T+j}\right|\right]^p\right)^{\frac{1}{p}}$$
$$\leq \left(E\left[\sum_{k=0}^{\infty}\left|\sum_{j=1}^{T}\psi_{kT+\nu-j}(\nu)\hat{Z}_{(i-k)T+j}\right|\right]^p\right)^{\frac{1}{p}} \leq \sum_{k=0}^{\infty}\left(E\left|\sum_{j=1}^{T}\psi_{kT+\nu-j}(\nu)\hat{Z}_{(i-k)T+j}\right|^p\right)^{\frac{1}{p}}.$$

The inequality $|f+g|^p \leq 2^{p-1}(|f|^p + |g|^p)$ gives

$$\begin{aligned}
&E\left[\left|\sum_{j=1}^{T}\psi_{kT+\nu-j}(\nu)\hat{Z}_{(i-k)T+j}\right|^p\right] \\
&\leq 2^{p-1}\left|\psi_{kT+\nu-1}(\nu)\right|^p E\left|\hat{Z}_{(i-k)T+1}\right|^p + 2^{2(p-1)}\left|\psi_{kT+\nu-2}(\nu)\right|^p E\left|\hat{Z}_{(i-k)T+2}\right|^p \\
&\quad + \cdots + 2^{T(p-1)}\left|\psi_{kT+\nu-T}(\nu)\right|^p E|\hat{Z}_{(i-k)T+T}|^p \\
&\leq \kappa^p r^{p(kT+\nu-1)} \max_{1\leq j\leq T} E\left[|\hat{Z}_{(i-k)T+j}|^p\right] \frac{2^{p-1}(2^{T(p-1)}-1)}{2^{p-1}-1} \\
&\leq L^p r^{p(kT+\nu-1)}.
\end{aligned}$$

Here, L is some constant, we have used the bound $|\psi_k| \leq \kappa r^k$, for some $r < 1$ and finite κ. Hence,

$$\left[E|\hat{X}_{iT+\nu}|^p\right]^{\frac{1}{p}} \leq \sum_{k=0}^{\infty}\left[L^p r^{p(kT+\nu-1)}\right]^{\frac{1}{p}} \leq L\frac{r^{(\nu-1)}}{1-r^T} < \infty.$$

For the $(2+\eta)$th moment of the partial sum in each cycle, we will apply Minkowski's integal inequality again and Hólder's inequality

$$E[|XY|] \leq \left(E[|X|^{p_1}]\right)^{\frac{1}{p_1}}\left(E[|Y|^{q_1}]\right)^{\frac{1}{q_1}},$$

where $\frac{1}{p_1} + \frac{1}{q_1} = 1$. Let $p_2 = 2 + \eta$ and $p_1 p_2 = p$. Then by Minkowski's integal inequality

$$\begin{aligned}\left(E\left|\sum_{i=1}^{C} \hat{X}_{iT+\nu}\right|^{p_2}\right)^{\frac{1}{p_2}} &= \left(E\left|\sum_{i=1}^{\infty} \hat{X}_{iT+\nu} 1_{(C>i)}\right|^{p_2}\right)^{\frac{1}{p_2}} \\ &\leq \sum_{i=1}^{\infty}\left(E\left[|\hat{X}_{iT+\nu}|^{p_2} 1_{(C>i)}\right]\right)^{\frac{1}{p_2}} \\ &= \sum_{i=1}^{\infty}\left(E\left[|\hat{X}_{iT+\nu}|^{p_1 p_2}\right]\right)^{\frac{1}{p_1 p_2}} P\left(C>i\right)^{\frac{1}{q_1 p_2}}.\end{aligned}$$

Notice that $\frac{1}{q_1 p_2} < 1$ and recall that $E[|\hat{X}_{iT+\nu}|^p] < \infty$ as proved above. Now $\sum_{i=1}^{\infty} P(C > i)^{\frac{1}{q_1 p_2}} < \infty$, because

$$P(\tau_{\mathcal{B}} > i) \leq \left[1 - \Pi_{\mathbf{\Phi}}(\mathcal{B})\right]^{i-1} \Pi_{\mathbf{\Phi}}(\mathcal{B}),$$

implying

$$\begin{aligned}&\sum_{i=1}^{\infty} P(C>i)^{\frac{1}{q_1 p_2}} \\ &= \sum_{i=1}^{\infty} P(\tau_{\mathcal{B}} > i)^{\frac{1}{q_1 p_2}} \\ &= \sum_{i=1}^{\infty}\left[(1 - \Pi_{\mathbf{\Phi}}(\mathcal{B}))^{i-1} \Pi_{\mathbf{\Phi}}(\mathcal{B})\right]^{\frac{1}{q_1 p_2}} < \infty,\end{aligned}$$

Hence, $E[|\hat{S}_{1,\nu}|^{2+\eta}] < \infty$ as required.

□

3.3 Asymptotic Independence

We now construct approximating sample sums and maxima from $\{\hat{\mathbf{X}}_t\}$ above. Set

$$\hat{\mathbf{S}}_k = (\hat{S}_{k,1}, \cdots, \hat{S}_{k,T}) = \left(\sum_{i \in \mathcal{C}_k} \hat{X}_{iT+1}, \cdots, \sum_{i \in \mathcal{C}_k} \hat{X}_{iT+T}\right)$$

and

$$\hat{\mathbf{M}}_k = (\hat{M}_{k,1}, \cdots, \hat{M}_{k,T}) = \left(\max_{i \in \mathcal{C}_k} \hat{X}_{iT+1}, \cdots, \max_{i \in \mathcal{C}_k} \hat{X}_{iT+T} \right).$$

Now fix a threshold $\mathbf{d} \in R^T$ and let

$$B = \{\mathbf{x} \in R^T | \ \ x_j > d_j, \ \ \text{for some} \ \ j = 1, \cdots, T\}.$$

Set $k^+(0) = k^-(0) = 0$ and for $i = 1, 2, \cdots$, set

$$k^+(i) = \inf\{k > k^+(i-1); \hat{\mathbf{M}}_k \in B\},$$

$$k^-(i) = \inf\{k > k^-(i-1); \hat{\mathbf{M}}_k \notin B\},$$

$$N_{\mathbf{d}}(n) = \sup\{i; k^+(i) \leq n\}.$$

With this setup, we can extract the asymptotic independence between maximums and sample sums from the original series. Before stating our main theorem, we need a Lemma.

Lemma 3 Under the above construction, if $E[\hat{S}^2_{k^+(i)\nu}] \to 0$ for all $\nu \in \{1, 2, \cdots, T\}$ as $d \to \infty$, then $\frac{1}{\sqrt{n}} \sum_{k=1}^{N(n)} \hat{\mathbf{S}}_k \xrightarrow{D} N(\mathbf{0}, \Sigma)$ as $n \to \infty$. Here, $N(\mathbf{0}, \Sigma)$ denotes the marginal distribution of partial sums of $\{\mathbf{X}_i\}$.

Proof. For the sequence $\{\hat{\mathbf{S}}_{k^-(i)}\}$, by the central limit theorem, we know that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N(n)-N_d(n)} \left(\hat{\mathbf{S}}_{k^-(i)} - E[\hat{\mathbf{S}}_{k^-(i)}] \right) \xrightarrow{D} N(\mathbf{0}, \Sigma_d),$$

since $\frac{N_{d(n)}}{n} = \frac{N_{d(n)}}{N(n)} \frac{N(n)}{n} \to \frac{P(\hat{\mathbf{M}}_1 \in B)}{E(C_1)}$ by the renewal theorem. Here, Σ_d is the covariance matrix of $\{\hat{\mathbf{S}}_{k^-(i)}\}$. Also observe that $N(\mathbf{0}, \Sigma_d)$ converges in distribution to $N(\mathbf{0}, \Sigma)$ as $\mathbf{d} \to \infty$.

By Theorem 4.2 in Billingsley [7], we need to show that

$$\lim_{n\to\infty}\lim_{d\to\infty} P\left(\left|\frac{1}{\sqrt{n}}\left(\sum_{k=1}^{N(n)}\hat{\mathbf{S}}_k - \sum_{i=1}^{N(n)-N_d(n)}(\hat{\mathbf{S}}_{k^-(i)} - E[\hat{\mathbf{S}}_{k^-(i)}])\right)\right| > \epsilon\right) = 0.$$

Since

$$P\left(\left|\frac{1}{\sqrt{n}}\left(\sum_{k=1}^{N(n)}\hat{\mathbf{S}}_k - \sum_{i=1}^{N(n)-N_d(n)}(\hat{\mathbf{S}}_{k^-(i)} - E[\hat{\mathbf{S}}_{k^-(i)}])\right)\right| > \epsilon\right)$$
$$= P\left(\left|\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{N_d(n)}(\hat{\mathbf{S}}_{k^+(i)} - E[\hat{\mathbf{S}}_{k^+(i)}])\right)\right| > \epsilon\right),$$

and

$$P(|\mathbf{X}| > \epsilon) := P(\|\mathbf{X}\|_2 > \epsilon) = P\left(\sqrt{X_1^2 + \cdots + X_T^2} > \epsilon\right)$$
$$\leq P\left[\cup_{\nu=1}^{T}\left(X_\nu^2 > \frac{\epsilon^2}{T}\right)\right] \leq \sum_{\nu=1}^{T} P\left(X_\nu^2 > \frac{\epsilon^2}{T}\right),$$

we have

$$P\left(\left|\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{N_d(n)}(\hat{\mathbf{S}}_{k^+(i)} - E[\hat{\mathbf{S}}_{k^+(i)}])\right)\right| > \epsilon\right)$$
$$\leq \sum_{\nu=1}^{T} P\left(\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{N_d(n)}(\hat{S}_{k^+(i)\nu} - E[\hat{S}_{k^+(i)\nu}])\right]^2 > \frac{\epsilon^2}{T}\right)$$
$$= \sum_{\nu=1}^{T} P\left(\sum_{k=1}^{N_d(n)}\left|\hat{S}_{k^+(i)\nu} - E[\hat{S}_{k^+(i)\nu}]\right| > \frac{\sqrt{n}\epsilon}{\sqrt{T}}\right)$$

For any $\nu = 1, 2, \cdots, T$,

$$
\begin{aligned}
& P\left(\sum_{k=1}^{N_d(n)} \left|\hat{S}_{k^+(i)\nu} - E[\hat{S}_{k^+(i)\nu}]\right| > \frac{\sqrt{n}\epsilon}{\sqrt{T}}\right) \\
& \leq \frac{T}{n\epsilon^2} E\left[\sum_{k=1}^{N_d(n)} (\hat{S}_{k^+(i)\nu} - E[\hat{S}_{k^+(i)\nu}])\right]^2 \\
& = \frac{T}{n\epsilon^2} \mathrm{var}\left(\sum_{k=1}^{N_d(n)} \hat{S}_{k^+(i)\nu}\right) \\
& = \frac{T}{n\epsilon^2}\left(\mathrm{var}(\hat{S}_{k^+(i)\nu}) E[N_d(n)] + \mathrm{var}(N_d(n)) E[\hat{S}_{k^+(i)\nu}]^2\right).
\end{aligned}
$$

Since $\frac{N_d(n)}{n} = \frac{N_d(n)}{N(n)} \frac{N(n)}{n} \to \frac{P(\hat{\mathbf{M}}_1 \in B)}{E(C_1)}$, $\frac{E(N_d(n))}{n} \to 0$ as $d \to \infty$ and $n \to \infty$.

By Theorem 4.2 in Billingsley [7], we get the result:

$$
\frac{1}{\sqrt{n}} \sum_{k=1}^{N(n)} \hat{\mathbf{S}}_k \xrightarrow{D} N(\mathbf{0}, \Sigma)
$$

as $n \to \infty$. □

Theorem 1 Suppose that $\{\mathbf{X}_t\}$ is a PARMA series with $E[Z_\nu^\eta] < \infty$ for some $\eta > 2$, for all $\nu \in \{1, 2, \cdots, T\}$. Also suppose that for every kth cycle and some $\delta > 0$,

$$
E\left[\left(\sup_{s \in \mathcal{C}_k} \left|\sum_{i=C_k+1}^{s} \hat{\mathbf{X}}_i\right|\right)^{2+\delta}\right] < \infty.
$$

Then as $n \to \infty$,

$$
\left(\frac{\max_{1 \leq k \leq n} M_k - b_{n1}}{a_{n1}}, \cdots, \frac{\max_{1 \leq k \leq n} M_k - b_{nT}}{a_{nT}}, \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbf{S}_k\right) \xrightarrow{D} (G(\cdot), N(0, \Sigma)).
$$

Here, $G(\cdot)$ represents the distribution of the limiting extremes of the T seasons.

Proof. For $\mathbf{x} > \mathbf{d}$,

$$\begin{aligned}
&nP(\cup_{\nu=1}^{T} a_{n\nu}^{-1}(\hat{M}_{k^+(1)\nu} - b_{n\nu}) > x_\nu) \\
&= nP(\cup_{\nu=1}^{T} a_{n\nu}^{-1}(\hat{M}_{1\nu} - b_{n\nu}) > x_\nu)/P(\hat{\mathbf{M}}_1 \in B) \\
&\to \mu((-\infty, \mathbf{x}]^c)/P(\hat{\mathbf{M}}_1 \in B),
\end{aligned}$$

where μ is chosen so that

$$G(\mathbf{x}) = \exp\{-\mu((-\infty, \mathbf{x}]^c)\},$$

that is, G is the extreme value limit distribution for the original series. Since

$$P\left(\cup_{\nu=1}^{T} \frac{\max_{i=0}^{N(n)+1} \hat{M}_{k^+(i)\nu} - b_{n\nu}}{a_{n\nu}} \le x_\nu\right) = \exp\{-\mu((-\infty, \mathbf{x}]^c)/P(\hat{\mathbf{M}}_1 \in B)\},$$

and $\{N_{\mathbf{d}}(n), n \ge 1\}$ is a renewal function, $N_{\mathbf{d}}(n)/N(n) \to P(\hat{\mathbf{M}}_1 \in B)$ almost surely in R^T. Hence,

$$\begin{aligned}
&\left(\frac{\max_{0\le i\le N(n)+1} \hat{M}_{k^+(i)1} - b_{n1}}{a_{n1}}, \cdots, \frac{\max_{0\le i\le N(n)+1} \hat{M}_{k^+(i)T} - b_{nT}}{a_{nT}}, N_{\mathbf{d}}(n)/N(n)\right) \\
&\qquad \xrightarrow{D} (G(\cdot)/P(\hat{\mathbf{M}}_1 \in B), P(\hat{\mathbf{M}}_1 \in B)),
\end{aligned}$$

composing two components gives

$$\left(\frac{\max_{1\le i\le N_{\mathbf{d}}(n)} \hat{M}_{k^+(i)1} - b_{n1}}{a_{n1}}, \cdots, \frac{\max_{1\le i\le N_{\mathbf{d}}(n)} \hat{M}_{k^+(i)T} - b_{nT}}{a_{nT}}\right) \xrightarrow{D} G(\cdot).$$

The Découpage de Lévy Theorem gives $\{N_{\mathbf{d}}(n), n \ge 1\}$, $\{\hat{\mathbf{S}}_{k^-(i)}\}$ and $\{\hat{\mathbf{M}}_{k^+(i)}\}$ are independent. And by Lemma 3, we have

$$\left(\frac{\max_{1\le i\le N_{\mathbf{d}}(n)} \hat{M}_{k^+(i)1} - b_{n1}}{a_{n1}}, \cdots, \frac{\max_{1\le i\le N_{\mathbf{d}}(n)} \hat{M}_{k^+(i)T} - b_{nT}}{a_{nT}}, \frac{1}{\sqrt{n}} \sum_{k=1}^{N(n)} \hat{\mathbf{S}}_k\right) \xrightarrow{D} (G(\cdot), N(0, \Sigma)).$$

Since $E\left[\sup_{s\in\mathcal{C}_k} \left|\sum_{i=C_k+1}^{s} \hat{\mathbf{X}}_i\right|\right]^{2+\delta} < \infty$, by Theorem 2.3.8 in Whitt [41], we have

$\sum_{k=0}^{N(n)+1} \hat{\mathbf{S}}_k \to N(\mathbf{0}, \Sigma)$. By construction of $\{\hat{\mathbf{X}}_n\}$, $\lim_{m\to\infty} |\hat{\mathbf{X}}_n - \mathbf{X}_n| \to 0$ and

$$\max_{1\le k\le n} \mathbf{M}_k = \max_{0\le i\le N(n)+1} \hat{\mathbf{M}}_{k^+(i)} = \max_{1\le i\le N_{\mathbf{d}}(n)} \hat{\mathbf{M}}_{k^+(i)}.$$

Thus,

$$\lim_{n\to\infty} \left| \sum_{k=1}^{n} \mathbf{S}_k - \sum_{k=0}^{N(n)+1} \hat{\mathbf{S}}_{\mathbf{k}} \right| = 0.$$

Therefore, we obtain the limiting joint distribution

$$\left(\frac{\max_{1\le k\le n} M_k - b_{n1}}{a_{n1}}, \cdots, \frac{\max_{1\le k\le n} M_k - b_{nT}}{a_{nT}}, \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbf{S}_k \right) \xrightarrow{D} (G(\cdot), N(0, \Sigma)) .$$

□

3.4 Remarks

Combining the ϵ-coupling method, the regerative processes are applied to our series such that it can be divided into IID cycles. The point process technique gives us the idea that $'$large$'$ points and $'$small$'$ points should be independent. The Découpage de Lévy Theorem gives techinical work about the asymptotic independence between maximums and partial sums under some conditions.

Chapter 4

Conclusions and Discussion

4.1 Conclusions

Many environmental time series have seasonal structures, including means, variances, and autocorrelations. We used these features to help identify the truly rare events that happened. This seemed to work better than threshold modeling techniques that adopted a time-constant threshold and examined the raw times and magnitudes of the threshold exceedences.

Our approach identified the rare events of periodic time series based on residuals. The work used a PARMA model to construct a set of one-step-ahead prediction residuals that gauged how extreme each new observation is relative to the series' past. Although the true residuals can not be recovered from the raw series data, we prove that the estimated residuals can be used in place of the true residuals without asymptotic loss of precision. Essentially, the PARMA residuals justify the application of classical peaks over threshold method. The residuals serve to set a seasonal threshold and decluster runs of extremes in the raw time series. Another overarching theme of our approach is that a seasonal POT model works well with periodic independent residuals. A seasonal POT model is fitted from estimated residuals with a non-seasonal threshold. This fitted seasonal POT model has a constant shape parameter but a seasonal scale parameter. The seasonal POT model is

combined with the tail of residuals at each season to compute an overall probability (rarity). Here, rare events are judged by the overall probability of exceedence. The residual approach meshes well with a seasonal POT paradigm and this proposed strategy is also useful in the stationary setting.

For the asymptotic independence between the sum and the maxima for a periodic time series, the approach here used a regenerative stochastic process to model the periodic time series. This is accomplished by showing that the noise sequence can be approximated by a Markov chain. The time series is a function of the Markov chain and hence can be viewed as a regenerative process. This allows the time series to be partitioned into IID cycles. The maximum of the time series is the maximum of the maximum in each cycle; the sum is the sum of each cycle. Hence, the problem is essentially reduced to showing that the maximum and the sum are independent when one has IID random variables. The proof is inspired by point process techniques, specifically from Poisson process, we consider the maximums as the successes of trials, they should be independent of the failures. Here, the central limit theorem says the maximum has no role to play. Moreover, the idea of Découpage de Lévy yields the asymptotic independence between maximums and partial sums.

4.2 Future Work

Extreme problems in periodic time series produce interesting and broad questions. We consider the heavy-tailed cases. Here the distribution for each cycle is heavy-tailed with the same index as the errors, but with different error constants. Point process techniques can still be used here. We suspect that the sum is independent of maximums for sequences in this case. We will also try to identify the specific joint limit distribution for maximums and partial sums, and elaborate on the constants involved. Statistical inference will also be considered in the near future.

Appendices

Appendix A Proof of Theorem 3.1

This appendix proves Theorem 3.1. A proof for a PAR series is not difficult to construct; the general PARMA case presented here entails considerably more work.

To aid calculations, we introduce a Banach algebra $\Im$ that contains $T \times T$ matrices whose individual elements are complex-valued functions defined on the complex unit circle $\{z : |z| = 1\}$. Specifically, for $F = \{F_{i,j}\}_{i,j=1}^{T}$, we say that $F \in \Im$ if

$$\|F\|_{\Im} := \sum_{i=1}^{T} \sum_{j=1}^{T} \|F_{i,j}\|_{W_+} < \infty.$$

Here, the W_+ norm on the components is $||f||_{W_+} = \sum_{n=0}^{\infty} |a_n|$ for any complex-valued function f of form

$$f(e^{i\alpha}) = \sum_{n=0}^{\infty} a_n e^{in\alpha} \tag{1.1}$$

with $\sum_{n=0}^{\infty} |a_n| < \infty$. In fact, the set of all complex-valued functions on the complex unit disc satisfying (1.1) with norm W_+ is a unital commutative Banach algebra with pointwise multiplication (MacCluer, 2009, page 152). Moreover, the invertible elements of W_+ are precisely those f for which $\sum_{n=0}^{\infty} a_n \lambda^n \neq 0$ for all λ in the closed unit disc $\{z : |z| \leq 1\}$.

To write the PARMA model in vector form, let $\vec{X}_n = (X_{nT+1}, \ldots, X_{nT+T})'$ and $\vec{Z}_n = (Z_{nT+1}, \ldots, Z_{nT+T})'$ denote the observation and error vectors for the nth cycle. The vector autoregressive moving-average (VARMA) representation of the PARMA model is

$$\Phi_0 \vec{X}_n - \sum_{k=1}^{p^*} \Phi_k \vec{X}_{n-k} = \Theta_0 \vec{Z}_n + \sum_{k=1}^{q^*} \Theta_k \vec{Z}_{n-k}.$$

The VARMA orders p^* and q^* and $T \times T$ matrix coefficients $\Phi_0, \ldots, \Phi_{p^*}; \Theta_0, \ldots, \Theta_{q^*}$ are listed in Vecchia (1985) and Basawa and Lund (2001). Their form is not overly important here; however, we note that each entry of the matrix VARMA coefficients is either zero, one, or a $\phi_k(\nu)$ or $\theta_k(\nu)$ coefficient.

The VARMA autoregressive and moving-average polynomials are

$$\Phi(z) = \Phi_0 - \sum_{k=1}^{p^*} \Phi_k z^k; \qquad \Theta(z) = \Theta_0 + \sum_{k=1}^{q^*} \Theta_k z^k.$$

Causality and invertibility imply that Φ and Θ belong to $\Im$. Standard VARMA theory affords the representation $\vec{Z}_n = \sum_{k=0}^{\infty} \Pi_k \vec{X}_{n-k}$, where $\Pi(z) = \sum_{k=0}^{\infty} \Pi_k z^k = \Theta^{-1}(z)\Phi(z)$. Since the model is invertible, we have $\Theta^{-1} \in \Im$ and, hence, $\Pi \in \Im$.

We have $\sqrt{d}$-asymptotically normal (and hence consistent) estimators $\hat{\phi}_k(\nu)$ and $\hat{\theta}_k(\nu)$ for each season ν and autoregressive and moving-average coefficient. By PARMA asymptotics, we know that these estimators can be chosen so that the model is causal and invertible for every d. Basawa and Lund 2001 and Proposition 4.4.2 of Brockwell and Davis (1991) elaborate on this aspect. We also know that $\hat{\sigma}^2(\nu) \xrightarrow{\mathcal{P}} \sigma^2(\nu)$ for each season ν. Now construct the $T \times T$ matrices $\hat{\Phi}_k$, $k = 0, \ldots, p^*$ and $\hat{\Theta}_k$, $k = 0, \ldots, q^*$ by replacing all quantities by their estimators. The estimated VARMA polynomials are

$$\hat{\Phi}(z) = \hat{\Phi}_0 - \sum_{k=1}^{p^*} \hat{\Phi}_k z^k; \qquad \hat{\Theta}(z) = \hat{\Theta}_0 + \sum_{k=1}^{q^*} \hat{\Theta}_k z^k.$$

Define $\hat{\Pi}(z) = \hat{\Theta}^{-1}(z)\hat{\Phi}(z)$. Since $\hat{\Theta}^{-1}(z)$ is invertible, we can write $\hat{\Pi}(z) = \sum_{n=0}^{\infty} \hat{\Pi}_n z^n$. We now state and prove the following convergence in probability.

Lemma: Under the setup above, $\|\hat{\Pi} - \Pi\|_{\Im} = O_{\mathcal{P}}(d^{-\frac{1}{2}})$, as $d \to \infty$. Hence, for all i, j, ν,

$$\sum_{k=0}^{\infty} |(\hat{\Pi}_k)_{ij} - (\Pi_k)_{ij}| = O_{\mathcal{P}}(d^{-\frac{1}{2}}), \qquad \sum_{k=0}^{\infty} |\hat{\pi}_k(\nu) - \pi_k(\nu)| = O_{\mathcal{P}}(d^{-\frac{1}{2}}).$$

Proof of Lemma: By asymptotic normality of the PARMA coefficients, for each $\epsilon > 0$, there is a finite M such that

$$P(\sqrt{d} \max(||\hat{\Phi} - \Phi||_{\Im}, ||\hat{\Theta} - \Theta||_{\Im}) > M) < \epsilon,$$

for all $d \geq 1$. So fix $\epsilon > 0$, let M be as above, and choose n_0 so that $M \leq \sqrt{d}/[2||\Theta^{-1}||_{\Im}]$, for

all $d \geq n_0$. Let $A_d = \{\sqrt{d}\max(||\hat{\Phi} - \Phi||_{\Im}, ||\hat{\Theta} - \Theta||_{\Im}) \leq M\}$. Then for $d \geq n_0$ and $\omega \in A_d$ (ω here refers to the underlying probability space), $||\hat{\Theta} - \Theta||_{\Im} \leq Md^{-1/2} \leq [2||\Theta^{-1}||_{\Im}]^{-1}$. Appealing to Lax (2002) gives, for $d \geq n_0$,

$$\begin{aligned}
||\hat{\Theta}^{-1}||_{\Im} &\leq \|\sum_{k=0}^{\infty}[\Theta^{-1}(\Theta - \hat{\Theta})]^k\Theta^{-1}\|_{\Im} \\
&\leq \sum_{k=0}^{\infty}\|(\Theta^{-1}[\Theta - \hat{\Theta}])^k\Theta^{-1}\|_{\Im} \\
&\leq \sum_{k=0}^{\infty}(\|\Theta^{-1}\|_{\Im}\|\Theta - \hat{\Theta}\|_{\Im})^k\|\Theta^{-1}\|_{\Im} \\
&\leq \sum_{k=0}^{\infty}\left(\|\Theta^{-1}\|_{\Im}\frac{1}{2\|\Theta^{-1}\|_{\Im}}\right)^k\|\Theta^{-1}\|_{\Im} \\
&= 2\|\Theta^{-1}\|_{\Im}.
\end{aligned}$$

The bound above allows us to handle $\hat{\Theta}$ through the fixed quantity Θ and is a consequence of the introduced Banach algebra.

For $\omega \in A_d$,

$$\begin{aligned}
||\hat{\Pi} - \Pi||_{\Im} &= \|\hat{\Theta}^{-1}\hat{\Phi} - \hat{\Theta}^{-1}\Phi + \hat{\Theta}^{-1}\Phi - \Theta^{-1}\Phi\|_{\Im} \\
&\leq \|\hat{\Theta}^{-1}\hat{\Phi} - \hat{\Theta}^{-1}\Phi\|_{\Im} + \|\hat{\Theta}^{-1}\Phi - \Theta^{-1}\Phi\|_{\Im} \\
&\leq \|\hat{\Theta}^{-1}\|_{\Im}\|\hat{\Phi} - \Phi\|_{\Im} + \|\hat{\Theta}^{-1} - \Theta^{-1}\|_{\Im}\|\Phi\|_{\Im} \\
&\leq 2\|\Theta^{-1}\|_{\Im}\|\hat{\Phi} - \Phi\|_{\Im} + \|\hat{\Theta}^{-1} - \Theta^{-1}\|_{\Im}\|\Phi\|_{\Im}. \qquad (1.2)
\end{aligned}$$

But for $\omega \in A_d$ and $d \geq n_0$, we have from (1.2) that

$$\|\hat{\Theta}^{-1} - \Theta^{-1}\|_{\Im} = \|\Theta^{-1}(\Theta - \hat{\Theta})\hat{\Theta}^{-1}\|_{\Im} \leq 2\|\Theta^{-1}\|_{\Im}^2\frac{M}{\sqrt{d}}. \qquad (1.3)$$

Thus,

$$\sqrt{d}||\hat{\Pi} - \Pi||_{\Im} \leq 2M||\Theta^{-1}||_{\Im} + 2M\|\Theta^{-1}\|_{\Im}^2\|\Phi\|_{\Im} := M^*,$$

for all $d \geq n_0$ and $\omega \in A_d$. Hence, $P(\sqrt{d}||\hat{\Pi} - \Pi||_{\Im} < M^*) \geq P(A_d) \geq 1 - \epsilon$ for all $d \geq n_0$; hence, $||\hat{\Pi} - \Pi||_{\Im}$ is $O_{\mathcal{P}}(d^{-1/2})$. □

Proof of Theorem 3.1: Observe that

$$\max_{\ell(d)T<t\leq dT} |R_t - \hat{R}_t| \leq \max_{\ell(d)\leq n\leq d-1} \max_{1\leq\nu\leq T} \left| \frac{\hat{Z}_{nT+\nu}}{\hat{\sigma}(\nu)} - \frac{Z_{nT+\nu}}{\sigma(\nu)} \right|.$$

For a fixed season ν,

$$\begin{aligned}
\max_{\ell(d)\leq n\leq d-1} \left| \frac{\hat{Z}_{nT+\nu}}{\hat{\sigma}(\nu)} - \frac{Z_{nT+\nu}}{\sigma(\nu)} \right| &= \max_{\ell(d)\leq n\leq d-1} \left| \frac{Z_{nT+\nu} - \hat{Z}_{nT+\nu}}{\hat{\sigma}(\nu)} - Z_{nT+\nu}\left(\frac{1}{\hat{\sigma}(\nu)} - \frac{1}{\sigma(\nu)}\right) \right| \\
&\leq \max_{\ell(d)\leq n\leq d-1} \left| \frac{Z_{nT+\nu} - \hat{Z}_{nT+\nu}}{\hat{\sigma}(\nu)} \right| \\
&+ \max_{\ell(d)\leq n\leq d-1} |Z_{nT+\nu}| \left| \frac{1}{\hat{\sigma}(\nu)} - \frac{1}{\sigma(\nu)} \right|. \qquad (1.4)
\end{aligned}$$

Hence, our work is done if we show that the two bottom terms in (1.4) converge to zero in probability for each season ν.

We first consider the quantity in the middle line of (1.4). Since $\hat{\sigma}^2(\nu) \xrightarrow{\mathcal{P}} \sigma^2(\nu) > 0$, we need only show that $\max_{\ell(d)\leq n\leq d-1} |Z_{nT+\nu} - \hat{Z}_{nT+\nu}| = o_{\mathcal{P}}(1)$ for each season ν. Standard VARMA theory provides

$$\hat{Z}_{nT+\nu} = \sum_{k=0}^{nT+\nu-1} \hat{\pi}_k(\nu) X_{nT+\nu-k}.$$

Using this and (2.4) gives

$$Z_{nT+\nu} - \hat{Z}_{nT+\nu} = \sum_{k=0}^{nT+\nu-1} [\pi_k(\nu) - \hat{\pi}_k(\nu)] X_{nT+\nu-k} + \sum_{k=nT+\nu}^{\infty} \pi_k(\nu) X_{nT+\nu-k}. \qquad (1.5)$$

Now note that for each $\epsilon > 0$,

$$P\left(\max_{\ell(d)\le n\le d-1}\left|\sum_{k=nT+\nu}^{\infty}\pi_k(\nu)X_{nT+\nu-k}\right| > \epsilon\right) \le P\left(\sum_{n=\ell(d)}^{d-1}\sum_{k=nT+\nu}^{\infty}|\pi_k(\nu)X_{nT+\nu-k}| > \epsilon\right). \tag{1.6}$$

Applying Markov's and Cauchy-Schwarz's inequalities, we bound the right hand side of (1.6) by

$$\frac{\max_{1\le\nu\le T}\gamma_\nu(0)}{\epsilon}\sum_{n=\ell(d)}^{d-1}\sum_{k=nT+\nu}^{\infty}|\pi_k(\nu)|.$$

Now apply $\ell(d)\to\infty$, as $d\to\infty$, and the invertibility implication, $\sum_{k=0}^{\infty}k|\pi_k(\nu)|<\infty$, for all ν, to see that the maximum of the rightmost term in (1.5) over all ν, is $o_{\mathcal{P}}(1)$.

For the first sum on the right hand side of (1.5), we bound it via

$$\max_{\ell(d)\le n\le d-1}\left|\sum_{k=0}^{nT+\nu-1}[\pi_k(\nu)-\hat{\pi}_k(\nu)]X_{nT+\nu-k}\right| \le F_1F_2,$$

where

$$F_1=\sqrt{d}\sum_{k=0}^{dT}\max_{1\le\nu\le T}|\pi_k(\nu)-\hat{\pi}_k(\nu)|; \quad F_2=\frac{\max_{0\le n\le d-1}\ \max_{1\le\nu\le T}|X_{nT+\nu}|}{\sqrt{d}}.$$

By the Lemma, F_1 is bounded in probability; hence, it is sufficient to show that $F_2=o_{\mathcal{P}}(1)$. To do this, let $\epsilon>0$ and note that

$$\begin{aligned}P\left(\max_{1\le\nu\le T}\ \max_{0\le n\le d-1}\frac{|X_{nT+\nu}|}{\sqrt{d}}>\epsilon\right) &= P\left(\max_{1\le\nu\le T}\left(\bigcup_{n=0}^{d-1}(|X_{nT+\nu}|>\sqrt{d}\epsilon)\right)\right)\\ &\le \sum_{\nu=1}^{T}P\left(\bigcup_{n=0}^{d-1}(|X_{nT+\nu}|>\sqrt{d}\epsilon)\right).\end{aligned}$$

Strict stationarity of $\{X_{nT+\nu}\}_{n=0}^{\infty}$ for each fixed ν and Markov's inequality give

$$P\left(\bigcup_{n=0}^{d-1}|X_{nT+\nu}|>\sqrt{d}\epsilon\right)\leq dP(|X_\nu|>\sqrt{d}\epsilon)\leq\frac{E[|X_\nu|^{2+\delta}]}{\epsilon^{2+\delta}d^{\delta/2}}\to 0.$$

Here, we have used that $Z_{nT+\nu}$ having a finite $(2+\delta)$th moment implies that X_ν has a finite $(2+\delta)$th moment. Combining the last two relations shows that $F_2=o_{\mathcal{P}}(1)$.

Finally, write the last term in (1.4) as

$$\max_{1\leq\nu\leq T}\left[\max_{\ell(d)\leq n\leq d-1}\frac{|Z_{nT+\nu}|}{\sqrt{d}}\right]\sqrt{d}\left|\frac{1}{\hat{\sigma}(\nu)}-\frac{1}{\sigma(\nu)}\right|.$$

For each season ν, $\sqrt{d}[\hat{\sigma}(\nu)^{-1}-\sigma(\nu)^{-1}]$ is asymptotically normal and hence tight; that is, it is $O_{\mathcal{P}}(1)$. Thus, it suffices to show that $\max_{\ell(d)\leq n\leq d-1}|Z_{nT+\nu}|d^{-1/2}=o_{\mathcal{P}}(1)$ for all seasons ν. Since $\{Z_{nT+\nu}\}_{n=0}^{\infty}$ are IID with a finite $(2+\delta)$th moment, periodic identical distributions and Markov's inequality finishes our work via

$$P\left(\max_{\ell(d)\leq n\leq d-1}|Z_{nT+\nu}|>\sqrt{d}\epsilon\right)\leq dP\left(|Z_\nu|>\sqrt{d}\epsilon\right)\leq\frac{E[|Z_\nu|^{2+\delta}]}{\epsilon^{2+\delta}d^{\delta/2}}\to 0. \tag{1.7}$$

Bibliography

[1] C. W. Anderson and K. F. Turkman. Sums and maxima of stationary sequences with heavy tailed distributions. *Sankhyā Ser. A*, 57(1):1–10, 1995.

[2] CW Anderson and KF Turkman. The joint limiting distribution of sums and maxima of stationary sequences. *Journal of applied probability*, pages 33–44, 1991.

[3] Paul L Anderson, Yonas Gebeyehu Tesfaye, and Mark M Meerschaert. Fourier-parma models and their application to river flows. *Journal of Hydrologic Engineering*, 12(5):462–472, 2007.

[4] M Antón, A Serrano, ML Cancillo, and JA García. Total ozone and solar erythemal irradiance in southwestern spain: Day-to-day variability and extreme episodes. *Geophysical Research Letters*, 35(20), 2008.

[5] August A Balkema and Laurens De Haan. Residual life time at great age. *The Annals of Probability*, pages 792–804, 1974.

[6] Rocco Ballerini and William P McCormick. Extreme value theory for processes with periodic variances. *Stochastic Models*, 5(1):45–61, 1989.

[7] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[8] Peter Bloomfield, Harry L Hurd, and Robert B Lund. Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, 15(2):127–150, 1994.

[9] Jenkins Box and Gwilym M Jenkins. Reinsel. time series analysis, forecasting and control, 1994.

[10] Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

[11] Te Lin Chow and Jozef L Teugels. The sum and the maximum of iid random variables. In *Proceedings of the 2nd Prague Symposium on Asymptotic Statistics*, pages 81–92, 1978.

[12] Stuart Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[13] Stuart G. Coles and Jonathan A. Tawn. Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B*, 53(2):377–392, 1991.

[14] Stuart G Coles and Jonathan A Tawn. Statistical methods for multivariate extremes: an application to structural design. *Applied Statistics*, pages 1–48, 1994.

[15] Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442, 1990.

[16] Emma F Eastoe and Jonathan A Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45, 2009.

[17] Emma F Eastoe and Jonathan A Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45, 2009.

[18] Paul Embrechts. *Modelling extremal events: for insurance and finance*, volume 33. Springer, 1997.

[19] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:180–190, 4 1928.

[20] Janos Galambos. *The asymptotic theory of extreme order statistics*. Robert E. Krieger Publishing Co. Inc., Melbourne, FL, second edition, 1987.

[21] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):pp. 423–453, 1943.

[22] Hwai-Chung Ho and Tailen Hsing. On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of applied probability*, pages 138–145, 1996.

[23] Hwai-Chung Ho and William P McCormick. Asymptotic distribution of sum and maximum for gaussian processes. *Journal of applied probability*, 36(4):1031–1044, 1999.

[24] Joel Horowitz. Extreme values from a nonstationary stochastic process: an application to air quality analysis. *Technometrics*, 22(4):469–478, 1980.

[25] Tailen Hsing. A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *The Annals of Probability*, 23(2):938–947, 1995.

[26] Barry James, Kang James, and Yongcheng Qi. Limit distribution of the sum and maximum from multivariate Gaussian sequences. *J. Multivariate Anal.*, 98(3):517–532, 2007.

[27] Arthur F Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, 1955.

[28] G Koch, H Wernli, C Schwierz, J Staehelin, and T Peter. A composite study on the structure and formation of ozone miniholes and minihighs over central europe. *Geophysical research letters*, 32(12):L12810, 2005.

[29] MR Leadbetter and G Lindgren. H. rootzen 1983: Extremes and related properties of random sequences and series.

[30] Torgny Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc. A Wiley-Interscience Publication, 1992.

[31] Robert Lund, Qin Shao, and Ishwar Basawa. Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics*, 48(1):33–47, 2006.

[32] WP McCormick and Y Qi. Asymptotic distribution for the sum and maximum of gaussian processes. *Journal of applied probability*, pages 958–971, 2000.

[33] Z Peng and Saralees Nadarajah. On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory of Probability & Its Applications*, 47(4):706–709, 2003.

[34] James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

[35] Sidney I Resnick. *Adventures in stochastic processes*. Springer, 1992.

[36] Sidney I Resnick. *Extreme values, regular variation, and point processes*. Springer, 2007.

[37] Richard L Smith. Extreme value theory. *Handbook of applicable mathematics*, 7:437–471, 1990.

[38] Hermann Thorisson. *Coupling, stationarity, and regeneration*. Springer New York, 2000.

[39] AV Vecchia. Periodic autoregressive-moving average (parma) modeling with applications to water resources1. *JAWRA Journal of the American Water Resources Association*, 21(5):721–730, 1985.

[40] Richard Von Mises. La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique*, 1(1), 1936.

[41] Ward Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.