EXTREME VALUE ESTIMATORS FOR VARIOUS NON-NEGATIVE TIME SERIES WITH HEAVY-TAIL INNOVATIONS

by

ANDREW CHRISTOPHER BARTLETT

(Under the direction of William McCormick)

ABSTRACT

Extreme value theory is a branch of statistics that is devoted to studying the phenomena governed by extremely rare events. The modeling of such phenomena are tail dependent, therefore we consider a class of heavy-tail distributions, which are characterized by regular variation in the tails. While many articles have considered regular variation at one endpoint (particularly the left endpoint), the idea of regular variation at both endpoints has not be addressed.

In this dissertation, we propose extreme value estimators for various non-negative time series, where the second or higher moments does not exist and the innovations are positive random variables with regular variation at both the right endpoint infinity and the positive left endpoint. This contrasts with traditional estimators whose asymptotic behavior depends on the central part of the innovation distribution. For certain estimation problems, the presence of heavy tails can provide the setting for exceedingly accurate estimates.

Within each model, we provide estimates for the model parameters with respect to an extreme value criteria. Through the use of regular variation and point processes the limit distributions for the proposed estimators are obtained while weak convergence results for asymptotically independent joint distribution are derived. A simulation study is performed to first assess the small sample size behavior and reliability of our proposed estimates and secondly to compare the performance of our extreme value estimation procedure and that of traditional and alternative estimation procedures.

The main goal of all proposed methods, is to capitalize on the behavior of extreme value estimators over traditional estimators when the regular varying exponent is between zero and two. In this heavy-tail regime, extreme value estimators converge at a rate faster than square root n. If a practitioner can entertain an infinite variance time series model, then methods such as the one proposed in this dissertation should receive consideration and even more so if an infinite mean time series model were deemed to be acceptable.

INDEX WORDS: nonnegative time series, autoregressive processes, autoregressive-moving-average processes, bifurcating autoregressive processes, extreme value estimator, regular variation, point process

Extreme Value Estimators for Various Non-Negative Time Series with Heavy-tail Innovations

by

Andrew Christopher Bartlett

B.S., Coastal Carolina University, 2004

M.S., College of Charleston, 2005

M.S., University of Georgia, 2010

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2013

Extreme Value Estimators for Various Non-Negative Time Series with Heavy-tail Innovations

by

Andrew Christopher Bartlett

Approved:

Major Professor: William McCormick

Committee: Lynne Seymour
T.N. Sriram
Gauri Datta
Cheolwoo Park

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

Dedication

To my wife (Jung Ae), parents (Mom and Moose), brothers (Brian and Tim), grandparents(June and Ken), and the several special friends that made this possible.

Special Dedication

In loving memory of my grandfather Ken Chandler.

Acknowledgments

First and foremost, I express my deepest gratitude to the faculty, for believing in me during my darkest hour; without your understanding and kindness this would never have been possible. Secondly, my sincere thanks goes to my major professor, Dr. William McCormick, for his insightful direction, constant encouragement, and help throughout the entire duration of my dissertation research. I am especially grateful for his patience and support that goes beyond the advisement needed for the completion of my dissertation.

I am also greatly thankful to my committee members, Dr. Lynne Seymour, Dr. T.N Sriram, Dr. Gauri Datta, and Dr. Cheolwoo Park; I truly appreciate your time spent in reading my work, as well as your excellent questions and recommendations that helped to make this a scholarly work and most of all, your friendship over the past five years.

My many thanks go to Professors Olcay Akman, Andrew Incognito, Ray Moye, and Amy Langville for their encouragement, guidance, and help along the way. It is because of them that I became interested in pursuing a Ph.D. degree. Sincere thanks to the nice time spent with my team buddies and gamers; Walker, Stevie, Delaney, and Chris. The time on the tennis courts and Saturday night board games provided me with recreation in the midst of my academic stress.

Finally, special thanks are due to my mom and dad for their endless love and long-term support. It was because of their hardwork and sacrifices that I was able to pursue my dreams and receive a good education. On a side note, I would like to say thanks to Dr. Dew, the chipmunks, tigers, bears, rabbits, and mice for their unconditional love and company. Last but not least, I want to thank my grandparents June and Ken for their love and financial support; I couldn't have done this without them.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction And Literature Review

1.1 Introduction

Extreme Value Theory (EVT) is a branch of statistics that is devoted to studying the phenomena governed by extremely rare events. The modeling and statistics of such phenomena are tail dependent and vary greatly from classical modeling and statistical analysis, which give primacy to central moments, averages, and the normal density which has a light tail and large observations are typically treated as an outlier or even ignored. Traditionally, the uses for extreme value theory has been limited to answering questions relating to the distribution of extremes (e.g., what is the probability that a snowfall exceeding x inches will occur in a given location during a given year?) or the inverse problem of return levels (e.g., what height of a river will be exceeded with probability 1/100 in a given year?). However, during the past 40 years or so, many new techniques have been developed to assist with the problems such as the exceedance over high thresholds, the dependence among extreme events in various types of stochastic processes, and multivariate extremes (e.g, estimate the extreme quantile of the profit-and-loss density, once the density is estimated).

There is a fairly large body of literature concerning the theory, computation and application of non–negative time series models. These models form a prominent class to model the phenomena under study by a dependent process. Over the past four decades or so, a variety of estimation approaches have been adopted for nonnegative time series models. These include; least squares, maximum likelihood, linear programming, and extreme value theory method.

An excellent presentation of the classical theory concerning these models can be found, for example, in Brockwell and Davis (1987). A few practical applications for such autoregressive models with positive innovations are described in Collings (1975), where the sequence in an AR(1) process describes the input process for dams. In Hutton (1990), for example, river flows were studied by such a model. More recent developments have focused on some specialized features of the model, e.g. heavy tail innovations or nonnegativity of the model. An elegant approach to studying heavy tail linear models is to examine the behavior of traditional estimates under conditions leading to non-Gaussian limits. For example, the standard approach for parameter estimation within an autoregressive process of order p (AR(p)) is through the Yule-Walker estimator;

$$\hat{\phi}_{YW} = \frac{\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}, \quad \text{where } \bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{1.1.1}$$

If the AR(1) process has a finite variance, then $\phi = Corr(X_t, X_{t-1})$ and the Yule-Walker estimator is an asymptotically normal random variable with mean ϕ and asymptotic variance $n^{-1}(1-\phi^2)$, i.e.

$$\sqrt{n}(\hat{\phi}_{YW} - \phi) \Rightarrow N(0, 1-\phi^2).$$

Davis and Resnick (1986) presented a slightly different approach where they established the weak limit behavior for the sample autocorrelation function under an assumption that the innovations have a regularly varying tail with index α and are tailed balanced where α ranges from greater than 0 to at most 4. However, the analytic difficulties for the Davis-Resnick estimates are further complicated by the fact that the limiting distribution involves certain stable distributions whose exact parameter values are not completely explicit.

On a similar note, Anderson et al. (2008) take an approach analogous to Davis and Resnick (1986) but utilize the innovations algorithm applied to periodically stationary time series and in particular to PARMA models, i.e. periodic ARMA models. A least absolute deviation estimation and other related methods are carried out in Calder and Davis (1998) while a weighted least squares method to estimate the parameters of a heavy tailed ARMA model is presented in Markov (2009).

An interesting approach worth mentioning that was unexpected in the time series setting, is the maximum likelihood, since the likelihood function is generally particularly intractable and intricate. In this case finding the MLE amounts to solving a constrained maximization problem with linear constraints.

With these considerations in mind, Feigin and Resnick (1994) developed linear programming estimates for AR(p) processes with nonnegative innovations having 0 as its left endpoint and satisfying one of two types of regular variation property on the innovation distribution.

McCormick and Mathew (1993) developed consistent estimators with a linear programming estimate for not only the autocorrelation coefficient ϕ but for an unknown location parameter θ under certain optimization constraints. Feigin et al. (1996) continued the study of linear programming estimate for the case of a nonnegative moving average process. The technique has also been applied to positive nonlinear time series in Brown et al (1996) and Datta et al (1998).

In addition to their relation to linear programming estimation procedures, nonnegative time series have also been considered in Anděl (1989), Anděl (1991), and Datta and McCormick (1995). Modeling issues in connection to causal nonnegative time series are addressed in Tsai and Chan (2006). As remarked in Calder and Davis (1998), second-order based estimation methods for the ARMA model parameters perform well when the innovations are heavy-tailed. Our estimation procedure is more in keeping with the second-order estimates developed in Davis and Resnick (1986) with which we make a comparison through a simulation study.

A similar but relatively unknown model in the time series family is called a bifurcating autoregressive process (BAR, for short). It was first introduced by Cowan and Staudte (1986) for analyzing cell lineage data, where each individual in one generation gives rise to two offspring in the next generation. Over the past three decade's there has been many extensions of bifurcating autoregressive models. For example, the driven noise $(\epsilon_{2t}, \epsilon_{2t+1})$ was originally assumed to be independent and identically distributed normal random variables. However, in the past decade or so, model's that allow for correlation between sisters cells and cousin's cells have been studied.

Interestingly enough, there is only one paper in the literature by Zhang (2011) which applies a point process technique to a first-order bifurcating process.

In this dissertation, we focus squarely on the use of extreme value theory to develop alternative estimators for a variety of time series models. The proposed estimation procedure relies on the criteria that the innovations follow a heavy–tail distribution. More specifically, the innovations are positive random variables with regular variation at both the right endpoint (infinity), and the positive left endpoint θ. Our extreme value methods enjoy several benefits over traditional estimation methods such as Least-squares or linear programming method when the regular varying exponent, β is $0 < \beta < 2$. For example, our estimation procedure is especially easy to implement. That is, our estimators are nothing more than a ratio of two sample values, chosen respectively to specify an extreme value criteria. Furthermore, our extreme value method does not rely on a finite second or higher moment to exist. Instead we capitalize on the behavior of extreme value estimators, which in a heavy-tail regime, converge at a faster rate than the square root n. Thus, for a time series model with an infinite variance, our method becomes a very attractive alternative. Thereby providing a competitive alternative to traditional procedures in the literature. Finally, the limiting distribution for our estimators are explicit and tractable, whereas the limiting distribution for traditional methods such as Least-square or maximum likelihood are typically complex and unpractical. The derivation for some of our limiting distributions relies heavily upon the use of point processes, where the essential idea is to first establish the convergence of a sequence of point processes based on simple quantities and then apply the continuous mapping theorem to obtain convergence of the desired statistics.

1.2 Review of Mathematical Tools in Extreme Value Theory

In the previous section we discussed the advances and expansions of a number of established statistical estimation procedures such as; Least–squares, Maximum Likelihood, and Linear programming for various time series models. In this section, we present a brief review of the mathematical building blocks used with or in the development of our extreme value estimators. This includes heavy–tail distributions, weak convergence, regular variation, and point processes. Among many excellent books on this subject, Resnick (1987) gives a survey of the mathematical, probabilistic, and statistical tools used in extreme value theory, Resnick (2007) on Heavy-Tail Phenomena gives a comprehensive survey of theory as well as applications, and Finkenstadt (2004) provides an in depth applications in Environment, insurance, and finance.

1.2.1 A Conceptual Understanding of Extreme Value Estimators

At first, Extreme Value Theory was nothing more than a mere intellectual enthusiasm, but over the past three decades the subject has flourished in applications for many real situations. Additionally, the modeling and statistics of extreme events are tail dependent and differ from classical modeling and statistical analysis, which rely on central moments, averages, and the normal density. For example, consider a first–order autoregressive process

$$X_t = \phi X_{t-1} + Z_t, \tag{1.2.1}$$

where the innovations are nonnegative random variables having a distribution function F for which $\bar{F} = 1 - F$ is regularly varying at infinity with index $-\beta$. The motivation for an extreme value estimator of ϕ is based on the positivity of the innovations and the fact when the observation X_{t-1} is large, equation (1.2.1) implies

$$0 \leq \phi \leq X_t / X_{t-1}. \tag{1.2.2}$$

Therefore, by minimizing the ratio in (1.2.2) we expect $\hat{\phi}_n = \bigwedge_{t=1}^{n} \frac{X_t}{X_{t-1}}$, where $x \wedge y = \min(x, y)$ to be a reasonably good estimator of ϕ. The actual precision of our extreme value estimator depends directly upon the index of regular variation β, which controls the tail behavior of the innovation distribution. Thus, when the regular varying index is small (less than 1) imagine a sample path which remains relatively constant except for a few random spikes at different times. From this illustration presented in Figures 1.1 and 1.2 we can draw upon two key observations; first, the type of outcome expected from a process where the innovations have a heavy–tail distribution and secondly, how our estimators capitalizes on large innovations.

Figure 1.1: Sample Path for AR(1) process with $\beta = .8$

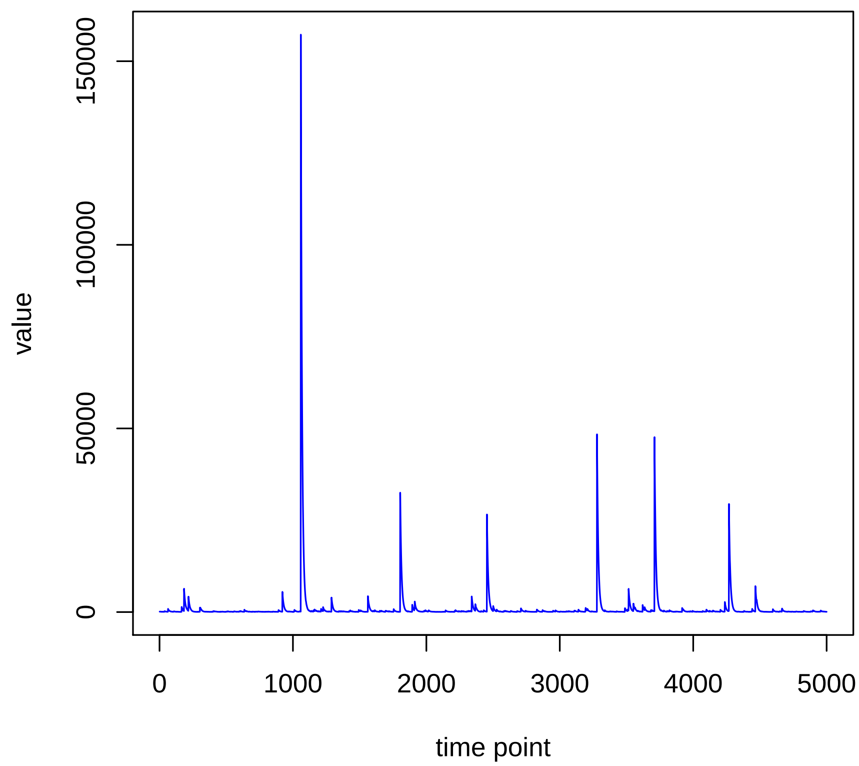
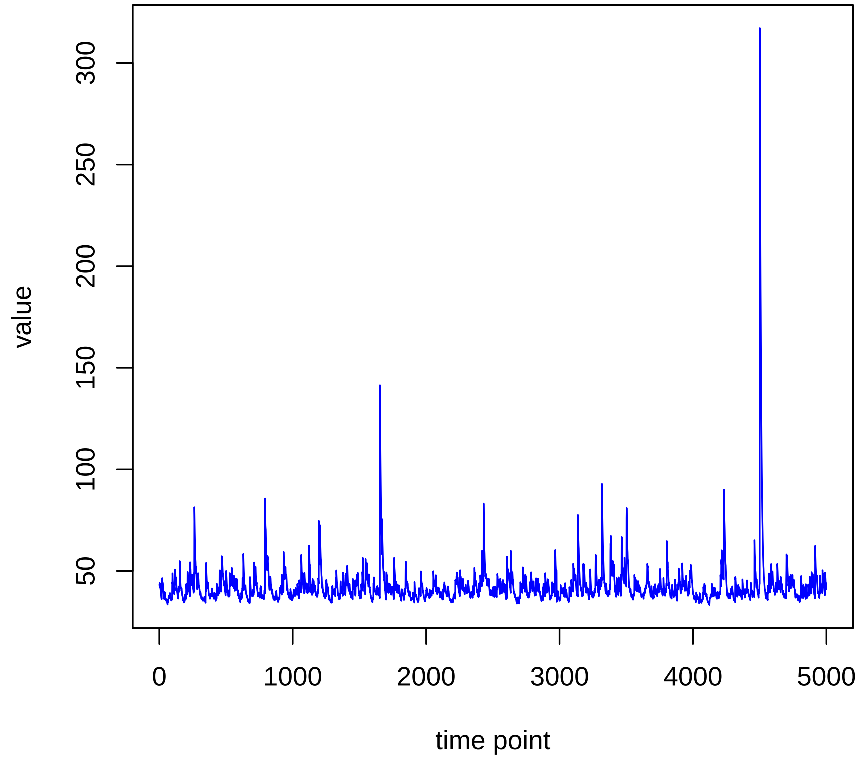


Figure 1.2: Sample Path for AR(1) process with $\beta = 1.5$



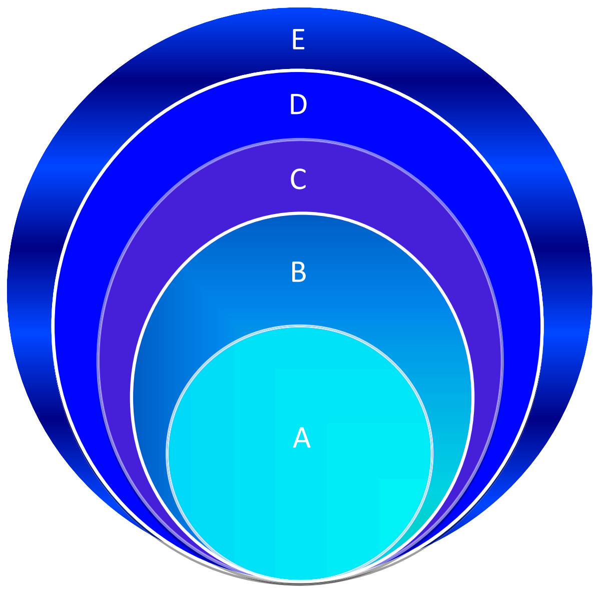
1.2.2 Heavy–tails: The Underlying Component of Extreme Value Theory

Throughout the literature, there does not seem to be a general rule to classify a distribution as heavy–tailed. The difficulty lies in the name, that is, the word "heavy" is not well-defined, thus a precise definition is not easy. For example, sometimes heavy–tailed distributions are called "fat–tailed", "thick–tailed", or even "long–tailed". A common used definition of heavy–tailness is based on the fourth central moment. If X is a random variable with mean μ and standard deviation σ, then X is called heavy–tailed if

$$E\left[\frac{(X-\mu)^4}{\sigma}\right] > 3.$$

This property is called excess kurtosis because the fourth moment of the normal distribution is 3. Clearly, this definition is only helpful if the fourth moment of a random variable actually exists. Figure 1.3 below demonstrates five classes of distributions that are nested. [1]

Figure 1.3: Different Classes of Heavy–tailed Distributions



E: nonexistence of exponential moments

D: subexponential distributions

C: regular variation with tail index $\alpha > 0$

B: pareto tails with $\alpha > 0$

A: stable (non-normal) distributions

The broadest class E encompasses all distributions with

$$E(e^X) = \infty.$$

[1] This classification is borrowed from Bamberg and Dorfleitner (2001).

With this distinction, the normal distribution is clearly not contained in this class as its tail probability declines faster than exponentially. A distribution is subexponential distribution if

$$\lim_{n\to\infty} \frac{P((X_1 + \ldots + X_n) > x)}{P(\max(X_1, \ldots, X_n) > x)} = 1. \tag{1.2.3}$$

Therefore, a class D distribution has the interpretation that the sum of n i.i.d. subexponential random variables is likely to be large if and only if their maximum is likely to be large. This is often known as the principle of the single big jump. Another way to express (1.2.3) is

$$\lim_{t\to\infty} \frac{\bar{F}(t)}{e^{-\epsilon t}} = \infty \quad \forall\, \epsilon > 0.$$

As the name suggests, the tails of this class decrease slower than any exponential distribution.

In this dissertation, we focus on the class C distributions, which are characterized by regular variation in the tails. They form a subclass of the subexponential distributions and satisfy the condition

$$\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}.$$

Hence, far out in the tail ($t \to \infty$) the distribution in this class behaves like a Pareto distribution. As a consequence, the tail probabilities $P(X > x)$ decline according to a power function. As we have discussed above, the parameter α is called "regular varying exponent" and is used to measure the (heaviness) of the tail.

In contrast, distributions in class B have exact Pareto tails. The tail probability $1-F(x) = \bar{F}(x)$ of class B distributions is therefore $a^{\alpha}x^{-\alpha}$ where $x > a$ and $a > 0$. Understanding better the relationship between the regularly varying index α and moments of a distribution with Pareto tails, will help bridge the gap of understanding with the relatively unknown class A, or as it is called $\alpha-$stable distributions. Consider the k^{th} moment for a distribution in Class B, that is,

$$E[X^k] = \alpha a^{\alpha} \int_a^{\infty} x^{k-\alpha-1} dx.$$

It then follows that only $k-$moments with $k < \alpha$ are bounded. This property states that a distribution of class A will have Pareto tails with $\alpha < 2$, which implies infinite variance and, as a consequence, very heavy tails. Despite this restriction, class A is of great importance because asymptotic theory similar to the central limit laws is possible. Unfortunately, it is only possible to represent a stable distribution through its characteristic function. That is, the density function can only be computed by numerical approximation. Another common way to define a heavy-tailed distribution is if its moment generating function doesn't exist on the positive real line, i.e.

$$\int e^{tx} dF(x) = \infty \quad \text{for any } t > 0.$$

Thus we can simply say a distribution is heavy-tailed if and only if it has a heavier tail than any exponential distribution. Under this definition many distributions can be classified, such as Student's t, F, Cauchy, Pareto, log-normal, log-gamma, Weibull as heavy-tailed.

1.2.3 Background

Since there is a nice connection between class C and classical extreme value theory, let us look deeper into class C. Probably the fundamental topic in extreme value theory is modeling the variation of the sample maxima. Suppose we have an i.i.d. sequence of random variables, $X_1, X_2, \ldots,$ whose common cumulative distribution function is F, i.e.,

$$F(x) = P(X_i \leq x).$$

Then if $F(x) < 1$, we have $P(M_n \leq x) = F(x)^n \to 0$, where $M_n = \bigvee_{i=1}^{n} X_i = \max\{X_1, \ldots, X_n\}$. This result is of no immediate interest, since it simply says that for any fixed x for which $F(x) < 1$, we have that the probability of the nth sample maximum of the process staying below x converges to 0. However, if we set $x_0 = \sup\{x : F(x) < 1\} \leq \infty$, then

$$M_n \xrightarrow{a.s.} x_0. \tag{1.2.4}$$

Just as the normal distribution is a useful approximation to the distribution of $\sum_{i=1}^{n} X_i$, we seek a limit distribution to act as an approximation to F^n. The relation (1.2.4) makes it clear that a non-degenerate (i.e. , a distribution function which does not put all its mass at a single point) limit distribution will not exist unless we normalize M_n. It is common to use an affine normalization, which is also the most practical in statistical estimation problems. It was not until the mid 20th century before the founding fathers Fisher and Tippett (1928) originally stated without detailed mathematical proof, if the properly normalized maximum converges to a non-degenerate distribution G, then this distribution belongs to one of the following three distributions defined in Proposition (1.2.1). Later derived rigorously by Gnedenko (1943).

Proposition 1.2.1. *(Gnedenko, 1943) Suppose there exist $b_n > 0, a_n \in \Re, n \geq 1$ such that*

$$P[b_n^{-1}(M_n - a_n) \leq x] = F^n(b_n x + a_n) \to G(x),$$

weakly as $n \to \infty$ where G is assumed non-degenerate. Then G is of the type of one of the following three classes called the Domains of Attraction.

$$\textit{Fréchet: } \Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \end{cases}$$

$$\textit{Weibull: } \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x \leq 0 \\ 0 & x > 0, \end{cases}$$

$$\textit{Gumbel: } \Lambda(x) = \exp(-e^{-x}), \ \textit{for all } x,$$

where α is a positive parameter for the first two cases.

In this dissertation we will be focusing more on the class of distributions that belongs to the domain of attraction of the Fréchet distribution. A distribution belongs to this class if and only if its tails are regularly varying. This is exactly corresponds to class C distribution where the heaviness of the tails depends negatively on the tail index α.

Example: Suppose $\{X_i, i \geq 1\}$ is iid from $F(x) = 1 - e^{-x}, x > 0$ and $Y \sim \Phi(x)$. In the case $a_n = 1$ and $b_n = \ln(n)$,

$$\bigvee_{i=1}^{n} X_i - \ln(n) \Rightarrow Y.$$

Remark. This example is a situation where the right end point $x_0 = \sup\{s : F(s) < 1\} = \infty$.

1.2.4 Weak Convergence

Asymptotic properties of statistics in extreme value theory are understood with an interpretation which comes from weak convergence of probability measures on metric spaces. See Billingsley, (1968) for more details. Utilizing the power of weak convergence allows for a unified treatment of the one-dimensional and higher-dimensional cases of heavy-tail phenomena. Before we discuss some important properties of weak convergence, we must first familiarize ourself with some key definitions.

Definition 1.2.1. *A set A in S is compact if each open cover of A contains a finite subcover. In addition, a set A is complete if each fundamental sequence in A converges to some point of A.*

Tightness proves important in both the theory of weak convergence and in its applications. This condition helps insure that the probability mass does not escape the state space.

Definition 1.2.2. *A family Π of probability measures on the general metric space S is said to be tight if for every $\epsilon > 0$ there exists a compact set K such that $P(K) > 1 - \epsilon$ for all $P \in \Pi$.*

A main proof in Chapter 2 of this dissertation, makes use of dissecting covering rings. Let E denote a fixed locally compact second–countable Hausdorff space with Borel $\sigma-$algebra $\mathcal{B}$.

Let ζ denote a ring consisting of all bounded (i.e. relatively compact) sets in Borel $\sigma-$algebra $\mathcal{B}$.

Definition 1.2.3. *A DC–*ring *(D for dissecting, C for covering) is a ring $\mathcal{U} \subset \zeta$ with the property that, given any $B \in \zeta$ and any $\epsilon > 0$, there exists some finite cover of B by $\mathcal{U}$-sets of diameter less than ϵ (in any fixed metrization of E).*

Definition 1.2.4. *A DC–*semiring *is a class $\mathcal{J}$ of sets which is closed under finite intersections and such that any proper difference between $\mathcal{J}$-sets may be written as a finite disjoint union of sets in $\mathcal{J}$.*

In this dissertation, DC–rings and DC–semirings are families of intervals and interval unions respectively.

Definition 1.2.5. *A class $\mathcal{C} \subset \zeta$ is* covering *if every set $B \in \zeta$ has a finite cover of $\mathcal{C}$-sets.*

Here we define the actual definition of weak convergence, later we will discuss methods to actually prove weak convergence.

Definition 1.2.6. *Let F be the Borel σ-algebra of subsets of S. If we have a given sequence $\{X_n\}$ of random elements of S, there is a corresponding sequence of distributions on F, $P_n = P[X_n \in \cdot], n \geq 0$, where P_n is called the distribution of X_n. Then X_n converges weakly to X (written $X_n \Rightarrow X$) if whenever $f \in C(S)$, the class of bounded, continuous, real-valued functions on S, we have*

$$E[f(X_n)] = \int_S f(x)P_n(dx) \to E[f(X)] = \int_S f(x)P(dx).$$

Remark: If f is a measurable function on S, then, by the change-of-variable formula,

$$\int_{\Omega} f(X_n)dP = \int_{S} f(x)P_n(dx).$$

The following results are essential analytical tools used in proofs for this dissertation. For notational purposes, weak convergence can be expressed in either of the following ways

$$(i)P_n \Rightarrow P$$

$$(ii)X_n \xrightarrow{D} X$$

$$(iii)X_n \xrightarrow{w} X.$$

For the purpose of the following theorem, let ℓ denote the class of Borel sets as the σ-field generated by the open sets, which is the same thing as the σ-field generated by the closed sets.

Theorem 1.2.1. *(Billingsley, 1968) Let κ be a subclass of ℓ such that (i) κ is closed under finite intersections and (ii) each open set in S is a finite or countable union of elements of κ. If $P_n(A) \to P(A)$ for every $A \in \kappa$, then $P_n \Rightarrow P$.*

Many of the concepts and results in this dissertation deal with verifying convergence in probability.

Definition 1.2.7. *Let ρ denote the metric for a metric space S. If, for an element a of S, $P[\rho(X_n, a) \geq \epsilon] \to 0$ for each $\epsilon > 0$, we say X_n converges in probability to a and write*

$$X_n \xrightarrow{p} a.$$

Theorem 1.2.2. *(Billingsley, 1968) [continuous mapping theorem]. Let* $(S_i, \rho_i), i = 1, 2$ *be two metric spaces and suppose* $\{X_n, n \geq 0\}$ *are random elements of* (S_1, ℓ_1) *and* $X_n \Rightarrow X$. *If* $f : S_1 \to S_2$ *satisfies* $P[X \in D_f] = 0$, *where*

$$D_f = \{s_1 \in S_1 : \ f \ \textit{is discontinuities at} \ s_1 \ \}.$$

then,

$$f(X_n) \Rightarrow f(X), \quad \textit{in } S_2.$$

Many weak convergence results can be obtained through the continuous mapping theorem. Another method that can be applied given weak convergence is Slutsky's theorem.

Theorem 1.2.3. *(Resnick, 2007) [Slutsky's theorem]. Suppose* $\{X_n, X, Y_n, n \geq 1\}$ *are random elements of a separable metric space* S *with metric* $\rho(\cdot, \cdot)$. *If* $X_n \Rightarrow X$ *and* $\rho(X_n, Y_n) \xrightarrow{p} 0$, *then* $Y_n \Rightarrow X$.

Theorem 1.2.4. *(Billingsley, 1968) [second converging together theorem]. Suppose that* $\{X_{un}, X_u, Y_n, X; n \geq, u \geq 1\}$ *are random elements that have a common domain and that* S *is separable. Assume for each* u, $X_{un} \xrightarrow{D} X_u$ *as* $n \to \infty$ *and that* $X_u \xrightarrow{D} X$ *as* $u \to \infty$. *Suppose further that*

$$\lim_{u \to \infty} \limsup_{n \to \infty} P[\rho(X_{un}, Y_n) \geq \epsilon] = 0$$

for each $\epsilon > 0$. *Then* $Y_n \xrightarrow{D} X$ *as* $n \to \infty$.

1.2.5 Regular Variation

The theory of regular varying functions is an essential analytical tool for dealing with heavy tails, long-range dependence, and domains of attraction. Roughly speaking, regular varying functions are those functions which behave asymptotically like power functions.

Definition 1.2.8. *A real measurable function $U : \Re_+ \mapsto \Re_+$ is regular varying at infinity with index $-\beta \in \Re$ (denoted $U \in RV_{-\beta}$) if for $x > 0$,*

$$\lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^{-\beta}. \tag{1.2.5}$$

Note: We call β the exponent of variation. For a more detailed discussion, see Bingham, Goldie and Teugels (1987), and L. de Haan (1970).

For example, suppose X has a common distribution F, and consider for a fixed $x > 0$ the probability $P(X > tx)$. Then for large t, this probability is approximately $x^{-\beta} P(X > t)$ with $\beta \geq 0$. If this statement is true for all $x > 0$, then F is said to have a regularly varying right tail.

If $\beta = 0$, then we call U slowly varying. Slowly varying functions are generically denoted by $\ell(x)$. The significance of slowly varying functions is that we can express the tail of a distribution as a multiple of a slowly varying function, i.e., $U(x) = x^{-\beta}\ell(x)$.

The next theorem concerns global bounds for $\ell(y)/\ell(x)$.

Theorem 1.2.5. *(Bingham, Goldie, and Teugels, 1987) [Potter's Theorem]. (i) If ℓ is slowly varying then for any chosen constants $A > 1, \delta > 0$ there exists $X = X(A, \delta)$ such that*

$$\ell(y)/\ell(x) \leq A \left\{ (\frac{y}{x})^{\delta} \vee (\frac{y}{x})^{-\delta} \right\} \quad (x \geq X, y \geq X).$$

(ii) If, further, ℓ is bounded away from 0 and ∞ on every compact subset of $[0, \infty)$, then for every $\delta > 0$ there exists $A' = A'(\delta) > 1$ such that

$$\ell(y)/\ell(x) \leq A' \{ (\frac{y}{x})^{\delta} \vee (\frac{y}{x})^{-\delta} \} \quad (x \geq 0, y \geq 0).$$

(iii) If f is regularly varying of index ρ, then for any chosen $A > 1, \delta > 0$ there exists $X = X(A, \delta)$ such that

$$f(y)/f(x) \le A\{(\frac{y}{x})^{\rho+\delta} \vee (\frac{y}{x})^{\rho-\delta}\} \quad (x \ge X, y \ge X).$$

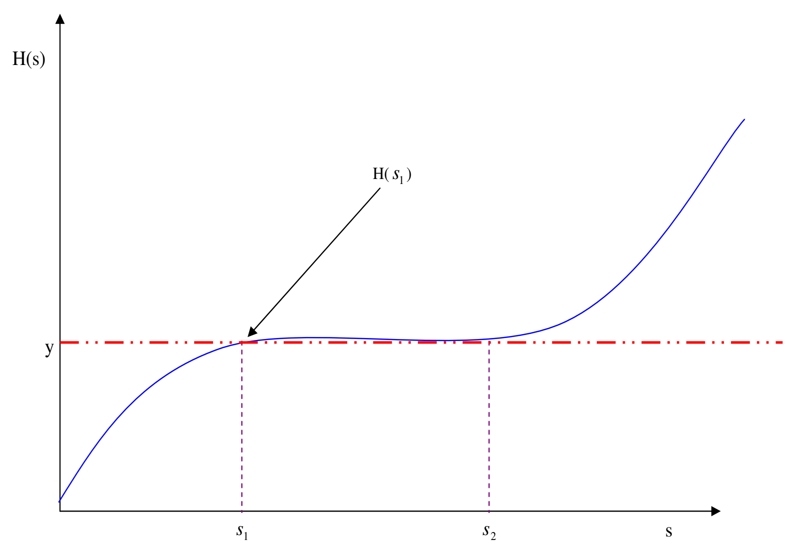
Having established regular variation and possible limit laws for normalized maxima we are ready now to characterize the normalizing constant $\{b_n\}$. First, recall the definition for the inverse of a monotone function.

Definition 1.2.9. *Suppose $H : \Re \mapsto (a, b)$ is a nondecreasing function on $\Re$ with range (a, b), where $-\infty \le a < b \le \infty$. Then the (left-continuous) inverse $H^{\leftarrow} : (a, b) \mapsto \Re$ of H is*

$$H^{\leftarrow}(y) = \inf\{s : H(s) \ge y\}.$$

Figure 1.4 below gives an illustration of this definition.

Figure 1.4: Inverse of a Monotone Function



The inverse at y is the foot of the left dotted perpendicular.

The following corollary provides the normalizing constants for M_n in Proposition 1.2.1 and for an extreme value estimates, which in turn can improve the rate of convergence on

$\sqrt{n}$. Note that here and elsewhere that we use the notation

$$f(x) \sim g(x), x \to \infty, \text{ as shorthand for } \lim_{x\to\infty} \frac{f(x)}{g(x)} = 1,$$

for two real functions f, g.

Corollary 1.2.1. *(Resnick, 1987) Let $F(x)$ be a distribution with a regular varying tail at ∞ with index $-\alpha$ for some $\alpha > 0$. Set $U(x) = 1/(1-F(x))$, then $b_n = U^{\leftarrow}(n) = F^{\leftarrow}(1-1/n)$ determines b_n as the normalizing constant in Proposition 1.2.1 with $a_n = 0$. Then, using the fact that $b_n \to \infty$ we have for $x > 0$,*

$$\lim_{n\to\infty} \frac{1-F(b_n x)}{1-F(b_n)} = x^{-\alpha}. \tag{1.2.6}$$

1.2.6 Point Processes

The use of point processes as a statistical approach in extreme value theory was introduced by Smith (1989). Although a few statisticians have used point processes, this technique is becoming popular. In particular, Davis and McCormick (1989) derived the limiting distribution of their proposed estimators for a First-Order Autoregressive process using point processes. Books by Leadbetter, et al (1983) and Resnick (1987) contain excellent information on point processes used in this dissertation.

In this approach, instead of considering times at which high-threshold exceedances occur and the excess values over the threshold as two separate processes, they are combined into one process based on the two-dimensional plot of exceedance times and values. The asymptotic theory of threshold exceedances shows that under suitable normalization, this process behaves like a nonhomogeneous Poisson process.

Definition 1.2.10. *A nonhomogeneous Poisson process on a domain D is defined by an intensity $\lambda(x), x \in D$, such that if $A_1, \ldots, A_k$ are disjoint subsets of a measurable subset of D and $N(A)$ denotes the number of points in A, where $N(A_1), \ldots, N(A_k)$ are independent* *random variables, then $N(A)$ has a Poisson distribution with mean*

$$\mu(A) = \int_A \lambda(x)dx.$$

Definition 1.2.11. *A point process N is a Poisson random measure denoted by (PRM(μ) with mean measure μ, if*

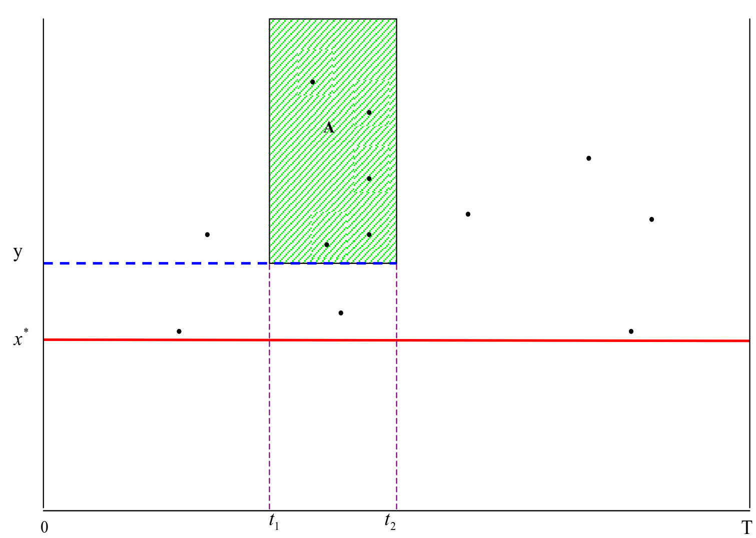
$$P[N(A) = k] = e^{-(\mu(A))}\mu(A)^k/k!, \quad k = 0, 1, \ldots$$

for $A \in D$ with $\mu(A) < \infty$.

That is, a point process is nothing more than a random distribution of points in space. Since point processes are such an essential tool in our dissertation, we will begin by explaining how point processes can be applied to extreme value theory.

Assume the process is observed over a time interval $[0, T]$, and that all observations above a threshold level $x^\star$ are recorded, and denoted by $x = (t, y)$ where t is time, and $y \geq x^\star$ is the value of the process, $D = [0, T] \times [x^\star, \infty]$. These points are marked on a two-dimensional scatterplot and Figure 1.5 below provides an illustration of a point process.

Figure 1.5: Illustration of Point Process Approach



In this dissertation, the state space where the points live will be denoted by E, unless otherwise mentioned we will assume E to be a fixed locally compact second–countable Hausdorff state space with the Borel $\sigma-$algebra $\mathcal{B}$. For us, E will typically be a subset of Euclidean space of finite dimension.

Definition 1.2.12. *For $x \in E$, we define the measure ε_x by*

$$\varepsilon_x(A) = \begin{cases} 1 & x \in A \\ \\ 0 & x \in A^c. \end{cases}$$

Definition 1.2.13. *Let C be compact subsets of E. A measure μ is called Radon if*

$$\mu(K) < \infty \textit{ for all } K \in C$$

Thus compact sets are known to have finite $\mu-$mass.

Definition 1.2.14. *A point measure on E is a measure μ which can be represented as follows. If $\{x_i, i \geq 1\}$ is a countable collection of (not necessarily distinct) points of E, then*

$$\mu := \sum_{i=1}^{\infty} \varepsilon_{x_i}. \tag{1.2.7}$$

We can think of $\{x_i\}$ as the atoms and μ as the function that counts how many atoms fall in a set. The set $M_p(E)$ is the set of all Radon point measures of the form (1.2.7). That is, $M_p(E)$ is the class of non–negative integer–valued Radon measures on E. Hence, $M_p(E)$ is a closed subset of $M_+(E)$, where $M_+(E) = \{\mu : \mu > 0$ measure on E and μ is Randon$\}$.

Definition 1.2.15. *A point measure on E is considered a measurable map from a probability space $(\Omega, \Im, P)$ into $(M_p(E), \mathfrak{M}_p(E))$, where $\mathfrak{M}_p(E)$ is the σ-algebra generated by the vague topology.*

A small but important detail to notice throughout the literature in extreme value theory and this dissertation is with infinite measures in $M_+(E)$, we cannot just integrate a bounded function to get something finite. However, we know our measures are also Radon, therefore we can try to use functions that vanish on complements of compact sets.

Definition 1.2.16. *Let*

$$C_K^+(E) = \{f : E \mapsto R_+ : \ f \text{ is continuous with compact support}\}.$$

If $\mu_n \in M_+(E)$ for $n \geq 0$, then μ_n converges vaguely to μ, (written $\mu_n \xrightarrow{v} \mu$), if for all $f \in C_K^+(E)$, we have

$$\mu_n(f) := \int_E f(x)\mu_n(dx) \to \mu(f) := \int_E f(x)\mu(dx) \text{ as } n \to \infty.$$

The next theorem makes the connection between regular variation and vague convergence. Since the vague topology renders $M_p(E)$ a complete separable metric space, we may speak of convergence in distribution of point processes which will be denoted by $\Rightarrow$.

Theorem 1.2.6. *(Resnick, 2007) Suppose $X_1 \geq 0$ is a random variable with distribution function $F(x)$. The following are equivalent:*
(i) $1 - F(x) \in RV_{-\beta}, \beta > 0$.
(ii) There exists a sequence $\{b_n\}$ with $b_n \to \infty$ such that

$$\lim_{n\to\infty} n(1 - F(b_n x)) = x^{-\beta}, \quad x > 0.$$

(iii) There exists a sequence $\{b_n\}$ with $b_n \to \infty$ such that

$$\mu_n(\cdot) := nP[b_n^{-1} X_1 \in \cdot] \xrightarrow{v} \nu_\beta.$$

in $M_+(0, \infty]$, where $\nu_\beta(x, \infty] = x^{-\beta}$.

Remark. Note that in (iii) the space $E = (0, \infty]$ has excluded 0 and included ∞. This is required since we need neighborhoods of infinity to be relatively compact. Within the vague topology, we can specify the notion of "distance" in $M_p(E)$ by putting a metric $d(\cdot, \cdot)$ on the space.

Definition 1.2.17. *If there exists some sequence of functions $f_i \in C_k^+(E)$ and $\mu_1, \mu_2 \in M_p(E)$, then the vague metric is defined as*

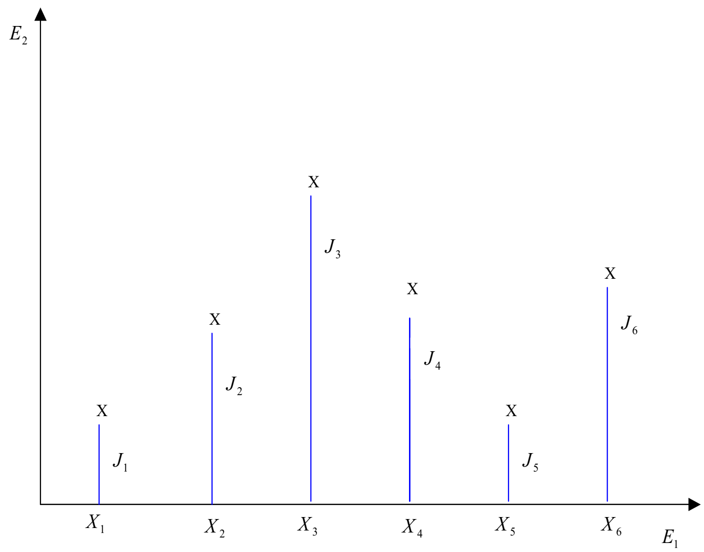
$$d(\mu_1, \mu_2) = \sum_{i=1}^{\infty} \frac{(1 - e^{-|\mu_1(f_i) - \mu_2(f_i)|})}{2^i}, \text{ for } i = 1, \dots, n. \tag{1.2.8}$$

Proposition 1.2.2. *(Resnick, 1987) Suppose $\{X_n\}$ are random elements of a nice space E_1 such that $\sum_n \varepsilon_{X_n}$ is PRM(μ). Suppose $\{J_n\}$ are iid random elements of a second nice space E_2 with common probability distribution F, and suppose the Poisson process and the sequence $\{J_n\}$ are defined on the same probability space and are independent. Then the point process on $E_1 \times E_2$, is given by*

$$\sum_n \varepsilon_{(X_n, J_n)} \text{ is PRM } (\mu \times F).$$

This procedure is described by saying we give to point X_n the mark J_n. See figure 1.6 below for an illustration.

Figure 1.6: Illustration of Marking an Poisson Process



The next proposition has far-reaching implications and provides the link between regular variation and point processes.

Proposition 1.2.3. *(Resnick, 1987) For each n suppose $\{X_{n,j}, j \geq 1\}$ are iid random elements of E and μ is a Radon measure on E. Define $N_n := \sum_{j=1}^{\infty} \varepsilon_{(jn^{-1}, X_{n,j})}$ and suppose N is PRM($dt \times d\mu$) on $[0, \infty) \times E$. Then $N_n \Rightarrow N$ in $M_p([0, \infty) \times E)$ iff*

$$nP[X_{n,1} \in \cdot] \xrightarrow{v} \mu \qquad \text{on } E.$$

The following condition is necessary for Breiman's Theorem which deals with the case where the multiplier has a relatively thin tail:

$$\int x^{\beta} F(dx) < \infty \quad \text{for some } \beta > \alpha. \tag{1.2.9}$$

Proposition 1.2.4. *(Resnick, 2007) [Breiman's Theorem]. Let Z and Y be two positive and independent random variables with distribution functions F and G, respectively. Suppose F satisfies* (1.2.9) *and $EY^\beta < \infty$ for some $\beta > \alpha$. Then*

$$\lim_{x\to\infty} \frac{P[YZ > x]}{P[Z > x]} = EY^\alpha.$$

1.3 Literature Review

1.3.1 Estimation for an First–order Autoregressive Processes

In this dissertation we consider a stationary AR(1) process $\{X_t\}$ satisfying the difference equations

$$X_t = \phi X_{t-1} + Z_t, t = 0, \pm 1, \dots, \tag{1.3.1}$$

where $|\phi| < 1$ and $\{Z_t\}$ is an iid sequence of random variables with common distribution F. A common estimator known as the Yule-Walker estimator (see Brockwell and Davis, (1987)) for the autocorrelation parameter $\phi = corr(X_t, X_{t-1})$ when $\{Z_t\}$ has finite variance is

$$\hat{\phi} = \sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X}) / \sum_{t=1}^{n}(X_t - \bar{X})^2, \quad \bar{X} = 1/n \sum_{t=1}^{n} X_t.$$

This estimator is an asymptotically normal random variable with mean ϕ and asymptotic variance $(1 - \phi^2)/n$, i.e.,

$$\sqrt{n}(\hat{\phi} - \phi) \Rightarrow N(0, 1 - \phi^2).$$

Davis and McCormick (1989) lay the foundation for our dissertation. They made two major contributions; one, considering an estimator

$$\hat{\phi} = \bigwedge_{t=1}^{n} X_t / X_{t-1} \tag{1.3.2}$$

when the innovation distribution F is regularly varying at the lower endpoint 0, which under certain conditions can be vastly superior to the Yule-Walker estimator. Secondly, considering when F is bounded between $[-1, 1]$. In this situation, they determined the limiting distribution for their estimators as well as an insightful example. Below are some of the results from this article. The first proposition and corollary are handy tools for weak convergence of point processes.

Proposition 1.3.1. *(Davis and McCormick, 1989) Let $\{X_t\}$ be the stationary AR(1) process* (1.3.1)*, where F is regularly varying at 0 with variation index α. Let N_n and N be the point processes on the space $E = [0, \infty) \times [0, \infty)$ defined by*

$$N_n = \sum_{t=1}^{n} \varepsilon_{(a_n^{-1} Z_t, X_{t-1})} \quad \text{and} \quad N = \sum_{k=1}^{\infty} \varepsilon_{(j_k, Y_k)}$$

where $a_n = F^{\leftarrow}(1/n) := inf\{x : F(x) \geq 1/n\}$, $\sum_{k=1}^{\infty} \varepsilon_{j_k}$ is PRM($\alpha x^{\alpha-1}dx$) and $\{Y_k\}$ is an iid sequence of random variables, independent of $\sum_{k=1}^{\infty} \varepsilon_{j_k}$, with $Y_1 = X_1$. In other words, N is PRM($\alpha x^{\alpha-1}dx \times G(dy)$), where $G(y) = P[Y_1 \leq y]$. Then in $M_p(E)$,

$$N_n \Rightarrow N.$$

Corollary 1.3.1. *(David and McCormick, 1989) Under the conditions of Proposition 1.3.1. Define the point process η_n and η on $[0, \infty)$ by*

$$\eta_n = \sum_{t=1}^{n} \varepsilon_{a_n^{-1}(X_t/X_{t-1} - \phi)} \quad \text{and } \eta = \sum_{t=1}^{\infty} \varepsilon_{j_k}/Y_k$$

where $\sum_{k=1}^{\infty} \varepsilon_{j_k}$ and $\{Y_k\}$ are as defined in the statement of Proposition 1.3.1. In particular η is PRM $\left((EX_1^{\alpha})\, \alpha x^{\alpha-1}dx\right)$. Then in $M_p([0, \infty))$,

$$\eta_n \Rightarrow \eta.$$

Davis and McCormick (1989) used the above results produced from point processes to obtain the limit distribution for $\hat{\phi}$.

Theorem 1.3.1. *(Davis and McCormick, 1988) With* $\hat{\phi} := \bigwedge_{t=1}^{n} X_t/X_{t-1}$*, we have*

$$\lim_{n\to\infty} P[a_n^{-1}(\hat{\phi} - \phi)c_\alpha \leq x] = 1 - \exp\{-x^\alpha\}, \quad x > 0,$$

where $c_\alpha = [E(X_1^\alpha)]^{1/\alpha}$ *and* $\hat{\phi} \to \phi$ *a.s. .*

In the following corollary, they show the same limit law holds if $\{X_t\}$ is replaced by the non-stationary solution $\{\tilde{X}_t\}$ defined by

$$\tilde{X}_t = \begin{cases} 0 & \text{if } t = 0, \\ \phi\tilde{X}_t + Z_t & \text{if } t \geq 1, \end{cases}$$

which has the representation

$$\tilde{X}_t = \sum_{j=1}^{t-1} \phi^j Z_{t-j}. \tag{1.3.3}$$

Corollary 1.3.2. *(Davis and McCormick, 1989) Let* $\{\tilde{X}_t, t = 0, 1, \ldots\}$ *be the nonstationary AR(1) process given by 1.3.3 and define* $\tilde{\phi} = \bigwedge_{j=2}^{n} \tilde{X}_j/\tilde{X}_{j-1}$*. Then*

$$\lim_{n\to\infty} P[a_n^{-1}(\tilde{\phi} - \phi)c_\alpha \leq x] = 1 - exp\{-x^\alpha\}, \quad x > 0$$

where c_α *is as specified in Theorem 1.3.1 and* $\tilde{\phi} \xrightarrow{a.s.} \phi$*.*

Davis and McCormick (1989) also considered the case when F is supported on $[-1, 1]$, is regularly varying at its endpoints and satisfies a balancing condition at the two endpoints.

Theorem 1.3.2. *Let $\{X_t\}$ be the stationary AR(1) process given by* (1.3.1) *where the distribution of the innovation sequence satisfies*

$$\lim_{t \downarrow 0} \frac{P[Z_t \leq -1 + tx]}{P[Z_t \leq -1 + t] + P[Z_t > 1 - t]} = qx^{\alpha},$$

and

$$\lim_{t \downarrow 0} \frac{P[Z_t > 1 - tx]}{P[Z_t \leq -1 + t] + P[Z_t > 1 - t]} = px^{\alpha},$$

where $p, q \geq 0$ and $p + q = 1$. Define the sequence of positive constants a_n by $a_n = g^{\leftarrow}(n^{-1})$ where g is the nondecreasing function $g(t) = P[Z_t \leq -1 + t] + P[Z_t > 1 - t]$. Define

$$T_{1n} = \bigvee_{t=1}^{n} \left[\frac{X_t}{X_{t-1}} - \frac{1}{|X_{t-1}|} \right] \quad \text{and } T_{2n} = \bigwedge_{t=1}^{n} \left[\frac{X_t}{X_{t-1}} + \frac{1}{|X_{t-1}|} \right].$$

Then, for all $x > 0$ and $y > 0$,

$$\lim_{n \to \infty} P[a_n^{-1}(T_{1n} - \phi) \leq -x, a_n^{-1}(T_{2n} - \phi) > y] = \exp\{-c_1 x^{\alpha} - c_2 y^{\alpha}\}. \tag{1.3.4}$$

Here the endpoints are known, therefore they simply add/subtract a known constant $1/|X_{t-1}|$ to $\hat{\phi}$ defined in (1.3.2) and apply the result when F is regular varying at zero. McCormick and Mathew (1993) considered an AR(1) process with an unknown location parameter, but unlike Davis and McCormick (1989) they did not consider a finite upper bound. However, they proposed estimates of (θ, ϕ), which were obtained as the solution to a linear programming problem, where θ represents the unknown lower endpoint and ϕ the correlation coefficient. In addition, they did a simulation study to determine the performance of their estimators for (θ, ϕ). This help demonstrated the dependence of the estimators on the model and how if that model is not correct, then there is not much robustness in the estimators. This result verifies that their estimators for (θ, ϕ) are strongly consistent.

Theorem 1.3.3. *(McCormick and Mathew, 1993) Let $\{Z_t, t = 0, \pm 1, \pm 2, \ldots\}$ be an i.i.d. sequence of nonnegative random variables having common distribution function F. Define a stationary sequence by*

$$X_t = \theta + \phi X_{t-1} + Z_t \quad (t \geq 1) \tag{1.3.5}$$

where

$$X_0 = \theta/(1-\phi) + \sum_{i=0}^{\infty} \phi^i Z_{-i} \tag{1.3.6}$$

and $\theta \geq 0, \quad 0 \leq \phi < 1$. If $(\hat{\theta}, \hat{\phi})$ are estimates defined as the value of (θ, ϕ) maximizing $\bar{X}\phi + \theta$ subject to $X_t - \theta - \phi X_{t-1} \geq 0, \quad 1 \leq t \leq n$, then

$$(\hat{\theta}, \hat{\phi}) \xrightarrow{a.s.} (\theta, \phi) \qquad \text{as } n \to \infty.$$

A particular result from McCormick and Mathew (1993) is that the minimum $\{X_t\}$ converges almost surely as $n \to \infty$.

Corollary 1.3.3. *(McCormick and Mathew, 1993) Under the assumptions of Theorem 1.3.3 we have that*

$$\bigwedge_{k=1}^{n} X_k \xrightarrow{a.s.} \frac{\theta}{1-\phi} \qquad \text{as } n \to \infty.$$

The next result is very useful when trying to prove weak convergence of a point process in the situation when X_t is not independent such as Proposition 1.2.3.

Proposition 1.3.2. *Let $\{X_t\}$ be the AR(1) process given by $X_t = \theta + \phi X_{t-1} + Z_t \quad (t \geq 1)$ with error distribution F satisfying (1.2.6). If $a_n = F^{\leftarrow}(1 - 1/n)$ and*

$$M_n = \sum_{k=1}^{n} \varepsilon_{((k/n), a_n^{-1} X_k, Z_{k+1})}$$

then in $M_p(E)$ we have

$$M_n \Rightarrow M$$

where $M = \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \varepsilon_{(t_k, \phi^i j_k, V_k)}$.

There has been a good bit of work done on the AR(1) process, but little work for higher order autoregressive process has been successful. Anděl, (1989) considered the case of $p > 1$, but his straightforward generalization of $\hat{\phi}$ defined in (1.3.2) did not perform well. Under the AR(2) model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$ he suggested two estimators of (ϕ_1, ϕ_2). One of which was based on a maximum likelihood argument. Using simulation he found that this estimator also converged at a faster rate than the Yule-Walker estimator. Anděl's finding is explained in Feigin and Resnick (1992), whom establish a rate of consistency for the estimators of ϕ_1 and ϕ_2 in the case $p = 2$. Feigin and Resnick (1994) used an MLE argument when the innovations have an exponential distribution. In their article they considered a stationary autoregressive processes of order p and derive a limit distribution for $\hat{\phi}$ defined in (1.3.2) which has positive innovations using a linear program for their consistent estimators. Feigin and Resnick (1992) assumed that $\{\phi_1, \ldots, \phi_p\}$ were nonnegative. Based on the literature this is suitable for most modeling applications of data such as, stream flows, interarrival times, teletraffic applications (video conference scenes), etc. However, Feigin and Resnick (1992) discovered that the nonnegative assumption can be dropped, adding flexibility to fitting models to data. The aim of Feigin and Resnick (1999) was to evaluate what happens when the $AR(p)$ model is not an accurate description of the structure in the time series. They considered alternatives such as a model with contamination of large outliers, a non-linear time series, and a moving average $MA(q)$. Another major result in their article is that the common distribution F can have regular variation of either the left or right tail instead of just regular variation at one tail. Below is a standard but necessary condition used in many articles that used a linear program to obtain there estimates.

Condition S (stationarity): The coefficients $\phi_1, \ldots, \phi_p$ satisfy the stationarity condition that the autoregressive polynomial $\Phi(z) \equiv 1 - \sum_{i=1}^{p} \phi_i z^i$ has no roots in the unit disk $\{z : |z| \leq 1\}$. Furthermore, assume $\Phi(1) > 0; i.e., \sum_{i=1}^{p} \phi_i < 1$.

1.3.2 Estimation for Bifurcating Autoregressive Processes

Bifurcating autoregressive (BAR) processes are an adaptation of autoregressive (AR) processes to binary tree structured data. They were first introduced by Cowan and Staudte (1986) for cell lineage data, where each individual in one generation gives birth to two offspring in the next generation. Cell lineage data typically consist of observations of some quantitative characteristic of the cells over several generations of descendants from an initial cell. BAR processes take into account both inherited and environmental effects to explain the evolution of the quantitative characteristic under study. More precisely, the original BAR process is defined as follows. The initial cell is labeled 1, and the two offspring of cell t are labeled $2t$ and $2t+1$. Denoted by X_t the quantitative characteristic of individual t. Then, the first-order BAR process is given, for all $t \geq 1$, by

$$X_t = \phi X_{\lfloor t/2 \rfloor} + \epsilon_t, \tag{1.3.7}$$

where $0 \leq \phi < 1$, $\lfloor x \rfloor$ is the largest integer which does not exceed x. Thus, we first arrange the observations $(X_t, t = 1, \dots, n)$ into k generations, the jth generation consisting of 2^{j-1} observations, $j = 1, \dots, k$, and $n = 2^k - 1$ total number of observations. Let A_j denote the set of observations in the jth generation. Thus, if $t \in A_j$ then $j = \log_2(t) + 1$. Furthermore, we define B_j to be the set of all observations contained in the first j generations.

The driven noise $(\epsilon_{2t}, \epsilon_{2t+1})$ was originally supposed to be independent and identically distributed with normal distribution. However, two sister cells being in the same environment early in their lives, (ϵ_{2t} and ϵ_{2t+1}) are allowed to be correlated, inducing a correlation between sister cells distinct from the correlation inherited from their mother. Huggins and Basawa (1999) proposed bifurcating ARMA(p, q) models to accommodate for correlation between the environmental effects of relatives more distant than sisters.

There are several results on statistical inference and asymptotic properties of estimators for BAR models in the literature. Huggins and Basawa (2000) discussed maximum likelihood estimation (mle) for a Gaussian BAR(p) process and established the consistency and asymptotic normality of the mle of the model parameters. Recently, Zhou and Basawa (2004) introduced non-Gaussian bifurcating autoregressive models. They also considered the asymptotic properties of the least-squares estimator (LSE) of parameters in a BAR(p) process, see Zhou and Basawa (2005a). In all these papers, the process is supposed to be stationary. Consequently, $\{X_t\}$ has a time-series representation involving an holomorphic function. The goal of this dissertation is to improve and extend the previous results in the non-stationary situation. As previously done by Zhou and Basawa (2004, 2005a, 2005b), we shall make use of the strong law of large numbers as well as the central limit theorem for martingales. See Hall and Heyde (1980) for more details on martingales. Based on Bercu, Saporta and Petit (2009) it should allow us to relax some of the standard assumptions on the driven noise (ϵ_t) found in articles on this subject. Below are some keystone results from various articles mention above.

Theorem 1.3.4. *(Zhuo and Basawa, 2005b) Suppose the marginal distributions of* ϵ_{2t} *and* ϵ_{2t+1} *are both exponential with mean* λ *and correlation* ρ. *Let* A_j *denote the set of observations in the* j*th generation. If* $\hat{\phi}_{ML} = \bigwedge_{t=2}^{n} \frac{X_t}{X_{\lfloor t/2 \rfloor}}$. *Then we have*

$$\lim_{k\to\infty} \alpha_k(\hat{\phi_{ML}} - \phi) \xrightarrow{d} \exp(1), \qquad \textit{for all } \phi \geq 0$$

where

$$\alpha_k = \begin{cases} \frac{2^k}{(1+\rho)(1-\phi)}, & \textit{for } 0 \leq \phi < 1 \ \textit{(stationary)} \\ \frac{2^k k}{1+\rho}, & \textit{for } \phi = 1 \ \textit{(critical)} \\ \frac{2^k \phi^{k-1} W}{\lambda(1+\rho)(2\phi-1)}, & \textit{for } \phi > 1, \ \textit{(explosive)} \end{cases}$$

with W *being a positive random variable defined by* $W = \sum_{j=2}^{\infty} \phi^{-(j-1)} \bar{\epsilon_j} + X_1$, *with* $\bar{\epsilon_j} = 2^{j-1)} \sum_{t \in A_j} \epsilon_t$ *the average of* ϵ_t*s corresponding to the* j*th generation.*

Below is a fundamental result for least-squares estimation of $\phi = (\phi_0, \ldots, \phi_p)'$ for a BAR(p) model, where $Y_t = (1, X_{[t/2]}, \ldots, X_{[t/2^p]})'$ without imposing any specific distributional assumption on ϵ_t.

Theorem 1.3.5. *(Zhou and Basawa, 2005a)Let* $\hat{\phi} = (\sum_{t=2^p}^{n} Y_t Y_t')^{-1} \sum_{t=2^p}^{n} Y_t X_t$. *Under sufficient conditions, we have*

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2(1+\rho)A^{-1}) \quad \text{as } n \to \infty.$$

where A is a positive definite matrix.

Consider the BAR(1) model defined in (1.3.7) where the LS estimators are given by

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{n} U_t (X_t - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2} \text{ where } U_t = \frac{\epsilon_{2t} + \epsilon_{2t+1}}{2} \text{ and } \bar{X} = \sum_{t=1}^{n} X_t,$$
$$\hat{\phi}_0 = \bar{U} - \hat{\phi}_1 \bar{X} \text{ where } \bar{U} = \sum_{t=1}^{n} U_t.$$

Corollary 1.3.4. *(Zhou and Basawa, 2005a) The limit distribution for the LS estimator $\hat{\phi}_1$ is*

$$\sqrt{n}(\hat{\phi}_1 - \phi_1) \xrightarrow{d} N(0, (1+\rho)(1-\phi_1^2)).$$

The following result provides the limiting distribution for the correlation parameter $\hat{\phi}_n$ in the first-order bifurcating autoregressive process where we assume $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ is a sequence of independent random vectors such that the distribution of $Y_t = \epsilon_{2t} \wedge \epsilon_{2t+1}$, F_Y, is regularly varying at 0 with index α, i.e. for all $x > 0$

$$\lim_{t \to 0^+} \frac{F_Y(tx)}{F_Y(t)} = x^{\alpha}.$$

Furthermore, we define $b^{-1}(t) = \inf\{x : F_{Y_1}(x) \geq 1/t\}$ for $t > 1$.

The following proposition is one particular result Zhang (2011).

Proposition 1.3.3. *Let* $\hat{\phi}_n = \bigwedge_{t=2}^{n} \frac{X_t}{X_{\lfloor t/2 \rfloor}}$. *For any* $x > 0$,

$$P[(E(X_1)^\alpha)^{1/\alpha} b(n)(\hat{\phi}_n - \phi) > x] \to \exp(-x^\alpha).$$

1.4 Outline of Dissertation

In Chapter 2, we consider a first-order autoregressive processes $X_t = \phi X_{t-1} + Z_t$, where the innovations are nonnegative random variables with regular variation at both the right endpoint infinity and the unknown left endpoint θ. We propose estimates for the autocorrelation parameter ϕ and the unknown location parameter θ. The joint limit distribution of the proposed estimators is derived using point process techniques. A simulation study is provided to examine the small sample size behavior of these estimates.

In Chapter 3, we consider an infinite order moving average processes $X_t = \sum_{i=0}^{\infty} c_i Z_{t-i}$ where the coefficients are nonnegative and the innovations are positive random variables with a regularly varying tail at infinity, we provide estimates for the coefficients. We then apply this result to obtain estimates for the parameters of nonnegative ARMA models. Weak convergence results for the joint distribution of our estimates are established and a simulation study is provided to examine the small sample size behavior of these estimates.

In Chapter 4, we consider a first-order bifurcating autoregressive processes $X_t = \phi X_{\lfloor t/2 \rfloor} + \epsilon_t$, where the innovations are nonnegative random variables with regular variation at both the right endpoint infinity and the unknown left endpoint θ. We provide estimates for the autocorrelation parameter ϕ and unknown location parameter θ. Using a moment generating technique, the limit distributions of the proposed estimators are directly derived and compared with an extensive simulation study.

Each chapter is self-contained in terms of describing and highlighting the performance of the above mentioned methods.

1.5 References

[1] Anděl, J. (1989). Non-negative autoregressive processes. *J. Time Ser. Anal.* **10**, 1-11.

[2] Anděl, J. (1991). Non-negative linear processes. *Appl. of Math.* **36**, 277-283.

[3] Anderson, P.L., Kavalieris, L., Meerschaert, M.M. (2008). Innovations algorithm asymptotics for periodically stationary time series with heavy tails. *J. Multivariate Anal.* **99**, 94-116.

[4] Bamberg, G. and Dorfleitner, G. (2001). *Fat Tails and Traditional Capital Market Theory, mineo*, University of Augsburg, Germany.

[5] Billingsley P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[6] Bingham N.H., Goldie C.M., and Teugels J.L., (1987). *Regular Variation.* Cambridge University Press, Cambridge, UK.

[7] Brockwell, P.J., Davis, R.A. (1987). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

[8] Brown, T.C., Feigin, P.D., Pallant, D.L. (1996). Estimation for a class of positive nonlinear time series models. *Stochastic Process. Appl.* **63**, 139-152.

[9] Calder, M., Davis, R.A. (1998). In R.J. Adler, R.E. Feldman, M.S. Taqqu, editors, *A Practical Guide To Heavy Tails: Statistical Techniques and Applications* pages 159-176, Birkhäuser, Boston.

[10] Collings P.S. (1975). Dams with autoregressive inputs. *J. Appl. Probab.* **12**, 533 - 541.

[11] Cowan, R. Staudte, R. G. (1986). The bifurcating autoregressive model in cell lineage studies. *Bioetrics*, **42**, 769 - 783.

[12] Datta, S., McCormick, W.P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *J. Amer. Statist. Assoc.* **90**, 1289-1300.

[13] Datta, S., Mathew, G., McCormick, W.P. (1998). Nonlinear autoregression with positive innovations. *Austral. & New Zealand J. Statist.* **40**, 229-239.

[14] Davis, R.A., Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.

[15] Davis, R.A., McCormick, W.P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Process. Appl.* **31**, 237–250.

[16] Feigin P. and Resnick S.I., (1992). Estimation for autoregressive processes with positive innovations. *Stochastic Models*, **8**, 479 - 498.

[17] Feigin, P.D., Resnick, S.I. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51**, 135-165.

[18] Feigin, P.D., Kratz, M.F., Resnick, S.I. (1996). Parameter estimation for moving averages with positive innovations. *Ann. Appl. Probab.* **6**, 1157-1190.

[19] Feigin, P. and Resnick, S.I., (1999). Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes*, **1**, 391 - 422.

[20] Finkenstadt B. and Rootzen H., (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. CRC Press LLC, New York.

[21] Fisher R.A. and Tippett L.H.C., (1928). Limiting forms of the frequency distributions of largest or smallest member of a sample. *Proc. Cambridge Philos.Soc.*, **24**, 180 - 190.

[22] Gnedenko B.V., (1943). Sur la distribution limite du terme d' une serie aleatoire. *Ann. Math*, **44**, 423 - 453.

[23] Haan L. de, (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematisch Centrum, Amsterdam.

[24] Hall P. and Heyde C.C., (1980). *Martingale limit theory and its application.* Academic Press Inc., New York.

[25] Huggins, R.M., Basawa, I.V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36**, 1225 - 1233.

[26] Huggins, R.M., Basawa, I.V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42**, 423 - 432.

[27] Hutton, J. (1990). Non-negative time series models for dry river flow. *J.Appl. Probab.* **27**, 171–182.

[28] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

[29] McCormick W.P. and Mathew, G., (1993). Estimation for Nonnegative autoregressive Processes with an Unknown Location Parameter. *J. Time Series Anal.* **14**, 71–92.

[30] Markov, A.S. (2009). Estimation of the autoregressive parameter with infinite dispersion of noise. *Automation and Remote Control* **70**, 92-106.

[31] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[32] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

[33] Smith R.L., (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statist. Sc.*, **4**, 367 - 393.

[34] Tsai, H., Chan, K.S. (2006). A note on non-negative ARMA processes. *J. Time Ser. Anal.* **28**, 350-360.

[35] Zhang, C. (2011). Parameter Estimation for First-order Bifurcating Autoregressive Processes with Weibull Type Innovations. *Statistics and Probability Letters*, **81**, 1961 - 1969.

[36] Zhou, J. Basawa, I.V. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. Stochastic methods and their applications. *J.Appl. Probab.*, **41A**, 55 - 64.

[37] Zhou, J. Basawa, I.V. (2005a). Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74**, 77 - 88.

[38] Zhou, J. Basawa, I.V. (2005b). Maximum Likelihood Estimation for a First-Order Bifurcating Autoregressive Process with Exponential Errors. *J. Time Series Anal.*, **26**, 825 - 842.

Chapter 2

Estimation for Nonnegative first-order autoregressive processes with an unknown location parameter and heavy-tail innovations [2]

[2]Bartlett A. and McCormick W.P. 2012. *Applied Mathematics*, **3 No. 12A**, 2133 - 2147. Reprinted here with permission of the publisher.

Abstract

Consider a first-order autoregressive processes $X_t = \phi X_{t-1} + Z_t$, where the innovations are nonnegative random variables with regular variation at both the right endpoint infinity and the unknown left endpoint θ. We propose estimates for the autocorrelation parameter ϕ and the unknown location parameter θ by taking the ratio of two sample values chosen with respect to an extreme value criteria for ϕ and by taking the minimum of $X_t - \hat{\phi}_n X_{t-1}$ over the observed series, where $\hat{\phi}_n$ represents our estimate for ϕ. The joint limit distribution of the proposed estimators is derived using point process techniques. A simulation study is provided to examine the small sample size behavior of these estimates.

AMS 1980 *subject classifications.* Primary 62M10; secondary 62E20, 60F05.

Keywords: nonnegative time series, autoregressive processes, extreme value estimator, regular variation, point processes

2.1 Introduction

In many applications, the desire to model the phenomena under study by nonnegative dependent processes has increased. An excellent presentation of the classical theory concerning these models can be found, for example, in Brockwell and Davis (1987). Recently, advancements in such models have shifted focus to some specialized features of the model, e.g. heavy tail innovations or nonnegativity of the model. In this chapter we examine the behavior of traditional estimates under conditions leading to non-Gaussian limits. For example, the standard approach to parameter estimation within the AR(1) process is through the Yule-Walker estimator;

$$\hat{\phi}_{LS} = \frac{\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}, \quad \text{where } \bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{2.1.1}$$

A slightly different approach presented in McCormick and Mathew (1993) used linear programming to obtain estimates for ϕ and θ under certain optimization constraints. While there are many established methods to estimate the autocorrelation coefficient in an AR(1) model, there are just a few approaches on estimating the unknown location parameter in an AR(1) model. One, was mentioned in McCormick and Mathew (1993) where they considered

$$\hat{\theta}_{range} = X_{j^*} - \hat{\phi}_{range} X_{j^*-1}, \quad \text{where} \quad \hat{\phi}_{range} = \frac{X_{t^*+1} - X_{j^*}}{X_{t^*} - X_{j^*-1}},$$

t^* and j^* provides the index of the maximal and minimal X_i respectively for $1 \leq i \leq n$.

In this paper we examine estimation questions and asymptotic properties of alternative estimates for ϕ and θ respectively, relating to the model

$$X_t = \phi X_{t-1} + Z_t, \quad t \geq 1, \tag{2.1.2}$$

where $0 < \phi < 1$, $\theta > 0$ and $\{Z_t\}$ is an i.i.d. sequence of nonnegative random variables whose innovation distribution F is assumed to be regularly varying at infinity with index

$-\beta$ and regularly varying at θ with index α, where θ, denotes the unknown but positive left endpoint. As a result of not restricting the innovations $\{Z_t\}$ to be bounded on a finite range, we can first estimate the autoregressive parameter ϕ through regular variation at infinity and then estimate the positive but unknown location parameter through regular variation at θ, the left endpoint.

While we have mentioned a few established estimation procedures, one notable exception was that of maximum likelihood. Although typically intractable and intricate in the time series setting, when the innovations in the AR(1) model are exponential, the maximum likelihood procedure had a major contribution on the estimation of positive heavy tailed time series. With these considerations in mind, Raftery (1980) determined the limiting distribution of the maximum likelihood estimate for the autocorrelation coefficient ϕ. As a result, the estimator

$$\hat{\phi}_n = \bigwedge_{t=1}^{n} \frac{X_t}{X_{t-1}}, \tag{2.1.3}$$

was considered. The realization of this estimator was the stepping stone for the work done in this paper along with Davis and McCormick (1989) which first considered this alternative estimator and used a point process approach to obtain the asymptotic distribution of the natural estimator $\hat{\phi}_n$. This was done in the context that the innovations distribution F varies regularly at 0, the left endpoint, and satisfies some moment condition.

The work presented in this paper is an extension of the work done in Davis and McCormick (1989) including the following contributions to dependent time series with heavy-tail innovations. The first contribution involves the development of estimates for the autocorrelation coefficient and unknown location parameter under regular variation at both endpoints, with a rate of convergence $n^{1/\beta}\ell(n)$, where ℓ is slowly varying function. The second contribution involves using an extreme value method, e.g. point processes to establish the asymptotic distribution of the proposed estimators and weak convergence for the asymptotically independent joint distribution. An initial observation is that our estimation procedure is especially easy to implement for both ϕ and θ. That is, the autoregressive coefficient ϕ in the causal AR(1) process is estimated by taking the minimum of the ratio of two sample values while estimation for the unknown location parameter θ was achieved through minimizing $X_t - \hat{\phi}_n X_{t-1}$ over the observed series.

This naturally motivates a comparison between the estimation procedure presented in this paper and the standard linear programming estimates mentioned above, since within a nonnegative AR(1) model the linear programming estimate reduces to the estimate proposed, namely, $\min_{1 \leq t \leq n}(X_t/X_{t-1})$, where $\{X_t\}$ denotes the AR(1) process. This comparison along with the comparison between McCormick and Mathew (1993) optimization method and Bartlett and McCormick (2012) extreme value method was performed through simulation and is presented in Section 2.3. The results found appear to demonstrate a favorable performance for our extreme value method over the three alternative estimators.

The main proofs in this paper rely heavily on point process methods from extreme value theory. The essential idea is to first establish the convergence of a sequence of point processes based on simple quantities and then apply the continuous mapping theorem to obtain convergence of the desired statistics. More background information on point processes, regular variation, and weak convergence can be found in Resnick (1986). Also, a nice survey on linear programming estimation procedures and nonnegative time series can be found in Anděl (1989), Anděl (1991), and Datta and McCormick (1995), whereas more applications on modeling the phenomena with heavy tailed distributions and ensuing estimation issues can be found in Resnick (2007).

The rest of the paper is organized as follows: asymptotic limit results for the autocorrelation parameter ϕ, unknown location parameter θ, and joint distribution of (ϕ, θ) are presented in Section 2.2, while Section 2.3 is concerned with the small sample size behavior of these estimates through simulation.

2.2 Asymptotics

The following point process limit result presented in Lemma 2.2.1 is fundamental. It involves the convergence in distribution of a sequence of point processes, which is ultimately used to derive the limit distribution for $\hat{\phi}_n$. In turn, this will allow us under certain conditions, to increase the rate of convergence from the square root of n to b_n. The significance of the normalizing constant b_n is that it not only can help improve the rate of convergence of our estimator, but ensures that the limit distribution is non-degenerate. The following proof was greatly simplified by applying Theorem 4.7 of Kallenberg (1976), which makes use of dissecting covering semi–rings denoted by *DC-semiring*. In this paper, *DC-semiring* are essentially a family of intervals on the real line. In this regime, we assume that $\{X_t, t \geq 0\}$ is a unique stationary solution to (2.1.2) and is given by

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

By assuming that F is assumed to be regularly varying at infinity with index $-\beta$, we are considering time series with heavy-tailed errors and within certain time series applications a better model is achieved. Thus, our goal is to capitalize on the behavior of extreme value estimators over traditional estimators when $0 < \beta < 2$.

Lemma 2.2.1. *Let $\boldsymbol{Z}_t^{(m)} = (Z_t, Z_{t-1}, \ldots, Z_{t-m+1})$ and consider the point processes $\mathcal{I}_n^{(m)}$ and $\mathcal{I}^{(m)}$ on the space $E = (0, \infty]^m \times [\theta, \infty)$ defined by*

$$\mathcal{I}_n^{(m)} = \sum_{t=1}^{n} \epsilon_{(b_n^{-1}\boldsymbol{Z}_t^{(m)}, Z_{t+1})} \quad \text{and} \quad \mathcal{I}^{(m)} = \sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \epsilon_{(j_k \cdot \boldsymbol{e}_i, Z_{i,k})},$$

where $\sum_{k=1}^{\infty} \epsilon_{j_k}$ is $PRM(\nu)$ with $\nu[x, \infty) = x^{-\beta}, x > 0$, $\boldsymbol{e}_0 = (1, \ldots, 0), \ldots, \boldsymbol{e}_{m-1} = (0, \ldots, 1)$ are the usual basis vectors in $\mathbb{R}^m$ and $\{Z_{i,k}\}$ are i.i.d. with $Z_{i,k} \stackrel{d}{=} Z_1$ and further $\{Z_{i,k}\}$ is *independent of* $\{j_k, k \geq 1\}$*. Then in* $M_p(E)$ *we have*

$$\mathcal{I}_n^{(m)} \Rightarrow \mathcal{I}^{(m)}.$$

Proof. Suppose that $U \in \mathcal{U}$ where $\mathcal{U}$ is a *DC-ring* of bounded continuity sets. That is, $P[\mathcal{I}^{(m)}(\partial U) = 0] = 1$. Further suppose $A \in \mathcal{J}$, where $\mathcal{J}$ is a *DC-semiring* of bounded continuity sets with respect to $\mathcal{I}^{(m)}$. Since $\mathcal{I}^{(m)}$ is almost surely simple, we can establish weak convergence by applying Theorem 4.7 of Kallenberg (1976). With this result, it now suffices to show as $n \to \infty$

$$(C1) \quad P[\mathcal{I}_n^{(m)}(U) = 0] \to P[\mathcal{I}^{(m)}(U) = 0], \quad \text{for all } U \in \mathcal{U},$$

$$(C2) \quad E\mathcal{I}_n^{(m)}(A) \to E\mathcal{I}^{(m)}(A), \quad \text{for all } A \in \mathcal{J}.$$

To establish (C2), consider a vector $\mathbf{x} = (x_0, \ldots, x_{m-1})$ and real y. We write $\mathbf{z} = (\mathbf{x}, y)$ to denote $\mathbf{z} \in \mathbb{R}^{m+1}$. Now observe that the mean measure $E\mathcal{I}^{(m)}$ is concentrated on the set $S = \bigcup_{i=0}^{m-1}(\mathbb{R}_+ \cdot \mathbf{e}_i, [\theta, \infty))$. Thus, if $\mathbf{z} \in S$ then $\mathbf{z} \in \mathbb{R}^{m+1}$, and $z_i = 0$ for all $0 \leq i \leq m-1$ except if the two nonzero coordinates z_{i_0} and z_m with $0 \leq i_0 \leq m-1$ are both positive. Now let $A \subset S$ be Borel measurable. Then by writing $A = \bigcup_{i=0}^{m-1} A_i$ with $A_i = A \cap C_i$, where $C_i = \{(x \cdot \mathbf{e}_i, y) : x \in \mathbb{R}_+, y \in [\theta, \infty)\}$, for $0 \leq i \leq m-1$, we have

$$E\mathcal{I}^{(m)}(A) = \sum_{i=0}^{m-1} \int_{\pi_i(A)} d\nu \times dF,$$

where $\pi_i(A) = \{(x, y) : (0, \ldots, x, \ldots, 0, y) = (x \cdot \mathbf{e}_i, y) \in A\}$. Now, to compute the mean measure $E\mathcal{I}_n^{(m)}$ notice that $\mathcal{I}_n^{(m)}$ is a point process such that the limit point process has no points off the axes. Therefore, we consider taking the one-dimensional j_k's and laying them down on the axis $\mathbf{e}_0$, and then repeating deterministically this pattern on each axis $\mathbf{e}_1, \ldots, \mathbf{e}_{m-1}$. Let us begin by considering a set $A \subset \mathbb{R}_+^m \backslash \{\mathbf{0}\} \times [\theta, \infty)$ of the form

$$A = [x_0, \infty) \times \ldots \times [x_{m-1}, \infty) \times [\theta, y],$$

where $x_i \geq 0$ and $\sum_{i=0}^{m-1} x_i > 0$ and $y > \theta$. For the case when exactly one component of $\mathbf{x}$ is positive, we assume without loss of generality (WLOG), that A takes on the form $[x, \infty) \times [0, \infty) \times \ldots \times [0, \infty) \times [\theta, y]$. Then

$$\begin{aligned} E\mathcal{I}_n^{(m)}(A) &= nP[\mathbf{Z}_1^{(m)} \in [b_n x, \infty) \times [0, \infty) \times \ldots \times [0, \infty), Z_2 \leq y] \\ &\sim \frac{P[Z_1 \geq b_n x]}{P[Z_1 \geq b_n]} F(y) \\ &\sim \nu[x, \infty) F(y). \end{aligned} \tag{2.2.1}$$

For the case when exactly two components of $\mathbf{x}$ are positive, we assume (WLOG) that A takes on the form $[x_1, \infty) \times [x_2, \infty) \times [0, \infty) \times \ldots \times [0, \infty) \times [\theta, y]$. Then

$$\begin{aligned} E\mathcal{I}_n^{(m)}(A) &= nP[\mathbf{Z}_1^{(m)} \in [b_n x_1, \infty) \times [b_n x_2, \infty) \times [0, \infty) \times \ldots \times [0, \infty), Z_3 \leq y] \\ &\leq F(y) \frac{P[Z_1 \geq b_n s]}{P[Z_1 \geq b_n]} P[Z_2 \geq b_n s] \\ &\to F(y) s^{-\beta} \cdot 0 = 0, \end{aligned}$$

where $s = \min(x_1, x_2)$. Finally observe that

$$E\mathcal{I}^{(m)}(A) = \sum_{i=0}^{m-1} \int_{\pi_i(A)} d\nu \times dF = \int_{[x,\infty)\times[0,\infty)\times\ldots\times[0,\infty)\times[\theta,y]} d\nu \times dF = \nu[x, \infty) F(y).$$

Therefore (C2) holds since $E\mathcal{I}_n^{(m)}(A) \to E\mathcal{I}^{(m)}(A),$ for all $A \in \mathcal{J}$.

To establish (C1) we consider sets of the same form

$$U = \bigcup_{l=1}^{k} A_l, \quad \text{where} \quad A_l = [x_{(l,1)}, \infty) \times \ldots \times [x_{(l,m)}, \infty) \times [\theta, y_{(l,m+1)}]$$

are disjoint sets for $1 \leq l \leq k$. As we have seen in the proof for (C2), the limit process can have no points off the axes. Thus, we consider the case when exactly one component of $\mathbf{x}$ is positive. For simplicity, we suppose (WLOG) that U takes on the form

$$U = [0, \infty) \times [0, \infty) \times \ldots \times [x, \infty) \times [\theta, y].$$

Then

$$(\mathcal{I}_n^{(m)}(U) = 0) = \bigcap_{t=1}^{n} \left((Z_t, Z_{t+1}) \notin [b_n x, \infty) \times [\theta, y]\right). \tag{2.2.2}$$

In order to calculate the probability in (2.2.2) we consider the following blocking argument. Since the vectors (Z_t, Z_{t+1}) are 1-dependent we need only separate the blocks by 1 to achieve independence between blocks. Therefore, we define $r_n = \lfloor n/k \rfloor$ for large k, and $h = \lfloor n/r_n \rfloor$. Then for $i = 1, \ldots, h$ set

$$J_i = [(i-1)r_n + 1, \ldots, ir_n - 1] \quad \text{and} \quad J_i^{'} = \{ir_n\}.$$

Now we define the events

$$\chi_i = \{(Z_t, Z_{t+1}) \notin [b_n x, \infty) \times [\theta, y], \text{ for all } t \in J_i\}$$

and

$$\chi_i^{'} = \{(Z_{ir_n}, Z_{ir_n+1}) \notin [b_n x, \infty) \times [\theta, y]\}.$$

In this way, the events $\chi_i, i = 1, \ldots, h$ are independent and equiprobable. For notational purposes, we define $\tilde{E} = [b_n x, \infty) \times [\theta, y]$. Then we have

$$P[\chi_1] = 1 - P[\chi_1^c],$$

where $P[\chi_1^c] = P[(Z_t, Z_{t+1}) \in \tilde{E}, \text{ for some } t \in J_1]$. Now applying the Bonferroni inequality, we have

$$P[\chi_1^c] \leq (r_n - 1)P[(Z_1, Z_2) \in \tilde{E}],$$

which from (2.2.1) is asymptotically equivalent to $\frac{1}{k}x^{-\beta}F(y)$ as $n \to \infty$. $P[\chi_1^c]$ is also bounded below by

$$(r_n - 1)P[(Z_1, Z_2) \in \tilde{E}] - \sum_{1 \le s < t \le r_n - 1} P[\{(Z_s, Z_{s+1}) \in \tilde{E}\} \cap \{(Z_t, Z_{t+1}) \in \tilde{E}\}]$$

which as $n \to \infty$ is asymptotically equivalent to

$$\begin{aligned}
&\sim \frac{1}{k}x^{-\beta}F(y) - \sum_{s=1}^{r_n-2} P[\{(Z_s, Z_{s+1}) \in \tilde{E}\} \cap \{(Z_{s+1}, Z_{s+2}) \in \tilde{E}\}] \\
&\sim \frac{1}{k}x^{-\beta}F(y) - \frac{1}{2}(r_n - 3)(r_n - 2)P^2[(Z_1, Z_2) \in \tilde{E}] \\
&\sim \frac{1}{k}x^{-\beta}F(y) - \frac{1}{2k^2}x^{-2\beta}F^2(y).
\end{aligned}$$

Now recall that we are interested in calculating the probability that the point process $\mathcal{I}_n^{(m)}(U)$ contains no points, which from (2.2.2) is equivalent to $P[\bigcap_{t=1}^n (Z_t, Z_{t+1}) \notin \tilde{E}]$. Therefore, to complete the proof it suffices to show that

$$P\left[\bigcap_{t=1}^{n} (Z_t, Z_{t+1}) \notin \tilde{E}\right] = P[\chi_1]^h + o(1), \qquad \text{as} \qquad n \to \infty.$$

However,

$$\begin{aligned}
\bigcap_{t=1}^{n} (Z_t, Z_{t+1}) \notin \tilde{E} &= \bigcap_{i=1}^{h} \chi_i \cap \bigcap_{i=1}^{h} \chi_i' \cap \{(Z_t, Z_{t+1}) \notin \tilde{E}, \text{ for } t = r_n h + 1, \ldots, n\} \\
&\subset \bigcap_{i=1}^{h} \chi_i.
\end{aligned}$$

Thus, the difference of the two sets above is contained in

$$\bigcup_{i=1}^{h} \chi_i'^{c} \cup \{(Z_t, Z_{t+1}) \in \tilde{E}, \text{ for some } t = r_n h + 1, \ldots, n\}.$$

But this implies that the latter set has probability at most $2n/r_n P[(Z_1, Z_2) \in \tilde{E}]$. Thus, as $n \to \infty$ this set is asymptotically equivalent to

$$\sim \frac{2}{r_n} x^{-\beta} F(y) \to o(1).$$

Hence,

$$\begin{aligned}
\lim_{k\to\infty} \lim_{n\to\infty} P[\mathcal{I}_n^{(m)}(U) = 0] &= \lim_{k\to\infty} \lim_{n\to\infty} P\left[\bigcap_{t=1}^{n} (Z_t, Z_{t+1}) \notin \tilde{E}\right] \\
&= \lim_{k\to\infty} P\left[\bigcap_{i=1}^{k} \chi_i\right] + o(1) \\
&= \lim_{k\to\infty} P[\chi_1]^k + o(1) \\
&= \lim_{k\to\infty} \left(1 - \frac{1}{k} x^{-\beta} F(y) + o\left(\frac{1}{k}\right)\right)^k \\
&= e^{-x^{-\beta} F(y)}.
\end{aligned}$$

Therefore (C1) holds since

$$e^{-x^{-\beta} F(y)} = P[\mathcal{I}^{(m)}(U) = 0].$$

$\square$

Lemma 2.2.2. *Let ξ_n and ξ be the point processes on the space $E = (0, \infty] \times [\theta, \infty)$ defined by*

$$\xi_n = \sum_{t=1}^{n} \epsilon_{(b_n^{-1} X_{t-1}, Z_t)} \quad \textit{and} \quad \xi = \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \epsilon_{(\phi^i j_k, Z_{i,k})}$$

where $\sum_{k=1}^{\infty} \epsilon_{j_k}$ and $\{Z_{i,k}\}$ are as defined in the statement of Lemma 2.2.1. Then in $M_p(E)$,

$$\xi_n \Rightarrow \xi.$$

Proof. First observe that the points defined in ξ_n are not independently identically distributed. Therefore, to handle the dependence, we begin by fixing an integer m and defining

$$X_t^{(m)} = \sum_{i=0}^{m-1} \phi^i Z_{t-i}$$

as an approximation of X_t. Now we can think of $X_t^{(m)}$ as a simple functional of the vector $\mathbf{Z}_t^{(m)}$. Applying Lemma 2.2.1 to the point processes

$$\tilde{\mathcal{I}}_n^{(m)} = \sum_{t=1}^{n} \varepsilon_{(b_n^{-1}\mathbf{Z}_{t-1}^{(m)}, Z_t)} \quad \text{and} \quad \mathcal{I}^{(m)} = \sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \varepsilon_{(j_k \cdot \mathbf{e}_i, Z_{i,k})}.$$

we have that

$$\tilde{\mathcal{I}}_n^{(m)} \Rightarrow \mathcal{I}^{(m)}.$$

We now consider using a continuous mapping argument to obtain convergence for our point process ξ_n and then apply Slutsky's lemma to remove m. Hence, we define the map $T : (0,\infty]^m \times [\theta,\infty) \to (0,\infty] \times [\theta,\infty)$ given by

$$T(z,y) = \begin{cases} (\sum_{i=0}^{m-1} \phi^i z_{m-1-i}, y) & \text{if } \bigvee_{i=0}^{m-1} z_i < \infty; \\ (\infty, y) & \text{otherwise .} \end{cases}$$

Now if $q \in M_p(E)$ is a point measure of ξ_n, then the mapping $\hat{T} : M_p((0,\infty]^m \times [\theta,\infty)) \to M_p((0,\infty] \times [\theta,\infty))$ defined by

$$\hat{T}(q) = q \circ T^{-1}$$

can be applied to (Proposition 3.18 in Resnick (1987)) in order to obtain

$$\begin{aligned}
\sum_{t=1}^{n} \varepsilon_{(b_n^{-1} X_{t-1}^{(m)}, Z_t)} &= \sum_{t=1}^{n} \varepsilon_{(b_n^{-1} \sum_{i=0}^{m-1} \phi^i Z_{t-1-i}, Z_t)} \\
&= \hat{T}\left(\sum_{t=1}^{n} \varepsilon_{(b_n^{-1} \mathbf{Z}_{t-1}^{(m)}, Z_t)} \right) \\
&\Rightarrow \hat{T}\left(\sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \varepsilon_{(j_k \cdot \mathbf{e}_i, Z_{i,k})} \right) \\
&= \sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \varepsilon_{T(j_k \cdot \mathbf{e}_i, Z_{i,k})} \\
&\stackrel{d}{=} \sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \varepsilon_{(\phi^i j_k, Z_{i,k})}.
\end{aligned} \tag{2.2.3}$$

Now notice as $m \to \infty$

$$\sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \varepsilon_{(\phi^i j_k, Z_{i,k})} \to \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \varepsilon_{(\phi^i j_k, Z_{i,k})} \tag{2.2.4}$$

pointwise in the vague metric. In order to complete the proof, we need to show that the point process

$$\sum_{t=1}^{n} \varepsilon_{(b_n^{-1} X_{t-1}^{(m)}, Z_t)} \quad \text{is equivalent to} \quad \sum_{t=1}^{n} \varepsilon_{(b_n^{-1} X_{t-1}, Z_t)}.$$

Equations (2.2.3), (2.2.4), and Lemma 4.25 in Resnick (1987) show that it is enough to prove for any $\eta > 0$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left[d\left(\sum_{t=1}^{n} \varepsilon_{(b_n^{-1} X_{t-1}^{(m)}, Z_t)}, \sum_{t=1}^{n} \varepsilon_{(b_n^{-1} X_{t-1}, Z_t)} \right) > \eta \right] = 0 \tag{2.2.5}$$

where d is the vague metric described in (1.2.8) as a complete separable metric space. Now, applying Theorem 4.2 in Billingsley (1968) and the definition of the vague metric, proving

(2.2.5) is equivalent to checking for all $\eta > 0$ and $f \in C_K^+(E_1)$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left[\sum_{t=1}^{n} f(b_n^{-1}X_{t-1}^{(m)}, Z_t) - \sum_{t=1}^{n} f(b_n^{-1}X_{t-1}, Z_t) > \eta\right] = 0. \tag{2.2.6}$$

To verify this, let $\gamma > 0$ and without loss of generality suppose the compact support of f is contained in $\{x \in (0,\infty] : x > \gamma\} \times [\theta, y]$. Since f has compact support it is uniformly continuous and given $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x,z) - f(w,z)| \leq \epsilon$$

whenever $|x - w| < \delta$. Thus, $\epsilon \to 0$ as $\delta \to 0$. After all the first component of f is of interest, hence we consider decomposing the probability in (2.2.6) according to whether

$$V := b_n^{-1} \bigvee_{t=1}^{n} |X_{t-1}^{(m)} - X_{t-1}| < \delta \quad \text{or} \quad V^c := b_n^{-1} \bigvee_{t=1}^{n} |X_{t-1}^{(m)} - X_{t-1}| \geq \delta$$

occurs.

Case 1: Suppose that V occurs

Assuming $\delta < \gamma/2$, if $b_n^{-1}X_{t-1}^{(m)} \leq \gamma/2$ then $f(b_n^{-1}X_{t-1}^{(m)}, Z_t) = f(b_n^{-1}X_{t-1}, Z_t) = 0$. However, if $b_n^{-1}X_{t-1}^{(m)} > \gamma/2$, then the probability in (2.2.6) is bounded above by

$$P\left[\epsilon \sum_{t=1}^{n} \varepsilon_{(b_n^{-1}X_{t-1}^{(m)}, Z_t)}(\{x : x > \gamma/2\} \times [\theta, y]) > \eta\right],$$

and as a result of (2.2.3) as $n \to \infty$ this converges to

$$P\left[\sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \varepsilon_{(\phi^i j_k, Z_{i,k})}(\{x : x > \gamma/2\} \times [\theta, y]) > \eta/\epsilon\right].$$

Therefore, as $m \to \infty$ we have

$$P\left[\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\varepsilon_{(\phi^i j_k, Z_{i,k})}(\{x : x > \gamma/2\} \times [\theta, y]) > \eta/\epsilon\right].$$

By choosing $\epsilon > 0$ small, the preceding probability goes to zero as $\delta \to 0$ since the point process N has radon measure on the compact set describe above.

Case 2: Suppose that V^c occurs

Note by stationarity

$$\begin{aligned}
P[V^c] &\leq nP\left[|X_{t-1}^{(m)} - X_{t-1}| > b_n\delta\right] \\
&= nP\left[\sum_{i=m}^{\infty}\phi^i Z_{t-1-i} > b_n\delta\right].
\end{aligned}$$

Now applying Lemma 4.24 in Resnick (1987) this is asymptotic equivalent to

$$\sim \delta^{-\beta}\sum_{i=m}^{\infty}\phi^{i\beta}.$$

Therefore,

$$\begin{aligned}
&\lim_{m\to\infty}\limsup_{n\to\infty} P\left[\sum_{t=1}^{n} f(b_n^{-1}X_{t-1}^{(m)}, Z_t) - \sum_{t=1}^{n} f(b_n^{-1}X_{t-1}, Z_t) > \eta; V^c\right] \\
&\leq \lim_{m\to\infty}\delta^{-\beta}\sum_{i=m}^{\infty}\phi^{i\beta} = 0.
\end{aligned}$$

This completes the proof since the point process

$$\sum_{t=1}^{n}\varepsilon_{(b_n^{-1}X_{t-1}^{(m)}, Z_t)} \text{ is equivalent to } \sum_{t=1}^{n}\varepsilon_{(b_n^{-1}X_{t-1}, Z_t)}.$$

□

Remark. Unlike the situation when the points on different axes are independent this situation is very different since the pattern of points on each axis is the same, creating dependence among points. Therefore, the limit process ξ is not Poisson random measure and the reason for this has to do more with the point process $\xi_n^{(m)}$ than ξ. That is, the continuous mapping theorem bridged the gap between $\mathcal{I}_n^{(m)}$ and ξ_n, where $\mathcal{I}^{(m)}$ was obtained by taking the one-dimensional j_k's and laying them down on axis $\mathbf{e}_0$ and then repeating deterministically this pattern on each axis $\mathbf{e}_1, \dots, \mathbf{e}_{m-1}$. Therefore, if the largest innovation Z_t occurs on axis $\mathbf{e}_{m-1}$, then we have a mark j_k in the $m-1$ location for the point Z_{t+1}, but if the large innovation occurs on axis $\mathbf{e}_{m-2}$, then we have the same mark j_k in the $m-2$ location for a different point Z_{t+2}. Hence, we obtain a collection of points in ξ on different axes which are not independent. To resolve this problem, we consider the point process $\mathcal{N} = \sum_{k=1}^{\infty} \varepsilon_{(j_k, W_k)}$, which is $PRM(\nu_1 \times G)$ and $W_k = \bigwedge_{i=0}^{\infty} \phi^{-i} Z_{i,k}$ is an i.i.d. sequence of random variables, independent of $\sum_{k=1}^{\infty} \varepsilon_{j_k}$ with $G(w) = P[W_1 \leq w]$. This then implies for all sets of the form $Q_r = \{(g,h) : \frac{h}{g} \leq r, g, h > 0\}$ that

$$P[\xi(Q_r) = 0] \equiv P[\mathcal{N}(Q_r) = 0].$$

Theorem 2.2.1. *Let $\{X_t, t \geq 0\}$ be the stationary solution to the AR(1) recursion $X_t = \phi X_{t-1} + Z_t$ and consider the estimator $\hat{\phi}_n = \bigwedge_{t=1}^{n} \frac{X_t}{X_{t-1}}$ for ϕ. Under the assumption that $0 \leq \phi < 1$ and the innovation distribution F has regularly varying right tail with index $-\beta$ and finite positive left endpoint θ, then*

$$\lim_{n \to \infty} P[b_n(\hat{\phi}_n - \phi) > x] = e^{-x^{\beta} E W^{-\beta}} \qquad \text{for all } x > 0,$$

where $W = \bigwedge_{i=0}^{\infty} \frac{Z_i}{\phi^i}$ and $b_n = F^{\leftarrow}(1 - 1/n)$.

Proof. Since

$$P[b_n(\hat{\phi}_n - \phi) > x] = P\left[\bigwedge_{t=1}^{n} \frac{Z_t}{b_n^{-1} X_{t-1}} > x\right], \tag{2.2.7}$$

let us define on a subset of $\mathbb{R}_+^2$, $Q_x = \{(x_1, x_2) : \frac{x_2}{x_1} \leq x, x_1, x_2 > \theta\}$. Then it suffices to show that there are no points t that satisfies the condition in Q_x. Thus to complete the proof, let $x_2 = Z_t$ and $x_1 = b_n^{-1} X_{t-1}$, then notice that (2.2.7) is equivalent to $P[\xi_n(Q_x) = 0]$. Furthermore, observe that Q_x is a bounded set in $E = (0, \infty] \times [\theta, \infty)$ provided $\theta > 0$. Therefore assuming $\phi > 0$ and applying Lemma 2.2.2 we have

$$\begin{aligned} P[\xi_n(Q_x) = 0] \Rightarrow P[\xi(Q_x) = 0] \\ &= P\left[\bigwedge_{i=0}^{\infty} \frac{Z_{i,k}}{\phi^i j_k} > x\right] \\ &= P[\mathcal{N}(Q_x) = 0] \\ &= P\left[\sum_{k=1}^{\infty} \varepsilon_{(j_k, W_k)}(Q_x) = 0\right]. \end{aligned}$$

But $\mathcal{N} = \sum_{k=1}^{\infty} \varepsilon_{(j_k, W_k)} \sim PRM(\nu_1 \times G)$, where

$$\begin{aligned} \nu_1 \times G(Q_x) = \int_\theta^\infty \nu_1[w/x, \infty) dG(w) &= \int_\theta^\infty \left(\frac{w}{x}\right)^{-\beta} dG(w) \\ &= x^\beta \int_\theta^\infty w^{-\beta} dG(w) \\ &= x^\beta E W^{-\beta}. \end{aligned}$$

Therefore,

$$\begin{aligned} \lim_{n\to\infty} P[b_n(\hat{\phi}_n - \phi) > x] &= \lim_{n\to\infty} P[\xi_n(Q_x) = 0] \\ &= P[\xi(Q_x) = 0] \\ &= P[\mathcal{N}(Q_x) = 0] \\ &= e^{-x^\beta E W^{-\beta}}. \end{aligned}$$

$\square$

The previous proof relied on point processes and dissecting covering semi–rings to establish weak convergence of ξ_n and the limiting distribution of $\hat{\phi}_n$. The following alternative proof, makes the necessary adjustments to the proof presented in Theorem 2.4 in Davis and Resnick (1985) to achieve the same limiting distribution for $\hat{\phi}_n$. Unlike the previous result, which makes no use of an ARMA structure and only applies to autoregressive processes of order one, this result applies to general linear models subject to usual summability conditions on the coefficients. In that regard for this result, we assume that $\{X_n, n \geq 0\}$ is the stationary linear process given by

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}$$

with $\sum_{j=0}^{\infty} |c_j|^\delta < \infty$ for some $0 < \delta < \beta, \delta \leq 1$. Furthermore for this result we may relax our assumptions on the innovation distribution and we require that $|Z_1|$ has a regularly varying tail distribution, i.e. $P(|Z_1| > x) = x^{-\beta}\ell(x), x > 0$ for a slowly varying function ℓ and the innovation distribution is tail balanced

$$\frac{P(Z_1 > x)}{P(|Z_1| > x)} \to p \quad \text{and} \quad \frac{P(Z_1 \leq -x)}{P(|Z_1| > x)} \to q \quad \text{as} \quad x \to \infty.$$

Define point processes

$$\xi_n = \sum_{k=1}^{n} \epsilon_{(b_n^{-1} X_{k-1}, Z_k)}, \quad n \geq 1$$

and let $\sum_{k \geq 1} \epsilon_{j_k}$ denote PRM(ν) on $\mathbb{R}_0 = \mathbb{R} \backslash \{0\}$ where ν has Radon-Nikodym derivative with respect to Lebesgue measure

$$\frac{d\nu}{d\lambda}(x) = p\beta x^{-\beta-1} \mathbb{1}_{(0,\infty)}(x) + q\beta |x|^{-\beta-1} \mathbb{1}_{(-\infty,0)}(x).$$

Let $\{Z_{i,k}, i \geq 0, k \geq 1\}$ be an iid array with $Z_{i,k} \stackrel{d}{=} Z_1$ and independent of $\sum_{k=1}^{\infty} \epsilon_{j_k}$. Define

$$\xi = \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \epsilon_{(c_i j_k, Z_{i,k})}.$$

Our basic result is to show that $M_p(\mathbb{R}_0 \times \mathbb{R})$ equipped with the topology of vague convergence

$$\xi_n \xrightarrow{d} \xi \text{ in } M_p$$

which is close in statement and spirit to Theorem 2.4 in Davis and Resnick (1985). In view of the commonality of the two results, we present only the needed changes to the Davis and Resnick proof to accommodate the current setting. Aside from keeping track of the time when points occur, i.e. large jumps, the difference in the point processes considered here with those in Davis and Resnick (1985) is the inclusion of marks, i.e. the second component of the point $(b_n^{-1}X_{k-1}, Z_k)$. This complication induces an additional weak dependence in the points which is addressed in Lemma 2.2.4 through a straight forward blocking argument. First, we establish weak convergence of marked point processes of a normalized vector of innovations. For a positive integer m define

$$\mathbf{Z}_k^{(m)} = (Z_k, Z_{k-1}, \ldots, Z_{k-m+1})$$

and point process

$$\mathcal{I}_n^{(m)} = \sum_{k=1}^{n} \epsilon_{(b_n^{-1}\mathbf{Z}_k^{(m)}, Z_{k+1})}, \quad n \geq 1.$$

Let $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0), 1 \leq i \leq m$ denote the standard basis vectors for $\mathbb{R}^m$. Define an associated marked point process with the first component placed on an axis by

$$\tilde{\mathcal{I}}_n^{(m)} = \sum_{k=1}^{n} \sum_{i=1}^{m} \epsilon_{(b_n^{-1}Z_k\mathbf{e}_i, Z_{k+i})}.$$

In the following Lemma, we show that $\mathcal{I}_n^{(m)}$ and $\tilde{\mathcal{I}}_n^{(m)}$ are asymptotically indistinguishable in the following sense. Let $\mathbb{R}_0^m = \mathbb{R}^m \backslash \{\mathbf{0}\}$ and $E = \mathbb{R}_0^m \times \mathbb{R}$. Consider the class of rectangles

$$\mathcal{S} = \{R \times (-\infty, x] : R = \bigtimes_{i=1}^{m} (c_i, d_i], \;\; x < \infty \text{ and } \bar{R} \times (-\infty, x] \subset E\}.$$

Lemma 2.2.3. *As n tends to infinity,* $\mathcal{I}_n^{(m)}(B) - \tilde{\mathcal{I}}_n^{(m)}(B) \xrightarrow{p} 0 \quad$ *for all* $B \in \mathcal{S}$.

Proof. Following the proof presented in Proposition 2.1 of Davis and Resnick (1985), suppose that $B \in \mathcal{S}$ is such that for some $1 \leq i' \leq m, B \cap (\mathbb{R}\mathbf{e}_{i'} \times \mathbb{R}) \neq \varnothing$. As noted in Davis and

Resnick (1985) for all $i \neq i'$, one has $c_i < 0 < d_i$. Observe that

$$\begin{aligned} \mathcal{I}_n^{(m)}(B) &\leq \mathcal{I}_n^{(m)}(\mathbb{R} \times \ldots \times [c_{i'}, d_{i'}] \times \ldots \times \mathbb{R} \times (-\infty, x]) \\ &= \sum_{k=-i'+2}^{n-i'+1} \epsilon_{(b_n^{-1} Z_k \mathbf{e}_{i'}, Z_{k+i'})}(\mathbb{R} \times \ldots \times [c_{i'}, d_{i'}] \times \ldots \times \mathbb{R} \times (-\infty, x]). \end{aligned} \tag{2.2.8}$$

Similarly let $c = \bigvee_{i \neq i'} c_i < 0$ and $d = \bigwedge_{i \neq i'} d_i > 0$. Then

$$\begin{aligned} \mathcal{I}_n^{(m)}(B) &\geq \mathcal{I}_n^{(m)}([c,d] \times \ldots \times [c_{i'}, d_{i'}] \times \ldots \times [c,d] \times (-\infty, x]) \\ &\geq \mathcal{I}_n^{(m)}(\mathbb{R} \times \ldots \times [c_{i'}, d_{i'}] \times \ldots \times \mathbb{R} \times (-\infty, x]) - \sum_{i \neq i'} \mathcal{I}_n^{(m)}(E_i \times (-\infty, x]) \end{aligned} \tag{2.2.9}$$

where $\mathbf{y} = (y_1, \ldots, y_m) \in E_i$ according to $y_i \notin [c,d]$ and $y_{i'} \in [c_{i'}, d_{i'}]$. Note that

$$E[\mathcal{I}_n^{(m)}(E_i \times (-\infty, x])] \leq nP\left(\frac{1}{b_n}|Z_1| > |c| \wedge d, \frac{1}{b_n}|Z_1| > |c_{i'}| \wedge |d_{i'}|\right) = o(1). \tag{2.2.10}$$

Thus from (2.2.8) - (2.2.10) we obtain

$$\begin{aligned} E|\mathcal{I}_n^{(m)}(B) - \tilde{\mathcal{I}}_n^{(m)}(B)| &\leq 2(i'-1)P\left(\frac{1}{b_n} Z_1 > |c_{i'}| \wedge |d_{i'}|\right) + o(1) \\ &= o(1) \quad \text{as } n \to \infty. \end{aligned} \tag{2.2.11}$$

Then the result follows as in Davis and Resnick (1985), Proposition 2.1, completing the proof. $\square$

Lemma 2.2.4. *Let $\mathcal{V}_n^{(m)}$ and $\mathcal{V}^{(m)}$ be the point processes on the space $E = \mathbb{R}_0 \times \mathbb{R}^m$ defined by*

$$\mathcal{V}_n^{(m)} = \sum_{k=1}^{n} \epsilon_{(b_n^{-1} Z_k, \mathbf{Z}_{k+m}^{(m)})} \quad \text{and} \quad \mathcal{V}^{(m)} = \sum_{k=1}^{\infty} \epsilon_{(j_k, \mathbf{Y}_k^{(m)})}$$

where $\{\mathbf{Y}_k^{(m)}, k \geq 1\}$ is an iid sequence with $\mathbf{Y}_1^{(m)} \stackrel{d}{=} (Z_1, \ldots, Z_m)$ and is independent of $\sum_{k=1}^{\infty} \epsilon_{j_k}$. Then in $M_p(E)$,

$$\mathcal{V}_n^{(m)} \xrightarrow{d} \mathcal{V}^{(m)}.$$

Proof. We employ a blocking argument to establish this result. Let r_n be a sequence of integers such that $r = r_n \to \infty$ as $n \to \infty$ and $r = o(n)$. Let $h = \lfloor n/r \rfloor$ and $l \geq m$. Define blocks

$$I_{n,s} = [r(s-1)+1, \ldots, rs-l], \qquad J_{n,s} = [rs-l+1, \ldots, rs] \text{ for } 1 \leq s \leq h$$

and

$$J_{n,h+1} = [rh+1, \ldots, n].$$

Then it is clear that for $s \neq t$

$$\sigma(Z_{k+i}, k \in I_{n,s}, 0 \leq i \leq m) \perp\!\!\!\perp \sigma(Z_{k+i}, k \in I_{n,t}, 0 \leq i \leq m).$$

Write

$$\mathcal{V}_{n,s}^{(m)} = \sum_{k \in I_{n,s}} \epsilon_{(b_n^{-1} Z_k, \mathbf{Z}_{k+m}^{(m)})} \quad \text{and} \quad \tilde{\mathcal{V}}_{n,t}^{(m)} = \sum_{k \in J_{n,t}} \epsilon_{(b_n^{-1} Z_k, \mathbf{Z}_{k+m}^{(m)})}.$$

Then

$$\mathcal{V}_n^{(m)} = \sum_{s=1}^{h} \mathcal{V}_{n,s}^{(m)} + \sum_{t=1}^{h+1} \tilde{\mathcal{V}}_{n,t}^{(m)}. \tag{2.2.12}$$

Let $B \subset E$ be a disjoint union of rectangles

$$B = \bigcup_{i=1}^{j} B_i \tag{2.2.13}$$

where $B_i = [c_i, d_i] \times R_i$ with $R_i = \bigtimes_{l=1}^{m} [x_{il}, y_{il}]$. Let $\mu = \nu \times F \times \ldots \times F$ denote the mean measure of $\mathcal{V}^{(m)}$ which is PRM(μ) on E. To complete the proof we first show that for all sets B of the form given in (2.2.13) that

$$\lim_{n \to \infty} P\left(\mathcal{V}_n^{(m)}(B) = 0\right) = \exp\left(-\mu(B)\right). \tag{2.2.14}$$

The above limit result follows from the easily verifiable relations:

$$P\left(\mathcal{V}_{n,1}^{(m)}(B) = \ldots = \mathcal{V}_{n,h}^{(m)}(B) = 0\right) = P^h\left(\mathcal{V}_{n,1}^{(m)}(B) = 0\right); \tag{2.2.15}$$

$$P\left(\mathcal{V}_{n,s}^{(m)}(B_k) \wedge \mathcal{V}_{n,s}^{(m)}(B_l) \geq 1\right) = O\left(\frac{r^2}{n^2}\right) \text{ for } 1 \leq k \neq l \leq j;$$

$$P\left(\mathcal{V}_{n,s}^{(m)}\left(\bigcup_{i=1}^{j} B_i\right) \geq 1\right) = \sum_{i=1}^{j} P\left(\mathcal{V}_{n,s}^{(m)}(B_i) \geq 1\right) + O\left(\frac{r^2}{n^2}\right); \tag{2.2.16}$$

$$P\left(\mathcal{V}_{n,1}^{(m)}(B_i) \geq 1\right) = \frac{r}{n}\mu(B_i)(1 + o(1)); \tag{2.2.17}$$

and

$$\sum_{t=1}^{h+1} \tilde{\mathcal{V}}_{n,t}^{(m)}(B) \xrightarrow{p} 0 \text{ as } n \to \infty. \tag{2.2.18}$$

Indeed, in view of (2.2.12) and (2.2.18), (2.2.14) is equivalent to showing

$$\lim_{n\to\infty} P\left(\sum_{s=1}^{h} \mathcal{V}_{n,s}^{(m)}(B) = 0\right) = \exp\left(-\mu(B)\right) \tag{2.2.19}$$

and the above relation holds by (2.2.15), (2.2.16), and (2.2.17), viz.

$$\begin{aligned} P\left(\sum_{s=1}^{h} \mathcal{V}_{n,s}^{(m)}(B) = 0\right) &= \left(1 - P\left(\mathcal{V}_{n,1}^{(m)}(B) \geq 1\right)\right)^h \\ &= \left(1 - \sum_{i=1}^{j} \frac{r}{n}\mu(B_i) + o(r/n)\right)^h \\ &\to \exp\left(-\sum_{i=1}^{j} \mu(B_i)\right). \end{aligned}$$

It is immediate that for a rectangle $B = [c,d] \times \bigtimes_{i=1}^{m} [x_i, y_i] \subset E$ we have

$$\lim_{n\to\infty} E[\mathcal{V}_n^{(m)}(B)] = \mu(B). \tag{2.2.20}$$

Therefore the result is seen to hold by (2.2.14) and (2.2.20) by application of Theorem 4.7 in Kallenberg (1976). $\square$

Lemma 2.2.5. *Let $\mathcal{I}_n^{(m)}$ and $\mathcal{I}^{(m)}$ be point processes on the space $E = \mathbb{R}_0^{(m)} \times \mathbb{R}$*

$$\mathcal{I}_n^{(m)} = \sum_{k=1}^{n} \epsilon_{(b_n^{-1}\mathbf{Z}_k^{(m)}, Z_{k+1})}, \quad n \geq 1 \quad \text{and} \quad \mathcal{I}^{(m)} = \sum_{k=1}^{\infty}\sum_{i=1}^{m} \epsilon_{(j_k \mathbf{e}_i, Z_{i,k})}.$$

Then in $M_p(E)$,

$$\mathcal{I}_n^{(m)} \xrightarrow{d} \mathcal{I}^{(m)}.$$

Proof. We begin by applying the argument used in Theorem 2.2 of Davis and Resnick (1985) with the modification that the relevant composition of maps of point processes is given by

$$\begin{aligned}\sum_{k\geq 1} \epsilon_{(u_k, v_{mk}, \ldots, v_{1k})} &\mapsto \left(\sum_{k\geq 1} \epsilon_{(u_k, v_{mk})}, \ldots, \sum_{k\geq 1} \epsilon_{(u_k, v_{1k})} \right) \\ &\mapsto \left(\sum_{k\geq 1} \epsilon_{(u_k \mathbf{e}_1, v_{1k})}, \ldots, \sum_{k\geq 1} \epsilon_{(u_k \mathbf{e}_m, v_{mk})} \right) \mapsto \sum_{k\geq 1}\sum_{i=1}^{m} \epsilon_{(u_k \mathbf{e}_i, v_{ik})}.\end{aligned}$$

Each map being continuous, the composition is a continuous map from $M_p(\mathbb{R}_0 \times \mathbb{R}^m)$ to $M_p(\mathbb{R}_0^{(m)} \times \mathbb{R})$ with each space being equipped with the topology of vague convergence. Therefore by the continuous mapping theorem and Lemma 2.2.4 we obtain

$$\tilde{\mathcal{I}}_n^{(m)} = \sum_{k=1}^{\infty}\sum_{i=1}^{m} \epsilon_{(b_n^{-1} Z_k \mathbf{e}_i, Z_{k+i})} \xrightarrow{d} \sum_{k=1}^{\infty}\sum_{i=1}^{m} \epsilon_{(j_k \mathbf{e}_i, Z_{i,k})} = \mathcal{I}^{(m)}. \tag{2.2.21}$$

Finally we complete the proof by Lemma 2.2.3 and (2.2.21) arguing as in Davis and Resnick (1985). $\square$

We are now ready to present our fundamental result.

Theorem 2.2.2. *Let ξ_n and ξ be the point processes on the space $E = \mathbb{R}_0 \times \mathbb{R}$ defined by*

$$\xi_n = \sum_{k=1}^{n} \epsilon_{(b_n^{-1} X_{k-1}, Z_k)}, \quad n \geq 1 \quad \text{and} \quad \xi = \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \epsilon_{(c_i j_k, Z_{i,k})}$$

where $\sum_{k=1}^{\infty} \epsilon_{j_k}$ is $PRM(\nu)$ and $\{Z_{i,k}, i \geq 0, k \geq 1\}$ is an iid array with $Z_{i,k} \stackrel{d}{=} Z_1$ and independent of $\sum_{k=1}^{\infty} \epsilon_{j_k}$. Then in $M_p(E)$,

$$\xi_n \stackrel{d}{\to} \xi.$$

Proof. **Remark.** Apart from considering a time coordinate and restricting the process to an AR(1) process, the above Theorem 2.2.2 and Theorem 3.1 in McCormick and Mathew (1993) consider essentially the same point process limit result. However, their result gave a wrong limit point process. This error is corrected in the current paper.

Observe that the map

$$(z_k, z_{k-1}, \ldots, z_{k-m+1}) \mapsto \sum_{i=0}^{m-1} c_i z_{k-i}$$

induces a continuous map on point processes given by

$$\sum_{k=1}^{\infty} \epsilon_{(z_k, z_{k-1}, \ldots, z_{k-m+1}, z_{k+1})} \mapsto \sum_{k=1}^{\infty} \epsilon_{(\sum_{j=0}^{m-1} c_j z_{k-j}, z_{k+1})}.$$

Thus we obtain from Lemma 2.2.5 that

$$\sum_{k=1}^{n} \epsilon_{(b_n^{-1} \sum_{i=0}^{m-1} c_i Z_{k-i}, Z_{k+1})} \stackrel{d}{\to} \sum_{k=1}^{\infty} \sum_{i=0}^{m-1} \epsilon_{(c_i j_k, Z_{i,k})}. \tag{2.2.22}$$

The result now follows from (2.2.22) by the same argument in Davis and Resnick (1985) to finish their Theorem 2.4. ∎

Returning to the AR(1) model under discussion in this paper and the estimate $\hat{\phi}_n$ given in (2.1.3), we obtain the following asymptotic limit result.

Theorem 2.2.3. *Let $\{X_t, t \geq 0\}$ be the stationary solution to the AR(1) recursion $X_t = \phi X_{t-1} + Z_t$ and $\hat{\phi}_n = \bigwedge_{t=1}^n \frac{X_t}{X_{t-1}}$. Under the assumptions that $0 < \phi < 1$ and the innovation distribution F has regularly varying right tail with index $-\beta$ and finite positive left endpoint θ,*

$$\lim_{n\to\infty} P\big(b_n(\hat{\phi}_n - \phi) > x\big) = e^{-x^\beta EW^{-\beta}} \qquad \textit{for all } x > 0,$$

where $W = \bigwedge_{i=0}^\infty \frac{Z_i}{\phi^i}$ and $b_n = F^{\leftarrow}(1 - 1/n)$.

Proof. For $x > 0$ define a subset

$$Q_x = \{(x_1, x_2) : \frac{x_2}{x_1} \leq x, x_1 > 0, x_2 \geq \theta\}.$$

Then note that for the point processes $\xi_n = \sum_{k=1}^n \epsilon_{(b_n^{-1}X_{k-1}, Z_k)}$, we have

$$\Big\{\xi_n\big(Q_x\big) = 0\Big\} = \Big\{\bigwedge_{k=1}^n \frac{Z_k}{b_n^{-1}X_{k-1}} > x\Big\} = \Big\{b_n\big(\hat{\phi}_n - \phi\big) > x\Big\}.$$

Applying Theorem 2.2.2 in the case of an AR(1) process so that $c_i = \phi^i$, $i \geq 0$, we have

$$\xi_n \xrightarrow{d} \xi = \sum_{k=1}^\infty \sum_{i=0}^\infty \epsilon_{(\phi^i j_k, Z_{i,k})}.$$

Note that as a subset of $E = (0, \infty] \times [\theta, \infty)$, the set Q_x is a bounded continuity set with respect to the limit point process ξ so that

$$\begin{aligned}\lim_{n\to\infty} P\Big(b_n\big(\hat{\phi}_n - \phi\big) > x\Big) &= P\Big(\xi\big(Q_x\big) = 0\Big) \\ &= P\Big(\bigwedge_{k=1}^\infty \bigwedge_{i=0}^\infty \frac{Z_{i,k}}{\phi^i j_k} > x\Big) = P\Big(\bigwedge_{k=1}^\infty \frac{W_k}{j_k} > x\Big)\end{aligned}$$

where $(W_k = \bigwedge_{i=0}^{\infty} \phi^{-i} Z_{i,k},\ k \geq 1)$ is an i.i.d. sequence independent of $(j_k,\ k \geq 1)$. Let

$$\tilde{\xi} = \sum_{k \geq 1} \epsilon_{(j_k, W_k)}. \tag{2.2.23}$$

Then by Proposition 5.6 in Resnick (2007), we have that if G denotes the distribution of W_1, then $\tilde{\xi}$ is a Poisson random measure on E with mean measure $\mu = \nu \times G$, where $\nu(dx) = \beta x^{-\beta-1} \mathbb{1}_{[0,\infty)}(x)dx$. Using (2.2.23) we can write

$$P\Big(\bigwedge_{k=1}^{\infty} \frac{W_k}{j_k} > x\Big) = P\Big(\tilde{\xi}(Q_x) = 0\Big) = \exp\left(-\mu(Q_x)\right).$$

Since $\mu(Q_x) = x^{\beta} E W^{-\beta}$, the result follows. □

Corollary 2.2.1. *Under conditions given in Theorem 2.2.3,*

$$\hat{\phi}_n \xrightarrow{a.s} \phi.$$

Proof. Since $b_n = F^{\leftarrow}(1 - 1/n) \to \infty$ we have $\hat{\phi}_n \xrightarrow{p} \phi$. But this implies $\hat{\phi}_n \xrightarrow{a.s.} \phi$ since $\hat{\phi}_n \geq \phi$ and is non-increasing. □

Let us now define our estimate of θ:

$$\hat{\theta}_n = \bigwedge_{t \in I_n} (X_t - \hat{\phi}_n X_{t-1}),$$

where we define the index set $I_n = \{t : 1 \leq t \leq n \text{ and } X_{t-1} \leq (a_n b_n)^{\rho}\}$ where $0 < \rho < 1$ is a fixed value.

Lemma 2.2.6. *Under the assumptions that F is regularly varying with index α at its positive left endpoint θ and $\overline{F}$ is regularly varying with index $-\beta$ at infinity, its right endpoint, and $\alpha > \beta$, then*

$$a_n^{-1}\Big(\hat{\theta}_n - \bigwedge_{t \in I_n} Z_t\Big) \xrightarrow{p} 0, \text{ as } n \to \infty,$$

where $a_n = F^{\leftarrow}(1/n) - \theta$. *Furthermore, for any* $y \geq 0$

$$\lim_{n\to\infty} P\Big(a_n^{-1}(\hat{\theta}_n - \theta) > y\Big) = \lim_{n\to\infty} P\Big(a_n^{-1}(\bigwedge_{t=1}^{n} Z_t - \theta) > y\Big) = e^{-y^{\alpha}}.$$

Proof. Since $\alpha > \beta$, we have $\lim_{n\to\infty} a_n b_n = \infty$. Therefore since $\big(b_n(\hat{\phi}_n - \phi), n \geq 1\big)$ is a tight sequence by Theorem 2.2.3 and since $\max_{t\in I_n} X_{t-1}/(a_n b_n)^{\rho} \leq 1$ with $0 < \rho < 1$, we have

$$a_n^{-1}(\hat{\phi}_n - \phi) \bigvee_{t\in I_n} X_{t-1} \xrightarrow{p} 0.$$

The first statement now follows since

$$a_n^{-1}\Big|\hat{\theta}_n - \bigwedge_{t\in I_n} Z_t\Big| \leq a_n^{-1}|\hat{\phi}_n - \phi| \bigvee_{t\in I_n} X_{t-1}.$$

For the second statement observe

$$\begin{aligned}
0 &\leq P\Big(a_n^{-1}(\bigwedge_{t\in I_n} Z_t - \theta) > y\Big) - P\Big(a_n^{-1}(\bigwedge_{t=1}^{n} Z_t - \theta) > y\Big) \\
&= P\Big(a_n^{-1}(\bigwedge_{t=1}^{n} Z_t - \theta) \leq y < a_n^{-1}(\bigwedge_{t\in I_n} Z_t - \theta)\Big) \\
&\leq P\Big(\bigcup_{1\leq t\leq n} \big(X_{t-1} > (a_n b_n)^{\rho} \text{ and } a_n^{-1}(Z_t - \theta) \leq y\big)\Big) \\
&\leq nP(X_0 > (a_n b_n)^{\rho})P(Z_1 \leq \ a_n y + \theta) = o(1). \qquad (2.2.24)
\end{aligned}$$

The result for the second statement now follows from (2.2.24) and the first part of the lemma. Finally, the identification of the limit distribution is well known. ☐

A useful observation follows from this lemma which we state as a corollary.

Corollary 2.2.2. *Under the assumptions of Lemma 2.2.6 for any* $x, y > 0$

$$\lim_{n\to\infty} \Big\{P\Big(b_n(\hat{\phi}_n - \phi) > x, a_n^{-1}(\hat{\theta}_n - \theta) > y\Big) - P\Big(b_n(\hat{\phi}_n - \phi) > x, a_n^{-1}(\bigwedge_{t=1}^{n}(Z_t - \theta) > y\Big)\Big\} = 0.$$

Proof. By Lemma 2.2.6 we have

$$\lim_{n\to\infty}\left\{P\Big(b_n(\hat{\phi}_n-\phi)>x, a_n^{-1}(\hat{\theta}_n-\theta)>y\Big)-P\Big(b_n(\hat{\phi}_n-\phi)>x, a_n^{-1}(\bigwedge_{t\in I_n} Z_t-\theta)>y\Big)\right\}=0.$$

The result then follows from

$$\begin{aligned}0 &\le E\mathbb{1}\{b_n(\hat{\phi}_n-\phi)>x\}\Big(\mathbb{1}\{a_n^{-1}(\bigwedge_{t\in I_n} Z_t-\theta)>y\}-\mathbb{1}\{a_n^{-1}(\bigwedge_{t=1}^{n} Z_t-\theta)>y\}\Big)\\ &\le P\Big(a_n^{-1}(\bigwedge_{t\in I_n} Z_t-\theta)>y\Big)-P\Big(a_n^{-1}(\bigwedge_{t=1}^{n} Z_t-\theta)>y\Big).\end{aligned}$$

□

Corollary 2.2.2 allows a simplification in determining the joint asymptotic behavior of $(\hat{\phi}_n, \hat{\theta}_n)$ by allowing us to replace $\hat{\theta}_n$ with $\min_{1\le t\le n} Z_t$. The next lemma will provide another useful simplification - this time on $\hat{\phi}_n$.

For a positive integer m define

$$X_t^{(m)}=\sum_{i=0}^{m-1}\phi^i Z_{t-i}.$$

Lemma 2.2.7. *Let $U_n^{(m)}$ and U_n be defined as*

$$U_n^{(m)}=\bigwedge_{t=1}^{n}\frac{Z_t}{b_n^{-1}X_{t-1}^{(m)}} \qquad \textit{and} \qquad U_n=\bigwedge_{t=1}^{n}\frac{Z_t}{b_n^{-1}X_{t-1}}.$$

Then for any $\epsilon>0$

$$\lim_{m\to\infty}\lim_{n\to\infty}P(|U_n^{(m)}-U_n|>\epsilon)=0.$$

Proof. We first note for any positive M that

$$\begin{aligned}P(|U_n^{(m)}-U_n|>\epsilon) &= P\left(U_n^{(m)}U_n\left|\frac{1}{U_n^{(m)}}-\frac{1}{U_n}\right|>\epsilon\right)\\ &\le P\left(\left|\frac{1}{U_n^{(m)}}-\frac{1}{U_n}\right|>\epsilon/M^2\right)+P(U_n^{(m)}>M).\end{aligned}$$

In order to calculate $P(|\frac{1}{U_n^{(m)}} - \frac{1}{U_n}| > \epsilon)$ we partition X_t. That is, we write

$$X_t = X_t^{(m)} + X_t^{\prime(m)}$$

where $X_t^{\prime(m)} = \sum_{j=m}^{\infty} \phi^j Z_{t-j}$, so that

$$0 \leq \bigvee_{t=1}^{n} \frac{X_{t-1}}{Z_t} - \bigvee_{t=1}^{n} \frac{X_{t-1}^{(m)}}{Z_t} \leq \bigvee_{t=1}^{n} \frac{X_{t-1}^{\prime(m)}}{Z_t}.$$

Define point processes

$$\xi_n^{(m)} = \sum_{t=1}^{n} \epsilon_{(b_n^{-1} X_{t-1}^{\prime(m)}, Z_t)} \quad \text{and} \quad \xi^{(m)} = \sum_{k=1}^{\infty} \sum_{i=m}^{\infty} \epsilon_{(\phi^i j_k, Z_{i,k})}$$

where $\{j_k,\ k \geq 1\}$ and $\{Z_{i,k}, k \geq 1, i \geq 0\}$ have the distributional properties given in Theorem 2.2.3. Applying Theorem 2.2.3 with $c_i = \phi^i$ for $i \geq m$ and c_i equal to 0 otherwise, we obtain

$$\xi_n^{(m)} \xrightarrow{d} \xi^{(m)}.$$

Then letting for $x > 0$

$$\mathcal{R}_x = \{(u, v) : u > 0, v > \theta, \text{ and } \frac{u}{v} > x\},$$

we have

$$\lim_{n \to \infty} P\Big(\bigvee_{t=1}^{n} b_n^{-1} X_{t-1}^{\prime(m)} / Z_t \leq x\Big) = \lim_{n \to \infty} P\Big(\xi_n^{(m)}(\mathcal{R}_x) = 0\Big) = P\Big(\xi^{(m)}(\mathcal{R}_x) = 0\Big).$$

Setting

$$V_k^{(m)} = \bigvee_{i=m}^{\infty} \frac{\phi^i}{Z_{i,k}} \quad \text{and} \quad \zeta^{(m)} = \sum_{k=1}^{\infty} \epsilon_{(j_k, V_k^{(m)})}$$

we have

$$P\Big(\xi^{(m)}(\mathcal{R}_x)=0\Big)=P\Big(\bigvee_{k=1}^{\infty} j_k V_k^{(m)} \le x\Big)=P\Big(\zeta^{(m)}(\mathcal{S}_x)=0\Big),$$

where

$$\mathcal{S}_x=\{(u,v): u>0, v>\theta, \text{ and } uv>x\}.$$

Since $\zeta^{(m)}$ is Poisson random measure with mean measure $\mu_m=\nu\times H_m$ where $V_1^{(m)}\sim H_m$ and

$$\mu_m(\mathcal{S}_x)=x^{-\beta}E\Big(V_1^{(m)}\Big)^{\beta},$$

we obtain

$$\lim_{n\to\infty} P\Big(\bigvee_{t=1}^{n} b_n^{-1}X_{t-1}^{\prime(m)}/Z_t\le x\Big)=\exp\Big(-x^{-\beta}E\big(V_1^{(m)}\big)^{\beta}\Big).$$

Next since

$$\left|\frac{1}{U_n^{(m)}}-\frac{1}{U_n}\right|\le\bigvee_{t=1}^{n}\frac{b_n^{-1}X_{t-1}^{'(m)}}{Z_t},$$

we have for large n that

$$\begin{aligned}
P\left(\left|\frac{1}{U_n^{(m)}}-\frac{1}{U_n}\right|>\epsilon/M^2\right) &\le P\left(\bigvee_{t=1}^{n}\frac{b_n^{-1}X_{t-1}^{'(m)}}{Z_t}>\epsilon/M^2\right)\\
&\le 2\Big(1-\exp\big(-\epsilon^{-\beta}M^{2\beta}E\big(V_1^{(m)}\big)^{\beta}\big)\Big)\\
&\le 2\epsilon^{-\beta}M^{2\beta}E\big(V_1^{(m)}\big)^{\beta}.
\end{aligned}$$

Therefore, since $V_1^{(m)}\le\phi^m/\theta$,

$$\lim_{m\to\infty}\lim_{n\to\infty}P\left(\left|\frac{1}{U_n^{(m)}}-\frac{1}{U_n}\right|>\epsilon/M^2\right)=0.$$

Next, note that from the limit law for the maximum obtained above, by replacing $X_{t-1}^{\prime(m)}$ with $X_{t-1}^{(m)}$ and by taking reciprocals, we derive the limit law for minimum,

$$\lim_{n\to\infty}P\Big(U_n^{(m)}=\bigwedge_{t=1}^{n}\frac{Z_t}{b_n^{-1}X_{t-1}^{(m)}}\le x\Big)=1-\exp\big(-x^{\beta}E\tilde{W}_m^{-\beta}\big) \tag{2.2.25}$$

where $\tilde{W}_m$ has the distribution of $\wedge_{i=0}^{m-1} Z_{i,k}/\phi^i$. Thus, for any integer $m \geq 1$,

$$\lim_{n\to\infty} P(U_n^{(m)} \leq x) \geq 1 - \exp\big(-(x/\theta)^\beta\big).$$

Thus for any $\epsilon > 0$, we have for M large enough that

$$\limsup_{n\to\infty} \sup_{m\geq 1} P(U_n^{(m)} > M) < \epsilon$$

completing the proof. □

With Lemma 2.2.7 in hand we can now focus our attention to the limiting joint distribution of $(U_n^{(m)}, \bigwedge_{t=1}^n Z_t)$. This will be accomplished by a blocking argument. To that end for a fixed positive integer k, let $r_n = \lfloor n/k \rfloor$ and define blocks for $i = 1, \ldots, \lfloor n/r_n \rfloor$ by

$$J_i = [(i-1)r_n + 1, \ldots, ir_n - q], \quad \text{and} \quad J_i^{'} = [ir_n - q + 1, \ldots, ir_n]$$

where q is a positive integer greater than m. Furthermore, let

$$J_0^{'} = [r_n \lfloor n/r_n \rfloor + 1, \ldots, n].$$

Now we define the events

$$\chi_i = \left\{ \exists\ l \in J_i : \left(\frac{Z_l}{b_n^{-1} X_{l-1}^{(m)}} \right) \leq x \text{ or } a_n^{-1}(Z_l - \theta) \leq y \right\}, \quad i = 1, \ldots, l_n$$

and

$$\chi_i^{'} = \left\{ \exists\ l \in J_i^{'} : \left(\frac{Z_l}{b_n^{-1} X_{l-1}^{(m)}} \right) \leq x \text{ or } a_n^{-1}(Z_l - \theta) \leq y \right\}, \quad i = 0, \ldots, l_n$$

where $l_n = \lfloor n/r_n \rfloor$. We begin by showing that the events $\chi_i^{'}$ are negligible.

Lemma 2.2.8. *For any $x, y > 0$*

$$\lim_{n\to\infty} P\big(\bigcup_{i=0}^{l_n} \chi_i^{'}\big) = 0.$$

Proof. Observe that

$$nP\Big(\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}} \leq x\Big) \leq nP\Big(\frac{\theta}{b_n^{-1}X_{l-1}^{(m)}} \leq x\Big) \sim x^{\beta}\theta^{-\beta}\sum_{i=0}^{m-1}\phi^{i\beta} \leq \frac{x^{\beta}}{\theta^{\beta}(1-\phi^{\beta})} \qquad (2.2.26)$$

and

$$nP\Big(a_n^{-1}(Z_l - \theta) \leq y\Big) \sim y^{\alpha}. \qquad (2.2.27)$$

Thus for some constant c and any $n \geq 1$,

$$P\big(\bigcup_{i=0}^{l_n} \chi_i^{'}\big) \leq ck/n$$

establishing the lemma. □

Define events A_i and B_i by

$$A_i = \left\{\exists\ l \in J_i : \left(\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}}\right) \leq x\right\} \text{ and } B_i = \{\exists\ l \in J_i : a_n^{-1}(Z_l - \theta) \leq y\}.$$

The following result provides the asymptotic behavior of the probability of these events.

Lemma 2.2.9. *For any $x > 0$, $y > 0$, we have as $k \to \infty$*

$$\lim_{n\to\infty} P(A_i) \sim \frac{1}{k}x^{\beta}E\tilde{W}_m^{-\beta} \quad \text{and} \quad \lim_{n\to\infty} P(B_i) \sim \frac{y^{\alpha}}{k}.$$

Proof. Since the events A_i are independent, we have

$$P\left(\bigcap_{i=1}^{k}\left(\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}} > x, \forall\, l \in J_i\right)\right) = \left(P\left(\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}} > x, \forall\, l \in J_1\right)\right)^k.$$

Using Lemma 2.2.8 we have that

$$P\left(\bigwedge_{l=1}^{n}\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}}>x\right)=P\left(\bigcap_{i=1}^{k}\left(\left(\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}}\right)>x,\forall\, l\in J_i\right)\right)+O\left(\frac{1}{n}\right)$$
$$=\left(P\left(\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}}>x,\forall\, l\in J_1\right)\right)^k+O\left(\frac{1}{n}\right).$$

From (2.2.25) we showed that

$$\lim_{n\to\infty}P\Big(U_n^{(m)}=\bigwedge_{t=1}^{n}\frac{Z_t}{b_n^{-1}X_{t-1}^{(m)}}\leq x\Big)=1-\exp\big(-x^{\beta}E\tilde{W}_m^{-\beta}\big).$$

Hence using this limit law on $\wedge_{t=1}^{n}(Z_t/b_n^{-1}X_{t-1}^{(m)})$, we obtain

$$\lim_{k\to\infty}\lim_{n\to\infty}\left(P\left[\frac{Z_l}{b_n^{-1}X_{l-1}^{(m)}}>x,\forall\ l\in J_1\right]\right)^k=\exp(-x^{\beta}E\tilde{W}_m^{-\beta}).$$

Thus,

$$\lim_{n\to\infty}P[A_i]\sim\frac{x^{\beta}}{k}E\tilde{W}_m^{-\beta},\qquad\text{as } k\to\infty. \tag{2.2.28}$$

Similarly using the result of Lemma 2.2.6, we obtain

$$\lim_{n\to\infty}P[B_i]\sim\frac{y^{\alpha}}{k},\qquad\text{as } k\to\infty. \tag{2.2.29}$$

Hence the lemma holds. □

Lemma 2.2.10. *For some constant c*

$$P(A_i\cap B_i)\leq c\,P(A_i)P(B_i).$$

Proof. **Remark.** Since the cardinality of J_i depends on r_n which depends on k and the events A_i and B_i depend on n, $P(A_i)$ and $P(B_i)$ depend on k and n. The conclusion of this lemma provides that for all k and n, there is a constant dependent on no parameters for which the inequality stated there holds.

To calculate the intersection we define the following sets

$$K_1 = K_{1,i} = \{(l_1, l_2) : l_1, l_2 \in J_i \quad \text{and} \quad l_2 \notin [l_1 - 1 - m, l_1]\}$$

and

$$K_2 = K_{2,i} = \{(l_1, l_2) : l_1, l_2 \in J_i \quad \text{and} \quad l_2 \in [l_1 - 1 - m, l_1]\}.$$

Now for $l_1 - 1 - m \leq l_2 \leq l_1 - 1$ and n sufficiently large,

$$\left\{ \left(\frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \right) \leq x \right\} \cap \{a_n^{-1}(Z_{l_2} - \theta) \leq y\} \subset \{ \sum_{\substack{i \neq l_1 - l_2 - 1 \\ 0 \leq i \leq m-1}} \phi^i Z_{l_1-1-i} \geq b_n \frac{Z_{l_1}}{x} - \theta - 1\} \cap \{Z_{l_2} \leq \theta + a_n y\}.$$

It then follows from (2.2.26), (2.2.27), and independence that

$$P\Big(\{ \sum_{\substack{i \neq l_1 - l_2 - 1 \\ 0 \leq i \leq m-1}} \phi^i Z_{l_1-1-i} \geq b_n \frac{Z_{l_1}}{x} - \theta - 1\} \cap \{Z_{l_2} \leq \theta + a_n y\}\Big) = O\left(\frac{1}{n^2}\right).$$

If $l_1 = l_2 = l$, we have

$$P\left(\left\{ \frac{Z_l}{b_n^{-1} X_{l-1}^{(m)}} \leq x \right\} \cap \{a_n^{-1}(Z_l - \theta) \leq y\}\right) = \int_\theta^{\theta + a_n y} P\left(X_{l-1}^{(m)} \geq \frac{z}{b_n^{-1} x}\right) dF(z) = O\left(\frac{1}{n^2}\right).$$

Therefore, for some constant c

$$P\left(\bigcup_{(l_1,l_2) \in K_2} \left\{ \frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \leq x \right\} \cap \{a_n^{-1}(Z_{l_2} - \theta) \leq y\} \right) = c/nk.$$

In order to handle set K_1, observe from construction of the blocks J_i and set K_1 that if $(l_1, l_2) \in K_1$ then the events

$$\left\{ \frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \leq x \right\} \quad \text{and} \quad \{a_n^{-1}(Z_{l_2} - \theta) \leq y\}$$

are independent. Thus, if we define $\{Z_i^{'}, i \in \mathbb{Z}\}$ as an independent copy of $\{Z_i, i \in \mathbb{Z}\}$, then

$$\begin{aligned}
&P\left(\bigcup_{(l_1,l_2)\in K_1} \left\{ \frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \leq x \right\} \cap \{a_n^{-1}(Z_{l_2} - \theta) \leq y\} \right) \\
&= P\left(\bigcup_{(l_1,l_2)\in K_1} \left\{ \frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \leq x \right\} \cap \{a_n^{-1}(Z_{l_2}^{'} - \theta) \leq y\} \right) \\
&\leq P(A_i \cap B_i^{'}) = P(A_i)P(B_i) \leq c/k^2
\end{aligned}$$

where $B_i^{'} = \{\exists\ l \in J_i : a_n^{-1}(Z_l^{'} - \theta) \leq y\}$ and where we used Lemma 2.2.9 in the last step. Thus, we have that for some constant c

$$\begin{aligned}
P\left(A_i \cap B_i\right) = P(&\bigcup_{(l_1,l_2)\in K_1} \left\{ \frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \leq x \right\} \cap \{a_n^{-1}(Z_{l_2} - \theta) \leq y\} \\
&\cup \bigcup_{(l_1,l_2)\in K_2} \left\{ \frac{Z_{l_1}}{b_n^{-1} X_{l_1-1}^{(m)}} \leq x \right\} \cap \{a_n^{-1}(Z_{l_2} - \theta) \leq y\}) \\
&\leq c/k^2 = O((P(A_i)P(B_i))\,,
\end{aligned}$$

which completes the proof in view of Lemma 2.2.9. □

Lemma 2.2.11. *For any $x > 0,\ y > 0$,*

$$\lim_{n\to\infty} P\left(\bigwedge_{t=1}^{n} \left(\frac{Z_t}{b_n^{-1} X_{t-1}^{(m)}} \right) > x, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y \right) = \exp\left(-x^{\beta} E\tilde{W}_m^{-\beta} - y^{\alpha}\right).$$

Proof. First from Lemma 2.2.8, we have as $n \to \infty$ that

$$P\left(\bigwedge_{t=1}^{n}\left(\frac{Z_t}{b_n^{-1}X_{t-1}^{(m)}}\right) > x, a_n^{-1}\bigwedge_{t=1}^{n}(Z_t-\theta) > y\right) = P\left(\bigcap_{i=1}^{k}\chi_i^c\right) + o(1).$$

Next by (2.2.28), (2.2.29), and Lemma 2.2.10 we obtain that as k tends to infinity

$$\lim_{n\to\infty} P(\chi_i) \sim \frac{1}{k}\left(x^{\beta}E\tilde{W}_m^{-\beta} + y^{\alpha}\right).$$

Therefore, we obtain

$$\lim_{k\to\infty}\lim_{n\to\infty} P\left(\bigcap_{i=1}^{k}\chi_i^c\right) = \lim_{k\to\infty}\left(1-\frac{1}{k}(x^{\beta}E\tilde{W}_m^{-\beta} + y^{\alpha})\right)^k = \exp(-(x^{\beta}E\tilde{W}_m^{-\beta} + y^{\alpha})).$$

Hence

$$\lim_{n\to\infty} P\left(\bigwedge_{t=1}^{n}\left(\frac{Z_t}{b_n^{-1}X_{t-1}^{(m)}}\right) > x, a_n^{-1}\bigwedge_{t=1}^{n}(Z_t-\theta) > y\right) = \exp(-x^{\beta}E\tilde{W}_m^{-\beta} - y^{\alpha}).$$

□

Theorem 2.2.4. *Let $\{X_t, t \geq 1\}$ denote the stationary AR(1) process such that the innovation distribution F satisfies*

$$\bar{F} \text{ is } RV_{-\beta} \text{ at infinity and } F \text{ is } RV_{\alpha} \text{ at } \theta.$$

If $\alpha > \beta$ then for any $x > 0, \quad y > 0$ we have

$$\lim_{n\to\infty} P(b_n(\hat{\phi}_n - \phi) > x, a_n^{-1}(\hat{\theta}_n - \theta) > y) = e^{-x^{\beta}EW^{-\beta} - y^{\alpha}},$$

where $W = \bigwedge_{j=0}^{\infty} Z_j/\phi^j$.

Proof. Let us first observe that for $\epsilon > 0$

$$P\Big(U_n^{(m)} > x + \epsilon, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y\Big) - P\Big(|U_n^{(m)} - U_n| > \epsilon\Big)$$
$$\leq P\Big(U_n > x, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y\Big) \leq P\Big(U_n^{(m)} > x, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y\Big).$$

Thus by Lemma 2.2.11 we obtain

$$\exp\{-(x+\epsilon)^\beta E\tilde{W}_m^{-\beta} - y^\alpha\} - \limsup_{n\to\infty} P\Big(|U_n - U_n^{(m)}| > \epsilon\Big)$$
$$\leq \liminf_{n\to\infty} P\Big(U_n > x, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y\Big)$$
$$\leq \limsup_{n\to\infty} P\Big(U_n > x, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y\Big)$$
$$\leq \exp(-x^\beta E\tilde{W}_m^{-\beta} - y^\alpha).$$

Letting m tend to infinity in the above and then ϵ tend to 0, we obtain from Lemma 2.2.7 and $\lim_{m\to\infty} E\tilde{W}_m^{-\beta} = EW^{-\beta}$ that

$$\lim_{n\to\infty} P\Big(U_n > x, a_n^{-1} \bigwedge_{t=1}^{n} (Z_t - \theta) > y\Big) = \exp(-x^\beta EW^{-\beta} - y^\alpha).$$

The theorem now follows from this and Corollary 2.2.2. □

2.3 Simulation Study

In this section we assess the reliability of our extreme value estimation method through a simulation study. This included a comparison between our estimation procedure and that of three alternative estimation procedures for both the autocorrelation coefficient ϕ and the unknown location parameter θ under two different innovation distributions. Additionally, the degree of approximation for the empirical probabilities of $\hat{\phi}_{min}$ and $\hat{\theta}_{min}$ to its respective limiting distribution was reported.

To study the performance of the estimators $\hat{\phi}_{min} = \bigwedge_{t=1}^{n} \frac{X_t}{X_{t-1}}$ and $\hat{\theta}_{min} = \bigwedge_{t \in I_n}(X_t - \hat{\phi}_{min} X_{t-1})$ respectively, we generated 5,000 replications for the nonnegative time series $(X_0, X_1, \ldots, X_n)$ for two different sample sizes $(500, 1000)$, where $\{X_t\}$ is an AR(1) process satisfying the difference equation

$$X_t = \phi X_{t-1} + Z_t, \text{ for } (1 \leq t \leq n) \text{ and } Z_t \geq \theta.$$

The autoregressive parameter ϕ is taken to be in the range from 0 to 1 guaranteeing a nonnegative time series and the unknown location parameter θ is positive when the innovations Z_t are taken to be

$$F(z) = \begin{cases} c(z-\theta)^{\alpha} & \text{if } \theta < z < \theta + 1, \\ 1 - d(z-\theta)^{-\beta} & \text{if } \theta + 1 < z < \infty. \end{cases}$$

For this innovation distribution let c and d be nonnegative constants such that $c+d = 1$, then this distribution is regularly varying at both endpoints with index of regular variation $-\beta$ at infinity and index of regular variation α at θ. For this simulation study two distributions were considered: (i) $F_1, c = 0, d = 1,$ (ii) $F_2, c = .5, d = .5.$

Now observe in case (i) the innovation distribution F_1 is a Pareto distribution with a regular varying tail distribution at ∞ with index of regular variation $-\beta$ and regular varying at $\theta + 1$ with a fixed index $\alpha = 1$, whereas in case (ii) the innovation distribution F_2 is regular varying at ∞ and θ with no restriction on α or β.

First we examine the simulation results for $\phi = .9$ under F_1 for each of the six different β values considered by computing $5,000$ estimates using $\hat{\phi}_{min} = \bigwedge_{t=1}^{n} \frac{X_t}{X_{t-1}}$, $\hat{\phi}_{max} = \frac{X_{t^*+1}}{X_{t^*}}$, $\hat{\phi}_{range} = \frac{X_{t^*+1} - X_{j^*}}{X_{t^*} - X_{j^*-1}}$, where t^* and j^* provides the index of the maximal and minimal X_i respectively for $1 \leq i \leq n$, and

$$\hat{\phi}_{LS} = \begin{cases} \sum_{t=1}^{n-1} X_t X_{t+1} / \sum_{t=1}^{n} X_t^2, & \text{if } 0 < \beta < 1 \\ \sum_{t=1}^{n} (X_t - \bar{X})(X_{t+1} - \bar{X}) / \sum_{t=1}^{n} (X_t - \bar{X})^2, & \text{if } 1 < \beta < 3 \end{cases}$$

where $\bar{X} = \sum_{t=1}^{n} X_t/n$. The means and standard deviations (written below in parentheses), of these estimates are reported in Table 2.1 along with the average length for a 95 percent empirical confidence intervals with exact coverage. Since the main purpose of this section is to compare our estimator $\hat{\phi}_{min}$ to Bartlett and McCormick (2012) estimator $\hat{\phi}_{max}$, McCormick and Mathew (1993) estimator $\hat{\phi}_{range}$, and Davis and Resnick's (1986) estimator $\hat{\phi}_{LS}$, the confidence intervals were directly constructed from the empirical distributions of $n^{1/\beta}(\hat{\phi}_{min} - \phi)$, $n^{1/\beta}(\hat{\phi}_{max} - \phi)$, $n^{1/\beta}(\hat{\phi}_{range} - \phi)$, and $(n/\log n)^{1/\beta}(\hat{\phi}_{LS} - \phi)$, respectively.

Table 2.1: Comparison of Estimators for $\phi = .9$ under F_1

						95% C.I. Avg. Length			
β	n	$\hat{\phi}_{min}$	$\hat{\phi}_{max}$	$\hat{\phi}_{range}$	$\hat{\phi}_{LS}$	Min est.	Max est.	Range est.	LS est.
.2	500	.9000	.9002	.9002	.8988	$<$ .0001	$<$ .0001	$<$ .0001	.0288
		($<$ .0001)	(.0015)	(.0016)	(.0219)	-	-	-	-
	1000	.9000	.9001	.9002	.8997	$<$ .0001	$<$ .0001	$<$ .0001	.0091
		($<$ .0001)	(.0009)	(.0009)	(.0112)	-	-	-	-
.8	500	.9004	.9026	.9032	.9231	.0046	.0139	.0228	.0657
		(.0016)	(.0083)	(.0083)	(.0251)	-	-	-	-
	1000	.9002	.9020	.9014	.9158	.0023	.0072	.0133	.0529
		(.0008)	(.0064)	(.0064)	(.0186)	-	-	-	-
1.2	500	.9049	.9072	.9067	.8923	.0271	.0449	.0548	.0623
		(.0078)	(.0146)	(.0139)	(.0176)	-	-	-	-
	1000	.9029	.9117	.9047	.8964	.0187	.0325	.0389	.0439
		(.0056)	(.0117)	(.0109)	(.0117)	-	-	-	-
1.8	500	.9191	.9202	.9136	.8947	.0562	.0763	.0846	.0695
		(.0622)	(.0673)	(.0612)	(.0381)	-	-	-	-
	1000	.9113	.9208	.9116	.8989	.0501	.0690	.0743	.0485
		(.0313)	(.0367)	(.0384)	(.0223)	-	-	-	-
2.2	500	.9236	.9313	.9176	.8917	.0670	.0847	.0944	.0709
		(.0909)	(.1050)	(.0244)	(.0281)	-	-	-	-
	1000	.9207	.9293	.9151	.8961	.0623	.0821	.0874	.0494
		(.0412)	(.0456)	(.0217)	(.0228)	-	-	-	-
2.8	500	.9475	.9505	.9205	.8873	.0767	.0905	.1050	.0737
		(.1378)	(.1471)	(.0723)	(.0691)	-	-	-	-
	1000	.9437	.9484	.9187	.8906	.0735	.0885	.0956	.0516
		(.0834)	(.1163)	(.0317)	(.0331)	-	-	-	-

To evaluate and compare the performance of four location estimators, six different scenarios for α and β are presented in Table 2.2 under F_2. When $\theta = 2$, 5,000 estimates for each estimator; $\hat{\theta}_{min} = \bigwedge_{t \in I_n}(X_t - \hat{\phi}_{min}X_{t-1})$, $\hat{\theta}_{range} = X_{j^*} - \hat{\phi}_{range}X_{j^*-1}$, $\hat{\theta}_{e1} = (1 - \hat{\phi}_{range}) \bigwedge_{t=1}^{n} X_t$, and $\hat{\theta}_{e2} = (1 - \hat{\phi}_{LS}) \bigwedge_{t=1}^{n} X_t$ were obtained. The exponent ρ inside the index set $I_n = \{t : 1 \leq t \leq n \text{ and } X_{t-1} \leq (a_n b_n)^{\rho}\}$, was set to .9. The means and standard deviations (written below in parentheses), of these estimates are reported in Table 2.2 along with the average length for a 95 percent empirical confidence intervals. For convenience, the empirical distributions of $n^{-1/\alpha}(\hat{\theta}_{min} - \theta)$, $n^{-1/\alpha}(\hat{\theta}_{range} - \theta)$, $q_n(\hat{\theta}_{e1} - \theta - (1 - \hat{\phi}_{range})w_n)$, and $q_n(\hat{\theta}_{e2} - \theta - (1 - \hat{\phi}_{LS})w_n)$

were respectively used, where the normalizing constants q_n and w_n are obtained through equations (3.12 - 3.16) of McCormick and Mathew (1993).

Table 2.2: Comparison of Estimators for $\theta = 2$ under F_2

							95% C.I. Avg. Length			
β	α	n	$\hat{\theta}_{min}$	$\hat{\theta}_{range}$	$\hat{\theta}_{e1}$	$\hat{\theta}_{e2}$	Min est.	Range est.	E1(range) est.	E2(LS) est.
.5	.6	500	2.00	2.44	3.16	3.98	$<$ .0001	2.34	2.09	3.02
			(.0087)	(.4068)	(.6775)	(.7657)	-	-	-	-
		1000	2.00	2.39	3.77	3.67	$<$ .0001	1.89	1.69	2.60
			(.0002)	(.3605)	(.4381)	(.4960)	-	-	-	-
	1.6	500	1.94	2.68	3.34	4.10	.0273	2.37	3.18	3.66
			(.0879)	(.3566)	(.6656)	(.7503)	-	-	-	-
		1000	1.99	2.63	3.94	3.81	.0152	1.87	2.23	2.79
			(.0503)	(.3199)	(.4311)	(.4780)	-	-	-	-
	2.6	500	1.87	2.77	3.41	4.16	.1422	2.19	2.85	2.38
			(.4697)	(.7553)	(.6733)	(.7632)	-	-	-	-
		1000	1.89	2.73	3.99	3.87	.0988	1.74	2.21	2.03
			(.4045)	(.3143)	(.4276)	(.4822)	-	-	-	-
1.5	.6	500	NA	1.72	2.35	2.15	NA	1.57	1.28	2.21
			(-)	(.9410)	(.5327)	(.3706)	-	-	-	-
		1000	NA	1.80	2.19	2.65	NA	1.20	1.25	1.88
			(-)	(.9141)	(.4381)	(.2565)	-	-	-	-
	1.6	500	2.80	1.95	2.54	2.26	1.72	2.23	2.18	1.94
			(.3863)	(.9569)	(.5298)	(.3815)	-	-	-	-
		1000	2.38	2.03	2.37	2.84	1.28	2.10	1.78	1.31
			(.2263)	(.9308)	(.4461)	(.2701)	-	-	-	-
	2.6	500	2.99	2.16	2.79	2.35	1.83	2.47	2.21	2.09
			(.3604)	(.9443)	(.5271)	(.3982)	-	-	-	-
		1000	2.68	2.12	2.27	2.93	1.26	2.10	1.81	1.42
			(.2956)	(.9010)	(.4398)	(.2837)	-	-	-	-

Remark. In the case that $0 < \beta < 1$, $\hat{\theta}_{e1}$ converge at a faster rate than $\hat{\theta}_{e2}$ and in the case that $1 < \beta < 3$, $\hat{\theta}_{e2}$ converges at a faster rate than $\hat{\theta}_{e1}$. Lastly, since the McCormick and Mathew (1993) paper has the restriction that $Var(Z_1) < \infty$, only when $\alpha > 2$ can the estimators $\hat{\theta}_{e1}$ and $\hat{\theta}_{e2}$ be fairly compared, whereas only when $\alpha > \beta$ is our estimator applicable.

Now observe for the selected β values being considered, Table 2.1 shows that our estimator performs at least as well as the three other alternative estimators. This is particularly true under the heavier tail models, i.e. when $0 < \beta < 2$. In this regime our estimate shows little bias and the average lengths of the confidence intervals are smaller than the other three estimates, sometimes by a wide margin. In particular, when $\beta = .8$ and $n = 1000$ the 95% confidence interval average length for our method is $3.13, 5.78$, and 23 times smaller than the three alternative estimators respectively. This is in part due to the use of one-sided confidence intervals since $\hat{\phi}_{min} \geq \phi$, for all $t \geq 1$. Naturally, when $1 < \beta < 3$, Davis and Resnick least square estimator is more efficient than all three extreme value estimators. While our estimator $\hat{\phi}_{min}$ will always perform slightly better than the $\hat{\phi}_{max}$ estimator, Bartlett's and McCormick (2012) estimator $\hat{\phi}_{max}$ main advantage lies with its versatility to perform well for various nonnegative time series, including but not restricted to higher order autoregressive models, along with ARMA models.

Table 2.2 reveals that our estimator for θ generally performs better than the three alternative estimators for $\alpha = .6, 1.6, 2.6$ when $0 < \beta < 1$. This is particularly true when comparing average confidence interval lengths. Although all three estimators $\hat{\theta}_{min}$, $\hat{\theta}_{range}$, and $\hat{\theta}_{e1}$ converge to the true value of the parameter θ as n tends to infinity respectively, in this setting they may not compete asymptotically with, say, a conditional least square estimator $\hat{\theta}_{e2}$ when $\beta > 1$. Nonetheless for small sample sizes our simulation study favors $\hat{\theta}_{range}$ over the other three estimators. The difficulty for a least square estimate is that a small negative bias for the estimate of the autocorrelation parameter ϕ gives rise to a much larger positive bias in the estimate of $\hat{\theta}_{e2}$. While the affect is not as great, the positive bias found in our estimator $\hat{\phi}_{min}$ and the others for ϕ has a significant effect on the estimate for θ.

Figures 2.1,2.2, 2.3, and 2.4 below show a comparison between the probability that estimators $\hat{\phi}_{min}$, $\hat{\phi}_{max}$, $\hat{\phi}_{range}$, and $\hat{\phi}_{LS}$ are within .01 of the true autocorrelation parameter value, respectively. With a sample size of 500, these figures plotted the sample fraction of estimates which fell within a bound of $\epsilon = .01$ of the true value. Good performance with respect to this measure is reflected in curves near to 1.0 with diminishing good behavior as curves approach 0.0. When $0 < \beta < 1$, the figures seem to show that our estimator compared to the other three produced a higher fraction of precise estimates, especially compared to Davis and Resnick estimator. When the regular variation index value is closer to 2, we see a higher fraction of the Davis-Resnick estimates showing better accuracy by this measure. The figures also indicate that McCormick and Mathew's range estimator produced a consistent high fraction of precise estimates when $\beta > 2$.

Figure 2.1: $P[|\hat{\phi}_{Min} - \phi| < .01]$ for $RV_{-\beta}$

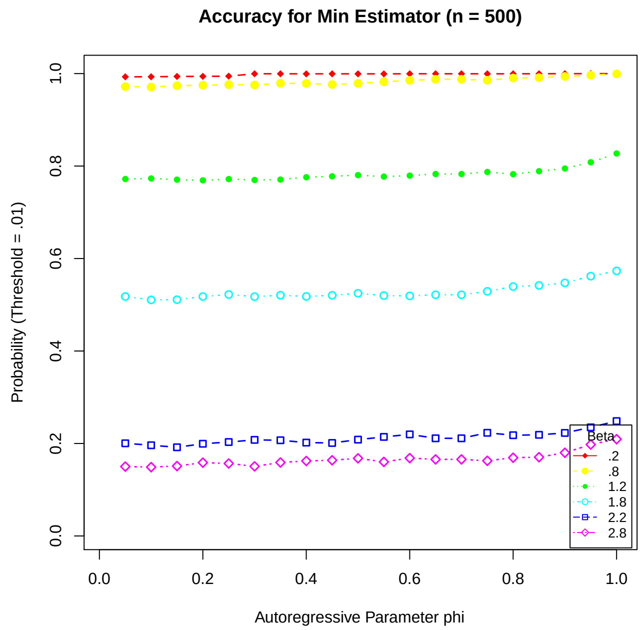


Figure 2.2: $P[|\hat{\phi}_{Max} - \phi| < .01]$ for $RV_{-\beta}$

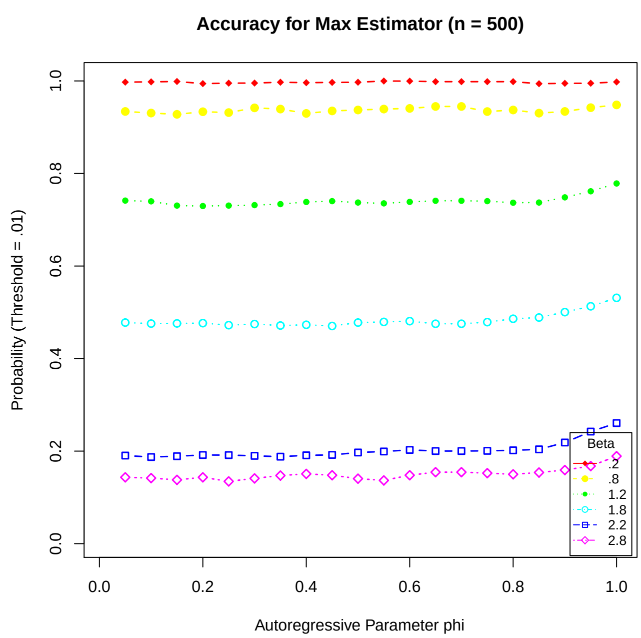


Figure 2.3: $P[|\hat{\phi}_{Range} - \phi| < .01]$ for $RV_{-\beta}$

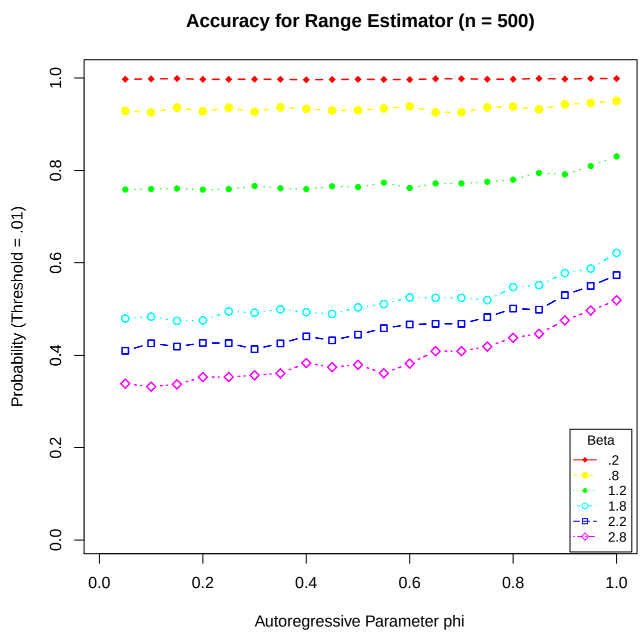
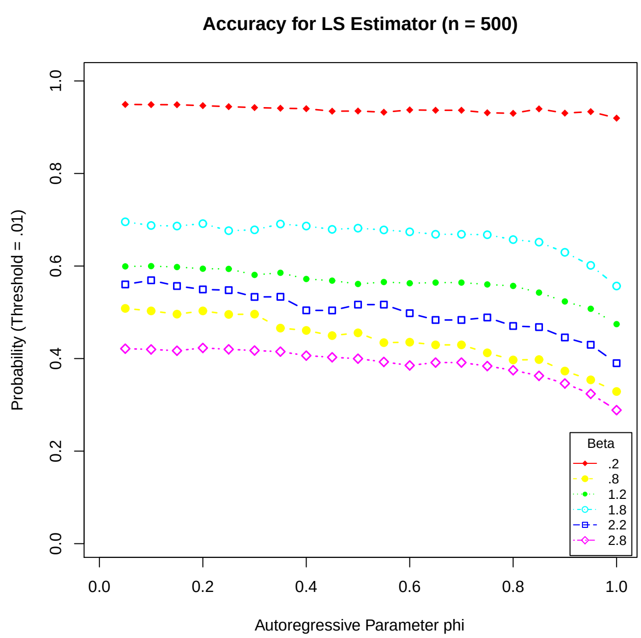
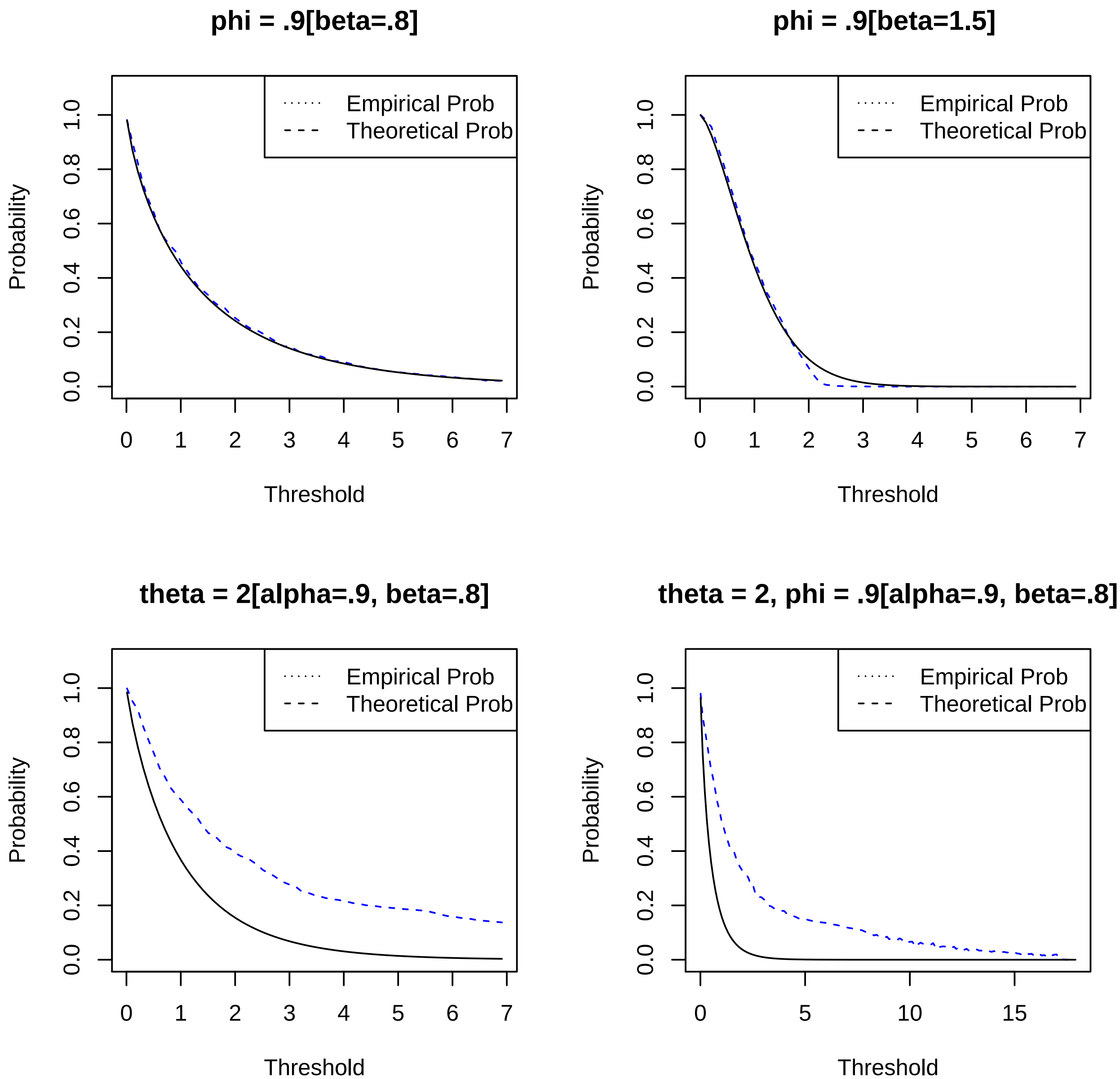


Figure 2.4: $P[|\hat{\phi}_{LS} - \phi| < .01]$ for $RV_{-\beta}$



Lastly, we performed a Monte Carlo simulation to study the degree of approximation for the empirical probability $P[b_n(\hat{\phi}_{min} - \phi) > x]$, $P[a_n^{-1}(\hat{\theta}_{min} - \theta) > y]$, and $P[b_n(\hat{\phi}_{min} - \phi) > x, a_n^{-1}(\hat{\theta}_{min} - \theta) > x]$ to its limiting values $e^{-x^\beta EW^{-\beta}}$, e^{-y^α}, and $e^{-x^\beta EW^{-\beta} - y^\alpha}$ respectively. The empirical distributions were calculated from 5,000 replications of the nonnegative time series $(X_0, X_1, \ldots, X_n)$ for a sample size of $5,000$, where $EW^{-\beta} = 1 - \sum_{i=0}^{M} \frac{\phi^{(\beta i(i+3)/2)}}{i+2}(1 - \phi^{\beta(i+2)})$, and M was set to 500. Additionally, we restricted $\alpha > \beta$. The top two plots in Figure 2.5 below shows the performance when $Z_t \sim F_1$ and the autocorrelation coefficient ϕ is .9 for $\alpha = 1$ and β equal to $.8, 1.5$ respectively. Observe for $0 < x < 7$ that the empirical tail probability $b_n(\hat{\phi}_{min} - \phi) > x$ mirrors the theoretical probability quite nicely. The lower left plot in Figure 2.5 displays the asymptotic performance when $Z_t \sim F_2$ and the location parameter θ is 2 for $\alpha = .9, \beta = .8$. Notice that the convergence rate of the empirical probability to the theoretical probability is extremely slow. This is not surprising since on average our estimate falls more than .1 from the true value when $\beta = .8$. The lower right plot in Figure 2.5 displays the asymptotic performance when $Z_t \sim F_1$ for the joint distribution of $(\hat{\phi}_{min}, \hat{\theta}_{min})$. Observe that this plot solidifies the asymptotic independence between $b_n(\hat{\phi}_{min} - \phi)$ and $a_n^{-1}(\hat{\theta}_{min} - \theta)$.

Figure 2.5: Empirical vs. Theoretical Probability



2.4 References

[1] Anděl, J. (1989). Non-negative autoregressive processes. *J. Time Ser. Anal.* **10**, 1-11.

[2] Anděl, J. (1991). Non-negative linear processes. *Appl. of Math.* **36**, 277-283.

[3] Bartlett, A., McCormick W.P. (2012). Estimation for Non-negative First order Autoregressive Process with an Unknown Location Paramter. *Applied Mathematics*, **3 No. 12A**, 2133 - 2147.

[4] Billingsley P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[5] Brockwell, P.J., Davis, R.A. (1987). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

[6] Datta, S., McCormick, W.P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *J. Amer. Statist. Assoc.* **90**, 1289-1300.

[7] Davis, R.A., McCormick, W.P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Process. Appl.* **31**, 237–250.

[8] Davis, R.A., Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann Probab.* **13**, 179-195.

[9] Davis, R.A., Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.

[10] Kallenberg, O. (1976). *Random Measures*, Akademie-Verlag, Berlin.

[11] McCormick, W.P., Mathew, G. (1989). Asymptotic results for an extreme value estimator of the autocorrelation coefficient for a first order autoregressive sequence. *In Lecture Notes in Statistics 51, Extreme Value Theory (eds Husler and Reiss)*, Springer-Verlag, Berlin, 166–180.

[12] McCormick W.P., Mathew, G. (1993). Estimation for Nonnegative autoregressive Processes with an Unknown Location Parameter. *J. Time Series Anal.* **14**, 71–92.

[13] Raftery, A.E., (1980). Estimation eficace pour un processus autoregressif exponentiel a densite discontinue. *Pub. Inst. Statist.* **25**, 65 – 91.

[14] Resnick, S.I., (1986). Point Processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66 – 138.

[15] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[16] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

Chapter 3

Estimation for Nonnegative Time Series with Heavy-tail Innovations [3]

[3]Bartlett A. and McCormick W.P. 2013. *J. Time Series Anal*, **34**, 96 - 115.
Reprinted here with permission of the publisher.

Abstract

For moving average processes $X_t = \sum_{i=0}^{\infty} c_i Z_{t-i}$ where the coefficients are nonnegative and the innovations are nonnegative random variables with a regularly varying tail at infinity, we provide estimates for the coefficients based on the ratio of two sample values chosen with respect to an extreme value criteria. We then apply this result to obtain estimates for the parameters of nonnegative ARMA models. Weak convergence results for the joint distribution of our estimates are established and a simulation study is provided to examine the small sample size behavior of these estimates.

AMS 1980 *subject classifications.* Primary 62M10; secondary 62E20, 60F05.

Key words: Nonnegative time series, ARMA processes, extreme value estimator, regular variation, point processes

3.1 Introduction

Linear time series models form a prominent class of models for dependent time series data. An excellent presentation of the classical theory concerning these models can be found, for example, in Brockwell and Davis (1987). More recent developments have focused on some specialized features of the model, e.g. heavy tail innovations or nonnegativity of the model. An elegant approach to studying heavy tail linear models is to examine the behavior of traditional estimates under conditions leading to non-Gaussian limits. For example, for AR(p) models, the standard approach to parameter estimation is through the Yule-Walker estimates, i.e. using the Yule-Walker equations relating covariances to the model parameters. In Davis and Resnick (1986), the authors establish the weak limit behavior for the sample autocorrelation function under an assumption that the innovations have a regularly varying tail with index α and are tailed balanced where α ranges from greater than 0 to at most 4. When $0 < \alpha < 2$ so that variances do not exist, a natural analogue to the correlation function in the linear model setting, which reduces to the correlation function when a second moment exists, appears in the limit results. These results may then be applied in the usual way to obtain estimates of the AR(p) parameters and their limiting behavior. On a similar note, Anderson et al. (2008) take an approach analogous to the Davis and Resnick (1986) work but utilize the innovations algorithm applied to periodically stationary time series and in particular to PARMA models, i.e. periodic ARMA models. In Davis (1996), the author analyzes parameter estimation for heavy tailed ARMA models through application of Gauss-Newton type estimators and M-estimators. Least absolute deviation estimation and other related methods are carried out in Calder and Davis (1998) while a weighted least squares method to estimate the parameters of a heavy tailed ARMA model is presented in Markov (2009).

While a number of established statistical estimation procedures were enumerated in the list of papers cited above, one notable exception was that of maximum likelihood. This is not to be unexpected since in the time series setting, the likelihood function is generally particularly intractable and intricate. However, there is an exception to this general rule which has particular bearing on estimation of positive heavy tailed time series and that is the case of exponential innovations in an AR(p) process. In this case the likelihood function is easily obtained and finding the MLE amounts to solving a constrained maximization problem with linear constraints, in other words, a linear programming problem. With these considerations in mind, Feigin and Resnick (1994) develops linear programming estimates for AR(p) processes with nonnegative innovations having 0 as its left endpoint and satisfying one of two types of regular variation property on the innovation distribution. They allow innovation distributions F which are regularly varying at 0, the left endpoint, and satisfy some moment condition or the tail of F is regularly varying at infinity and some reciprocal moment condition, $EZ^{-\beta}$ finite for suitable positive β, where Z has distribution function F. Feigin et al. (1996) continues this study with a linear programming estimate for the case of a nonnegative moving average process. The technique has also been applied to positive nonlinear time series in Brown et al (1996) and Datta et al (1998). In addition to their relation to linear programming estimation procedures, nonnegative time series have also been considered in Anděl (1989), Anděl (1991), and Datta and McCormick (1995). Furthermore, modeling issues in connection to causal nonnegative time series are addressed in Tsai and Chan (2006). We mention that a very good survey of modeling phenomena with heavy tailed stochastic structures and ensuing estimation issues may be found in Resnick (1997) and Resnick (2007) where many references to applications of these methods may be found.

We make a comparison with the estimation procedure developed in this paper with the linear programming estimates. This is an apt comparison for the following reason. For a nonnegative AR(1) model, the linear programming estimate reduces to the estimate proposed in Davis and McCormick (1989), namely $\min_{1\leq t<n}(X_{t+1}/X_t)$, where X_t denotes the AR(1) process. The estimate proposed in this paper is a natural extension to the Davis and McCormick (1989) estimate based on extreme value considerations so that both the linear programming and the current estimation procedure represent higher order extensions of the same estimate for the AR(1) model but based on different rationales.

A first observation is that our estimation procedure is especially easy to implement. Indeed, the coefficients in the causal representation are estimated by a ratio of two sample values, as simple an estimate as one could imagine. The parameters of the ARMA model are then estimated as solutions to a system of linear equations. This compares with the linear programming estimates which require setting up and solving a linear programming problem, a more difficult task. Furthermore, the limiting distribution of our estimates is explicit and tractable, whereas the limiting distribution for the linear programming estimates is given as the solution to a random linear programming problem, rendering finding quantiles of the limiting distribution to be a nontrivial simulation problem, since analytically obtaining the asymptotic behavior for tail areas of the limit distribution appears to be a significantly challenging problem.

As remarked in Calder and Davis (1998), second-order based estimation methods for the ARMA model parameters perform well when the innovations are heavy-tailed. Our estimation procedure is more in keeping with the second-order estimates developed in Davis and Resnick (1986) with which we make a comparison through a simulation study. This is presented in section 3.5 and demonstrates a favorable performance for the extreme value method presented in this paper.

3.2 Estimation for nonnegative time series

Although the main purpose in this paper is to study extreme value based estimates of the parameters of ARMA models, it is more natural to begin with general infinite order moving average processes and apply the results gained in that setting to ARMA models. Therefore consider the model

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j} \tag{3.2.1}$$

where the i.i.d. innovations Z_t are nonnegative random variables having distribution function F for which $\bar{F} = 1 - F$ is regularly varying at infinity with index $-\beta$, written $\bar{F} \in RV_{-\beta}$. Our concern is with nonnegative time series and so we will assume that the coefficients c_i are nonnegative and further that $\sum_{i \geq 0} c_i^\delta < \infty$ for some $0 < \delta < \beta \wedge 1$. Under these assumptions the series converges almost surely. See Resnick (1987) Section 4.5. The almost sure convergence of such series can be established under weaker conditions (see, e.g., Mikosch and Samorodnitsky (2000)) but this suffices for our application to ARMA models. Our goal is to capitalize on the behavior of extreme value estimators over traditional estimators when $0 < \beta < 2$. In this heavy-tail regime, extreme value estimators converge at a rate faster than square root n. This contrasts with estimators whose asymptotic behavior depends on the central part of the innovation distribution when a second or higher moment is finite. To motivate the form of the estimate, let us define the index $1 \le t^* = t^*_n \le n$ such that

$$X_{t^*} = \max_{1 \le i \le n} X_i. \tag{3.2.2}$$

For the purpose of motivation we will assume F to be continuous thereby ensuring t^* is almost surely well defined. We shall assume in addition to our other assumptions on the c_i that $c_0 = 1 > c_i, i > 0$. It is not essential that the leading coefficient is the maximal coefficient and its value is 1 but these assumptions make bookkeeping easier. It is intuitive that in the heavy-tail regime that the maximal observation should be associated with the observation having the maximal innovation receiving the highest weight. Then if the weights decay appropriately

$$X_{t^*} \text{ approximately equals } Z_{t^*} \tag{3.2.3}$$

and if $1 \le k^* = k^*_n \le n$ is such that

$$Z_{k^*} = \max_{1 \le i \le n} Z_i, \tag{3.2.4}$$

then with high probability

$$t^* = k^*. \tag{3.2.5}$$

These relations will be made rigorous in section 3.3. From (3.2.3) one then has

$$X_{t^*+i} \approx c_i Z_{t^*}$$

which leads to our estimate

$$\hat{c}_i = \frac{X_{t^*+i}}{X_{t^*}}, \ i \ge 1. \tag{3.2.6}$$

Remark: In proposing the estimators $\hat{c}_i$ in (3.2.6), we have done so for the model (3.2.1) with $c_0 = 1$ being the unique maximal coefficient among the nonnegative coefficients c_i. Notice that the estimators $\hat{c}_i$ being homogeneous functions of the data of degree 0 are invariant with respect to a change of scale. That is, for any $a > 0$

$$\hat{c}_i(X) = \hat{c}_i(aX)$$

where $X = (X_n)_{n\geq 1}$. Thus for the estimation scheme to succeed in the infinite order moving average case one constraint on the coefficients needs to be imposed to remove the ambiguity resulting from the $\hat{c}_i$ being estimates of the parameters up to a scale factor. The simplest such constraint is to have a fixed known value of one coefficient. Actually, two pieces of information are needed to implement this method for the MA(∞) model: a known value for c_{k_0} for some k_0 and knowing the index l_0 of the maximal coefficient. In this more general case the estimator in (3.2.6) should be replaced by

$$\hat{c}_i = c_{k_0}\frac{X_{t^*-l_0+i}}{X_{t^*-l_0+k_0}}. \tag{3.2.7}$$

In the ARMA(p, q) case, one constraint is automatically satisfied. The coefficient $c_0 = 1$. Therefore, if the location of the maximal coefficient is l_0, determined by say applying standard time series estimation to the data $X_1 - \bar{X}, \ldots, X_n - \bar{X}$, e.g. least squares estimation (Brockwell and Davis (1987) section 8.7), then the extreme value estimates become

$$\hat{c}_i = \frac{X_{t^*-l_0+i}}{X_{t^*-l_0}}, \quad i \geq 1. \tag{3.2.8}$$

3.3 Asymptotics

In the following it is assumed that

$$c_0 = 1 > c_i \geq 0, i \geq 1 \text{ and } \sum_{i=0}^{\infty} c_i^{\delta} < \infty, \tag{3.3.1}$$

for some $0 < \delta < \beta \wedge 1$.

$$F(0) = 0 \text{ and } 1 - F = \bar{F} \in RV_{-\beta}. \tag{3.3.2}$$

Let X_n be the model defined in (3.2.1) with innovations having distribution function F. Define

$$M_n = \max_{1 \le i \le n} X_i \text{ and } W_n = \max_{1 \le i \le n} Z_i. \tag{3.3.3}$$

From Resnick (1987), Proposition 4.29, we have that M_n and W_n share the same weak limit behavior. The following lemma provides a strengthening to that.

Lemma 3.3.1. *Under assumptions* (3.3.1) *and* (3.3.2),

$$\lim_{n \to \infty} \frac{M_n}{W_n} = 1 \qquad \textit{in probability.}$$

Proof. Define

$$b_n = F^{\leftarrow}\left(1 - \frac{1}{n}\right) = \inf\{t : F(t) \ge 1 - \frac{1}{n}\}. \tag{3.3.4}$$

For a real a and $\mathbf{v} \in \mathbb{R}^m$ we define $a\mathbf{v}$ to be the vector $(av_1, \ldots, av_m)$ where $\mathbf{v} = (v_1, \ldots, v_m)$. Next, we define a sequence of point processes $\mathfrak{N}_{n,m}, n \ge 1$ in $E = (0, \infty]^{m+2}$ by

$$\mathfrak{N}_{n,m} = \sum_{t=1}^{n} \epsilon_{((Z_{t-m}, Z_{t-m+1}, \ldots, Z_t, Z_t)/b_n)}. \tag{3.3.5}$$

Notice that the last two coordinates of the points of $\mathfrak{N}_{n,m}$ are equal and ϵ_x denotes the degenerate measure at x. The argument used to prove Theorem 2.2 in Davis and Resnick (1985) can be applied to show that

$$\mathfrak{N}_{n,m} \xrightarrow{d} \mathfrak{N}_m = \sum_{k=1}^{\infty} \epsilon_{(0,\ldots,0,j_k,j_k)} + \sum_{k=1}^{\infty} \epsilon_{(j_k,0,\ldots,0)} + \ldots + \sum_{k=1}^{\infty} \epsilon_{(0,\ldots,0,j_k,0,0)} \tag{3.3.6}$$

where $\sum_{k=1}^{\infty} \epsilon_{j_k}$ is PRM(μ), Poisson random measure with mean measure μ determined by

$$\mu((x, \infty]) = x^{-\beta}, x > 0. \tag{3.3.7}$$

The limit point process $\mathfrak{N}_m$ is obtained by taking the points of $\sum_{k=1}^{\infty} \epsilon_{j_k}$ and for each point j_k create $m+1$ points in $(0,\infty)^{m+2}$ with j_k in one of the first m coordinates with all others being 0 and another point where the last two coordinate values are j_k with the others being 0. Applying the continuous mapping theorem to (3.3.6), we obtain the following weak convergence of point process

$$\mathfrak{M}_{n,m} = \sum_{t=1}^{n} \epsilon_{(b_n^{-1} \sum_{i=0}^{m} c_i Z_{t-i}, b_n^{-1} Z_t)} \xrightarrow{d} \mathfrak{M}_m = \sum_{k=1}^{\infty} \epsilon_{(j_k, j_k)} + \sum_{k=1}^{\infty} \epsilon_{(c_1 j_k, 0)} + \ldots + \sum_{k=1}^{\infty} \epsilon_{(c_m j_k, 0)}. \quad (3.3.8)$$

By setting

$$X_{t,m} = \sum_{i=0}^{m} c_i Z_{t-i},$$

it then follows from (3.3.8) that for $x, y > 0$

$$\begin{aligned}
&\lim_{n\to\infty} P(b_n^{-1} \max_{1\le t\le n} X_{t,m} \le x, b_n^{-1} \max_{1\le t\le n} Z_t \le y) \\
&\quad = \lim_{n\to\infty} P\Big(\mathfrak{M}_{n,m}\big(([0,x]\times[0,y])^c\big) = 0\Big) \\
&\quad = P\Big(\mathfrak{M}_m\big(([0,x]\times[0,y])^c\big) = 0\Big) \\
&\quad = P\Big(\sum_{k=1}^{\infty} \epsilon_{(j_k,j_k)}\big(([0,x]\times[0,y])^c\big) = 0\Big) \\
&\quad = P\Big(\sum_{k=1}^{\infty} \epsilon_{j_k}\big([0, x\wedge y]^c\big) = 0\Big) \\
&\quad = \exp\{-(x\wedge y)^{-\beta}\}, \qquad (3.3.9)
\end{aligned}$$

where we used the fact that $0 \le c_i < c_0 = 1$ for $i \ge 1$ so that the condition $(j_k, j_k) \in [0,x]\times[0,y]$ implies $(c_i j_k, 0) \in [0,x]\times[0,y]$ for $1 \le i \le m$.

Summarizing with $M_{n,m} = \max_{1\le i\le n} X_{i,m}$ we have that

$$b_n^{-1}(M_{n,m}, W_n) \xrightarrow{d} H \quad (3.3.10)$$

where $H(x,y) = \exp\{-(x \wedge y)^{-\beta}\},\ x, y > 0$. We note two things concerning this limit law. Firstly, the limit does not depend on m which is a consequence of the assumption $c_0 = 1$ is maximal and, secondly, H has all its mass on the diagonal $\{(x,x) : x \geq 0\}$. Letting $\xi \sim \Phi_\beta$ with $\Phi_\beta(x) = \exp\{-x^{-\beta}\}$, we have that

$$b_n^{-1}(M_{n,m}, W_n) \xrightarrow{d} (\xi, \xi). \tag{3.3.11}$$

Let

$$X_{t,m}^* = X_t - X_{t,m} = \sum_{i=m+1}^{\infty} c_i Z_{t-i},$$

and set $\gamma_m = \max_{i>m} c_i$. Then by Resnick (1987), Proposition 4.29,

$$\lim_{n\to\infty} P(b_n^{-1} \max_{1\leq t\leq n} X_{t,m}^* \leq x) = \exp\{-\gamma_m^\beta x^{-\beta}\}. \tag{3.3.12}$$

Since $X_{t,m}^*$ is a nonnegative process, we have by (3.3.9)

$$\limsup_{n\to\infty} P(b_n^{-1} M_n \leq x, b_n^{-1} W_n \leq y) \leq \lim_{n\to\infty} P(b_n^{-1} \max_{1\leq t\leq n} X_{t,m} \leq x, b_n^{-1} W_n \leq y) = \Phi_\beta(x \wedge y). \tag{3.3.13}$$

Furthermore,

$$\begin{aligned}
&\liminf_{n\to\infty} P(b_n^{-1} M_n \leq x, b_n^{-1} W_n \leq y) \\
&\qquad \geq \lim_{n\to\infty} P(b_n^{-1} \max_{1\leq t\leq n} X_{t,m} \leq x - \epsilon, b_n^{-1} W_n \leq y) - \lim_{n\to\infty} P(b_n^{-1} \max_{1\leq t\leq n} X_{t,m}^* > \epsilon) \\
&\qquad = \Phi_\beta((x - \epsilon) \wedge y) - (1 - \Phi_\beta(\epsilon/\gamma_m)).
\end{aligned}$$

Letting m tend to infinity and then ϵ tend to zero, we obtain

$$\liminf_{n\to\infty} P(b_n^{-1} M_n \leq x, b_n^{-1} W_n \leq y) \geq \Phi_\beta(x \wedge y), \tag{3.3.14}$$

so that putting (3.3.13) and (3.3.14) together

$$b_n^{-1}(M_n, W_n) \xrightarrow{d} (\xi, \xi). \tag{3.3.15}$$

Hence, for $\epsilon > 0$ letting

$$D_\epsilon = \{(x, y) : x/y > 1 + \epsilon, x, y > 0\},$$

we deduce by (3.3.15) that

$$\lim_{n\to\infty} P(M_n/W_n > 1 + \epsilon) = \lim_{n\to\infty} P(b_n^{-1}(M_n, W_n) \in D_\epsilon) = P((\xi, \xi) \in D_\epsilon) = 0$$

establishing the lemma. □

As indicated in section 2.2, our extreme value estimator is motivated by the closeness of X_{t^*} and Z_{t^*}, where t^* provides the index of the maximal X_i for $1 \le i \le n$. The following lemma makes that relationship precise. We set

$$t^* = t_n^* = \min\{i : 1 \le i \le n \text{ and } X_i = M_n\}.$$

Lemma 3.3.2. *Under assumptions* (3.3.1) *and* (3.3.2)

$$\lim_{n\to\infty} \frac{X_{t^*}}{Z_{t^*}} = 1 \quad \textit{in probability .}$$

Proof. The proof is accomplished by a blocking argument. The idea of the proof is that for any $\epsilon > 0$ the expected number of $Z_i, 1 \le i \le n$ to exceed ϵb_n remains bounded as n tends to infinity and therefore these large innovations will tend to fall in separate blocks allowing us to capitalize on this feature of isolated large values. Let m_n and l_m be two sequences of positive integers tending to infinity with n such that

$$m_n = o(n) \quad \text{and} \quad l_m = o(m).$$

To ease notation we suppress the subscript n in the following. Define disjoint blocks

$$I_k = [(k-1)m+1, km-l], \quad J_k = [km-l+1, km]$$

and

$$J_{\lfloor n/m \rfloor+1} = \big[\lfloor n/m \rfloor m+1, n\big]$$

where $k = 1, \ldots, \lfloor n/m \rfloor$ and where, for integers $i \leq j$, $[i,j]$ denotes the integer interval $\{i, i+1, \ldots, j\}$. Fix $\epsilon > 0$ and define counting variables for $1 \leq k \leq \lfloor n/m \rfloor$

$$n_k = \sum_{i \in I_k} \mathbf{1}[Z_i > \epsilon b_n], \quad n'_k = \sum_{i \in J_k} \mathbf{1}[Z_i > \epsilon b_n],$$

and $n'_{\lfloor n/m \rfloor+1} = \sum_{i \in J_{\lfloor n/m \rfloor+1}} \mathbf{1}[Z_i > \epsilon b_n]$.

Our first step is to show that with high probability the big blocks, I_k, contain at most one large innovation and the small blocks, J_k, contain none. We assert that

$$P\Big(\bigcup_{k=1}^{\lfloor n/m \rfloor} (n_k \geq 2)\Big) = o(1), P\Big(\bigcup_{k=1}^{\lfloor n/m \rfloor} (n'_k \geq 1)\Big) = o(1), \text{ and } P\big(n'_{\lfloor n/m \rfloor+1} \geq 1\big) = o(1), \tag{3.3.16}$$

as $n \to \infty$. To establish (3.3.16) observe that

$$P(n_1 \geq 2) \leq m^2(1 - F(\epsilon b_n))^2 = O\left(\frac{m^2}{n^2}\right)$$

so that

$$P\Big(\bigcup_{k=1}^{\lfloor n/m \rfloor} (n_k \geq 2)\Big) = O\left(\frac{m}{n}\right) = o(1).$$

Similarly,

$$P(n'_1 \geq 1) = O\left(\frac{l}{n}\right)$$

which implies

$$P\Big(\bigcup_{k=1}^{\lfloor n/m \rfloor}(n'_k \geq 1)\Big) = O\left(\frac{l}{m}\right) = o(1).$$

Finally,

$$P(n'_{\lfloor n/m \rfloor+1} \geq 1) = O\left(\frac{m}{n}\right) = o(1).$$

Thus we see (3.3.16) holds. Our next step is to show that the index t^* providing the index of the maximal observation is with high probability contained in one of the big blocks. Specifically, we show

$$P\Big(t^* \in \bigcup_{k=1}^{\lfloor n/m \rfloor+1} J_k\Big) = o(1) \tag{3.3.17}$$

as n tends to infinity. This follows since

$$\begin{aligned} P\Big(t^* \in \bigcup_{k=1}^{\lfloor n/m \rfloor+1} J_k\Big) &\leq P\Big(M_n = \max\{X_t : t \in \bigcup_{k=1}^{\lfloor n/m \rfloor+1} J_k\}\Big) \\ &\leq P(b_n^{-1}M_n \leq \epsilon) + P\Big(\max\{X_t : t \in \bigcup_{k=1}^{\lfloor n/m \rfloor+1} J_k\} > \epsilon b_n\Big) \\ &\leq P(b_n^{-1}M_n \leq \epsilon) + \left(\frac{n}{m}l + m\right) P\left(X_1 > \epsilon b_n\right). \end{aligned}$$

Hence

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} P\Big(t^* \in \bigcup_{k=1}^{\lfloor n/m \rfloor+1} J_k\Big) \leq \lim_{\epsilon \to 0} \Phi_\beta(\epsilon) = 0.$$

Thus (3.3.17) holds. Define the event

$$E_n = \bigcap_{k=1}^{\lfloor n/m \rfloor}(n_k \leq 1) \cap \bigcap_{k=1}^{\lfloor n/m \rfloor+1}(n'_k = 0) \cap (t^* \notin \bigcup_{k=1}^{\lfloor n/m \rfloor+1} J_k).$$

By (3.3.16) and (3.3.17)

$$\lim_{n \to \infty} P(E_n) = 1. \tag{3.3.18}$$

On E_n suppose $t^* \in I_k$. Since on E_n, $(n_k \le 1)$, three cases are possible. These are:

$$\begin{cases} \text{case (1): } Z_{t^*} > \epsilon b_n \text{ and } Z_t \le \epsilon b_n & \text{for } t \in I_k \setminus \{t^*\} \\ \text{case (2): } Z_{t^*} \le \epsilon b_n \text{ and } Z_t > \epsilon b_n & \text{for some unique } t \in I_k \setminus \{t^*\} \\ \text{case (3): } Z_t \le \epsilon b_n & \text{for all } t \in I_k \ . \end{cases}$$

Before discussing these cases, we note that since $0 \le c_i < c_0 = 1$ for all $i \ge 1$,

$$0 < \bar{c} = \sum_{i=0}^{\infty} c_i \le \sum_{i=0}^{\infty} c_i^{\delta} < \infty$$

where $\delta < 1$ is chosen so as to satisfy (3.3.1). Let

$$\tilde{X}_{t,m} = \sum_{i=1}^{m} c_i Z_{t-i} = X_{t,m} - Z_t.$$

Then under case (1), we find providing t^* is greater than l

$$0 \le \tilde{X}_{t^*,l} \le \bar{c}\epsilon b_n$$

because for $1 \le i \le l$, if $t^* > l$, then either $t^* - i \in I_k$ or $t^* - i \in J_{k-1}$ if $k \ge 2$.

Fix $\eta > 0$ and choose $0 < \epsilon < \eta/\bar{c}$. Then on $E_n \cap (t^* > l)$, we have under case 1

$$\tilde{X}_{t^*,l} < \eta b_n. \tag{3.3.19}$$

Before analyzing case (2) we make the observation that

$$0 \le \zeta = \sup_{i>0} c_i < 1.$$

This is obvious from the assumption that $0 \le c_i < c_0 = 1$ for $i \ge 1$ and c_n tend to zero as n tends to infinity. We remark that it is only in this part of the proof that the assumption that the coefficients c_i have a unique maximal value is used.

Let us note that case 2 divides into two subcases. We first consider the case that for some $t^{'} \in I_k$ with $t^{'} < t^*$, we have $Z_{t'} > \epsilon b_n$. Then observe that in view of the decomposition

$$X_{t^*} = X_{t^*,t^*-t'-1} + c_{t^*-t'} Z_{t'} + X^*_{t^*,t^*-t'} \geq Z_{t'},$$

we have

$$\begin{aligned}
&(t^* \in I_k, Z_{t'} > \epsilon b_n \text{ for some } t^{'} \in I_k \text{ with } t^{'} < t^*) \\
&\quad \subset \bigcup_{t'<t, t', t \in I_k} \Big(X_{t,t-t'-1} + X^*_{t,t-t'} > (1-\zeta)\epsilon b_n, Z_{t'} > \epsilon b_n \Big).
\end{aligned}$$

Hence

$$P\Big(t^* \in I_k, Z_{t'} > \epsilon b_n \text{ for some } t^{'} \in I_k \text{ with } t^{'} < t^*) \leq m^2 P^2(X_1 > (1-\zeta)\epsilon b_n\Big) = O\left(\frac{m^2}{n^2}\right).$$

Thus

$$\lim_{n\to\infty} P\Big(\bigcup_{k=1}^{\lfloor n/m \rfloor} (t^* \in I_k, Z_{t'} > \epsilon b_n \text{ for some } t^{'} \in I_k \text{ with } t^{'} < t^*)\Big) = 0. \tag{3.3.20}$$

For the other subcase of case 2, we consider the event that for some $t^{'} \in I_k$ with $t^{'} > t^*$, we have $Z_{t'} > \epsilon b_n$. Then simply

$$\tilde{X}_{t^*,l} = \sum_{i=1}^{l} c_i Z_{t^*-i} \leq \bar{c}\epsilon b_n < \eta b_n. \tag{3.3.21}$$

Under case 3, we have for the same reasons as in case 1 that provided $t^* > l$

$$\tilde{X}_{t^*,l} < \eta b_n. \tag{3.3.22}$$

Thus we find that by (3.3.19), (3.3.21), (3.3.22) and our analysis of case 2 that

$$E_n \cap (t^* > l) \subset (b_n^{-1}\tilde{X}_{t^*,l} < \eta) \cup \bigcup_{k=1}^{\lfloor n/m \rfloor} \left(t^* \in I_k, Z_{t'} > \epsilon b_n \text{ for some } t^{'} \in I_k \text{ with } t^{'} < t^*\right).$$

Therefore in view of (3.3.18), (3.3.20) and for any $\epsilon > 0$

$$P(t^* \leq l) \leq P(b_n^{-1} M_n \leq \epsilon) + l P(X_1 > \epsilon b_n)$$

showing that

$$\lim_{n\to\infty} P(t^* \leq l) = 0,$$

we have for any $\eta > 0$

$$\lim_{n\to\infty} P(b_n^{-1}\tilde{X}_{t^*,l} < \eta) = 1.$$

We record this result as

$$\lim_{n\to\infty} b_n^{-1}\tilde{X}_{t^*,l} = 0 \quad \text{in probability.} \tag{3.3.23}$$

Fix a positive integer m. Then since $l = l_m$ tends to infinity, for large enough n we have $l > m$. Thus by nonnegativity for all n large enough

$$X_{t,l}^* \leq X_{t,m}^* \quad \text{for all } t \geq 1. \tag{3.3.24}$$

Thus for any $\epsilon > 0$ using (3.3.12) and (3.3.24)

$$\lim_{n\to\infty} P(b_n^{-1} \max_{1\leq t\leq n} X_{t,l}^* > \epsilon) \leq 1 - \Phi_\beta(\epsilon/\gamma_m).$$

Letting m tend to infinity we see that

$$\lim_{n\to\infty} b_n^{-1} \max_{1\leq t\leq n} X_{t,l}^* = 0 \quad \text{in probability .} \tag{3.3.25}$$

From (3.3.23) and (3.3.25) we obtain the main step in the proof of lemma 3.3.2, namely that

$$\lim_{n\to\infty} b_n^{-1} \sum_{i=1}^{\infty} c_i Z_{t^*-i} = \lim_{n\to\infty} b_n^{-1} X^*_{t^*,0} = 0 \quad \text{in probability .} \tag{3.3.26}$$

The proof is now completed by noting that

$$\frac{Z_{t^*}}{X_{t^*}} = 1 - \frac{X^*_{t^*,0}}{X_{t^*}} = 1 - \left(\frac{X_{t^*}}{b_n}\right)^{-1} \left(\frac{X^*_{t^*,0}}{b_n}\right) \xrightarrow{p} 1, \quad \text{as } n \to \infty,$$

using (3.3.15) and (3.3.26). □

In view of Lemmas 3.3.1 and 3.3.2 the following result is to be expected.

Lemma 3.3.3. *Assuming* (3.3.1) *and* (3.3.2) *hold,*

$$t^* - k^* \xrightarrow{a.s} 0, \quad \text{as } n \to \infty.$$

Proof. Since the variables $t^* = t_n^*$ and $k^* = k_n^*$ are discrete it suffices to show convergence to zero in probability. Denote the first n order statistics for the Z_i by $\{Z_{1,n} \leq \ldots \leq Z_{n,n}\}$. Thus for any $0 < \epsilon < 1$

$$\begin{aligned} P(|t^* - k^*| > \epsilon) &\leq P(|t^* - k^*| \geq 1) \\ &\leq P(Z_{t^*} \leq Z_{n-1,n}). \end{aligned} \tag{3.3.27}$$

By Lemmas 3.3.1 and 3.3.2

$$\lim_{n\to\infty} \frac{Z_{k^*}}{Z_{t^*}} = 1 \quad \text{in probability .} \tag{3.3.28}$$

Hence

$$\lim_{n\to\infty} P(b_n^{-1} Z_{k^*} \leq x, b_n^{-1} Z_{t^*} \leq y) = \Phi_\beta(x \wedge y). \tag{3.3.29}$$

By Leadbetter, Lindgren, and Rootzén (1983), Theorem 2.3.2,

$$\lim_{n\to\infty} P(b_n^{-1}Z_{n,n} \le x, b_n^{-1}Z_{n-1,n} \le y) = P(\xi_1 \le x, \xi_2 \le y), \tag{3.3.30}$$

where (ξ_1, ξ_2) has an absolutely continuous joint distribution such that $P(\xi_1 \le \xi_2) = 0$. Note by (3.3.29)

$$\lim_{n\to\infty} b_n^{-1}(Z_{k^*} - Z_{t^*}) = 0 \quad \text{in probability}$$

and therefore

$$\begin{aligned}\lim_{n\to\infty} P(|t^* - k^*| > \epsilon) &\le \lim_{n\to\infty} P(b_n^{-1}Z_{k^*} - b_n^{-1}(Z_{k^*} - Z_{t^*}) \le b_n^{-1}Z_{n-1,n}) \\ &= P(\xi_1 \le \xi_2) = 0.\end{aligned}$$

□

A straightforward calculation shows for $\hat{c}_i$ defined in (3.2.6) under the model (3.2.1) that

$$\hat{c}_i - c_i = \frac{1}{X_{t^*}}\Big(\sum_{j=1}^{\infty}(c_{j+i} - c_i c_j)Z_{t^*-j} + \sum_{j=1}^{i} c_{i-j}Z_{t^*+j}\Big), \quad i \ge 1.$$

For integers $i, k \ge 1$ define

$$U_{i,k} = \sum_{j=1}^{\infty}(c_{j+i} - c_i c_j)Z_{k-j} \quad \text{and} \quad V_{i,k} = \sum_{j=1}^{i} c_{i-j}Z_{k+j}.$$

Then

$$b_n(\hat{c}_1 - c_1, \ldots, \hat{c}_m - c_m) = \frac{b_n}{X_{t^*}}\left(U_{1,t^*} + V_{1,t^*}, \ldots, U_{m,t^*} + V_{m,t^*}\right). \tag{3.3.31}$$

The following lemma will be useful in analyzing the asymptotics of the sequence defined in (3.3.31). Let

$$U_i = U_{i,1} \quad \text{and} \quad V_i = V_{i,1}, \quad i \ge 1.$$

Lemma 3.3.4. *Under* (3.3.1) *and* (3.3.2),

$$\left(\frac{Z_1}{b_n}, U_1, \dots, U_m, Z_2, \dots, Z_{m+1} | k^* = 1\right) \xrightarrow{d} (\xi, U_1, \dots, U_m, Z_2, \dots, Z_{m+1})$$

as $n \to \infty$ *where* $\xi \sim \Phi_\beta$ *and is independent of*

$$(U_1, \dots, U_m, Z_2, \dots, Z_{m+1}).$$

Proof. We first make the observation that

$$\lim_{n\to\infty} nP(W_n \le z) = \lim_{n\to\infty} nF^n(z) = 0. \tag{3.3.32}$$

Then

$$\begin{aligned}
&P\Big(\frac{Z_1}{b_n} \le z_1, U_1 \le u_1, \dots, U_m \le u_m, Z_2 \le z_2, \dots Z_{m+1} \le z_{m+1} | k^* = 1\Big)\\
&\qquad = nP(U_1 \le u_1, \dots, U_m \le u_m) P\Big(\frac{Z_1}{b_n} \le z_1, Z_2 \le z_2, \dots, Z_{m+1} \le z_{m+1}, W_n \le Z_1\Big)\\
&\qquad = nP(U_1 \le u_1, \dots, U_m \le u_m) P\Big(\frac{1}{b_n} \max_{j\in\{1,m+2,\dots,n\}} Z_j \le z_1, \max_{m+2\le j\le n} Z_j \le Z_1\Big)\\
&\qquad\qquad \cdot P(Z_2 \le z_2) \dots P(Z_{m+1} \le z_{m+1}) + o(1)
\end{aligned} \tag{3.3.33}$$

where we used $U_i \in \sigma(Z_{-j}, j \ge 0), i \ge 1$ and where we used (3.3.32) which allowed us to replace W_n in the next to last inequality with $\max_{m+2\le j\le n} Z_j$. The result follows from (3.3.33) and

$$\lim_{n\to\infty} nP(b_n^{-1} \max_{j\in\{1,m+2,\dots,n\}} Z_j \le z_1, \max_{m+2\le j\le n} Z_j \le Z_1) = \lim_{n\to\infty} P(b_n^{-1} W_n \le z_1) = \Phi_\beta(z).$$

□

The first result describes the weak limiting behavior of our estimates for the moving average coefficients. The asymptotic distribution of our parameter estimates then follows from this result by the usual delta method.

Theorem 3.3.1. *Under* (3.3.1) *and* (3.3.2),

$$b_n(\hat{c}_1 - c_1, \ldots, \hat{c}_m - c_m) \xrightarrow{d} \frac{1}{\xi}(U_1 + V_1, \ldots, U_m + V_m)$$

where $\xi \sim \Phi_\beta$ *and* ξ *is independent of* $(U_1, \ldots, U_m, V_1, \ldots, V_m)$.

Proof. Using (3.3.31), Lemmas 3.3.1 and 3.3.3 and setting for any $\epsilon > 0$ and real y_i

$$z_i = z_{i,\epsilon} = (1-\epsilon)y_i \vee (1+\epsilon)y_i,$$

we have

$$\begin{aligned}
&\limsup_{n\to\infty} P(b_n(\hat{c}_i - c_i) \le y_i, 1 \le i \le m) \\
&\qquad = \limsup_{n\to\infty} P\Big(\frac{b_n}{X_{t^*}}(U_{i,t^*} + V_{i,t^*}) \le y_i, 1 \le i \le m\Big) \\
&\qquad \le \limsup_{n\to\infty} P\Big(\frac{b_n}{Z_{k^*}}(U_{i,t^*} + V_{i,t^*}) \le z_i, 1 \le i \le m\Big) \\
&\qquad = \limsup_{n\to\infty} P\Big(\frac{b_n}{Z_{k^*}}(U_{i,k^*} + V_{i,k^*}) \le z_i, 1 \le i \le m\Big). \qquad (3.3.34)
\end{aligned}$$

Next we note that

$$\begin{aligned}
P\Big(\frac{b_n}{Z_{k^*}}(U_{i,k^*} + V_{i,k^*}) \le z_i, 1 \le i \le m\Big) &= \sum_{k=1}^{n} P\Big(\frac{b_n}{Z_k}(U_{i,k} + V_{i,k}) \le z_i, 1 \le i \le m, k^* = k\Big) \\
&= nP\Big(\frac{b_n}{Z_1}(U_i + V_i) \le z_i, 1 \le i \le m, k^* = 1\Big) \\
&= P\Big(\frac{b_n}{Z_1}(U_i + V_i) \le z_i, 1 \le i \le m \Big| k^* = 1\Big). \qquad (3.3.35)
\end{aligned}$$

Now observe that

$$\frac{b_n}{Z_1}(U_1+V_1,\ldots,U_m+V_m)=h\Big(\frac{Z_1}{b_n},U_1,\ldots,U_m,Z_2,\ldots,Z_{m+1}\Big)$$

for

$$h(w,u_1,\ldots,u_m,z_2,\ldots,z_{m+1})=\frac{1}{w}(u_1+z_2,u_2+c_1z_2+z_3,\ldots,u_m+\sum_{j=1}^{m}c_{m-j}z_{1+j}).$$

Thus from (3.3.34), (3.3.35), Lemma 3.3.4 and the continuous mapping theorem

$$\limsup_{n\to\infty}P(b_n(\hat{c}_i-c_i)\le y_i,1\le i\le m)\le P\Big(\frac{1}{\xi}(U_i+V_i)\le z_i,1\le i\le m\Big).$$

Letting ϵ tend to zero, we obtain

$$\limsup_{n\to\infty}P(b_n(\hat{c}_i-c_i)\le y_i,1\le i\le m)\le P\Big(\frac{1}{\xi}(U_i+V_i)\le y_i,1\le i\le m\Big).$$

The corresponding lower bound is proved similarly, establishing the theorem. □

3.4 ARMA Models

The autoregressive moving average model, ARMA(p,q), can be written as

$$\Phi(B)X_t=\Theta(B)Z_t$$

where

$$\Phi(z)=1-\phi_1z-\ldots-\phi_pz^p\text{ and }\Theta(z)=1+\theta_1z+\ldots+\theta_qz^q$$

and B denotes the backward shift operator. With regard to this model, we make the following assumptions. The polynomials Φ and Θ have no common zeros and Φ has no roots in the closed unit disc. We further assume that the Z_t are i.i.d. nonnegative innovations having distribution function F for which $\bar{F} = 1 - F$ is regularly varying at infinity with index $-\beta$, written $\bar{F} \in RV_{-\beta}$.

By the assumptions on Φ, we have that $\xi(z) = \Theta(z)/\Phi(z)$ is analytic in a neighborhood of 0 with radius of convergence greater than 1 and writing

$$\xi(z) = \sum_{j=0}^{\infty} c_j z^j, \quad |z| \leq 1,$$

we have that for some constants k and $r > 1$

$$|c_j| \leq kr^{-j}, j \geq 0.$$

Therefore by Resnick (1987) section 4.5 the series

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j} \tag{3.4.1}$$

converges almost surely. It then follows that X_n satisfies $\Phi(B)X_n = \Theta(B)Z_n$ and so X_n given by (3.4.1) provides the stationary solution to the ARMA(p, q) recursion.

Let us observe that, if the parameters $\phi_i, \theta_j, 1 \leq i \leq p, 1 \leq j \leq q$ are nonnegative, then since we have assumed additionally that the innovations are nonnegative, it is evident that X_n is a nonnegative process. Furthermore, this entails that all the coefficients c_j in the MA(∞) process defined in (3.4.1) must be nonnegative. Indeed, if $c_i < 0$ for some $i \geq 0$, then noting that $c_0 = 1$ and, if $i \geq 1$ is the first index such that $c_i < 0$, we obtain

$$0 \leq \theta_i = c_i - \sum_{j=1}^{i \wedge p} \phi_j c_{i-j}$$

which leads to the contradiction that $c_i \geq 0$, where θ_i denotes a coefficient of the polynomial Θ if $i \leq q$ and where we take $\theta_i = 0$ for $i > q$. However, it is not necessary that all the causal ARMA model parameters be nonnegative for the process to be nonnegative. We refer to Tsai and Chan (2006) for information on this question.

We now apply Theorem 3.3.1 to the ARMA(p, q) model. From the relation

$$\Theta(z) = \xi(z)\Phi(z) = \Phi(z) \sum_{j=0}^{\infty} c_j z^j$$

where Φ and Θ are the ARMA(p, q) polynomials, we obtain the relations

$$\theta_i = c_i - \sum_{j=1}^{i \wedge p} \phi_j c_{i-j}, \quad i \geq 1 \tag{3.4.2}$$

where $\theta_i = 0$ for $i > q$ and otherwise equals the coefficient of z^i in Θ. Assuming that the first $p+q$ equations in (3.4.2) provide an invertible transformation between $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and $(c_1, \ldots, c_{p+q})$, we may solve for the ARMA parameters in terms of the c_i. Let

$$\phi_i = \psi_i(c_1, \ldots, c_{p+q}) \quad \text{and} \quad \theta_j = \psi_{p+j}(c_1, \ldots, c_{p+q})$$

denote these solutions to the linear equations in (3.4.2). Define the estimates

$$\hat{\phi}_i = \psi_i(\hat{c}_1, \ldots, \hat{c}_{p+q}), 1 \leq i \leq p \quad \text{and} \quad \hat{\theta}_j = \psi_{p+j}(\hat{c}_1, \ldots, \hat{c}_{p+q}), 1 \leq j \leq q. \tag{3.4.3}$$

In order to state the next result, it will be convenient to introduce some notation. We denote the parameter vectors by,

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)' \quad \text{and} \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$$

with estimates

$$\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)' \quad \text{and} \quad \hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)'.$$

Define the transformation $\Psi : \mathbb{R}^{p+q} \to \mathbb{R}^{p+q}$ by

$$\Psi(c_1, \ldots, c_{p+q}) = \left(\psi_1(c_1, \ldots, c_{p+q}), \ldots, \psi_{p+q}(c_1, \ldots, c_{p+q})\right)', \tag{3.4.4}$$

where the ψ_i are as in (3.4.3).

We take vectors to be column vectors. Let the differential of Ψ be denoted $\mathrm{D}\Psi : \mathbb{R}^{p+q} \to L(\mathbb{R}^{p+q}, \mathbb{R}^{p+q})$. Since a cleaner result exists in the AR(p) case, we present this case separately. Firstly, we discuss the estimates $\hat{\phi}_i$ in this case and the differential map $\mathrm{D}\Psi$ of the map in (3.4.4). From (3.4.2) we obtain

$$\begin{aligned} \phi_1 &= c_1 \\ \phi_i &= c_i - \sum_{j=1}^{i-1} c_{i-j}\phi_j, \quad 2 \leq i \leq p. \end{aligned} \tag{3.4.5}$$

Therefore the $\hat{\phi}_i$ are defined recursively by

$$\hat{\phi}_i = \hat{c}_i - \sum_{j=1}^{i-1} \hat{c}_{i-j}\hat{\phi}_j, \quad 1 \leq i \leq p. \tag{3.4.6}$$

Let $\partial_j\phi_i = \frac{\partial\phi_i}{\partial c_j}$. Then from (3.4.5) we obtain

$$\partial_k\phi_l = -\phi_{l-k} - c_{l-k} - \sum_{i=1}^{l-k-1} c_i\partial_k\phi_{l-i}, \quad k < l, l = 2, \ldots, p. \tag{3.4.7}$$

The matrix $\mathrm{D}\Psi = (\partial_j\phi_i)_{1\leq i,j\leq p}$ has its diagonal elements equal to 1 and is lower triangular. Its nonzero elements $\partial_j\phi_i$ with $i > j$ may be found recursively from the equations in (3.4.7). Notice that $\partial_k\phi_k = 1$ and for $k < l$ the expression for $\partial_k\phi_l$ involves only the expressions $\partial_k\phi_{k+1}, \ldots, \partial_k\phi_{l-1}$.

Corollary 3.4.1. *Let X_t be an autoregressive process with parameters $\phi_i \geq 0, 1 \leq i \leq p$ such that $\phi_1 + \ldots + \phi_p < 1$. Then under the usual conditions on the innovations distribution F*

$$b_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \frac{1}{\xi} D\Psi(c_1, \ldots, c_p)(U_1 + V_1, \ldots, U_p + V_p)'.$$

Proof. From (3.4.5) we find $c_i < 1, i = 1, \ldots, p$. This is clear for c_1, suppose the statement is true for c_l with $1 \leq l \leq i-1$. Then

$$c_i = \sum_{j=1}^{i \wedge p} c_{i-j}\phi_j \leq \sum_{j=1}^{p} \phi_j < 1.$$

Hence the statement is true for all $i \geq 1$. Therefore we may apply Theorem 3.3.1 to get

$$b_n(\hat{c}_1 - c_1, \ldots, \hat{c}_p - c_p) \xrightarrow{d} \frac{1}{\xi}(U_1 + V_1, \ldots, U_p + V_p). \tag{3.4.8}$$

Since $(\hat{\phi}_1, \ldots \hat{\phi}_p) = \Psi(\hat{c}_1, \ldots, \hat{c}_p)$, the result follows by the delta-method applied to (3.4.8). □

Example 1: Consider an AR(2) model. We find

$$\phi_1 = c_1, \quad \text{and} \quad \phi_2 = c_2 - c_1^2.$$

Hence

$$\mathrm{D}\Psi(c_1, c_2) = \begin{pmatrix} 1 & 0 \\ -2c_1 & 1 \end{pmatrix}.$$

Furthermore,

$$U_1 = \sum_{j=1}^{\infty} (c_{j+1} - c_1 c_j) Z_{1-j}, \quad V_1 = Z_2$$

and

$$U_2 = \sum_{j=1}^{\infty} (c_{j+2} - c_2 c_j) Z_{1-j}, \quad V_2 = c_1 Z_2 + Z_3.$$

Using the relation

$$c_{j+1} - c_1 c_j = \phi_2 c_{j-1}, \quad j \geq 1,$$

we can write

$$U_1 = \phi_2 \sum_{j=0}^{\infty} c_j Z_{-j} = \phi_2 X_0.$$

Next using the relation

$$c_{j+2} - c_2 c_j = \phi_1 \phi_2 c_{j-1},$$

we have

$$U_2 = \phi_1 \phi_2 X_0.$$

We find after some simplification

$$\begin{bmatrix} 1 & 0 \\ -2c_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} U_1 + V_1 \\ U_2 + V_2 \end{bmatrix} = \begin{bmatrix} \phi_2 X_0 + Z_2 \\ -\phi_1 \phi_2 X_0 - \phi_1 Z_2 + Z_3 \end{bmatrix}.$$

Hence

$$b_n(\hat{\phi}_1 - \phi_1, \hat{\phi}_2 - \phi_2) \xrightarrow{d} \frac{1}{\xi}(\phi_2 X_0 + Z_2, -\phi_1 \phi_2 X_0 - \phi_1 Z_2 + Z_3).$$

For a positive ARMA(p, q) process while $c_0 = 1$ the location of the maximal coefficient needs to be known or obtained from a preliminary estimate. We assume the location l_0 is known. The estimate of the coefficients c_i are now given in (3.2.8) and the parameters estimates for the ARMA(p, q) model are given in (3.4.3). With regard to the ARMA(p, q) model X_t, we recall our standing assumptions. The model is given by $\Phi(B)X_t = \Theta(B)Z_t$ where the model parameters $\phi_i, \theta_j, 1 \leq i \leq p, 1 \leq j \leq q$ are nonnegative. Φ and Θ have no common roots and Φ has no root in the closed unit disc. The innovations have distribution F with support on the positive half line and tail which is regularly varying at infinity with index $-\beta$.

Corollary 3.4.2. *Let X_t be an ARMA(p,q) process satisfying the above conditions. Let*

$$(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})' = \Psi(\hat{c}_1, \ldots, \hat{c}_{p+q}).$$

Then

$$b_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \frac{1}{\xi} D\Psi(c_1, \ldots, c_{p+q})(U_1 + V_1, \ldots, U_{p+q} + V_{p+q})'$$

where $\xi \sim \Phi_\beta$ and is independent of $(U_1, \ldots, U_{p+q}, V_1, \ldots, V_{p+q})$.

Proof. The result is a direct application of the delta-method once the convergence of $b_n(\hat{c}_1 - c_1, \ldots, \hat{c}_{p+q} - c_{p+q})$ to the appropriate limit law is established. This follows by Theorem 3.3.1, the proof of which can be adapted to the case of a location of the maximal coefficient to be something other than 0. In that regard, the main changes are that

$$X_{t^*-l_0} \quad \text{is approximately } Z_{t^*-l_0}$$

and

$$k^* \quad \text{is approximately } t^* - l_0.$$

Further, observe that

$$\hat{c}_i - c_i = \frac{1}{X_{t^*-l_0}} \Big(\sum_{j=1}^{\infty} (c_{j+i} - c_i c_j) Z_{t^*-l_0-j} + \sum_{j=1}^{i} c_{i-j} Z_{t^*-l_0+j} \Big)$$

and one can imitate the steps of the proof of Theorem 3.3.1 to obtain the weak limit of $b_n(\hat{c}_i - c_i)_{1 \le i \le m}$ in this case with the same limit law resulting. □

Example 2: Consider the ARMA$(1,1)$ process

$$X_n = \phi_1 X_{n-1} + Z_n + \theta_1 Z_{n-1}, \quad n \ge 1.$$

We assume that $0 < \phi_1 < 1$ and $\phi_1 + \theta_1 > 1$. One readily checks that the stationary solution to this recursion is given by

$$X_n = Z_n + \sum_{i=1}^{\infty} (\phi_1 + \theta_1)\phi_1^{i-1} Z_{n-i}.$$

For this model $l_0 = 1$ and so we estimate the coefficients $c_i = (\phi_1 + \theta_1)\phi_1^{i-1}, \quad i \geq 1$ by

$$\hat{c}_i = \frac{X_{t^*-1+i}}{X_{t^*-1}}, \quad i \geq 1.$$

Solving (3.4.2) we find

$$\phi_1 = \frac{c_2}{c_1} \quad \text{and} \quad \theta_1 = \frac{c_1^2 - c_2}{c_1}$$

yielding

$$\mathrm{D}\Psi = \begin{bmatrix} \partial_1\phi_1 & \partial_2\phi_1 \\ \partial_1\theta_1 & \partial_2\theta_1 \end{bmatrix} = \begin{bmatrix} -c_2/c_1^2 & 1/c_1 \\ 1 + c_2/c_1^2 & -1/c_1 \end{bmatrix} = \frac{1}{\phi_1 + \theta_1} \begin{bmatrix} -\phi_1 & 1 \\ 2\phi_1 + \theta_1 & -1 \end{bmatrix}.$$

Next from the relations obtained from (3.4.2)

$$c_{j+1} - c_1 c_j = -\theta_1 c_j, \quad j \geq 1$$

and

$$c_{j+2} - c_2 c_j = -\phi_1 \theta_1 c_j, \quad j \geq 1$$

we obtain

$$U_1 = \sum_{j=1}^{\infty} (c_{j+1} - c_1 c_j) Z_{1-j} = \theta_1 (Z_1 - X_1)$$

and

$$U_2 = \sum_{j=1}^{\infty} (c_{j+2} - c_2 c_j) Z_{1-j} = \phi_1 \theta_1 (Z_1 - X_1).$$

Moreover,

$$V_1 = Z_2 \quad \text{and} \quad V_2 = (\phi_1 + \theta_1)Z_2 + Z_3.$$

$$\mathrm{D}\Psi(c_1, c_2)\begin{bmatrix} U_1 + V_1 \\ U_2 + V_2 \end{bmatrix} = \frac{1}{\phi_1 + \theta_1}\begin{bmatrix} \theta_1 Z_2 + Z_3 \\ \theta_1(\phi_1 + \theta_1)(Z_1 - X_1) + \phi_1 Z_2 - Z_3 \end{bmatrix}.$$

Thus we obtain

$$b_n(\hat{\phi}_1 - \phi_1, \hat{\theta}_1 - \theta_1) \xrightarrow{d} \frac{1}{\xi}\frac{1}{\phi_1 + \theta_1}(\theta_1 Z_2 + Z_3, \theta_1(\phi_1 + \theta_1)(Z_1 - X_1) + \phi_1 Z_2 - Z_3).$$

3.5 Simulation Study

In this section we study the performance of our extreme value estimation method through the following three models: AR(1), AR(2), and ARMA(3,1). For the first two models, a simulation study compared our estimation procedure with that of Davis and Resnick (1986). Whereas the third model was chosen as it represents a nonnegative time series where not all parameters are positive.

3.5.1 AR(1) Estimation Study

To study the performance of our estimator $\hat{\phi}_{BM} = \frac{X_{t^*+1}}{X_{t^*}}$, we generated 5,000 replications of the nonnegative time series $(X_0, X_1, \ldots, X_n)$ for two different sample sizes $(100, 500)$, where (X_t) is an AR(1) process satisfying the difference equation

$$X_t = \phi X_{t-1} + Z_t, \text{ for } 1 \leq t \leq n. \tag{3.5.1}$$

The autoregressive parameter ϕ is taken to be in the range from 0 to 1 guaranteeing a nonnegative time series when the innovations Z_t are taken with a Pareto distribution,

$$F_\beta(x) = \begin{cases} 1 - x^{-\beta}, & \text{if } x \geq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.5.2}$$

The Pareto distribution has a regularly varying tail distribution at infinity with index $-\beta$. For each of the six different β values considered, we computed $5,000$ estimates for $\phi = .2$ using $\hat{\phi}_{BM}$ and

$$\hat{\phi}_{DR} = \begin{cases} \sum_{t=1}^{n-1} X_t X_{t+1} / \sum_{t=1}^{n} X_t^2, & \text{if } 0 < \beta < 1 \\ \sum_{t=1}^{n} (X_t - \bar{X})(X_{t+1} - \bar{X}) / \sum_{t=1}^{n} (X_t - \bar{X})^2, & \text{if } 1 < \beta < 2 \end{cases}$$

where $\bar{X} = \sum_{t=1}^{n} X_t / n$. The means and standard deviations (written below in parentheses), of these estimates are reported in Table 3.1 along with the average length for a 90, 95, and 99 percent empirical confidence intervals with exact coverage. Since the main purpose of this section is to compare our estimator to Davis and Resnick's (1986) estimator, the confidence intervals were directly constructed from the empirical distributions of $n^{1/\beta}(\hat{\phi}_{BM} - \phi)$ and $(n/\log n)^{1/\beta}(\hat{\phi}_{DR} - \phi)$, respectively. While the limiting distribution of the parameter estimates by our method have an explicit expression, there is still a difficulty in obtaining percentiles for the limiting distribution needed to construct asymptotic confidence intervals in that the limiting distribution may involve the stationary distribution for the underlying ARMA model which is analytically intractable requiring such percentiles to be obtained empirically through a simulation. The analytic difficulties for the Davis-Resnick estimates are further complicated by the fact that the limiting distribution involves certain stable distributions whose exact parameter values are not made completely explicit in their paper. For these reasons, the approach taken for our simulation study was to obtain confidence intervals for the parameter estimates by obtaining directly the empirical distributions for the estimates. In this way, our confidence intervals have been constructed to have nearly exact coverage probabilities.

Table 1 shows that under the parameter values being considered our estimate does as least as well and more often better than the Davis-Resnick estimate. This is particularly true under the heavier tail models, i.e. when β is small. In that regime our estimate shows little bias and the average lengths of the confidence intervals are smaller than for the Davis-Resnick estimate, sometimes by a wide margin. While Table 1 generally shows a better performance for our estimate over the Davis-Resnick estimate, our simulations showed that with respect to certain criteria there are advantages to the Davis-Resnick estimate over our estimate. If one considered a loss function of the form: lose one unit if the relative error, $|\hat{\phi}-\phi|/\phi$, exceeds a threshold m and nothing otherwise, then our estimate did not perform as well under this loss function as the Davis-Resnick estimate when the threshold was high. For example, for a sample size of 100 and a threshold $m = 1$, i.e. a relative error at least 100%, the risk expressed as a percentage, viz. $P(|\hat{\phi} - \phi|/\phi > 1)\,100\%$, for the Davis-Resnick estimate at $\phi = 0.2$ and for choices of β equal to 0.2, 0.8, and 1.8 were 0.6, 11, and 1.5, respectively. Whereas for our estimates, these risks were 0.4, 2.4, and 16. When the threshold increased to $m = 2$, the risks for the Davis-Resnick estimates under these parameter combinations were 0.0, 0.5, and 0.9, respectively. While for our method, they were 0.1, 0.9, and 3.3, respectively.

Reflection on the form of our estimator in the AR(1) case suggests that when the innovation following the largest sample value is also large that tends to deteriorate the accuracy of our estimate and this effect is amplified the smaller the autocorrelation parameter. These observations are reinforced by consideration of the limit distribution of our estimate in the AR(1) case. The tail limiting distribution is given by $P(Z/\eta > x)$ where Z and η are independent random variables having a Pareto(β) and Frechet distribution with parameter β, respectively. A direct calculation shows the tail area to be asymptotically equivalent to $x^{-\beta}$ as x tends to infinity so that one finds $P(|\hat{\phi}-\phi|/\phi > m) \sim (mb_n\phi)^{-\beta}$ where $b_n = F^{\leftarrow}(1-1/n)$. In the case of a Pareto(β) innovation distribution, this becomes $(nm^{\beta}\phi^{\beta})^{-1}$ showing that the chance of a relative error at least m increases with decreasing ϕ for fixed β and for fixed ϕ increases with increasing β. These observations appear to be substantiated in simulations and may be helpful in refining our estimate.

Finally, Table 1 reveals that the average lengths of 99% confidence intervals are considerably wider than for 95% confidence intervals and this observation applies to both estimates. The reason for this behavior is the heavy tail feature of the sampling distribution of the estimates. For the Davis-Resnick estimate, the limit distribution for their estimate is of the form S_1/S_0 where S_0 and S_1 are independent stable random variables with indices $\beta/2$ and β, respectively. For our estimate, as noted above the limit distribution is of the form Z/η where Z and η are independent random variables having a Pareto(β) and Frechet distribution with parameter β. In both cases, these limit distributions are heavy tail distributions.

Now observe from Table 3.1 that our estimator on average is significantly more accurate in estimating the true autoregressive parameter $\phi = .2$ than Davis and Resnick's (1986) estimator when $0 < \beta < 1$. While the accuracy of our estimator remains consistent as the regular varying index approaches 1, Davis and Resnick's estimator performs drastically worse. An obvious distinction in performances of both estimators can be seen in the average length of the confidence intervals. While our method was able to use one-sided confidence intervals since $\hat{\phi}_{BM} \geq \phi, \quad \text{for all } t \geq 1$, Davis and Resnick's methods couldn't. Within every comparison (expect $\beta = 1.8$) our method of estimation out performs Davis and Resnick's estimator in precision. In particular, when $\beta = .2, n = 500$ the 95% confidence interval average length for our method is 58 times smaller than our competitors. Furthermore, the average length of Davis and Resnick's estimator for the 95% confidence interval is approximately .3 (one-third of the maximum coverage for ϕ) when the sample size is small i.e. ($n = 100$), whereas ours is approximately 3 times smaller. Due to the fact that any extreme value method of estimation becomes extremely difficult for a sample size i.e. ($n = 100$), this result is quite astonishing.

While the average length of our 95% confidence intervals remain significantly smaller, the average length of our 99% confidence intervals becomes increasingly larger as β increases, whereas Davis and Resnick estimator average length remains consistently poor. A possible explanation for the significant difference in length between the 95% and 99% confidence interval for our method lies within the occurrence of a few extreme estimates. That is, our extreme value method of estimation depends heavily on obtaining large innovations. Thus, it can be shown when the regular varying index is small (less than 1) the largest innovation Z_t will be extremely large, and only in the situation when the next innovation is also large does our estimator behave badly. Whereas if β takes on higher values like 1.5 or 1.8, then the largest innovation is not likely to be nearly as large, thus the chance that we get a bad estimate increases since now it only takes a moderately above normal innovation to produce an extreme estimate. Furthermore, the effect of the autoregressive parameter ϕ is that if it takes on a value near its lower bound, then the largest innovation is considerably reduced in value, thus allowing a large spectrum of moderate to normal values for the next innovation to cause bad estimates. Therefore, we can expect with small probability some extreme estimates from our estimator, which in turn affects the length for a 99% confidence interval but not the length for a 95% confidence interval. Additionally observe with $Z \sim F_\beta$ and $\eta \sim \Phi_\beta$ we have from Corollary 3.4.1 that $b_n(\hat{\phi}_{BM} - \phi) \xrightarrow{D} \frac{Z}{\eta}$, where Z and η are independent and $b_n = n^{1/\beta}$. For $y > 0$

$$P\left(\frac{Z}{\eta} > y\right) = \Phi_\beta(1/y) + \frac{1}{y^\beta}\int_0^{y^\beta} xe^{-x}dx.$$

Hence

$$P\left(\frac{Z}{\eta} > y\right) \sim \frac{1}{y^\beta}, \text{ as } y \to \infty.$$

Thus, we find for some positive constant $m > 0$ that

$$P((\hat{\phi}_{BM} - \phi) > m\phi) = P(n^{1/\beta}(\hat{\phi}_{BM} - \phi) > m\phi n^{1/\beta}) \sim P\left(\frac{Z}{\eta} > m\phi n^{1/\beta}\right) \sim \frac{1}{(m\phi)^\beta n}.$$

In the special case that $n = 500, m = 2, \phi = .2$, and $\beta = .8$ we find

$$P\left(\hat{\phi}_{BM} > .6\right) \approx \frac{1}{(.4)^{(.8)}500} = .0041.$$

Although the small probability of an extreme estimate appears alarming, in reality, they are merely a part of our extreme value method and despite this, Table 3.1 verifies when $0 < \beta < 1$ that our estimator produces a smaller standard deviation than Davis and Resnick's estimator. Even with the few extreme estimates, the smaller standard deviation just verifies the amount of precision and accuracy of our estimator. For example, when $\beta = .5, n = 500$

only 35 estimates were more than .02 away from $\phi = .2$ and of those only 5 were more than 2 times as large as ϕ.

With regard to bias, it is easily checked that the mean of the limiting distribution of our estimator $\hat{\phi}_{BM}$ is $\frac{\beta}{\beta-1}\Gamma(\frac{1}{\beta}+1) = \frac{1}{\beta-1}\Gamma(\frac{1}{\beta})$, where Γ denotes the gamma function. Noting that $b_n = n^{1/\beta} = F^{\leftarrow}(1-\frac{1}{n})$, we see that the estimate

$$\tilde{\phi}_{BM} = \hat{\phi}_{BM} - \frac{1}{\beta-1}\Gamma\left(\frac{1}{\beta}\right) n^{-1/\beta} \tag{3.5.3}$$

is asymptotically unbiased when $\beta > 1$.

Table 3.1: Comparison of Estimators for the AR(1) model when $\phi = .2$.

β	n	$\hat{\phi}_{DR}$	$\hat{\phi}_{BM}$	avgLength$_{DR}$			avgLength$_{BM}$		
				90%	95%	99%	90%	95%	99%
$\beta = .2$	$n = 100$	.2160	.2022	.0154	.0544	.2473	.0001	.0008	.0760
		(.0291)	(.0257)	-	-	-	-	-	-
	$n = 500$	.2010	.2003	.0094	.0176	.0715	$<$ .0001	.0003	.0021
		(.0131)	(.0081)	-	-	-	-	-	-
$\beta = .5$	$n = 100$	.2243	.2089	.1185	.1765	.3233	.0097	.0302	.2174
		(.0565)	(.0491)	-	-	-	-	-	-
	$n = 500$	.2049	.2014	.0204	.0408	.1927	.0004	.0014	.0250
		(.0201)	(.0134)	-	-	-	-	-	-
$\beta = .8$	$n = 100$	.2804	.2237	.2751	.3241	.3845	.0564	.0982	.3345
		(.0874)	(.0668)	-	-	-	-	-	-
	$n = 500$	.2263	.2054	.0836	.1149	.2109	.0077	.0179	.1036
		(.0346)	(.0306)	-	-	-	-	-	-
$\beta = 1.2$	$n = 100$	.1862	.2205	.2242	.3028	.4401	.1471	.2188	.4188
		(.0705)	(.0721)	-	-	-	-	-	-
	$n = 500$	.1978	.2158	.0866	.1214	.2432	.0392	.0671	.1932
		(.0328)	(.0343)	-	-	-	-	-	-
$\beta = 1.5$	$n = 100$	.1858	.2210	.2424	.3082	.4464	.1957	.2617	.3702
		(.0735)	(.1002)	-	-	-	-	-	-
	$n = 500$	.1974	.2157	.1058	.1373	.2409	.0729	.1175	.2147
		(.0355)	(.0426)	-	-	-	-	-	-
$\beta = 1.8$	$n = 100$	.1842	.2445	.2535	.3158	.4464	.2388	.3005	.4138
		(.0755)	(.0969)	-	-	-	-	-	-
	$n = 500$	.1976	.2274	.1169	.1469	.2222	.1108	.1623	.2307
		(.0381)	(.0582)	-	-	-	-	-	-

Figures 3.1 and 3.2 below show a comparison between the probability that estimators $\hat{\phi}_{BM}$ and $\hat{\phi}_{DR}$ are within .01 of the true autocorrelation parameter value, respectively. The figures show how the performance of our extreme value method and Davis and Resnick's Yule-Walker method varies with respect to size of the autocorrelation parameter and index of regular variation. With a sample size of 500 these figures plotted the sample fraction of estimates which fell within a bound of $\epsilon = .01$ of the true value. Good performance with respect to this measure is reflected in curves near to 1.0 with diminishing good behavior as curves approach 0.0. The figures show that as the autocorrelation parameter ϕ increases,

for each fixed value of β the accuracy of our method increases while the accuracy of the Davis and Resnick estimator decreases. When $0 < \beta < 1$, our estimator compared to theirs produced a higher fraction of precise estimates as measured by being within the bound ϵ of the true autoregressive parameter ϕ. When the regular variation index value is nearer to 2, we see a higher fraction of the Davis-Resnick estimates showing better accuracy by this measure. The figures also indicate another interesting difference between the two methods. Whereas the accuracy of our method increases with decreasing value of β, the Davis-Resnick estimate shows no such monotonicity with the most challenging values for them occurring for β near 1.

Figure 3.1: $P[|\hat{\phi}_{BM} - \phi| < .01]$ for $RV_{-\beta}$

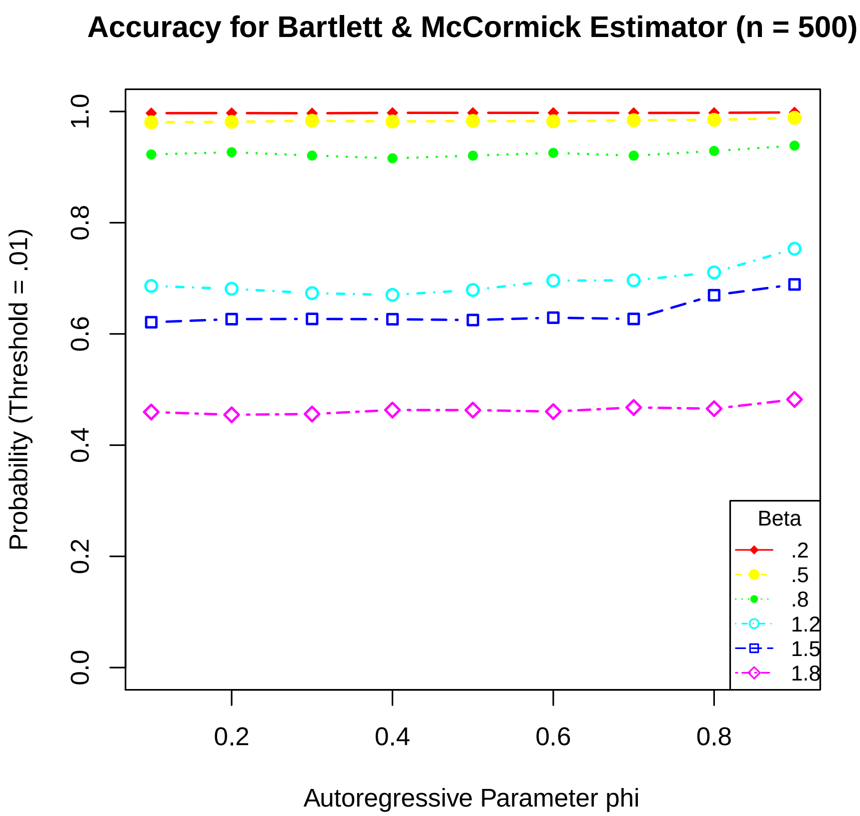
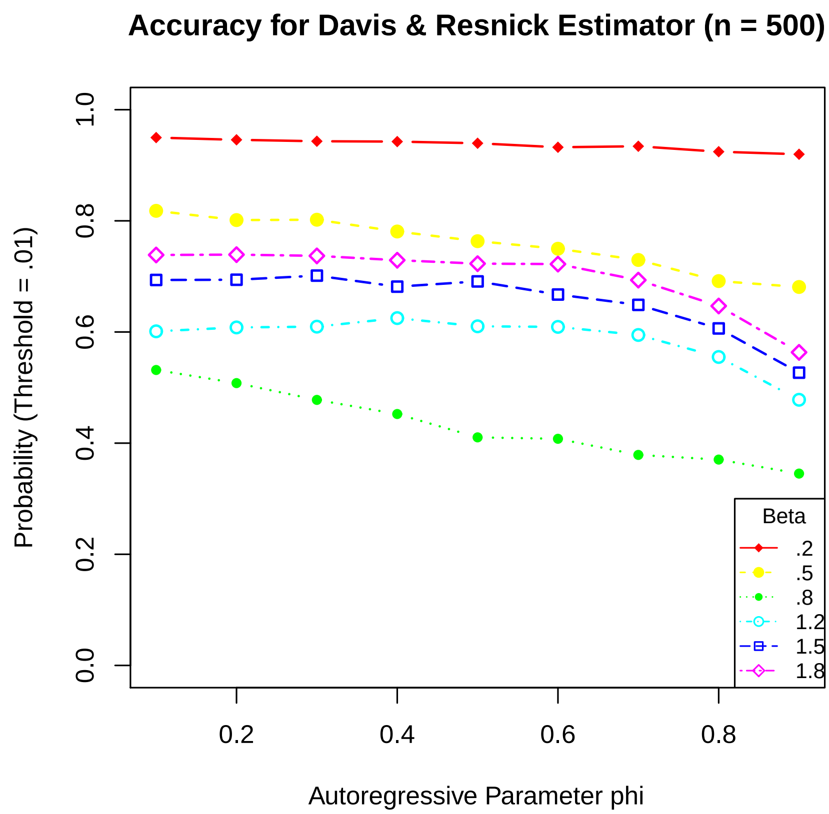


Figure 3.2: $P[|\hat{\phi}_{DR} - \phi| < .01]$ for $RV_{-\beta}$



3.5.2 AR(2) Simulation Study

The simulation study performed for the AR(2) model consisted of two parameter regimes. For the simulation study presented in Tables 3.2 and 3.3, the AR(2) model had an autoregressive polynomial with root of smaller norm close to 1 whereas for the simulation study presented in Tables 3.4 and 3.5, the smaller root was fairly distant from 1. In particular, the smaller root in norm for the example in Tables 3.2 and 3.3 was 1.045, whereas the smaller root for the model reported in Tables 3.4 and 3.5 was 2.0. In both examples, our simulation followed the same guidelines as AR(1) simulation. That is, we generated 5,000 replications of the nonnegative time series $(X_0, X_1, \ldots, X_n)$ for sample sizes of 100 and 500. In its infinite order moving average MA(∞) representation, the AR(2) process (X_t) takes the form

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \qquad \text{where } c_0 = 1, c_1 = \phi_1, c_i = \phi_1 c_{i-1} + \phi_2 c_{i-2}, \text{ for } i = 2, 3, \ldots.$$

The Pareto distribution defined in (3.5.2) was used for the nonnegative innovation distribution F.

A monte carlo simulation was performed in each of the four cases where we obtained $5,000$ estimates for (ϕ_1, ϕ_2), using $(\hat{\phi}_{1(DR)}, \hat{\phi}_{2(DR)}, \hat{\phi}_{1(BM)}, \hat{\phi}_{2(BM)})$. The means and standard deviations (written below in parentheses), of these estimates are reported in Tables 3.2 and 3.3 for the first example and in Tables 3.4 and 3.5 for the second example, along with the average length for a 95 and 99 percent empirical confidence intervals with exact coverage. More precisely, the endpoints of the 95% confidence interval, which depend on the .025 and .975 percentiles of the sampling distribution of the pivot, were determined empirically through simulation rather than obtained from a limit distribution. Thus, the probability that the confidence interval contains the true parameter was nearly perfect throughout all four cases.

For the situation when there is a root of norm near one and $0 < \beta < 1$, Tables 3.2 and 3.3 demonstrate a better performance both in terms of smaller bias and smaller length of confidence intervals of our estimate over the Davis-Resnick estimate. In the range of $1 < \beta < 2$ the Davis-Resnick estimate is less biased but has greater variability than our estimate.

For the situation when the roots of the autoregressive polynomial are farther in norm from one and $0 < \beta < 1$, Tables 3.4 and 3.5 indicate a similar comparison between the estimates as shown in Tables 3.2 and 3.3. In the range $1 < \beta < 2$, there is an improvement in bias for the Davis-Resnick estimate over ours and an improvement in both bias and length of confidence interval for $\beta = 1.8$, the highest value for β considered in our study.

Table 3.2: Estimator Comparison of ϕ_1 when a root has norm near one

		$\phi_1 = .80$		avgLength$_{DR}$		avgLength$_{BM}$	
β	n	$\hat{\phi}_{1DR}$	$\hat{\phi}_{1BM}$	95%	99%	95%	99%
$\beta = .2$	$n = 100$	.8453	.8262	.3902	.4603	.0567	.2591
		(.2104)	(.0752)	-	-	-	-
	$n = 500$	.8263	.8003	.1673	.2873	.0167	.0968
		(.1056)	(.0302)	-	-	-	-
$\beta = .8$	$n = 100$	.8776	.8162	.5081	.6323	.0835	.1829
		(.1622)	(.0565)	-	-	-	-
	$n = 500$	.8458	.8084	.3304	.3914	.0285	.1397
		(.3097)	(.0397)	-	-	-	-
$\beta = 1.2$	$n = 100$	.8047	.8357	.3200	.4404	.1136	.1913
		(.0660)	(.0481)	-	-	-	-
	$n = 500$	.7995	.8179	.2303	.3485	.0988	.1667
		(.1801)	(.0424)	-	-	-	-
$\beta = 1.8$	$n = 100$	.7836	.8529	.2253	.2798	.1659	.2262
		(.0818)	(.0333)	-	-	-	-
	$n = 500$	.7988	.8277	.1502	.2195	.1481	.1996
		(.2963)	.0388)	-	-	-	-

Table 3.3: Estimator Comparison of ϕ_2 when a root has norm near one

		$\phi_2 = .15$		avgLength$_{DR}$		avgLength$_{BM}$	
β	n	$\hat{\phi}_{2DR}$	$\hat{\phi}_{2BM}$	95%	99%	95%	99%
$\beta = .2$	$n = 100$	.1276	.1386	.2428	.3953	.0018	.0284
		(.2518)	(.0724)	-	-	-	-
	$n = 500$	.1388	.1489	.2155	.3302	$< .0001$	.0201
		(.1404)	(.0577)	-	-	-	-
$\beta = .8$	$n = 100$	.1071	.1427	.4161	.5224	.0179	.0445
		(.3045)	(.0556)	-	-	-	-
	$n = 500$	.1290	.1492	.3110	.4478	$< .0001$	.0228
		(.2262)	(.0391)	-	-	-	-
$\beta = 1.2$	$n = 100$	.1135	.1287	.3108	.5024	.0375	.0795
		(.1804)	(.0474)	-	-	-	-
	$n = 500$	.1479	.1408	.2076	.3358	.0187	.0309
		(.0648)	(.0387)	-	-	-	-
$\beta = 1.8$	$n = 100$	.1264	.1142	.2372	.4162	.0939	.1587
		(.0747)	(.0347)	-	-	-	-
	$n = 500$	.1438	.1323	.0603	.1339	.0478	.0759
		(.0411)	(.0379)	-	-	-	-

Table 3.4: Estimator Comparison of ϕ_1 when roots have norm far from one

		$\phi_1 = .20$		avgLength$_{DR}$		avgLength$_{BM}$	
β	n	$\hat{\phi}_{1DR}$	$\hat{\phi}_{1BM}$	95%	99%	95%	99%
$\beta = .2$	$n = 100$	.2453	.2220	.2554	.7056	.0817	.2791
		(.2571)	(.0924)	-	-	-	-
	$n = 500$	.2199	.2014	.0186	.3565	.0022	.0782
		(.1437)	(.0412)	-	-	-	-
$\beta = .8$	$n = 100$	.2597	.2287	.3781	.6724	.0994	.2998
		(.1930)	(.0826)	-	-	-	-
	$n = 500$	.2239	.2039	.1076	.2941	.0590	.1756
		(.1255)	(.0391)	-	-	-	-
$\beta = 1.2$	$n = 100$	.1955	.2227	.1649	.3411	.1364	.3429
		(.1927)	(.0923)	-	-	-	-
	$n = 500$	.1973	.2105	.1350	.2641	.1246	.2467
		(.1039)	(.0496)	-	-	-	-
$\beta = 1.8$	$n = 100$	.1855	.2403	.1197	.4049	.1744	.4250
		(.2243)	(.1152)	-	-	-	-
	$n = 500$	.1989	.2298	.0837	.2350	.1399	.2962
		(.0574)	(.0644)	-	-	-	-

Table 3.5: Estimator Comparison of ϕ_2 when roots have norm far from one

		$\phi_2 = .15$		avgLength$_{DR}$		avgLength$_{BM}$	
β	n	$\hat{\phi}_{2DR}$	$\hat{\phi}_{2BM}$	95%	99%	95%	99%
$\beta = .2$	$n = 100$	.1682	.1377	.2102	.6957	.0737	.3177
		(.2145)	(.0717)	-	-	-	-
	$n = 500$	.1535	.1481	.0339	.2696	.0004	.0621
		(.1247)	(.0303)	-	-	-	-
$\beta = .8$	$n = 100$	.2264	.1585	.2651	.7957	.0981	.3636
		(.2787)	(.0871)	-	-	-	-
	$n = 500$	.1885	.1519	.1223	.2904	.0453	.1915
		(.1106)	(.0392)	-	-	-	-
$\beta = 1.2$	$n = 100$	.1348	.1678	.1827	.4921	.1677	.3927
		(.1783)	(.0974)	-	-	-	-
	$n = 500$	.1444	.1585	.1504	.2337	.1232	.2422
		(.1135)	(.0522)	-	-	-	-
$\beta = 1.8$	$n = 100$	.1358	.1754	.1515	.3983	.1911	.4579
		(.2366)	(.1189)	-	-	-	-
	$n = 500$	.1456	.1637	.1105	.2057	.1336	.2678
		(.0413)	(.0707)	-	-	-	-

As with the AR(1) case, accounting for the bias in our estimates, it is easily checked that the mean of the limiting distribution of our estimator $\hat{\phi}_{1(BM)}$ is $\frac{1}{\beta-1}\Gamma(\frac{1}{\beta})\frac{1-\phi_1}{1-\phi_1-\phi_2}$ when $\beta > 1$. Thus, the estimate

$$\tilde{\phi}_{1(BM)} = \hat{\phi}_{1(BM)} - \frac{1}{\beta-1}\Gamma\left(\frac{1}{\beta}\right)\frac{1-\hat{\phi}_{1(BM)}}{1-\hat{\phi}_{1(BM)}-\hat{\phi}_{2(BM)}}n^{-1/\beta}$$

is asymptotically unbiased estimate for ϕ_1. Similarly

$$\tilde{\phi}_{2(BM)} = \hat{\phi}_{2(BM)} - \frac{1}{\beta-1}\Gamma\left(\frac{1}{\beta}\right)\left(\frac{(1-\hat{\phi}_{1(BM)})^2-\hat{\phi}_{2(BM)}}{1-\hat{\phi}_{1(BM)}-\hat{\phi}_{2(BM)}}\right)n^{-1/\beta}$$

is asymptotically unbiased estimate for ϕ_2.

We conclude this subsection with a comment on the shape of the confidence intervals for either of these estimation procedures. The point estimate falls towards the left end point of the intervals so that they are asymmetrical about the estimate and may be said to be right skewed. Unlike your typically confidence intervals where the parameter estimate is located in the center of the interval,the nature of the confidence intervals obtained from both methods in this simulation were non-symmetric. The non-symmetric confidence intervals arises from the respected empirical distributions. That is, the lower critical values found at the .025 or .005 percentiles are typically much smaller than the .975 or .995 percentiles respectively. For example, denoting c_u as the .975 percentile and c_l as the .025 percentile of the empirical distribution $b_n(\hat{\phi}_{1(BM)} - \phi_1)$ we have

$$P[c_l \leq b_n(\hat{\phi}_{1(BM)} - \phi_1) \leq c_u] = .95.$$

Consequently,

$$P[\hat{\phi}_{1(BM)} - c_u/b_n \leq \phi_1 \leq \hat{\phi}_{1(BM)} - c_l/b_n] = .95$$

and as a result in general, we get non-symmetrical 95% confidence intervals since c_u/b_n and c_l/b_n are not equal in length.

3.5.3 ARMA(3,1) Simulation Study

In this section we consider simulating the ARMA(3,1) process (X_t) satisfying the equations

$$X_t = .1044X_{t-1} + .0559X_{t-2} + .0068X_{t-3} + Z_t - .1Z_{t-1}, \quad \text{for } 1 \leq t \leq n$$

with ARMA polynomials given by

$$\phi(z) = 1 - .1044z - .0559z^2 - .0068z^3 \quad \text{and} \quad \theta(z) = 1 - .1z.$$

As with the AR(1) and AR(2) simulations, we considered two different sample sizes $(100, 500)$ and take the nonnegative innovation distribution to be the Pareto distribution defined in (3.5.2). Since the autoregressive polynomial $\phi(z) = 1 - .1044z - .0559z^2 - .0068z^3$ has zeros at 3.00016 and $-5.61037 \pm 4.18816\,i$, which are located outside the unit circle, we have a causal ARMA process. Lastly, the MA polynomial $\theta(z) = 1 - .1z$ has a zero at $z = 10$, which is also located outside the unit circle. This implies that (X_t) is invertible. What makes this example so interesting is that we are using a nonnegative time series with nonnegative innovations and the casual representation

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

where $c_0 = 1, c_1 = \phi_1 + \theta_1, c_2 = \phi_1 c_1 + \phi_2, c_i = \phi_1 c_{i-1} + \phi_2 c_{i-2} + \phi_3 c_{i-3},$ for $i = 3, 4, \ldots$ are such that all coefficients c_j are nonnegative in order to estimate the three positive autoregressive parameters and the negative moving average coefficient θ_1. For this model $l_0 = 1$, and so we estimate the coefficients $c_j, j \geq 1$ by

$$\hat{c}_i = \frac{X_{t^*-1+i}}{X_{t^*-1}}, \quad i \geq 1.$$

Then solving (3.4.2) we find

$$\hat{\phi}_1 = \frac{\hat{c}_4 - \hat{c}_2^2 - \hat{c}_1\hat{c}_3 + \hat{c}_1^2\hat{c}_2}{\hat{c}_3 - 2\hat{c}_1\hat{c}_2 + \hat{c}_1^3}, \hat{\phi}_2 = \hat{c}_2 - \hat{\phi}_1\hat{c}_1, \hat{\phi}_3 = \hat{c}_3 - \hat{\phi}_1\hat{c}_2 - \hat{\phi}_2\hat{c}_1, \hat{\theta}_1 = \hat{c}_1 - \hat{\phi}_1 \tag{3.5.4}$$

to be estimators for $(\phi_1, \phi_2, \phi_3, \theta_1)$, respectively.

A monte carlo method was performed for $\beta = (.2, .8, 1.2, 1.8)$ in which we obtained $5,000$ estimates for $(\phi_1, \phi_2, \phi_3, \theta_1)$, using $(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\theta}_1)$ in each case. The means and standard deviations (written below in parentheses) of these estimates are reported in Table's 3.6 and 3.7 along with the average length for a 95 and 99 percent empirical confidence intervals with exact coverage.

As seen from (3.5.4), the estimate for ϕ_1 is a complicated function of four estimates of coefficients from the moving average representation of the process and further the other parameter estimates involve this estimate in their definitions, suggestive therefore that variability of the estimates may be high for this problem. This is seen to have occurred in the estimation of ϕ_1 and in the estimation of the moving average parameter θ_1.

Observe from (3.5.4) that ϕ_1 is the most complicated parameter to estimate and the other estimates depend directly upon $\hat{\phi}_1$ with the negative moving average estimate $\hat{\theta}_1$ being affected the most. Hence, it should come to little surprise that θ_1 was the hardest parameter to estimate. However, Tables 3.6 and 3.7 clearly shows based on accuracy, variation, and precision that our method of estimation in the ARMA(3,1) model performed extremely well when the regular variation index β is smaller than 1 despite the heavy reliance among estimates. Not only was the accuracy for the three positive autoregressive parameter estimates within .01 of their true value when $n = 500$ and $0 < \beta < 1$, but remarkability the estimate for the negative moving average parameter was also extremely accurate (within .01) of θ_1. Furthermore, seven out of the eight average lengths of the 95% confidence intervals were smaller than .02 for all four parameters, while all six intervals for the three positive autoregressive parameters were smaller than .05 when $n = 100$. Within the AR(1) and AR(2) models the few extreme estimates have had the most impact on the 99% confidence intervals, however this was not the case for the ARMA(3,1) model as seven out of eight average lengths were smaller than .10 and of those seven, five were smaller than .05. Therefore our method of estimation not only got validation for ARMA models, but the ability with extreme precision to estimate negative parameters with a nonnegative time series with nonnegative innovations.

A further analysis of the variability of the estimates is provided in Figure 3.3. From these histograms we see that the sampling distributions for the estimates of ϕ_1 and θ_1 have long tails whereas the sampling distribution of the estimate of ϕ_2 is particularly concentrated at the true parameter value. More generally for all four estimates, the histogram spikes steeply at the true parameter and most of the mass of the sampling distribution is concentrated near the true parameter value.

The amount of variation seen in Table 3.6 for $\hat{\phi}_1$ appears to be a concern, but further analysis reveals that this variation is due to a few bad estimates which based on the discussion in section 4.2 is to be expected. Figure 3.3 confirms two statements about the empirical distribution of our estimators: Firstly a large spike within .001 of the true value and small spikes farther from the true value which skews the mean and standard deviation and secondly, the few bad estimates for ϕ_1 directly produce bad estimates for ϕ_2, ϕ_3, and θ_1, which is explained by the heavy dependence that the estimators $(\hat{\phi}_2, \hat{\phi}_3, \hat{\theta}_1)$ have with $\hat{\phi}_1$.

Table 3.6: ARMA(3,1) Confidence Intervals for (ϕ_1, ϕ_2)

.

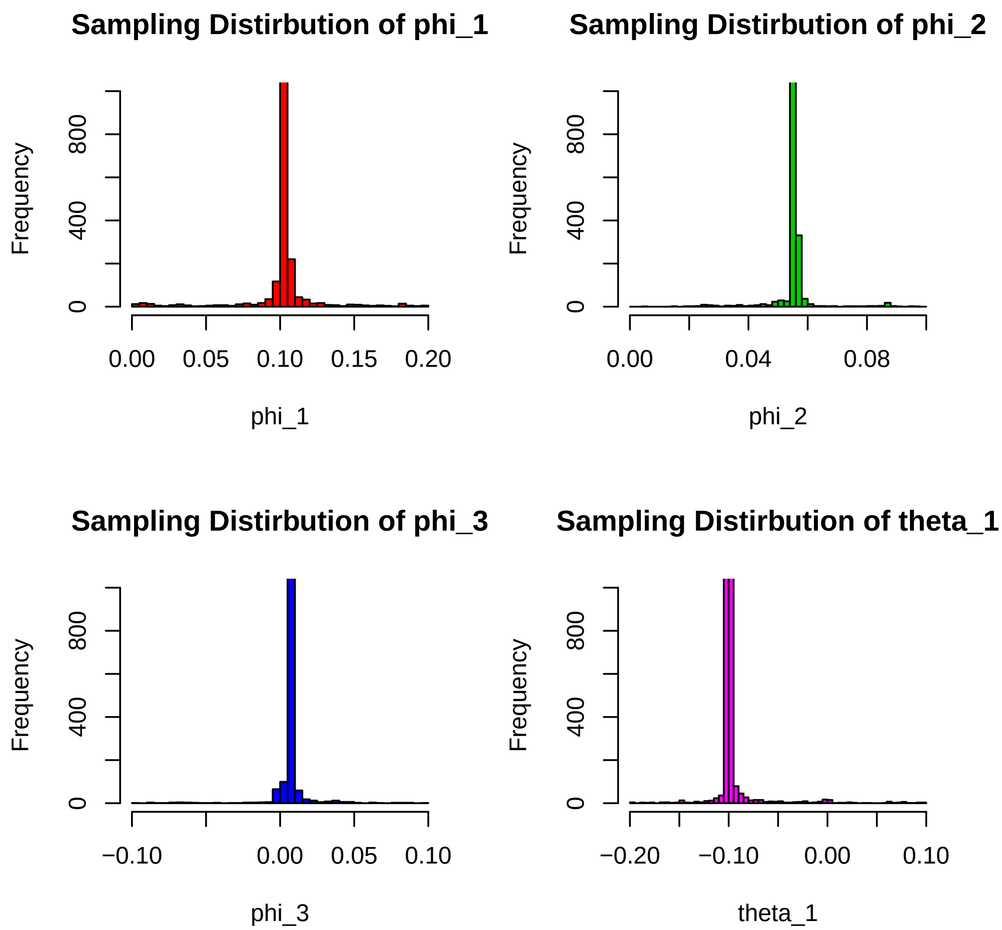
		$\phi_1 = .1044$	avgLength		$\phi_2 = .0559$	avgLength	
β	n	$\hat{\phi}_1$	95%	99%	$\hat{\phi}_2$	95%	99%
$\beta = .2$	$n = 100$	.1064	.0143	.1144	.0517	.0224	.1491
		(.2965)	-	-	(.0374)	-	-
	$n = 500$	.1045	.0019	.0524	.0559	.0001	.0131
		(.1706)	-	-	(.0079)	-	-
$\beta = .8$	$n = 100$	.1097	.0450	.1547	.0649	.0433	.1567
		(.4173)	-	-	(.0139)	-	-
	$n = 500$	.1048	.0164	.0768	.0556	.0167	.0338
		(.2441)	-	-	(.0149)	-	-
$\beta = 1.2$	$n = 100$	.1175	.1967	.3772	.0381	.0938	.2186
		(.5229)	-	-	(.0122)	-	-
	$n = 500$	.1104	.0673	.1512	.0536	.0287	.1028
		(.3748)	-	-	(.0225)	-	-
$\beta = 1.8$	$n = 100$	.1447	.2821	.4315	.0272	.1173	.2812
		(.6321)	-	-	(.0316)	-	-
	$n = 500$	.1205	.1184	.2478	.0529	.0501	.1438
		(.5148)	-	-	(.0497)	-	-

Table 3.7: ARMA(3,1) Confidence Intervals for (ϕ_3, θ_1)

.

		$\phi_3 = .0068$	avgLength		$\theta_1 = -0.1$	avgLength	
β	n	$\hat{\phi}_3$	95%	99%	$\hat{\theta}_1$	95%	99%
$\beta = .2$	$n = 100$	.0114	.0212	.1183	-.1277	.0776	.2247
		(.0217)	-	-	(.2764)	-	-
	$n = 500$	.0068	$< .0001$	.0282	-.1001	.0040	.0439
		(.0161)	-	-	(.1832)	-	-
$\beta = .8$	$n = 100$	.0131	.0112	.1486	-.1303	.1159	.2790
		(.0521)	-	-	(.4673)	-	-
	$n = 500$	.0086	.0027	.0386	-.1097	.0264	.1170
		(.0305)	-	-	(.2542)	-	-
$\beta = 1.2$	$n = 100$	.0217	.0173	.1592	-.1579	.1973	.3772
		(.1237)	-	-	(.5931)	-	-
	$n = 500$	.0101	.0151	.0435	-.1103	.0607	.1812
		(.0603)	-	-	(.3920)	-	-
$\beta = 1.8$	$n = 100$	.0469	.0221	.1821	-.1934	.2181	.4308
		(.1667)	-	-	(.6821)	-	-
	$n = 500$	.0326	.0319	.0944	-.1121	.1214	.2179
		(.1133)	-	-	(.4908)	-	-

Figure 3.3: Histograms for ARMA(3,1) estimates when $\beta = .2$



Our estimation procedure depends on knowing the index of the maximal coefficient in the moving average representation. We conclude the simulation study with an analysis of an ad hoc method for estimating this index. If the index of the maximal coefficient were l_0, then a rough approximation to the process above a high threshold is given by $(X_t, X_{t+1}, \ldots, X_{t+l_0}) \approx (Z_t, c_1 Z_t, \ldots, c_{l_0} Z_t)$ where Z_t represents a large innovation. This observation suggests a method for estimating l_0. Observe the start of a cluster of large values and note the position of the largest value within the cluster. Estimate l_0 to be that position minus one. We tried this procedure where a cluster was deemed large if all its values were as large or exceeded the 90th percentile of the sample values. We applied this procedure to two ARMA(1,1) models where for both models the largest coefficient in the moving average representation was given by c_1, so that our estimate of the index would be correct if the largest observation in a high values cluster occurred with the second observation. Table 3.8 reports the probability of correct selection of the index using this method. The procedure was applied to ARMA(1,1) models with a series length of 500 and 100 replications were performed. The results were poor and so we would recommend using standard statistical estimation procedures to obtain initial estimates of the parameter values and based on these estimates choose the value for l_0 needed for our procedure.

Table 3.8: Index Estimate for the Moving Average Largest Coefficient

.

β	.2	.8	1.2	1.8
$\phi = .8$, $\theta = .5$	.41	.46	.53	.56
$\phi = .3$, $\theta = .8$	.54	.76	.79	.86

3.6 References

[1] Anděl, J. (1989). Non-negative autoregressive processes. *J. Time Ser. Anal.* **10**, 1-11.

[2] Anděl, J. (1991). Non-negative linear processes. *Appl. of Math.* **36**, 277-283.

[3] Anderson, P.L., Kavalieris, L., Meerschaert, M.M. (2008). Innovations algorithm asymptotics for periodically stationary time series with heavy tails. *J. Multivariate Anal.* **99**, 94-116.

[4] Bartlett A., McCormick W.P. (2013). Estimation for Nonnegative Time Series with Heavy-tail Innovations, *J. Time Series Anal*, **34**, 96 - 115.

[5] Brockwell, P.J., Davis, R.A. (1987). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

[6] Brown, T.C., Feigin, P.D., Pallant, D.L. (1996). Estimation for a class of positive non-linear time series models. *Stochastic Process. Appl.* **63**, 139-152.

[7] Calder, M., Davis, R.A. (1998). In R.J. Adler, R.E. Feldman, M.S. Taqqu, editors, *A Practical Guide To Heavy Tails: Statistical Techniques and Applications* pages 159-176, Birkhäuser, Boston.

[8] Datta, S., McCormick, W.P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *J. Amer. Statist. Assoc.* **90**, 1289-1300.

[9] Datta, S., Mathew, G., McCormick, W.P. (1998). Nonlinear autoregression with positive innovations. *Austral. & New Zealand J. Statist.* **40**, 229-239.

[10] Davis, R.A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Process. Appl.* **63**, 75-95.

[11] Davis, R.A., McCormick, W.P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Process. Appl.* **31**, 237-250.

[12] Davis, R.A., Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann Probab.* **13**, 179-195.

[13] Davis, R.A., Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.

[14] Feigin, P.D., Resnick, S.I. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51**, 135-165.

[15] Feigin, P.D., Kratz, M.F., Resnick, S.I. (1996). Parameter estimation for moving averages with positive innovations. *Ann. Appl. Probab.* **6**, 1157-1190.

[16] Kallenberg, O. (1976). *Random Measures*, Akademie-Verlag, Berlin.

[17] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

[18] Markov, A.S. (2009). Estimation of the autoregressive parameter with infinite dispersion of noise. *Automation and Remote Control* **70**, 92-106.

[19] Mikosch, T., Samorodnitsky, G. (2000). The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Applied Probab.* **10**, 1025-1064.

[20] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[21] Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25**, 1805-1869.

[22] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

[23] Tsai, H., Chan, K.S. (2006). A note on non-negative ARMA processes. *J. Time Ser. Anal.* **28**, 350-360.

Chapter 4

Estimation for First-Order Bifurcating Autoregressive Processes with an Unknown Location Parameter

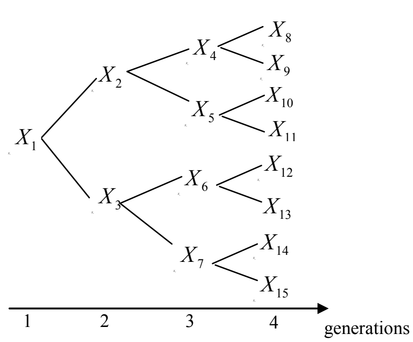
4.1 Introduction

The bifurcating autoregressive process (BAR, for short) was introduced by Cowan and Staudte (1986) for analyzing cell lineage data, where each individual in one generation gives rise to two offspring in the next generation. More precisely, the first- order bifurcating autoregressive process, BAR(1), is defined by the equation

$$X_t = \phi X_{\lfloor t/2 \rfloor} + \epsilon_t, \qquad \text{for } 2 \le t \le 2^{(k+1)} - 1, \tag{4.1.1}$$

where k represents the number of generations, $0 < \phi < 1$, and $\lfloor \cdot \rfloor$ denotes the greatest integer function, so that one can write recursively $X_2 = \phi X_1 + \epsilon_2, X_3 = \phi X_1 + \epsilon_3, X_4 = \phi X_2 + \epsilon_4$, etc. For illustration, the data structure with four generations is given below.

Figure 4.1: Illustration with four Generations in a Binary Tree



Over the past two decade's there have been many extensions of bifurcating autoregressive models. For example, the driven noise $(\epsilon_{2t}, \epsilon_{2t+1})$ was originally assumed to be independent and identically distributed normal random variables. However, two sister cells being in the same environment early in their lives are allowed to be correlated, inducing a correlation between sister cells distinct from the correlation inherited from their mother. Cowan and Staudte (1986) proposed the first BAR(1) model which views each line of descent as a first-order autoregressive AR(1) process with the added complication that the observations on the two sister cells who share the same parent are allowed to be correlated. In this section we will focus mainly upon the correlation between mother and daughter cells, however, we do verify through simulation that estimates for ϕ are not affected when the two sister cells are correlated.

In Zhang (2011), the author applies an point process technique to a first-order bifurcating process. Within this paper the author defined $Y_t = \min(\epsilon_{2t}, \epsilon_{2t+1})$ and assumed that the marginal distribution of this random variable was regularly varying at the left endpoint 0 with index α. Additionally, the author derived the joint limiting distribution of the estimator of ϕ and the tail index α under some regularity conditions. Bartlett and McCormick (2012) obtained the limit distribution of $\hat{\phi}_n^* = \min_{1\leq t\leq n} \frac{\tilde{X}_t}{\tilde{X}_{t-1}}$ for an AR(1) model defined in (4.2.4) when the innovation distribution is regularly varying at the unknown left endpoint θ and right endpoint infinity. If $0 < \phi < 1$, then $\hat{\phi}_n^* \xrightarrow{a.s.} \phi$ and there exists a sequence of constants $b_n = F^{\leftarrow}(1 - 1/n)$ such that

$$P[b_n(\hat{\phi}_n^* - \phi) > x] \to e^{-x^{\beta} EW^{-\beta}}, \tag{4.1.2}$$

where $W = \min_{0\leq i\leq\infty} \frac{\epsilon_i}{\phi^i}$. Since bifurcating processes are typically used to model each line of descent in a binary tree as a standard AR(1) process, it seems natural to expect the limiting distribution for a BAR(1) to be similar with (4.1.2) as the number of generations and henceforth the number of observations tends to infinity.

While most papers consider the finite variance case ($\beta > 2$) in which the model parameter ϕ can be interpreted as the correlation between the mother and daughter cells, in this paper, we consider the infinite variance case ($0 < \beta < 2$) and concentrate on modeling the correlation between mother and daughter cells and do not consider the large correlations between more distant relatives observed by some authors. That is, in Section 4.2 we propose an estimate for the correlation between the mother and daughter where we assume the innovation sequence $(\epsilon_{2t}, \epsilon_{2t+1}), t \geq 1$, is a sequence of independently and identically distributed positive bivariate random variables. Whereas, in Section 4.3 we propose an alternative approach to derive the limiting law for $\hat{\phi}_n = \bigwedge_{t=2}^{n} X_t / X_{\lfloor t/2 \rfloor}$ that removes the complexity and difficulty presented in Theorem 2 of Zhou and Basawa (2005b) under the specified bivariate exponential innovation distribution. The choice of this distribution allows us to consider the correlation between sisters that was not dealt with in Section 4.2. The motivation from a biological rationale for this model is that the sister cells grow in a similar environment, particularly early in their lives, and hence one expects the correlation between sisters to at least exist. In contrast, other more distant relatives, for example cousins, share less of their environment and it seems reasonable to suppose that their environmental effects are independent. Here the process is extended, first by considering alternative estimates for ϕ and θ rather than the typical least-square or maximum likelihood estimate such that innovations $\{\epsilon_t\}$ follow an non-gaussian distribution F and secondly assuming F to be regularly varying at both end-points.

An initial observation is that our estimation procedure relies heavily upon the large innovations, and because of this, the complex dependency that exists within this process becomes less of an issue compared to an maximum likelihood approach presented in other papers. Secondly, our approach is straightforward for both ϕ and θ. That is, the autocorrelation coefficient ϕ in the BAR(1) process is estimated by taking the minimum of the ratio of two sample values while estimation for the unknown location parameter θ was achieved through minimizing $X_t - \hat{\phi}_n X_{\lfloor t/2 \rfloor}$ over the observed series.

The rest of the paper is organized as follows: asymptotic limit results for the autocorrelation parameter ϕ and unknown location parameter θ, are presented in Section 4.2, Section 4.3 verifies that an extreme value theory method produces the same limit law as Theorem 2 in

Zhou and Basawa (2005b), while Section 4.4 is concerned with the small sample size behavior of these estimates through simulation. Additionally, we will investigate the performance of our heuristic approach.

4.2 Results

Suppose $\{X_t\}$ is a sequence of observations of some characteristic of individual t. Beginning with the positive starting value X_1, the first order bifurcating autoregressive process, BAR(1), with a positive left endpoint θ is defined as,

$$X_t = \phi X_{\lfloor t/2 \rfloor} + \epsilon_t, \qquad \text{for } 2 \leq t \leq 2^{(k+1)} - 1, \tag{4.2.1}$$

where $0 < \phi < 1$ and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. We assume that the innovations $\{\epsilon_t\}$ are such that the offspring $(\epsilon_{2t}, \epsilon_{2t+1}), t \geq 1$ are a sequence of independently and identically distributed nonnegative bivariate random vectors with $(\epsilon_{2t}, \epsilon_{2t+1}) \sim F_1$. Additionally, $\{\epsilon_t\}$ is assumed to have the same marginal distribution F such that $\theta = \inf\{x : F(x) > 0\}$ and $\sup\{x : F(x) < 1\} = \infty$. Lastly, we assume F is regularly varying with index α at its positive left endpoint θ, abbreviated $F \in RV_\alpha$, and $\bar{F} = 1 - F$ is regularly varying with index $-\beta$ at infinity, its right endpoint. That is, there exists $\beta > 0$ such that

$$\lim_{t \to \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\beta}, \text{ for all } x > 0. \tag{4.2.2}$$

In this paper we will consider the situation where k generations evolved and the offspring from that generation are included. That is,

$$1 + 2 + 2^2 + \ldots + 2^k = 2^{k+1} - 1.$$

In order to notationally clarify the transition from the k^{th} generation to the t^{th} individual in the BAR(1) process $\{X_t\}$, let

$$n = 2^{k+1} - 1.$$

That is, the total number of individuals in this process including the offspring during the k^{th} generation is n. Now by defining

$$t = \begin{cases} 2\tilde{t} & \text{if } t \text{ is even (male);} \\ 2\tilde{t} + 1 & \text{if } t \text{ is odd (female),} \end{cases}$$

observe that we can now express the process in (4.2.1) from the offspring perspective,

$$\begin{cases} X_{2\tilde{t}} = \phi X_{\tilde{t}} + \epsilon_{2\tilde{t}}, & \text{for } 1 \leq \tilde{t} \leq \lfloor n/2 \rfloor; \\ X_{2\tilde{t}+1} = \phi X_{\tilde{t}} + \epsilon_{2\tilde{t}+1}, & \text{for } 1 \leq \tilde{t} \leq \lfloor n/2 \rfloor. \end{cases} \tag{4.2.3}$$

The process in (4.2.3) allows for a better interpretation of the bifurcating autoregressive process, where the model parameter ϕ represents the strength of correlation between the mother and its offspring under the assumption that the variance is finite. Now denoting $x \wedge y = \min(x, y)$, we have

$$Y_{\tilde{t}} = \epsilon_{2\tilde{t}} \wedge \epsilon_{2\tilde{t}+1} = X_{2\tilde{t}} \wedge X_{2\tilde{t}+1} - \phi X_{\tilde{t}}.$$

Then $Y_{\tilde{t}} \sim G(x) = 2F(x) - F_1(x, x)$, with $G(\theta) = 0$. Finally, the first-order autoregressive process, AR(1), is

$$X_t^* = \phi X_{t-1}^* + \epsilon_t^*, \tag{4.2.4}$$

where $\{\epsilon_t^*\}$ is a sequence of i.i.d. random variables with the same marginal distribution as $\{\epsilon_t\}$.

By assuming (4.2.2), we are considering a time series with heavy-tailed errors and within certain time series applications a better model is achieved. Thus, our goal is to capitalize on the behavior of extreme value estimators over traditional estimators when $0 < \beta < 2$. This contrasts with estimators whose asymptotic behavior depends on the central part of the innovation distribution when a second or higher moment is finite. Since the estimate for θ depends on the estimate for the autocorrelation coefficient, we begin studying the asymptotic properties of $\hat{\phi}_n$ and then move onto asymptotic properties for $\hat{\theta}_n$. The motivation for the natural estimator $\hat{\phi}_n$, comes from the observation when $X_{\lfloor t/2 \rfloor}$ is large, equation (4.2.1) implies

$$0 \leq \phi \leq X_t / X_{\lfloor t/2 \rfloor}. \tag{4.2.5}$$

Therefore, by minimizing the ratio in (4.2.5) we expect

$$\hat{\phi}_n = \bigwedge_{t=2}^{n} \frac{X_t}{X_{\lfloor t/2 \rfloor}}$$

to be a reasonable estimator for ϕ. In this paper, we will implement the same notation on the probability space as presented in Davis and McCormick (1989). That is, we define E to be a two dimensional euclidean space with Borel σ-algebra $\mathcal{B}$. In E, we define ν to be a Radon measure on E and let $M_p(E)$ denote the class of nonnegative integer-valued Radon measures on E, where $\mathfrak{M}_p(E)$ is the σ-algebra generated by the vague topology, and $C_k^+(E)$ consists of all nonnegative continuous functions with compact support. We consider a point measure on E to be a measurable map from a probability space $(\Omega, \Im, P)$ into $(M_p(E), \mathfrak{M}_p(E))$. That is, $M_p(E)$ is topologized by vague convergence and since the vague topology renders $M_p(E)$ a complete separable metric space, we may speak of convergence in distribution of point processes which will be denoted by $\Rightarrow$. Finally, we write N is $PRM(\nu)$ to indicate that N is a Poisson process with intensity measure ν where by a Poisson random measure (PRM(ν))

with mean measure ν, we mean a point process N satisfying

$$P[N(A)=j]=e^{-(\nu(A))}\nu(A)^j/j!, \quad j=0,1,\dots$$

for $A\in E$ with $\nu(A)<\infty$, and furthermore, for any $j\geq 1$ and disjoint measurable sets $A_1,\dots,A_j$, we have that $N(A_1),\dots,N(A_j)$ are independent random variables. Hence, for a point $x\in E$, let $\epsilon_x(\cdot)$ denote the degenerate measure at x.

We now turn to showing that $b_n(\hat{\phi}_n-\phi)$ converges in distribution where

$$b_n=F^{\leftarrow}(1-\frac{1}{n}):=\inf\{x:F(x)\geq(1-1/n)\}.$$

Let $\bar{F}(x)=1-F(x)$. Then we have

$$\lim_{n\to\infty}\frac{\bar{F}(b_nx)}{\bar{F}(b_n)}=x^{-\beta},\ \text{for all } x>0. \tag{4.2.6}$$

Now define $a_n=G^{\leftarrow}(1/n)-\theta$. Then

$$\lim_{n\to\infty}\frac{G(\theta+a_ny)}{G(\theta+a_n)}=y^{\alpha},\ \text{for all } y>0. \tag{4.2.7}$$

First observe that the stationary solution is

$$X_t=\epsilon_t+\phi\epsilon_{\lfloor t/2\rfloor}+\phi^2\epsilon_{\lfloor t/2^2\rfloor}+\dots+\phi^m\epsilon_{\lfloor t/2^m\rfloor}+\dots.$$

Thus we begin with a truncation of X_t's by defining

$$X_t^{(m)}=\sum_{j=0}^{m}\phi^j\epsilon_{\lfloor t/2^j\rfloor}, \quad m\geq 1, t\geq 2^m$$

as an approximation to X_t. Furthermore,

$$X_1^{*(m)} = \sum_{j=0}^{m} \phi^j \epsilon_{2^j} \stackrel{d}{=} X_t^{(m)}, \quad m \geq 1 \quad X_1^* = X_1^{*(\infty)}. \tag{4.2.8}$$

Now using the fact that $Y_{\tilde{t}} = \epsilon_{2\tilde{t}} \wedge \epsilon_{2\tilde{t}+1}$, we can determine the necessary point process, since

$$\begin{aligned}
P[b_n(\hat{\phi}_n - \phi) > x] &= P\left[\bigwedge_{t=2}^{n} \left(\frac{X_t - \phi X_{\lfloor t/2 \rfloor}}{b_n^{-1} X_{\lfloor t/2 \rfloor}}\right) > x\right] \\
&= P\left[\bigwedge_{t=2}^{n} \left(\frac{\epsilon_t}{b_n^{-1} X_{\lfloor t/2 \rfloor}}\right) > x\right] \\
&= P\left[\bigwedge_{\tilde{t}=1}^{\lfloor n/2 \rfloor} \left(\frac{\epsilon_{2\tilde{t}} \wedge \epsilon_{2\tilde{t}+1}}{b_n^{-1} X_{\tilde{t}}}\right) > x\right] \\
&= P\left[\bigwedge_{\tilde{t}=1}^{\lfloor n/2 \rfloor} \left(\frac{Y_{\tilde{t}}}{b_n^{-1} X_{\tilde{t}}}\right) > x\right].
\end{aligned}$$

Thus we define the following point process:

$$\mathcal{I}_n = \sum_{\tilde{t}=1}^{\lfloor n/2 \rfloor} \varepsilon_{(Y_{\tilde{t}}, b_n^{-1} X_{\tilde{t}})} \quad \text{and} \quad \mathcal{I}_n^{(m)} = \sum_{\tilde{t}=1}^{\lfloor n/2 \rfloor} \varepsilon_{(Y_{\tilde{t}}, b_n^{-1} X_{\tilde{t}}^{(m)})}.$$

Observe that the point process $\mathcal{I}_n$ consists of two independent components, where the first component consists of the marks for the minimum of the offspring individuals $Y_{\tilde{t}}$ and the second component consists of the points from the parents $b_n^{-1} X_{\tilde{t}}$.

Since we are looking at the first order bifurcating process from the natural perspective of (4.2.3), we will let $t = \tilde{t}$, so that $1 \leq t \leq \lfloor n/2 \rfloor$.

Now we consider establishing convergence of the point process $\mathcal{I}_n^{(m)}$ by first defining rectangles

$$R_i = [a_i, b_i] \times [a_i^{'}, b_i^{'}], 1 \leq i \leq q. \tag{4.2.9}$$

Thus, we need to show for any $q \geq 1$ that the $q-$dimensional distribution converges. That is,

$$(\mathcal{I}_n^{(m)}(R_1), \ldots, \mathcal{I}_n^{(m)}(R_q)) \xrightarrow{d} Pois(\lambda_1) \times \ldots \times Pois(\lambda_q) \quad \text{as } n \to \infty, \tag{4.2.10}$$

where $\lambda_i \equiv \lambda_i^{(m)} = \lim_{n\to\infty} E[\mathcal{I}_n^{(m)}(R_i)]$ and $Pois(\lambda)$ denotes a Poisson distribution with parameter λ, while $X \times Y$ means that X and Y are independent.

We will prove (4.2.10) for the case of $q = 2$. That is, we will show that if $a_1^{'} < b_1^{'} < a_2^{'} < b_2^{'}$ so that $R_1 \cap R_2 = \emptyset$, then

$$(\mathcal{I}_n^{(m)}(R_1), \mathcal{I}_n^{(m)}(R_2)) \Rightarrow Pois(\lambda_1) \times Pois(\lambda_2) \quad \text{as } n \to \infty.$$

The proof of the other cases is similar, thus omitted. Now suppose that we have constructed arbitrary blocks $Q_{l,s}$ in such a way that $X_t^{(m)}$ and $X_{t'}^{(m)}$ are independent for all t if $t \in Q_{l,s}$ and $t' \in Q_{l',s'}$ such that $(l,s) \neq (l',s')$. Under this assumption. we proceed with the argument and define the following indicator function's

$$I_{(l,s)}^i = \begin{cases} 1 & \text{if } \sum_{t\in Q_{l,s}} \mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i] \geq 1 \text{ for } i = 1,2, \\ 0 & \text{otherwise }, \end{cases} \tag{4.2.11}$$

and

$$\tilde{I}_{(l,s)}^i = \begin{cases} 1 & \text{if } \sum_{t\in Q_{l,s}} \mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i] \geq 2 \text{ for } i = 1,2, \\ 0 & \text{otherwise }. \end{cases} \tag{4.2.12}$$

Now let $I_{(l,s)} = (I_{(l,s)}^1, I_{(l,s)}^2)$ and $\tilde{I}_{(l,s)} = (\tilde{I}_{(l,s)}^1, \tilde{I}_{(l,s)}^2)$.

Now observe that $\{I_{(l,s)} : 1 \leq s \leq \lfloor 2^l/r \rfloor, l = l_0, \ldots, k\}$ are i.i.d. Bernoulli random vectors. The following lemma shows if there exists at least two different indices in the same interval $Q_{l,s}$ such that the points in $\mathcal{I}_n^{(m)}$ fall within either region R_1, R_2 or both for some (l,s) then the event is negligible as n tends to infinity.

First observe since $X_t^{(m)}$ and $X_{t'}^{(m)}$ are independent if $t \in Q_{l,s}$ and $t' \in Q_{l',s'}$ with $(l, s) \neq (l', s')$ then $\sum_{t \in Q_{l,s}} \mathbb{1}[(Y_t, b_n^{-1} X_t^{(m)}) \in R_i]$ and $\sum_{t' \in Q_{l',s'}} \mathbb{1}[(Y_{t'}, b_n^{-1} X_{t'}^{(m)}) \in R_i]$ are independent for $i = 1, 2$.

Lemma 4.2.1. *Suppose $t \in Q_{l,s}$ and $t' \in Q_{l',s'}$ such that $(l, s) \neq (l', s')$. Under the assumptions that $0 < \phi < 1, \theta > 0, \bar{F} \in RV_{-\beta}$ we have*

$$P\left[\sum_{l=l_0}^{k} \sum_{s=1}^{\lfloor 2^l/r \rfloor} \tilde{I}_{(l,s)} \neq (0,0)\right] \to 0, \;\; \textit{as } k \to \infty.$$

Remark 1. By definition as k tends to infinity $n = 2^{k+1} - 1$ tends to infinity and therefore we can speak of regular variation in terms of k.

The next lemma shows the probability that there exists at least one index in the same interval $Q_{l,s}$ for each component of $I_{(l,s)}$, such that the point or points in $\mathcal{I}_n^{(m)}$ fall within region R_i is zero for $i = 1, 2$ and some (l, s).

Lemma 4.2.2. *Under the conditions specified in Lemma 4.2.1 for some $l_0 \leq l \leq k, 1 \leq s \leq \lfloor 2^l/r \rfloor$, we have*

$$P[I_{(l,s)} = (1,1)] = o\left(\frac{r}{2^{k+1}}\right).$$

The following lemma calculates the probability that there exists at least one index in the same interval $Q_{l,s}$, such that the point or points in $\mathcal{I}_n^{(m)}$ fall within exactly one region for some (l, s).

Lemma 4.2.3. *Let $Y_1 \sim G$. Then under the conditions specified in Lemma 4.2.1 for $r = o(2^k)$ and some $l_0 \leq l \leq k, 1 \leq s \leq \lfloor 2^l/r \rfloor$, we have*

$$\frac{2^{k+1}}{r} P[I_{(l,s)} = (j, 1-j)] \sim \gamma_{2-j}, \quad \textit{for } j = 0, 1, \quad \textit{and} \quad k \textit{ large}$$

where $\gamma_i = (G(b_i) - G(a_i)) \cdot \kappa_i$ and $\kappa_i = (b_i^{'-\beta} - a_i^{'-\beta})(1 - \phi^{(m\beta)})/(1 - \phi^\beta)$ for $i = 1, 2$.

The next lemma determines the limiting distribution for our i.i.d. random vector $I_{(l,s)}$ through the use of the moment generating function.

Lemma 4.2.4. *Under the conditions specified in Lemma 4.2.1 for $r = o(2^k)$ we have*

$$P[I_{(l,s)} = (i,j)] = \begin{cases} \frac{r}{2^{k+1}}\gamma_{(i,j)}(1+o(1)) & \text{if } 0 \le i,j \le 1, i+j=1 \ , \\ 1 - \frac{r}{2^{k+1}}(\gamma_{(1,0)} + \gamma_{(0,1)})(1+o(1)) & \text{if } i=j=0 \ , \end{cases}$$

where $\gamma_{(i,j)} = (G(b_{2-i}) - G(a_{2-i}))(b_{2-i}^{'-\beta} - a_{2-i}^{'-\beta})(1-\phi^{(m\beta)})/(1-\phi^{\beta})$ for $i+j=1$. Additionally,

$$\sum_{l=l_0}^{k} \sum_{s=1}^{\lfloor 2^l/r \rfloor} I_{(l,s)} = (I^1_{(l,s)}, I^2_{(l,s)}) \Rightarrow Pois(\lambda_1) \times Pois(\lambda_2), \qquad \text{as } k \to \infty,$$

where $\lambda_i = (G(b_i) - G(a_i))(b_i^{'-\beta} - a_i^{'-\beta})(1-\phi^{(m\beta)})/(1-\phi^{\beta})$ for $i = 1,2$.

Applying the previous lemma's yields our first main result.

Theorem 4.2.1. *Consider the stationary BAR(1) process $\{X_t\}$ from* (4.2.1) *where F satisfies* (4.2.2). *Then for any $m \ge 1$ and disjoint sets, $R_i := [a_i, b_i] \times [a_i^{'}, b_i^{'}]$ for $i = 1,2$ we have*

$$(\mathcal{I}_n^{(m)}(R_1), \mathcal{I}_n^{(m)}(R_2)) \Rightarrow Pois(\lambda_1) \times Pois(\lambda_2), \qquad \text{as } n \to \infty,$$

where $\lambda_i \equiv \lambda_i^{(m)} = \lim_{n\to\infty} E[\mathcal{I}_n^{(m)}(R_i)] = (G(b_i) - G(a_i))(b_i^{'-\beta} - a_i^{'-\beta})(1-\phi^{(m\beta)})/(1-\phi^{\beta})$, for $i = 1,2$.

Recall that if n is the total number of observations in k generations, we have $n = 2^{k+1} - 1$ or $k = \log_2(n+1) - 1$. The following corollary produces the limiting distribution for $\hat{\phi}_n$ as the number of generations tends to infinity.

Corollary 4.2.1. *Consider the estimator of ϕ, $\hat{\phi}_n = \bigwedge_{t=2}^{n} \frac{X_t}{X_{\lfloor t/2 \rfloor}}$. Suppose $0 < \phi < 1, \theta > 0, \bar{F} \in RV_{-\beta}$ and $EY^{-\gamma} < \infty$ for some $\gamma > \beta$, then*

$$\lim_{n\to\infty} P[b_n(\hat{\phi}_n - \phi) > x] = e^{-x^{\beta} EY^{-\beta}(1-\phi^{\beta})^{-1}}, \qquad \text{for all } x > 0,$$

where Y has the stationary distribution G for the process (4.2.1).

We now shift our attention to the positive unknown location parameter θ. The motivation for an estimator of θ arrives from the observation that, $X_t - \hat{\phi}_n X_{\lfloor t/2 \rfloor}$ can be expressed as $-(\hat{\phi}_n - \phi)X_{\lfloor t/2 \rfloor} + \epsilon_t$. Now since $\bigwedge_{t=2}^n \epsilon_t \xrightarrow{a.s.} \theta$ and $\hat{\phi}_n \xrightarrow{p} \phi$ as $n \to \infty$ allows us to define our estimator for θ:

$$\hat{\theta}_n = \bigwedge_{t \in I_n} (X_t - \hat{\phi}_n X_{\lfloor t/2 \rfloor}), \tag{4.2.13}$$

where we define the index set $I_n = \{t : 2 \leq t \leq n \text{ and } X_{\lfloor t/2 \rfloor} \leq (a_n b_n)^\rho\}$ where $0 < \rho < 1$ is a fixed value.

Now to determine the limiting distribution for $\hat{\theta}$ observe that,

$$\hat{\theta}_n - \theta = [(\hat{\theta}_n - \bigwedge_{t \in I_n} Y_t) + (\bigwedge_{t \in I_n} Y_t - \theta)]. \tag{4.2.14}$$

The following lemma show that the first term in (4.2.14) goes to zero in probability, thus allowing us to focus only on the second term.

Lemma 4.2.5. *Under the assumptions that G is regularly varying with index α at its positive left endpoint θ and $\bar{F}$ is regularly varying with index $-\beta$ at infinity, its right endpoint, and $\alpha > \beta$, then*

$$a_n^{-1}\Big(\hat{\theta}_n - \bigwedge_{t \in I_n} Y_t\Big) \xrightarrow{p} 0, \ \text{as } n \to \infty,$$

where $a_n = G^{\leftarrow}(1/n) - \theta$.

Now we use point processes to show that the second term in (4.2.14) converges (weakly) to a Poisson point process with mean measure $y^\alpha/2$.

Theorem 4.2.2. *Consider the stationary BAR(1) process $\{X_t\}$ from* (4.2.1) *where G satisfies* (4.2.7). *Let $\mathcal{V}_n$ and $\mathcal{V}$ be the point processes on the space $E_2 = [0, \infty)$ defined by*

$$\mathcal{V}_n = \sum_{t=1}^{\lfloor n/2 \rfloor} \varepsilon_{(a_n^{-1}(Y_t - \theta))} \quad \text{and} \quad \mathcal{V} = \sum_{p=1}^{\infty} \varepsilon_{j_p},$$

where $\sum_{p=1}^{\infty} \varepsilon_{j_p}$ *is* $PRM(\nu)$ *with* $\nu[0,y] = y^\alpha/2,\ y > 0$. *Then in* $M_p(E_2)$,

$$\mathcal{V}_n \Rightarrow \mathcal{V}.$$

The following lemma is an alternative approach to show that the point process used in Theorem 4.2.2 suffices.

Lemma 4.2.6. *Consider the stationary BAR(1) process* $\{X_t\}$ *from* (4.2.1) *where* G *satisfies* (4.2.7). *Let* $\tilde{\mathcal{V}}_n$ *and* $\mathcal{V}_n$ *be the point processes on the space* $E_3 = [0,\infty) \times [0,\infty)$ *defined by*

$$\tilde{\mathcal{V}}_n = \sum_{t=1}^{\lfloor n/2 \rfloor} \varepsilon_{(a_n^{-1}(Y_t - \theta), c_n)} \ \text{and} \quad \mathcal{V}_n = \sum_{t=1}^{\lfloor n/2 \rfloor} \varepsilon_{(a_n^{-1}(Y_t - \theta), 0)},$$

where $c_n = a_n^{-1}|(\hat{\phi}_n - \phi)|X_t$. *Then for* $t \in I_n$

$$d(\tilde{\mathcal{V}}_n, \mathcal{V}_n) \xrightarrow{p} 0,$$

where d *is the vague metric on* $M_p(E_3)$.

The following corollary uses the continuous mapping theorem to obtain the limiting distribution for $\hat{\theta}_n$ as n tends to infinity.

Corollary 4.2.2. *Consider the estimator of* θ, $\hat{\theta}_n = \bigwedge_{t \in I_n} (X_t - \hat{\phi}_n X_{\lfloor t/2 \rfloor})$. *Suppose* $\theta > 0$ *and* F *is* RV_α *at* θ. *If* $\alpha > \beta$ *then for any* $y > 0$ *we have*

$$\lim_{n \to \infty} P[a_n^{-1}(\hat{\theta}_n - \theta) > y] = e^{-y^\alpha/2}.$$

4.3 The Elegance of an Extreme Value Approach

In this section we use an extreme value approach to obtain the limit law for $\hat{\phi}_n$ under the assumption that $Y_t = \epsilon_{2t} \wedge \epsilon_{2t+1}$ is positive and the marginal distribution G is regularly varying at $\theta = 0$ with index α. In the previous section, we obtain the limit law for $\hat{\phi}_n$ under the assumption that the innovation distribution for $\{\epsilon_t\}$ was regularly varying at infinity with index $-\beta$. The goal in this section is not only designed to fill in the gaps for the proof of Proposition 2 in Zhang (2011), but present an alternative derivation that under the correct parameterization obtains the same limit law found in Theorem 2 of Zhou and Basawa (2005b). Naturally this approach may seem like over kill, but it demonstrates the pure power and elegance that an extreme value approach can have as we will show that complexity and difficulty found in Zhou and Basawa (2005b) approach is unnecessary. We continue in this section with the usual first-order bifurcating autoregressive process defined by

$$X_t = \phi X_{\lfloor \frac{t}{2} \rfloor} + \epsilon_t, \qquad \text{for } 2 \leq t \leq n, \tag{4.3.1}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

For the model in (4.3.1) we assume that $0 < \phi < 1$ and the innovations ϵ_t are such that $(\epsilon_{2t}, \epsilon_{2t+1})$ are i.i.d with $(\epsilon_{2t}, \epsilon_{2t+1}) \sim F_1$ and ϵ_t has the same marginal distribution F satisfying $F(\theta) = 0$. By defining $G(x) = 2F(x) - F_1(x,x)$, we assume G is regularly varying at $\theta = 0$ with index α. The following Theorem generalizes the result presented in Zhang (2011) where the author assumes Weibull type innovations.

Theorem 4.3.1. *Let $d_n = G^{\leftarrow}(1/n)$ and consider the estimator of ϕ, $\hat{\phi}_n = \bigwedge_{t=2}^{n} X_t / X_{\lfloor t/2 \rfloor}$. Under the assumption that $G \in RV_\alpha$ and $EX_1^\gamma < \infty$ for some $\gamma > \alpha$, we have*

$$\lim_{n\to\infty} P[d_n^{-1}(\hat{\phi}_n - \phi) > y] = e^{-y^\alpha} EX_1^\alpha / 2,$$

where X_1 has the stationary distribution H for the process in (4.3.1).

Remark 2. The stationary distribution H is the same as the stationary distribution of the AR(1) sequence $\tilde{X}_t = \phi\tilde{X}_{t-1} + \tilde{\epsilon}_t$, where the $\tilde{\epsilon}_t$ are i.i.d. with the same marginal distribution as ϵ_t.

The following Corollary is a verification that our extreme value method under the specified bivariate exponential innovation distribution is in agreement with the limit law presented in Theorem 2 of Zhou and Basawa (2005b). That is, suppose the joint distribution of $(\epsilon_{2t}, \epsilon_{2t+1})$ is specified by

$$\bar{F}_1(x_1, x_2) = P[\epsilon_{2t} > x_1, \epsilon_{2t+1} > x_2] = \exp\left(-\alpha(x_1 + x_2) - \beta(x_1 \vee x_2)\right), \quad x_1, x_2 > 0, \tag{4.3.2}$$

where α and β are the model parameters satisfying $\alpha > 0, \beta > 0$.

Observe that the marginal distribution of ϵ_{2t} and ϵ_{2t+1} are exponential with mean $(\alpha+\beta)^{-1}$ and correlation $\rho = \beta(2\alpha + \beta)^{-1}$.

Now we consider the parameterization

$$\alpha = \frac{1-\rho}{(1+\rho)\lambda} \quad \text{and} \quad \beta = \frac{2\rho}{(1+\rho)\lambda}, \tag{4.3.3}$$

where $\lambda > 0$ and $0 \le \rho < 1$. With this parameterization, the marginal distributions of ϵ_{2t} and ϵ_{2t+1} are both exponential with mean λ and correlation ρ. The correlation parameter ρ represents the correlation between sisters in the bifurcating process. Observe when $\rho = 0$, the innovations $\{\epsilon_t\}$ in (4.3.1) will be independent and identically exponential distributed random variables.

Corollary 4.3.1. *Suppose* $(\epsilon_{2t}, \epsilon_{2t+1}) \sim F_1$*, where* F_1 *is specified in* (4.3.2)*. Then with the parameterization defined in* (4.3.3) *we have*

$$\lim_{n\to\infty} P\left[\frac{n}{(1+\rho)(1-\phi)}(\hat{\phi}_n - \phi) > y\right] = e^{-y}.$$

4.4 Simulation Study

In this section, we assess the reliability of our extreme value estimation method through a simulation study, which examined the finite sample properties of $(\hat{\phi}_n, \hat{\theta}_n)$. Additionally, the degree of approximation for the empirical probabilities of $\hat{\phi}_n$ and $\hat{\theta}_n$ to its respective limiting distribution is reported to verify asymptotically that the dependency among $X_t^{(m)}$, which was not proven theoretically in Section 4.2 is negligible as the total number of observations tends to infinity.

To study the performance of the estimators $\hat{\phi}_n = \bigwedge_{t=2}^{n} \frac{X_t}{X_{\lfloor t/2 \rfloor}}$ and $\hat{\theta}_n = \bigwedge_{t \in I_n} (X_t - \hat{\phi}_n X_{\lfloor t/2 \rfloor})$, we generated 5,000 independent samples of size $n = 2^{k+1} - 1$, where k is the number of generations and $\{X_t\}$ is an BAR(1) process satisfying the difference equation

$$X_t = \phi X_{\lfloor t/2 \rfloor} + \epsilon_t, \quad \text{for } 2 \leq t \leq n \quad \text{and } \epsilon_t \geq \theta.$$

The autoregressive parameter ϕ is taken to be in the range from 0 to 1 guaranteeing a nonnegative time series and the unknown location parameter θ is positive. In order to perform this simulation the following ad hoc approach was adopted to generate bivariate random variables from a distribution that is regularly varying at both endpoints. First, let Z_1 and Z_2 be independent random variables that are taken from

$$F(z) = \begin{cases} c(z-\theta)^{\alpha} & \text{if } \theta < z < \theta + 1, \\ 1 - d(z-\theta)^{-\beta} & \text{if } \theta + 1 < z < \infty. \end{cases}$$

For this innovation distribution let c and d be nonnegative constants such that $c + d = 1$, then this distribution is regularly varying at both endpoints with index of regular variation $-\beta$ at infinity and index of regular variation α at θ. Now define

$$\epsilon_{2t} = a_1 Z_1 + b_1 Z_2 \quad \text{and} \quad \epsilon_{2t+1} = a_2 Z_1 + b_2 Z_2,$$

where a_i and b_i are nonnegative constants for $i = 1, 2$.

For this simulation study two distributions were considered: (i) $F_1, c = .5, d = .5, a_1 = 1, a_2 = 0, b_1 = 0, b_2 = 1$ $\quad (ii)$ $F_2, c = .5, d = .5, a_1 = .4, a_2 = .6, b_1 = .3, b_2 = .7$.

Observe in case (i) that $(\epsilon_{2t}, \epsilon_{2t+1})$ are i.i.d. with a regular varying tail distribution at infinity with index $-\beta$ and regular varying at θ with index α, whereas in case (ii) $(\epsilon_{2t}, \epsilon_{2t+1})$ are correlated with a regular varying tail with index -2β at infinity and index 2α at θ.

Choosing $k = 9, 10, 11, 12$ and $n = 2^{k+1} - 1$, the sample mean and standard error (s.e.) of the estimates are given in Table 1 and 2 for $\phi = .2$, $\theta = 1$, respectively. Additionally, the average lengths for 95 percent empirical confidence intervals with exact coverage are also reported. Note that the confidence intervals were directly constructed from the empirical distributions of $n^{1/\beta}(\hat{\phi}_n - \phi)$ and $n^{-1/\alpha}(\hat{\theta}_n - \theta)$ respectively, while the exponent ρ inside the index set $I_n = \{t : 2 \leq t \leq n \text{ and } X_{\lfloor t/2 \rfloor} \leq (a_n b_n)^\rho\}$, was set to .9.

We first examine the simulation results in Table 1 for different number of generations. As k increases, the standard errors and biases of $\hat{\phi}_n$ and $\hat{\theta}_n$ decrease. In particular, when $\beta = .2$ and k increases from 9 to 11, the standard error becomes 693 times smaller. Similarly, the 95% confidence interval average length is 118 times smaller.

Next we look at the behavior of the estimators as β increases. Not surprisingly, the standard errors and bias get larger as β increases. This is expected since our extreme value method of estimation depends heavily on obtaining large innovations. Thus, it can be shown when the regular varying index is small (less than 1) the largest innovation ϵ_t will be extremely large, and only in the situation when the next innovation is also large does our estimator behave badly. Whereas, if β takes on values larger than one, then the largest innovation is not likely to be nearly as large, thus the chance that we get a bad estimate increases since now it only takes a moderately above normal innovation to produce an extreme estimate. Furthermore, the effect of the autoregressive parameter ϕ is that if it takes on a value near its lower bound 0, then the largest innovation is considerably reduced in value, thus allowing a large spectrum of moderate to normal values for the next innovation to cause bad estimates.

Therefore, we can expect with small probability some extreme estimates from our estimator, which ultimately affect the length for a 99% confidence interval more than the length for a 95% confidence interval.

Table 4.1: Performance of $(\hat{\phi}_n, \hat{\theta}_n)$ with $(\phi = .2, \theta = 1)$ and $\alpha = 1$ under F_1

		$\hat{\phi}_n$		$\hat{\theta}_n$		95% C.I. Avg. Length	
k	β	mean	s.e.	mean	s.e.	$\hat{\phi}_n$	$\hat{\theta}_n$
9	.2	.20001	(1.38×10^{-9})	1.0026	(.0051)	1.37×10^{-4}	.0184
	.8	.20029	(1.24×10^{-3})	1.017	(.0131)	1.12×10^{-3}	.0347
10	.2	.20008	(4.17×10^{-11})	1.0043	(.0047)	3.6×10^{-4}	.0112
	.8	.20012	(1.43×10^{-5})	1.013	(.0125)	5.3×10^{-2}	.0238
11	.2	.200007	(1.99×10^{-12})	1.0037	(.0041)	1.16×10^{-6}	.0077
	.8	.20037	(6.4×10^{-6})	1.011	(.0116)	2.3×10^{-4}	.0146
12	.2	.200002	(1.03×10^{-13})	1.0024	(.0028)	4.72×10^{-8}	.0052
	.8	.20079	(2.8×10^{-7})	1.007	(.0103)	9.8×10^{-5}	.0113

We now turn our attention to Table 2. The purpose of this table is to see whether or not the correlation between the sisters $(\epsilon_{2t}, \epsilon_{2t+1})$ affects our estimates for the autocorrelation between the mother and daughter. The results are expected from a biological viewpoint, as one expects the environmental correlation between the sisters to be distinct from the environmental correlations inherited from the mother. Hence, the results seem to suggest that a cell's attributes can be explained by inheritance from its mother, suggesting that a BAR(1) model for a single line of descent is appropriate.

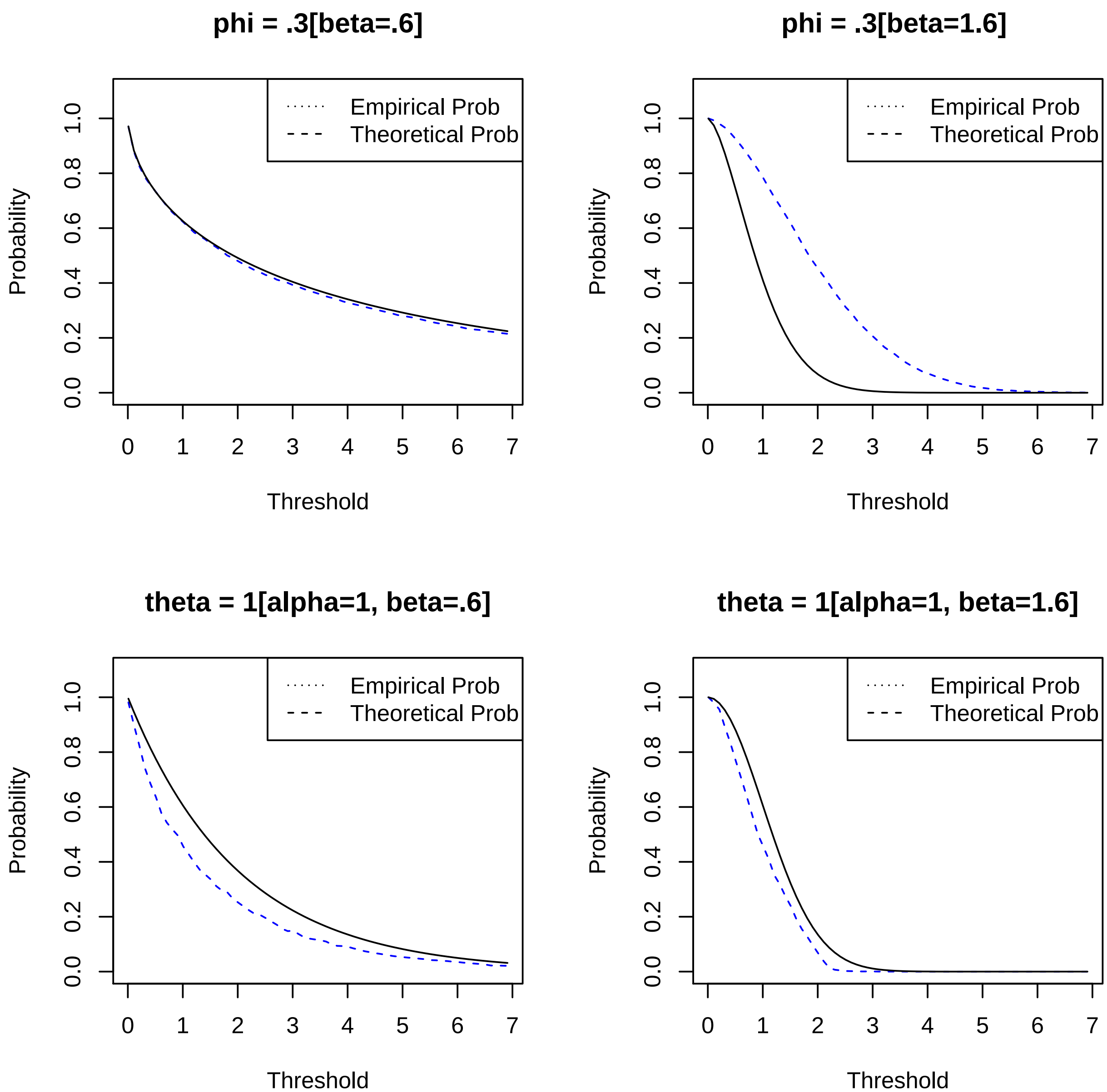
Table 4.2: Performance of $(\hat{\phi}_n, \hat{\theta}_n)$ with $(\phi = .2, \theta = 1)$ and $\alpha = 2$ under F_2

		$\hat{\phi}_n$		$\hat{\theta}_n$		95% C.I. Avg. Length	
k	β	mean	s.e.	mean	s.e.	$\hat{\phi}_n$	$\hat{\theta}_n$
9	.4	.2008	(5.85×10^{-9})	1.0139	(.0149)	1.39×10^{-4}	.0218
	1.2	.2005	(9.9×10^{-4})	1.0348	(.0345)	2.3×10^{-3}	.0715
10	.4	.20017	(1.74×10^{-10})	1.0214	(.0162)	4.56×10^{-5}	.0478
	1.2	.2002	(3.07×10^{-4})	1.0372	(.04061)	7.2×10^{-3}	.0536
11	.4	.20009	(3.28×10^{-10})	1.0187	(.00978)	2.81×10^{-6}	.0325
	1.2	.20048	(5.58×10^{-4})	1.0368	(.0679)	4.8×10^{-4}	.0419
12	.4	.200083	(7.62×10^{-11})	1.0145	(.01033)	8.14×10^{-8}	.0301
	1.2	.20052	(2.34×10^{-4})	1.0282	(.0402)	1.45×10^{-6}	.0376

Lastly, we performed a Monte Carlo simulation to study the degree of approximation for the empirical probability $P[b_n(\hat{\phi}_n - \phi) > x]$, $P[a_n^{-1}(\hat{\theta}_n - \theta) > y]$ to its limiting values $e^{-x^\beta EY^{-\beta}(1-\phi^\beta)^{-1}}$ and $e^{-y^\alpha/2}$ respectively. The empirical distributions were calculated from 5,000 replications of the nonnegative time series $(X_0, X_1, \ldots, X_n)$ for a sample size of $2^{10+1} - 1 = 2,047$. Additionally, we restricted $\alpha > \beta$. The top two plots in Figure 4.2 below shows the performance when $\epsilon_t \sim F_1$ and the autocorrelation coefficient ϕ is .3, α is equal to $(1, 2)$ and β is equal to $(.6, 1.6)$ respectively. Observe for $0 < x < 7$ that the empirical tail probability $b_n(\hat{\phi}_n - \phi) > x$ mirrors the theoretical probability quite nicely, leading us to believe that indeed the dependency among $X_t^{(m)}$ is negligible as the total number of observations tends to infinity.

The bottom two plots in Figure 4.2 displays the asymptotic performance when $\epsilon_t \sim F_1$ and the location parameter θ is 1 for $\alpha = 1, \beta = .6$ and $\alpha = 2, \beta = 1.6$ respectively. Observe that the lower left plot solidifies the asymptotic result presented in Corollary 2.2. However, the lower right plot shows that the convergence rate of the empirical probability to the theoretical probability is extremely slow when β is greater than one. This is not surprising since on average our estimate falls more than .1 from the true value when $\beta = 1.6$.

Figure 4.2: Empirical vs. Theoretical Probability



4.5 Proofs

Prior to proving any results, our first objective is to look at the dependency among $X_t^{(m)}$. Upon doing so, we realized that the dependency within each tree segment is more delicate than anticipated, hence the following argument should be considered heuristic. The thought process to obtain independence was to determine how much distance was needed between observations. With this in mind, we begin by partitioning the index set $[1, 2^{k+1} - 1]$.
Thus, for the l^{th} generation we consider intervals

$$Z_{l,s} = [2^l + (s-1)r, 2^l + sr - 1],$$

for $s = 1, \ldots, \lfloor 2^l/r \rfloor, l = l_0, \ldots, k$, where l_0 is such that $2^{l_0 - 1} < r \leq 2^{l_0}$ and $r \geq 2^{m+1} + 1$. Notice that the number of indices in each interval is at most

$$2^{l+1} - 1 - 2^l - r(2^l/r - 1) + 1 = r.$$

Now we consider trimming the intervals by 2^{m+1} in hopes of achieving the necessary independence. Thus, we define

$$Q_{l,s} = [2^l + (s-1)r, 2^l + sr - 1 - 2^{m+1}].$$

Then

$$dist(Q_{l,s}, Q_{l,s+1}) = 2^l + sr - (2^l + sr - 1 - 2^{m+1}) = 2^{m+1} + 1 > 2^{m+1}.$$

Therefore,

$$X_t^{(m)} \text{ is independent of } X_{t'}^{(m)}, \quad \text{for all} \quad t \in Q_{l,s} \quad \text{and} \quad t' \in Q_{l',s'}$$

provided $(l, s) \neq (l', s')$. This is true by construction when $l = l'$ and $s \neq s'$. In the case $l \neq l'$, lets look at $Q_{l,s}$ and $Q_{l-1,\lfloor 2^{l-1}/r \rfloor}$. These blocks will be the closest when $l = l'$, namely, the first block in the l^{th} generation and the last block in the $(l-1)^{th}$ generation. For these, we have

$$dist(Q_{l,s}, Q_{l-1,\lfloor 2^{l-1}/r \rfloor}) = 2^l - (2^{l-1} + \lfloor 2^{l-1}/r \rfloor - 1 - 2^{m+1} \geq 2^{m+1} + 1 > 2^{m+1}.$$

Remark 1. As stated above, it is not the case that $X_t^{(m)}$ will be independent of $X_{t'}^{(m)}$ for all t and t'. That is, most of the time the distance between indices t and t' will be large enough so that observations $X_t^{(m)}$ and $X_{t'}^{(m)}$ will be independent, but there are a few scenarios where this is not true. While this is a concern, we will show that asymptotically this dependency does not affect the limit laws. Additionally, the reason for setting the initial point of $Q_{l,s}$ to 2^l, is due to the observations of negligible events outside the set $Q_{l,s}$.

Proof of Lemma 4.2.1.

Proof. Here we sketch how a proof would go. Assuming we have independence between $X_{t'}^{(m)}$ and $X_t^{(m)}$ we could use (4.2.8) to complete the proof, since applying Lemma 4.24 in Resnick (1987) for large k gives us

$$\begin{aligned}
P\left[\sum_{l=l_0}^{k} \sum_{s=1}^{\lfloor 2^l/r \rfloor} \tilde{I}^i_{(l,s)} \geq 1\right] &\leq \frac{2^{k+1}}{r} P[b_n^{-1} X_{t'}^{(m)} \geq a_i^{'}, b_n^{-1} X_t^{(m)} \geq a_i^{'}, t > t' \geq r] \\
&\leq \frac{2^{k+1}}{r} r^2 P^2[X_1^{*(m)} \geq b_n a_i^{'}] \\
&\leq \frac{r}{2^{k+1}} \left[P[\sum_{j=0}^{m} \phi^j \epsilon_{2^j} \geq b_n a_i^{'}] / \bar{F}(b_n) \right]^2 \\
&\leq \frac{r}{2^{k+1}} a_i^{' -2\beta} \sum_{j=0}^{m} \phi^{(2\beta)j} \\
&= o(1).
\end{aligned}$$

□

Proof of Lemma 4.2.2.

Proof. Observe that $r = o(2^k)$ for large k. Now using the Bonferroni inequality and Lemma 4.24 in Resnick (1987) we have

$$\begin{aligned}
P\left[\bigcup_{l,s}\left(I_{(l,s)} = (1,1)\right)\right] &\leq \sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r\rfloor} P\left[I_{(l,s)} = (1,1)\right] \\
&\leq \frac{2^{k+1}}{r} P[a_1^{'} \leq b_n^{-1}X_{t'}^{(m)} \leq b_1^{'}, a_2^{'} \leq b_k^{-1}X_t^{(m)} \leq b_2^{'},\ \text{for some } t > t' \geq r] \\
&\leq r2^{k+1}P^2[X_1^{*(m)} \geq b_n a_1^{'}] \\
&= \frac{r}{2^{k+1}}\left[P[X_1^{*(m)} \geq b_n a_1^{'}]/\bar{F}(b_n)\right]^2 \\
&\leq \frac{r}{2^{k+1}} a_1^{'-2\beta}\sum_{j=0}^{m}\phi^{(2\beta)j} \\
&= o\left(\frac{r}{2^{k+1}}\right).
\end{aligned}$$

$\square$

Proof of Lemma 4.2.3.

Proof. With out loss of generality suppose $j = 1$. Then applying Lemma 4.24 in Resnick (1987) for $\epsilon > 0$ and k large we have

$$\begin{aligned}
P[I_{(l,s)} = (1,0)] &\leq r2^{k+1}P[a_1 \leq Y_1 \leq b_1]P[b_n a_1^{'} \leq X_1^{*(m)} \leq b_n b_1^{'}] \\
&\leq (1+\epsilon)r2^{k+1}(G(b_1) - G(a_1))P[b_n a_1^{'} \leq \sum_{j=0}^{m}\phi^j\epsilon_{2^j} \leq b_n b_1^{'}] \\
&= (1+\epsilon)\frac{r}{2^{k+1}}(G(b_1) - G(a_1))(b_1^{'-\beta} - a_1^{'-\beta})\frac{1-\phi^{(m\beta)}}{1-\phi^{\beta}}.
\end{aligned}$$

Thus,

$$\limsup_{k\to\infty}\frac{2^{k+1}}{r}P[I_{(l,s)} = (1,0)] \leq \gamma_1.$$

Now to obtain a lower bound observe that

$$\begin{aligned}
P[I_{(l,s)} = (1,0)] &\geq (r - 2^{m+1})P[a_1 \leq Y_1 \leq b_1]P[b_n a_1^{'} \leq X_1^{*(m)} \leq b_n b_1^{'}] \\
&\quad - r^2 P^2[b_n a_1^{'} \leq X_1^{*(m)} \leq b_n b_1^{'}] \\
&\geq (1-\epsilon)\frac{r}{2^{k+1}}\gamma_1 - 2\left(\frac{r}{2^{k+1}}\right)^2 (b_1^{'-2\beta} - a_1^{'-2\beta}) \sum_{j=0}^{m} \phi^{(2\beta)j}.
\end{aligned}$$

Then

$$\liminf_{k\to\infty} \frac{2^{k+1}}{r} P[I_{(l,s)} = (1,0)] \geq \gamma_1,$$

which completes the proof. ☐

Proof of Lemma 4.2.4.

Proof. To begin we consider using the moment–generating function

$$M_n(t', t) := E\left[\exp\left(t' \sum_{l=l_0}^{k} \sum_{s=1}^{\lfloor 2^l/r \rfloor} I_{(l,s)}^1 + t \sum_{l=l_0}^{k} \sum_{s=1}^{\lfloor 2^l/r \rfloor} I_{(l,s)}^2\right)\right], \quad t', t \in \mathbb{R},$$

where $I_{(l,s)} = (I_{(l,s)}^1, I_{(l,s)}^2)$. Now observe that our indicator function $I_{(l,s)}$ are i.i.d. bernoulli random vectors. Then from Lemma's 4.2.2 and 4.2.3 we have

$$\begin{aligned}
P[I_{(l,s)} = (1,1)] &= o\left(\frac{r}{2^{k+1}}\right) \\
P[I_{(l,s)} = (1,0)] &= \gamma_{(1,0)} \frac{r}{2^{k+1}} \\
P[I_{(l,s)} = (0,1)] &= \gamma_{(0,1)} \frac{r}{2^{k+1}} \\
P[I_{(l,s)} = (0,0)] &= 1 - (\gamma_{(1,0)} + \gamma_{(0,1)}) \frac{r}{2^{k+1}}.
\end{aligned}$$

Therefore,

$$\lim_{k\to\infty} E\left[\exp\left((t',t)\sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r\rfloor} I_{(l,s)}\right)\right] = \lim_{k\to\infty} E\left[\exp^{(t',t)I_{(l,s)}}\right]^{2^{k+1}/r}$$

$$= \lim_{k\to\infty}(e^{(0,0)}\cdot P[I_{(l,s)}=(0,0)] + e^{(t',0)}\cdot P[I_{(l,s)}=(1,0)]$$

$$+ e^{(0,t)}\cdot P[I_{(l,s)}=(0,1)] + e^{(t',t)}\cdot P[I_{(l,s)}=(1,1)])^{2^{k+1}/r}$$

$$= \lim_{k\to\infty}\left(1-(\gamma_{(1,0)}+\gamma_{(0,1)}-\gamma_{(1,0)}e^{t'}-\gamma_{(0,1)}e^{t})/2^{k+1}/r\right)^{2^{k+1}/r}$$

$$= e^{(\gamma_{(1,0)}(e^{t'}-1))}\cdot e^{(\gamma_{(0,1)}(e^{t}-1))}.$$

Setting $\lambda_1 = \gamma_{(1,0)}$ and $\lambda_2 = \gamma_{(0,1)}$ completes the second part of the proof, since the moment–generating function for our random vector $I_{(l,s)} = (I^1_{(l,s)}, I^2_{(l,s)})$ is equal to the product of the individual moment–generating functions which corresponds to a Poisson distribution. □

Proof of Theorem 4.2.1.

Proof. We begin with a partition of the index set $[1, 2^{k+1}]$ into two types of sub-intervals. That is,

$$[1, 2^{k+1}] = \bigcup_{l=l_0}^{k}\bigcup_{s=1}^{\lfloor 2^l/r\rfloor} Q_{l,s}\cup Q^c_{l,s},$$

where

$$Q^c_{l,s} = [1, 2^{l_0}-1]\cup\bigcup_{l=l_0}^{k}[2^l + r\lfloor 2^l/r\rfloor - 2^{m+1},\ldots,2^{l+1}-1]\cup\{2^{k+1}\}.$$

We now determine the cardinality of $Q_{l,s}^c$ denoted by $|Q_{l,s}^c|$ using the fact that $2^{l_0-1} < r \leq 2^{l_0}$, then

$$\begin{aligned}
|Q_{l,s}^c| &= 2^{l_0} - 1 + \sum_{l=l_0}^{k}(2^l - r\lfloor 2^l/r \rfloor) + \sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r \rfloor} 2^{m+1} + 2^{k+1} \\
&\leq 2^{l_0} + r(k - l_0 + 1) + \frac{2^{k+1}}{r}2^{m+1} + 2^{k+1} \\
&\leq r + r(k+1+1) + 2^{k+1}\left(\frac{2^{m+1}}{r} + 1\right) \\
&\leq r(k+3) + 2^{k+1}\left(\frac{2^{m+1}}{r} + 1\right) \\
&\leq 2\left(kr + 2^{k+1}\left(\frac{2^{m+1}}{r} + 1\right)\right).
\end{aligned}$$

The partition of the index set allows us to break the point process $\mathcal{I}_n^{(m)}$ into two parts. The first part deals with the situation when the index t falls into one of the $Q_{l,s}$ intervals, whereas the second part represents the situation when the index t falls into one of the $Q_{l,s}^c$ intervals. Thus, for $i = 1, 2$ we define

$$W_1^i = \sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r \rfloor} I_{(l,s)}^i$$

$$W_2^i = \sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r \rfloor}\sum_{t \in Q_{l,s}^c} \mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i]$$

Hence on the set $\{\sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r \rfloor} \tilde{I}_{(l,s)} = (0,0)\}$, we have

$$\begin{aligned}
\mathcal{I}_n^{(m)}(R_i) &= \sum_{t=1}^{2^k} \mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i] \\
&= \sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r \rfloor}\left(\sum_{t \in Q_{l,s}} \mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i] + \sum_{t \in Q_{l,s}^c} \mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i]\right) \\
&= W_1^i + W_2^i \quad \text{ for } i = 1, 2.
\end{aligned}$$

Now observe for k large and $i = 1, 2$ that

$$
\begin{aligned}
E[W_2^i] &= E\left[\sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r\rfloor}\sum_{t\in Q_{l,s}^c}\mathbb{1}[(Y_t, b_n^{-1}X_t^{(m)}) \in R_i]\right] \\
&\le |Q_{l,s}^c| P[Y_1, b_n^{-1}X_1^{*(m)} \in R_i] \\
&\le |Q_{l,s}^c| P^2[X_1^{*(m)} \ge b_n a_i^{'}] \\
&\le 2\left(kr + 2^{k+1}\left(\frac{2^{m+1}}{r} + 1\right)\right)(2^{-(k+1)})^2 a_i^{'-2\beta}\sum_{j=0}^{m}\phi^{(2\beta)j} \\
&= o_p(1).
\end{aligned}
$$

Applying Lemma 4.2.4 completes the proof since

$$
(\mathcal{I}_n^{(m)}(R_1), \mathcal{I}_n^{(m)}(R_2)) = \sum_{l=l_0}^{k}\sum_{s=1}^{\lfloor 2^l/r\rfloor} I_{(l,s)} + o_p(1) \Rightarrow Pois(\lambda_1) \times Pois(\lambda_2).
$$

□

Proof of Corollary 4.2.1.

Proof. First observe that

$$
P[b_n(\hat{\phi}_n - \phi) > x] = P\left[\bigwedge_{t=1}^{\lfloor n/2\rfloor}\left(\frac{Y_t}{b_n^{-1}X_t}\right) > x\right]. \tag{4.5.1}
$$

Now define a subset of $\mathbb{R}_+^2$ by $A_x = \{y_1, y_2 : y_1/y_2 \le x, y_1, y_2 > 0\}$. Then it suffices to show that there are no points t that satisfies the condition in A_x. Thus, if we let $y_1 = Y_t$ and $y_2 = b_n^{-1}X_t$, then notice that (4.5.1) is equivalent to $(\mathcal{I}_n(A_x) = 0)$. Furthermore, observe that A_x is a bounded set in $E = [\theta, \infty) \times (0, \infty]$ provided $\theta > 0$. Therefore, assuming $\phi > 0$ and applying Lemma 2.5 in Bartlett and McCormick (2012) we have that the point process $\mathcal{I}_n^{(m)} = \sum_{t=1}^{\lfloor n/2\rfloor}\varepsilon_{(Y_t, b_n^{-1}X_t^{(m)})}$ is equivalent to $\mathcal{I}_n = \sum_{t=1}^{\lfloor n/2\rfloor}\varepsilon_{(Y_t, b_n^{-1}X_t)}$. Hence using Theorem 4.2.1

and (4.5.1), we have that

$$
\begin{aligned}
\lim_{n\to\infty} P[b_n(\hat{\phi}_n - \phi) > x] &= \lim_{n\to\infty} P[\mathcal{I}_n(A_x) = 0] \\
&= \exp\left(-(1-\phi^\beta)^{-1} \int_\theta^\infty \left(\frac{y_1}{x}\right)^{-\beta} \bar{G}(y_1/x) dy_1 \right) \\
&= \exp\left(-(1-\phi^\beta)^{-1} x^\beta \int_\theta^\infty y_1^{-\beta} \bar{G}(y_1/x) dy_1 \right) \\
&= e^{-x^\beta E Y^{-\beta} (1-\phi^\beta)^{-1}}.
\end{aligned}
$$

□

Proof of Lemma 4.2.5.

Proof. Since $\alpha > \beta$, we have $\lim_{n\to\infty} a_n b_n = \infty$. Therefore since $\big(b_n(\hat{\phi}_n - \phi), n \geq 1\big)$ is a tight sequence by Corollary 4.2.1 and $\max_{t\in I_n} X_t/(a_n b_n)^\rho \leq 1$ with $0 < \rho < 1$, we have

$$a_n^{-1} |\hat{\phi}_n - \phi| \bigvee_{t\in I_n} X_t \xrightarrow{p} 0.$$

This completes the proof since

$$a_n^{-1} \Big| \hat{\theta}_n - \bigwedge_{t\in I_n} Y_t \Big| \leq a_n^{-1} |\hat{\phi}_n - \phi| \bigvee_{t\in I_n} X_t.$$

□

Proof of Theorem 4.2.2.

Proof. First observe from (4.2.7) we have

$$nP[a_n^{-1}(Y_1 - \theta) \in \cdot] \xrightarrow{v} \nu \quad \text{in} \quad E_2$$

where

$$\nu[0,y] = \lim_{n\to\infty} P[a_n^{-1}(Y_t-\theta)\le y] = \lim_{n\to\infty} n/2P[Y_1 \le (\theta + a_n y)]$$
$$= \lim_{n\to\infty} 1/2\frac{G(\theta+a_n y)}{G(\theta+a_n)}$$
$$= y^\alpha/2.$$

The result now follows from the fact that $\{Y_t, t = 1, \ldots, \lfloor n/2 \rfloor\}$ are i.i.d. random elements of $(E_2, \mathcal{B})$ where E_2 is locally compact, $\mathcal{B}$ is the Borel σ-algebra, and ν is a Radon measure on $(E_2, \mathcal{B})$. Therefore, by (proposition 3.21 in Resnick (1987)) we have

$$\mathcal{V}_n \Rightarrow \mathcal{V}.$$

□

Proof of Lemma 4.2.6.

Proof. As a result of Theorem 4.2 in Billingsley (1968) and the definition of the vague metric it suffices to show that for all $\eta > 0$ and $f \in C_n^+(E_3)$, $f \le 1$,

$$\limsup_{n\to\infty} P\left[\left|\sum_{t=1}^{\lfloor n/2\rfloor} f(a_n^{-1}(Y_t-\theta), c_n) - \sum_{t=1}^{\lfloor n/2\rfloor} f(a_n^{-1}(Y_t-\theta), 0)\right| > \eta\right] = 0. \qquad (4.5.2)$$

Now suppose the support of f is contained in the compact set $[0,s]\times[0,\infty]$. Since f is uniformly continuous, given $\epsilon > 0$ there exists a $\delta > 0$ s.t,

$$|f(z,x) - f(z,0)| < \epsilon, \text{ whenever } |x - 0| < \delta.$$

Let $b_n(\hat{\phi}_n - \phi) \Rightarrow U$, where U is an non-degenerate distribution. Now intersecting the event in (4.5.2) with the set

$$\bigcap_{j=1}^{\lfloor n/2\rfloor} \left(\{a_n^{-1}(Y_j-\theta) > s\} \cup \{c_n < \delta\}\right)$$

and its complement, the probability in (4.5.2) is bounded above by

$$
\begin{aligned}
&P\left[\bigcup_{j=1}^{\lfloor n/2\rfloor}\{(a_n^{-1}(Y_j-\theta)\le s)\cap(c_n\ge\delta)\}\right]+P\left[\epsilon\sum_{j=1}^{\lfloor n/2\rfloor}\varepsilon_{(a_n^{-1}(Y_j-\theta))}([0,s])>\eta\right]\\
&\le\frac{n}{2}P[a_n^{-1}(Y_1-\theta)\le s]P[c_n\ge\delta]+P\left[\sum_{j=1}^{\infty}\varepsilon_{(a_n^{-1}(Y_j-\theta))}([0,s])>\eta/\epsilon\right]\\
&\le\frac{n}{2}P[Y_1\le(\theta+a_ns)]P[b_n(\hat{\phi}_n-\phi)(a_nb_n)^{\rho}\ge a_nb_n\delta]+P\left[\sum_{j=1}^{\infty}\varepsilon_{(a_n^{-1}(Y_j-\theta))}([0,s])>\eta/\epsilon\right]\\
&\le\frac{1}{2}\frac{G(\theta+a_ns)}{G(\theta+a_n)}\bar{U}((a_nb_n)^{1-\rho}\delta)+P\left[\sum_{j=1}^{\infty}\varepsilon_{(a_n^{-1}(Y_j-\theta))}([0,s])>\eta/\epsilon\right]\\
&\to\frac{s^{\alpha}}{2}e^{-(((a_nb_n)^{1-\rho}\delta)^{\beta}EY_1^{-\beta}(1-\phi^{\beta})^{-1})}+P[\varepsilon_{([0,s]\times[0,\infty])}>\eta/\epsilon].
\end{aligned}
$$

Choosing $\epsilon>0$ completes the proof since this bound can be made arbitrarily small as $n\to\infty$. □

Proof of Corollary 4.2.2.

Proof. By Lemma 4.2.5 and (4.2.14) we have

$$
\lim_{n\to\infty}P[a_n^{-1}(\hat{\theta}-\theta)>y]=\lim_{n\to\infty}P[a_n^{-1}(\bigwedge_{t\in I_n}Y_t-\theta)>y]+o(1).
$$

Now observe that

$$
\begin{aligned}
0&\le P\Big[a_n^{-1}(\bigwedge_{t\in I_n}Y_t-\theta)>y\Big]-P\Big[a_n^{-1}(\bigwedge_{t=1}^{\lfloor n/2\rfloor}Y_t-\theta)>y\Big]\\
&=P\Big[a_n^{-1}(\bigwedge_{t=1}^{\lfloor n/2\rfloor}Y_t-\theta)\le y<a_n^{-1}(\bigwedge_{t\in I_n}Y_t-\theta)\Big]\\
&\le P\Big[\bigcup_{1\le t\le\lfloor n/2\rfloor}\big(X_t>(a_nb_n)^{\rho}\text{ and }a_n^{-1}(Y_t-\theta)\le y\big)\Big]\\
&\le nP[X_1^*>(a_nb_n)^{\rho}]P[Y_1\le a_ny+\theta]=o(1).
\end{aligned}
\tag{4.5.3}
$$

It then follows from (4.5.3) that

$$\lim_{n\to\infty} P\left[\bigwedge_{t\in I_n}^{\lfloor n/2\rfloor}\left(\frac{Y_t-\theta}{a_n}\right)>y\right]=\lim_{n\to\infty} P\left[\bigwedge_{t=1}^{\lfloor n/2\rfloor}\left(\frac{Y_t-\theta}{a_n}\right)>y\right]+o(1).$$

Now observe from Lemma 4.2.6 that the point process $\mathcal{V}_n$ suffices. Therefore, if we define the subset $B_y=\{z: z\le y, z>0\}$, then $\bigwedge_{t=1}^{\lfloor n/2\rfloor}\left(\frac{Y_t-\theta}{a_n}\right)>y$ is equivalent to $(\mathcal{V}_n(B_y)=0)$. The result now follows from Theorem 4.2.2 since

$$\begin{aligned}\lim_{n\to\infty} P[a_n^{-1}(\hat{\theta}_n-\theta)>y] &= \lim_{n\to\infty} P\left[\bigwedge_{t=1}^{\lfloor n/2\rfloor} a_n^{-1}(Y_t-\theta)>y\right]+o(1)\\ &= \lim_{n\to\infty} P[\mathcal{V}_n(B_y)=0]\\ &= P[\mathcal{V}(B_y)=0]=e^{-y^\alpha/2}.\end{aligned}$$

□

Proof of Theorem 4.3.1.

Proof. Let $Y_t=\epsilon_{2t}\wedge\epsilon_{2t+1}$ and define a sequence of point processes

$$\mathcal{N}_n=\sum_{t=1}^{\lfloor n/2\rfloor}\epsilon_{(d_n^{-1}Y_t,X_t)}.$$

Then since Y_t are i.i.d. with distribution G, we have that

$$\mathcal{N}_n\Rightarrow\eta, \tag{4.5.4}$$

where η is Poisson Random Measure with mean measure given by ν_2 satisfying for $x,\ y>0$

$$\nu_2([0,x]\times[0,y])=\lim_{n\to\infty}\frac{n}{2}P[d_n^{-1}Y_1\le x]P[X_1\le y]=\frac{1}{2}x^\alpha H(y).$$

Now define a subset $\mathbb{R}_+^2$ by $A_y = \{x_1, x_2 : x_1/x_2 \leq y, x_1, x_2 > 0\}$. Then observe that

$$(\mathcal{N}_n(A_y) = 0) = (d_n^{-1}(\hat{\phi}_n - \phi) > y). \tag{4.5.5}$$

Thus, by (4.5.4) and (4.5.5), we have that

$$\begin{aligned}
\lim_{n\to\infty} P[d_n^{-1}(\hat{\phi}_n - \phi) > y] &= \lim_{n\to\infty} P[\mathcal{N}_n(A_y) = 0] \\
&= P[\eta(A_y) = 0] \\
&= \exp\left(-1/2 \int_0^\infty \alpha x_1^{\alpha-1} \bar{H}(x_1/y) dx_1\right) \\
&= e^{-y^\alpha E X_1^\alpha / 2}.
\end{aligned}$$

□

Proof of Corollary 4.3.1.

Proof. This follows from Theorem 4.3.1 upon noting that $\bar{F}(x) = e^{-(\alpha+\beta)x}$ so that

$$G(x) = 2F(x) - F_1(x, x) \sim (2\alpha + \beta)x \quad \text{ as } x \to 0$$

and observing that $E[X_1] = ((1-\phi)(\alpha+\beta))^{-1}$. Thus,

$$\lim_{n\to\infty} P\left[\frac{n(2\alpha+\beta)}{2(\alpha+\beta)(1-\phi)}(\hat{\phi}_n - \phi) > y\right] = e^{-y}.$$

Plugging α and β into the above equation completes the proof. □

4.6 References

[1] Bartlett, A., McCormick W.P. (2012). Estimation for Non-negative First order Autoregressive Process with an Unknown Location Paramter. *Applied Mathematics*, **3 No. 12A**, 2133 - 2147.

[2] Bercu, B., Saporta DE, Gegout-Petit B. (2009). Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.,* **14**, 2492 - 2526.

[3] Billingsley P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[4] Cowan, R. Staudte, R. G. (1986). The bifurcating autoregressive model in cell lineage studies. *Bioetrics*, **42**, 769 - 783.

[5] Davis, R.A., McCormick, W.P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Process. Appl.* **31**, 237–250.

[6] Huggins, R.M., Basawa, I.V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36**, 1225 - 1233.

[7] Huggins, R.M., Basawa, I.V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42**, 423 - 432.

[8] Hwang, S.Y., Basawa, I.V. (2006). Local asymptotic normaility for bifurcating autoregressive processes and related asymptotic inference. *Statistical Methodology.* **6**, 61 - 69.

[9] Kallenberg, O. (1976). *Random Measures*, Akademie-Verlag, Berlin.

[10] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[11] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

[12] Staudte, R.G., Zhang, J., Huggins, R.M., Cowan, R (1996). A reexamination of the cell lineage data of E.O. Powell. *Biometrics*, **52**, 1214 - 1222.

[13] Zhang, C. (2011). Parameter Estimation for First-order Bifurcating Autoregressive Processes with Weibull Type Innovations. *Statistics and Probability Letters*, **81**, 1961 - 1969.

[14] Zhou, J. Basawa, I.V. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. Stochastic methods and their applications. *J.Appl. Probab.*, **41A**, 55 - 64.

[15] Zhou, J. Basawa, I.V. (2005a). Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74**, 77 - 88.

[16] Zhou, J. Basawa, I.V. (2005b). Maximum Likelihood Estimation for a First-Order Bifurcating Autoregressive Process with Exponential Errors. *J. Time Series Anal.*, **26**, 825 - 842.

Chapter 5

Conclusion

In this chapter we provide a few closing remarks regarding the performance of our estimators within various nonnegative time series where the innovation distribution is assumed to be regularly varying at both the lower endpoint θ and upper endpoint infinity.

5.1 Summary

In Chapter 2 we studied an estimation problem for the autoregressive parameter and the unknown location parameter of a nonnegative AR(1) model. Our estimators were obtained by taking the ratio of two sample values chosen with respect to an extreme value criteria for ϕ and taking the minimum of $X_t - \hat{\phi}_n X_{t-1}$ over the observed series for θ. In the case that $0 < \beta < 2$ our estimators, compared to Davis and Resnick's (1986) sample correlation estimator, performed extremely well in terms of bias, accuracy, and the average length of confidence intervals. However, when $\beta > 2$ our estimator was outperformed by Davis and Resnick's estimator. This result is not surprising as our procedure relies heavily upon extremely large innovations and when this does not occur, our estimators produce a small fraction of estimates with a high relative error. Additionally, the estimation procedure presented in this chapter lacks the ability to extend to nonnegative autoregressive models of higher order. Hence, further investigation towards a refinement of the estimation procedure studied in this chapter is required.

In Chapter 3 we studied an estimation problem for higher order time series models with heavy tail innovations. We first began by estimating the nonnegative coefficients in an infinite order moving average process. This was accomplished by considering the location of the maximum observation rather than minimizing the ratio of two sample values. The idea being, within the heavy-tail regime the maximal observation should be associated with the observation having the maximal innovation. Using estimates for the coefficients within an MA(∞) process, we could then obtain estimates for the parameters of nonnegative ARMA models by expressing them as a function of the estimated coefficients. Our extreme value estimators for nonnegative ARMA models converged at a rate faster than square root n and performs significantly better than Davis and Resnick's estimator when $0 < \beta < 2$. While the nonnegative restriction of innovations and coefficients may appear alarming, based upon our simulation results this does not appear to be an issue since our estimation procedure has the ability to accurately estimate negative parameters of various time series models. On a side note, the estimator defined in Chapter 2 performed slightly better in terms of bias and accuracy than our estimator presented in this chapter.

In Chapter 4, we applied the same estimator introduced in Chapter 2 to a nonnegative BAR(1) process. The limiting distributions obtained heuristically for our estimators turned out to be extremely similar to those found in Chapter 2. This was expected since the bifurcating process is typically used to model each line of descent in a binary tree as an AR(1) process. Additionally, the simulation results appear to suggest that our estimation procedure has the advantage of being resistent to the complex dependency that exists within this process. Hence, further investigation into rigourously proving this observation is required.

In conclusion, our estimation procedure is easy to implement and capitalizes on the behavior of extreme value estimators over traditional estimators when the regularly varying exponent is less than 2. Finally, our extreme value procedure provides an alternative and attractive method to obtain the limiting distribution for estimators in various nonnegative time series models that are both explicit and tractable, whereas the limiting distribution for traditional methods such as a maximum likelihood or linear programming are complex and unpractical.

5.2 Future Research

In this dissertation we have addressed various nonnegative time series models such as; AR(1), AR(p), MA(q), ARMA(p, q), and BAR(1). Within each model there are several interesting directions and questions that can be investigated in the future. First, we have defined $\hat{\theta}_n$ as a minimum over a restricted index I_n, but in the definition of I_n we have defined sequences a_n and b_n which are defined in terms of F, and so are unknown. Therefore, we need to determine estimators for a_n and b_n so that $\hat{\theta}_n$ is a statistic. Secondly, our models mentioned up to this point have a major limitation in common; they only capture short-range dependence. A model that can handle both short-range and long-range dependence simultaneously would be a natural stepping stone. Such a model that could be studied is known as a FARIMA (fractional autoregressive integrated moving average) model, where one does a fractional difference to the data and then models with an ARMA model. They are the natural generalizations of standard ARIMA (p, d, q) processes when the degree of differencing d is allowed to take nonintegeral values. For example, consider a process of the form if B denotes the backward shift operator an ARIMA(p, d, q) model and X_t is such that $Y_t = (1 - B)^d X_t$ then Y_t is an ARMA(p, q) model. This would bring us into the domain of non-stationary time series models which has hardly been studied from the extreme value prospective. Xue et al. (1999) presented an article on traffic tracing based on FARIMA models. A particular application of interest is to model traffic flow by fitting a FARIMA model.

Another application of particular interest is to apply an extreme value approach to estimate the location of change points within a time series model. When trying to perform regression on a time series for temperature, a change point is the single most important factor for obtaining accurate trend estimates. For performing the change point test, we consider starting with an ARMA(1,2) process and using an extreme value approach to estimate the location and number of change points on the Chula Vista, CA station data set. This data set contains 936 data points recorded monthly over the years 1919 - 1996. Additionally, we hope to perform a Bayesian analysis to compare performance between the two approaches. Our analysis will implement various Bayesian methods, which produce a probability distribution for the location and number of change points, where we treat the number and location of change points as unknown parameters to be estimated using a specified prior distribution. The advantage of treating the number and location of the change points as parameters allows for a more flexible and realistic model in which the model through the data will determine the number and location of the change points, instead of assuming that a change point exists and using an ad-hoc method by comparing Goodness of Fit statistics (GOF) for each time point. That is, we begin with a simple frequentist approach that computes the goodness of fit at each time point and the point that maximizes GOF is the change point. This method can only find the location of one change point, but can give us a starting place for our Bayesian methods. We will then try different Bayesian models that use prior information to find and locate change points. This includes expanding techniques to improve our posterior probabilities for a change point.

Bibliography

[1] Anděl, J. (1989). Non-negative autoregressive processes. *J. Time Ser. Anal.* **10**, 1-11.

[2] Anděl, J. (1991). Non-negative linear processes. *Appl. of Math.* **36**, 277-283.

[3] Anderson, P.L., Kavalieris, L., Meerschaert, M.M. (2008). Innovations algorithm asymptotics for periodically stationary time series with heavy tails. *J. Multivariate Anal.* **99**, 94-116.

[4] Bamberg, G. and Dorfleitner, G. (2001). *Fat Tails and Traditional Capital Market Theory, mineo*, University of Augsburg, Germany.

[5] Bartlett, A., McCormick W.P. (2012). Estimation for Non-negative First order Autoregressive Process with an Unknown Location Paramter. *Applied Mathematics*, **3 No. 12A**, 2133 - 2147.

[6] Bartlett A., McCormick W.P. (2013). Estimation for Nonnegative Time Series with Heavy-tail Innovations, *J. Time Series Anal*, **34**, 96 - 115.

[7] Billingsley P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[8] Bingham N.H., Goldie C.M., and Teugels J.L., (1987). *Regular Variation.* Cambridge University Press, Cambridge, UK.

[9] Brockwell, P.J., Davis, R.A. (1987). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

[10] Brown, T.C., Feigin, P.D., Pallant, D.L. (1996). Estimation for a class of positive nonlinear time series models. *Stochastic Process. Appl.* **63**, 139-152.

[11] Calder, M., Davis, R.A. (1998). In R.J. Adler, R.E. Feldman, M.S. Taqqu, editors, *A Practical Guide To Heavy Tails: Statistical Techniques and Applications* pages 159-176, Birkhäuser, Boston.

[12] Collings P.S. (1975). Dams with autoregressive inputs. *J. Appl. Probab.* **12**, 533 - 541.

[13] Cowan, R. Staudte, R. G. (1986). The bifurcating autoregressive model in cell lineage studies. *Bioetrics*, **42**, 769 - 783.

[14] Datta, S., McCormick, W.P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *J. Amer. Statist. Assoc.* **90**, 1289-1300.

[15] Datta, S., Mathew, G., McCormick, W.P. (1998). Nonlinear autoregression with positive innovations. *Austral. & New Zealand J. Statist.* **40**, 229-239.

[16] Davis, R.A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Process. Appl.* **63**, 75-95.

[17] Davis, R.A., Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann Probab.* **13**, 179-195.

[18] Davis, R.A., Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.

[19] Davis, R.A., McCormick, W.P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Process. Appl.* **31**, 237–250.

[20] Feigin P. and Resnick S.I., (1992). Estimation for autoregressive processes with positive innovations. *Stochastic Models*, **8**, 479 - 498.

[21] Feigin, P.D., Resnick, S.I. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51**, 135-165.

[22] Feigin, P.D., Kratz, M.F., Resnick, S.I. (1996). Parameter estimation for moving averages with positive innovations. *Ann. Appl. Probab.* **6**, 1157-1190.

[23] Feigin, P. and Resnick, S.I., (1999). Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes*, **1**, 391 - 422.

[24] Finkenstadt B. and Rootzen H., (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. CRC Press LLC, New York.

[25] Fisher R.A. and Tippett L.H.C., (1928). Limiting forms of the frequency distributions of largest or smallest member of a sample. *Proc. Cambridge Philos.Soc.*, **24**, 180 - 190.

[26] Gnedenko B.V., (1943). Sur la distribution limite du terme d' une serie aleatoire. *Ann. Math*, **44**, 423 - 453.

[27] Haan L. de, (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematisch Centrum, Amsterdam.

[28] Hall P. and Heyde C.C., (1980). *Martingale limit theory and its application*. Academic Press Inc., New York.

[29] Huggins, R.M., Basawa, I.V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36**, 1225 - 1233.

[30] Huggins, R.M., Basawa, I.V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42**, 423 - 432.

[31] Hutton, J. (1990). Non-negative time series models for dry river flow. *J.Appl. Probab.* **27**, 171–182.

[32] Kallenberg, O. (1976). *Random Measures*, Akademie-Verlag, Berlin.

[33] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

[34] McCormick, W.P., Mathew, G. (1989). Asymptotic results for an extreme value estimator of the autocorrelation coefficient for a first order autoregressive sequence. *In Lecture Notes in Statistics 51, Extreme Value Theory (eds Husler and Reiss)*, Springer-Verlag, Berlin, 166–180.

[35] McCormick W.P. and Mathew, G., (1993). Estimation for Nonnegative autoregressive Processes with an Unknown Location Parameter. *J. Time Series Anal.* **14**, 71–92.

[36] Markov, A.S. (2009). Estimation of the autoregressive parameter with infinite dispersion of noise. *Automation and Remote Control* **70**, 92-106.

[37] Mikosch, T., Samorodnitsky, G. (2000). The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Applied Probab.* **10**, 1025-1064.

[38] Raftery, A.E., (1980). Estimation eficace pour un processus autoregressif exponentiel a densite discontinue. *Pub. Inst. Statist.* **25**, 65 – 91.

[39] Resnick, S.I., (1986). Point Processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66 – 138.

[40] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[41] Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25**, 1805-1869.

[42] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

[43] Smith R.L., (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statist. Sc.*, **4**, 367 - 393.

[44] Tsai, H., Chan, K.S. (2006). A note on non-negative ARMA processes. *J. Time Ser. Anal.* **28**, 350-360.

[45] F. Xue, J. Liu, Y. Shu, L.Zhang, O. Yang, (1999). *Electrical and Computer Engineering.* IEEE Canadian Conference, **1**, 162 - 167.

[46] Zhang, C. (2011). Parameter Estimation for First-order Bifurcating Autoregressive Processes with Weibull Type Innovations. *Statistics and Probability Letters*, **81**, 1961 - 1969.

[47] Zhou, J. Basawa, I.V. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. Stochastic methods and their applications. *J.Appl. Probab.*, **41A**, 55 - 64.

[48] Zhou, J. Basawa, I.V. (2005a). Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74**, 77 - 88.

[49] Zhou, J. Basawa, I.V. (2005b). Maximum Likelihood Estimation for a First-Order Bifurcating Autoregressive Process with Exponential Errors. *J. Time Series Anal.*, **26**, 825 - 842.