Studies of Several Curious Probabilistic Phenomena: Unobservable Tail Exponents in Random Difference Equations, and Confusion Between Models of Long-Range Dependence and Changes in Regime

Changryong Baek

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2010

Approved by

Vladas Pipiras, advisor

Amarjit Budhiraja, reader

Chuanshu Ji, reader

Vidyadhar G. Kulkarni, reader

M. Ross Leadbetter, reader

ABSTRACT

CHANGRYONG BAEK: Studies of Several Curious Probabilistic Phenomena: Unobservable Tail Exponents in Random Difference Equations, and Confusion Between Models of Long-Range Dependence and Changes in Regime
(Under the direction of Vladas Pipiras)

The dissertation is centered on two research topics. The first topic concerns reduction of bias in estimation of tail exponents in random difference equations (RDEs). The bias is due to deviations from the exact power-law tail, which are quantified by proving a weaker form of the so-called second-order regular variation of distribution tails of RDEs. In particular, the latter suggests that the distribution tails of RDEs have an explicitly known second-order power-law term. By taking this second-order term into account, a number of successful bias-reduced tail exponent estimators are proposed and examined. The second topic concerns the confusion between long-range dependent (LRD) time series and several nonstationary alternatives, such as changes in local mean level superimposed by short-range dependent series. Exploratory and informal tools based on the so-called unbalanced Haar transformation are first suggested and examined to assess the adequacy of LRD models in capturing changes in local mean in real time series. Second, formal statistical procedures are proposed to distinguish between LRD and alternative models, based on estimation of LRD parameter in time series after removing changes in local mean level. Basic asymptotic properties of the tests are studied and applications to several real time series are also discussed.

ACKNOWLEDGEMENTS

I am truly indebted to my advisor, Professor Vladas Pipiras, for his guidance, encouragement and patience during my doctoral studies. It was my great fortune and privilege to work with him.

My sincere thanks go to the other members of the dissertation committee, Professors Amarjit Budhiraja, Chuanshu Ji, Vidyadhar G. Kulkarni and M. Ross Leadbetter, for their careful reading, insightful discussions and stimulating suggestions which significantly improved this dissertation.

I should thank again Professors M. Ross Leadbetter and Amarjit Budhiraja for their sincere and valuable help with the job market.

I am grateful to Professor Patrice Abry for his warm hospitality and support during my visit at ENS Lyon, and for raising questions which form the basis of part of this dissertation.

I thank Professor Richard L. Smith for his elucidating comments and suggestions.

I am also deeply indebted to Professor Edward Carlstein for his thorough guidance to become a better instructor and a number of suggestions on giving a presentation.

I am thankful to all STOR members - faculty, staff and fellow graduate students for wonderful time in Chapel Hill.

For my family side, I would like to express special thanks to my wife, Jihee Yum, for her dedication and endless love. Also, many thanks to my parents for their love and support. For our dear little one, Tate Gunnou Baek, we love you!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

The core of the dissertation consists of Chapters 2-5. Chapters 2 and 3 are related to heavy tail phenomenon, and Chapters 4 and 5 contribute to a better understanding of long-range dependent time series. Each chapter is written and could be read somewhat separately from the others. Here, we shall provide a softer introduction to Chapters 2-5, explain our motivation and discuss our results.

The work concerning heavy tail phenomenon (Chapters 2 and 3) was motivated by extremal behavior of multiplicative cascades (MCs) which are widely used in modeling of physical phenomena (such as turbulence, rain clouds). Briefly, an MC is a limiting random measure defined by multiplying positive random variables in a binary tree structure. Letting $\lambda_\infty(0,1]$ be the MC measure on a unit interval, it is known (Kahane and Peyrière (1976), Guivarc'h (1990)) that, under mild assumptions, this measure has a power-law tail in the sense that

$$P(\lambda_\infty(0,1] > x) \sim cx^{-\alpha}, \quad \text{as } x \to \infty, \tag{1.1}$$

for some positive α, which is called a tail exponent. We were interested in estimation of the tail exponent α in MCs, especially when α is larger as suggested by real life data. To our great puzzlement, classical estimators of the tail exponent, such as the Hill estimator (see, for example, de Haan and Ferreira (2006), Embrechts, Klüppelberg and Mikosch (1997), Leadbetter, Lindgren and Rootzén (1983)), never found the true α however large a sample size we took!

In trying to understand what was going on, we were naturally led to RDEs. In one dimension, RDE in a stationary solution form is given by

$$X \stackrel{d}{=} AX + B, \tag{1.2}$$

where (A, B) is typically assumed to be independent of X. MCs turn out to be one example of RDEs. Another, and probably a better known example is that of the squares of ARCH(1) models used in Finance. According to a celebrated result of Kesten (1973), under mild assumptions, the tail distribution of an RDE X also has the power-law tail

$$P(X > x) \sim cx^{-\alpha}, \tag{1.3}$$

where $\alpha > 0$ satisfies $EA^{\alpha} = 1$. The tail behavior (1.1) for MCs is a (non-trivial) consequence of Kesten's result (1.3). Focusing on RDEs, we found that the same phenomenon seemed characteristic to all RDEs, not just MCs. In other words, classical estimators of tail exponents were also extremely biased for larger values of tail exponents of RDEs.

We began to suspect that the bias problem might be due to deviations from the exact power law in the tail regions of practical interest. A natural framework to examine these issues is that of the so-called second-order regular variation or 2RV in short (see, for example, de Haan and Ferreira (2006)). The distribution tail $P(X > x)$ is 2RV with first-order parameter $\alpha > 0$ and second-order parameter $\rho < 0$, if for suitable $g(x)$ and any $a > 0$,

$$\lim_{x\to\infty} \frac{x^{\alpha}P(X > x) - (ax)^{\alpha}P(X > ax)}{g(x)} = \frac{a^{\rho} - 1}{\rho}. \tag{1.4}$$

Alternatively, the second-order regular variation can be thought as

$$P(X > x) = c_1 x^{-\alpha} + c_2 x^{-\alpha+\rho} + o(x^{-\alpha+\rho}), \tag{1.5}$$

where the second term $c_2 x^{-\alpha+\rho}$ is responsible for deviations from the exact power-law behavior.

Proving 2RV for RDEs appears an open and difficult problem. Nevertheless, in Chapter 2, we are able to show a weaker form of the second-order regular variation of distribution tails of

RDEs, namely,

$$\int_x^\infty \Big(P(X>u) - P(AX>u)\Big)du \sim Cx^{-\alpha}. \tag{1.6}$$

Relation (1.6) suggests that

$$P(X>x) - P(AX>x) \sim C\alpha x^{-\alpha-1} \tag{1.7}$$

or

$$\lim_{x\to\infty} \frac{x^\alpha P(X>x) - x^\alpha (EA^\alpha)^{-1} P(X>A^{-1}x)}{x^{-1}} = C\alpha \tag{1.8}$$

which can be viewed as 2RV at random $a = 1/A$. From a more practical perspective, we show that (1.5) is consistent with (1.6) only when $\rho = -1$, suggesting that, for RDEs X,

$$P(X>x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1} + o(x^{-\alpha-1}). \tag{1.9}$$

In other words, deviations from the exact power-law tail occur through the second-order term with *known* parameter $\rho = -1$.

With established deviations from the exact power-law tail in RDEs, the next natural question is how to remove the resulting bias. Various authors proposed bias-reduced estimators under the assumption (1.5) with *unknown* ρ (see, for example, Feuerverger and Hall (1999), Beirlant, Dierckx, Goegebeur and Matthys (1999), Gomes and Martins (2002)). In Chapter 3, we examine a number of possible bias-reduced tail exponent estimators with explicitly known second-order parameter ρ. Estimators based on least squares, generalized jackknife and conditional maximum likelihood are considered and their basic asymptotics are established.

As an example of least squares method, observe that taking the log transformation in (1.9) gives

$$\log P(X>x) \approx \log c_1 - \alpha \log x + c_2/c_1 x^{-1}.$$

Then, replacing x by ith upper order statistics $X_{(n-i+1)}$ gives

$$\log(i/n) \approx \log c_1 - \alpha \log X_{(n-i+1)} + c_2/c_1 X_{(n-i+1)}^{-1},$$

and hence least squares methods can be used to estimate the tail exponent α. We call the resulting estimator *rank-based*. Another least squares-based estimator, called the QQ-estimator, is obtained by reversing the roles of $\log P(X > x)$ and $\log x$.

Conditional maximum likelihood estimators are based on the exact form of the distribution tail $P(X > x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1}$. In contrast to other maximum likelihood estimators, for example, Feuerverger and Hall (1999) and Smith (1987), our estimator overcomes numerical instability issues raised by many authors while keeping the same asymptotic efficiency. Monte Carlo simulations study shows that our proposed estimators successfully provide bias correction. In particular, rank-based estimator performs best and is also very simple to calculate.

The other part of the dissertation (Chapters 4 and 5) concerns long-range dependence, and distinguishing it from some nonstationary alternatives. Long-range dependent (LRD) time series models are commonly defined as weakly stationary time series with power-law decaying autocorrelation function for large time lags, namely,

$$\mathrm{Corr}(X_0, X_k) = L(k)k^{2d-1}, \quad d \in (0, .5), \tag{1.10}$$

as $k \to \infty$, for a slowly varying function L at infinity. The parameter d appearing in (4.2) is called the LRD parameter. LRD models capture persistent dependence between observations and are used in a wide range of areas such as Hydrology, Physics, Economics and Finance, and Internet traffic modelling. Extensive reviews on LRD can be found in Doukhan, Oppenheim and Taqqu (2003), Park and Willinger (2000), and Robinson (2003).

The basic question about LRD studied in the dissertation can be described as follows. One of the main features of LRD series is that they exhibit apparent changes in local mean level over a wide range of larger scales. Artificially removing several of these most significant changes in local mean level would immediately make the autocorrelation function of the remaining residuals decay to zero more rapidly. From a different angle, taking a short-range dependent or SRD series (with fast decaying correlations) and superimposing several deterministic changes in local mean level would lead to slowly decaying correlations, suggesting mistakenly that LRD is present.

This confusion between LRD and nonstationary models (such as SRD series superimposed by deterministic changes in local mean level) have caused and still causes much debate in all areas where LRD models are used. See, for example, Klemeš (1974) in Hydrology, Veres and Boda (2000) in Internet traffic modeling, Granger and Hyung (1999), Diebold and Inoue (2001), Mikosch and Stărică (2004), Smith (2005) in Economics and Finance. More recently, a number of statistical tests were proposed with the aim of distinguishing between LRD and some of the nonstationary models. See, for example, Berkes, Horváth, Kokoszka and Shao (2006), Jach and Kokoszka (2008), Ohanissian, Russell and Tsay (2008), Bisaglia and Gerolimetto (2009), Qu (2009).

Chapters 4 and 5 contribute to a better understanding of the basic problem above. In Chapter 4, we take the following more elementary and applied approach. If the main goal is to model the data series exhibiting changes in local mean level, then a natural first question is whether LRD models do a good job in capturing the characteristics of an observed series when looking through the prism of changes in local mean level.

A natural tool to study changes in local mean level is the unbalanced Haar wavelet transformation (UHT) (see, for example, Fryzlewicz (2007)). For a given series, UHT produces a collection of UHT detail coefficients and break points which capture best, according to a particular criterion used, changes in local mean level through their sizes and locations, respectively, at different time scales. From a different angle, having the sequence of all UHT detail coefficients and break points, one can reconstruct the original time series and hence this sequence is ideally suited to study the characteristics of changes in local mean level.

With UHT detail coefficients and break points, the idea is now simple. First, examine some characteristics of these coefficients and break points on LRD models. Then, see how these characteristics compare to those on a number of real time series exhibiting LRD features. More precisely, we focus on times between break points and the respective local mean levels (related to the UHT coefficients) at a particular resolution level, where the latter is obtained after the procedure referred to as denoising. For LRD models, we find empirically that these times are nearly independent and exponentially distributed, and that the series of local mean levels is the only part inheriting LRD features. We then compare these findings for LRD models with those on a number of real time series considered in Chapter 4, with varying conclusions on how well LRD models capture changes in local mean level.

In contrast to Chapter 4, Chapter 5 concerns formal statistical testing to distinguish between LRD and nonstationary models. These tests are generally divided into two types. First, there are the tests where LRD appears as the null hypothesis. See, for example, Dolado, Gonzalo and Mayoral (2005), Shimotsu (2006), Ohanissian et al. (2008), Qu (2009). Second, there are the tests where the null hypothesis consists of a nonstationary model. Our work falls in the latter category and takes as starting point the test developed in Berkes et al. (2006), BHKS test in short, where the null consists of SRD series superimposed by deterministic changes in local mean level.

We argue that the BHKS test suffers from a very low power against LRD alternatives, and explain this through the fact that the BHKS test statistic involves estimation of the sum of covariances of the underlying series. We argue that such estimation would not be necessary if a suitable regression procedure were used instead, and that the resulting test would already considerably improve the power. The regression procedure is in the spirit of standard estimation methods for LRD parameter, applied to the series of residuals obtained after removing changes in local mean level, that is, applied to

$$R_i = X_i - \widehat{X}_i,$$

where X_i is the original time series and $\widehat{X}_i$ is the series of estimated local mean levels. The LRD estimation method itself resembles the popular R/S estimation method (see, for example, Beran (1994)).

The R/S method is not the best available method to estimate LRD parameter. For example, the well-known and widely studied Geweke and Porter-Hudak (GPH; after Geweke and Porter-Hudak (1983)) and local Whittle estimators (Robinson (1995*a*)) perform much better. This naturally suggests introducing tests similar to BHKS but based on these estimation methods. A simulation study shows that the resulting tests have better power than when using R/S-like estimation in the spirit of BHKS and the BHKS test itself. In Chapter 5, we also establish theoretical properties of the tests based on GPH and local Whittle estimators. The corresponding proofs make one of the more technical parts of the dissertation. The tests are also applied to several real data sets.

CHAPTER 2

Second-order properties of distribution tails of random difference equations

2.1 Introduction

We are interested in tail exponents of random difference equations (RDEs, in short), also known as random recurrence equations, autoregressive models with random coefficients. In one dimension, RDE is given by

$$X_n = A_n X_{n-1} + B_n, \quad n \geq 1, \tag{2.1}$$

where (A_n, B_n) are typically assumed to be i.i.d. vectors and X_0 is some starting position. Important examples of RDEs include autoregressive conditionally heteroscedastic (ARCH) processes used in Finance or multiplicative cascades of interest in Physics. Several examples are introduced in detail in Section 2.2. We shall focus throughout on one dimensional RDE (2.1) though the multidimensional case will also be considered (see Section 2.4 below).

Under mild assumptions, the series $\{X_n\}$ in (2.1) has a stationary solution X satisfying the equation (which we also call RDE)

$$X \stackrel{d}{=} AX + B, \tag{2.2}$$

where $(A, B) \stackrel{d}{=} (A_1, B_1)$ is independent of X, and the tail distribution of X has a power-law tail. This result was first shown by Kesten (1973) and studied further by many authors, for example, Goldie (1991), Grintsyavichyus (1981) to name a few. It is stated in the following theorem (analogous result for the multidimensional case is given in Theorem 2.4.1).

Theorem 2.1.1. *(Kesten (1973), Theorem 5) Let $\{X_n\}_{n\geq 1}$ be defined by (2.1). Suppose that*

(A_n, B_n), $n \geq 1$, are i.i.d. random vectors such that

$$E \log |A_1| < 0, \tag{2.3}$$

and that, for some $\alpha > 0$,

$$E|A_1|^\alpha = 1, \tag{2.4}$$

$$E|A_1|^\alpha \log^+ |A_1| < \infty, \quad 0 < E|B_1|^\alpha < \infty. \tag{2.5}$$

If, in addition, $\log |A_1|$ does not have a lattice distribution and B_1 is not a constant times $(1 - X_1)$, then

$$X_n \stackrel{d}{\to} X, \tag{2.6}$$

where

$$X \stackrel{d}{=} \sum_{k=1}^{\infty} A_1 \dots A_{k-1} B_k, \tag{2.7}$$

and the series on the right-hand side of (2.7) converges a.s. Moreover,

$$P(X < -x) \sim c_- x^{-\alpha} \text{ and } P(X > x) \sim c_+ x^{-\alpha}, \quad \text{as } x \to \infty, \tag{2.8}$$

where at least one of c_- and c_+ is nonzero.

We are interested in questions concerning estimation of the tail exponent α appearing in (2.8). A common estimation method is based on a Hill estimator (see, for example, Embrechts, Klüppelberg and Mikosch (1997)). If $Y_1, \dots, Y_n$ are n given observations with a common distribution of Y (independent or not, depending on the context) and

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

are the observations in the increasing order, the Hill estimator $\widehat{\alpha}_H$ is defined as

$$\widehat{\alpha}_H^{-1} = \frac{1}{k} \sum_{i=1}^{k} (\log Y_{(n-i+1)} - \log Y_{(n-k)}), \tag{2.9}$$

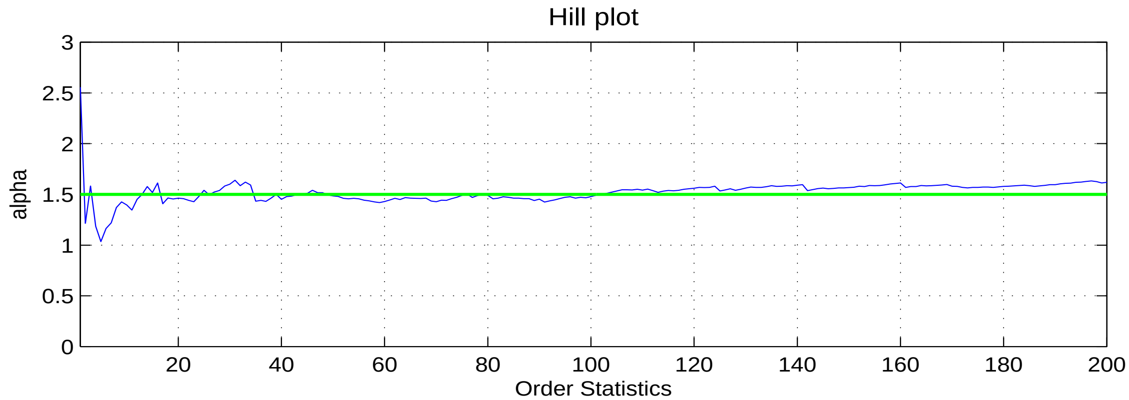

Figure 2.1: Hill estimator from Pareto distribution with $\alpha = 1.5$.

where k is a threshold. If the underlying distribution of Y has a power-law tail

$$P(Y > y) \sim cy^{-\alpha}, \qquad \text{as } y \to \infty,\ \alpha > 0, \tag{2.10}$$

the Hill estimator (2.9) of α is known to have nice theoretical properties such as consistency, asymptotic normality under fairly mild assumptions. In practice, the presence of heavy tails is assessed by examining the so-called Hill plot. This plot is produced by plotting $\widehat{\alpha}_H$ as a function of threshold k (from the smallest k to larger k). An example is given in Figure 2.1 based on i.i.d. observations from the Pareto distribution with $\alpha = 1.5$. If the distribution tail has power law, as in Figure 2.1, the Hill plot levels off in a region of small k and that level is taken as the Hill estimate for the power-tail exponent.

Statement of the problem. When tail exponent α is *large*, Hill plots for RDEs show tail exponent estimates surprisingly biased. For example, Figure 2.2 shows a typical Hill plot for an ARCH(1) model (see Section 2.2.1) having tail exponent $\alpha = 10$, based on 5,000 independent realizations of the process at a chosen, fixed time. Observe from the figure how far the Hill plot is from the true value of α. Perhaps even more surprising is that estimation improves only slightly by taking any reasonably larger sample size. For example, Figure 2.3 also shows the Hill plot for a million independent realizations. The basic goal of this Chapter is to understand why above estimation of tail exponents fails in RDEs for larger values of exponents and how this can be remedied.

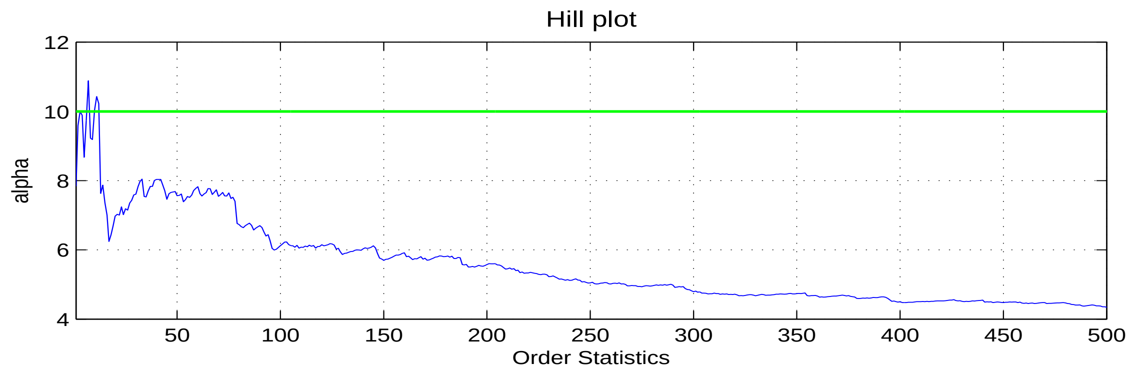

Figure 2.2: Hill plot for 5000 independent realizations of ARCH(1) model with $\alpha = 10$.

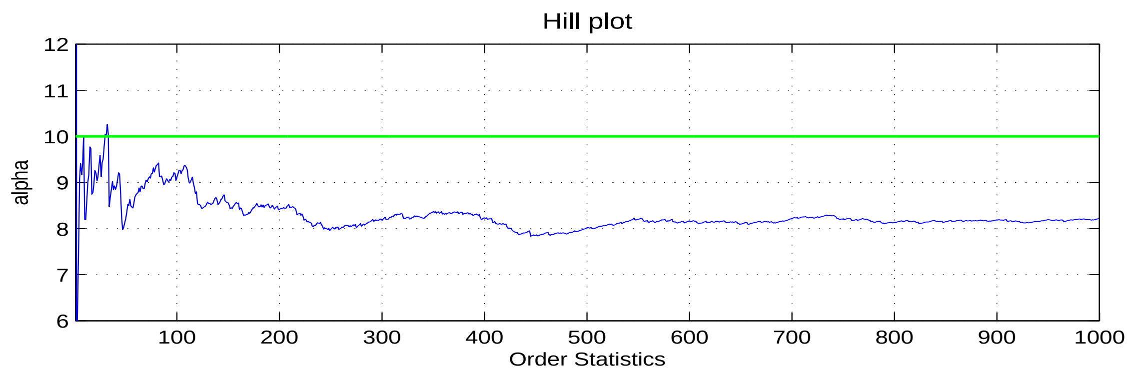

Figure 2.3: Hill plot for a million independent realizations of ARCH(1) model with $\alpha = 10$.

Here are some further important comments about the above problem.

- *Why do we consider RDEs?* The problem described above seems to be characteristic to all RDEs. We illustrate this in Section 2.2 through simulations in a number of different RDE models. One of our main goals is also to explain this in theory.

- *Why should one care about larger tail exponent?* The problem described above and supporting simulations involve larger values of tail exponent α. It is important to ask then why one should care about larger α. Several points should be made in this regard. First, in some applications of RDE models, larger values of α are, in fact, expected. This is the case, for example, with multiplicative cascades and other so-called multifractal models of interest in Physics. Second, observed bias in estimation of α becomes larger with increasing α and is still present (though smaller) for smaller α. If one believes that RDE models are appropriate for data at hand, this should be taken into account for either larger or smaller tail exponent α. Moreover, it would be desirable to have an estimation method that takes into account the possibility of bias, and which performs well for both larger and smaller values of α. Proposed estimation methods will be discussed in Chapter 3.

- *Independent versus dependent observations.* Two types of observations can be considered in regard to the problem stated above. First, one may suppose that observations are obtained from the RDE (2.1) and hence dependent in time. Second, one may suppose given independent copies of X_N for large fixed N (which can be thought as independent copies of the stationary solution X). For simplicity, we shall focus throughout on the second case. Perhaps surprising but this case is also relevant in practice (for example, in the context of multiplicative cascades) and the problem stated above is as equally relevant. Moreover, for dependent data given by (2.1), tail exponent estimation problems are known to get only worse. See Section 3.5.2 for a related discussion.

Possible explanations for the problem. Since we have removed temporal dependence from observations, two explanations seem plausible for the above problem:

1. Convergence to stationary solution (2.2) is so slow that the observation X_N is still far from the stationary solution X.

2. The result (2.8) is asymptotic in nature. It can happen that the region where (2.8) actually happens is too far in the tail to be observed for practical purposes. In other words, even with a huge number of data points, there are significant deviations from the exact power-law tail in practice.

In fact, in the context of RDEs, using their Markov structure, one expects that underlying measure P_N induced by RDE (2.1) converges to its invariant measure P_∞ induced by (2.2) at a geometrically fast rate. The latter fact is known as geometric ergodicity. Basrak, Davis and Mikosch (2002*b*) and Stelzer (2009) show geometric ergodicity for RDEs $\{X_n\}$. We also summarize their result in Theorem A.0.1 in Appendix A to the reader's convenience.

Theoretical properties of tail distribution. We therefore suspect that the tail exponent is not observed because of the second explanation above. As clearly pointed out in Resnick

(1997), all nice theoretical properties of Hill and other related estimators are valid only when the underlying distribution is close to power-law distribution. If the underlying distribution deviates from power-law distribution, bias is inevitable.

There is a general theoretical framework, called second-order regular variation or 2RV, in short (see, for example, de Haan and Ferreira (2006)), that allows one to study bias in first-order regular variation such as (2.8). The tail distribution $\overline{F}(x) = P(X > x)$ is second-order regularly varying with first-order parameter $\alpha > 0$ and second-order parameter $\rho < 0$ (denoted as $\overline{F} \in 2RV(-\alpha, \rho)$) if there exists a function $G(x) \to 0$ as $x \to \infty$ which ultimately has constant sign such that, for any $a > 0$,

$$\lim_{x\to\infty} \frac{\frac{\overline{F}(ax)}{\overline{F}(x)} - a^{-\alpha}}{G(x)} = ca^{-\alpha}\frac{a^{\rho}-1}{\rho}, \tag{2.11}$$

for some constant $c \neq 0$. Second-order regular variation can be thought as

$$\overline{F}(x) - c_1 x^{-\alpha} \sim l_2(x) x^{-\alpha+\rho}, \tag{2.12}$$

where $l_2(x)$ is a slowly varying function at infinity (Bingham, Goldie and Teugels (1989), Theorem 3.6.6, p. 158). The right-hand side of (2.12) is thought as a bias. For practical (estimation) questions, the slowly varying function in (2.12) is taken as

$$l_2(x) = c_2$$

for some constant c_2, that is,

$$\overline{F}(x) - c_1 x^{-\alpha} \sim c_2 x^{-\alpha+\rho}. \tag{2.13}$$

Remark. It is important to note that 2RV is asymptotic in nature. Even if proved for RDEs, it does not yield the exact region where (2.12) holds. Hence, without further analysis, establishing 2RV, in principle, does not completely address the problem raised in this chapter. Despite these limitations, 2RV at least indicates that there exists a bias and that it should be taken into account, for example, in questions of estimation.

To the best of our knowledge and understanding, 2RV is still an open and difficult problem for any larger class of RDEs. Instead of trying to prove 2RV, we shall focus on its weaker forms by considering the asymptotics of

$$P(X > x) - P(AX > x), \tag{2.14}$$

where A is the multiplier appearing in (2.2), and more specifically that of

$$\int_x^\infty (P(X > u) - P(AX > u))\, du. \tag{2.15}$$

We will show under mild assumptions that, as $x \to \infty$,

$$\int_x^\infty (P(X > u) - P(AX > u))\, du \sim c x^{-\alpha}, \tag{2.16}$$

which also suggests that, as $x \to \infty$,

$$P(X > x) - P(AX > x) \sim c\alpha x^{-\alpha-1}. \tag{2.17}$$

The expressions (2.14) and (2.15) are much easier to consider than (2.11) because AX can be related back to X by using the RDE structure (2.2). In fact, as seen from Section 2.3, (2.16) follows just by using the RDE structure (2.2) and the Kesten's result itself. A particularly simple case of RDE is considered in the beginning of Section 2.3.

How is (2.17) related to 2RV in (2.11)? Observe that with $g(x) = -x^\alpha G(x)\overline{F}(x)$, (2.11) can be rewritten as

$$\lim_{x\to\infty} \frac{x^\alpha P(X > x) - (ax)^\alpha P(X > ax)}{g(x)} = c\frac{a^\rho - 1}{\rho}. \tag{2.18}$$

Relation (2.17), on the other hand, can be expressed as

$$\lim_{x\to\infty} \frac{x^\alpha P(X > x) - x^\alpha (EA^\alpha)^{-1} P(X > A^{-1}x)}{x^{-1}} = c\alpha, \tag{2.19}$$

since $EA^\alpha = 1$ by (2.4). Hence, (2.19) can be viewed as 2RV (2.18) at random $a = 1/A$.

From a practical perspective, relation (2.17) says that there is a bias in (2.8) (if there is no bias in (2.8), then $P(X > x) - P(AX > x) = 0$). Moreover, if one believes that $\overline{F}(x)$ satisfies

(2.13), then necessarily $\rho = -1$ and

$$P(X > x) - c_1 x^{-\alpha} \sim c_2 x^{-\alpha-1}, \tag{2.20}$$

as $x \to \infty$ (Proposition 2.3.1 below).

Discussion on estimation. If one believes in RDE model and that the model has 2RV, it is natural to estimate tail exponent by taking 2RV into account. Tail exponent estimation based on 2RV has been studied by a number of authors. In particular, Peng (1998) shows asymptotic bias of Hill estimator under 2RV and proposes linear estimator considering second order parameter to adjust for the asymptotic bias. More recently, Gomes, de Hann and Rodrigues (2008) consider weighted Hill estimator where the weights are determined by 2RV parameters. Feuerverger and Hall (1999) utilize normalized log-spacings of order statistics

$$i(\log Y_{(n-i+1)} - \log Y_{(n-i)}),$$

which are known to follow Exponential distribution with mean 1 by Rényi's representation theorem for order statistics. Under the relation (2.13), these authors derive the maximum likelihood estimators of parameters α, ρ. Estimation questions under the framework (2.20) are extensively studied in Chapter 3. Several least squares estimators, generalizing rank-based and QQ-estimators, and conditional maximum likelihood estimators, based on (2.20), are introduced and their basic asymptotics are established.

The rest of Chapter 2 is organized in the following way. Several examples of RDEs and some simulation study with unobservable exponents for RDEs can be found in Section 2.2. In Section 2.3, we prove the weaker form (2.16) of second-order regular variation in RDEs. A multidimensional extension is discussed in Section 2.4.

2.2 Examples of RDEs and simulation study

In this section, several examples of one dimensional RDEs (2.1) are given. These include autoregressive conditionally heteroscedastic processes of order 1, an example with an explicit stationary distribution and multiplicative cascades. We also report here a further simulation study supporting the statement of the problem discussed in Section 2.1. The simulations are to show that the problem seems prevalent for all RDEs.

2.2.1 ARCH(1) models

A particular example of RDEs is a popular autoregressive conditionally heteroscedastic (ARCH(1)) model of order 1, defined by

$$\xi_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \beta + \eta \xi_{t-1}^2, \tag{2.21}$$

where $\{\epsilon_t\}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables and coefficients β and η are strictly positive. The squares of ARCH(1) series can be written as

$$\xi_t^2 = \eta \epsilon_t^2 \xi_{t-1}^2 + \beta \epsilon_t^2, \tag{2.22}$$

which is the RDE (2.1) with

$$X_t = \xi_t^2, \quad A_t = \eta \epsilon_t^2, \quad B_t = \beta \epsilon_t^2. \tag{2.23}$$

By Theorem 2.1.1, the distribution ξ^2 has the tail exponent α satisfying

$$\Gamma(\alpha + 1/2) = \sqrt{\pi}(2\sigma^2\eta)^{-\alpha}. \tag{2.24}$$

The equation (2.24) does not have a closed-form solution. For example, if $\sigma^2 = 1$, numerical calculations yield:

α	2	3	4	5	6	7	8	9	10
η	.577	.406	.312	.254	.214	.185	.163	.145	.105.

Note also that, by symmetry, the tail exponent of ξ is

$$\alpha_\xi = 2\alpha \tag{2.25}$$

because

$$P(\xi > x) = 1/2 P(\xi^2 > x^2) \sim \frac{c}{2} x^{-2\alpha}. \tag{2.26}$$

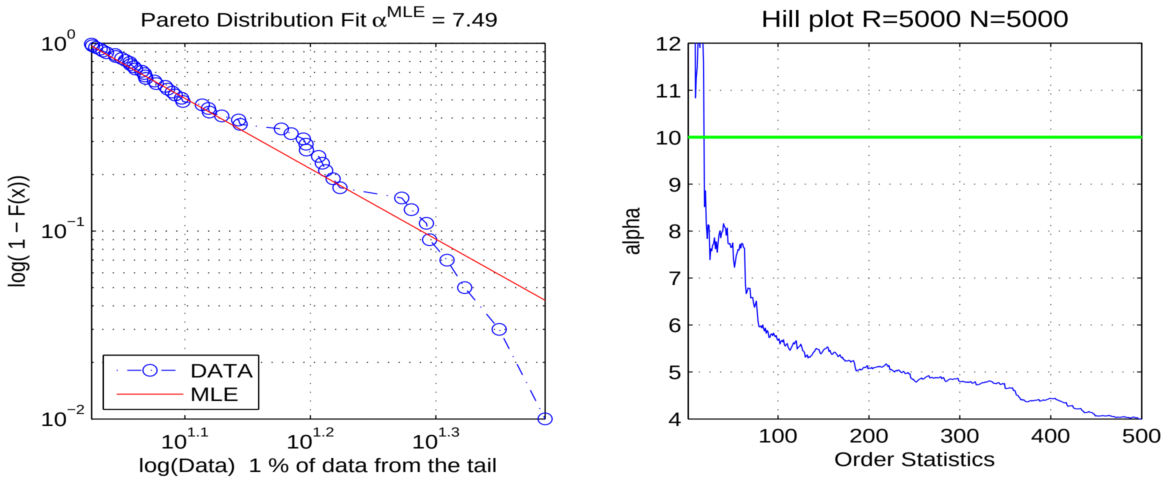

Figure 2.4: ARCH(1) with $\alpha_\xi = 10$.

Consider the ARCH(1) series (2.21) with $\eta = .254$, $\beta = 1$, $\epsilon_t \overset{d}{=} \mathcal{N}(0,1)$. From the above table, the tail exponent is

$$\alpha_\xi = 10.$$

In this simulation, we generated $R = 5,000$ independent observations with $N = 5,000$ iterations of ARCH(1) series. Figure 2.4 shows tail exponent estimation, and again it does not find the true tail exponent $\alpha_\xi = 10$. (Pareto distribution used in Figure 2.4 has an exact power-law tail.)

Remark. Empirical observations for ARCH(1) models similar to those above can also be found in Beirlant et al. (1999) (see Figure 12 on p. 195). Though this was the only place in the literature that we found to make such observations.

2.2.2 Examples of RDEs with explicit power-tail distributions

The following appears to be the only known family of RDEs for which a stationary solution has a power-tail distribution in closed form. Consider the so-called beta prime distribution $\beta(a, b)$ given by the density

$$\frac{1}{B(a,b)} x^{a-1}(1+x)^{-a-b} 1_{\{x>0\}}, \quad a, b > 0. \tag{2.27}$$

where $B(a,b)=\int_0^1 u^{a-1}(1-u)^{b-1}du$ is the beta function. Simple calculations shows that

$$\beta(a,b) \stackrel{d}{=} \frac{1}{Z} - 1, \tag{2.28}$$

where $Z \stackrel{d}{=} B(b,a)$ follows the beta distribution with the density

$$\frac{1}{B(b,a)} u^{b-1}(1-u)^{a-1}1_{(0,1)}(u).$$

Now fix $k \in \mathbb{N}$ and let $a_1, \ldots, a_k, b$ be positive reals. Denote $a_{k+1} = a_1$ and set

$$A = Y_1 \ldots Y_k, \quad B = Y_1 \ldots Y_k + \ldots + Y_{k-1}Y_k + Y_k, \tag{2.29}$$

where $Y_j \stackrel{d}{=} \beta(a_{j+1}, a_j + b)$ for $j = 1, \ldots, k$. Then, as shown in Goldie (1991), Chamayou and Letac (1991),

$$X \stackrel{d}{=} \beta(a_1, b) \tag{2.30}$$

satisfies RDE (2.2) with A, B in (2.29).

Observe that the distribution tail of random variable X in (2.30) satisfies

$$\begin{aligned} P(X > x) &= \int_x^\infty \frac{1}{B(a_1,b)} u^{a_1-1}(1+u)^{-a_1-b}du \\ &= \frac{1}{B(a_1,b)} \int_x^\infty u^{-b-1}\left(1+\frac{1}{u}\right)^{-a_1-b} du \\ &= \frac{1}{B(a_1,b)} \int_x^\infty u^{-b-1}\left(1-(a_1+b)u^{-1}+o(u^{-1})\right) du. \end{aligned} \tag{2.31}$$

Therefore, we have

$$P(X > x) - \frac{1}{bB(a_1,b)}x^{-b} \sim \frac{-(a_1+b)}{B(a_1,b)} \int_x^\infty u^{-b-2}du = \frac{-(a_1+b)}{B(a_1,b)(b+1)}x^{-b-1}, \tag{2.32}$$

as $x \to \infty$, that is, the tail exponent for X is b and the second-order term has the exponent $b+1$.

Consider the above example with $k = 2$, $a_1 = a_2 = 1$ and $b = 9$. The tail exponent in this case is $b = 9$. The simulations here are based on 5,000 independent realizations of X_N

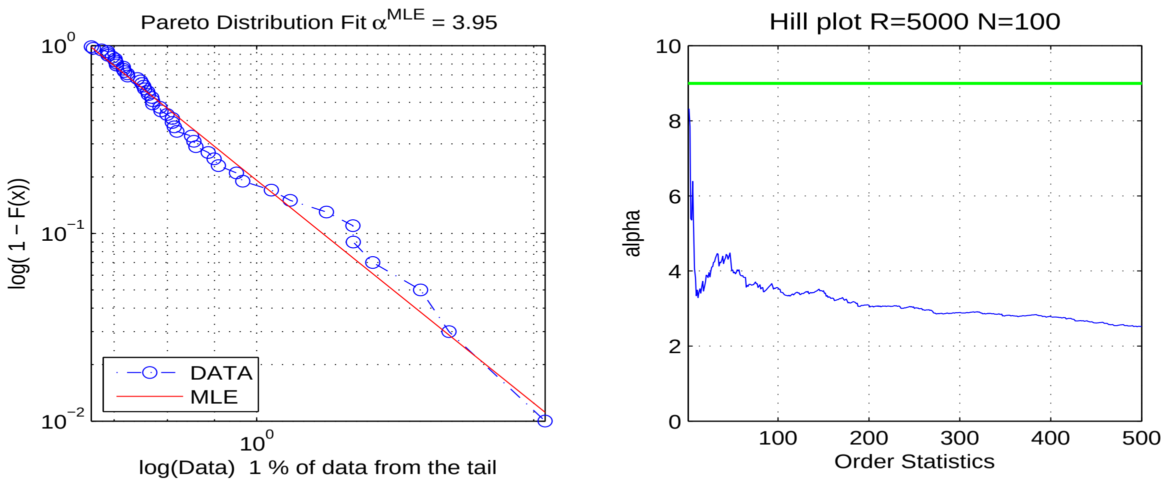

Figure 2.5: Explicit example of Section 2.2.2 with $a_1 = a_2 = 1$ and $b = 9$. True exponent $\alpha = 9$.

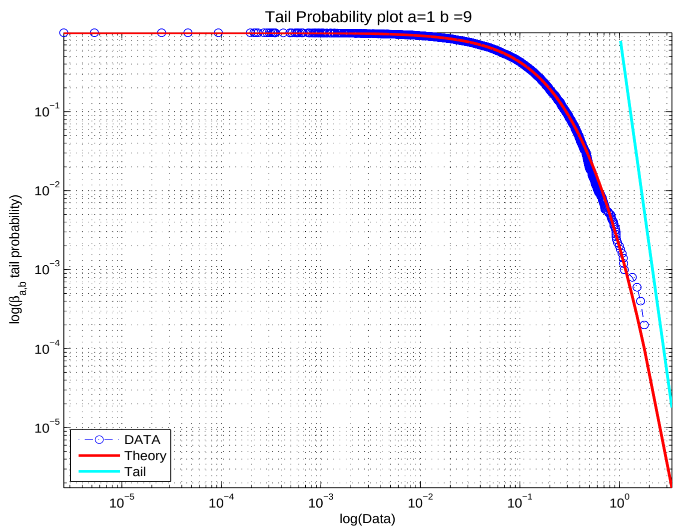

Figure 2.6: Explicit example of Section 2.2.2 with $a_1 = a_2 = 1$ and $b = 9$. Empirical tail distribution stays far below the asymptotic tail behavior.

with $N = 100$ iterations. Figure 2.5 shows tail exponent estimation. In addition, Figure 2.6 shows tails of empirical and theoretical distributions over the range of data. The theoretical power-law tail is also plotted and it is seen from Figure 2.6 that the theoretical tail is not yet in the asymptotic range (2.8).

The fact that the asymptotic range (2.8) is not observed here with data, can be explained in theory because the corresponding distribution has a closed form. Using (2.32) and comparing the ratio of distribution tail and the power law x^{-b}, suppose that

$$\left|\frac{P(X>x)}{x^{-b}/\left(bB(a_1,b)\right)}-1\right| = \left|\frac{P(Z<1/(1+x))}{x^{-b}/\left(bB(a_1,b)\right)}-1\right| \le \epsilon, \tag{2.33}$$

where $Z \stackrel{d}{=} B(b,a_1)$ from the relation (2.28). For example, with $a_1 = 1$, $b = 9$ and $\epsilon = .5$, numerical computations show that (2.33) holds for $x > 12.49$. The probability of having observation greater than 12.49 is 6.76×10^{-11}, which is too small from a practical perspective.

2.2.3 Multiplicative cascades

Let $T = [0,1)$. For $k_i \in \{0,1\}$, $i \ge 1$, denote

$$I_{k_1,\dots,k_n} = \Big[\frac{l}{2^n}, \frac{l+1}{2^n}\Big), \quad l = k_1 2^0 + \ldots + k_n 2^{n-1}, \tag{2.34}$$

subintervals of T obtained by splitting in a dyadic fashion. Let also $\{W_{k_1,\dots,k_i}, k_i \in \{0,1\}, i \ge 1\}$ be a family of i.i.d., nonnegative, mean 1 random variables, called multipliers. Define a random measure λ_n on $\mathcal{B}(T)$ by

$$\lambda_n(E) = \int_E f_n(t)dt, \quad \text{with } f_n(t) = \sum_{k_1,\dots,k_n \in \{0,1\}} \left(\prod_{i=1}^n W_{k_1,\dots,k_i}\right) 1_{I_{k_1,\dots,k_n}}(t). \tag{2.35}$$

Note, in particular, that

$$\lambda_n(I_{k_1,\dots,k_n}) = 2^{-n} \prod_{i=1}^n W_{k_1,\dots,k_i}. \tag{2.36}$$

(For example, $\lambda_1[0,1/2) = 2^{-1}W_0$, $\lambda_3([1/8,2/8)) = 2^{-3}W_0W_{0,0}W_{0,0,1}$ and so on.) Provided $E(W\log_2 W) < 1$, one can show that the sequence λ_n converges weakly to a random measure λ_∞ on $\mathcal{B}(T)$ almost surely, that is,

$$\lambda_n \Rightarrow \lambda, \text{ on } \mathcal{B}(T) \text{ a.s.} \tag{2.37}$$

where $\Rightarrow$ indicates weak convergence. The limiting random measure λ_∞ is known as a multiplicative cascade (MC, in short). See, for example, Mandelbort (1974), Ossiander and Waymire (2000). The following theorem is a well-known fact about the existence of moments of λ_∞ and related results. Let

$$\chi_2(h) = \log_2 E(W^h 1_{\{W>0\}}) - (h-1) \tag{2.38}$$

be the so-called structure function associated with a multiplier W.

Theorem 2.2.1. *(Kahane and Peyrière (1976), Guivarc'h (1990)) The following statements hold:*

i) $E\lambda_\infty(T) = 1$ *if and only if* $\chi_2'(1-) < 0$.

ii) $E(\lambda_\infty(T))^h < \infty$ *for* $0 \le h \le 1$ *and if*

$$\alpha := \sup\{h \ge 1 : \chi_2(h) \le 0\} > 1, \tag{2.39}$$

then $E\lambda_\infty^h(T) < \infty$ *for* $1 < h < \alpha$.

iii) Furthermore, if the cascade (multiplier) is non-lattice, then

$$P(\lambda_\infty(T) > x) \sim cx^{-\alpha}, \quad \text{as } x \to \infty. \tag{2.40}$$

The tail behavior in (2.40) of interest here can be proved by using Theorem 2.1.1 in the following way. Denote

$$M_n = \lambda_n[0,1), \quad M = \lambda_\infty[0,1). \tag{2.41}$$

By "separating" the multipliers W_1 and W_2 at the first generation, one can see that

$$M_n \stackrel{d}{=} \frac{W_1}{2} M_{n-1}^{(1)} + \frac{W_2}{2} M_{n-1}^{(2)}, \quad n \ge 1, \tag{2.42}$$

where W_1, W_2, $M_{n-1}^{(1)}$, $M_{n-1}^{(2)}$ are all independent, $M_{n-1}^{(1)} \stackrel{d}{=} M_{n-1}^{(2)} \stackrel{d}{=} M_{n-1}$, $W_1 \stackrel{d}{=} W_2 \stackrel{d}{=} W$ (the general multipliers in (2.36)), and by convention, $M_0 \equiv 1$. Similarly, the limiting measure satisfies the equation

$$M \stackrel{d}{=} \frac{W_1}{2} M^{(1)} + \frac{W_2}{2} M^{(2)}. \tag{2.43}$$

The equation (2.42) resembles the RDE (2.1) when considered in distribution by setting $A_n = W_1/2$, $B_n = W_2 M_{(n-1)}^{(2)}/2$. The key difference is that in (2.1), it is supposed that (A_n, B_n) are i.i.d. vectors which is not the case for MC because the distribution of $B_n = W_2 M_{n-1}^{(2)}/2$ depends on n. Equation (2.43), on the other hand, can be thought as a special case of RDE (2.2).

Though (2.43) is RDE, Theorem 2.1.1 cannot be applied directly to it. Indeed, in view of (2.43) and (2.2), supposing $A \stackrel{d}{=} W_1/2$, $B \stackrel{d}{=} W_2 M/2$, the assumption (2.5) requires that

$$E|B|^\alpha = E\left|\frac{W}{2}\right|^\alpha E|M|^\alpha < \infty. \tag{2.44}$$

But one expects M to have the tail exponent α and hence one cannot expect that (2.44) is satisfied. Despite this, however, there is still a way that Theorem 2.1.1 can be applied to obtain the tail behavior of M. The trick can be found in Guivarc'h (1990), Liu (2000) and others (though seems to be originally due to Guivarc'h (1990)).

The basic idea is as follows. Let $\widetilde{M}$ be a random variable with distribution $P_{\widetilde{M}}(dx) = xP_M(dx)$. Note that $EM = 1$, so $xP_M(dx)$ is a probability measure. Equation (2.43) can be rewritten in terms of characteristic functions as

$$\phi(t) = E(e^{itM}) = E\left(e^{it(A_1 M^{(1)} + A_2 M^{(2)})}\right) = \left(E\left(\phi(A_1 t)\right)\right)^2 \tag{2.45}$$

(for the shortness of notation, we denote $A_i = W_i/2, i = 1, 2$). Consider the random vector $(\widetilde{A}, \widetilde{B})$, independent of $\widetilde{M}$, with the distribution given by

$$Eh(\widetilde{A}, \widetilde{B}) = E\left(A_1 h(A_1, A_2 M^{(2)}) + A_2 h(A_2, A_1 M^{(1)})\right) = 2E\left(A_1 h(A_1, A_2 M)\right). \tag{2.46}$$

Note also that the characteristic function of $\widetilde{M}$ is

$$\widetilde{\phi}(t) = E(e^{it\widetilde{M}}) = \int e^{itx} x P_M(dx) = E(M e^{itM}) = -i\phi'(t), \tag{2.47}$$

where $\phi'(t)$ is the derivative of $\phi(t)$.

Observe that

$$E\left(e^{it(\widetilde{A}\widetilde{M} + \widetilde{B})}\right) = E\left(e^{it\widetilde{B}} e^{it\widetilde{A}\widetilde{M}}\right) = E\left(e^{it\widetilde{B}} \widetilde{\phi}(\widetilde{A}t)\right).$$

By applying (2.46) with $h(a,b) = e^{itb}\widetilde{\phi}(at)$ and (2.47), it becomes

$$2E\left(A_1 e^{itA_2M}\widetilde{\phi}(A_1t)\right) = 2E\left(A_1\widetilde{\phi}(A_1t)\right)E\left(e^{itA_2M}\right) = -2iE\left(A_1\phi'(A_1t)\right)E\left(\phi(A_2t)\right). \quad (2.48)$$

Note that differentiating the right-hand side of (2.45) gives

$$\phi'(t) = 2E\left(\phi'(A_1t)A_1\right)E\left(\phi(A_1t)\right).$$

Therefore, one can conclude that

$$E(e^{it\widetilde{M}}) = E(e^{it(\widetilde{A}\widetilde{M}+\widetilde{B})})$$

or

$$\widetilde{M} \stackrel{d}{=} \widetilde{A}\widetilde{M} + \widetilde{B}, \quad (2.49)$$

where $(\widetilde{A}, \widetilde{B})$ is independent of $\widetilde{M}$.

Now, consider the solution $\widetilde{M}$ of (2.49) instead of M of (2.43). The conditions in Theorem 2.1.1 become

$$E|\widetilde{A}|^{p-1} = 2E|A_1|^p = 2^{\chi_2(p)},$$

$$E|\widetilde{B}|^{p-1} = 2EA_1|A_2M|^{p-1} = 2EA_1E|A_2|^{p-1}E|M|^{p-1}.$$

Since $\chi_2(\alpha) = 0$ by (2.39) and $E|M|^{\alpha-1} < \infty$ is expected, RDE (2.49) should now have a stationary solution with the tail exponent $(\alpha - 1)$.

Establishing the tail behavior of $\widetilde{M}$ leads naturally to that of M. Observe that

$$P(M > x) = \int_x^\infty F_M(dy) = \int_x^\infty \frac{1}{y} yF_M(dy) = \int_x^\infty \frac{1}{y} F_{\widetilde{M}}(dy).$$

Integration by part gives,

$$P(M > x) = -\frac{1}{x}F_{\widetilde{M}}(x) + \int_x^\infty y^{-2}F_{\widetilde{M}}(y)dy,$$

where $F_{\widetilde{M}}(y) = P(\widetilde{M} \leq y)$, or

$$xP(M > x) = P(\widetilde{M} > x) - x \int_x^\infty y^{-2} P(\widetilde{M} > y) dy. \tag{2.50}$$

Rewriting (2.50) gives

$$\frac{x^\alpha P(M > x)}{x^{\alpha-1} P(\widetilde{M} > x)} = 1 - \frac{x^\alpha \int_x^\infty y^{-2} P(\widetilde{M} > y) dy}{x^{\alpha-1} P(\widetilde{M} > x)}. \tag{2.51}$$

Since the tail of random variable $\widetilde{M}$ is expected as

$$P(\widetilde{M} > x) \sim \widetilde{c} x^{-(\alpha-1)},$$

for some positive constant $\widetilde{c}$, the right-hand side of (2.51) converges to $1 - 1/\alpha$, as $x \to \infty$. This yields

$$P(M > x) \sim c x^\alpha,$$

where $c = \widetilde{c}(\alpha - 1)/\alpha$.

To illustrate our problem through simulations, consider the case of multiplicative cascade with log-normal multipliers $LN(-\sigma^2/2, \sigma^2)$ where the latter choice of parameters ensures mean 1. Simulations are based on i.i.d. copies of M_N^r, $r = 1, \ldots, R$, where $N = 13$ and $R = 1,000$. The parameter is taken as $\sigma^2 = .2 \log 2$. According to a small calculation found in Appendix B, the corresponding tail exponent is given by

$$\alpha = \frac{2 \log 2}{\sigma^2} = \frac{2 \log 2}{.2 \log 2} = 10. \tag{2.52}$$

Figure 2.7 shows the corresponding tail distribution plot with Pareto distribution fit (left) and Hill plot (right). The tail appears power-law from the tail plot. However, as seen from the Hill plot, theoretical tail exponent (2.52) is far from any reasonable estimate of the tail.

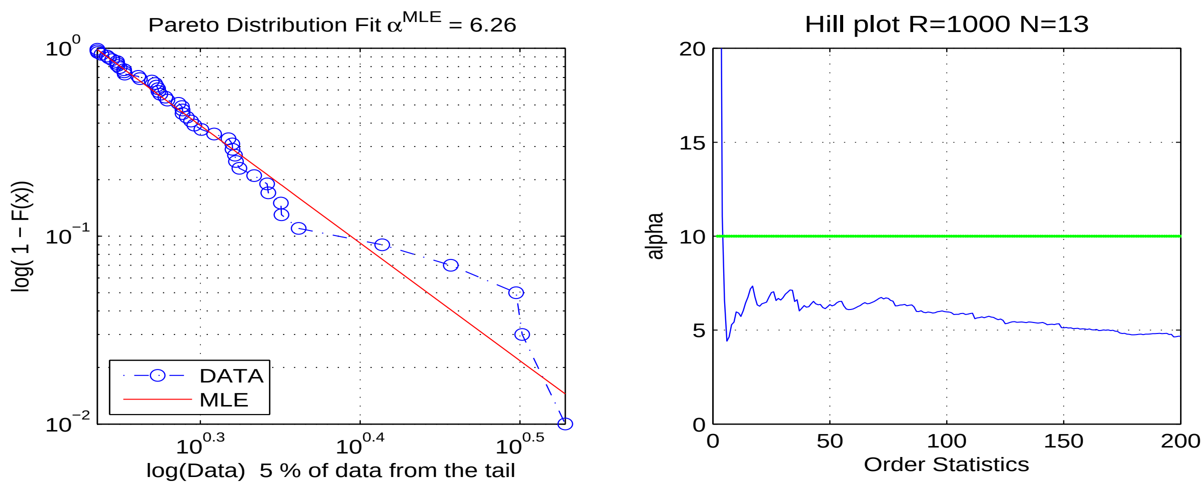

Figure 2.7: Multiplicative cascade with log-normal multipliers. Theoretical tail exponent 10 is far from the estimated tail exponent.

2.3 Second-order properties of distribution tails of RDEs

In this section, we show that weaker form (2.16) of second-order regular variation holds for RDEs, that is,

$$\int_x^\infty (P(X > u) - P(AX > u))\, du \sim cx^{-\alpha}, \tag{2.53}$$

as $x \to \infty$. We first illustrate (2.53) in a simple example, and then extend our proof to more general RDEs.

Example 2.3.1. Consider RDE defined as

$$X \stackrel{d}{=} AX + 1.$$

Then,

$$\int_x^\infty (P(X > u) - P(AX > u))\, du = \int_x^\infty (P(X > u) - P(X > u+1))\, du$$

$$= \int_x^\infty P(X > u)du - \int_{x+1}^\infty P(X > u)du = \int_x^{x+1} P(X > u)du.$$

Since $P(X > u)$ is monotone decreasing, we have

$$P(X > x+1) \le \int_x^{x+1} P(X > u)du \le P(X > x). \tag{2.54}$$

By Theorem 2.1.1, both sides of (2.54) behave as $c_+x^{-\alpha}$. Therefore, as $x \to \infty$,

$$\int_x^\infty (P(X > u) - P(AX > u))\, du \sim c_+x^{-\alpha}.$$

As the following theorem shows, the relation (2.53) holds for a large class of RDEs. We first consider the case when A, B and X in (2.2) are all nonnegative. The general case is considered later in the section.

Theorem 2.3.1. *Let $A \geq 0$, $B \geq 0$ and $X \geq 0$ a.s. and (A, B) be independent of X. Suppose that the assumptions of Theorem 2.1.1 hold. In addition, if*

$$EX < \infty, \quad EA^\alpha B < \infty, \tag{2.55}$$

$$x^\alpha EA1_{\{B>Cx\}} \to 0, \quad \textit{for any } C > 0, \textit{ as } x \to \infty, \tag{2.56}$$

$$x^\alpha EA1_{\{A>Cx\}} \to 0, \quad \textit{for any } C > 0, \textit{ as } x \to \infty, \tag{2.57}$$

$$x^\alpha \int_x^\infty P(B > z)dz \to C_+, \quad \textit{as } x \to \infty, \tag{2.58}$$

where $C_+ \geq 0$, then X in RDE (2.2) satisfies

$$\int_x^\infty (P(X > u) - P(AX > u))\, du \sim x^{-\alpha} (C_+ + c_+E(A^\alpha B)), \tag{2.59}$$

with positive constant c_+ defined in (2.8).

Proof. Observe that

$$\begin{aligned}
&\int_x^\infty (P(X > u) - P(AX > u))\, du \\
&= \int_x^\infty \int_0^\infty \int_0^\infty (P(aX > u - b) - P(aX > u))\, F_{A,B}(da, db)du \\
&= \int_0^\infty \int_0^\infty \int_{x-b}^x P(aX > u)duF_{A,B}(da, db).
\end{aligned} \tag{2.60}$$

For fixed $c < 1$, we can further rewrite (2.60) by splitting the range of b into $(0, cx)$ and (cx, ∞)

$$\int_{cx}^\infty \int_0^\infty \int_{x-b}^x P(aX > u)duF_{A,B}(da, db)$$

$$+\int_0^{cx}\int_0^{\infty}\int_{x-b}^{x} P(aX>u)du F_{A,B}(da,db) =: I+L. \tag{2.61}$$

For the integral I, write $I = I_1 + I_2$, where

$$\begin{aligned} I_1 &= \int_{cx}^{x}\int_0^{\infty}\int_{x-b}^{x} P(aX>u)du F_{A,B}(da,db), \\ I_2 &= \int_{x}^{\infty}\int_0^{\infty}\int_{x-b}^{x} P(aX>u)du F_{A,B}(da,db). \end{aligned}$$

The integral I_1 can be bounded as

$$I_1 \le \int_{cx}^{\infty}\int_0^{\infty} E(aX)F_{A,B}(da,db) = EXEA1_{\{B>cx\}},$$

and by the assumptions (2.55) and (2.56), we have

$$x^{\alpha} I_1 \to 0. \tag{2.62}$$

Observe for I_2 that

$$\begin{aligned} I_2 &= \int_{x}^{\infty}\int_0^{\infty}\left(\int_{x-b}^{0} P(aX>u)du + \int_0^{x} P(aX>u)du\right) F_{A,B}(da,db) \\ &= \int_{x}^{\infty}\int_0^{\infty}\left((b-x) + \int_0^{x} P(aX>u)du\right) F_{A,B}(da,db) \\ &= \int_{x}^{\infty}\int_0^{\infty}(b-x)F_{A,B}(da,db) + \int_{x}^{\infty}\int_0^{\infty}\int_0^{x} P(aX>u)du F_{A,B}(da,db) =: I_{2,1}+I_{2,2}. \end{aligned}$$

Note that, by (2.58),

$$x^{\alpha}I_{2,1} = x^{\alpha}E(B-x)_+ = x^{\alpha}\int_0^{\infty} P((B-x)_+>y)dy = x^{\alpha}\int_x^{\infty} P(B>z)dz \to C_+. \tag{2.63}$$

By the assumptions (2.55) and (2.56) with $C=1$, we have

$$x^{\alpha}I_{2,2} \le x^{\alpha}EXEA1_{\{B\ge x\}} \to 0. \tag{2.64}$$

Combining (2.62), (2.63) and (2.64) yields

$$x^{\alpha} I \to C_{+}. \tag{2.65}$$

We now turn to the integral L in (2.61). By Theorem 2.1.1, we can select x_0 such that, for any given $\epsilon > 0$,

$$|x^{\alpha} P(X > x) - c_{+}| \leq \epsilon, \text{ for all } x \geq x_0, \tag{2.66}$$

where c_{+} is a constant described in (2.8). For such x_0, write the integral L as

$$L = \int_0^{cx} \int_{(x-b)/x_0}^{\infty} \int_{x-b}^{x} P(aX > u) du F_{A,B}(da, db)$$

$$+ \int_0^{cx} \int_0^{(x-b)/x_0} \int_{x-b}^{x} P(aX > u) du F_{A,B}(da, db) =: J + K. \tag{2.67}$$

For fixed x_0, the integral J can be bounded as

$$J \leq EXE(A 1_{\{0<B<cx\}} 1_{\{A>(x-B)/x_0\}}) \leq EXE\left(A 1_{\{A>(1-c)x/x_0\}}\right),$$

since $B \in (0, cx)$ implies $(x - B)/x_0 \geq (1 - c)x/x_0$. Hence, (2.57) implies that

$$x^{\alpha} J \to 0. \tag{2.68}$$

Consider now the integral K in (2.67). Since $P(aX > u)$ is monotone decreasing, the integral K satisfies,

$$x^{\alpha} K_1 := x^{\alpha} \int_0^{cx} \int_0^{(x-b)/x_0} bP\left(aX > x\right) F_{A,B}(da, db) \leq x^{\alpha} K$$

$$\leq x^{\alpha} \int_0^{cx} \int_0^{(x-b)/x_0} bP\left(aX > x - b\right) F_{A,B}(da, db) =: x^{\alpha} K_2. \tag{2.69}$$

Write the integral $x^{\alpha} K_2$ as

$$c_{+} \int_0^{cx} \int_0^{(x-b)/x_0} ba^{\alpha} \left(1 - \frac{b}{x}\right)^{-\alpha} \frac{P\left(X > \frac{x-b}{a}\right)}{c_{+} \left(\frac{x-b}{a}\right)^{-\alpha}} F_{A,B}(da, db).$$

Since $b \in (0, cx)$ and $a \in (0, (x-b)/x_0)$,

$$\left(1 - \frac{b}{x}\right)^{-\alpha} \frac{P\left(X > \frac{x-b}{a}\right)}{c_+ \left(\frac{x-b}{a}\right)^{-\alpha}} \le (1-c)^{-\alpha} D_1,$$

where D_1 is some constant determined by (2.66). The assumption (2.55) and the dominated convergence theorem yield

$$x^\alpha \int_0^{cx} \int_0^{(x-b)/x_0} bP\left(aX > x - b\right) F_{A,B}(da, db) \to c_+ E(A^\alpha B). \tag{2.70}$$

Similarly, $x^\alpha K_1$ becomes

$$c_+ \int_0^{cx} \int_0^{(x-b)/x_0} ba^\alpha \frac{P\left(X > \frac{x}{a}\right)}{c_+ \left(\frac{x}{a}\right)^{-\alpha}} F_{A,B}(da, db)$$

and $a \le (x-b)/x_0$ implies

$$\frac{P\left(X > \frac{x}{a}\right)}{c_+ \left(\frac{x}{a}\right)^{-\alpha}} \le D_2,$$

for some constant D_2. Again, by the assumption (2.55) and the dominated convergence theorem,

$$\int_0^{cx} \int_0^{(x-b)/x_0} bP\left(aX > x\right) F_{A,B}(da, db) \to c_+ E(A^\alpha B). \tag{2.71}$$

Hence, (2.70) and (2.71) imply

$$x^\alpha K \to c_+ E(A^\alpha B). \tag{2.72}$$

Finally, combining (2.65), (2.68) and (2.72) yields (2.59). $\square$

Remarks

1. The relation (2.16) implies (2.17), for example, when $P(X > u) - P(AX > u)$ is ultimately monotone (see, for example, Bingham et al. (1989), p. 39). Whether the latter monotonicity holds is still an open question.

2. Note that, for $\delta > 0$,

$$EA1_{\{B > Cx\}} \le x^{-\alpha-\delta} C^{-\alpha-\delta} EAB^{\alpha+\delta}.$$

Hence, if

$$EAB^{\alpha+\delta} < \infty \tag{2.73}$$

for some $\delta > 0$, then (2.56) is satisfied. Similarly, the conditions (2.57) and (2.58) hold if

$$EA^{\alpha+1+\delta} < \infty \text{ and } EB^{\alpha+1+\delta} < \infty, \tag{2.74}$$

for some $\delta > 0$, respectively. In particular, $EB^{\alpha+1+\delta} < \infty$ implies $C_+ = 0$.

Example 2.3.2. (ARCH(1) model) Recall the discussion on ARCH(1) model found in Section 2.2.1. The model satisfies the assumptions of Theorem 2.3.1 for $\kappa = \alpha/2 > 1$. Indeed, for such κ, $E\xi_t^2 < \infty$ and obviously $EA_t^\kappa B_t = E(\lambda\epsilon_t^2)^\kappa(\beta\epsilon_t^2) < \infty$ so that (2.55) holds. The conditions (2.73)–(2.74) hold (with $C_+ = 0$) in the second remark above because $A_t = \lambda\epsilon_t^2$, $B_t = \beta\epsilon_t^2$ have all their moments finite for normal error terms ϵ_t. Hence, by Theorem 2.3.1, if $\kappa = \alpha/2 > 1$,

$$\int_x^\infty \big(P(\xi^2 > u) - P(\lambda\epsilon^2\xi^2 > u)\big)\, du \sim c_+\lambda^{2\alpha}\beta E(\epsilon^{2\alpha+2})x^{-\alpha/2}$$

or, by symmetry and a change of variables,

$$\int_x^\infty \Big(P(\xi > v) - P(\sqrt{\lambda}\epsilon\xi > v)\Big)\, vdv \sim \frac{c_+}{4}\lambda^{2\alpha}\beta E(\epsilon^{2\alpha+2})x^{-\alpha}.$$

Example 2.3.3. (Multiplicative cascades with lognormal multipliers) Consider RDE (2.49) with tail exponent $\alpha - 1$ and log-normal multipliers. If $\alpha > 2$ or $\alpha - 1 > 1$, then the condition (2.55) in Theorem 2.3.1 is satisfied because

$$E\widetilde{M} = EM^2 < \infty, \quad E\widetilde{A}^{\alpha-1}\widetilde{B} = 2EA_1^\alpha EA_2 EM = \frac{1}{2} < \infty$$

since $EM = 1$ by Theorem 2.2.1, i). Condition (2.56) can be easily checked by observing that

$$x^{\alpha-1}E\widetilde{A}1_{\{\widetilde{B}>Cx\}} = x^{\alpha-1}2E(A_1^2 1_{\{A_2M>Cx\}}) = x^{\alpha-1}2EA_1^2 P(A_2M > Cx) \to 0,$$

since Breiman's theorem (Breiman (1965)) implies $P(A_2M > Cx) \sim \frac{1}{2}c_+(Cx)^{-\alpha}$. The condition

(2.57) can be verified through the first condition in (2.74) with $\delta = 1$,

$$E\widetilde{A}^{\alpha+1} = 2EA_1^{\alpha+2} < \infty,$$

since log-normal distribution has all its moments finite. Note next that the condition (2.58) becomes

$$x^{\alpha-1} \int_x^\infty P(\widetilde{B} > z)dz = x^{\alpha-1} \int_x^\infty P\left(A_2 M > z\right) dz. \tag{2.75}$$

Applying Breiman's theorem again, we have

$$P\left(A_2 M > z\right) \sim \frac{1}{2} c_+ z^{-\alpha},$$

and for sufficiently large x, (2.75) leads to

$$x^{\alpha-1} \int_x^\infty P\left(A_2 M > z\right) dz \sim x^{\alpha-1} \int_x^\infty \frac{1}{2} c_+ z^{-\alpha} dz = \frac{c_+}{2(\alpha-1)}.$$

Hence, if $\alpha > 2$, by Theorem 2.3.1, MC with log-normal multiplier satisfies the relation

$$x^{\alpha-1} \int_x^\infty \left(P(\widetilde{M} > u) - P(\widetilde{A}\widetilde{M} > u)\right) du \sim \frac{c_+(\alpha+1)}{2(\alpha-1)}. \tag{2.76}$$

By using the relationship between $\widetilde{M}$, $\widetilde{A}$ and M, A found in Section 2.2.3, the relation (2.76) can be rewritten as

$$x^{\alpha-1} \int_x^\infty \left(\int_u^\infty y P_M(dy) - E \int_{u/\widetilde{A}}^\infty y P_M(dy) \right) du \sim \frac{c_+(\alpha+1)}{2(\alpha-1)}. \tag{2.77}$$

Furthermore, if (2.76) implies

$$x^\alpha \left(P(\widetilde{M} > x) - P(\widetilde{A}\widetilde{M} > x)\right) \sim \frac{c_+(\alpha+1)}{2}, \tag{2.78}$$

then this could be translated back to M, A as follows. Similarly to (2.50) one has

$$xP(AM > x) = \frac{1}{2} P(\widetilde{A}\widetilde{M} > x) - \frac{x}{2} \int_x^\infty y^{-2} P(\widetilde{A}\widetilde{M} > y) dy. \tag{2.79}$$

By (2.50) and (2.79),

$$\frac{x\left(P(M>x)-2P(AM>x)\right)}{P(\widetilde{M}>x)-P(\widetilde{A}\widetilde{M}>x)} = 1-\frac{x\int_x^\infty y^{-2}\left(P(\widetilde{M}>y)-P(\widetilde{A}\widetilde{M}>y)\right)dy}{P(\widetilde{M}>x)-P(\widetilde{A}\widetilde{M}>x)}. \tag{2.80}$$

The right-hand side of (2.80) converges to $1-1/(\alpha+1)=\alpha/(\alpha+1)$, as $x\to\infty$. Therefore, using (2.78),

$$x^{\alpha+1}\left(P(M>x)-2P(AM>x)\right)\sim\frac{c_+\alpha}{2}. \tag{2.81}$$

Theorem 2.3.2 below generalizes second-order properties to real-valued X, A and B satisfying RDE (2.2).

Theorem 2.3.2. *Suppose that the assumptions of Theorem 2.1.1 hold. If*

$$E|X|<\infty, \quad E|A|^\alpha|B|<\infty, \tag{2.82}$$

$$x^\alpha E|A|1_{\{|A|>Cx\}}\to 0, \quad \text{for any } C>0, \text{ as } x\to\infty, \tag{2.83}$$

$$x^\alpha E|A|1_{\{|B|>Cx\}}\to 0, \quad \text{for any } C>0, \text{ as } x\to\infty, \tag{2.84}$$

$$x^\alpha E(B-x)_+1_{\{A>0\}}\to C_+^1, \quad x^\alpha E(B-x)_+1_{\{A<0\}}\to C_+^2, \quad \text{as } x\to\infty, \tag{2.85}$$

where C_+^1 and C_+^2 are nonnegative constants. Then, X in RDE (2.2) satisfies

$$x^\alpha\int_x^\infty\left(P(X>u)-P(AX>u)\right)du\sim C_+^1+C_+^2+c_+E(A_+^\alpha B)+c_-E(A_-^\alpha B), \tag{2.86}$$

where constants c_+ and c_- are defined in (2.8).

Proof. We sketch the proof as it is similar to that of Theorem 2.3.1. Split the integral in (2.86) according to the sign of random variables A and B as

$$\int_x^\infty\left(P(X>u)-P(AX>u)\right)du=\left(\int_0^\infty\int_0^\infty\int_x^\infty+\int_0^\infty\int_{-\infty}^0\int_x^\infty+\right.$$

$$\left.\int_{-\infty}^0\int_0^\infty\int_x^\infty+\int_{-\infty}^0\int_{-\infty}^0\int_x^\infty\right)\left(P(aX+b>u)-P(aX>u)\right)duF_{A,B}(da,db)$$

$$=: J_1+J_2+J_3+J_4.$$

Applying the proof of Theorem 2.3.1 with A_+ and B_+ gives

$$x^{\alpha} J_1 \to C_+^1 + c_+ E(A_+^{\alpha} B_+).$$

Note that J_2 can be related to J_1 by rewriting it as

$$J_2 = \int_0^{\infty} \int_0^{\infty} \int_x^{\infty} \left(P\left(aX < -(u-b)\right) - P\left(aX < -u\right)\right) du F_{-A,B}(da, db).$$

Hence,

$$x^{\alpha} J_2 \to C_+^2 + c_- E A_-^{\alpha} B_+.$$

From the Kesten's result, there is x_0 such that, for all $x > x_0$ and given $\epsilon > 0$,

$$|x^{\alpha} P(X > x) - c_+| \leq \epsilon. \tag{2.87}$$

For such x_0, J_3 can be rewritten as

$$J_3 = \int_{-\infty}^{0} \int_0^{\infty} \int_{x-b}^{x} P(aX > u) du F_{A,B}(da, db)$$

$$= \int_{-\infty}^{0} \int_0^{x/x_0} \int_{x-b}^{x} P(aX > u) du F_{A,B}(da, db) + \int_{-\infty}^{0} \int_{x/x_0}^{\infty} \int_{x-b}^{x} P(aX > u) du F_{A,B}(da, db)$$

$$=: J_{3,1} + J_{3,2}.$$

Second term $J_{3,2}$ does not contribute to the asymptotics because (2.82) and (2.83) imply

$$x^{\alpha} |J_{3,2}| \leq x^{\alpha} \int_{-\infty}^{0} \int_{x/x_0}^{\infty} \int_0^{\infty} P(aX > u) du F_{A,B}(da, db) = x^{\alpha} E X_+ E A 1_{\{A > x/x_0\}} 1_{\{B<0\}} \to 0.$$

First term $J_{3,1}$ satisfies

$$x^{\alpha} \int_{-\infty}^{0} \int_0^{x/x_0} b P\left(aX > x\right) F_{A,B}(da, db) \leq x^{\alpha} J_{3,1}$$

$$\leq x^{\alpha} \int_{-\infty}^{0} \int_0^{x/x_0} b P\left(aX > x - b\right) F_{A,B}(da, db). \tag{2.88}$$

The right-hand side of (2.88) can be rewritten as

$$c_+ \int_{-\infty}^{0} \int_{0}^{x/x_0} ba^{\alpha} \left(1 - \frac{b}{x}\right)^{-\alpha} \frac{P\left(X > \frac{x-b}{a}\right)}{c_+ \left(\frac{x-b}{a}\right)^{-\alpha}} F_{A,B}(da, db).$$

Relation (2.87) and $1 - b/x > 1$ imply further that

$$\left(1 - \frac{b}{x}\right)^{-\alpha} \frac{P\left(X > \frac{x-b}{a}\right)}{c_+ \left(\frac{x-b}{a}\right)^{-\alpha}} \leq D_1,$$

for some constant D_1. Arguing similarly for the left-hand side of (2.88) and applying the dominated convergence theorem lead to

$$x^{\alpha} J_{3,1} \to -c_+ E A_+^{\alpha} B_-.$$

Observe for J_4 that

$$J_4 = \int_0^{\infty} \int_0^{\infty} \int_x^{\infty} \left(P\left(aX < -(u+b)\right) - P\left(aX < -u\right)\right) du F_{-A,-B}(da, db).$$

As for J_3, this leads to

$$x^{\alpha} J_4 \to -c_- E(A_-^{\alpha} B_-).$$

Gathering the results for J_1, J_2, J_3 and J_4 leads to the desired result. □

Finally, we show that if distribution tail behaves as (2.13) and satisfies the weaker form of 2RV (2.16), then $\rho = -1$. For simplicity, we only consider the case of nonnegative A, B and X. We need the following lemma relating distribution tails $P(X > x)$ and $P(AX > x)$.

Lemma 2.3.1. *Suppose $A \geq 0$ a.s., $EA^{\alpha-\rho} < \infty$ for $\rho < 0$ and*

$$P(X > x) - c_1 x^{-\alpha} \sim c_2 x^{-\alpha+\rho}. \tag{2.89}$$

Then

$$P(AX > x) - c_1 x^{-\alpha} \sim c_2 E A^{\alpha-\rho} x^{-\alpha+\rho}. \tag{2.90}$$

Proof. Observe that

$$P(AX > x) - c_1 x^{-\alpha} = \int_0^\infty \left(P\left(X > \frac{x}{a}\right) - c_1 \left(\frac{x}{a}\right)^{-\alpha} \right) F_A(da), \tag{2.91}$$

since $EA^\alpha = 1$. Relation (2.89) implies that for any $\epsilon > 0$, there is x_0 such that

$$|x^{\alpha-\rho}(P(X > x) - c_1 x^{-\alpha}) - c_2| \leq \epsilon, \text{ for all } x > x_0. \tag{2.92}$$

For such chosen x_0, we have

$$\frac{P(AX > x) - c_1 x^{-\alpha}}{c_1 EA^{\alpha-\rho} x^{-\alpha+\rho}} = \frac{\int_0^{x/x_0} \left(P\left(X > \frac{x}{a}\right) - c_1 \left(\frac{x}{a}\right)^{-\alpha} \right) F_A(da)}{c_1 EA^{\alpha-\rho} x^{-\alpha+\rho}}$$

$$+ \frac{\int_{x/x_0}^\infty \left(P\left(X > \frac{x}{a}\right) - c_1 \left(\frac{x}{a}\right)^{-\alpha} \right) F_A(da)}{c_1 EA^{\alpha-\rho} x^{-\alpha+\rho}} =: I + J. \tag{2.93}$$

Note first that the second term J in (2.93) does not contribute to the asymptotics. Indeed,

$$\left| \int_{x/x_0}^\infty \left(P\left(X > \frac{x}{a}\right) - c_1 \left(\frac{x}{a}\right)^{-\alpha} \right) F_A(da) \right| \leq \int_{x/x_0}^\infty \left(P\left(X > \frac{x}{a}\right) + c_1 \left(\frac{x}{a}\right)^{-\alpha} \right) F_A(da).$$

$$\leq E 1_{\{A \geq x/x_0\}} + c_1 x^{-\alpha} E(A^\alpha 1_{\{A > x/x_0\}})$$

$$\leq E\left(\left(\frac{A}{x/x_0}\right)^{\alpha-\rho} 1_{\{A \geq x/x_0\}} \right) + c_1 x^{-\alpha} E\left(A^\alpha \left(\frac{A}{x/x_0}\right)^{-\rho} 1_{\{A > x/x_0\}} \right)$$

yields

$$|J| \leq \frac{x^{-\alpha+\rho} x_0^{\alpha-\rho} E\left(A^{\alpha-\rho} 1_{\{A > x/x_0\}}\right) + c_1 x^{-\alpha+\rho} x_0^{-\rho} E(A^{\alpha-\rho} 1_{\{A > x/x_0\}})}{c_1 EA^{\alpha-\rho} x^{-\alpha+\rho}} \to 0,$$

as $x \to \infty$ since $EA^{\alpha-\rho} < \infty$.

The first term I, on the other hand, can be bounded using relation (2.92) as

$$\frac{\int_0^{x/x_0} (c_2 - \epsilon)(x/a)^{-\alpha+\rho} F_A(da)}{c_2 EA^{\alpha-\rho} x^{-\alpha+\rho}} \leq I \leq \frac{\int_0^{x/x_0} (c_2 + \epsilon)(x/a)^{-\alpha+\rho} F_A(da)}{c_2 EA^{\alpha-\rho} x^{-\alpha+\rho}}.$$

By taking limit $x \to \infty$ and $\epsilon \downarrow 0$, dominated convergence theorem implies

$$I \to 1,$$

since $EA^{\alpha-\rho} < \infty$. $\square$

Proposition 2.3.1. *If $A \geq 0$ a.s., $EA^{\alpha-\rho} < \infty$ for some $\rho < 0$, and*

$$P(X > x) - c_1 x^{-\alpha} \sim c_2 x^{-\alpha+\rho}, \tag{2.94}$$

$$\int_x^\infty \left(P(X > u) - P(AX > u)\right) du \sim c x^{-\alpha}, \tag{2.95}$$

then $\rho = -1$ and $c = c_2(1 - EA^{\alpha-\rho})/\alpha$.

Proof. Lemma 2.3.1 implies that

$$P(AX > x) - c_1 x^{-\alpha} \sim c_2 EA^{\alpha-\rho} x^{-\alpha+\rho}. \tag{2.96}$$

Therefore, we have

$$\int_x^\infty \left(P(X > u) - P(AX > u)\right) du = \int_x^\infty \left(c_2(1 - EA^{\alpha-\rho}) u^{-\alpha+\rho} + o(u^{-\alpha+\rho})\right) du$$

$$= \frac{-c_2(1 - EA^{\alpha-\rho})}{-\alpha+\rho+1} x^{-\alpha+\rho+1} + \int_x^\infty o(u^{-\alpha+\rho}) du.$$

$$= \frac{-c_2(1 - EA^{\alpha-\rho})}{-\alpha+\rho+1} x^{-\alpha+\rho+1} + o(x^{-\alpha+\rho+1}).$$

Finally, assumption (2.95) gives $\rho = -1$ and $c = c_2(1 - EA^{\alpha-\rho})/\alpha$. $\square$

Remark. Our results suggest for RDE that

$$P(X > x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1} + o(x^{-\alpha-1}). \tag{2.97}$$

This imposes conditions on c_1 and c_2 in the following sense. One expects that

$$P(X > x) = P(AX + B > x) = \int_0^\infty \int_0^\infty P\left(X > \frac{x-b}{a}\right) F_{A,B}(da, db)$$

$$= \int_0^\infty \int_0^\infty \left\{ c_1 \left(\frac{x-b}{a}\right)^{-\alpha} + c_2 \left(\frac{x-b}{a}\right)^{-\alpha-1} + o\left(\left(\frac{x-b}{a}\right)^{-\alpha-1}\right) \right\} F_{A,B}(da, db)$$

$$= \int_0^\infty \int_0^\infty \left(c_1 a^\alpha x^{-\alpha} + c_1 a^\alpha \left((x-b)^{-\alpha} - x^{-\alpha}\right) + c_2 a^{\alpha+1} x^{-\alpha-1} + o(x^{-\alpha-1})\right) F_{A,B}(da, db)$$

$$= c_1 x^{-\alpha} + \left(c_1 \alpha E(A^\alpha B) + c_2 E A^{\alpha+1}\right) x^{-\alpha-1} + o(x^{-\alpha-1}), \tag{2.98}$$

by using the relation

$$(x-b)^{-\alpha} - x^{-\alpha} = x^{-\alpha}\left(\left(1 - \frac{b}{x}\right)^{-\alpha} - 1\right) = b\alpha x^{-\alpha-1} + o(x^{-\alpha-1}).$$

Therefore, (2.97) and (2.98) are consistent only when

$$c_2 = \frac{c_1 \alpha E(A^\alpha B)}{1 - E A^{\alpha+1}}. \tag{2.99}$$

2.4 Multidimensional extension

In this subsection, we extend our results to multidimensional RDEs,

$$\mathbf{X}_n = \mathbf{A}_n \mathbf{X}_{n-1} + \mathbf{B}_n, \quad n \in \mathbb{Z}, \tag{2.100}$$

where $(\mathbf{A}_n, \mathbf{B}_n)$ is an i.i.d. sequence of $d \times d$ random matrices $\mathbf{A}_n$ and d-dimensional random vectors $\mathbf{B}_n$. We consider only the case when the entries of $\mathbf{A}_n$, $\mathbf{B}_n$ are nonnegative. Under mild conditions, multidimensional RDE has a stationary solution,

$$\mathbf{X} \stackrel{d}{=} \mathbf{A}\mathbf{X} + \mathbf{B}, \tag{2.101}$$

where $(\mathbf{A}, \mathbf{B}) \stackrel{d}{=} (\mathbf{A}_1, \mathbf{B}_1)$ is independent of $\mathbf{X}$. We recall next the result of Kesten (1973) for multidimensional RDEs. (Generalizations of Kesten's result can be found in Basrak, Davis and Mikosch (2002a), de Saporta, Guivarc'h and Le Page (2004), Klüppelberg and Pergamenchtchikov (2004), Guivarc'h (1990) to name but a few.) Denote the Euclidean norm as $\|\cdot\|$ and the operator norm as $\|\cdot\|_{op}$, namely,

$$\|\mathbf{A}\|_{op} = \sup_{\|\mathbf{y}\|=1} \|\mathbf{A}\mathbf{y}\|.$$

Let also $S_+ = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| = 1, \mathbf{z} > 0\}$ and, for $\mathbf{w} \in \mathbb{R}^d$,

$$\mathbf{w}^{\#} = \frac{\mathbf{w}}{\|\mathbf{w}\|}.$$

In particular, for $\mathbf{w} \in \mathbb{R}^d$ with nonnegative entries, note that $\mathbf{w}^{\#} \in S_+$.

Theorem 2.4.1. *(Kesten (1973), Theorems 3 and 4) Let $(\mathbf{A}_n, \mathbf{B}_n)$ be a sequence of i.i.d. $d \times d$ matrices $\mathbf{A}_n$ and $d \times 1$ vectors $\mathbf{B}_n$ with nonnegative entries. Assume that the following conditions hold:*

(A1) For some $\epsilon > 0$, $E\|\mathbf{A}_1\|_{op}^{\epsilon} < 1$.

(A2) $\mathbf{A}_1$ has no zeros rows a.s.

(A3) The group generated by

$$\{\log \rho(\pi) : \pi = \mathbf{a}_n \dots \mathbf{a}_1 > 0, n \geq 1, \mathbf{a}_n \dots \mathbf{a}_1 \in \textit{support of } \mathbf{A}_1\}$$

is dense in $\mathbb{R}$, where $\rho(\pi)$ denotes the largest positive eigenvalue, known as Frobenius eigenvalue and $\pi > 0$ means that all entries of this matrix are positive.

(A4) There exists $\kappa_0 > 0$ such that

$$E\left(\min_{i=1,\dots,d} \sum_{j=1}^{d} A_1(i,j)\right)^{\kappa_0} \geq d^{\kappa_0/2},$$

where $A_1(i,j)$ is a (i,j) entry of matrix $\mathbf{A}_1$ and

$$E\left(\|\mathbf{A}_1\|_{op}^{\kappa_0} \log^+ \|\mathbf{A}_1\|_{op}\right) < \infty.$$

Then the following statements hold:

(R1) There exists a unique solution $\alpha \in (0, \kappa_0]$ to the equation

$$0 = \lim_{n\to\infty} \frac{1}{n} E \log \|\mathbf{A}_n \ldots \mathbf{A}_1\|_{op}^{\alpha}.$$

(R2) There exists a unique solution $\mathbf{X}$ to the RDE (2.101).

(R3) If $E\|\mathbf{B}_1\|^\alpha < \infty$, then, for all $\mathbf{z} \in S_+$,

$$P(\mathbf{z}'\mathbf{X} > x) \sim x^{-\alpha} c_+ r(\mathbf{z}), \quad \text{as } x \to \infty, \tag{2.102}$$

where c_+ is a positive constant and $r(\mathbf{z})$ is a continuous and strictly positive function on S_+ satisfying

$$r(\mathbf{z}) = E\|\mathbf{z}'\mathbf{A}\|^\alpha r(\mathbf{z}'\mathbf{A}^{\#}). \tag{2.103}$$

The following result extends Theorem 2.3.1 to the multidimensional case. The proof is similar to that of Theorem 2.3.1 (with an additional technical difficulty reflected by assumptions (2.108) and (2.109) below). We denote by μ a Haar measure on S_+.

Theorem 2.4.2. *Suppose that the assumptions of Theorem 2.4.1 hold and let $(\mathbf{A}, \mathbf{B}) \stackrel{d}{=} (\mathbf{A}_1, \mathbf{B}_1)$. In addition, suppose that for $\mathbf{z} \in S_+$,*

$$E\mathbf{X} < \infty, \quad E\|\mathbf{z}'\mathbf{A}\|^\alpha \mathbf{z}'\mathbf{B} < \infty, \tag{2.104}$$

$$x^\alpha E(\mathbf{z}'\mathbf{A} 1_{\{\mathbf{z}'\mathbf{B} > Cx\}}) \to \mathbf{0}, \quad \text{for any } C > 0, \quad \text{as } x \to \infty, \tag{2.105}$$

$$x^\alpha E(\mathbf{z}'\mathbf{A} 1_{\{\|\mathbf{z}'\mathbf{A}\| > Cx\}}) \to \mathbf{0}, \quad \text{for any } C > 0, \quad \text{as } x \to \infty, \tag{2.106}$$

$$x^\alpha \int_x^\infty P(\mathbf{z}'\mathbf{B} > u) du \to C_+(\mathbf{z}), \tag{2.107}$$

for some function $C_+(\mathbf{z}) \geq 0$. Assume also that either

$$\mathbf{z}'\mathbf{A} \text{ is a discrete, finite random vector, or} \tag{2.108}$$

$$P\left((\mathbf{z}'\mathbf{A}^{\#}) \in E\right) \to 0, \text{ as } \mu(E) \to 0. \tag{2.109}$$

Then, the stationary solution $\mathbf{X}$ *of (2.101) satisfies*

$$\int_x^\infty \big(P(\mathbf{z}'\mathbf{X} > u) - P(\mathbf{z}'\mathbf{A}\mathbf{X} > u)\big)\, du \sim x^{-\alpha}\left(C_+(\mathbf{z}) + c_+ E\left(r(\mathbf{z}'\mathbf{A}^{\#})\|\mathbf{z}'\mathbf{A}\|^{\alpha}\mathbf{z}'\mathbf{B}\right)\right). \quad (2.110)$$

Proof. Observe that

$$\int_x^\infty \big(P(\mathbf{z}'\mathbf{X} > u) - P(\mathbf{z}'\mathbf{A}\mathbf{X} > u)\big)\, du$$

$$= \int_x^\infty \int_{\mathbf{a},\mathbf{b}} \big(P((\mathbf{z}'\mathbf{a})\mathbf{X} + \mathbf{z}'\mathbf{b} > u) - P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)\big)\, F_{\mathbf{A},\mathbf{B}}(d\mathbf{a}, d\mathbf{b})du$$

$$= \int_{\mathbf{a},\mathbf{b}} \int_{x-\mathbf{z}'\mathbf{b}}^{x} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)duF_{\mathbf{A},\mathbf{B}}(d\mathbf{a}, d\mathbf{b}) = \int_{\mathbf{a},\overline{\mathbf{b}}} \int_{x-\overline{\mathbf{b}}}^{x} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}),$$

where $\overline{\mathbf{b}} = \mathbf{z}'\mathbf{b}$ is a scalar and $\overline{\mathbf{B}} = \mathbf{z}'\mathbf{B}$. By splitting the range of $\overline{\mathbf{b}}$ with fixed $c < 1$, this can further be written as

$$\int_{cx}^\infty \int_{\mathbf{a}} \int_{x-\overline{\mathbf{b}}}^{x} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}})$$

$$+ \int_0^{cx} \int_{\mathbf{a}} \int_{x-\overline{\mathbf{b}}}^{x} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) =: I + L.$$

For the integral I, write it as $I = I_1 + I_2$, where

$$I_1 = \int_{cx}^{x} \int_{\mathbf{a}} \int_{x-\overline{\mathbf{b}}}^{x} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}),$$

$$I_2 = \int_{x}^{\infty} \int_{\mathbf{a}} \int_{x-\overline{\mathbf{b}}}^{x} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}).$$

For the integral I_1, we have

$$x^\alpha I_1 \le x^\alpha \int_{cx}^{x} \int_{\mathbf{a}} E(\mathbf{z}'\mathbf{a}\mathbf{X})F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) \le x^\alpha E\left(\mathbf{z}'\mathbf{A}1_{\{\overline{\mathbf{B}} > cx\}}\right) E\mathbf{X} \to 0, \quad (2.111)$$

by (2.105).

Observe for I_2 that

$$I_2 = \int_{x}^{\infty} \int_{\mathbf{a}} \left\{ \int_{x-\overline{\mathbf{b}}}^{0} P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)du + \int_0^x P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)du \right\} F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}})$$

$$= \int_{x}^{\infty} \int_{\mathbf{a}} \left\{ (\overline{\mathbf{b}} - x) + \int_0^x P((\mathbf{z}'\mathbf{a})\mathbf{X} > u)du \right\} F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) =: I_{2,1} + I_{2,2}.$$

Note that, by (2.107),

$$x^\alpha I_{2,1} = x^\alpha E(\mathbf{z}'\mathbf{B} - x)_+ = x^\alpha \int_x^\infty P(\mathbf{z}'\mathbf{B} > u)du \to C_+(\mathbf{z}). \tag{2.112}$$

Also note from (2.105) with $C = 1$ that

$$x^\alpha I_{2,2} \le x^\alpha E\left(\mathbf{z}'\mathbf{A}1_{\{\mathbf{z}'\mathbf{B}>x\}}\right) E\mathbf{X} \to 0. \tag{2.113}$$

Combining (2.111), (2.112) and (2.113) yields

$$x^\alpha I \to C_+(\mathbf{z}). \tag{2.114}$$

For the integral L, for some x_0 to be determined later, write it as

$$L = \int_0^{cx} \int_{\mathbf{a}} \int_{x-\overline{\mathbf{b}}}^{x} P(\|\mathbf{z}'\mathbf{a}\|(\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}})$$

$$= \int_0^{cx} \int_{\|\mathbf{z}'\mathbf{a}\|>(x-\overline{\mathbf{b}})/x_0} \int_{x-\overline{\mathbf{b}}}^{x} P(\|\mathbf{z}'\mathbf{a}\|(\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}})$$

$$+ \int_0^{cx} \int_{\|\mathbf{z}'\mathbf{a}\|\le(x-\overline{\mathbf{b}})/x_0} \int_{x-\overline{\mathbf{b}}}^{x} P(\|\mathbf{z}'\mathbf{a}\|(\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} > u)duF_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) =: J + K.$$

For fixed x_0, observe that

$$x^\alpha J \le x^\alpha E\left(\mathbf{z}'\mathbf{A}1_{\{\|\mathbf{z}'\mathbf{A}\|>(1-c)x/x_0\}}\right) E\mathbf{X} \to 0, \tag{2.115}$$

using (2.106), since $\overline{\mathbf{b}} \in (0, cx)$ implies $(x - \overline{\mathbf{b}})/x_0 > (1 - c)x/x_0$.

We now show how one can deal with the integral K. We consider only the case (2.109). (The case (2.108) is easier and can be proved as below.) Let μ denote a Haar measure on S_+ as in the statement of the theorem. Since $\mu(S_+) < \infty$, Theorem 2.4.1 and Egoroff's theorem imply that, for any $\epsilon > 0$, there is $E_\epsilon \subset S_+$ such that $\mu(E_\epsilon) < \epsilon$ and

$$\sup_{\mathbf{w}\in S_+\setminus E_\epsilon} \left|x^\alpha P(\mathbf{w}'\mathbf{X} > x) - c_+ r(\mathbf{w})\right| \to 0, \tag{2.116}$$

as $x \to \infty$. Now write the integral K as

$$K = \int_0^{cx} \int_{\|\mathbf{z}'\mathbf{a}\| \le (x-\overline{\mathbf{b}})/x_0} \int_{x-\overline{\mathbf{b}}}^{x} P(\|\mathbf{z}'\mathbf{a}\|(\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} > u) du F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}})$$

$$\cdot \left(1_{\{(\mathbf{z}'\mathbf{a}^{\#}) \in E_\epsilon\}} + 1_{\{(\mathbf{z}'\mathbf{a}^{\#}) \in S_+ \backslash E_\epsilon\}} \right) =: K_1 + K_2. \tag{2.117}$$

For K_2, observe that

$$x^\alpha K_{2,1} := x^\alpha \int_0^{cx} \int_F \overline{\mathbf{b}} P(\|\mathbf{z}'\mathbf{a}\|(\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} > x) F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) \le x^\alpha K_2$$

$$\le x^\alpha \int_0^{cx} \int_F \overline{\mathbf{b}} P(\|\mathbf{z}'\mathbf{a}\|(\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} > x - \overline{\mathbf{b}}) F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) =: x^\alpha K_{2,2},$$

where $F = \{\mathbf{a} : \|\mathbf{z}'\mathbf{a}\| \le (x - \overline{\mathbf{b}})/x_0, (\mathbf{z}'\mathbf{a}^{\#}) \in S_+ \backslash E_\epsilon\}$. Write the integral $x^\alpha K_{2,2}$ as

$$x^\alpha K_{2,2} = \int_0^{cx} \int_F \overline{\mathbf{b}} c_+ r(\mathbf{z}'\mathbf{a}^{\#}) \|\mathbf{z}'\mathbf{a}\|^\alpha \left(1 - \frac{\overline{\mathbf{b}}}{x}\right)^{-\alpha} \frac{P\left((\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} \ge \frac{x-\overline{\mathbf{b}}}{\|\mathbf{z}'\mathbf{a}\|}\right)}{c_+ r(\mathbf{z}'\mathbf{a}^{\#}) \left(\frac{x-\overline{\mathbf{b}}}{\|\mathbf{z}'\mathbf{a}\|}\right)^{-\alpha}} F_{\mathbf{A},\overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}).$$

By condition (2.116), the term

$$\left(1 - \frac{\overline{\mathbf{b}}}{x}\right)^{-\alpha} \frac{P\left((\mathbf{z}'\mathbf{a}^{\#})\mathbf{X} \ge \frac{x-\overline{\mathbf{b}}}{\|\mathbf{z}'\mathbf{a}\|}\right)}{c_+ r(\mathbf{z}'\mathbf{a}^{\#}) \left(\frac{x-\overline{\mathbf{b}}}{\|\mathbf{z}'\mathbf{a}\|}\right)^{-\alpha}}$$

is bounded on F for large enough x_0, and converges to 1 as $x \to \infty$. The dominated convergence theorem implies that

$$x^\alpha K_{2,2} \to c_+ E r(\mathbf{z}'\mathbf{A}^{\#}) \|\mathbf{z}'\mathbf{A}\|^\alpha 1_{\{(\mathbf{z}'\mathbf{A}^{\#}) \in S_+ \backslash E_\epsilon\}} \mathbf{z}'\mathbf{B}.$$

The same asymptotics holds for $x^\alpha K_{2,1}$, and we can conclude that

$$x^\alpha K_2 \to c_+ E r(\mathbf{z}'\mathbf{A}^{\#}) \|\mathbf{z}'\mathbf{A}\|^\alpha 1_{\{(\mathbf{z}'\mathbf{A}^{\#}) \in S_+ \backslash E_\epsilon\}} \mathbf{z}'\mathbf{B}. \tag{2.118}$$

For K_1, observe that

$$x^\alpha K_1 \leq x^\alpha \int_0^{cx} \int_{\|\mathbf{z}'\mathbf{A}\| \leq (x-\overline{\mathbf{b}})/x_0, (\mathbf{z}'\mathbf{a}^\#) \in E_\epsilon} \overline{\mathbf{b}} P \left(\|\mathbf{z}'\mathbf{a}\| \sqrt{d} \mathbf{z}_0' \mathbf{X} > x - \overline{\mathbf{b}} \right) F_{\mathbf{A}, \overline{\mathbf{B}}}(d\mathbf{a}, d\overline{\mathbf{b}}) =: x^\alpha K_{1,1}, \tag{2.119}$$

where $\mathbf{z}_0 = (1, \ldots, 1)/\sqrt{d} \in S_+$. The argument as above yields in the same way that

$$x^\alpha K_{1,1} \to r(\mathbf{z}_0) d^{\alpha/2} E \|\mathbf{z}'\mathbf{A}\|^\alpha 1_{\{(\mathbf{z}'\mathbf{A}^\#) \in E_\epsilon\}} \mathbf{z}'\mathbf{B}. \tag{2.120}$$

Using assumption (2.109), since ϵ is arbitrarily small, we conclude from (2.118), (2.119) and (2.120) that

$$x^\alpha K \to c_+ E r(\mathbf{z}'\mathbf{A}^\#) \|\mathbf{z}'\mathbf{A}\|^\alpha \mathbf{z}'\mathbf{B}. \tag{2.121}$$

The conclusion follows from (2.114), (2.115) and (2.121). $\square$

Example 2.4.1. Generalized autoregressive conditionally heteroscedastic process $\{\xi_t\}_{t \in \mathbb{Z}}$ of order (p, q) with $p, q \geq 0$ (GARCH(p, q)) is given as

$$\xi_t = \sigma_t \epsilon_t, \tag{2.122}$$

$$\sigma_t^2 = \beta + \sum_{i=1}^p \lambda_i \xi_{t-i}^2 + \sum_{j=1}^q \phi_j \sigma_{t-j}^2, \tag{2.123}$$

where $\{\epsilon_t\}$ are i.i.d. normal random variables, and $\beta > 0$, $\lambda_i \geq 0$, $\phi_i \geq 0$, with the convention that $\lambda_p > 0$ if $p \geq 1$ and $\phi_q > 0$ if $q \geq 1$. (See, for example, Bollerslev (1986) and Embrechts et al. (1997)). The squares of the GARCH model can be expressed as a multidimensional RDE

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t, \tag{2.124}$$

where

$$\mathbf{X}_t = (\xi_t^2, \xi_{t-1}^2, \ldots, \xi_{t-p+2}^2, \xi_{t-p+1}^2, \sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-q+2}^2, \sigma_{t-q+1}^2)',$$

$$\mathbf{A}_t = \begin{pmatrix} \lambda_1\epsilon_t^2 & \lambda_2\epsilon_t^2 & \dots & \lambda_{p-1}\epsilon_t^2 & \lambda_p\epsilon_t^2 & \phi_1\epsilon_t^2 & \phi_2\epsilon_t^2 & \dots & \phi_{q-1}\epsilon_t^2 & \phi_q\epsilon_t^2 \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ \lambda_1 & \lambda_2 & \dots & \lambda_{p-1} & \lambda_p & \phi_1 & \phi_2 & \dots & \phi_{q-1} & \phi_q \\ 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix},$$

$$\mathbf{B}_t = (\beta\epsilon_t^2, 0, \dots, 0, 0, \beta, 0, \dots, 0, 0)'.$$

(The matrix $\mathbf{A}_t$ has to be interpreted with care when either p or q is zero. In this case, one should take $p = 1$ and $\lambda_1 = 0$ or $q = 1$ and $\phi_1 = 0$ respectively.) It can be seen that assumption (2.108) or (2.109) holds for the squares of a GARCH process with continuous innovations ϵ_t.

CHAPTER 3

Estimation of parameters in heavy-tailed distribution when its second-order tail parameter is known

3.1 Introduction

Heavy tails refer to a slow, power-like decay of a tail of a distribution function. This phenomenon is observed in a wide range of applications, for example, distributions of log-returns in Finance, transmitted file sizes (in packets) in Telecommunications to name but a few (see, for example, Embrechts, Klüppelberg and Mikosch (1997), de Haan and Ferreira (2006), Leadbetter, Lindgren and Rootzén (1983), Resnick (1997) and Reiss and Thomas (2007)). In mathematical terms, a heavy-tailed distribution of a random variable X is supposed to have a regularly varying tail in the sense that

$$\overline{F}(x) = P(X > x) = L(x)x^{-\alpha}, \quad \alpha > 0, \tag{3.1}$$

where $L(x)$ is a slowly varying function at infinity (see, for example, Bingham, Goldie and Teugels (1989)). The parameter α is called a tail exponent. From an estimation perspective, (3.1) is typically replaced by

$$\overline{F}(x) = P(X > x) = c_1 x^{-\alpha} + o(x^{-\alpha}), \quad \alpha > 0,\ c_1 > 0, \tag{3.2}$$

as $x \to \infty$, that is, by modelling the slowly varying function L in (3.1) as a constant c_1. Note that (3.2) implies $\log \overline{F}(x) \approx \log c_1 - \alpha \log x$ for large x. If given n i.i.d. observations X_i, $i = 1, \ldots, n$, of X, consider their order statistics $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n-1)} \leq X_{(n)}$. Since

$\overline{F}(X_{(n-i+1)}) \approx i/n$, the above suggests that

$$\log\left(\frac{i}{n}\right) \approx \log c_1 - \alpha \log X_{(n-i+1)} \tag{3.3}$$

and hence that α, in particular, can be estimated through least squares of $\log(i/n)$ on log $X_{(n-i+1)}$, $i = 1, \ldots, k$, where k denotes a threshold. The resulting estimator $\widehat{\alpha}_{RK}$ is referred to here as the rank-based estimator (though there is no widely accepted terminology). It is popular in applied literature (see, for example, Gabaix and Ibragimov (2007) and references therein) and has been considered more rigorously in, for example, Csörgő and Viharos (1997). The relation (3.3) also provides a simple way to detect heavy tails, namely, by considering a log-log plot of empirical distribution function against order statistics.

It is well known that (3.2) yields

$$F^{\leftarrow}\left(1 - y^{-1}\right) = (c_1 y)^{1/\alpha} + o(y^{1/\alpha}), \tag{3.4}$$

as $y \to \infty$, where $F^{\leftarrow}(z) = \inf\{x : F(x) \geq z\}$ is an inverse of the distribution function F. Hence, one similarly expects that

$$\log X_{(n-i+1)} \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log\left(\frac{i}{n}\right) \tag{3.5}$$

and that α, in particular, can be estimated through least squares of $\log X_{(n-i+1)}$ on $\log(i/n)$, $i = 1, \ldots, k$. The resulting estimator $\widehat{\alpha}_{QQ}$ is called the QQ-estimator and has been studied in detail by Kratz and Resnick (1996).

The rank-based and QQ estimators are not most efficient (supposing the strict power-tail or strict Pareto distribution) and more efficient estimators for tail exponent have been studied by many authors. One such popular estimator $\widehat{\alpha}_H$, the Hill estimator after Hill (1975) is defined as

$$\widehat{\alpha}_H^{-1} = \frac{1}{k}\sum_{i=1}^{k}(\log X_{(n-i+1)} - \log X_{(n-k)}) = \frac{1}{k}\sum_{i=1}^{k} i(\log X_{(n-i+1)} - \log X_{(n-i)}) =: \frac{1}{k}\sum_{i=1}^{k} U_i, \tag{3.6}$$

where k, as above, is the number of upper order statistics. The log-spacings of order statistics

$U_i = i(\log X_{(n-i+1)} - \log X_{(n-i)})$ are approximately (exactly for strict Pareto) i.i.d. exponential variables with mean $1/\alpha$ and (3.6) can be seen as its (conditional) maximum likelihood estimator. Some further relationships among QQ, Hill and the so-called kernel Hill estimators can be found in Beirlant, Vynckier and Teugels (1996), Aban and Meerschaert (2004) and others.

The above estimators work well and are designed for distributions that are strict or very close to those of strict Pareto. It is well known, however, in both theory and practice, that estimation can be seriously biased when distribution deviates from that of strict Pareto (see, for example, Martins, Gomes and Neves (1999)). A common way to quantify and to model such deviations is through the so-called second-order regular variation (see, for example, de Haan and Ferreira (2006)). The distribution tail $\overline{F}(x) = P(X > x)$ is second-order regularly varying with first-order tail parameter $\alpha > 0$ and second-order tail parameter $\rho < 0$ (denoted as $\overline{F} \in 2RV(-\alpha, \rho)$) if there is a suitable function $g(x)$ such that, for any $a > 0$,

$$\lim_{x\to\infty} \frac{x^{\alpha}\overline{F}(x) - (ax)^{\alpha}\overline{F}(ax)}{g(x)} = c\frac{a^{\rho}-1}{\rho}. \tag{3.7}$$

(The relation (3.1) is referred to as first-order regular variation.) From a practical (modelling) perspective, the condition (3.7) is often replaced by

$$\overline{F}(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha+\rho} + o(x^{-\alpha+\rho}), \quad \text{as } x \to \infty, \tag{3.8}$$

where $c_1 > 0$ and $\rho < 0$. The collection of distributions (3.8) is known as the class of Hall and Welsh (1985). This class and more generally (3.7) also naturally appear when proving asymptotic normality of the above estimators. These asymptotic normality results already indicate the presence of bias through nonzero mean in the limiting normal distributions.

In order to reduce bias, various authors proposed bias reduced estimators under the assumption (3.8). For example, Feuerverger and Hall (1999) and Beirlant, Dierckx, Guillou and Stărică (2002) approximated log-spacings of order statistics by normalized exponential distribution and derived estimators based on the maximum likelihood or regression with exponential responses. The generalized jackknife estimators accommodating bias are studied by Gomes, Martins and Neves (2000) and Gomes and Martins (2002), and asymptotically best linear unbiased estimator is proposed by Gomes, Figueiredo and Mendonça (2005). A nice extensive review of this research direction can be found in Reiss and Thomas (2007), Chapter 6.

Bias reduced estimators discussed above assume (3.8) with unknown second-order tail parameter ρ (and unknown α, c_1 and c_2). In this chapter, we are interested in estimation methods when the second-order tail parameter ρ is *known*. Note first that we may suppose without loss of generality that $\rho = -1$, that is,

$$\overline{F}(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1} + o(x^{-\alpha-1}), \quad \text{as } x \to \infty. \tag{3.9}$$

Indeed, if the distribution tail of X follows (3.8) with known $\rho < 0$, then one may consider instead the variable $X^{-\rho}$ which follows (3.9) (with α replaced by $-\alpha/\rho$, in particular).

Considering the specification (3.9), we are interested in possible ways to estimate unknown parameters in (3.9), basic properties of resulting estimators and comparison of possible estimators. A range of estimators is obviously possible adapting (taking $\rho = -1$ in) available estimation methods that suppose unknown ρ, for example, maximum likelihood estimator of Feuerverger and Hall (1999), and will be considered below. On the other hand, the specification (3.9) also suggests other simple estimators. For example, note that (3.9) implies that

$$\log\left(\frac{i}{n}\right) \approx \log c_1 - \alpha \log X_{(n-i+1)} + \frac{c_2}{c_1} X^{-1}_{(n-i+1)}. \tag{3.10}$$

In particular, the tail exponent α can be estimated by regressing $\log(i/n)$ on $(1, \log X_{(n-i+1)}, X^{-1}_{(n-i+1)})$, generalizing the rank-based estimator based on (3.3). The analogue of the QQ-estimator can also be introduced, and is considered below.

Remark. Note that, for the second-order tail parameter, we use the convention found in, for example, Smith (1987), Feuerverger and Hall (1999), Embrechts et al. (1997), p. 341. An alternative popular specification (see the papers by Beirlant, Gomes and others) is given by

$$\overline{F}(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha(1-\rho^*)} + o(x^{-\alpha(1-\rho^*)}) = c_1 x^{-\alpha}(1 + c_2/c_1 x^{\alpha\rho^*} + o(x^{\alpha\rho^*})). \tag{3.11}$$

The connection between (3.8) and (3.11) is

$$\rho^* = \frac{\rho}{\alpha}. \tag{3.12}$$

In the specification (3.11), we therefore assume that $\rho^* = -1/\alpha$. In particular, note that, with this specification, ρ^* is unknown! On another hand, estimation methods presented below for known $\rho = -1$ can also be adapted to the case of known ρ^*, say $\rho^* = -1$. This is briefly discussed in Section 3.2.7 below.

The assumption of known ρ is largely motivated by the result of Chapter 2, namely second-order properties of tails of the so-called random difference equations or RDEs for short. For example, the widely studied ARCH models are RDEs. In one dimension, a stationary solution X of RDE satisfies the relation $X \stackrel{d}{=} AX + B$, where (A, B) is a vector of variables that are independent of X. Since a celebrated result of Kesten (1973), it is well-known that, under mild assumptions, RDE X has a power-law tail with the tail exponent α determined by the variable A, namely, $EA^\alpha = 1$. In Theorem 2.3.1 of Chapter 2, it is shown that, under mild assumptions on RDE X,

$$\int_x^\infty (P(X > u) - P(AX > u))du \sim cx^{-\alpha}, \quad c > 0, \tag{3.13}$$

as $x \to \infty$. This can be considered as a weaker from of 2RV (3.7) in that (3.13) suggests (note from above that $EA^\alpha = 1$)

$$\lim_{x\to\infty} \frac{x^\alpha P(X > x) - x^\alpha (EA^\alpha)^{-1} P(X > A^{-1}x)}{x^{-1}} = c\alpha, \tag{3.14}$$

which can be viewed as (3.7) at random $a = A^{-1}$ (and, in fact, $\rho = -1$ as suggested by the term x^{-1} in the denominator in (3.14)). Moreover as a consequence of (3.13), it is shown in Proposition 2.3.1 of Chapter 2 that, under mild assumptions, RDE X can have the form (3.8), the only form for bias reduction used in practice, only when $\rho = -1$ and also $c_2 < 0$.

Existence of a second-order term in distribution tail does not, by itself, imply that significant estimation bias is present from a practical perspective (that is, for example, a second-order term could be "too far" in the tail from a practical perspective). In the case of RDEs, however, the estimation bias of, for example, Hill estimator appears extremely large for larger values of tail exponents α (say $\alpha = 5$ or 10). Taking the second order term with $\rho = -1$ in (3.9) into account leads to a satisfactory bias correction.

Considering known ρ should also not be surprising from the following angle. In the specification (3.11), it has been common to pay particular attention to the special case supposing known $\rho^* = -1$ (corresponding to unknown $\rho = -\alpha$ under the framework (3.9)). For example, a number of distributions such as symmetric α-stable ($\alpha \in (1, 2)$) and Fréchet have $\rho^* = -1$, and several estimators are designed specially for this case (see, for example, Gomes and Martins (2002), Gomes et al. (2005) and Section 3.2.7 below). Let us also add that popular generalized Pareto distributions (GPD) and generalized extreme value distributions (GEVD) also have a known second-order tail parameter $\rho = -1$, as in (3.9) (see Section 3.5.1 below).

The structure of Chapter 3 is as follows. In Section 3.2, we gather a number of possible estimators under the model (3.9). In Section 3.3, some of their properties are proved, focusing on estimators of the tail exponent α. In Section 3.4, we provide a simulation study comparing the proposed estimators. Several further issues are discussed in Section 3.5. Conclusions can be found in Section 3.6.

3.2 Estimation methods

In this section, we gather a number of possible estimators of parameters in the framework (3.9). Several estimators are based on least squares methods (Sections 3.2.1-3.2.3), generalized jackknife (Section 3.2.4) and others are maximum likelihood estimators (Sections 3.2.5-3.2.6). The estimators of Section 3.2.1, 3.2.2 and 3.2.6 depend particularly on the specific form (3.9) and hence could be considered new. The estimators of Sections 3.2.4 and 3.2.5, on the other hand, are rather adaptations of estimators available in the case when ρ is unknown. Finally, in Section 3.2.7, we briefly discuss estimation in the framework (3.11) assuming known $\rho^* = -1$.

3.2.1 Rank-based, least squares estimators

From (3.9), observe that

$$\overline{F}(x) = P(X > x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1} + o(x^{-\alpha-1}) = c_1 x^{-\alpha}\left(1 + \frac{c_2}{c_1} x^{-1}\right) + o(x^{-\alpha-1}), \quad (3.15)$$

as $x \to \infty$. By taking the logarithm, as $x \to \infty$,

$$\log(\overline{F}(x)) \approx \log c_1 - \alpha \log x + \log\left(1 + \frac{c_2}{c_1} x^{-1}\right) \approx \log c_1 - \alpha \log x + \frac{c_2}{c_1} x^{-1}. \tag{3.16}$$

Therefore, one expects that parameters α, c_1 and c_2 could be estimated by a linear regression of the logarithm of empirical distribution tail of X on $(1, \log x, 1/x)$, that is,

$$\underset{\beta_0,\alpha,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{k} \left(\log(i/n) - \beta_0 + \alpha \log X_{(n-i+1)} - \beta_1 / X_{(n-i+1)}\right)^2, \tag{3.17}$$

where $\beta_0 = \log c_1$ and $\beta_1 = c_2/c_1$. This approach generalizes the least squares tail estimator based on the first-order asymptotics (3.3). We denote the corresponding tail exponent estimator as $\widehat{\alpha}_{RK2}$.

Equivalently, problem (3.17) can be written as

$$\underset{\beta_0,\alpha,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{k} \left(\log\left(\frac{i}{k}\right) - \beta_0 + \alpha \log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \beta_1 \left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\right)^2, \tag{3.18}$$

where $\beta_0 = \log c_1 - \log(k/n) - \alpha \log X_{(n-k)}$, $\beta_1 = c_2/(c_1 X_{(n-k)})$. The tail exponent estimator, in particular, can be expressed as

$$\widehat{\alpha}_{RK2} = -\frac{A - B}{C - D}, \tag{3.19}$$

where

$$\begin{aligned} A &= \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2 - \left(\frac{1}{k}\sum_{i=1}^{k}\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2\right) \times \\ &\quad \left(\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right)\log\left(\frac{i}{k}\right) - \frac{1}{k^2}\sum_{i=1}^{k}\log\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\sum_{i=1}^{k}\log\left(\frac{i}{k}\right)\right), \end{aligned} \tag{3.20}$$

$$\begin{aligned} B &= \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \frac{1}{k^2}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\sum_{i=1}^{k}\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right)\right) \times \\ &\quad \left(\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{i}{k}\right)\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right) - \frac{1}{k^2}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\sum_{i=1}^{k}\log\left(\frac{i}{k}\right)\right), \end{aligned} \tag{3.21}$$

$$C = \left(\frac{1}{k}\sum_{i=1}^{k}\log^2\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right)\right)^2\right) \times$$

$$\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2 - \left(\frac{1}{k}\sum_{i=1}^{k}\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2\right), \tag{3.22}$$

$$D = \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \frac{1}{k^2}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\sum_{i=1}^{k}\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right)\right)^2. \tag{3.23}$$

Remark. Applying continuity correction for empirical distribution, namely replacing $\log(i/k)$ by $\log((i-.5)/k)$ in (3.18) improves estimation in smaller samples. In the context of (3.3), similar correction is well-known and is justified recently in Gabaix and Ibragimov (2007).

3.2.2 Analogues of the QQ-estimator

The estimators below can be viewed as generalizations of the QQ-estimators based on (3.5). Observe from (3.15) that an inverse function of $F(x)$ satisfies

$$F^{\leftarrow}\left(1-\frac{1}{y}\right) = (c_1 y)^{1/\alpha}\left(1+\frac{c_2}{\alpha c_1}(c_1 y)^{-1/\alpha}\right) + o(1), \quad y \to \infty \tag{3.24}$$

(this can be seen by making the change of variables, $x = 1 - c_1 y^{\alpha} - c_2 y^{\alpha+1}$ in (3.9) and using the approximation $(1+z)^{-\gamma} = 1 - \gamma z + o(z)$, as $z \to 0$). By replacing $1/y$ by i/n and taking the logarithm in (3.24) gives

$$\log X_{(n-i+1)} \approx \frac{1}{\alpha}\log c_1 - \frac{1}{\alpha}\log\left(\frac{i}{n}\right) + \log\left(1+\frac{c_2}{\alpha c_1}\left(c_1\frac{n}{i}\right)^{-1/\alpha}\right).$$

This suggests nonlinear least squares estimation

$$\underset{\beta_0,\beta_1,\alpha}{\operatorname{argmin}}\sum_{i=1}^{k}\left(\log X_{(n-i+1)} - \beta_0 + \frac{1}{\alpha}\log\left(\frac{i}{n}\right) - \log\left(1+\beta_1\left(\frac{i}{n}\right)^{1/\alpha}\right)\right)^2, \tag{3.25}$$

where $\beta_0 = \log c_1/\alpha$ and $\beta_1 = c_2/(\alpha c_1^{1+1/\alpha})$. The nonlinear minimization (3.25) can be reduced to that over β_1 and α. We will denote the tail exponent estimator based on (3.25) as $\widehat{\alpha}_{QQn}$.

Furthermore, observe from (3.24) that

$$F^{\leftarrow}\left(1-\frac{1}{y}\right) - \frac{c_2}{\alpha c_1} = (c_1 y)^{1/\alpha} + o(1). \tag{3.26}$$

Taking the logarithm gives

$$\log\left(X_{(n-i+1)} - \frac{c_2}{\alpha c_1}\right) \approx \frac{1}{\alpha}\log c_1 - \frac{1}{\alpha}\log\left(\frac{i}{n}\right) \tag{3.27}$$

and approximating the left-hand side yields

$$\log X_{(n-i+1)} \approx \frac{1}{\alpha}\log c_1 - \frac{1}{\alpha}\log\left(\frac{i}{n}\right) + \frac{c_2}{\alpha c_1}\frac{1}{X_{(n-i+1)}}. \tag{3.28}$$

Relation (3.28) suggests linear regression estimators through

$$\operatorname*{argmin}_{\beta_0,\beta_1,\alpha}\sum_{i=1}^{k}\left(\log X_{(n-i+1)} - \beta_0 + \frac{1}{\alpha}\log\left(\frac{i}{n}\right) - \beta_1\frac{1}{X_{(n-i+1)}}\right)^2,$$

where $\beta_0 = \log c_1/\alpha$, $\beta_1 = c_2/(\alpha c_1)$, or equivalently

$$\operatorname*{argmin}_{\beta_0,\beta_1,\alpha}\sum_{i=1}^{k}\left(\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \beta_0 + \frac{1}{\alpha}\log\left(\frac{i}{k}\right) - \beta_1\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2,$$

where $\beta_0 = \log c_1/\alpha - \log X_{(n-k)} - \log(k/n)/\alpha$, $\beta_1 = c_2/(\alpha c_1 X_{(n-k)})$. This is, in fact, the inverse regression of (3.18) by reversing the roles of log order statistics and empirical distribution. We denote the corresponding tail exponent estimator as $\widehat{\alpha}_{QQ2}$. Similar calculation as for (3.19) (or just reversing the roles of $\log X_{(n-i+1)}$ and $\log(i/k)$) gives an explicit form of $\widehat{\alpha}_{QQ2}$ as

$$\widehat{\alpha}_{QQ2}^{-1} = -\frac{A-B}{E-F}, \tag{3.29}$$

where A and B are given in (3.20) and (3.21), and

$$E = \left(\frac{1}{k}\sum_{i=1}^{k}\log^2\left(\frac{i}{k}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{i}{k}\right)\right)^2\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2 - \left(\frac{1}{k}\sum_{i=1}^{k}\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)^2\right), \tag{3.30}$$

$$F = \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-i+1)}}\right)\log\left(\frac{i}{k}\right) - \frac{1}{k^2}\sum_{i=1}^{k}\frac{X_{(n-k)}}{X_{(n-i+1)}}\sum_{i=1}^{k}\log\left(\frac{i}{k}\right)\right)^2. \tag{3.31}$$

3.2.3 Generalized least squares methods

In the first-order regular variation framework, Aban and Meerschaert (2004) examined the generalized least squares estimator based on (3.5). We briefly comment here on similar ideas, for the second-order framework (3.9). Recall that

$$-\log(1-F(X_{(n-i+1)})) \stackrel{d}{=} \frac{e_1}{n} + \frac{e_2}{n-1} + \ldots + \frac{e_{n+1-i}}{i} =: Y_{(n-i+1)},$$

where e_i are i.i.d. exponential random variables with mean 1. Based on the assumption (3.9), this leads to

$$-\log c_1 + \alpha \log X_{(n-i+1)} - \frac{c_2}{c_1}\frac{1}{X_{(n-i+1)}}$$

approximately having the same distribution as $Y_{(n-i+1)}$. The expectation and covariance of $Y_{(n-i+1)}$ are calculated as

$$\nu_i := EY_{(n-i+1)} = \frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{i},$$

$$\Sigma_{ij} := \mathrm{Cov}(Y_{(n-i+1)}, Y_{(n-j+1)}) = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{j^2}, \quad j \geq i.$$

This suggests generalized least squares estimator for known $\beta_1 = c_2/(\alpha c_1)$ as

$$\operatorname*{argmin}_{\beta_0,\alpha} \left(\log \mathbf{X} - \beta_1 1/\mathbf{X} - \beta_0 - \frac{1}{\alpha}E\mathbf{Y}\right)^T \alpha^2 \Sigma^{-1} \left(\log \mathbf{X} - \beta_1 1/\mathbf{X} - \beta_0 - \frac{1}{\alpha}E\mathbf{Y}\right), \qquad (3.32)$$

where $\log \mathbf{X} = (\log X_{(n)}\ \log X_{(n-1)}\ \ldots\ \log X_{(n-k+1)})^T$, $1/\mathbf{X} = (X_{(n)}^{-1}\ X_{(n-1)}^{-1}\ \ldots\ X_{(n-k+1)}^{-1})^T$, $E\mathbf{Y}^T = (\nu_1, \nu_2, \ldots, \nu_k)$ and $\beta_0 = \log c_1/\alpha$. The corresponding tail exponent estimator $\widehat{\alpha}_{GLS}$ can be expressed as in equation (2.9) of Aban and Meerschaert (2004),

$$\widehat{\alpha}_{GLS}^{-1} = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{(n-i+1)} - X_{(n-k+1)} - \beta_1\left(X_{(n-i+1)}^{-1} - X_{(n-k+1)}^{-1}\right)\right)$$

$$= \widehat{\alpha}_H^{-1} - \frac{\beta_1}{k}\sum_{i=1}^{k}\left(X_{(n-i+1)}^{-1} - X_{(n-k+1)}^{-1}\right). \qquad (3.33)$$

This estimator has nice properties according to Aban and Meerschaert (2004). However, in practice, β_1 needs to be estimated. One possible approach is to use β_1 from QQ2 estimator in Section 3.2.2, which is a GLS estimator with identity covariance matrix.

Remark. Assuming β_1 is unknown in (3.32) yields inconsistent estimators of tail exponent α. The same happens if one tried the generalized least squares regression

$$\underset{\beta_0,\alpha,\beta_1}{\operatorname{argmin}} \left(-E(\mathbf{Y}) - \beta_0 + \alpha \log \mathbf{X} - \beta_1 1/\mathbf{X}\right)^T \Sigma^{-1} \left(-E(\mathbf{Y}) - \beta_0 + \alpha \log \mathbf{X} - \beta_1 1/\mathbf{X}\right), \tag{3.34}$$

based on (3.17). The reason for inconsistency is that, for example, (3.17) is not a standard regression problem in that, for example, response variables are constant. (Moreover, there is nothing special here about the second-order framework: no consistency would be possible through (3.34) with $\beta_1 = 0$ in the first-order framework.)

3.2.4 Generalized jackknife estimators

We introduce here several generalized jackknife estimators accommodating bias. Following Gomes et al. (2000), consider two biased estimators $\widehat{\alpha}^{(1)}(k)$ and $\widehat{\alpha}^{(2)}(k)$ of α such that

$$E\widehat{\alpha}^{(1)}(k) = \alpha + \phi(\alpha)d_1(k), \quad E\widehat{\alpha}^{(2)}(k) = \alpha + \phi(\alpha)d_2(k).$$

The generalized jackknife estimator of α based on $\widehat{\alpha}^{(1)}(k)$ and $\widehat{\alpha}^{(2)}(k)$ is defined as

$$\widehat{\alpha}^{(G)}(\widehat{\alpha}^{(1)}, \widehat{\alpha}^{(2)}) = \frac{\widehat{\alpha}^{(1)} - q_k \widehat{\alpha}^{(2)}}{1 - q_k}, \tag{3.35}$$

where the weight q_k is given by

$$q_k = \frac{\mathrm{BIAS}(\widehat{\alpha}^{(1)})}{\mathrm{BIAS}(\widehat{\alpha}^{(2)})}. \tag{3.36}$$

For example, consider the following three biased estimators of tail exponent,

$$\widehat{\alpha}^{(1)} = \widehat{\alpha}_H, \quad \widehat{\alpha}^{(2)} = \frac{2}{M_2 \widehat{\alpha}_H}, \quad \widehat{\alpha}^{(3)} = \sqrt{\frac{2}{M_2}},$$

where $\widehat{\alpha}_H$ is the Hill estimator in (3.6) and M_2 is given by

$$M_2 = \frac{1}{k}\sum_{i=1}^{k} \log^2\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right).$$

Under the second-order framework, asymptotic behavior of these three estimators is characterized by

$$\sqrt{k}(\widehat{\alpha}^{(1)} - \alpha) \xrightarrow{d} \mathcal{N}\left(-\frac{\lambda\alpha^2}{1-\rho^*}, \alpha^2\right),$$

$$\sqrt{k}(\widehat{\alpha}^{(2)} - \alpha) \xrightarrow{d} \mathcal{N}\left(-\frac{\lambda\alpha^2}{(1-\rho^*)^2}, 2\alpha^2\right),$$

$$\sqrt{k}(\widehat{\alpha}^{(3)} - \alpha) \xrightarrow{d} \mathcal{N}\left(-\frac{\lambda\alpha^2(2-\rho^*)}{2(1-\rho^*)^2}, \frac{5}{4}\alpha^2\right),$$

where λ appears in (3.69) below (see, for example, Gomes et al. (2000)). Hence, the generalized jackknife estimator based on $(\widehat{\alpha}^{(1)}, \widehat{\alpha}^{(2)})$ is given by

$$\widehat{\alpha}^{(G)}(\widehat{\alpha}^{(1)}, \widehat{\alpha}^{(2)}) = \frac{\widehat{\alpha}^{(1)} - (1-\rho^*)\widehat{\alpha}^{(2)}}{\rho^*} \tag{3.37}$$

and is the same as an asymptotically unbiased estimator in Peng (1998). In our context, $\rho^* = -1/\alpha$ is unknown. Estimating ρ^* by $-1/\widehat{\alpha}_H$ gives the generalized jackknife estimator of Peng,

$$\widehat{\alpha}_P = -(\widehat{\alpha}_H)^2 + \left(\frac{1}{\widehat{\alpha}_H} + 1\right)\frac{2}{M_2}. \tag{3.38}$$

Similarly, the generalized jackknife estimator based on $\widehat{\alpha}^{(2)}$ and $\widehat{\alpha}^{(3)}$ is given by

$$\widehat{\alpha}^{(G)}(\widehat{\alpha}^{(2)}, \widehat{\alpha}^{(3)}) = \frac{(2-\rho^*)\widehat{\alpha}^{(2)} - 2\widehat{\alpha}^{(3)}}{-\rho^*}. \tag{3.39}$$

By replacing ρ^* by $-1/\widehat{\alpha}_H$, this yields a jackknife estimator

$$\widehat{\alpha}_{JK} = \left(2 + \frac{1}{\widehat{\alpha}_H}\right)\frac{2}{M_2} - 2\widehat{\alpha}_H\sqrt{\frac{2}{M_2}}. \tag{3.40}$$

Remark. Gomes et al. (2000) derive generalized jackknife estimators for $\gamma = 1/\alpha$. For example, replacing $\rho^* = -1/\widehat{\alpha}_H$ in Peng's jackknife estimator gives

$$\widehat{\gamma}_P = \frac{1}{\widehat{\alpha}^{(2)}} + \frac{\widehat{\alpha}^{(1)}}{\widehat{\alpha}^{(2)}} - 1.$$

In terms of α, this gives

$$1/\widehat{\gamma}_P = \frac{\widehat{\alpha}^{(2)}}{1 + \widehat{\alpha}^{(1)} - \widehat{\alpha}^{(2)}} = \frac{2}{M_2\widehat{\alpha}_H(1+\widehat{\alpha}_H) - 2},$$

which is different from $\widehat{\alpha}_P$ in (3.38). It can be easily checked, however, that the asymptotic normality results (Section 3.3) for $\widehat{\alpha}_P$ and $1/\widehat{\gamma}_P$ are the same.

3.2.5 Approximate normalized log-spacings

Feuerverger and Hall (1999) proposed parameter estimators based on normalized log-spacings of order statistics and their approximations by a normalized Exponential distribution. Under the assumption (3.9), consider the normalized log-spacings

$$U_i = i(\log X_{(n-i+1)} - \log X_{(n-i)})$$

and set $\delta(x) = -\alpha^{-1}c_1^{-(\alpha^{-1}+1)}c_2 x^{1/\alpha} = Dx^{1/\alpha}$. Then, one expects that

$$U_i \approx \alpha^{-1}(1+\delta(i/n))Z_i \approx \alpha^{-1}\exp\left(\delta(i/n)\right) Z_i, \tag{3.41}$$

where Z_i are independent Exponential random variables with mean 1. This suggests the maximum likelihood estimator of α based on maximizing

$$L(D,\alpha) = \sum_{i=1}^{k} \left\{ \log\alpha - D\left(\frac{i}{n}\right)^{\alpha^{-1}} - \alpha U_i \exp\left(-D\left(\frac{i}{n}\right)^{\alpha^{-1}}\right)\right\}.$$

We denote the corresponding estimator as $\widehat{\alpha}_{FH}$. A related estimator based on regression is the following. Observe from (3.41) that taking logarithm gives

$$\log U_i \approx -\log\alpha + \delta(i/n) + \log Z_i$$

$$=: -\log\alpha - E(\log(Z_i)) + \delta(i/n) + u_i = \theta + \delta(i/n) + u_i,$$

where $\theta = -\log\alpha - E(\log(Z_i))$ and u_i are i.i.d. random variables with mean zero and variance σ_1^2. The mean of $\log Z_i$ is known as Euler's constant ($E(\log(Z_i)) = .5772\ldots$) and $\sigma_1^2 = \text{Var}(\log(Z_i)) = \pi^2/6 = 1.644934$. This leads to the nonlinear regression estimator $\widehat{\alpha}_{FHn}$.

3.2.6 Conditional maximum likelihood estimators

We derive here conditional maximum likelihood estimators supposing that the distribution tail behaves exactly as

$$\overline{F}(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1}, \quad x > u, \tag{3.42}$$

for fixed known threshold u, or equivalently, the corresponding density is given as

$$f(x) = x^{-\alpha-1}(\alpha c_1 + (\alpha+1)c_2 x^{-1}), \quad x > u,$$

where

$$\alpha c_1 + (\alpha+1)c_2 x^{-1} > 0 \quad \text{ for all } x > u. \tag{3.43}$$

Several approaches are possible and are explained next.

First approach:

For given order statistics $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, suppose that k upper-order observations are above the threshold u. The joint density of k upper-order statistics (see, for example, Embrechts et al. (1997), p. 185) is

$$f(X_{(n-k+1)} = x_k, \ldots, X_{(n)} = x_1) = \frac{n!}{(n-k)!} F(x_k)^{n-k} f(x_k) f(x_{k-1}) \ldots f(x_1)$$

$$= \frac{n!}{(n-k)!}(1 - c_1 x_k^{-\alpha} - c_2 x_k^{-\alpha-1})^{n-k} \prod_{i=1}^{k} x_i^{-\alpha-1}(\alpha c_1 + (\alpha+1)c_2 x_i^{-1}). \tag{3.44}$$

The corresponding maximum likelihood estimators $(\widehat{\alpha}_{ML}, \widehat{c}_{1ML}, \widehat{c}_{2ML})$ are obtained by minimizing negative log-likelihood, namely,

$$\underset{\alpha,c_1,c_2}{\operatorname{argmin}}\left(\alpha\sum_{i=1}^{k}\log x_i-\sum_{i=1}^{k}\log(\alpha c_1+(\alpha+1)c_2x_i^{-1})-(n-k)\log(1-c_1x_k^{-\alpha}-c_2x_k^{-\alpha-1})\right),\tag{3.45}$$

subject to (3.43), where we denote $x_i = X_{(n-i+1)}$ for notational simplicity. This is equivalent to finding solutions to

$$(n-k)\frac{(c_1x_k+c_2)\log x_k}{x_k^{\alpha+1}-c_1x_k-c_2}-\sum_{i=1}^{k}\log x_i+\sum_{i=1}^{k}\frac{c_1x_i+c_2}{\alpha c_1x_i+(\alpha+1)c_2}=0,\tag{3.46}$$

$$(n-k)\frac{-x_k}{x_k^{\alpha+1}-c_1x_k-c_2}+\sum_{i=1}^{k}\frac{\alpha x_i}{\alpha c_1x_i+(\alpha+1)c_2}=0,\tag{3.47}$$

$$(n-k)\frac{-1}{x_k^{\alpha+1}-c_1x_k-c_2}+\sum_{i=1}^{k}\frac{\alpha+1}{\alpha c_1x_i+(\alpha+1)c_2}=0,\tag{3.48}$$

subject to the condition (3.43).

Substituting (3.47) and (3.48) into (3.46), we get that

$$\sum_{i=1}^{k}\frac{1}{\alpha+c_2/(c_1x_i+c_2)}=\sum_{i=1}^{k}\log\left(\frac{x_i}{x_k}\right).\tag{3.49}$$

Observe also that adding $c_1\times$(3.47) and $c_2\times$(3.48) gives

$$c_1x_k+c_2=\frac{k}{n}x_k^{\alpha+1}\tag{3.50}$$

and relations (3.47), (3.48), (3.49) and (3.50) give

$$\frac{c_1}{\alpha}\frac{n}{x_k^{\alpha}}+\frac{c_2}{\alpha+1}\frac{n}{x_k^{\alpha+1}}=\sum_{i=1}^{k}\log\left(\frac{x_i}{x_k}\right).\tag{3.51}$$

Solving linear equations (3.50) and (3.51) gives a closed-form expression for c_1 and c_2 in terms of α,

$$c_1=\frac{k}{n}x_k^{\alpha}\alpha\left(\frac{\alpha+1}{\widetilde{\alpha}_H}-1\right),\quad c_2=\frac{k}{n}x_k^{\alpha+1}\left(1-\alpha\left(\frac{\alpha+1}{\widetilde{\alpha}_H}-1\right)\right),\tag{3.52}$$

where

$$\widetilde{\alpha}_H^{-1} = \frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{x_i}{x_k}\right) = \frac{k-1}{k}\widehat{\alpha}_H^{-1}$$

is a version of the original Hill estimator. Finally, $\widehat{\alpha}_{ML}$ is the solution of nonlinear equation,

$$\sum_{i=1}^{k}\left(\frac{1}{\alpha + w_i(\alpha)} - \frac{1}{\widetilde{\alpha}_H}\right) = 0, \tag{3.53}$$

where weights are given by

$$w_i(\alpha) = \frac{x_k(\alpha+1)(\widetilde{\alpha}_H - \alpha)}{\alpha(\alpha+1-\widetilde{\alpha}_H)x_i + x_k(\alpha+1)(\widetilde{\alpha}_H - \alpha)}.$$

Remark. Observe from (3.49) and (3.50) that if $c_2 = 0$, then $\widehat{\alpha}_{ML}$ leads to conditional maximum likelihood estimator of Hill (1975) for the first order asymptotics,

$$\widehat{\alpha}_{ML} = \widetilde{\alpha}_H, \quad \widehat{c_1}_{ML} = \frac{k}{n}x_k^{\widetilde{\alpha}_H}.$$

In the case of RDEs, for example, one expects that $c_1 > 0$ and $c_2 < 0$. This leads to further restrictions on the solutions of (3.53). Note that the inequality (3.43) and (3.52) imply that

$$1 < \alpha((\alpha+1)/\widetilde{\alpha}_H - 1) < \alpha + 1. \tag{3.54}$$

This is equivalent to

$$\widetilde{\alpha}_H < \alpha < \widetilde{\alpha}_H - \frac{1}{2} + \sqrt{\widetilde{\alpha}_H^2 + \frac{1}{4}}. \tag{3.55}$$

From computational perspective, by using relation (3.50), maximum likelihood estimator $\widehat{\alpha}_{ML}$ is obtained by minimizing the negative log-likelihood

$$\underset{\alpha}{\operatorname{argmin}}\left(\alpha\sum_{i=1}^{k}\log x_i - \sum_{i=1}^{k}\log(\alpha c_1 + (\alpha+1)c_2 x_i^{-1})\right),$$

where c_1 and c_2 are functions of α as in (3.52), subject to restriction (3.55).

Second approach:

Alternatively, note that for given order statistics $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, the joint distribution of $(X_{(n-k+1)}, \ldots, X_{(n)})$ given $X_{(n-k)} = u$ is the same as the joint distribution of order statistics $Y_{(1)}, \ldots, Y_{(k)}$ of i.i.d. random variables from a distribution

$$F_u(y) = P(X \leq y | X > u) = \frac{F(y) - F(u)}{1 - F(u)}, \quad y \geq u \tag{3.56}$$

(see, for example, Lemma 3.4.1 of de Haan and Ferreira (2006)). Under (3.42) conditional density function becomes

$$f_u(y) = y^{-\alpha-2} u^{\alpha} (\alpha c y + (\alpha + 1)(1 - c)u), \tag{3.57}$$

where

$$c = \frac{c_1}{c_1 + c_2 u^{-1}} \tag{3.58}$$

and

$$\alpha c y + (\alpha + 1)(1 - c)u > 0, \quad y \geq u. \tag{3.59}$$

Therefore, maximum likelihood estimators $(\widehat{\alpha}_{ML2}, \widehat{c}_{ML2})$ are given by minimizing negative log-likelihood, namely,

$$l(\alpha, c) = -\log k! + \sum_{i=1}^{k} ((\alpha + 2)\log y_i - \alpha \log u - \log(\alpha c y_i + (\alpha + 1)(1 - c)u), \tag{3.60}$$

subject to (3.59), where we denote $y_i = Y_{(i)} = X_{(n-k+i)}$ for notational simplicity. By solving $\partial l / \partial \alpha = 0$ and $\partial l / \partial c = 0$, the maximum likelihood estimators are given as the solution of

$$\sum_{i=1}^{k} \frac{1}{\alpha + (1 - c)u/(y_i c + (1 - c)u)} = \sum_{i=1}^{k} \log\left(\frac{y_i}{u}\right), \tag{3.61}$$

$$c = \frac{\alpha(\alpha + 1 - \widehat{\alpha}_H)}{\widehat{\alpha}_H}. \tag{3.62}$$

Remark. The only difference between the two maximum likelihood estimators for α is what version of Hill estimator is used. First approach uses a version of original Hill estimator $\widetilde{\alpha}_H$, while the second approach uses unbiased Hill estimator $\widehat{\alpha}_H$. Note also that for the constant estimators, the first approach gives explicit estimators for two constant parameters c_1 and c_2, whereas the second approach is parameterized only by one constant c. However, replacing $\widetilde{\alpha}_H$ by $\widehat{\alpha}_H$ in $\widehat{c_1}_{ML}$ and $\widehat{c_2}_{ML}$ and plugging them into (3.58) with $u = X_{(n-k)}$ gives $\widehat{c}_{ML2}$.

Remark. Suppose that $c_1 > 0$ and $c_2 < 0$. Then, (3.58) and (3.59) with $y = u$ imply that

$$1 < c < \alpha + 1,$$

which leads further to

$$\widehat{\alpha}_H < \alpha < \widehat{\alpha}_H - \frac{1}{2} + \sqrt{\widehat{\alpha}_H^2 + \frac{1}{4}}. \tag{3.63}$$

This is exactly of the form (3.54) replacing $\widetilde{\alpha}_H$ by $\widehat{\alpha}_H$. In practice, maximum likelihood estimator is obtained by minimizing

$$\underset{\alpha}{\operatorname{argmin}} \left(\alpha \sum_{i=1}^{k} \log(y_i/u) - \sum_{i=1}^{k} \log(\alpha c + (\alpha+1)(1-c)(y_i/u)^{-1}) \right),$$

subject to (3.63).

3.2.7 Discussions on estimation for known ρ^*

We discuss here briefly estimation methods in the framework (3.11) where ρ^* is known. Without loss of generality, we may suppose that $\rho^* = -1$. Maximum likelihood estimators of Sections 3.2.5 and 3.2.6, and the generalized jackknife estimators (for example, Peng's estimator in (3.37)) can be adapted analogously to this case. The least squares regression estimators can be defined as outlined next.

As in (3.24), the inverse function of $F(x)$ can be written as

$$F^{\leftarrow}\left(1 - \frac{1}{y}\right) = (c_1 y)^{1/\alpha} \left(1 + \frac{c_2}{\alpha c_1}(c_1 y)^{-1}(1 + o(1))\right), \quad y \to \infty. \tag{3.64}$$

Then, the corresponding QQ-estimator can be based on

$$\log X_{(n-i+1)} \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log\left(\frac{i}{n}\right) + \frac{c_2}{\alpha c_1^2}\left(\frac{i}{n}\right). \tag{3.65}$$

To see how rank-based, least squares estimator can be defined, note that

$$\overline{F}(x) = c_1 x^{-\alpha}\left(1 + \frac{c_2}{c_1}x^{-\alpha} + o(x^{-\alpha})\right) = c_1 x^{-\alpha}\left(1 + \frac{c_2}{c_1^2}(c_1 x^{-\alpha}) + o(x^{-\alpha})\right)$$

$$\approx c_1 x^{-\alpha}\left(1 + \frac{c_2}{c_1^2}\overline{F}(x)\right). \tag{3.66}$$

Taking the logarithm in (3.66) yields

$$\log\left(\frac{i}{n}\right) \approx \log c_1 - \alpha \log X_{(n-i+1)} + \frac{c_2}{c_1^2}\left(\frac{i}{n}\right). \tag{3.67}$$

3.3 Theoretical properties of estimators

We examine here basic theoretical properties of proposed estimators. More specifically, we focus on estimators $\widehat{\alpha}_{RK2}$, $\widehat{\alpha}_{QQ2}$, $\widehat{\alpha}_{JK}$, $\widehat{\alpha}_{P}$ (Section 3.3.1) and $\widehat{\alpha}_{ML2}$, $\widehat{\alpha}_{FH}$ (Section 3.3.2) of tail exponent α, and their asymptotic normality. The proofs are similar to what can be found in the literature, and are only outlined. The other estimators are not considered for shortness sake and for being close relatives of the considered estimators.

The asymptotic normality results will be established under the second-order condition (3.7) with parameter $\rho = -1$, which is more convenient to write here as

$$\lim_{x\to\infty} \frac{\frac{\overline{F}(xa)}{\overline{F}(x)} - a^{-\alpha}}{G(x)} = a^{-\alpha}\frac{a^{-1}-1}{-1/\alpha^2}, \tag{3.68}$$

where $\alpha > 0$ and G is (ultimately) a positive or negative function with $\lim_{x\to\infty} G(x) = 0$. As common for similar results in related literature, we shall consider $k = k(n) \to \infty$, $k/n \to 0$ as $n \to \infty$, such that

$$\lim_{n\to\infty} \sqrt{k}G\left(\frac{n}{k}\right) = \lambda < \infty. \tag{3.69}$$

3.3.1 Least squares based and generalized jackknife estimators

We state here the asymptotic normality result of the rank-based, QQ and generalized jackknife estimators.

Theorem 3.3.1. *Let $\widehat{\alpha}$ denote one of the estimators $\widehat{\alpha}_{RK2}$, $\widehat{\alpha}_{QQ2}$, $\widehat{\alpha}_{JK}$ and $\widehat{\alpha}_{P}$. Under the* *assumptions (3.68)-(3.69) above, we have*

$$\sqrt{k}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{3.70}$$

where the respective asymptotic variances σ^2 *are given by*

$$\sigma^2_{RK2} = \sigma^2_{QQ2} = 2\alpha^2(\alpha + 1)^2, \tag{3.71}$$

$$\sigma^2_{JK} = \sigma^2_{P} = \alpha^2((\alpha + 1)^2 + \alpha^2). \tag{3.72}$$

Remark. All proposed estimators are aimed to reduce bias in the second order framework. Therefore, it should not be surprising that the mean is zero in the limiting normal distribution (3.70).

Remark. Note that the asymptotic normality result (3.70) implies consistency. However, consistency can also be established under milder assumptions of first order asymptotics. The idea is to use Potter's bound to have lower and upper bounds for each individual term in considered estimators, and apply Central Limit Theorem (see, for example, Lemma 3.2.3 in de Haan and Ferreira (2006), p. 71) for Renyi's representation of order statistics.

The proof of Theorem 3.3.1 is standard and will be mostly outlined. We consider only the estimator $\widehat{\alpha}_{RK2}$ (others can be dealt with in analogous way). The main idea is to establish joint normality of each individual term entering into estimator $\widehat{\alpha}_{RK2}$ and to apply the delta method. Let $\gamma = 1/\alpha$ to simplify the notation and further denote $Z_{(n-i+1)} = X_{(n-i+1)}/X_{(n-k)}$,

$$a = \frac{1}{k}\sum_{i=1}^{k} Z^{-2}_{(n-i+1)}, \quad b = \frac{1}{k}\sum_{i=1}^{k} Z^{-1}_{(n-i+1)}, \quad c = \frac{1}{k}\sum_{i=1}^{k}(\log(i/k) - \overline{\log(i/k)}) \log Z_{(n-i+1)},$$

$$d = \frac{1}{k}\sum_{i=1}^{k} Z^{-1}_{(n-i+1)} \log Z_{(n-i+1)}, \quad e = \frac{1}{k}\sum_{i=1}^{k} \log Z_{(n-i+1)},$$

$$f = \frac{1}{k}\sum_{i=1}^{k}(\log(i/k) - \overline{\log(i/k)}) Z^{-1}_{(n-i+1)}, \quad g = \frac{1}{k}\sum_{i=1}^{k}(\log Z_{(n-i+1)})^2,$$

where $\overline{\log(i/k)} = \sum_{i=1}^{k} \log\left(\frac{i}{k}\right)/k$. Then, the rank-based estimator can be rewritten as

$$\widehat{\alpha}_{RK2} = \frac{(d-be)f-(a-b^2)c}{(g-e^2)(a-b^2)-(d-be)^2} =: H(\boldsymbol{h}).$$

For asymptotic normality of $\boldsymbol{h} = (a, b, \ldots, g)$, observe first that under the second order condition (3.68), we can represent log-spacings of order statistics by tail empirical quantile process (see, for example, Theorem 2.4.8 and top of p. 76 of de Haan and Ferreira (2006)). For each $\epsilon > 0$,

$$\log Z_{(n-[ks]+1)} = -\gamma \log s + \frac{\gamma}{\sqrt{k}} s^{-1} B_n^0(s) + G_0\left(\frac{n}{k}\right)\left(\frac{s^\gamma - 1}{-\gamma} + o_P(1)s^{-1/2-\epsilon}\right), \tag{3.73}$$

where $G_0(s) \sim G(s)$ and $B_n^0(s) = B_n(s) - sB_n(1)$ is a Brownian bridge and $o_P(1)$ term tends to zero in probability uniformly for $0 < s \leq 1$. By using (3.73), it follows that

$$a = \frac{1}{2\gamma+1} - \frac{2\gamma}{\sqrt{k}} \int_0^1 s^{2\gamma-1} B_n^0(s) ds - 2G_0\left(\frac{n}{k}\right) \int_0^1 s^{2\gamma}\left(\frac{s^\gamma-1}{-\gamma} + o_P(1)s^{-1/2-\epsilon}\right) ds$$

and hence, as $k \to \infty$, using (3.69),

$$\sqrt{k}\left(a - \frac{1}{2\gamma+1}\right) = -2\gamma \int_0^1 s^{2\gamma-1} B_n^0(s) ds - 2\lambda \int_0^1 s^{2\gamma}\left(\frac{s^\gamma-1}{-\gamma}\right) ds + o_P(1).$$

Similar expansions for the rest of the terms are

$$\sqrt{k}\left(b - \frac{1}{\gamma+1}\right) = -\gamma \int_0^1 s^{\gamma-1} B_n^0(s) ds - \lambda \int_0^1 s^{\gamma}\left(\frac{s^\gamma-1}{-\gamma}\right) ds + o_P(1),$$

$$\sqrt{k}\left(c - (-\gamma)\right) = \gamma \int_0^1 s^{-1}(1+\log s) B_n^0(s) ds + \lambda \int_0^1 (1+\log s)\left(\frac{s^\gamma-1}{-\gamma}\right) ds + o_P(1),$$

$$\sqrt{k}\left(d - \frac{\gamma}{(\gamma+1)^2}\right) = \gamma \int_0^1 s^{\gamma-1}(1+\gamma\log s) B_n^0(s) ds + \lambda \int_0^1 s^{\gamma}(1+\gamma\log s)\left(\frac{s^\gamma-1}{-\gamma}\right) ds + o_P(1),$$

$$\sqrt{k}\left(e - \gamma\right) = \gamma \int_0^1 s^{-1} B_n^0(s) ds + \lambda \int_0^1 \left(\frac{s^\gamma-1}{-\gamma}\right) ds + o_P(1),$$

$$\sqrt{k}\left(f - 2\gamma^2\right) = -2\gamma^2 \int_0^1 s^{-1}\log s \; B_n^0(s) ds - 2\lambda\gamma \int_0^1 \log s\left(\frac{s^\gamma-1}{-\gamma}\right) ds + o_P(1),$$

$$\sqrt{k}(g - 2\gamma^2) = -2\gamma^2 \int_0^1 s^{-1} \log s B_n^0(s) ds - 2\gamma\lambda \int_0^1 \log s \left(\frac{s^\gamma - 1}{-\gamma} \right) ds + o_P(1),$$

by using the approximation $e^{-x} \approx 1 - x$ as $x \to 0$ and

$$\left(B_n^0(s)\right)^2 / \sqrt{k} \to 0, \quad B_n^0(s) G_0(n/k) \to 0, \quad \sqrt{k}(G_0(n/k))^2 \to 0,$$

as $k \to \infty$, $k/n \to 0$.

Since B_n^0 is a Gaussian process, the vector $(a, b, \ldots, g)$ is asymptotically multivariate normal, and $EB_n^0(s) = 0$ implies that the asymptotic mean comes from the integration related to λ. For example, for c, the change of variables $y = -\log s$ gives

$$\lambda \int_0^1 (1 + \log s) \left(\frac{s^\gamma - 1}{-\gamma} \right) ds = \frac{\lambda}{\gamma} \int_0^\infty (1 - y)(1 - e^{-\gamma y}) e^{-y} dy = \frac{-\lambda}{(\gamma + 1)^2}.$$

Note also that the limiting covariance function only involves the Brownian bridge. For example, the limiting covariance of a and b becomes

$$E \left(-2\gamma \int_0^1 s^{2\gamma-1} B_n^0(s) ds \right) \left(-\gamma \int_0^1 u^{\gamma-1} B_n^0(u) du \right)$$

$$= 2\gamma^2 \int_0^1 \int_0^1 s^{2\gamma-1} u^{\gamma-1} (s \wedge u - su) ds du = \frac{2\gamma^2}{(\gamma + 1)(2\gamma + 1)(3\gamma + 1)}.$$

After tedious calculations not reported here, we obtain that

$$\sqrt{k} \left(\begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{pmatrix} - \begin{pmatrix} \frac{1}{2\gamma+1} \\ \frac{1}{\gamma+1} \\ -\gamma \\ \frac{\gamma}{(\gamma+1)^2} \\ \gamma \\ \frac{\gamma}{(\gamma+1)^2} \\ 2\gamma^2 \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.74}$$

where

$$\boldsymbol{\mu}' = \left(\frac{-2\lambda}{(3\gamma + 1)(2\gamma + 1)}, \frac{-\lambda}{(2\gamma + 1)(\gamma + 1)}, \frac{-\lambda}{(\gamma + 1)^2}, \frac{-\lambda(\gamma^2 - \gamma - 1)}{(\gamma + 1)^2 (2\gamma + 1)^2}, \right.$$

$$\frac{\lambda}{\gamma+1}, \frac{\lambda(1-2\gamma^2)}{(\gamma+1)^2(2\gamma+1)^2}, \frac{\lambda 2\gamma(\gamma+2)}{(\gamma+1)^2}\right)$$

and the covariance matrix $\mathbf{\Sigma}$ is given by

$$\begin{pmatrix}
\frac{4\gamma^2}{(2\gamma+1)^2(4\gamma+1)} & \frac{2\gamma^2}{(\gamma+1)(2\gamma+1)(3\gamma+1)} & \frac{2\gamma^2}{(2\gamma+1)^3} & \frac{2\left(\gamma^4-2\gamma^3-\gamma^2\right)}{(\gamma+1)^2(2\gamma+1)(3\gamma+1)^2} & -\frac{2\gamma^2}{(2\gamma+1)^2} & \frac{2\left(3\gamma^4-\gamma^2\right)}{(\gamma+1)^2(2\gamma+1)(3\gamma+1)^2} & -\frac{8\gamma^3(\gamma+1)}{(2\gamma+1)^3} \\
\frac{2\gamma^2}{(\gamma+1)(2\gamma+1)(3\gamma+1)} & \frac{\gamma^2}{(\gamma+1)^2(2\gamma+1)} & \frac{\gamma^2}{(\gamma+1)^3} & \frac{\gamma^4-\gamma^3-\gamma^2}{(\gamma+1)^3(2\gamma+1)^2} & -\frac{\gamma^2}{(\gamma+1)^2} & \frac{2\gamma^4-\gamma^2}{(\gamma+1)^3(2\gamma+1)^2} & -\frac{2\gamma^3(\gamma+2)}{(\gamma+1)^3} \\
\frac{2\gamma^2}{(2\gamma+1)^3} & \frac{\gamma^2}{(\gamma+1)^3} & 2\gamma^2 & \frac{2\gamma^3-\gamma^2}{(\gamma+1)^4} & -\gamma^2 & \frac{\gamma^3-2\gamma^2}{(\gamma+1)^4} & -6\gamma^3 \\
\frac{2\left(\gamma^4-2\gamma^3-\gamma^2\right)}{(\gamma+1)^2(2\gamma+1)(3\gamma+1)^2} & \frac{\gamma^4-\gamma^3-\gamma^2}{(\gamma+1)^3(2\gamma+1)^2} & \frac{2\gamma^3-\gamma^2}{(\gamma+1)^4} & \frac{\gamma^2\left(2\gamma^4+2\gamma+1\right)}{(\gamma+1)^4(2\gamma+1)^3} & \frac{\gamma^2-\gamma^3}{(\gamma+1)^3} & \frac{2\gamma^6-3\gamma^5-5\gamma^4+\gamma^2}{(\gamma+1)^4(2\gamma+1)^3} & -\frac{2\left(\gamma^5+2\gamma^4-2\gamma^3\right)}{(\gamma+1)^4} \\
-\frac{2\gamma^2}{(2\gamma+1)^2} & -\frac{\gamma^2}{(\gamma+1)^2} & -\gamma^2 & \frac{\gamma^2-\gamma^3}{(\gamma+1)^3} & \gamma^2 & \frac{\gamma^2-\gamma^3}{(\gamma+1)^3} & 4\gamma^3 \\
\frac{2\left(3\gamma^4-\gamma^2\right)}{(\gamma+1)^2(2\gamma+1)(3\gamma+1)^2} & \frac{2\gamma^4-\gamma^2}{(\gamma+1)^3(2\gamma+1)^2} & \frac{\gamma^3-2\gamma^2}{(\gamma+1)^4} & \frac{2\gamma^6-3\gamma^5-5\gamma^4+\gamma^2}{(\gamma+1)^4(2\gamma+1)^3} & \frac{\gamma^2-\gamma^3}{(\gamma+1)^3} & \frac{\gamma^2\left(4\gamma^4-3\gamma^2+2\gamma+2\right)}{(\gamma+1)^4(2\gamma+1)^3} & -\frac{2\left(\gamma^5+\gamma^4-3\gamma^3\right)}{(\gamma+1)^4} \\
-\frac{8\gamma^3(\gamma+1)}{(2\gamma+1)^3} & -\frac{2\gamma^3(\gamma+2)}{(\gamma+1)^3} & -6\gamma^3 & -\frac{2\left(\gamma^5+2\gamma^4-2\gamma^3\right)}{(\gamma+1)^4} & 4\gamma^3 & -\frac{2\left(\gamma^5+\gamma^4-3\gamma^3\right)}{(\gamma+1)^4} & 20\gamma^4
\end{pmatrix}.$$

Finally, applying the delta method to the function $H(\boldsymbol{h})$ at the point value

$$\boldsymbol{h}_0 = \left(\frac{1}{2\gamma+1}, \frac{1}{\gamma+1}, -\gamma, \frac{\gamma}{(\gamma+1)^2}, \gamma, \frac{\gamma}{(\gamma+1)^2}, 2\gamma^2\right)$$

gives

$$\sqrt{k}\,(H(\boldsymbol{h}) - H(\boldsymbol{h}_0)) \xrightarrow{d} \mathcal{N}\left(\frac{\partial H}{\partial \boldsymbol{h}}(\boldsymbol{h}_0)\boldsymbol{\mu}, \frac{\partial H}{\partial \boldsymbol{h}}(\boldsymbol{h}_0)\mathbf{\Sigma}\frac{\partial H}{\partial \boldsymbol{h}}(\boldsymbol{h}_0)'\right).$$

Algebraic calculations give

$$\frac{\partial H}{\partial \boldsymbol{h}}(\boldsymbol{h}_0) = \left(0, \frac{(\gamma+1)^2(2\gamma+1)}{\gamma^4}, \frac{-(\gamma+1)^2}{\gamma^4}, \frac{-(\gamma+1)^2(2\gamma+1)}{\gamma^5}, \frac{(\gamma+1)\left(2\gamma^2+4\gamma+1\right)}{\gamma^5},\right.$$

$$\left.\frac{-(\gamma+1)^2(2\gamma+1)}{\gamma^4}, \frac{-(\gamma+1)^2}{\gamma^5}\right).$$

Hence, we have

$$\sqrt{k}\,(\widehat{\alpha}_{RK2} - \alpha) \xrightarrow{d} \mathcal{N}(0, 2\alpha^2(\alpha+1)^2).$$

3.3.2 Maximum likelihood estimators

We state and outline the proof for the asymptotic normality of maximum likelihood estimators $\widehat{\alpha}_{ML2}$ and $\widehat{\alpha}_{FH}$.

Theorem 3.3.2. *Under the assumption (3.42), we have*

$$\sqrt{k}(\widehat{\alpha}_{ML2} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{ML2}(u)), \tag{3.75}$$

where $\sigma^2_{ML2}(u)$ *is given in (3.81) below, and satisfies*

$$\lim_{u\to\infty} \sigma^2_{ML2}(u) = \alpha^2(\alpha+1)^2 =: \sigma^2_{ML2}. \tag{3.76}$$

Under the assumptions (3.68)-(3.69) above, we have

$$\sqrt{k}(\widehat{\alpha}_{FH} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{FH}), \tag{3.77}$$

where

$$\sigma^2_{FH} = \alpha^2(\alpha+1)^2.$$

The proof for $\widehat{\alpha}_{ML2}$ is based on the standard asymptotic normality result for maximum likelihood estimator. From the likelihood function l in (3.60), observe that

$$\frac{\partial^2 l}{\partial \alpha^2} = -\sum_{i=1}^{k} \frac{(cy_i + (1-c)u)^2}{(\alpha c y_i + (\alpha+1)(1-c)u)^2}, \tag{3.78}$$

$$\frac{\partial^2 l}{\partial \alpha \partial c} = \sum_{i=1}^{k} \frac{y_i u}{(\alpha c y_i + (\alpha+1)(1-c)u)^2}, \tag{3.79}$$

$$\frac{\partial^2 l}{\partial c^2} = -\sum_{i=1}^{k} \frac{(\alpha y_i - (\alpha+1)u)^2}{(\alpha c y_i + (\alpha+1)(1-c)u)^2}. \tag{3.80}$$

Tedious calculations for information matrix lead to (3.75) with

$$\sigma^2_{ML2}(u) = \frac{\alpha^2 L - 2\alpha(\alpha+1)M + (\alpha+1)^2 N}{(c^2 L + 2c(1-c)M + (1-c)^2 N)(\alpha^2 L - 2\alpha(\alpha+1)M + (\alpha+1)^2 N) - M^2}, \tag{3.81}$$

where

$$\begin{aligned} L &= {}_2F_1(1, \alpha, 1+\alpha, (1+\alpha)(c-1)/\alpha c)/(\alpha^2 c), \\ M &= {}_2F_1(1, 1+\alpha, 2+\alpha, (1+\alpha)(c-1)/\alpha c)/(\alpha(\alpha+1)c), \end{aligned}$$

$$N = {}_2F_1(1, 2+\alpha, 3+\alpha, (1+\alpha)(c-1)/\alpha c)/(\alpha(\alpha+2)c)$$

and ${}_2F_1(a,b,c,z) = \Gamma(c)(\Gamma(b)\Gamma(c-b))^{-1}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}dt$ denotes a hypergeometric function. (Note that the dependence on u in (3.81) is through c.) Finally, as $u \to \infty$, we have $c \to 1$, and all hypergeometric functions converging to 1. This yields (3.76).

Theoretical properties of $\widehat{\alpha}_{FH}$ with known second order tail parameter ρ^* are studied in Theorem 2.1 of Gomes and Martins (2002). Here, we briefly argue that the same asymptotic result holds for known $\rho = -1$ or unknown $\rho^* = -1/\alpha$. Denote $\gamma = 1/\alpha$ for notational simplicity, and suppose that the upper k-order normalized log-spacings exactly follow

$$U_i \stackrel{d}{=} \gamma \exp\left(D(i/n)^\gamma\right) Z_i,$$

as in relation (3.41). Then, log-likelihood becomes

$$l(\gamma, D) = \sum_{i=1}^{k}\left\{-\log\gamma - D\left(\frac{i}{n}\right)^\gamma - \frac{U_i}{\gamma}\exp\left(-D\left(\frac{i}{n}\right)^\gamma\right)\right\}.$$

From the asymptotic normality of maximum likelihood estimator,

$$\sqrt{k}\left(\frac{\widehat{\gamma}_{ML} - \gamma}{\widehat{\sigma}_{FH}(\gamma)}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\widehat{\sigma}^2_{FH}(\gamma) = \frac{\frac{1}{k}\operatorname{Var}(\frac{\partial l}{\partial D})}{\frac{1}{k}\operatorname{Var}(\frac{\partial l}{\partial \gamma})\frac{1}{k}\operatorname{Var}(\frac{\partial l}{\partial D}) - \left(\frac{1}{k}E\left(-\frac{\partial^2 l}{\partial\gamma\partial D}\right)\right)^2},$$

$$\frac{1}{k}\operatorname{Var}\left(\frac{\partial l}{\partial D}\right) = \left(\frac{k}{n}\right)^{2\gamma}\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{2\gamma},$$

$$\frac{1}{k}\operatorname{Var}\left(\frac{\partial l}{\partial \gamma}\right) = \frac{1}{k}\sum_{i=1}^{k}\gamma^2\left(\frac{1}{\gamma^2} + \frac{D}{\gamma}\left(\frac{i}{n}\right)^\gamma\log\left(\frac{i}{n}\right)\right)^2,$$

$$E\left(-\frac{\partial^2 l}{\partial\gamma\partial D}\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{n}\right)^\gamma\left(\frac{1}{\gamma} + D\left(\frac{i}{n}\right)\log\left(\frac{i}{n}\right)\right).$$

Using $k = k(n) \to \infty$, $k/n \to 0$ as $n \to \infty$, we have $\sum_{i=1}^{k}\left(\frac{i}{n}\right)^\gamma\log\left(\frac{i}{n}\right)/k \to 0$. Hence, by

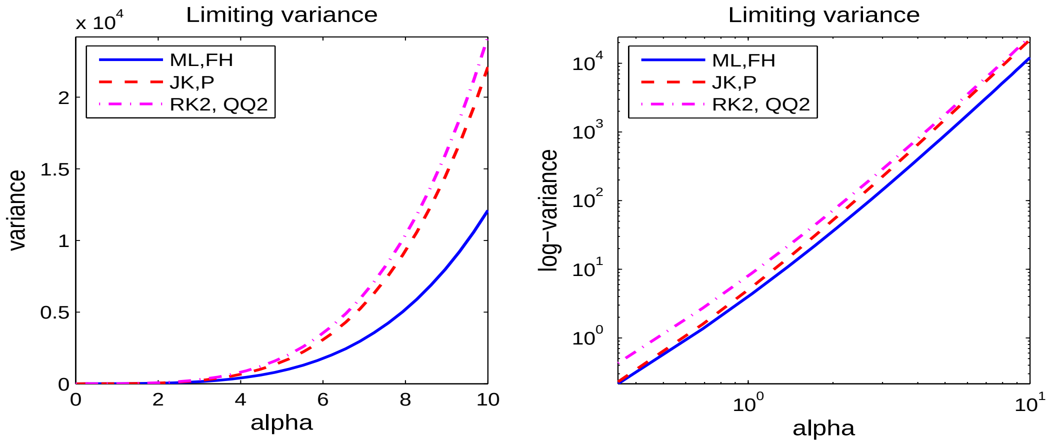

Figure 3.1: Comparison of theoretical variances.

factoring out $(k/n)^{2\gamma}$,

$$\widehat{\sigma}^2_{FH}(\gamma) \to \frac{\int_0^1 x^{2\gamma} dx}{\gamma^{-2}\int_0^1 x^{2\gamma}dx - \left(\int_0^1 x^{\gamma}\gamma^{-1}dx\right)^2} = (\gamma+1)^2,$$

which gives (3.77).

Remark. Figure 3.1 compares theoretical variances of tail exponent estimators. The following ordering takes place:

$$\widehat{\sigma}^2_{ML2} = \widehat{\sigma}^2_{FH} < \widehat{\sigma}^2_{JK} = \widehat{\sigma}^2_{P} < \widehat{\sigma}^2_{RK2} = \widehat{\sigma}^2_{QQ2}. \tag{3.82}$$

Remark. Just comparing variances is not the right way to measure performance of estimators. Bias can arise by introducing third order framework, and it can seriously affect asymptotic mean squared error. For example, under the simpler second order framework (3.68)-(3.69), the Hill estimator behaves as

$$\sqrt{k}(\widehat{\alpha}_H - \alpha) \xrightarrow{d} \mathcal{N}\left(-\frac{\lambda\alpha^3}{\alpha+1}, \alpha^2\right),$$

while the QQ-estimator behaves as

$$\sqrt{k}(\widehat{\alpha}_{QQ} - \alpha) \xrightarrow{d} \mathcal{N}\left(-\frac{\lambda\alpha^4}{(\alpha+1)^2}, 2\alpha^2\right).$$

That is the QQ-estimator has twice larger variance than the Hill estimator though it has smaller bias. The respective asymptotic mean squared errors (AMSE) become

$$AMSE(\widehat{\alpha}_H) = \frac{1}{k}\frac{\lambda^2\alpha^6}{(\alpha+1)^2} + \frac{\alpha^2}{k}, \quad AMSE(\widehat{\alpha}_{QQ}) = \frac{1}{k}\frac{\lambda^2\alpha^8}{(\alpha+1)^4} + \frac{2\alpha^2}{k},$$

and simple algebra gives that if

$$\lambda > \sqrt{\frac{(\alpha+1)^4}{\alpha^4(2\alpha+1)}},$$

then AMSE of the QQ-estimator is smaller than that of the Hill estimator. Similar observation is also made in Feuerverger and Hall (1999) by comparing FHn and FH estimators. FHn estimator produces 64% greater variance than $\widehat{\alpha}_{FH}$, but numerical studies indicate that $\widehat{\alpha}_{FHn}$ is sometimes less biased. Compared to $\widehat{\alpha}_{FHn}$, for example, which is based on nonlinear regression, our RK2 or QQ2 estimator has 22% greater variance. However, they are based on linear regression that is easy to implement and are free of initial estimators to solve nonlinear regression.

3.4 Simulation study

In this section, we present a simulation study examining performance of proposed estimators on several models, and discuss other issues. We consider three models for distribution of X: beta prime distributions (Section 3.4.1), ARCH(1) or autoregressive conditionally heteroscedastic models of order 1 (Section 3.4.2) and multiplicative cascades (Section 3.4.3). These are known examples of random difference equations and have also been considered in Chapter 2. Beta prime distributions have a closed form with the tail satisfying (3.9). The same behavior (3.9) is expected for ARCH(1) and multiplicative cascades models by the results of Chapter 2.

We shall examine seven estimators $\widehat{\alpha}_{RK2}$, $\widehat{\alpha}_{QQ2}$, $\widehat{\alpha}_{GLS}$, $\widehat{\alpha}_P$, $\widehat{\alpha}_{JK}$, $\widehat{\alpha}_{FH}$, $\widehat{\alpha}_{ML}$ of heavy tail exponent α. All simulations are based on 1,000 realizations. In some realizations, when applying FH method, estimates were very unstable (see below for further discussion on FH method). Therefore, we choose to present here robust measures of performance. We focus on median of tail exponent estimators and median absolute error given by

$$\text{MAE}(\widehat{\alpha}(k)) = \text{ median of } |\widehat{\alpha}(i,k) - \alpha|,$$

where k denotes the number of upper order statistics and i represents realization. Another

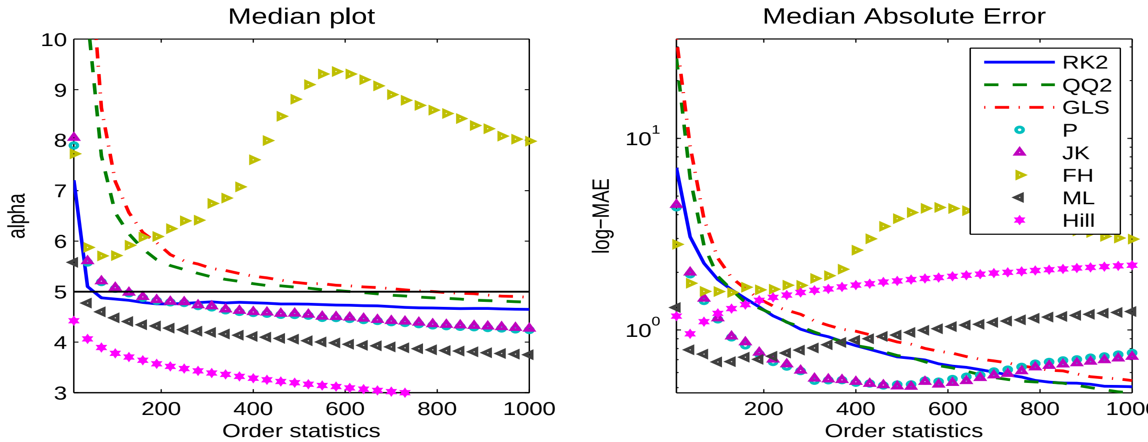

Figure 3.2: $\beta(9,5)$ with sample size $N = 5,000$.

obvious measure of interest is median absolute deviation given by the median of $|\widehat{\alpha}(i,k) - \text{median}(\widehat{\alpha}(i,k))|$. We do not report it for shortness sake though we find that its ordering among estimators corresponds to that from theoretical analysis (3.82) (with the exception of FH which is unstable).

3.4.1 Results for beta prime distribution

Beta prime distribution $\beta(a,b)$, also known as beta distribution of the second kind, is given by its density

$$\frac{1}{B(a,b)} x^{a-1}(1+x)^{-a-b} 1_{\{x>0\}}, \tag{3.83}$$

where $a > 0$, $b > 0$ are two parameters. Observe that the tail of beta prime distribution behaves as

$$\begin{aligned} \overline{F}(x) = P(X > x) &= \frac{1}{B(a,b)} \int_x^\infty u^{a-1}(1+u)^{-a-b} du \\ &= \frac{x^{-b}}{B(a,b)b} - \frac{(a+b)x^{-b-1}}{B(a,b)(b+1)} + o(x^{-b-1}), \end{aligned} \tag{3.84}$$

as $x \to \infty$, and hence satisfies the relation (3.9) with

$$\alpha = b, \quad c_1 = \frac{1}{B(a,b)b}, \quad c_2 = -\frac{a+b}{B(a,b)(b+1)}. \tag{3.85}$$

Figure 3.2 presents simulation results for $\beta(9,5)$ (hence $\alpha = 5$) distribution when the sample size is $N = 5,000$. Median plot suggests that least squares-based ($\widehat{\alpha}_{RK2}$, $\widehat{\alpha}_{QQ2}$, $\widehat{\alpha}_{GLS}$) and

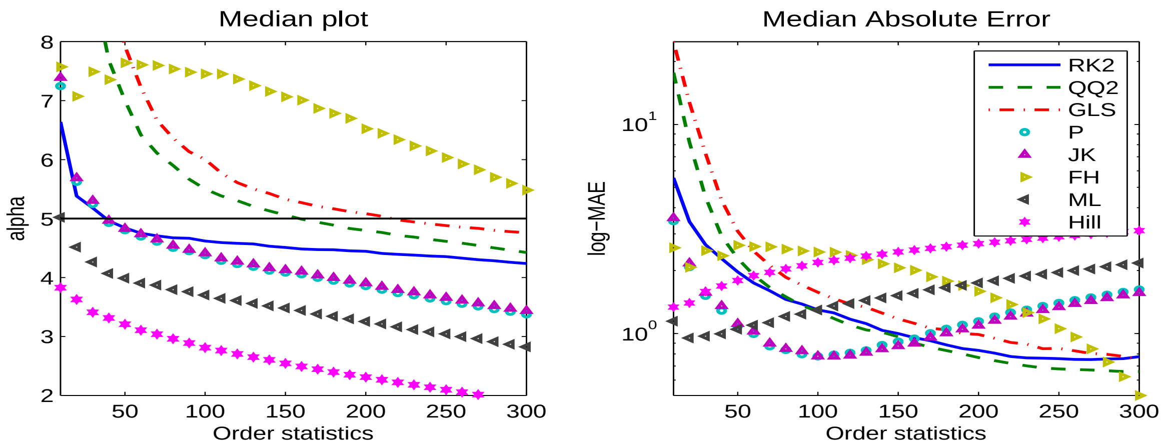

Figure 3.3: $\beta(9,5)$ with sample size $N = 500$.

generalized jackknife ($\widehat{\alpha}_P$, $\widehat{\alpha}_{JK}$) estimators perform well. In particular, $\widehat{\alpha}_P$, $\widehat{\alpha}_{JK}$ work well for moderate numbers of order statistics used, while $\widehat{\alpha}_{QQ2}$, $\widehat{\alpha}_{GLS}$ have the smallest bias over a relatively large k used and $\widehat{\alpha}_{RK2}$ yields stable estimates regardless of (reasonable) k used. By comparing MAE, simulation study shows that maximum likelihood, generalized jackknife and least squares estimators work well for a small, moderate and larger number of order statistics k used, respectively. In Figure 3.3, we also present similar plots for the same distribution when the sample size is $N = 500$. Analogues conclusions can be drawn in this case as well. By considering stability of median plot and ease of implementation, we would prefer the estimator $\widehat{\alpha}_{RK2}$ in these simulations. Note also that $\widehat{\alpha}_{FH}$ estimator is worst in this setting because of unstable numerical optimization.

Remark. Numerical instabilities associated with FH method are well known. See, for example, p. 770 in Feuerverger and Hall (1999), p. 187 in Beirlant et al. (1999) or p. 7 in Gomes and Martins (2002). More specifically, the instabilities result from a flat likelihood surface which can yield both larger (true) maximum likelihood values for α and local (as opposed to global) maximum values. The problems are especially pronounced when ρ or ρ^* is small (or α is large in our context). These numerical instabilities can and have been somewhat addressed. For example, Gomes and Martins (2002) use an approximate maximum likelihood which yields an explicit solution. Beirlant et al. (1999) use least squares method in the spirit of FHn method described in Section 3.2.5. Both methods (Gomes and Martins (2002), and Beirlant et al. (1999)) require a plug-in estimate for ρ. In our context, using the Hill estimator as the plug-in

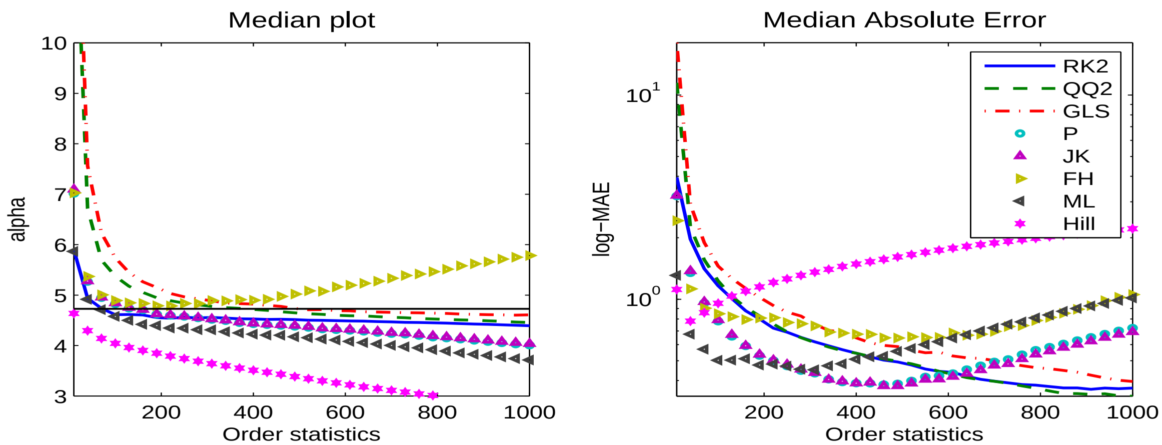

Figure 3.4: ARCH(1) with $\alpha_\xi = 4.73$ and sample size $N = 5,000$.

leads to satisfactory estimators in our simulations. In particular, we find that the estimator based on Gomes and Martins (2002) performs similarly to generalized jackknife estimator, and that based on Beirlant et al. (1999) similarly to least squares estimators. Finally, let us mention that, in contrast to FH method, conditional maximum likelihood estimator is stable in numerical optimization.

3.4.2 Results for ARCH model

Recall from Section 2.2.1 that ARCH(1) model is defined by

$$\xi_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \beta + \eta \xi_{t-1}^2, \tag{3.86}$$

where $\{\epsilon_t\}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables and $\beta > 0$, $\eta > 0$ are parameters. If ξ denotes a stationary distribution of (3.86), then from the results proved in Chapter 2, the distribution tail of ξ^2 is expected to satisfy (3.9) (with $c_2 < 0$ in particular). Moreover, if the tail distribution of ξ^2 satisfies (3.9), then that of ξ satisfies

$$P(\xi > x) = \frac{c_1}{2} x^{-2\alpha} + \frac{c_2}{2} x^{-2\alpha-2} + o(x^{-2\alpha-2}) \tag{3.87}$$

and corresponding to (3.8) with $\rho = -2$.

In simulations, we have chosen parameters $\eta = .5$, $\beta = 1$ and $\sigma^2 = 1$. This yields tail exponents $\alpha = 4.73/2$ and $\alpha_\xi = 4.73$ for ξ^2 and ξ, respectively. The simulations are based on

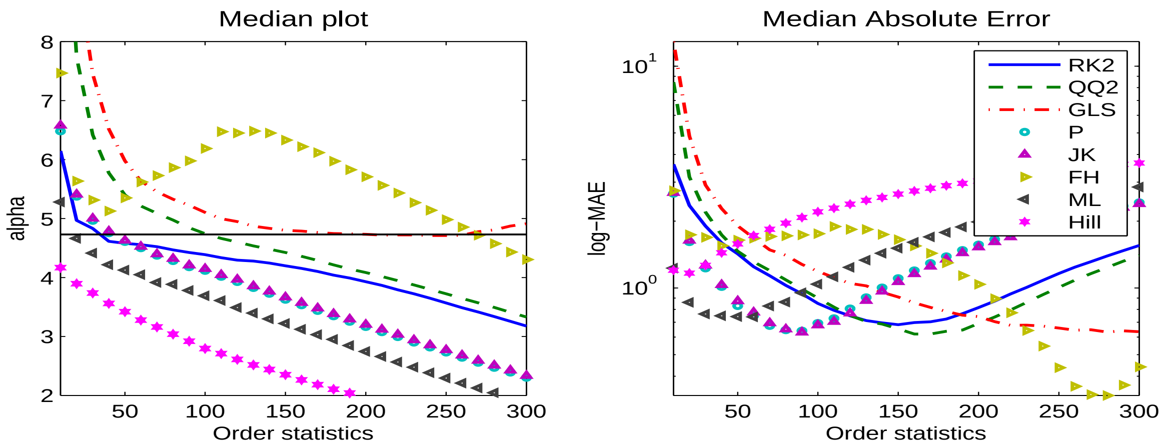

Figure 3.5: ARCH(1) with $\alpha_\xi = 4.73$ and sample size $N = 500$.

independent copies of ξ^2, not the time series data (see also remark below). Since one is more interested in the series itself, we report performance of estimator for α_ξ (which are obtained from those of α through (2.25)). The results are plotted in Figures 3.4 and 3.5 for sample sizes $N = 5,000$ and $N = 500$, respectively. Analogues observations can be made here as in the case of beta prime distribution. Small differences are that $\widehat{\alpha}_{FH}$ performs well around $k = 200$ in Figure 3.4, and $\widehat{\alpha}_{GLS}$ works quite well in Figure 3.5.

Remark. Simulations for ARCH(1) model above are based on independent copies of a stationary solution. (In fact, we take independent copies of ξ_T in (3.86) for large T but these can be considered as good approximations to the stationary solution as discussed in Appendix A.) We removed temporal dependence in order not to confuse the effects of temporal dependence and second order terms of distribution tails. If temporal dependence is also taken into account, then all considered estimators generally perform worse than when observations are independent. This is also briefly discussed in Section 3.5.2.

3.4.3 Results for multiplicative cascades

Recall from Section 2.2.3 that a multiplicative cascade is a random measure defined on a unit interval in the following manner. Denote a unit interval as $T = [0, 1)$. For $k_i \in \{0, 1\}$, $i \geq 1$, subintervals of T obtained by splitting in a dyadic fashion denote by

$$I_{k_1,\dots,k_n} = \Big[\frac{l}{2^n}, \frac{l+1}{2^n}\Big), \quad l = k_1 2^0 + \dots + k_n 2^{n-1}.$$

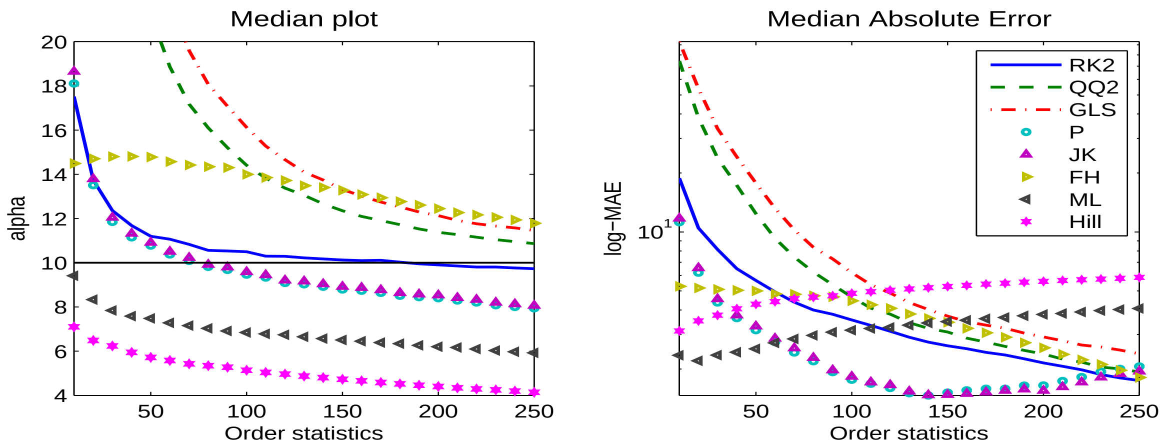

Figure 3.6: MC with $\alpha = 10$ and sample size $N = 1000$.

Let also $\{W_{k_1,\ldots,k_i}, k_i \in \{0,1\}, i \geq 1\}$ be a family of i.i.d., nonnegative, mean 1 random variables, called multipliers. Define a random measure λ_n on $\mathcal{B}(T)$ by

$$\lambda_n(E) = \int_E f_n(t)dt, \quad \text{with } f_n(t) = \sum_{k_i \in \{0,1\}} \left(\prod_{i=1}^{n} W_{k_1,\ldots,k_i} \right) 1_{I_{k_1,\ldots,k_n}}(t).$$

Note, in particular, that

$$\lambda_n(I_{k_1,\ldots,k_n}) = 2^{-n} \prod_{i=1}^{n} W_{k_1,\ldots,k_i}.$$

Under mild assumptions on multipliers, it is known that the sequence λ_n converges weakly to a random measure λ_∞ on $\mathcal{B}(T)$ almost surely, that is,

$$\lambda_n \Rightarrow \lambda, \text{ on } \mathcal{B}(T) \text{ a.s.}$$

where $\Rightarrow$ indicates weak convergence. The limiting random measure λ_∞ is known as a multiplicative cascade (MC, in short). Distribution tail of MC has a power tail as described in Theorem 2.2.1.

For the simulations, we have chosen log-normal multipliers $LN(-\sigma^2/2, \sigma^2)$ with $\sigma^2 = .2 \log 2$ and sample size $N = 1000$. This gives the corresponding tail exponent $\alpha = 10$ from a small calculation in Appendix B by using Theorem 2.2.1. Figure 3.6 shows various tail exponent estimators discussed in this chapter. Similar observations can be made here as in Sections 3.4.2 and 3.4.3. Note, in particular, that RK2 estimator performs best in this simulation.

3.5 Other issues

In this section, we discuss generalized Pareto distribution and the effect of temporal dependence on the tail exponent estimation of RDEs. We contrast our estimation methods in correctly specified models to those that assume unknown ρ, and also examine our estimation methods on misspecified models. We also compare our estimation methods to that of Smith (1987).

3.5.1 Generalized Pareto distribution

One other popular family of distributions for power-law tail behavior consists of Generalized Pareto distributions (GPDs). Parameterized by the parameters $\alpha > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$, GPD(α, μ, σ) has distribution tail given by

$$\overline{F}(x) = \left(1 + \frac{x-\mu}{\alpha\sigma}\right)^{-\alpha}, \quad x > \mu. \tag{3.88}$$

It has the tail exponent α. Note also that, as in the case of RDEs discussed in Chapter 2, GPD has similar second order term, namely,

$$\log \overline{F}(x) = -\alpha \log\left(1 + \frac{x-\mu}{\alpha\sigma}\right)$$

$$= -\alpha \log\left(\frac{x-\mu}{\alpha\sigma}\right) - \frac{\alpha^2\sigma}{x-\mu} + o(x^{-1})$$

$$= \alpha \log(\alpha\sigma) - \alpha \log x - \frac{\alpha^2\sigma - \alpha\mu}{x} + o(x^{-1}). \tag{3.89}$$

This intriguers interesting question to ask whether GPD shares the same problems for larger values of α. If this is the case, then fitted values of large α should be interpreted with care. For example, if data were generated by exact GPD, the values of α fitted by MLE and that from the Hill plot would be quite different.

Similar to Section 2.2.2, consider deviations from the true Pareto tail as

$$\left|\frac{\overline{F}(x)}{((x-\mu)/\alpha\sigma)^{-\alpha}} - 1\right| = \left|\left(1 - \frac{\alpha\sigma}{x-\mu+\alpha\sigma}\right)^{\alpha} - 1\right| \leq \epsilon. \tag{3.90}$$

For example, taking $\alpha = 10$, $\sigma = 2$, $\mu = 1$ and $\epsilon = .5$ gives numerical solution to (3.90)

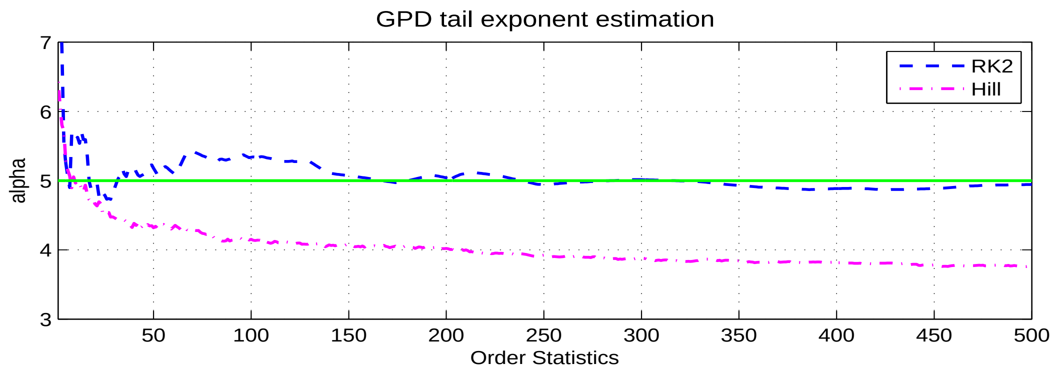

Figure 3.7: GPD (5, 1, 2) tail exponent estimation with $R = 200,000$ observations.

as $x > 279.05$. The probability of having GPD observations in this range is approximately 1.85×10^{-12}. If we increase σ to $\sigma = 3$, then $x > 418.92$ and the corresponding probability is approximately 3.95×10^{-14}. This shows that, for larger values of α, GPD has the Pareto-like region too far in the tail for practical purposes as well.

Figure 3.7 presents a Hill plot for $R = 200,000$ independent observations from GPD(5, 1, 2), and a similar, superimposed plot based on least squares (RK2) taking the second term in (3.89) into account. Note that, even for such large sample size, Hill estimate is very biased. The RK2 estimator performs much better.

3.5.2 Temporal dependence

In this subsection, we study the effect of temporal dependence in estimating tail exponent. In brief, with temporal dependence, estimation is worse than that for independent observations. Figure 3.8 shows tail exponent estimation in ARCH(1) model with $\alpha = 8$. We generated $20,000$ and $50,000$ dependent ARCH(1) observations from (2.22) with first 200 observations disregarded for convergence. For comparison, we also generated $20,000$ independent observations. Note from the figure that RK2 estimator works well in the independent case, and its performance is worse for dependent observations. Increasing the sample size to $50,000$, the dependent case resembles that with $20,000$ independent observations. Note also that the simple Hill estimation is poor in both independent and dependent cases, the dependent case being worse.

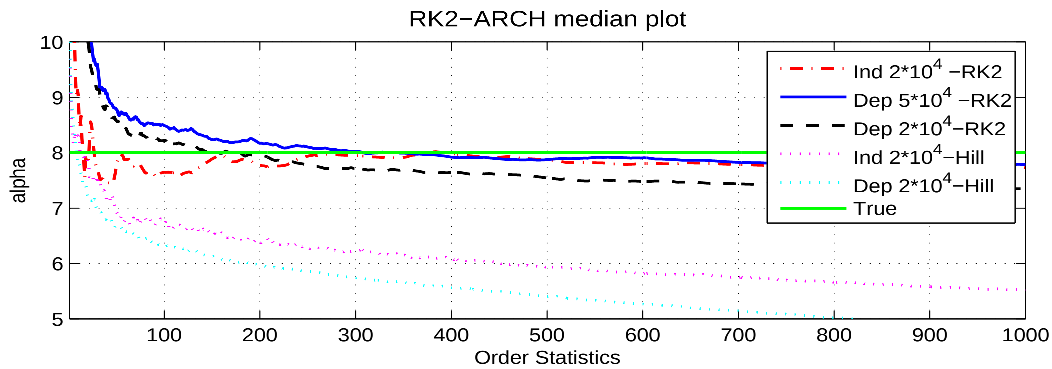

Figure 3.8: Tail exponent estimations for dependent and independent ARCH(1) series.

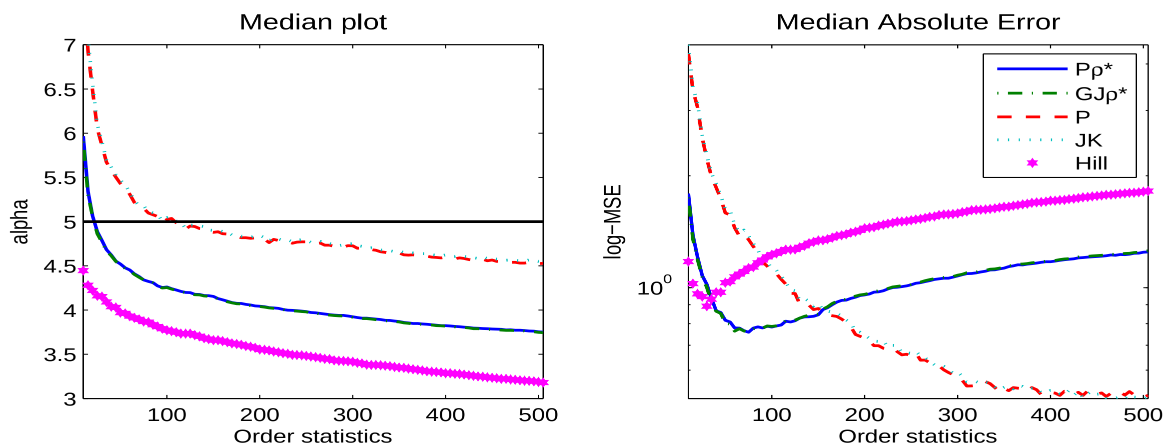

Figure 3.9: $\beta(9, 5)$ with sample size $N = 5,000$.

3.5.3 Estimators with estimated second order tail parameter ρ^*

We compare here the performance of our proposed estimator to those of estimators with estimated second order tail parameter ρ^*. Since FH estimator is numerically unstable, we compare generalized jackknife estimators instead. Gomes and Martins (2002) suggest to estimate ρ^* as

$$\widehat{\rho}^* = -\left|\frac{3(T(k_1) - 1}{(T(k_1) - 3)}\right|, \tag{3.91}$$

where

$$T(k_1) = \frac{M_1 - (M_2/2)^{1/2}}{(M_2/2)^{1/2} - (M_3/6)^{1/3}}, \quad M_j = \frac{1}{k_1}\sum_{i=1}^{k_1} \log^j \left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right),$$

with the choice of $k_1 = \min(n - 1, [2n/\log\log n])$.

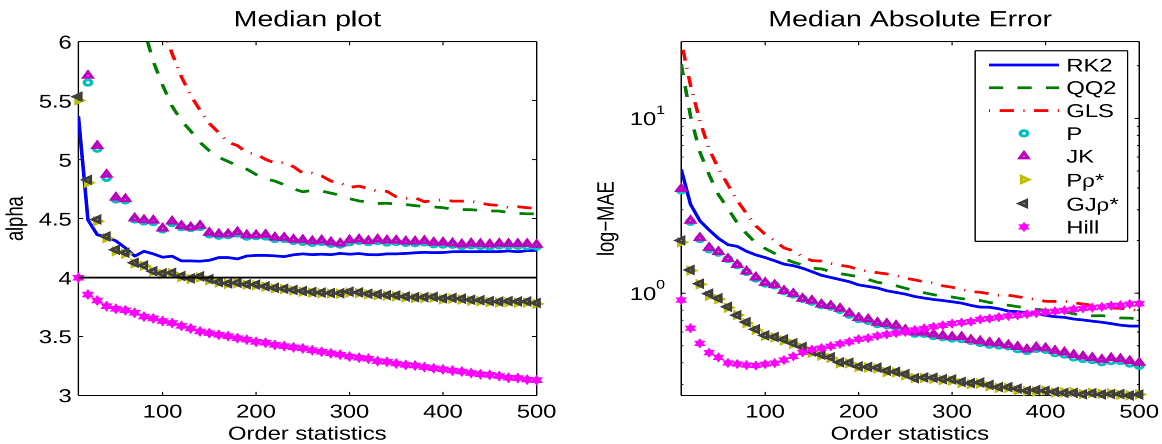

Figure 3.10: Burr$(4, -.5)$ with sample size $N = 5,000$ $(\rho = -2)$.

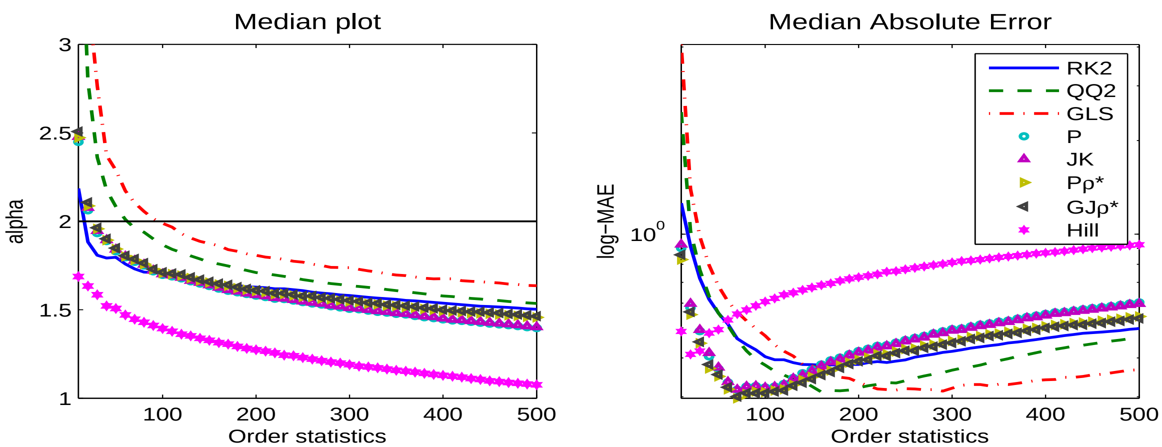

Figure 3.11: Burr$(2, -.25)$ with sample size $N = 5,000$ $(\rho = -.5)$.

Figure 3.9 compares performance of our suggested generalized jackknife estimators $\widehat{\alpha}_P$, $\widehat{\alpha}_{JK}$ and the generalized jackknife estimators based on estimated ρ^* in (3.37) and (3.39) as

$$\widehat{\alpha}_{P\rho^*} = \frac{\widehat{\alpha}^{(1)} - (1-\widehat{\rho}^*)\widehat{\alpha}^{(2)}}{\widehat{\rho}^*}, \quad \widehat{\alpha}_{GJ\rho^*} = \frac{(2-\widehat{\rho}^*)\widehat{\alpha}^{(2)} - 2\widehat{\alpha}^{(3)}}{-\widehat{\rho}^*}, \tag{3.92}$$

respectively. Generalized jackknife estimators with estimated ρ^* certainly perform better than the Hill estimator, but not as good as our proposed jackknife estimators taking into accounts an exact second order tail parameter.

On a different side, suppose that the second order tail parameter ρ is not -1. Consider, for example, the Burr model with the distribution tail

$$\overline{F}(x) = \left(1 + x^{-\rho^* \alpha}\right)^{1/\rho^*}, \quad x \geq 0, \quad \alpha > 0, \quad \rho^* < 0. \tag{3.93}$$

Figures 3.10 and 3.11 illustrate the performance of our proposed estimators and the generalized jackknife methods with estimated ρ^* for the Burr model with $\alpha = 2$, $\rho^* = -.5$ ($\rho = -2$) and $\alpha = 2$, $\rho = -.25$ ($\rho = -.5$). These plots are representatives of what to expect for $\rho < -1$ and $\rho > -1$. The generalized jackknife methods using estimated ρ^* performs best among all estimators. Performance of our estimators (especially rank-based and generalized jackknife) is less sensitive to misspecifications with $\rho < -1$.

3.5.4 Comparison with the estimator of Smith (1987)

Smith (1987) proposed tail exponent estimator based on the excesses over threshold. The conditional distribution function of excess $X - u$ given $X > u$ is

$$F(y|u) := \frac{F(u+y) - F(u)}{1 - F(u)}.$$

According to Pickands (1975), generalized Pareto distribution (GPD) provides a good approximation of $F(y|u)$ for fairly general class of distributions F, including those with a regularly varying tail as in (3.2). Smith (1987) replaces threshold u by upper order statistic $X_{(n-k)}$ and uses maximum likelihood method for GPD parameters to estimate the tail exponent α from the series

$$Y_{(n-i+1)} = X_{(n-i+1)} - X_{(n-k)}. \tag{3.94}$$

Denote the corresponding estimator of α by $\widehat{\alpha}_{GPD}$. In particular, only for the special case of $\rho = -1$ in the second order regular variation framework (3.7), $\widehat{\alpha}_{GPD}$ is asymptotically unbiased in the sense that

$$\sqrt{k}\left(\widehat{\alpha}_{GPD} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \alpha^2(\alpha^2+1)\right) \tag{3.95}$$

(see, Theorem 3.2 and (4.2) of Smith (1987)).

Figure 3.12 compares the performance of two estimators, $\widehat{\alpha}_{GPD}$ and $\widehat{\alpha}_{RK2}$. Simulation is for ARCH(1) model with $\alpha_\xi = 4.73$ and the sample size $N = 5,000$. It can be seen that

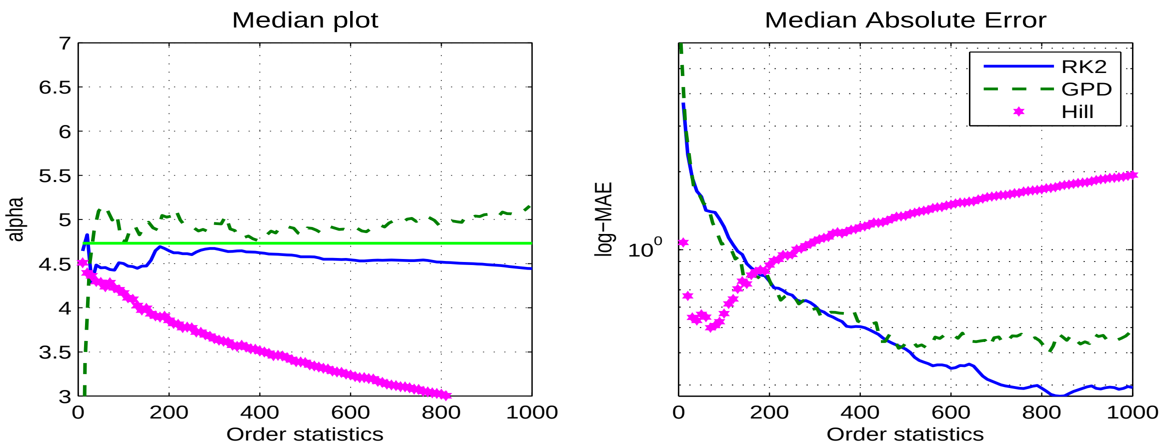

Figure 3.12: Median and MAE plot for ARCH(1) model with $\alpha_\xi = 4.73$ and sample size $N = 5,000$.

$\widehat{\alpha}_{GDP}$ also successfully reduces the bias. However, in terms of median absolute error, RK2 estimator achieves smaller MAE when sufficiently large number of upper order statistics are used in calculation.

3.6 Conclusions

In this chapter, we studied various tail parameter estimators in the second order framework with known second order tail parameter. Assuming known second order parameter was largely motivated by the result discussed in Chapter 2. Second order term in distribution tails can significantly affect the bias of common tail estimators and this chapter shows how the bias can satisfactorily be removed (supposing a known second order tail parameter).

Among various estimators considered, we generally found least squares estimators, especially the rank-based estimator, perform the best. They consistently show smallest bias over a large range of upper order statistics considered, and are easy to implement. These estimators would also lead to more reliable confidence intervals for larger number of upper order statistics, despite their asymptotic variance being the largest in theory (among the estimators considered). In the next order of our preference go generalized jackknife and conditional maximum likelihood estimators. The other, FH maximum likelihood estimator has nice theoretical properties but generally suffers from numerical instabilities in practice.

CHAPTER 4

Long range dependence, unbalanced Haar wavelet transformation and changes in local mean level

4.1 Introduction

Long range dependent (LRD) or long memory time series models are commonly defined as weakly stationary time series with autocorrelation function that decays slowly as a power function for large time lags. Dependence in LRD time series remains strong even between times that are far apart. See, for instance, the collection of articles in Doukhan, Oppenheim and Taqqu (2003). Long range dependence is closely related to self-similar processes of which fractional Brownian motion (FBM) is the best known example.

First LRD time series and related FBM were introduced by Mandelbrot and Van Ness (1968) as models explaining the Hurst phenomenon observed with data of river flows. Granger and Joyeux (1980), Hosking (1981) introduced LRD models known as FARIMA time series, generalizing the class of popular ARMA models of Box and Jenkins. These models and their extensions have become popular in Economics, Finance and other areas. More recently, LRD time series have become particularly relevant as data traffic models in modern communication networks such as Internet (Park and Willinger (2000)).

The top plot of Figure 4.1 depicts the celebrated time series of the level of the Nile river studied by Hurst (1951). The bottom plot of that figure depicts a simulated fractional Gaussian Noise (FGN) time series which was originally proposed as a model for the Nile river. (FGN is defined as the series of stationary increments of FBM.) The characteristic feature of these plots, readily apparent from Figure 4.1, is the presence of local cycles or changes in local mean level (across a wide range of scales), even though a time series itself does not exhibit a global

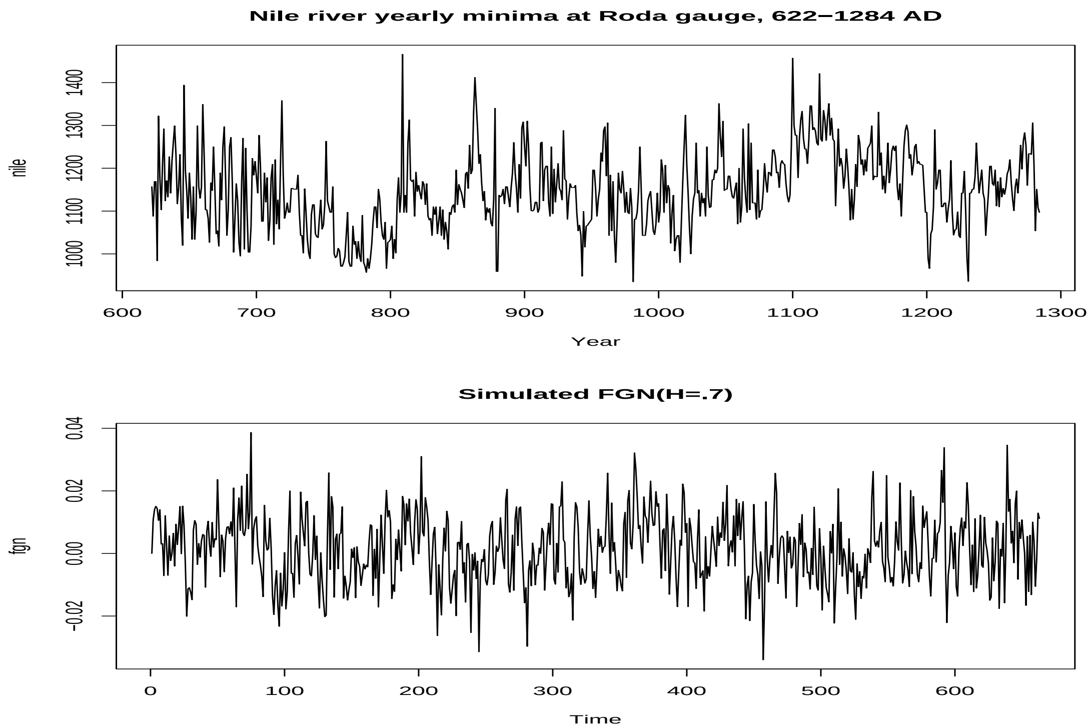

Figure 4.1: The Nile river yearly minima at Roda gauge (top) and simulated FGN (bottom).

trend. It is precisely this feature of an observed time series that long range dependence aims to capture.

Needless to say, these apparent nonstationarities and the idea of persistent memory have caused and still causes quite a debate in most of the areas where LRD time series are used. Related questions are studied in at least several directions. A number of authors have provided non-LRD-like models exhibiting LRD-like features (in particular, suggesting LRD through commonly used statistical estimation methods). See, for example, Klemeš (1974), Bos, Franses and Ooms (1999), Diebold and Inoue (2001), Veres and Boda (2000), Mikosch and Stărică (2004), Granger and Hyung (1999), to name but a few. Others have studied classical change point estimation methods on LRD time series (for example, Kuan and Hsu (1998), Wright (1998)). Yet others, more recently, have proposed ways to distinguish statistically between stationary LRD time series and various alternatives. See, for example, Jach and Kokoszka (2008), Hariz, Wylie and Zhang (2007), Berkes, Horváth, Kokoszka and Shao (2006). Despite all this work, the use of LRD models still causes much debate in applications, and there do not seem to be many conclusive answers.

The main purpose of this work is to contribute to understanding of the relation among LRD time series, changes in local mean level and non-LRD-like alternatives. We mentioned above that LRD time series aim to capture changes in local mean level across a wide range of scales observed in real time series. We will focus on the local mean level changes, in either real time series or their LRD models. Some basic questions we are interested in are:

- Do LRD time series models capture these changes in local mean level well? How can one measure this?
- What are the properties of changes in local mean level in LRD time series models? How do they compare with those in real time series?
- What differences are there between using non-LRD-like models for changes in local mean level and LRD time series models?

We are interested in these questions in the context of *long time series* such as those appearing in data traffic over Internet, not of time series of several hundred or few thousands observations encountered in Economics and other areas. Though some of our conclusions will also be relevant for the latter case.

A natural tool to address the questions above is Unbalanced Haar Transformation (UHT). UHT was recently popularized by Fryzlewicz (2007), though the concept goes back to earlier works. The idea behind UHT is simple. UHT decomposes time series in finer and finer regimes of changes in local mean level. It starts by identifying the first break point in a whole time series (by a suitably chosen criterion) and then proceeds recursively with the interval up to the first break point and that after the break point. UHT essentially consists of the collection of all break points and UHT detail coefficients that measure difference in level change from one regime to next at some scale. This collection allows to recover the original time series, and thus contains all the information about it. Importantly, this information is exactly related to changes in local mean level. UHT is described in detail in Section 4.2.2.

With UHT, partitioning into finer regimes continues, in principle, till no further division of interval is possible. By ignoring small UHT detail coefficients or denoising the time series in the UHT domain, one obtains a new time series (denoised time series) with longer intervals of constant local mean level. It can be thought as depicting changes in local mean level of the original time series. The nice feature of the denoised time series is that it is described by durations (inter-arrivals, runs) of changes in local mean level and mean levels (jumps from zero) themselves whose properties could now be studied.

By applying UHT described above, our main findings through simulations is that denoised LRD time series is described by inter-arrivals (runs) of nearly homogeneous Poisson process, and LRD jumps from zero. (Simulation are used because denoised LRD series are difficult, if not impossible to analyze in theory with available tools.) In particular, jumps from zero is the only feature of denoised series that inherits the original LRD structure. The difference between the original series and its denoised series, called residuals, are short range dependent (SRD). These findings are quite robust with respect to long range dependence parameter and denoising level used in UHT.

We also show in theory that, from the reverse angle and under mild assumptions, the time series generated by the above mechanism is LRD. More specifically (see Proposition 4.3.1 below), suppose $\{Y_n\}_{n\geq 0}$ is a stationary LRD time series, $\{X_n\}_{n\in\mathbb{Z}}$ is a stationary SRD time series, $\{U_k\}_{k\geq 1}$ is a sequence of i.i.d. inter-arrivals taking (positive) integer values and set $S_n = \sum_{k=0}^{n} U_k$. Then, under mild assumptions, the series

$$W_n = \sum_{k=0}^{\infty} Y_k 1_{\{S_k \leq n < S_{k+1}\}} + X_n, \quad n \geq 0, \tag{4.1}$$

is a stationary LRD time series, having the same LRD parameter as the series $\{Y_n\}_{n\geq 0}$.

We compare above findings based on simulations with several real data sets that include two Internet traffic packet series, the celebrated mount Campito tree ring data and the squared log returns of S&P 500 index. One of the Internet traffic time series exhibits almost identical features when compared with simulated data. The other Internet traffic time series and Campito tree ring data are less comparable with those of simulated data, though in favor of LRD models. The comparison is also made with several non-LRD-like models exhibiting LRD features for finite sample sizes previously considered in the literature. The analysis based on UHT reveals that these models are quite different from LRD models. The key difference is that jumps from zero (or its squares) no longer exhibit LRD. The same conclusion is also reached with the UHT analysis for the squared log returns of S&P 500 index, suggesting that LRD models may not be suitable for this series.

UHT is not the only way to define changes in local mean level. For example, these changes could be defined using simple kernel smoothing, based on the signs of the series or even simple orthogonal Haar transformation.

- Why then pay particular attention to UHT?

It turns out that this transformation seems to be the one among available methods that leads to simplest structure of inter-arrivals and jumps from zero of the series of changes in mean levels. This might explain why earlier efforts to look into runs were not pursued (see, for example Mandelbrot and Wallis (1969), p. 34).

Finally, of independent interest, we study basic properties of UHT for LRD time series. These include stationarity, asymptotics as the sample size increases, and multiresolution properties. For example, the aforementioned asymptotics are not too difficult to establish assuming a suitable functional central limit result for the underlying LRD time series. Multiresolution properties turn out to be more involved than analogous properties when using orthogonal wavelet transform instead of UHT.

Chapter 4 is organized as follows. LRD time series and UHT are discussed in Section 4.2. In that section, we also describe denoising in the UHT domain and its connection to changes in local mean level. In Section 4.3, we then examine this procedure on LRD time series, and compare our results for simulated data with examples of real and other simulated time series. Comparison with other ways to define runs can also be found in that section. Some properties of UHT for LRD time series are studied in Section 4.4. These include questions of stationarity, behavior across scales, dependence. Conclusions are summarized in Section 4.5.

4.2 Preliminaries

In this section, we define long range dependent (LRD) processes, self-similarity and unbalanced Haar transformation (UHT).

4.2.1 Long range dependence and self-similarity

Long range dependent (LRD) time series $X = \{X_n\}_{n\in\mathbb{Z}}$ is typically defined as a weakly stationary time series with an autocovariance function

$$\gamma_X(k) = \mathrm{Cov}(X_0, X_k) = L(k)k^{-\beta}, \quad \beta \in (0,1), \tag{4.2}$$

as $k \to \infty$, for a slowly varying function $L : \mathbb{R} \to \mathbb{R}$ at infinity (Beran (1994), Doukhan et al. (2003)). The decay of autocovariance function (4.2) is so slow that $\sum_k |\gamma_X(k)| = \infty$. LRD time series can also be defined in the Fourier domain through a diverging spectral density function at zero.

A typical example of LRD time series is a FARIMA$(0, s, 0)$, $s \in (0, 1/2)$, time series $X = \{X_n\}_{n\in\mathbb{Z}}$ defined as

$$X_n = (I - B)^{-s}\epsilon_n = \sum_{k=0}^{\infty} \psi_k \epsilon_{n-k}, \tag{4.3}$$

where B is a backshift operator, $I = B^0$ is an identity operator, and the coefficients ψ_k come from the Taylor expansion $(1-z)^{-s} = \sum_{k=0}^{\infty} \psi_k z^k$. The long memory parameter for these time series is

$$\beta = 1 - 2s \in (0,1). \tag{4.4}$$

We will often use the s-parametrization for LRD time series in which case β in (4.2) will be replaced by (4.4) and $s \in (0, 1/2)$. The uncorrelated innovations $\{\epsilon_n\}$ are often taken Gaussian in which case the resulting FARIMA time series is Gaussian as well.

LRD time series are closely related to self-similar processes. Recall that a stochastic process $X = \{X(t)\}_{t\in\mathbb{R}}$ is self-similar if, for all $c > 0$,

$$\{X(ct)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{c^H X(t)\}_{t\in\mathbb{R}}, \tag{4.5}$$

where $H > 0$, called the self-similarity parameter, and $\stackrel{d}{=}$ denotes equality of finite-dimensional distributions. Self-similar processes of particular interest also have stationary increments. The best known example of such processes is fractional Brownian motion (FBM). FBM $B_H = \{B_H(t)\}_{t\in\mathbb{R}}$ is the only (up to constant) Gaussian, zero mean process which has stationary increments and is self-similar with parameter $H \in (0,1)$. The stationary increments of FBM, known as fractional Gaussian noise (FGN), are LRD when $H \in (1/2, 1)$, with the corresponding parameter

$$\beta = 2 - 2H \in (0,1). \tag{4.6}$$

Another connection between LRD time series and FBM is the following. It is known that, under suitable but quite general assumptions on LRD time series $X = \{X_n\}_{n\in\mathbb{Z}}$,

$$\frac{1}{n^H \widetilde{L}(n)} \sum_{k=1}^{[nt]} X_k \to B_H(t), \tag{4.7}$$

where $\widetilde{L}$ is another slowly varying function at infinity. The convergence (4.7) holds, for example, for Gaussian, zero mean LRD time series such that $\mathrm{Var}(\sum_{k=1}^{n} X_k) = n^{2H}(\widetilde{L}(n))^2$, and it takes place in the sense of convergence of finite dimensional distributions (Doukhan et al. (2003), p. 17).

4.2.2 Unbalanced Haar wavelet transformation

We first consider the case of discrete time and adopt the notation used by Fryzlewicz (2007). Unbalanced Haar transformation (UHT) is based on the unbalanced Haar (UH) wavelet function

$$\psi_{s,b,e}(k) = \Big(\frac{1}{b-s+1} - \frac{1}{e-s+1}\Big)^{1/2} 1_{\{s\le k\le b\}} - \Big(\frac{1}{e-b} - \frac{1}{e-s+1}\Big)^{1/2} 1_{\{b+1\le k\le e\}}. \tag{4.8}$$

Here, s, b and e denote the starting, break and end points, respectively. Observe that $\sum_k \psi_{s,b,e}(k) = 0$ and $\sum_k(\psi_{s,b,e}(k))^2 = 1$. The function $\psi_{s,b,e}$ generalizes the usual Haar wavelet where $e - s + 1$ is a power of 2, and b corresponds to a midpoint.

For a finite time series $X = \{X_1, \dots, X_N\}$, let

$$\langle X, \psi_{s,b,e}\rangle = \Big(\frac{1}{b-s+1} - \frac{1}{e-s+1}\Big)^{1/2} \sum_{k=s}^{b} X_k - \Big(\frac{1}{e-b} - \frac{1}{e-s+1}\Big)^{1/2} \sum_{k=b+1}^{e} X_k \tag{4.9}$$

be its inner product with UH wavelet $\psi_{s,b,e}$, $1 \le s \le e \le N$. First, set $s_{0,1} = 1$, $e_{0,1} = N$. Since the inner product (4.9) measures closeness of UH wavelet to X, it is natural to define the first break point as

$$b_{0,1} = \underset{b}{\operatorname{argmax}} \left| \left\langle X, \psi_{s_{0,1},b,e_{0,1}} \right\rangle \right| . \tag{4.10}$$

Then, for $j \geq 0$ and $k \in \{1, \ldots, 2^j\}$, proceed recursively as follows. If $b_{j,k} - s_{j,k} \geq 1$, set $s_{j+1,2k-1} = s_{j,k}$ and $e_{j+1,2k-1} = b_{j,k}$. If $e_{j,k} - b_{j,k} \geq 2$, set $s_{j+1,2k} = b_{j,k} + 1$, $e_{j+1,2k} = e_{j,k}$. In either case, with $l = 2k - 1$ or $l = 2k$, set also

$$b_{j+1,l} = \underset{b}{\operatorname{argmax}} \left| \left\langle X, \psi_{s_{j+1,l},b,e_{j+1,l}} \right\rangle \right| . \tag{4.11}$$

This procedure can be continued as long as possible. In particular, for fixed j, some of $s_{j,k}$, $b_{j,k}$ and $e_{j,k}$ may not be defined. Let also

$$\psi_{j,k} = \psi_{s_{j,k},b_{j,k},e_{j,k}}, \tag{4.12}$$

$$d_{j,k} = \langle X, \psi_{j,k} \rangle \tag{4.13}$$

be the corresponding wavelet functions and detail (wavelet) coefficients. The above procedure is known as unbalanced Haar transformation (UHT), with a particular choice of break points (4.11). By convention, UHT detail coefficients are set to zero where there is no break.

UHT can be thought as having a multiscale (multiresolution) structure with small (large, respectively) j associated with coarse (fine, respectively) scales. Moreover, a finite time series X can be recovered from the UH wavelet function (4.12) and UHT detail coefficients (4.13) as

$$X_n = \sum_{j,k} d_{j,k} \psi_{j,k}(n), \quad n = 1, \ldots, N, \tag{4.14}$$

if, in addition, one sets $\psi_{-1,1}(k) = N^{-1/2} 1_{\{1 \leq k \leq N\}}$. Letting $\overline{X}_{m,n} = \frac{1}{n-m+1} \sum_{k=m}^{n} X_k$, one also has

$$\overline{X}_{s_{j,k},e_{j,k}} = \sum_{j' < j, k'} d_{j',k'} \psi_{j',k'}(n), \quad n = s_{j,k}, \ldots, e_{j,k}. \tag{4.15}$$

In view of (4.15) and the definition of (4.8), UHT is a natural procedure to decompose a finite time series through changes in local mean level at various scales j.

Remark. UHT is closely related to the Classification and Regression Tree (CART) procedure

(Hastie, Tibshirani and Friedman (2001)). Following the above presentation of UHT for comparison, CART works as follows. Let $s^c_{0,1} = 1$ and $e^c_{0,1} = N$. The first break point in CART is defined as

$$
\begin{aligned}
b^c_{0,1} &= \operatorname*{argmin}_b \left\{ \min_{c_1} \sum_{k=s^c_{0,1}}^{b} (X_k - c_1)^2 + \min_{c_2} \sum_{k=b+1}^{e^c_{0,1}} (X_k - c_2)^2 \right\} \\
&= \operatorname*{argmin}_b \left\{ \sum_{k=s^c_{0,1}}^{b} (X_k - \overline{X}_{s^c_{0,1},b})^2 + \sum_{k=b+1}^{e^c_{0,1}} (X_k - \overline{X}_{b+1,e^c_{0,1}})^2 \right\}, \qquad (4.16)
\end{aligned}
$$

where $\overline{X}_{m,n} = \frac{1}{n-m+1}\sum_{k=m}^{n} X_k$. Then if $b^c_{j,k} - s^c_{j,k} \geq 1$, set $s^c_{j+1,2k-1} = s^c_{j,k}$ and $e^c_{j+1,2k-1} = b^c_{j,k}$, and if $e^c_{j,k} - b^c_{j,k} \geq 2$, set $s^c_{j+1,2k} = b^c_{j,k} + 1$, $e^c_{j+1,2k} = e^c_{j,k}$. In either case, with $l = 2k-1$ or $l = 2k$, set also

$$
b^c_{j+1,l} = \operatorname*{argmin}_b \left\{ \sum_{k=s^c_{j+1,l}}^{b} (X_k - \overline{X}_{s^c_{j+1,l},b})^2 + \sum_{k=b+1}^{e^c_{j+1,l}} (X_k - \overline{X}_{b+1,e^c_{j+1,1}})^2 \right\}. \qquad (4.17)
$$

The CART procedure above can be reformulated by using UH wavelet in (4.8). Indeed, one can rewrite (4.17) as

$$
b^c_{j+1,l} = \operatorname*{argmax}_b \sum_{k=s^c_{j+1,l}}^{e^c_{j+1,l}} \Big(X_k - \overline{X}_{s^c_{j+1,l},e^c_{j+1,l}} - \langle X, \psi_{s^c_{j+1,l},b,e^c_{j+1,l}} \rangle \psi_{s^c_{j+1,l},b,e^c_{j+1,l}}(k) \Big)^2. \qquad (4.18)
$$

Moreover, the relation (4.15) holds with s, e replaced by s^c, e^c. In other words, the CART procedure is also UHT but where the break points are chosen according to (4.18). We shall work with UHT and break points chosen by (4.11) for notational simplicity. But analogous presentation can be developed for CART.

We shall also use the continuous version of UHT on an interval $[0, 1]$. We want to view it as a limiting transformation of UHT in discrete time. It is therefore natural to consider the signal $X(t), t \in [0, 1]$, as an integrator in the inner product below. The UH wavelet is now defined as

$$
\psi_{s,b,e}(t) = \Big(\frac{1}{b-s} - \frac{1}{e-s} \Big)^{1/2} 1_{\{s \leq t < b\}} - \Big(\frac{1}{e-b} - \frac{1}{e-s} \Big)^{1/2} 1_{\{b \leq t < e\}}, \qquad (4.19)
$$

where $0 \leq s < b \leq e \leq 1$, and the corresponding inner products are

$$\langle X, \psi_{s,b,e} \rangle = \int_0^1 \psi_{s,b,e}(t) dX(t) \tag{4.20}$$

(to be correct, the inner product should be denoted as $\langle \dot{X}, \psi_{s,b,e} \rangle$). In continuous time, end points $e_{j,k}$ and starting points $s_{j,k+1}$ can be identified. It is therefore simpler to consider a collection of $T_{j,k}$ defined as follows. For $j = 0, 1, \ldots$, $k = 0, 1, \ldots, 2^{j+1}$,

$$T_{j,0} = 0,\ T_{j,2^{j+1}} = 1,\ T_{j,2k} = T_{j-1,k}, \tag{4.21}$$

$$B_{j,k}(X) = T_{j,2k-1} := \underset{b}{\operatorname{argmax}} \left| \left\langle X, \psi_{T_{j,2k-2},b,T_{j,2k}} \right\rangle \right|, \quad k = 1, \ldots, 2^j. \tag{4.22}$$

Denote also the corresponding detail coefficients as

$$D_{j,k}(X) = \left\langle X, \psi_{T_{j,2k-2},T_{j,2k-1},T_{j,2k}} \right\rangle. \tag{4.23}$$

4.2.3 Denoising with UHT and changes in local mean level

UHT defined in Section 4.2.2 is applied, in principle, till no further splitting (breaks) are possible. Small UHT detail coefficients are associated with small changes in local mean level. If the goal is to describe the evolution of a local mean level in time series, those small detail coefficients are natural to disregard. The resulting procedure, known as denoising, leads to a new time series representing changes in local mean level. More specifically, this procedure can be defined as follows.

Let $X = \{X_1, \ldots, X_N\}$ be a time series, and $d_{j,k}$, $\psi_{j,k}$ be the corresponding detail coefficients and wavelet functions in its UHT. Small detail coefficients can be disregarded in several ways. For example, consider

$$\widetilde{d}_{j,k} = d_{j,k} 1_{\{|d_{j,k}| > \epsilon\}} = \begin{cases} d_{j,k}, & \text{if } |d_{j,k}| > \epsilon, \\ 0, & \text{if } |d_{j,k}| \leq \epsilon, \end{cases} \tag{4.24}$$

corresponding to a hard thresholding of the coefficients. The threshold ϵ is often taken in the wavelet denoising literature as

$$\epsilon = \epsilon_N^{(u)} = \hat{\sigma}\sqrt{2\log N} \tag{4.25}$$

where $\hat{\sigma}$ is Mean Absolute Deviation (MAD) of the sequence $2^{-1/2}|X_{i+1} - X_i|$, $i = 1, \dots, N-1$. The choice (4.25) is known as the universal threshold. The *denoised time series* $\widetilde{X}$ is defined as in (4.14) but using the coefficients $\widetilde{d}_{j,k}$ instead:

$$\widetilde{X}_n = \sum_{j,k} \widetilde{d}_{j,k}\psi_{j,k}(n), \quad n = 1, \dots, N. \tag{4.26}$$

The new series $\widetilde{X}_n$ can be thought as representing changes in local mean level of the original time series X.

The series $\widetilde{X}_n$ can also be represented as

$$\widetilde{X}_n = \sum_{k=0}^{m} \beta_k 1_{\{c_k \le n < c_{k+1}\}}, \tag{4.27}$$

where c_k are times of local mean level changes (with $c_0 = 0, c_m = N$) and β_k represent local mean levels in regimes $(c_k, c_{k+1}]$. We will refer to the lengths $c_{k+1} - c_k$ of the local mean level as *inter-arrival times (*or *inter-arrivals* or *runs).* The coefficients β_k will be referred to as *jump sizes (*or *jumps) from zero.* Alternatively, (4.27) can be rewritten as

$$\widetilde{X}_n = \widetilde{X}_{n-1} + \sum_{k=0}^{m-1} \gamma_k 1_{\{n=c_k\}}, \tag{4.28}$$

where γ_k are changes from the previous local mean level (*changes in mean level*). The series $X_n - \widetilde{X}_n$ will be referred to as the *residuals.*

4.3 Changes in local mean level for LRD models and data

Denoising with UHT is studied here for LRD models (Section 4.3.1), with other alternatives (Section 4.3.2), and for LRD data (Section 4.3.3).

4.3.1 Denoising with UHT of simulated LRD time series

We examine here a denoised time series obtained by UHT from a simulated FARIMA(0,.45,0) time series. Our main results are based on 50 realizations of FARIMA time series of length

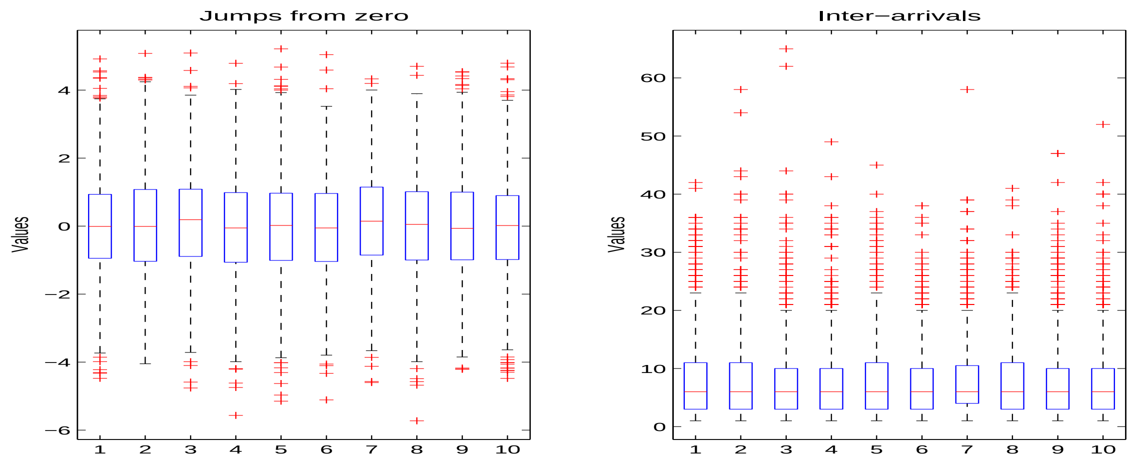

Figure 4.2: Boxplots of jumps from zero and inter-arrivals for the first 10 local mean levels.

$N = 75,000$, and the universal threshold in (4.25). We focus on the properties of inter-arrivals and jumps from zero of a denoised time series. We shall work as if their time series are stationary. This important assumption (or that of nearly stationarity) is difficult to verify in theory but is supported by simulations. For example, Figure 4.2 presents boxplots of first 10 jumps from zero and inter-arrivals obtained from 50 realizations, and shows apparent stationarity of the marginals.

Inter-arrivals: Figure 4.3 shows the histogram, exponential QQ-plot and ACF of inter-arrivals of denoised time series. The ACF was calculated by averaging sample ACF of 50 realizations. Several interesting features of inter-arrivals are seen from these plots. First, inter-arrivals follow closely a two-parameter exponential distribution $\mathrm{Exp}(\mu, \lambda)$ with the density

$$\frac{1}{\lambda} \exp\left(-\frac{x-\mu}{\lambda}\right) 1_{(\mu,\infty)}(x), \quad \lambda > 0,\ \mu \in \mathbb{R},$$

and estimated parameter $\lambda = 8$ for $\mu = 2$. The truncation parameter $\mu = 2$ is found from empirical considerations and this choice means, in particular, that with the universal threshold there are few inter-arrivals of size 1. Two-parameter exponential QQ-plot shows slightly heavier tails in inter-arrivals greater than 60. But the corresponding sample quantile is .9994 and this can be regarded as due to sampling variability. Second, the ACF plot suggests that inter-arrivals are decorrelated. (Decorrelation is also found when considering the series of squares and the absolute values of inter-arrivals.) In summary, these observations suggest, quite surprisingly,

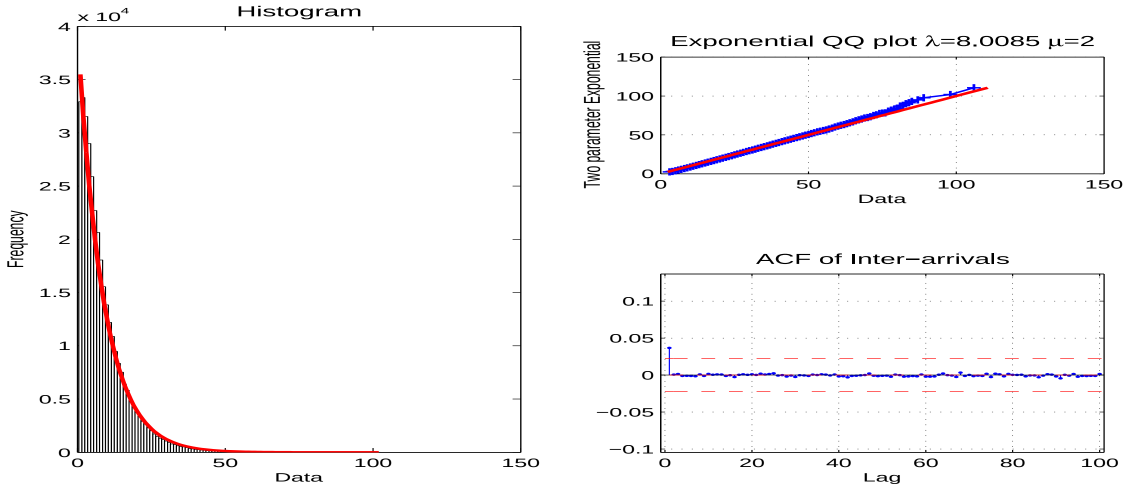

Figure 4.3: Histogram, Exponential QQ-plot and ACF of inter-arrivals.

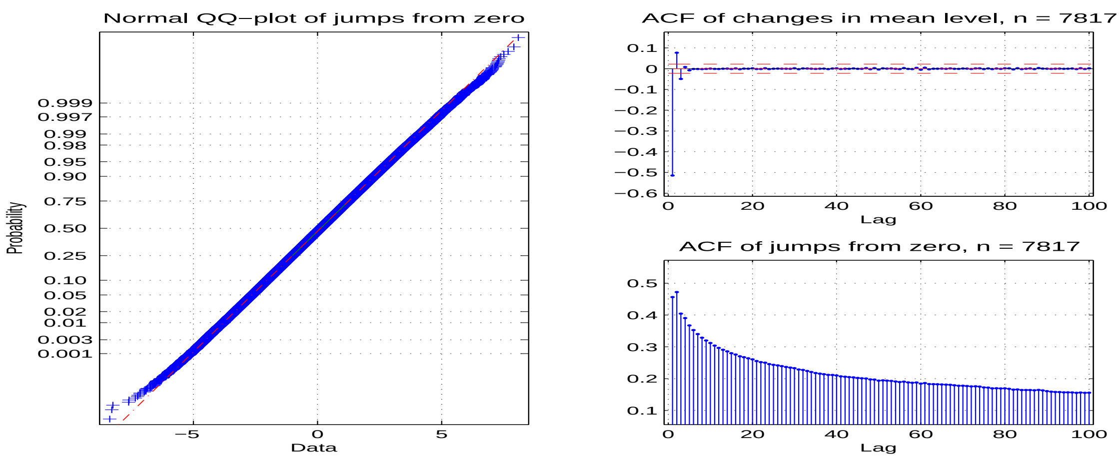

Figure 4.4: Normal QQ-plot and ACFs of changes in local mean level and jumps from zero.

that inter-arrivals occur according to a simple homogeneous Poisson process.

Jumps from zero: Figure 4.4 concerns jump sizes from zero and changes in local mean level. The left plot represents normal QQ-plot of jumps from zero and the plots on the right side present the ACF of changes in mean level and jumps from zero. As expected, there is a lag 1, negative correlation in ACF of changes in local mean level. One of the more important observations is that ACF of jumps from zero exhibits LRD. This is the only characteristic of denoised time series which inherits the LRD property from the original time series. The LRD parameter for jumps from zero was estimated (using the wavelet method of Veitch and Abry (1999)) on average as $\hat{s} = .443$ in 50 realizations, and is close to the original FARIMA LRD

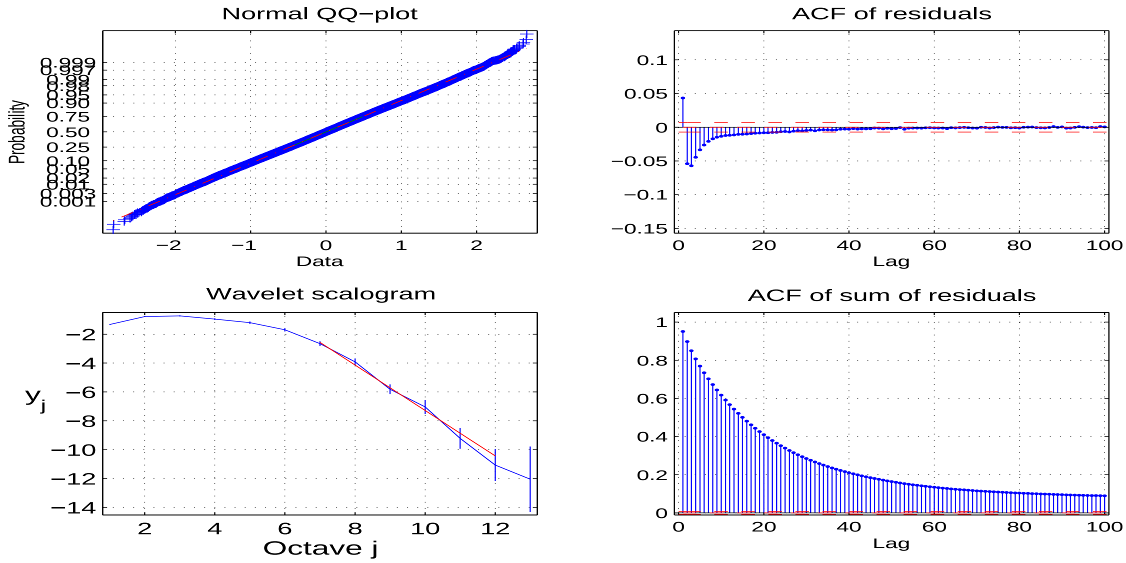

Figure 4.5: Normal QQ-plot, ACF and wavelet scalogram of residuals.

parameter $s = .45$.

Residuals: In Figure 4.5, we study the properties of residuals through the normal QQ-plot, ACF and wavelet scalogram. An interesting feature of the residuals is that ACF is mostly negative except at lag 1. The wavelet-based scaling parameter estimator (through a wavelet scalogram reported in Figure 4.5) yields parameter $\hat{s}$ from about -.9 to -.6. Since adding 1 to such $\hat{s}$ brings it back to the interval .1 to .4, the residuals can be interpreted as a first difference of LRD time series. ACF of the sum of residuals is also given in Figure 4.5.

The findings above suggest that LRD time series could be viewed as:

$$\begin{aligned} LRD\ time\ series &= nearly\ Poisson\ arrivals \\ &+ LRD\ jumps\ from\ zero \\ &+ SRD\ residuals. \end{aligned} \tag{4.29}$$

The decomposition (4.29) is obtained through UHT of LRD time series and denoising procedure.

Remarks

1. We have also performed the above analysis with larger thresholds ϵ, smaller sample sizes N and smaller LRD parameters s. The overall results are similar to the ones reported above

with the following exceptions. First, when increasing the threshold, $\mathrm{Exp}(\mu, \lambda)$ distribution fits inter-arrivals well with larger truncation parameter μ. For example, the thresholds $1.5\epsilon^{(u)}$ and $2\epsilon^{(u)}$ lead to the respective distributions $\mathrm{Exp}(7, 18.59)$ and $\mathrm{Exp}(16, 34.63)$ for inter-arrivals (keeping $N = 75,000$). Larger estimated λ is consistent with the fact that higher threshold leads to longer inter-arrival times. Taking larger threshold is also likely to lead to fewer small inter-arrivals which explains a satisfactory fit of $\mathrm{Exp}(\mu, \lambda)$ distribution only with larger μ. Second, we have found that analysis reported above is consistent in FARIMA$(0, s, 0)$ time series for LRD parameter from about $s = .2$. For $s = .1$, for example, the analysis finds that inter-arrivals have heavier tails and that jumps from zero are not as clearly LRD. Third, similar results as above were found with a smaller sample size $N = 5,000$.

2. As discussed in Section 4.2.2, CART can be viewed as UHT with different criteria in selecting break points. Slightly different but parallel to denoising with UHT, cost-complexity pruning (Hastie et al. (2001), p. 270) can be applied to find an optimal subtree of break points. CART with pruning procedure finds the denoised time series that minimizes square error loss and complexity of tree. Since pruning procedure is computationally more demanding due to cross-validation, we were able to study only a smaller sample size such as $N = 5,000$. With this particular N, we have found that CART gives quite similar results to the ones reported above when taking the threshold $1.7\epsilon^{(u)}$.

From a reverse angle, we show next that the right-hand side of (4.29) defines a LRD time series under fairly mild assumptions. Suppose that $Y = \{Y_n\}_{n\geq 0}$ is a zero mean stationary LRD time series with LRD parameter β in (4.2). Let $\{U_k\}_{k\geq 1}$ be a sequence of i.i.d. inter-arrivals taking (positive) integer values. To make a sequence of arrivals stationary, one needs to introduce a special first inter-arrival U_0 as

$$P(U_0 = k) = \frac{1}{\mu} P(U_1 \geq k+1), \quad k = 0, 1, 2, \dots, \tag{4.30}$$

where $\mu = EU_1$ (Resnick (1992), p. 225). This ensures that the point process consisting of the sequence of arrivals $S_n = \sum_{k=0}^{n} U_k$, $n = 0, 1, \dots$, is stationary. Let $N_n = \sup\{k : S_k \leq n\}$

be the number of arrivals up to time n. With $\widetilde{S}_n = \sum_{k=1}^n U_k$, let also $\widetilde{N}_n = \sup\{k : \widetilde{S}_k \le n\}$. Finally, let $X = \{X_n\}_{n\in\mathbb{Z}}$ be a stationary SRD time series independent of Y and consider the series

$$W_n = \sum_{k=0}^{\infty} Y_k 1_{\{S_k \le n < S_{k+1}\}} + X_n, \quad n \ge 0. \tag{4.31}$$

Proposition 4.3.1. *With the above notation, suppose in addition that*

$$\sup_{n\ge 1} n^p P(U_0 = n) < \infty, \text{ for some } p > 1, \tag{4.32}$$

$$\frac{\widetilde{N}_n}{n} \to \frac{1}{\mu} \quad a.s., \tag{4.33}$$

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{\widetilde{N}_n}{n} \le a\right) = -c < 0, \tag{4.34}$$

where $\mu > 0$, $a < 1/\mu$ and $c > 0$. Then, the series $\{W_n\}_{n\ge 0}$ in (4.31) is a stationary LRD time series, with the same LRD parameter β.

In Proposition 4.3.1, the quantities Y, X and S are associated with jumps from zero, residuals and arrivals in the simulations reported above. The model (4.31) can also be thought as a renewal-reward process with additive noise, where rewards Y_k are dependent, rewards are S_k and noise is X_k. We prefer to use our terminology of jumps from zero, inter-arrivals and residuals because the model (4.31) is suggested for changes in local mean level.

Remarks

1. Assumption (4.32) ensures, in particular, that the covariance function of $\{W_n\}_{n\ge 0}$ is not dominated by heavy-tailed inter-arrivals. For example, if inter-arrivals were heavy-tailed with parameter $\beta_0 < \beta$ (that is, $P(U_0 \ge h) = L_0(h)h^{-\beta_0}$), the proof of Proposition 4.3.1 shows (see relation (4.58)) that the series $\{W_n\}_{n\ge 0}$ would be LRD with the parameter β_0. Assumption (4.33) is a standard law of large numbers for renewals (see Resnick (1992) for independent inter-arrivals). Assumption (4.34) ia a large deviation principle for renewals. In the case of independent inter-arrivals, it is studied, for example, in Glynn and Whitt (1994) and Puhalskii and Whitt (1997).

2. Proposition 4.3.1 and (4.29) could be of independent interest for at least the following

reason. In simulation studies, LRD time series are not practical to be generated "on-the-fly". They are generated for a desired sample size and then being stored for later use. When large scale simulation studies need to be performed (large sample size and many replications), it may be difficult or impossible to store all the desired LRD data. In this case, LRD model (4.31) in Proposition 4.3.1 allows to reduce the storage requirements. With the model (4.31), only the LRD time series Y of jumps from zero needs to be stored and its average sample size is inversely proportional to the local mean length of inter-arrivals. Moreover, the discussion preceding Proposition 4.3.1 shows that such LRD model is representative of a larger class of LRD time series.

3. In simulations with UHT and LRD reported above, the sequence of jumps from zero $\{Y_k\}$ and inter-arrivals $\{U_k\}$ appear independent. But the sequence of residuals $\{X_n\}$ and the denoised series $\sum_k Y_k 1_{[S_k, S_{k+1})}(n)$ exhibit non zero cross-correlations across several lags.

The most surprising feature of (4.29) is that denoising with UHT decorrelates inter-arrivals. To better understand this, we further analyzed inter-arrivals with the following conclusions. We have found that inter-arrivals at fixed scales of UHT are strongly correlated. It is precisely the denoising procedure that makes them decorrelated. Denoising can be thought as of selecting only a position of break points. The selection is not random and closely depends on LRD process itself. For example, we have verified that randomly selecting break points at some scale does not actually lead to decorrelated inter-arrivals.

4.3.2 Comparison with other ways to define runs

Denoising with UHT is obviously not the only way to define changes in local mean level. We find it necessary to compare the results reported above for UHT with those obtained using other methods. We shall focus below only on durations of changes or runs. At least three other ways to define runs are the following:

- [SIGN] Define runs as the lengths of same sign in original time series.
- [SMOOTH] Locally smooth (e.g. kernel smooth) original time series, and define runs as the lengths of same sign in the smoothed time series.

- [OHT] Define changes in local mean level using denoising with orthogonal Haar transformation (OHT).

Note that SIGN and SMOOTH methods are local in nature, while OHT (as well as UHT) method is global.

A SMOOTH method seems to have been looked at in the context of LRD by Mandelbrot and coauthors, though without great success (e.g. Mandelbrot and Wallis (1969)). The SIGN method for LRD is related to analysis of recurrence times found in the Physics literature. See, for example, Altmann and Kantz (2005) and references therein. Finally, if the results with OHT were similar to those with UHT, one could question any need for the use of a more sophisticated UHT. The findings on simulated FARIMA time series for the three alternative methods above are summarized next.

SIGN: We have found that the distribution of runs in this case is well modeled by a Generalized Pareto Distribution (GPD) with the cumulative distribution function

$$F(x) = 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)_+^{-1/\xi}, \quad \xi,\ \sigma > 0,\ \mu \in \mathbb{R}$$

($y_+ = \max(y, 0)$). In particular, it has heavier tail than exponential distribution. A more distinct feature emerges after examining sample autocorrelation function, one such function depicted in Figure 4.6. Note that the functions decreases extremely slowly and in alternating fashion. The shape of the function is not very common.

SMOOTH: In this case, we applied locally weighted scatter smoothing (LOESS) to the time series under study. We were not able to fit well distribution of runs through a well known parametric distribution. Sample autocorrelation function exhibited a shape similar to that in Figure 4.6 observed in the case of SIGN method. This occurred across a wide range of the bandwidths used.

OHT: To compare OHT and UHT, we generated FARIMA(0,.45,0) time series of length $N = 2^{16} = 65,536$ and applied to it denoising based on OHT with the universal thresholding. The following are key differences in the results with OHT and UHT, as summarized in Figure 4.7. First, the inter-arrivals for OHT do not follow the exponential distribution. Because OHT

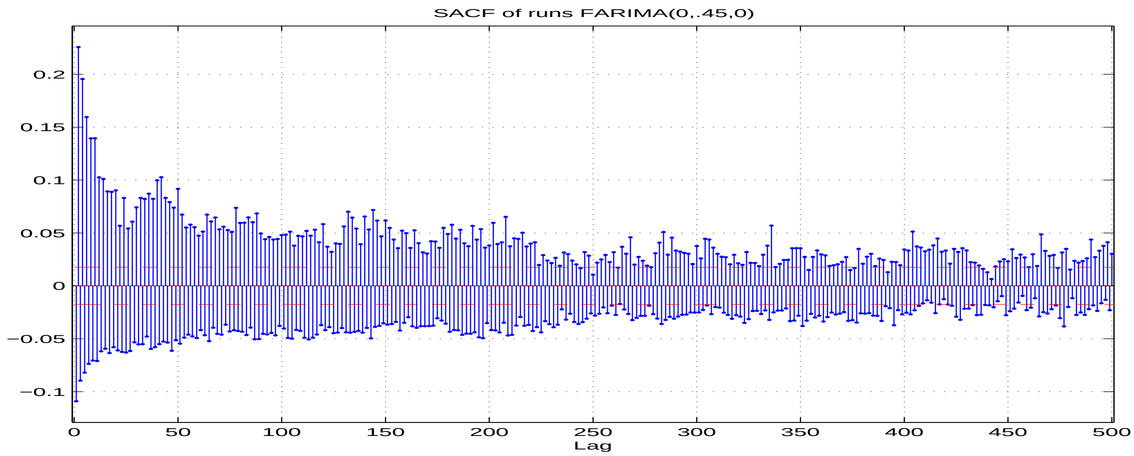

Figure 4.6: Sample ACF of [SIGN] runs for FARIMA(0,.45,0) time series.

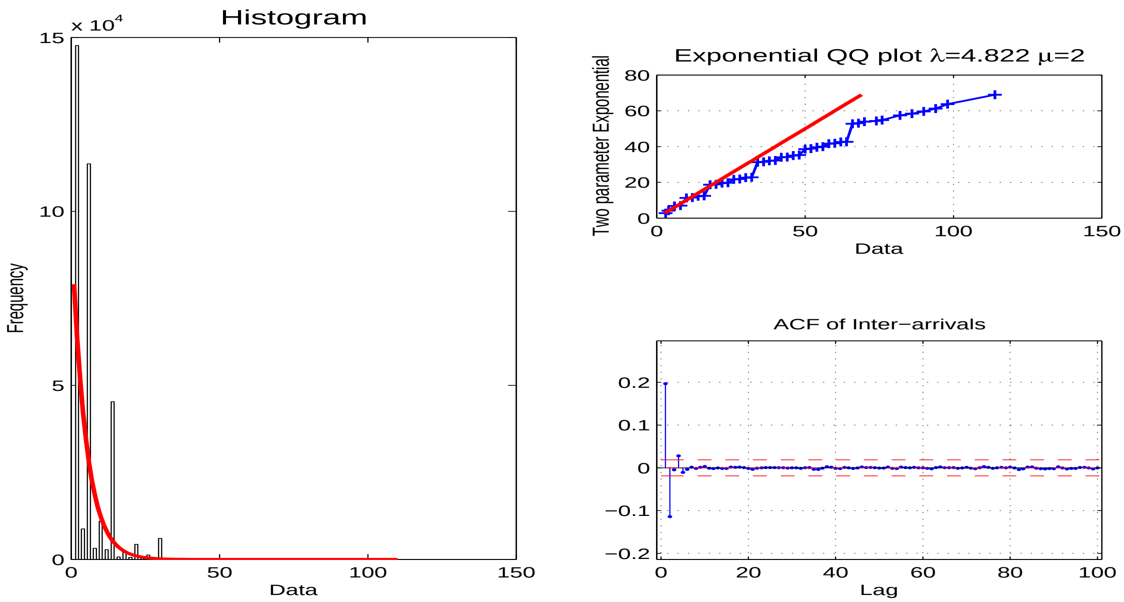

Figure 4.7: Inter-arrivals for OHT.

splits intervals in half, inter-arrivals are multiples of 2^j and their distribution does not have a clearly distinctive shape. Second, ACF of inter-arrivals is slightly larger at lags 1 and 2, showing that inter-arrivals with OHT are not as decorrelated. Third, we also found that the residuals for OHT have consistently larger variances than those for UHT.

Summarizing the findings above, UHT seems to be special in the sense that it leads to the components (runs, jumps from zero, residuals) having simplest structure.

One other issue that we largely pushed aside in Section 4.3.1 is the correlation structure of the residuals (that behave like the first difference of LRD series). It is interesting to ask here whether another denoising method could possibly lead to simpler structure of the residuals. (For example, one might seek the transformation where local mean levels are allowed to be linear rather than constant.) A desirable possibility might seem to have these residuals as white noise, uncorrelated with changes in mean level. We should note in this regard that this is not feasible with most of LRD models. To understand this point, suppose $X = \{X_n\}$ is LRD series such as FARIMA$(0, s, 0)$ series with $s \in (0, 1/2)$. Having

$$X_n = Y_n + \epsilon_n, \tag{4.35}$$

where $Y = \{Y_n\}$ is stationary and thought as representing changes in mean level, and $\epsilon = \{\epsilon_n\}$ is a white noise, uncorrelated with Y, is equivalent to the corresponding spectral density satisfying

$$f_X(\omega) = f_Y(\omega) + f_\epsilon(\omega) \tag{4.36}$$

or

$$f_Y(\omega) = f_X(\omega) - f_\epsilon(\omega) = f_X(\omega) - \frac{\sigma^2}{2\pi}. \tag{4.37}$$

The equation (4.37) defines the unique solution (solving for Y) to (4.35) as long as $f_Y(\omega) \geq 0$. The solution Y exists for most LRD models, e.g. FARIMA$(0, s, 0)$ series, with small enough chosen $\sigma > 0$. However, we cannot expect the corresponding Y to represent changes in mean level, that is, stay constant over periods of time. We will not try to formalize this statement here. We have looked at several LRD models and simulated the corresponding series Y. We found in all cases that the series $\{Y_n\}$ looks irregular. For the series to stay constant over periods of time, its spectral density and covariance function need to have a special form (which can be deduced form the proof of Proposition 4.3.1) and the operation (4.37) is not the one that ensures this.

4.3.3 Comparison with real and other simulated data

In this section, we apply the UHT analysis discussed in Section 4.3.1 to three real time series and few simulated data sets, all exhibiting LRD. The data sets are:

- [UNC-1], [UNC-2] These data sets represent Internet traffic of packets at a UNC link collected on April 9 and April 11, 2002. Both Internet data were aggregated up to 100 milliseconds and normalized by 1,000 to have smaller units. Moreover, we analyzed the series obtained by disregarding the outlying observations below 2.2 and above 3.4 for UNC-1 series, and above 2.1 for UNC-2 series. For example, the number of removed outliers was 193 for UNC-1 series, and the resulting series was of length $N = 73{,}405$. The wavelet-based LRD parameter estimate for UNC-1 series is $\hat{s} = .473$ with 95% confidence interval (.438, .5008).

- [Campito] This is the celebrated mount Campito yearly tree ring data representing the period from 3426 BC to 1946 AD.

- [D-I] This is the data simulated according to a Markov switching model proposed by Diebold and Inoue (2001) with transition probabilities $p_{00} = p_{11} = .999$ between means $\mu_0 = 0$, $\mu_1 = 1$, sample size $N = 5{,}000$, and the Gaussian noise variance $\sigma^2 = 1$. Averaging over 50 realizations, the wavelet-based LRD parameter estimate is $\hat{s} = .269$ with 95% confidence interval (.139, .398). This confirms, as argued by Diebold and Inoue, that for carefully chosen parameters (including sample size), Markov-switching model can be easily confused with LRD.

- [SP500] This is the data of squared log returns of the daily closing index of S&P 500 from Jan 3, 1918 to Dec 31, 2009 (21,675 observations). Squared log returns are defined for index P_t by

$$r_t = (100(\log P_t - \log P_{t-1}))^2.$$

  The wavelet-based LRD parameter estimate for the squared log returns of S&P500 series is $\hat{s} = .325$ with 95% confidence interval (.304, .346).

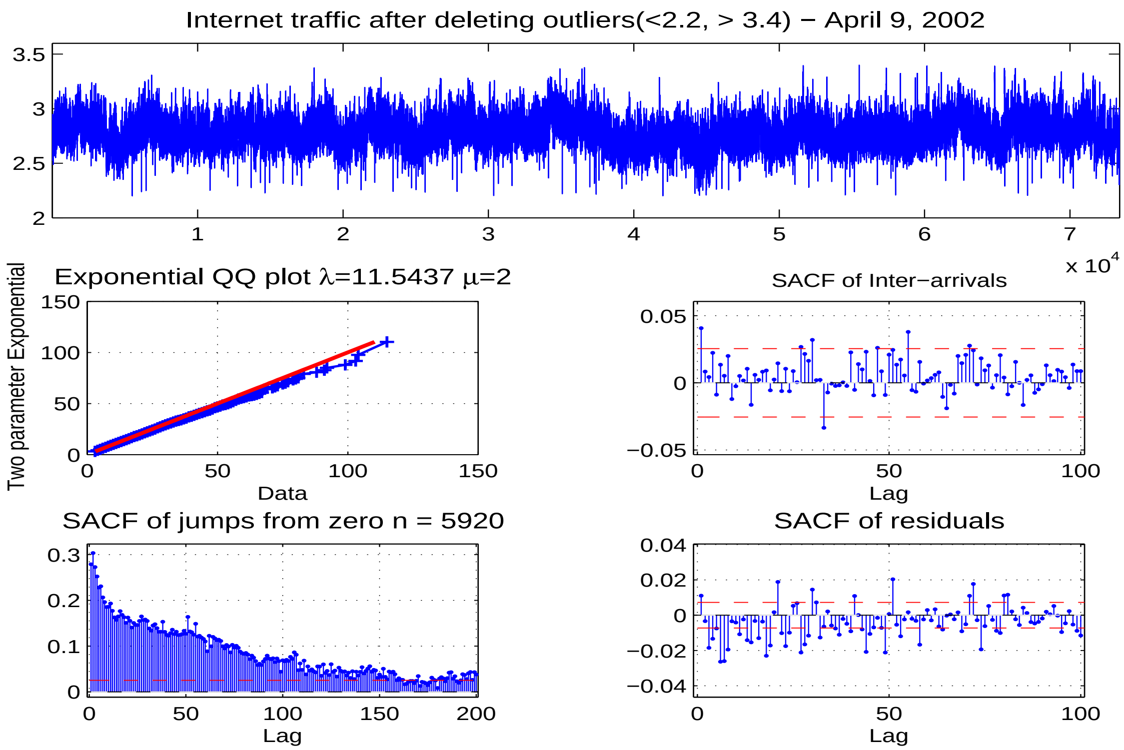

Figure 4.8: Internet traffic at UNC, April 9, 2002.

Five data sets are considered because findings are quite different across the sets. They are summarized next.

UNC-1: The UHT analysis of UNC-1 series is summarized in Figure 4.8. The results are surprisingly similar to what has been found with simulated FARIMA time series, with the exceptions that larger value of $\lambda = 11.54$ is found for inter-arrivals, the number of breaks (5920) is smaller, and the estimated $\hat{s}$ of the residuals is -.8245.

UNC-2: As seen from Figure 4.9, the key difference is that inter-arrivals now stay correlated over long lags. In this regard, it is interesting to note here the following. If inter-arrivals were also LRD, a closer examination of the proof of Proposition 3.1 shows that, under suitable modified assumptions, this would not affect the main conclusion of the result. This might explain why LRD emerges at jumps from zero only, and also suggests that available LRD models may not be able to capture dependence of inter-arrivals when there is such in real data.

Campito: The results for this data set are summarized in Figure 4.10. The difference here is that the sample ACF of jumps from zero does not exhibit LRD as clearly as in the

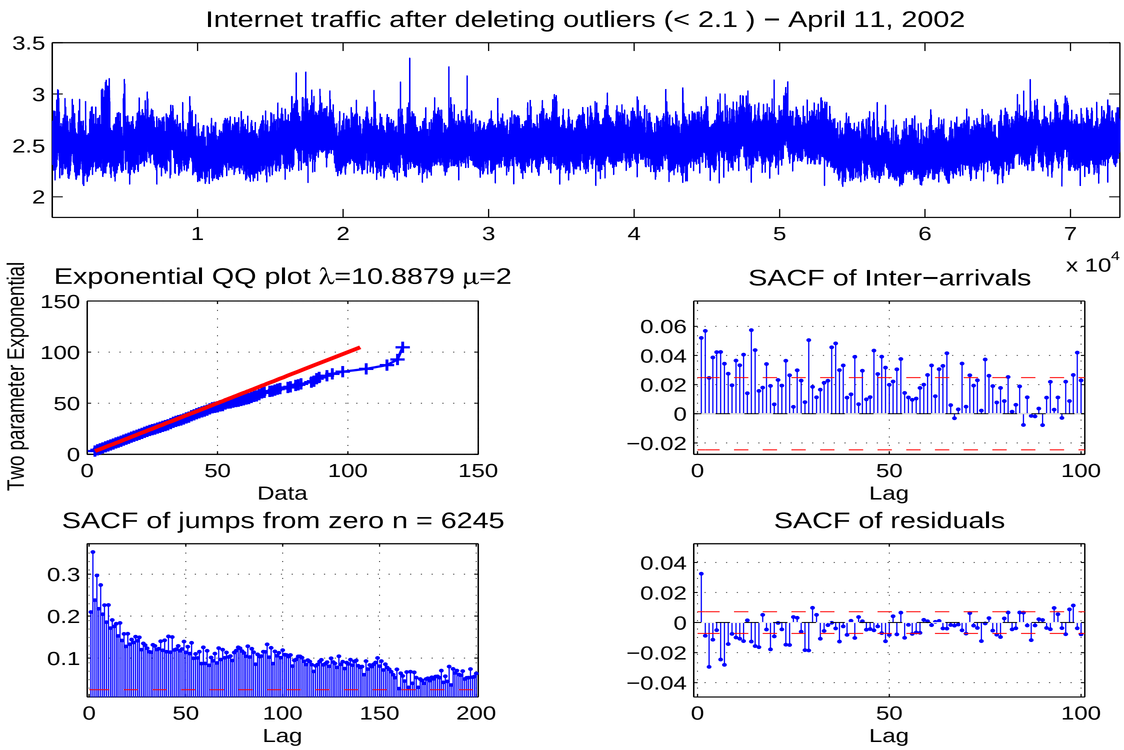

Figure 4.9: Internet traffic at UNC April 11, 2002.

case of simulated data above. For example, fitting AR(1) model $X_t = \varphi X_{t-1} + \epsilon_t$ to the series (using simple regression) yields the regression parameter estimate $\hat{\varphi} = .92$. In this regard, it is interesting to suggest the following non-LRD-like model exhibiting features of LRD.

As in (4.31), consider the model

$$W_n^T = \sum_{k=0}^{\infty} Y_k^T 1_{\{S_k \le n < S_{k+1}\}} + X_n, \quad n = 1, \dots, T, \tag{4.38}$$

where Y_k^T is a Gaussian AR(1) model

$$Y_k^T = \varphi_T Y_{k-1}^T + \epsilon_k, \quad \epsilon_k \sim \text{ i.i.d. } \mathcal{N}(0, \sigma^2), \tag{4.39}$$

with

$$\varphi_T = 1 - \frac{c}{T^\gamma}, \quad 0 < \gamma < 1/2, \; c > 0. \tag{4.40}$$

(As $T \to \infty$, $\varphi_T \to 1-$.) For large T, this model exhibits LRD in the following sense. The argument below is rigorous only in part. We will not pursue complete rigor for shortness sake.

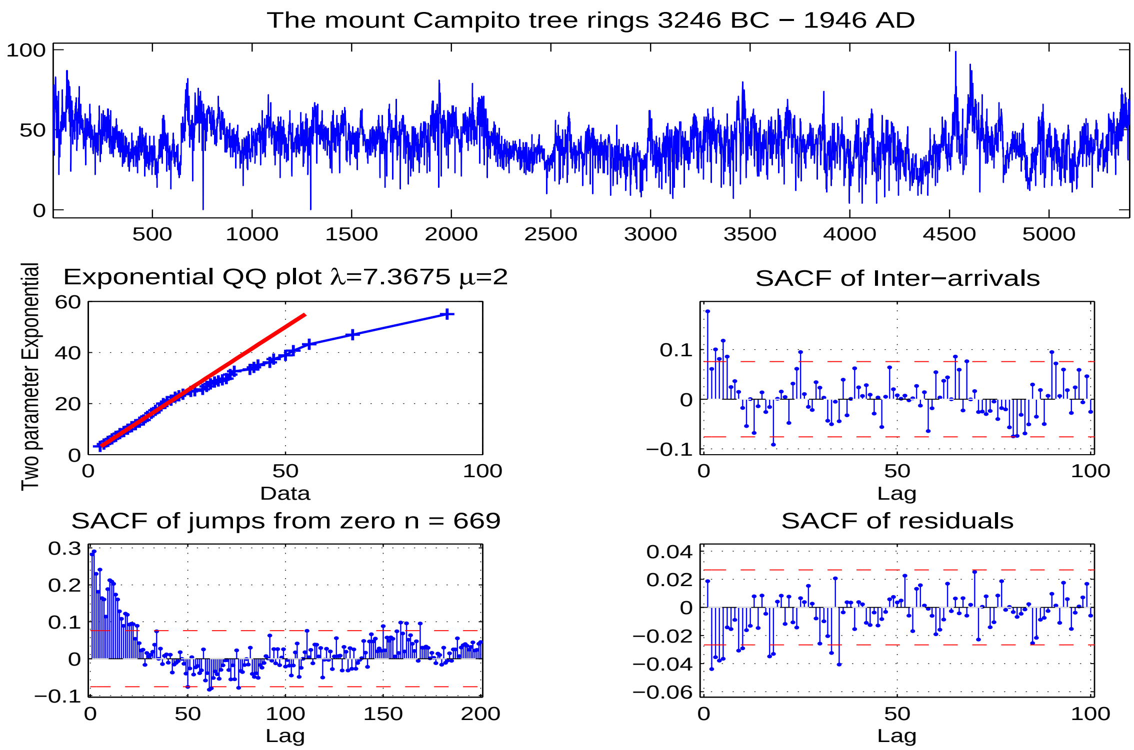

Figure 4.10: The mount Campito tree rings.

Let $S_T = \sum_{t=1}^{T} W_t^T$. We will argue that $\mathrm{Var}(S_T) = O(T^{2\gamma+1})$ and thus that W^T has LRD properties when $\gamma \in (0, 1/2)$. The proof of Proposition 4.3.1 in Section 4.6.1 (see (4.58)) suggests that, as $T \to \infty$, for $h = 1, \ldots, T$,

$$\gamma_{W^T}(h) \sim E\gamma_{Y^T}(\widetilde{N}_h) \sim \gamma_{Y^T}\left(\frac{h}{\mu}\right) \tag{4.41}$$

and therefore

$$\begin{aligned}\mathrm{Var}(S_{\mu T}) &= \sum_{h=-(\mu T-1)}^{\mu T-1} (\mu T - |h|)\gamma_{W^T}(h) \sim \sum_{h=-(\mu T-1)}^{\mu T-1} (\mu^T - |h|)\gamma_{Y^T}\left(\frac{h}{\mu}\right) \\ &= \mu \sum_{h=-(\mu T-1)}^{\mu T-1} (\mu T - |h|)\gamma_{Y^T}(h) = \mathrm{Var}(\widetilde{S}_T),\end{aligned} \tag{4.42}$$

where $\widetilde{S}_T = \sum_{t=1}^{T} Y_t^T$. The following result concern the asymptotic behavior of (4.42). It is proved in Section 4.6.2.

Proposition 4.3.2. *Let $\{Y_k^T\}$ be an AR(1) time series defined through (4.39)-(4.40), and*

T	10,000			20,000			30,000		
φ	scale	mean	std	scale	mean	std	scale	mean	std
.75	6-10	.7607	.0604	7-11	.6173	.0622	8-12	.5457	.0682
	7-10	.6370	.1111	8-11	.5470	.1266	9-12	.5108	.1281
	8-10	.5452	.2662	9-11	.5026	.2345	10-12	.5457	.2784
.9	6-10	1.0641	.0654	7-11	.8398	.0680	8-12	.6528	.0856
	7-10	.8898	.1182	8-11	.6879	.1102	9-12	.5690	.1335
	8-10	.7180	.1990	9-11	.5735	.2113	10-12	.5280	.3304

Table 4.1: Wavelet-based LRD parameter estimation on 50 realizations of the model (4.38).

$\widetilde{S}_T = \sum_{k=1}^T Y_k^T$. *Then, as* $T \to \infty$,

$$\mathrm{Var}(\widetilde{S}_T) \sim \frac{\sigma^2}{c^2} T^{2\gamma+1}. \tag{4.43}$$

The relations (4.43) and (4.42) now suggest that, as indicated above,

$$\mathrm{Var}(S_T) \sim \frac{\sigma^2}{c^2 \mu^{2\gamma+1}} T^{2\gamma+1}. \tag{4.44}$$

The relation (4.44) is illustrated through Table 4.1 where some statistics (mean and standard deviation) of wavelet-based LRD parameter estimation are reported based on 50 realizations of the model (4.38) with several choices of parameter φ, sample size T, and the scale (octave) range in wavelet-based estimation. (The inter-arrivals were taken according to Poisson process with parameter equal to 4, and $\sigma = 1$ in (4.39).)

An analysis similar to the one above is performed with various non-LRD-like models proposed in Diebold and Inoue (2001). As we argue below, however, their models are not realistic in capturing properties of LRD models beyond variance of partial sums.

D-I: UHT analysis for the Markov-switching model (combining 50 realizations) is summarized in Figure 4.11. Key differences from LRD models are the following. When the threshold parameter is small, we find that GPD fits better the distribution of inter-arrivals. However, if threshold parameter is large, Exponential distribution fit is satisfactory. A sharper difference is seen from sample ACF plots of jumps from zero and their squares and absolute values. In contrast to LRD models, these plots do not show slow decay of LRD. We have examined UHT with other choices of parameters in a Markov-switching model. In some of these cases, the sam-

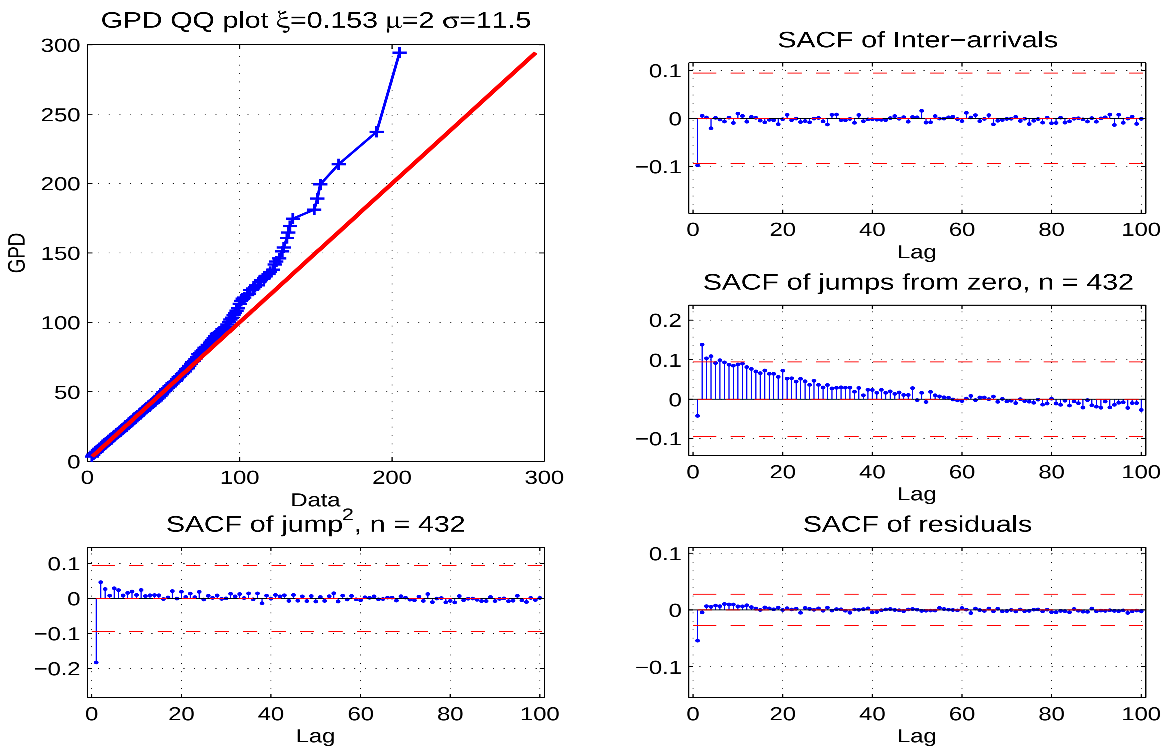

Figure 4.11: D-I Markov-switching model.

ple ACF of jumps from zero show slower decay but those of their squares and absolute values always exhibit decorrelation of Figure 4.11. These observations show that UHT can distinguish between Markov-switching and LRD models.

SP500: Figure 4.12 summarizes UHT analysis for the squared log returns of S&P 500 stock index. Surprisingly, it can be observed that sample ACF plots of jumps from zero does not show slow decay. Also, though not reported here, the same happens with the sample ACF of jumps squared. This shows that LRD models will not capture well some characteristics of this real series.

4.4 Properties of UHT of LRD Time Series

We argued in Section 4.3.2 that UHT leads to simplest structure of denoised series from LRD data. It is interesting to discuss here some basic properties of UHT of LRD time series and FBM. These properties are of independent interest and were not used directly in Section 4.3.

Let $X = \{X_1, \dots, X_N\}$ denote a time series vector. Several basic properties of UHT of X

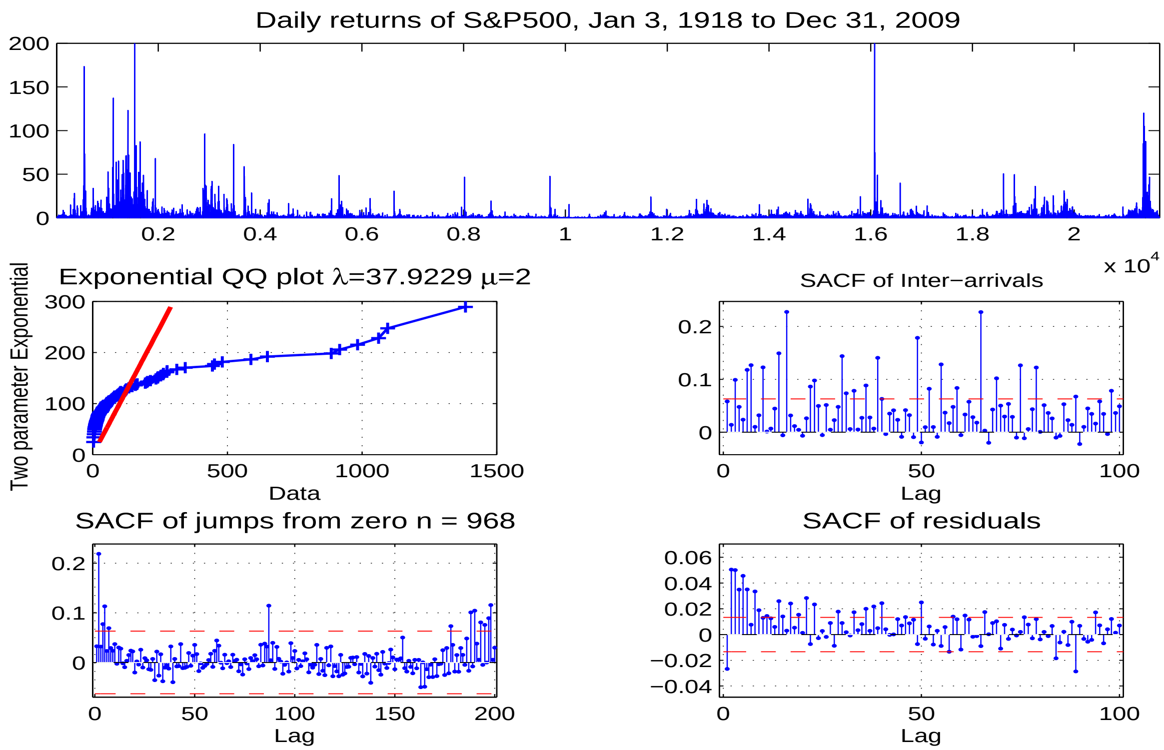

Figure 4.12: The squared log returns of S&P 500 index.

are:

- Symmetry, zero mean: If $X \stackrel{d}{=} -X$, then UHT detail coefficients are symmetric around zero. Moreover, if $E|X_n| < \infty$, these coefficients have zero mean.

- Reversibility: If $Y = \{X_N, \ldots, X_1\}$ denotes a reversed time series X and X is reversible in the sense that $Y \stackrel{d}{=} X$, then UHT detail coefficients and break points are reversible at each scale as well.

These properties follow easily from the definition of UHT and their proofs are omitted for shortness sake. In particular, note that the above properties hold for Gaussian FARIMA time series.

Observe also that reversibility does not imply stationarity of UHT coefficients except at scale $j = 1$. Even stationarity of marginals of coefficients is not implied by reversibility. To understand this, consider the scale $j = 2$. The reversibility implies that $d_{2,1} \stackrel{d}{=} d_{2,4}$, $d_{2,2} \stackrel{d}{=} d_{2,3}$. Stationarity of marginals, however, also requires that $d_{2,1} \stackrel{d}{=} d_{2,2}$. This would follow if reversing time series on an interval $[s_{1,1}, e_{1,1}]$ would not change its distribution. Since $e_{1,1}$ is random, this

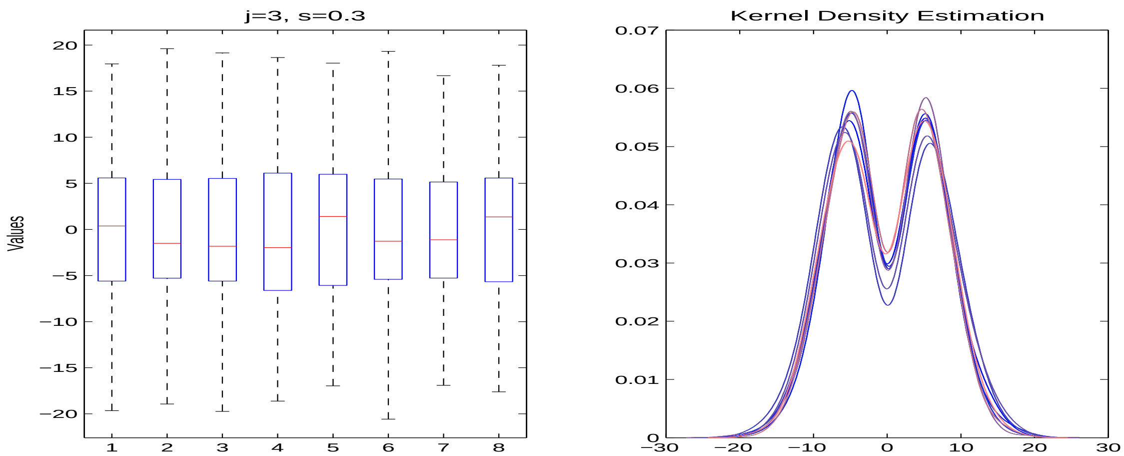

Figure 4.13: Boxplots (left) and kernel density estimation (right) of marginal distribution of UHT detail coefficients $d_{3,k}$, $k = 1, \ldots, 8$ for Gaussian FARIMA$(0, .3, 0)$ series of length 2^{14} based on 1,000 replications.

may not be necessarily true. Even though we do not have a theoretical result, a simulation study shows that UHT detail coefficients of FARIMA time series are close to stationary. For example, Figure 4.13 depicts boxplots and kernel density estimation of the marginal distributions of UHT detail coefficients for Gaussian FARIMA$(0, .3, 0)$ time series at scale $j = 3$. In Figure 4.14, we also examine dependence of UHT detail coefficients for fixed scale j through ACF plot. It is calculated empirically by taking average of 1,000 sample ACF of UHT detail coefficients from FARIMA$(0, .3, 0)$ time series of size 2^{14} when $j = 7$. (ACF plots for other j are similar to those for $j = 7$.) The left plot of Figure 4.14 represents ACF of UHT detail coefficients. For comparison, in the right plot of Figure 4.14, we also display ACF of the usual orthogonal Haar detail coefficients which has a quite different shape.

It is also quite easy to establish the asymptotics in UHT as $N \to \infty$. Suppose that $\{X_n\}_{n\in\mathbb{Z}}$ is a LRD time series with parameter $H \in (1/2, 1)$ such that

$$Y_{H,N}(t) = \frac{1}{N^H \widetilde{L}(N)} \sum_{n=1}^{[Nt]} X_n \xrightarrow{d} B_H(t), \quad t \in [0, 1], \tag{4.45}$$

in the space of functions equipped with the usual Skorohod J_1-topology, where $\widetilde{L}$ is a slowly varying function at infinity and B_H is FBM. Then, one can show that:

- Asymptotics as $N \to \infty$: For fixed J, with $\mathcal{I} = \{(j, k) : j = 0, 1, \ldots, J, k = 1, \ldots, 2^j\}$, as

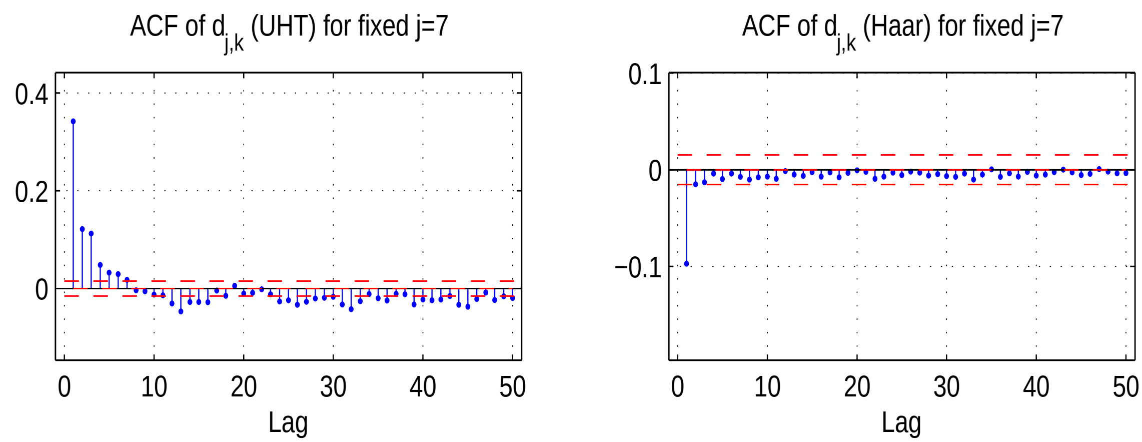

Figure 4.14: ACF plots of UHT detail coefficients (left) and the usual orthogonal Haar detail coefficients (right) $d_{j,k}, j = 7$, for Gaussian FARIMA$(0, .3, 0)$ sequences.

$N \to \infty$,

$$\left(\frac{1}{N}s_{j,k}, \frac{1}{N}b_{j,k}, \frac{1}{N}e_{j,k}, \frac{1}{N^{H-1/2}\widetilde{L}(N)}d_{j,k}\right)_{(j,k)\in\mathcal{I}}$$

$$\xrightarrow{d} \Big(T_{j,2k-2}(B_H), T_{j,2k-1}(B_H) = B_{j,k}(B_H), T_{j,2k}(B_H), D_{j,k}(B_H)\Big)_{(j,k)\in\mathcal{I}}. \qquad (4.46)$$

The proof of (4.46) is based on the fact that $s_{j,k}$, $b_{j,k}$ and $e_{j,k}$ can be expressed (recursively in j) in terms of $Y_{H,N}$ and hence one can pass to the limit by using (4.45). For example,

$$\frac{b_{j+1,2k-1}}{N} = \frac{1}{N}\underset{s_{j+1,2k-1}\le b\le e_{j+1,2k-1}}{\operatorname{argmax}} \left|\langle X, \psi_{s_{j+1,2k-1},b,e_{j+1,2k-1}}\rangle\right| = \frac{1}{N}\underset{s_{j,k}\le b\le b_{j,k}}{\operatorname{argmax}} \left|\langle X, \psi_{s_{j,k},b,b_{j,k}}\rangle\right|$$

$$= \frac{1}{N}\underset{s_{j,k}\le b\le b_{j,k}}{\operatorname{argmax}} \left| \left(\frac{1}{b - s_{j,k}+1} - \frac{1}{b_{j,k} - s_{j,k}+1}\right)^{1/2} \sum_{k=s_{j,k}}^{b} X_k \right.$$

$$\left. - \left(\frac{1}{b_{j,k} - b} - \frac{1}{b_{j,k} - s_{j,k}+1}\right)^{1/2} \sum_{k=b+1}^{b_{j,k}} X_k \right|$$

$$= \underset{\frac{s_{j,k}}{N}\le b\le \frac{b_{j,k}}{N}}{\operatorname{argmax}} \left| \left(\frac{1}{b - \frac{s_{j,k}}{N} + \frac{1}{N}} - \frac{1}{\frac{b_{j,k}}{N} - \frac{s_{j,k}}{N} + \frac{1}{N}}\right)^{1/2} \left(Y_{H,N}(b) - Y_{H,N}\left(\frac{s_{j,k}}{N} - \frac{1}{N}\right)\right) \right.$$

$$\left. - \left(\frac{1}{\frac{b_{j,k}}{N} - b} - \frac{1}{\frac{b_{j,k}}{N} - \frac{s_{j,k}}{N} + \frac{1}{N}}\right)^{1/2} \left(Y_{H,N}\left(\frac{b_{j,k}}{N}\right) - Y_H\left(b + \frac{1}{N}\right)\right) \right|$$

$$
\begin{aligned}
&\xrightarrow{d} \underset{T_{j,2k-2}\le b\le T_{j,2k-1}}{\operatorname{argmax}} \left| \left(\frac{1}{b-T_{j,2k-2}} - \frac{1}{T_{j,2k-1}-T_{j,2k-2}} \right)^{1/2} (B_H(b) - B_H(T_{j,2k-2})) \right. \\
&\qquad \left. - \left(\frac{1}{T_{j,2k-1}-b} - \frac{1}{T_{j,2k-1}-T_{j,2k-2}} \right)^{1/2} (B_H(T_{j,2k-1}) - B_H(b)) \right| \\
&= \underset{T_{j,2k-2}\le b\le T_{j,2k-1}}{\operatorname{argmax}} \left| \langle B_H, \psi_{T_{j,2k-2},b,T_{j,2k-1}} \rangle \right| = T_{j+1,4k-3} = B_{j+1,2k-1}.
\end{aligned}
$$

The same idea is used when dealing with R/S and related statistics for LRD time series. See, for example, Mandelbrot (1975), Giraitis, Kokoszka, Leipus and Teyssière (2003).

The asymptotic result (4.46) suggests that UHT of FBM plays an important role in dealing with LRD time series. Orthogonal wavelet decompositions for FBM have been extensively studied in the past (Flandrin (1992), Abry, Flandrin, Taqqu and Veitch (2003)). One of the key properties of orthogonal wavelet decompositions is that their detail coefficients inherit a scaling property from FBM, namely,

$$
d_{j,0} \stackrel{d}{=} 2^{-j(H+1/2)} d_{0,0}. \tag{4.47}
$$

Note that this yields, in particular,

$$
\log_2 E|d_{j,0}|^2 = -j(2H+1) + \log_2 E|d_{0,0}|^2. \tag{4.48}
$$

It is interesting to discuss here briefly whether similar relations also hold in the case of UHT of FBM.

For scale $j = 1$, the first UHT detail coefficient of FBM is given by

$$
|D_{1,1}(B_H)| = \max_{0\le u\le B_{0,1}} \left| \int_0^{B_{0,1}} \psi_{0,b,B_{0,1}}(s) dB_H(s) \right|
$$

$$
\begin{aligned}
&= \max_{0\le u\le B_{0,1}} \left| \int_0^{B_{0,1}} \left(\left(\frac{1}{b} - \frac{1}{B_{0,1}} \right)^{1/2} 1_{\{0\le s\le b\}} - \left(\frac{1}{B_{0,1}-b} - \frac{1}{B_{0,1}} \right)^{1/2} 1_{\{b\le s\le B_{0,1}\}} \right) dB_H(s) \right| \\
&= (B_{0,1})^{H-1/2} \max_{0\le u^*\le 1} \left| \int_0^1 \psi_{0,u^*,1}(s) dB_H^{(0)}(s) \right| = (B_{0,1})^{H-1/2} \left| D_{0,1}\left(B_H^{(0)}(s) \right) \right|,
\end{aligned} \tag{4.49}
$$

where, for $j = 0, 1, \ldots,$

$$
B_H^{(j)}(s) = \frac{B_H(B_{j,1}s)}{(B_{j,1})^H}, \quad s \in [0,1]. \tag{4.50}
$$

Similar argument for the break at scale $j = 1$ and $k = 1$ leads to

$$B_{1,1}(B_H) = B_{0,1} \operatorname*{argmax}_{0 \le u \le 1} \left| \int_0^1 \psi_{0,u,1}(s) dB_H^{(0)}(s) \right| =: B_{0,1} B_{0,1} \left(B_H^{(0)}(s) \right). \tag{4.51}$$

Note that (4.49) and (4.51) are written using the process $B_H^{(0)}(s), s \in [0,1]$. Using H-self-similarity of FBM, it is tempting to conclude that

$$B_H^{(0)} | B_{0,1} \stackrel{d}{=} B_H. \tag{4.52}$$

Since $B_{0,1}$ is random and depends on FBM B_H, however, this cannot be expected to be true. In fact, assuming this is true leads to conclusions not consistent with simulations. (The relation (4.52) is obviously incorrect for general $B_{0,1}$: just take $B_{0,1} = \inf\{t > 0 : B_H(t) = 1\}$ and compare the two sides of (4.52) at $s = 1$.)

Proceeding similarly as for (4.49) and (4.51) gives the following general formulae at scale j and $k = 1$,

$$|D_{j,1}| = (B_{j-1,1})^{H-1/2} \left| D_{0,1} \left(B_H^{(j-1)} \right) \right|, \tag{4.53}$$

$$B_{j,1} = B_{0,1} \prod_{m=0}^{j-1} B_{0,1} \left(B_H^{(m)} \right). \tag{4.54}$$

Substituting (4.54) into (4.53) and taking the logarithm and then expectation gives the following relationship.

- Multiresolution structure of UHT for FBM:

$$E \log_2 |D_{j,1}| = \left(H - \frac{1}{2} \right) E \log_2 B_{j-1,1} + E \log_2 \left| D_{0,1} \left(B_H^{(j-1)} \right) \right| \tag{4.55}$$

$$= \left(H - \frac{1}{2} \right) \left\{ E \log_2 B_{0,1} + \sum_{m=0}^{j-2} E \log_2 B_{0,1} \left(B_H^{(m)} \right) \right\} + E \log_2 \left| D_{0,1} \left(B_H^{(j-1)} \right) \right|. \tag{4.56}$$

Observe that relation (4.56) is more complex than (4.48), and there is no reason to suppose that it is linear in j. In fact, as reported in Figure 4.15, simulation study suggests the relation (4.56) is not linear in j but rather quadratic. To produce the figure, we generated FBM on an

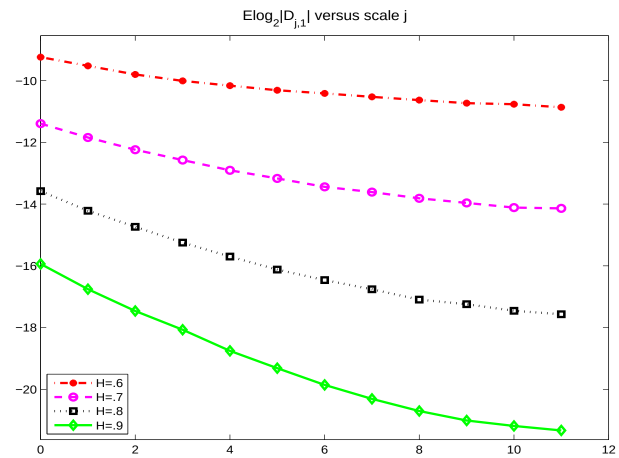

Figure 4.15: $E \log_2 |D_{j,1}|$ versus scale j.

interval $[0, 1]$ at 100,000 equally spaced points for self-similarity parameters $H = 0.6, 0.7, 0.8, 0.9$, and then applied continuous UHT. Expectations involving UHT detail coefficients and breaks are empirically calculated based on 1,000 replications.

Turn now to Figure 4.16. Its left plot shows the relationship between $E \log_2 B_{j,1}$, which is the first term in (4.55), and scale j. Observe that it has a pattern similar to that in Figure 4.15. The right plot of Figure 4.16 shows

$$E \log_2 B_{0,1} \left(B_H^{(j)} \right) = E \log_2 \frac{B_{j,1}}{B_{j-1,1}} = E \log_2 B_{j,1} - E \log_2 B_{j-1,1} \tag{4.57}$$

as a function of j. Since the plot appears linear, it confirms that the sum $E \log_2 B_{j-1}$ of increments (4.57) appears quadratic. Note also that Figure 4.16 essentially states that the ratio $B_{j,1}/B_{j-1,1}$, decreases on average with increasing j.

Figure 4.17 plots

$$E \log_2 \left| D_{0,1} \left(B_H^{(j-1)} \right) \right|,$$

which is the last term in (4.56), as a function of scale. Observe that this term decreases linearly with increasing j. Together with Figure 4.16, this shows that the nonlinear decrease in Figure 4.15 is due to the first term in (4.55).

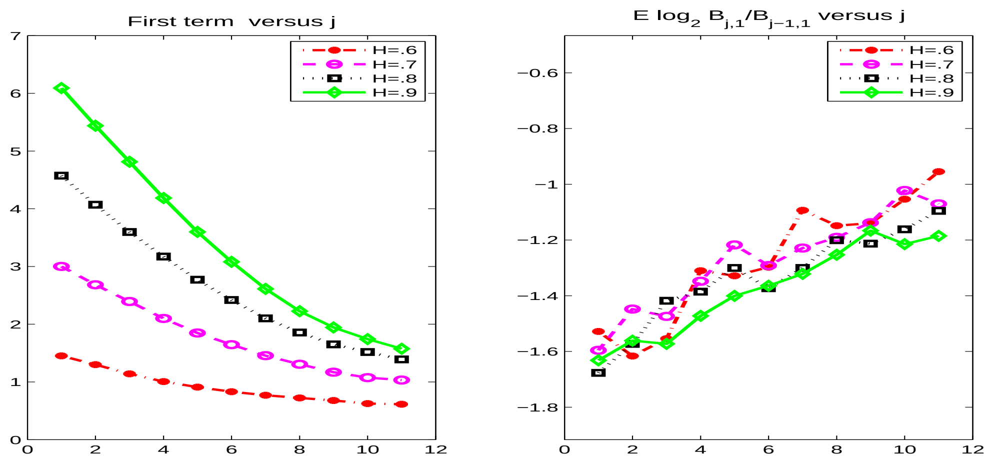

Figure 4.16: $(H - 1/2)E\log_2 B_{j-1,1}$ versus scale j (left plot), $E\log_2 B_{0,1}\left(B_H^{(j)}\right) = E\log_2 B_{j,1}/B_{j-1,1}$ versus j (right plot).

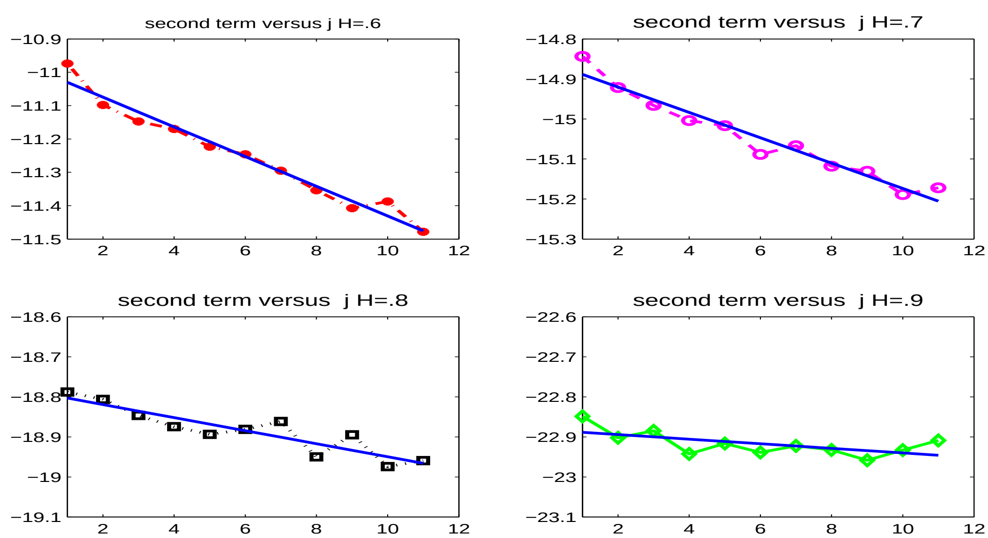

Figure 4.17: $E\log_2\left|D_{0,1}\left(B_H^{(j-1)}\right)\right|$ versus scale j.

4.5 Conclusions

In this chapter, we have used UHT to analyze changes in local mean level for LRD models and several real time series. We also argue that UHT is special among available ways to define changes in local mean level in the sense that it leads to them having simplest stochastic properties. The results of the paper shed light on LRD structure and, for example, allow to distinguish between LRD models and some non-LRD-like alternatives proposed earlier.

4.6 Proofs

4.6.1 The proof of Proposition 4.3.1

It is sufficient to establish the result assuming that $X \equiv 0$. We prove stationarity only between two fixed time points. The general case can be dealt with in a similar way. Stationarity of S ensures that $N = \{N_n\}_{n\geq 0}$ has stationary increments. Observe next that, for Borel sets A_1, A_2, and k_1, k_2, $h \geq 0$, $k_1 \leq k_2$

$$P(W_{k_1+h} \in A_1, W_{k_2+h} \in A_2) =$$

$$\sum_{n_1,n_2=0}^{\infty} P(W_{k_1+h} \in A_1, W_{k_2+h} \in A_2 | N_{k_1+h} = n_1, N_{k_2+h} = n_1 + n_2)$$

$$\times P(N_{k_1+h} = n_1, N_{k_2+h} = n_1 + n_2)$$

$$= \sum_{n_1,n_2=0}^{\infty} P(Y_{n_1} \in A_1, Y_{n_1+n_2} \in A_2) P(N_{k_1+h} = n_1, N_{k_2+h} - N_{k_1+h} = n_2)$$

$$= \sum_{n_1,n_2=0}^{\infty} P(Y_0 \in A_1, Y_{n_2} \in A_2) P(N_{k_1+h} = n_1) P(N_{k_2} - N_{k_1} = n_2)$$

$$= \sum_{n_2=0}^{\infty} P(Y_0 \in A_1, Y_{n_2} \in A_2) P(N_{k_2} - N_{k_1} = n_2) = P(W_{k_1} \in A_1, W_{k_2} \in A_2),$$

which proves stationarity.

We now prove that W_n is LRD with the parameter β. Observe that

$$\gamma_W(h) = \mathrm{Cov}(W_0, W_h) = EW_0 W_h$$

$$= EW_0 W_h 1_{\{S_0=0\}} + EW_0 W_h 1_{\{S_0 \geq h\}} + EW_0 W_h 1_{\{0<S_0<h\}}$$

$$= EY_1 Y_{\widetilde{N}_h} P(U_0 = 0) + EY_0^2 P(U_0 \geq h) + \sum_{k=1}^{h-1} EY_0 Y_{\widetilde{N}_{h-k}+1} P(U_0 = k)$$

$$= E\gamma_Y(\widetilde{N}_h) P(U_0 = 0) + \gamma_Y(0) P(U_0 \geq h) + \sum_{k=1}^{h-1} E\gamma_Y(\widetilde{N}_{h-k} + 1) P(U_0 = k), \tag{4.58}$$

where $\gamma_Y(\cdot)$ is a covariance function of Y. We will establish the result by showing that

$$\frac{\gamma_W(h)}{L(h)h^{-\beta}} \to const, \quad \text{as } h \to \infty. \tag{4.59}$$

First, we will show that under the conditions (4.33) and (4.34),

$$\frac{E\gamma_Y(\widetilde{N}_h)}{L(h)h^{-\beta}} = \frac{EL(\widetilde{N}_h)\widetilde{N}_h^{-\beta}}{L(h)h^{-\beta}} \to \mu^\beta, \quad \text{as } h \to \infty, \tag{4.60}$$

where $\gamma_Y(k) = L(k)k^{-\beta}$ with a slowly varying function L. We will argue (4.60) by a generalized dominated convergence theorem (Folland (1999), p. 59). Note that

$$f_h = \frac{L(\widetilde{N}_h)\widetilde{N}_h^{-\beta}}{L(h)h^{-\beta}} \to \mu^\beta \quad \text{ a.s.,}$$

since $L(\widetilde{N}_h)/L(h) \to 1$ a.s. as $h \to \infty$ (use (4.33)). By the Potter's bounds (Bingham et al. (1989), p. 25), for any $\epsilon > 0$, there is h_0 such that

$$\left| \frac{L(\widetilde{N}_h)}{L(h)} \right| \le 2 \left(\frac{\widetilde{N}_h}{h} \right)^{\pm\epsilon}, \quad \text{for } \widetilde{N}_h, h > h_0.$$

Hence, there is a function g_h such that

$$|f_h| \le g_h = C \left(\frac{h}{\widetilde{N}_h} \right)^{\beta\pm\epsilon} \to C\mu^{\beta\mp\epsilon} \quad \text{ a.s.,}$$

and a generalized dominated convergence theorem implies (4.60) if

$$\lim_{h\to\infty} E \left(\frac{h}{\widetilde{N}_h} \right)^{\beta\pm\epsilon} = \mu^{\beta\mp\epsilon}. \tag{4.61}$$

Without loss of generality, suppose that $\epsilon = 0$. Then,

$$E\left(\frac{h}{\widetilde{N}_h} \right)^\beta = E\left(\frac{h}{\widetilde{N}_h} \right)^\beta 1_{\{\widetilde{N}_h \le ha\}} + E\left(\frac{h}{\widetilde{N}_h} \right)^\beta 1_{\{\widetilde{N}_h > ha\}}. \tag{4.62}$$

The first term in (4.62) vanishes because for some sufficiently large constant $C > 0$,

$$E\left(\frac{h}{\widetilde{N}_h}\right)^{\beta} 1_{\{\widetilde{N}_h \le ha\}} \le h^{\beta} P\left(\frac{\widetilde{N}_h}{h} \le a\right) \le C h^{\beta} \exp(-ch) \to 0,$$

as $h \to \infty$ by (4.34). For the second term in (4.62), let $X_h = (h/\widetilde{N}_h)^{\beta} 1_{\{\widetilde{N}_h > ha\}}$. By (4.33), $X_h \to \mu^{\beta}$ a.s. and $0 \le X_h < a^{-\beta}$. The dominated convergence theorem implies that $\lim_{h\to\infty} EX_h = \mu^{\beta}$.

Note that (4.60) implies

$$\frac{E\gamma_Y(\widetilde{N}_h)P(U_0 = 0)}{L(h)h^{-\beta}} \to const, \tag{4.63}$$

and assumption (4.32) leads to

$$\frac{\gamma_Y(0)P(U_0 \ge h)}{L(h)h^{-\beta}} \to 0. \tag{4.64}$$

Finally, we want to argue that (replacing $\widetilde{N}_{h-k} + 1$ by $\widetilde{N}_{h-k}$ for notational simplicity in (4.58))

$$\frac{\sum_{k=1}^{h-1} E\gamma_Y(\widetilde{N}_{h-k})P(U_0 = k)}{L(h)h^{-\beta}} \to const. \tag{4.65}$$

Again, we will apply a generalized dominated convergence theorem. Let

$$f_h(k) = \begin{cases} \dfrac{E\gamma_Y(\widetilde{N}_{h-k})}{L(h-k)(h-k)^{-\beta}} \dfrac{L(h-k)(h-k)^{-\beta}}{L(h)h^{-\beta}} P(U_0 = k), & k = 1, \ldots, h-1, \\ \\ 0, & \text{otherwise.} \end{cases}$$

From the Potter's bounds, we have

$$|f_h(k)| \le g_h(k) = \begin{cases} CE\left(\dfrac{\widetilde{N}_{h-k}}{h-k}\right)^{-\beta\pm\epsilon} \left(\dfrac{h-k}{h}\right)^{\beta\pm\epsilon} P(U_0 = k), & k = 1, \ldots, h-1, \\ \\ 0, & \text{otherwise.} \end{cases}$$

Note also that $f_h(k) \to \mu^{\beta} P(U_0 = k)$ a.s., $g_h(k) \to C\mu^{\beta\mp\epsilon} P(U_0 = k)$ a.s. Therefore, (4.65) follows if

$$\sum_{k\in\mathbb{Z}} g_h(k) = C\sum_{k=1}^{h-1} E\left(\frac{\widetilde{N}_{h-k}}{h-k}\right)^{-\beta\pm\epsilon}\left(1-\frac{k}{h}\right)^{-\beta\pm\epsilon} P(U_0=k) \to C\sum_{k=1}^{\infty}\mu^{\beta\mp\epsilon}P(U_0=k), \quad (4.66)$$

as $h\to\infty$. Without loss of generality, suppose that $\epsilon=0$. For some small $\delta>0$, we can rewrite the left-hand side of (4.66) as

$$\sum_{k=1}^{(1-\delta)h} E\left(\frac{\widetilde{N}_{h-k}}{h-k}\right)^{-\beta}\left(1-\frac{k}{h}\right)^{-\beta} P(U_0=k)$$

$$+\sum_{k=(1-\delta)h+1}^{h-1} E\left(\frac{\widetilde{N}_{h-k}}{h-k}\right)^{-\beta}\left(1-\frac{k}{h}\right)^{-\beta} P(U_0=k). \quad (4.67)$$

By the dominated convergence theorem, the first term of (4.67) converges to

$$\sum_{k=1}^{\infty}\mu^{\beta}P(U_0=k), \quad \text{as } h\to\infty.$$

From (4.61) and assumption (4.32), the second term of (4.67) is bounded by

$$Ch^{1-p}\sum_{k=(1-\delta)h}^{h-1}\left(1-\frac{k}{h}\right)^{-\beta}\left(\frac{k}{h}\right)^{-p}\frac{1}{h}, \quad (4.68)$$

for some constant $C>0$. Therefore (4.68) converges to zero as $h\to\infty$ since

$$\sum_{k=(1-\delta)h}^{h-1}\left(1-\frac{k}{h}\right)^{-\beta}\left(\frac{k}{h}\right)^{-p}\frac{1}{h} \to \int_{1-\delta}^{1}(1-x)^{-\beta}x^{-p}dx<\infty,$$

and $p>1$. $\square$

4.6.2 The proof of Proposition 4.3.2

Observe that

$$\mathrm{Var}\left(\sum_{k=1}^{T} Y_k^T\right) = \sum_{k=-(T-1)}^{T-1}(T-|h|)\gamma_{Y^T}(h) = T\gamma_{Y^T}(0) + 2\sum_{k=1}^{T-1}\sum_{l=1}^{k}\gamma_{Y^T}(l)$$

$$= T\frac{\sigma^2}{1-\varphi_T^2} + 2\sum_{k=1}^{T-1}\sum_{l=1}^{k}\frac{\sigma^2}{1-\varphi_T^2}\varphi_T^l$$

$$= \frac{\sigma^2}{1-\varphi_T^2}\left(T + 2\sum_{k=1}^{T-1}\frac{\varphi_T(1-\varphi_T^k)}{1-\varphi_T}\right)$$

$$= \frac{\sigma^2}{1-\varphi_T^2}\left(T + \frac{2\varphi_T}{1-\varphi_T}(T-1) - 2\frac{\varphi_T}{1-\varphi_T}\frac{\varphi_T(1-\varphi_T^{T-1})}{1-\varphi_T}\right)$$

$$= \frac{\sigma^2}{1-\varphi_T^2}\left(\frac{1+\varphi_T}{1-\varphi_T} - \frac{2\varphi_T}{(1-\varphi_T)^2} + \frac{\varphi_T^{T+1}}{(1-\varphi_T)^2}\right).$$

Since $\varphi_T^T \to 0$ for $\gamma \in (0,1)$, and $\varphi_T \to 1$, $1-\varphi_T^2 \sim 2(1-\varphi_T)$, it follows that

$$\operatorname{Var}\left(\sum_{k=1}^{T} Y_k^T\right) \sim \sigma^2\left(\frac{T}{(1-\varphi_T)^2} - \frac{1}{(1-\varphi_T)^3} + \frac{o(1)}{2(1-\varphi_T)^3}\right) \sim \frac{\sigma^2}{c^2}T^{2\gamma+1}. \qquad \square$$

CHAPTER 5

Statistical tests for changes in mean against long range dependence

5.1 Introduction

Long range dependent (LRD) time series $\{X_j\}_{j\in\mathbb{Z}}$ are commonly defined as second order stationary time series models with a hyperbolically decaying covariance function,

$$\gamma(h) = \mathrm{Cov}(X_j, X_{j+h}) \sim Ch^{2H-2}, \quad \text{as } h \to \infty, \tag{5.1}$$

where $C > 0$ is a constant and

$$H \in \left(\frac{1}{2}, 1\right) \tag{5.2}$$

is a self-similarity (SS) parameter. The parameter

$$d = H - \frac{1}{2} \in \left(0, \frac{1}{2}\right) \tag{5.3}$$

is called long range dependence (LRD) parameter. LRD series are used to model real time series in many fields such as hydrology, economics, telecommunications, and have been studied extensively from a theoretical perspective. See Beran (1994), Park and Willinger (2000), Embrechts and Maejima (2002), Doukhan, Oppenheim and Taqqu (2003), Robinson (2003), Samorodnitsky (2006), Palma (2007).

Note that the decay of autocovariances in (5.1) is so slow that, for LRD series,

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \infty. \tag{5.4}$$

In contrast, when

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \tag{5.5}$$

a time series model is often referred to as being short range dependent (SRD). In the spectral domain (and under mild assumptions), LRD series have a spectral density $f(\lambda)$ satisfying

$$f(\lambda) \sim c\lambda^{1-2H}, \quad \text{as } \lambda \to 0. \tag{5.6}$$

SRD series also satisfy (5.6) with

$$H = \frac{1}{2},$$

which is one reason why this value of H is often associated with SRD.

The property (5.4) of "infinite memory" of LRD series has not been and may not be necessarily easy to accept, especially in the areas where physical models for LRD series are not available. Related to this, a characteristic feature of LRD series is that they exhibit apparent changes in local mean level over a wide range of large scales (see top left plot in Figure 5.1). This feature has suggested, in particular, that realizations of LRD series can be confused easily with those of suitable and simple nonstationary models for finite samples. One such natural nonstationary model is SRD series superimposed by changes in (local) mean level, namely, the series

$$X_j = \mu_j + \epsilon_j, \quad j = 1, \dots, n, \tag{5.7}$$

where $\{\epsilon_j\}_{j\in\mathbb{Z}}$ is SRD, and

$$\mu_j = \mu + \sum_{r=1}^{R} \Delta_r 1_{\{k_r < j \le n\}} \tag{5.8}$$

with Δ_r representing changes in mean level at R break times k_r, $r = 1, \dots, R$. For later reference, we will refer to (5.7)-(5.8) as changes in mean (CM) model. The left plots of Figure 5.1 depict single realizations of long range dependent series (FARIMA$(0, .4, 0)$ series corresponding to $d = .4$ and $H = .5 + .4 = .9$) and CM series with two breaks $R = 2$ and AR(1) SRD series $\{\epsilon_j\}_{j\in\mathbb{Z}}$. In the corresponding right plots, both sample autocorrelation functions decay slowly.

The confusion between LRD series and nonstationary alternatives has been documented well, and is raised in almost all applications of LRD series. See, for example, Klemeš (1974), Boes and

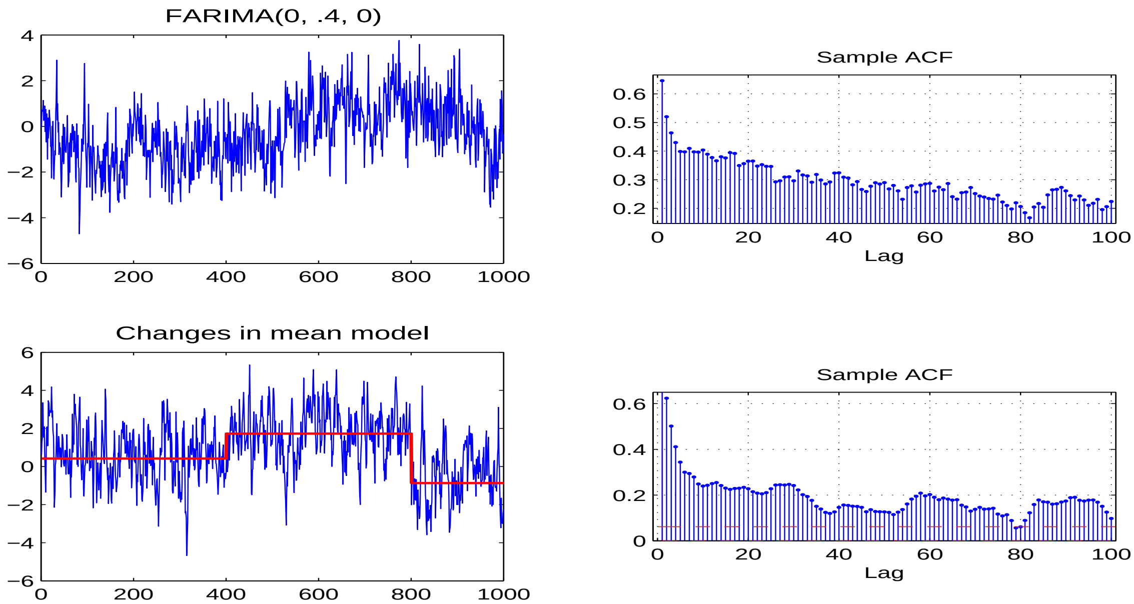

Figure 5.1: Time plots of long range dependent time series FARIMA$(0, .4, 0)$ and changes in mean model. Both time series exhibit slowly decaying sample autocorrelation functions.

Salas (1978) in hydrology, Roughan and Veitch (1999), Veres and Boda (2000), Karagiannis, Molle and Faloutsos (2004) in teletraffic, Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004), Smith (2005), Charfeddine and Guegan (2009) in economics and finance, Mills (2007) in climatology. This list is by far exhaustive. In a related direction, a number of authors have studied classical change point estimation methods on LRD series (e.g. Kuan and Hsu (1998), Wright (1998)).

More recently, a number of formal statistical procedures, and informal tests and exploratory tools were proposed to distinguish between LRD and nonstationary alternatives. For more informal procedures, see Bisaglia and Gerolimetto (2009), and Chapter 4 above. More formal statistical procedures can be divided into two classes of tests which have either

- LRD as null hypothesis, or
- Nonstationary model as null hypothesis.

For tests where LRD is null, see Shimotsu (2006), Ohanissian, Russell and Tsay (2008), Müller and Watson (2008), Qu (2009), Kuswanto (2009). For tests where a nonstationary model is null, see Berkes, Horváth, Kokoszka and Shao (2006), Jach and Kokoszka (2008). Some related works are Hariz, Wylie and Zhang (2007), Gil-Alana (2008).

This chapter concerns the class of tests where the null hypothesis involves a nonstationary model. As in Berkes, Horváth, Kokoszka and Shao (2006), the nonstationary model is the changes in mean (CM) model (5.7)-(5.8). The tests introduced here are devised to improve on the test by Berkes et al. (2006), to which we will refer as the BHKS test. In this regard, we show below (see Section 5.2) that the BHKS test suffers from low power against LRD alternatives. We argue that this is because the BHKS test statistic involves estimation of the sum of covariances of the underlying series. This is akin to issues around R/S-statistic and its modification (Lo (1991), Teverovsky, Taqqu and Willinger (1999)).

Estimation of the variance could be sidestepped by considering the BHKS test statistic for varying sample sizes and by performing a suitable regression (see Section 5.2). This procedure can be viewed as the BHKS-based estimation of SS (or LRD) parameter applied to the series after removing changes in mean. The resulting test would then check whether the parameter is in the SRD regime. The BHKS-based estimation is similar to popular R/S estimation method. As with the latter method, obtaining asymptotics of the BHKS-based estimator does not appear feasible. Moreover, better estimation methods for SS (or LRD) parameter are available such as GPH method (Geweke and Porter-Hudak (1983)) or local Whittle method (Robinson (1995a)). Adapting these methods leads to the tests considered in this chapter.

More specifically, our tests can be described as follows. As in Berkes et al. (2006), we focus on CM model with $R = 1$ break and a sequential procedure is used for testing CM model with $R = 1$ break against LRD. At the first stage, the test is for

H_0: SRD model

against CM model with $R = 1$ break, and LRD model. At the second stage, which is the focus of this chapter and that of Berkes et al. (2006), the hypotheses testing problem is

H_0: CM model with $R = 1$ break,

H_1: LRD model.

For this problem, given a series $X_j,\ j = 1, \dots, n$, define a new series obtained by removing one "break" from the series X_j,

$$R_j = X_j - \widehat{X}_j, \tag{5.9}$$

where

$$\widehat{X}_j = \begin{cases} \dfrac{1}{\widehat{k}} \displaystyle\sum_{s=1}^{\widehat{k}} X_s, & j = 1, \dots, \widehat{k}, \\ \dfrac{1}{n-\widehat{k}} \displaystyle\sum_{s=\widehat{k}+1}^{n} X_s, & j = \widehat{k}+1, \dots, n, \end{cases} \tag{5.10}$$

and $\widehat{k}$ is a suitable "break" point estimator (see Section 5.2). For example, in the test based on the local Whittle method, we consider the estimator $\widehat{H}_{lw}$ of SS parameter obtained by the method when applied to the series R_j. We show that, under suitable technical assumptions,

$$\sqrt{m}\left(\widehat{H}_{lw} - \frac{1}{2}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}\right) \tag{5.11}$$

under the hypothesis H_0, where m is the number of Fourier frequencies used in estimation, and that

$$\sqrt{m}\left(\widehat{H}_{lw} - \frac{1}{2}\right) \xrightarrow{p} +\infty \tag{5.12}$$

under H_1 (see Section 5.3). The results (5.11)-(5.12) lead to a consistent test for the hypothesis H_0 against H_1. We show numerically that the resulting test has much better power against LRD series than the BHKS test.

In the case of the test based on the GPH method, we prove the result analogous to (5.11). The result analogous to (5.12) is conjectured but not proved. The difficulties are well known in dealing with the GPH estimator, and here they are more pronounced because estimator involves the "break" point which is difficult to handle in the LRD regime.

In the first stage above, Berkes et al. (2006) use a well-known CUSUM statistic, and show that it diverges under the two alternatives of interest (CM model with $R = 1$ break and LRD model). For this stage and in the spirit of the second stage above, we also suggest to use the GPH or local Whittle estimator applied to the series itself (see Section 5.4). We prove the necessary theoretical results for the underlying estimators, for example, that (5.12) holds under CM model with $R = 1$ break, where $\widehat{H}_{lw}$ is the local Whittle estimator for the series itself. We show numerically that, at least for LRD alternatives, the resulting test has a much better power than the original CUSUM test.

Though the focus is on testing for CM model with just one break $R = 1$, we also briefly discuss obvious extension of the approach above to $R = 2, 3$, etc. number of breaks (see Section 5.4). The reason we do not formalize and pursue a general problem of an arbitrary number of breaks R is more practical. For moderate sample sizes of practical interest, it is too difficult to distinguish between LRD series and CM model with $R = 2, 3$ or more breaks or, said differently, the power of any test of CM model with 2, 3 or more breaks would be extremely low against LRD alternatives. We presume this was also the reason for the same focus in Berkes et al. (2006).

The structure of this chapter is as follows. We revisit the BHKS test in Section 5.2. The tests based on the GPH and local Whittle methods are considered in Section 5.3, in the case of CM model with $R = 1$ break. The situation of arbitrary number of breaks is discussed in Section 5.4. A simulation study and application to real data sets can be found in Sections 5.5 and 5.6, respectively. All the technical proofs are moved to Section 5.7.

5.2 BHKS test revisited

In this section, we revisit the BHKS test and explain why it has low power against long range dependent alternatives. The explanation will then naturally lead to the tests of this chapter.

5.2.1 Testing procedure

As in Berkes et al. (2006), we focus on testing procedure for the changes in mean (CM) model (5.7)-(5.8) with one break $R = 1$, namely,

$$X_j = \mu + \Delta 1_{\{k^* < j \le n\}} + \epsilon_j, \quad j = 1, \dots, n, \tag{5.13}$$

where μ, Δ and k^* are unknown, and $\{\epsilon_j\}$ is a SRD series (satisfying suitable assumptions below). The interest is in the hypothesis testing of:

$$\begin{aligned} &\mathrm{H}_0\text{: series } \{X_j\}_{j\in\mathbb{Z}} \text{ follows CM model (5.13),} \\ &\mathrm{H}_1\text{: series } \{X_j\}_{j\in\mathbb{Z}} \text{ is LRD.} \end{aligned} \tag{5.14}$$

The testing procedure involves a classical nonparametric change-point estimator

$$\widehat{k} = \min\left\{k : \max_{1\le l\le n}\left|\sum_{j=1}^{l} X_j - \frac{l}{n}\sum_{j=1}^{n} X_j\right| = \left|\sum_{j=1}^{k} X_j - \frac{k}{n}\sum_{j=1}^{n} X_j\right|\right\} \tag{5.15}$$

and the variables

$$T(0,\widehat{k}) = \frac{1}{s_{n,1}}\widehat{k}^{-1/2}\max_{1\le k\le \widehat{k}}\left|\sum_{j=1}^{k} X_j - \frac{k}{\widehat{k}}\sum_{j=1}^{\widehat{k}} X_j\right|, \tag{5.16}$$

$$T(\widehat{k},n) = \frac{1}{s_{n,2}}(n-\widehat{k})^{-1/2}\max_{\widehat{k}<k\le n}\left|\sum_{\widehat{k}<j\le k} X_j - \frac{k-\widehat{k}}{n-\widehat{k}}\sum_{\widehat{k}<j\le n} X_j\right|, \tag{5.17}$$

where

$$s_{n,1}^2 = \sum_{h=-q(\widehat{k})}^{q(\widehat{k})}\left(1-\frac{|h|}{q(\widehat{k})+1}\right)\widehat{\gamma}_1(h), \quad s_{n,2}^2 = \sum_{h=-q(n-\widehat{k})}^{q(n-\widehat{k})}\left(1-\frac{|h|}{q(n-\widehat{k})+1}\right)\widehat{\gamma}_2(h) \tag{5.18}$$

are the so-called Bartlett estimators of the sum of the time series covariances, with a bandwidth $q(\cdot)$ and

$$\widehat{\gamma}_1(h) = \frac{1}{\widehat{k}}\sum_{1\le j\le \widehat{k}-|h|}\left(X_j - \frac{1}{\widehat{k}}\sum_{1\le j\le \widehat{k}} X_j\right)\left(X_{j+h} - \frac{1}{\widehat{k}}\sum_{1\le j\le \widehat{k}} X_j\right),$$

$$\widehat{\gamma}_2(h) = \frac{1}{n-\widehat{k}}\sum_{\widehat{k}<j\le n-|h|}\left(X_j - \frac{1}{n-\widehat{k}}\sum_{\widehat{k}<j\le n} X_i\right)\left(X_{j+h} - \frac{1}{n-\widehat{k}}\sum_{\widehat{k}<j\le n} X_j\right).$$

The BHKS test statistic is now defined as

$$M_2 = \max\left\{T(0,\widehat{k}), T(\widehat{k},n)\right\}. \tag{5.19}$$

For the asymptotics of the BHKS test statistic M_2 under H_0, the following assumptions are made. For the series $\{\epsilon_j\}$, define its autocovariance by $\gamma(h) = \mathrm{Cov}(\epsilon_0, \epsilon_h)$ and its fourth order cumulant by $\kappa(h,r,s) = E\epsilon_0\epsilon_h\epsilon_r\epsilon_s - (\gamma(h)\gamma(r-s) + \gamma(r)\gamma(h-s) + \gamma(s)\gamma(h-r))$.

Assumption 5.2.1. The series $\{\epsilon_j\}$ is a zero mean, fourth-order stationary satisfying:

$$n^{-1/2}\sum_{1\le j\le nt}\epsilon_j \xrightarrow{d} \sigma B(t) \quad \text{in } D[0,1] \tag{5.20}$$

for some $\sigma > 0$ and standard Brownian motion $\{B(t)\}_{t\in[0,1]}$, and

$$\sum_h |\gamma(h)| < \infty, \quad \sup_h \sum_{r,s} |\kappa(h,r,s)| < \infty. \tag{5.21}$$

In addition, the following assumptions will be made:

(A1) $k^* = [n\theta]$ for some $0 < \theta < 1$,

(A2) $n\Delta^2 \to \infty$,

(A3) $\Delta^2|\widehat{k} - k^*| = O_p(1)$

and also $q(n) \to \infty$ is such that

$$\sup_{k\geq 0} \frac{q(2^{k+1})}{q(2^k)} < \infty, \quad q(n) \to \infty \quad \text{and} \quad q(n)(\log n)^4 = O(n). \tag{5.22}$$

Theorem 5.2.1. *(Berkes et al. (2006)) Suppose that Assumption 5.2.1, (A1), (A2), (A3) and (5.22) above hold. Then, under* H_0*,*

$$M_2 \xrightarrow{d} \max\left\{\sup_{0\leq t\leq 1} |\mathcal{W}^{(1)}(t)|, \sup_{0\leq t\leq 1} |\mathcal{W}^{(2)}(t)|\right\}, \tag{5.23}$$

where $\mathcal{W}^{(1)}$ *and* $\mathcal{W}^{(2)}$ *are independent Brownian bridges.*

For the asymptotics of M_2 under H_1, the following assumptions are made. Recall that a standard fractional Brownian motion $B_H = \{B_H(t)\}_{t\in[0,1]}$ with parameter H is a zero mean Gaussian process having covariance structure

$$EB_H(t)B_H(s) = \frac{1}{2}\left\{t^{2H} + s^{2H} - |t-s|^{2H}\right\}, \quad 0 \leq s,t \leq 1. \tag{5.24}$$

Assumption 5.2.2. The series $\{X_j\}$ is fourth order stationary satisfying

$$n^{-H} \sum_{1\leq j\leq nt} (X_j - EX_j) \xrightarrow{d} c_H B_H(t) \quad \text{in } D[0,1], \tag{5.25}$$

for some $c_H > 0$, and standard fractional Brownian motion B_H with

$$\frac{1}{2} < H < 1,$$

and the covariances $\gamma(h)$ and fourth order cumulants $\kappa(h, r, s)$ of the series $\{X_j\}$ satisfy

$$\gamma(h) \sim c_0 h^{2H-2}, \qquad \text{as } h \to \infty, \tag{5.26}$$

$$\sup_h \sum_{-n \le r,s \le n} |\kappa(h, r, s)| = O(n^{2H-1}), \tag{5.27}$$

for some $c_0 > 0$.

It will also be assumed that $q(n) \to \infty$ is such that

$$q(n) \uparrow, \quad \sup_{k \ge 0} \frac{q(2^{k+1})}{q(2^k)} < \infty, \quad q(n) = O(n(\log n)^{-7/(4-4H)}). \tag{5.28}$$

Define also a fractional Brownian bridge $\mathcal{W}_H = \{\mathcal{W}_H(t)\}_{t \in [0,1]}$ as $\mathcal{W}_H(t) = B_H(t) - tB_H(1)$ and let

$$\xi = \inf\left\{t \ge 0 : |\mathcal{W}_H(t)| = \sup_{0 \le s \le 1} |\mathcal{W}_H(s)|\right\}. \tag{5.29}$$

Theorem 5.2.2. *(Berkes et al. (2006)) Suppose that Assumption 5.2.2 and (5.28) above hold. Then, under* H_1*,*

$$\left[\left(\frac{q(\widehat{k})}{n}\right)^{H-1/2} T(0, \widehat{k}), \left(\frac{q(n-\widehat{k})}{n}\right)^{H-1/2} T(\widehat{k}, n)\right] \xrightarrow{d} [Z_1, Z_2], \tag{5.30}$$

where

$$Z_1 = \frac{1}{\sqrt{\xi}} \sup_{0 \le t \le \xi} \left| B_H(t) - \frac{t}{\xi} B_H(\xi)\right|,$$

$$Z_2 = \frac{1}{\sqrt{1-\xi}} \sup_{\xi \le t \le 1} \left| (B_H(t) - B_H(\xi)) - \frac{t-\xi}{1-\xi}(B_H(1) - B_H(\xi))\right|.$$

In particular, under H_1*,* $M_2 \xrightarrow{p} +\infty$*.*

	$n=500$		$n=1000$		$n=5000$	
	5%	10%	5%	10%	5%	10%
$d=.1$	0.7	3.5	1.6	5.9	12.6	24.8
$d=.2$	0.5	3.8	3.9	10.0	27.7	42.4
$d=.3$	0.2	3.6	4.2	13.1	46.2	59.1
$d=.4$	0.3	2.8	6.2	17.4	61.4	71.9

Table 5.1: Power of Berkes et al. (2006) for FARIMA(0,d,0) time series. Empirical power calculated from 1000 realizations.

5.2.2 Low power against LRD alternatives

Though the BHKS test has nice theoretical properties, its power against LRD series is very small for moderate sample sizes. We illustrate this on Gaussian FARIMA$(0,d,0)$ series which are LRD with

$$H = d + \frac{1}{2}, \quad d \in \left(0, \frac{1}{2}\right).$$

Table 5.1 shows empirical power of the BHKS test for these series at two significance levels 5% and 10%. We have selected

$$q(n) = 15 \log_{10} n,$$

which is the bandwidth used in Berkes et al. (2006). Observe that the power increases as LRD parameter or sample size increases. From a practical perspective, however, the power is too small for moderate sample sizes. For instance, the empirical power is only 6.2% for $d = .4$ with sample size $n = 1000$ and 5% significance level.

We argue next that the test has small power and is not very reliable because of the presence of estimators (5.18) of the sum of the time series covariances. It is shown in Theorem A.1, (ii), of Berkes et al. (2006) that, for LRD series,

$$q(n)^{1-2H} s_n^2 \to c^2, \qquad \text{a.s.}, \tag{5.31}$$

for some constant c. In view of (5.15), (5.16) and by using (5.25), this suggests that

$$T(0,\widehat{k}) \simeq \frac{1}{cq(\widehat{k})^{H-1/2}} \widehat{k}^{-1/2} \max_{1\le k\le \widehat{k}} \left| \sum_{j=1}^{k} X_j - \frac{k}{\widehat{k}} \sum_{j=1}^{\widehat{k}} X_j \right|$$

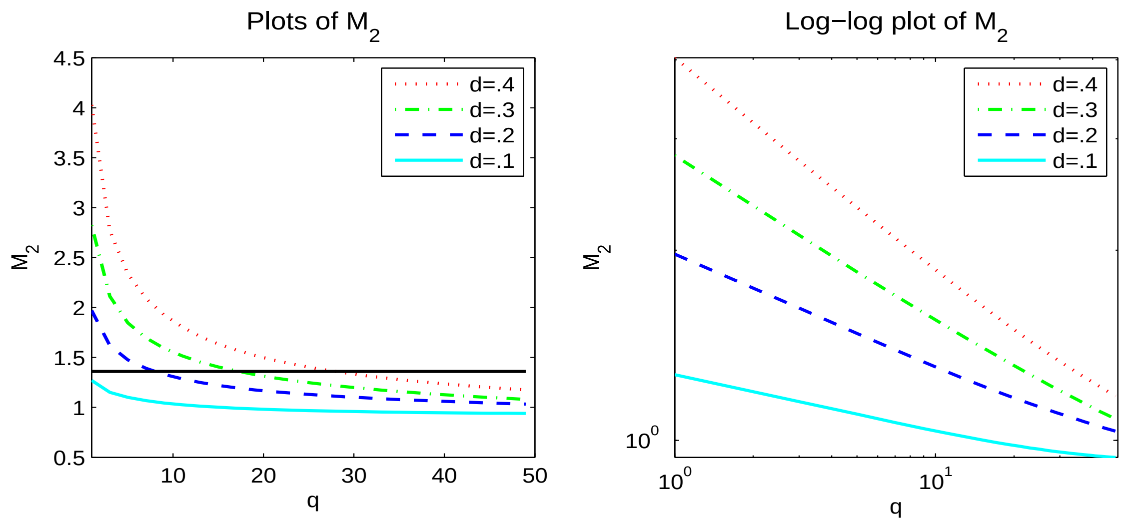

Figure 5.2: Plots of mean of test statistic M_2 as a function of q in the bandwidth $q(n) = q\log_{10} n$, with fixed n.

$$\simeq \frac{1}{c}\left(\frac{q(\widehat{k})}{\widehat{k}}\right)^{1/2-H} \widehat{k}^{-H} \max_{1\le k\le \widehat{k}} \left|\sum_{j=1}^{k} X_j - \frac{k}{\widehat{k}}\sum_{j=1}^{\widehat{k}} X_j\right| \simeq C_1 \left(\frac{q(\widehat{k})}{\widehat{k}}\right)^{1/2-H},$$

for a random variable $C_1 = c^{-1}\sup_{0\le u\le\xi} |B_H(u) - (u/\xi)B_H(\xi)|$, where ξ is given in (5.29). In particular, with the choice of bandwidth function

$$q(n) = q\log_{10} n, \tag{5.32}$$

we have

$$T(0,\widehat{k}) \simeq C_{1,n} q^{1/2-H},$$

where $C_{1,n} = C_1(\log_{10}\widehat{k}/\widehat{k})^{1/2-H}$. After a similar argument for $T(\widehat{k}, n)$, the BHKS test statistic behaves as

$$M_2 \simeq C_n q^{1/2-H}. \tag{5.33}$$

Figure 5.2 illustrates the relation (5.33). We have generated FARIMA$(0, d, 0)$ time series of length $n = 10,000$ for several values of d. The two plots are the mean values of the test statistic M_2, over 100 replications and as functions of q in (5.32) with n being fixed. A constant line in the left plot represents 10% asymptotic critical value for the test, which is 1.36. The right plot is the log of the mean values against $\log q$. In accord with the relation (5.33) approximately straight lines are observed and the slopes are close to $-d = 1/2 - H$.

Several conclusions can be drawn from the relation (5.33) and a supporting Figure 5.2. First, for a small sample size n and for LRD alternatives, the BHKS test statistic M_2 (and hence the test itself) will be highly sensitive to the choice of the bandwidth $q(n)$ or q in (5.32). Second, for SRD series under the null hypothesis, one would like to take larger bandwidth $q(n)$ and thus capture well short range correlations in the series. As $q \to \infty$, the right-hand side of (5.33) converges to zero and hence, for larger q, the test will rarely reject H_0 when the underlying series is LRD. This partly explain the low power of the BHKS test as reported in Table 5.1.

A discussion similar to that above can also be found in Teverovsky et al. (1999), and concerns the behavior of a modified R/S statistic proposed by Lo (1991). A classical R/S statistic is defined as

$$\frac{R}{S}(n) = \frac{1}{S(n)} \left\{ \max_{1 \le k \le n} - \min \left(\sum_{j=1}^{k} X_j - \frac{k}{n} \sum_{j=1}^{n} X_j \right) \right\}, \tag{5.34}$$

where $S^2(n)$ is the sample variance

$$S^2(n) = \frac{1}{n} \sum_{j=1}^{n} \left(X_j - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 .$$

Lo (1991) proposed a modified R/S-statistic to account for short range correlations. The Lo's modification replaces the sample variance by Bartlett estimator

$$S_q^2(n) = \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q+1} \right) \widehat{\gamma}(h),$$

where $\widehat{\gamma}(h)$ are sample auto covariances. The author then uses the modified statistic to devise a test for SRD against LRD.

As in the discussion above, for LRD series, Teverovsky et al. (1999) argue that

$$S_q^2(n) \simeq c_n q^{2H-1},$$

and hence that

$$\frac{1}{\sqrt{n}} \frac{R}{S_q}(n) \simeq C_{0,n} q^{1/2-H}. \tag{5.35}$$

This relation is exactly the same as (5.33). Regarding (5.35), Teverovsky et al. (1999) conclude similarly that the Lo's test is not very reliable and will have very small power against LRD alternatives.

5.2.3 Towards improving the BHKS test

Another important message in Teverovsky et al. (1999) is that a careful R/S-estimation of SS parameter (and deciding whether $H = 1/2$ or $H > 1/2$) may already do better than the test (of SRD against LRD) suggested by Lo. (In fairness to Lo (1991), the author also seemingly acknowledges this fact on p. 1296.) One reason for this is that R/S-estimation of SS parameter does not involve estimation of the sum of covariances as in (5.18). Since R/S statistic is closely related to the variables (5.16) and (5.17), this point can be explained here in the context of the BHKS test.

For example, consider

$$U_1 := s_{n,1} T(0, \widehat{k}) = \frac{1}{\sqrt{\widehat{k}}} \max_{1 \le k \le \widehat{k}} \left| \sum_{j=1}^{k} X_j - \frac{k}{\widehat{k}} \sum_{j=1}^{\widehat{k}} X_j \right|, \tag{5.36}$$

which is the variable $T(0, \widehat{k})$ without the estimator $s_{n,1}$. Under H_0 (that is, CM model with one break), one expects as in Theorem 5.2.1 that

$$U_1 \xrightarrow{d} \sigma Z_0, \tag{5.37}$$

where $Z_0 \stackrel{d}{=} \sup_{0 \le t \le 1} |\mathcal{W}^{(1)}(t)|$ and $\sigma^2 = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(\epsilon_0, \epsilon_h)$. Under H_1 (that is, LRD model), one expects as in Theorem 5.2.2 and (5.31) that

$$n^{-(H-1/2)} U_1 \xrightarrow{d} c Z_1, \tag{5.38}$$

where Z_1 appears in (5.30). Taking the logarithm in (5.37) and (5.38), we have

$$\log U_1 \simeq \log C + (H - 1/2) \log n, \tag{5.39}$$

with random C, under both H_0 corresponding to $H = 1/2$ and H_1 corresponding to $H > 1/2$.

The relation (5.39) naturally suggests an estimator of SS parameter in the spirit of R/S estimation, applied to the series $\{X_1, \dots, X_{\widehat{k}}\}$. More precisely, consider subsamples $\{X_t, X_{t+1}, \dots, X_{t+m-1}\}$ for suitable block size m and starting point t. For each subsample, calculate the statistic $U_1 = U_1(t,m)$. In view of (5.39), SS parameter can then be estimated by regressing $\log U_1(t,m)$ on $\log m$. Note that this procedure does not involve estimators of the sum of covariances, unlike the BHKS test.

As in Teverovsky et al. (1999), a new test based on above estimation of SS parameter would be expected more superior than the BHKS test. However, we shall not expand and not pursue this estimation method here. As with R/S statistic, the asymptotic results for the BHKS-based estimator of SS parameter would not be easy to derive. Moreover other estimation methods of SS parameter are now widely preferred. We turn next to two such methods, namely, the GPH and local Whittle, which are two popular methods in the spectral domain.

5.3 Tests based on estimation of SS parameter

As in Section 5.2, we consider hypothesis testing problem (5.14). We propose tests based on SS parameter estimation after removing one change in mean. Thus, let

$$R_j = X_j - \widehat{X}_j, \quad j = 1, \dots, n, \tag{5.40}$$

where $\widehat{X}_j$ is defined in (5.10). SS parameter estimation is in the spectral domain, and let

$$I_Y(\omega_l) = \frac{1}{2\pi n} \left| \sum_{j=1}^{n} Y_j e^{-ij\omega_l} \right|^2, \tag{5.41}$$

denote the periodogram of a general series $\{Y_j\}$ at the Fourier frequencies $\omega_l = 2\pi l/n$.

Remark. As in Section 5.2.3, another possibility is for the test to be based on time series X_j till (or after) the first break time $\widehat{k}$. In simulations (not reported here), our tests described below perform just slightly better when based on the series (5.40). This is why we work here with the series (5.40).

5.3.1 Test based on local Whittle method

The local Whittle estimator of SS parameter for the series $\{R_j\}$ is defined as

$$\widehat{H}_{lw} = \underset{H \in \Theta}{\operatorname{argmin}}\, R(H), \tag{5.42}$$

where $\Theta = [\Delta_1, \Delta_2]$ with $0 < \Delta_1 < \Delta_2 < 1$ and

$$R(H) = \log\left(\frac{1}{m}\sum_{l=1}^{m} \omega_l^{2H-1} I_R(\omega_l)\right) - (2H-1)\frac{1}{m}\sum_{l=1}^{m} \log \omega_l \tag{5.43}$$

with m denoting the number of low frequencies used in estimation (see Robinson (1995a)). Our test statistic is defined as

$$M_{lw} = \sqrt{m}\left(\widehat{H}_{lw} - \frac{1}{2}\right). \tag{5.44}$$

Its asymptotics for CM model with $R = 1$ break, that is, under H_0, is stated next.

Theorem 5.3.1. *Suppose that CM model (5.13) satisfies (5.20) and (A1)-(A3) from Section 5.2.1. Furthermore, suppose that the series $\{\epsilon_j\}$ and m satisfy the assumptions of Theorem 2 of Robinson (1995*a*) with*

$$H_0 = \frac{1}{2}.$$

Finally, assume that

$$\frac{m \log^2 m}{n\Delta^2} \to 0. \tag{5.45}$$

Then,

$$M_{lw} = \sqrt{m}\left(\widehat{H}_{lw} - \frac{1}{2}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}\right). \tag{5.46}$$

On the other hand, for LRD series, that is, under H_1, we have the following result.

Theorem 5.3.2. *Suppose that the assumptions of Theorem 1 in Robinson (1995*a*) hold for m and the time series $\{X_j\}$ with true $H_0 \in (1/2, 1)$. Suppose also that*

$$n^{-H_0} \sum_{1 \le j \le nt} (X_j - EX_j) \xrightarrow{d} \sigma B_{H_0}(t) \quad \textit{in } D[0,1], \tag{5.47}$$

where $\sigma > 0$ and B_{H_0} is a standard fractional Brownian motion. Then,

$$\widehat{H}_{lw} \xrightarrow{p} H_0. \tag{5.48}$$

In particular, $M_{lw} \xrightarrow{p} +\infty$.

5.3.2 Test based on GPH method

The GPH estimator of SS parameter for the series $\{R_j\}$ is defined through the regression of $\log I_R(\omega_l)$ on $\log \omega_l$, and is given by

$$\widehat{H}_{gph} = \frac{1}{2} + \sum_{l=1}^{m} a_l \log I_R(\omega_l) \tag{5.49}$$

with the regression weights $a_l = (z_l - \overline{z})/\sum_{l=1}^{m}(z_l - \overline{z})^2$, where $z_l = -2\log\omega_l$, $\overline{z} = \sum_{l=1}^{m} z_l/m$ (Geweke and Porter-Hudak (1983), Robinson (1995*b*)). As in (5.43), m is the number of low frequencies used in estimation. Our test statistic is defined as

$$M_{gph} = \sqrt{m}\left(\widehat{H}_{gph} - \frac{1}{2}\right). \tag{5.50}$$

Under H_0, it has the following asymptotics.

Theorem 5.3.3. *Suppose that CM model (5.13) satisfies assumptions (A1)-(A3) from Section 5.2.1. In addition, assume that the probability space on which $\{\epsilon_i\}$ are defined can be extended to a different probability space where there is a Brownian motion B with variance $EB(1)^2 = 2\pi f_\epsilon(0)$ such that*

$$\sup_{0\le t\le 1}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_j - B(t)\right| = o_p\left(\frac{1}{n^{1/2-\delta}}\right), \quad \delta > 0. \tag{5.51}$$

Suppose that

$$\frac{m^2\log^2 m}{n^{1/2-\delta}} \to 0, \quad \frac{m\log m}{(n\Delta^2)^{1/2-\delta_0}} \to 0, \tag{5.52}$$

for some $\delta_0 > 0$. Then,

$$M_{gph} = \sqrt{m}\left(\widehat{H}_{gph} - \frac{1}{2}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{24}\right). \tag{5.53}$$

The proof of Theorem 5.3.3 uses the approach of Phillips (2007) (see also Perron and Qu (2006)). In particular, sufficient conditions for (5.51) can be found in that work.

Under H_1, that is, for LRD series with true parameter H_0, we expect that

$$\widehat{H}_{gph} \xrightarrow{p} H_0 \tag{5.54}$$

and hence that $M_{gph} \xrightarrow{p} +\infty$. Proving (5.54), however, does not appear feasible with available tools. The reason is that the series $\{R_j\}$ involves the "break" point estimator $\widehat{k}$ of LRD series. How to address probabilistic questions about $\widehat{k}$ (or a related limiting variable ξ in (5.29)) remains an open question.

5.4 Tests for multiple breaks

When only an upper bound on the number of breaks is known, it is natural to consider a sequential testing procedure as described next. We follow the binary segmentation method of Vostrikova (1981), which is also considered in Berkes et al. (2006). At the first stage, consider testing of

H_0: series $\{X_j\}_{j\in\mathbb{Z}}$ is SRD

against LRD or CM model. Define

$$R_j^{(1)} = X_j - \overline{X}, \tag{5.55}$$

where $\overline{X} = 1/n\sum_{s=1}^{n} X_s$. Consider test statistics

$$M_{lw}^{(1)} = \sqrt{m}\left(\widehat{H}_{lw}^{(1)} - \frac{1}{2}\right), \quad M_{gph}^{(1)} = \sqrt{m}\left(\widehat{H}_{gph}^{(1)} - \frac{1}{2}\right), \tag{5.56}$$

where $\widehat{H}_{lw}^{(1)}$ and $\widehat{H}_{gph}^{(1)}$ are local Whittle and GPH estimators based on $\{R_j^{(1)}\}$, respectively.

Under suitable conditions, it is well known (Robinson (1995a, 1995b), Hurvich, Deo and Brodsky (1998)) that under H_0 (that is, SRD series),

$$M_{lw}^{(1)} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}\right), \quad M_{gph}^{(1)} \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{24}\right), \tag{5.57}$$

while for LRD series,

$$M_{lw}^{(1)} \xrightarrow{p} +\infty, \quad M_{gph}^{(1)} \xrightarrow{p} +\infty, \tag{5.58}$$

since $\widehat{H}_{lw}^{(1)}$ and $\widehat{H}_{gph}^{(1)}$ converge in probability to $H_0 > 1/2$.

Here is also the asymptotic result under the alternative of CM model with $R = 1$ break.

Theorem 5.4.1. *Suppose that CM model (5.13) satisfies assumptions (A1)-(A3). In addition, suppose that the series $\{X_j\}$ and m satisfy the assumptions of Theorem 2 of Robinson (1995*b*) and assume that*

$$\frac{m^2}{n\Delta^2} \to 0. \tag{5.59}$$

Then,

$$\widehat{H}_{lw}^{(1)} \xrightarrow{p} \Delta_2, \tag{5.60}$$

where Δ_2 enters $\Theta = [\Delta_1, \Delta_2]$ in (5.42). In particular, $M_{lw}^{(1)} \xrightarrow{p} +\infty$.

Regarding GPH estimator, we conjecture that for CM model with one break,

$$\widehat{H}_{gph}^{(1)} \xrightarrow{p} H_0(\theta) > 1/2, \tag{5.61}$$

where $H_0(\theta)$ is a function of the true break location θ, and hence that $M_{gph}^{(1)} \xrightarrow{p} +\infty$ as well. For example, consider the case where θ is a rational number between 0 to 1, namely, $\theta = p/q$ with $p, q \in \mathbb{Z}_+$ having no common devisor (larger than 1) and $p < q$.

By using (5.49) and (5.13), observe that

$$\widehat{H}_{gph}^{(1)} - \frac{1}{2} = \sum_{l=1}^{m} a_l \log\left(I_\epsilon(\omega_l) + 2\Re\left\{ \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{k^*} \epsilon_j e^{-ij\omega_l} \Delta c_l \right\} + \left| \frac{\Delta c_l}{\sqrt{2\pi n}} \right|^2 \right),$$

where

$$c_l = \sum_{j=1}^{k^*} e^{-ij\omega_l} = e^{-i\omega_l} \frac{e^{-i2\pi lp/q} - 1}{e^{-i\omega_l} - 1} =: \frac{b_l}{1 - e^{i\omega_l}}.$$

Note that b_l takes q values $e^{-2\pi s/q} - 1$, $s = 0, \ldots, q-1$, where $s = lp \bmod q$. Any value of $s = 0, \ldots, q-1$ is possible because p and q do not have common divisor (larger than 1). Moreover, letting $L_s = \{l = 1, \ldots, m : (lp \bmod q) = s\}$ and $|L_s|$ be the number of elements in

L_s, we have $|L_s| \sim m/q$, as $m \to \infty$. Denote b_l by g_s when $l \in L_s$. Write

$$\widehat{H}_{gph}^{(1)} - \frac{1}{2} =: T_1 + \sum_{s=1}^{q-1} T_{2,s},$$

where

$$T_1 = \sum_{L_0} a_l \log I_\epsilon(\omega_l),$$

$$T_{2,s} = \sum_{L_s} a_l \log \left(I_\epsilon(\omega_l) + 2\Re \left\{ \frac{1}{2\pi n} \sum_{j=1}^{k^*} \epsilon_j e^{-ij\omega_l} \frac{\Delta g_s}{1 - e^{i\omega_l}} \right\} + \frac{\Delta^2 |g_s|^2}{2\pi n |1 - e^{i\omega_l}|^2} \right).$$

As for the usual GPH estimator, we still expect that

$$T_1 \xrightarrow{p} 0.$$

On the other hand, we expect that $T_{2,s}$ is dominated by the last term in the logarithm and hence that

$$T_{2,s} = -2 \sum_{L_s} a_l \log |1 - e^{i\omega_l}| + o_p(1)$$

$$= \sum_{L_s} \frac{\log(l/m) + 1}{m} \log \left(\frac{l}{m} \right) + o_p(1) = \frac{1}{q} + o_p(1),$$

where we used the fact that $a_l = -(\log(l/m) + 1)/(2m) + o(1)$ (Hurvich et al. (1998), p. 38). Hence, we expect from above that

$$\widehat{H}_{gph}^{(1)} - \frac{1}{2} \xrightarrow{p} \frac{q-1}{q} = 1 - \frac{1}{q} =: H_0(\theta) - \frac{1}{2}.$$

The minimum in the limit is attained at $\theta = 1/2$ $(p = 1,\ q = 2)$. For irrational θ, we conjecture that $H_0(\theta) = 3/2$. We should also mention that theses conjectures, for both rational and irrational θ, are supported by simulations (not reported here).

If the null hypothesis is rejected at the first stage, then move to the second stage which is exactly that described in Section 5.3. If the null hypothesis is rejected at the second stage as well, move to the next stage. More precisely, consider the null hypothesis H_0: CM model with $R = 2$ breaks. To find the second break point, calculate $V(0, \widehat{k}_1)$ and $V(\widehat{k}_1, n)$, where $\widehat{k}_1$ is the first break and

$$V(a,b) := \max_{a+1\le k\le b}\left|\sum_{j=a+1}^{k} X_j - \frac{k}{b}\sum_{j=a+1}^{b} X_j\right|.$$

If, for example, $V(0,\widehat{k}_1) > V(\widehat{k}_1, n)$, then define the second break as

$$\widehat{k}_2 = \min\left\{k : \max_{1\le l\le \widehat{k}_1}\left|\sum_{j=1}^{l} X_j - \frac{l}{\widehat{k}_1}\sum_{j=1}^{\widehat{k}_1} X_j\right| = \left|\sum_{j=1}^{k} X_j - \frac{k}{\widehat{k}_1}\sum_{j=1}^{\widehat{k}_1} X_j\right|\right\} \tag{5.62}$$

and estimated local mean level as

$$\widehat{X}_j^{(3)} = \frac{1}{\widehat{k}_2}\sum_{s=1}^{\widehat{k}_2} X_s 1_{\{j\le \widehat{k}_2\}} + \frac{1}{\widehat{k}_1 - \widehat{k}_2}\sum_{s=\widehat{k}_2+1}^{\widehat{k}_1} X_s 1_{\{\widehat{k}_2 < j \le \widehat{k}_1\}} + \frac{1}{(n-\widehat{k}_1)}\sum_{s=\widehat{k}_1+1}^{n} X_s 1_{\{\widehat{k}_1 < j \le n\}}.$$

Now, consider the residual series

$$R_j^{(3)} = X_j - \widehat{X}_j^{(3)}$$

and perform the test as in Section 5.3 based on the SS parameter estimates of the series $R_j^{(3)}$. To continue if necessary, the third break point is determined by comparing $V(0,\widehat{k}_2)$, $V(\widehat{k}_2,\widehat{k}_1)$ and $V(\widehat{k}_1, n)$, and residuals are obtained by subtracting local mean levels separated by three break points.

Here is the summary of our sequential testing procedure for multiple breaks:

- Estimate SS parameter based on $R_j^{(1)} = X_j - \overline{X}$ and make a conclusion based on the test statistics in (5.56), that is,

$$M_{lw}^{(1)} = \sqrt{m}\left(\widehat{H}_{lw}^{(1)} - \frac{1}{2}\right), \quad M_{gph}^{(1)} = \sqrt{m}\left(\widehat{H}_{gph}^{(1)} - \frac{1}{2}\right).$$

- While the null hypothesis is rejected and in a sequential fashion, test for $R = r$ breaks using the test statistics

$$M_{lw}^{(r)} = \sqrt{m}\left(\widehat{H}_{lw}^{(r)} - \frac{1}{2}\right), \qquad M_{gph}^{(r)} = \sqrt{m}\left(\widehat{H}_{gph}^{(r)} - \frac{1}{2}\right),$$

where the local Whittle and GPH estimators $\widehat{H}_{lw}^{(r)}$ and $\widehat{H}_{gph}^{(r)}$ are obtained from residuals

$$R_j^{(r)} = X_j - \widehat{X}_j^{(r)}$$

with $\widehat{X}_j^{(r)}$ being local mean levels determined by $(r-1)$ breaks (as discussed above for $r=3$). The actual testing procedure is the same as in Section 5.3.

5.5 Simulation study

We report here on finite sample behavior of the proposed tests in simulations. We consider the CM model with $R = 1$ break,

$$X_j = \Delta 1_{\{k^* < j \le n\}} + \epsilon_j, \quad j = 1, \dots, n, \tag{5.63}$$

where jump size is randomly selected as

$$\Delta \sim \mathcal{N}(.5, 1)$$

and the break point k^* is selected at random from the interval $[.1n, .9n]$. SRD series $\{\epsilon_j\}_{j\in\mathbb{Z}}$ is such that

$$\epsilon_j \sim \text{ i.i.d. } \mathcal{N}(0,1), \qquad \text{or}$$

$$\epsilon_j = .7\epsilon_{j-1} + u_j, \quad u_j \sim \text{i.i.d. } \mathcal{N}(0,1),$$

and we refer to the corresponding model (5.63) as CM-WN and CM-AR, respectively. Gaussian FARIMA time series are considered for LRD model. The number of low frequencies used for both GPH and local Whittle estimators is $m = \sqrt{n}$. Empirical sizes and powers are calculated based on 1000 replications.

First, we examine the power of our test and CUSUM test at the first stage (Section 5.4), namely, testing for SRD series against CM model (5.63) or LRD series. Table 5.2 shows that our proposed tests are as powerful as CUSUM test when the true model is CM model (5.63) with one break. When it comes to LRD alternative, our proposed methods are much more powerful than CUSUM test. For example, the power of our test for FARIMA$(0, .4, 0)$ series

	CM-WN			CM-AR			FARIMA(0, .1 ,0)			FARIMA(0, .4, 0)		
	n=500, 1000, 5000			n=500, 1000, 5000			n=500, 1000, 5000			n=500, 1000, 5000		
CUSUM	.647	.828	.934	.305	.494	.797	.038	.069	.194	.186	.337	.766
LW	.786	.825	.918	.568	.570	.754	.260	.330	.486	.936	.977	1.00
GPH	.738	.784	.891	.507	.511	.676	.232	.276	.376	.877	.921	.999

Table 5.2: Empirical power of the first stage test for SRD series against CM model with one break or LRD series under 5% significance level.

	$n = 500$	$n = 1,000$	$n = 2,000$	$n = 5,000$
BHKS	1.1	1.3	3.6	5.8
LW	6.8	6.3	5.8	6.4
GPH	6.8	6.4	6.6	5.2

Table 5.3: Empirical size of the test for CM model with one break under 5% significance level.

		$n = 500$		$n = 1000$		$n = 2000$		$n = 5000$	
		5%	10%	5%	10%	5%	10%	5%	10%
$d = .1$	LW	11.6	16.5	12.0	18.9	18.2	27.2	28.2	38.8
	GPH	10.4	16.2	10.1	16.6	15.1	22.9	20.3	29.2
$d = .2$	LW	28.2	36.6	35.8	45.7	52.2	62.8	75.9	83.4
	GPH	21.4	31.1	25.1	36.2	38.7	51.8	61.6	72.4
$d = .3$	LW	49.6	58.5	68.1	76.4	84.3	89.1	97.4	98.5
	GPH	40.3	50.5	56.4	65.7	68.5	77.9	88.0	93.1
$d = .4$	LW	73.9	81.2	87.4	91.9	97.2	98.8	100	100
	GPH	61.9	72.2	76.9	84.8	89.9	94.3	99.3	99.8

Table 5.4: Power of our tests against FARIMA(0, d, 0) time series.

(corresponding LRD parameter $d = .4$ and SS parameter $H = .9$) with sample size $n = 1000$ is 92 % under 5% significance level, compared to only 34% when using CUSUM test.

Next, we report on the performance of our tests at the second stage, which is the focus of this chapter. Empirical sizes of the tests is reported in Table 5.3. Table 5.3 shows that the empirical sizes of our tests slightly exceed a given nominal significance level, while BHKS test seems to be too conservative for moderate sample size. Table 5.4 shows the power of our proposed tests against LRD series. The power increases as LRD parameter or sample size increases. In particular, compared to Table 5.1, the power of the tests has increased in all the cases considered. For example, the empirical power for FARIMA$(0, .4, 0)$ series with sample size $n = 1,000$ is now more than 76.9% while that of BHKS is only about 6%.

It may not be too appropriate to compare the power of two tests directly given that they have different empirical sizes. Hence, we also consider size adjusted power to correct for the

	BHKS	LW	GPH
$d = .1$	8.5	9.0	9.5
$d = .2$	10.8	32.8	26.1
$d = .3$	15.1	64.2	50.4
$d = .4$	19.3	86.6	74.9

Table 5.5: Size adjusted power of the second stage test for FARIMA(0,d,0) time series with sample size $n = 1,000$.

effect of different sizes of the tests. Based on the simulation study for CM-WN model with sample size $n = 1,000$ reported in Table 5.3, a critical value of BHKS test decreases from 1.48 to 1.34 (which corresponds to testing at significance level $\alpha = 10.72\%$). Critical values for our tests increase to 1.81 (corresponding to $\alpha = 3.52\%$) and 1.71 (corresponding to $\alpha = 4.31\%$) for local Whittle and GPH methods, respectively. Table 5.5 shows size adjusted power of the tests against LRD series. Still, our proposed tests are much more powerful compared to BHKS test. For example, when $d = .4$, the power is more than 74.9 %, compared to 19.3% for BHKS test.

5.6 Application to several real data sets

We first consider squared daily log returns

$$r_t = \left(100\left(\log P_t - \log P_{t-1}\right)\right)^2,$$

where P_t is the daily closing index of S&P 500 from the period of Jan 2, 1990 to Dec 31, 1999 (2,528 observations). Figure 5.3 shows a time series plot of squared daily log returns (top). Observe first that the sample autocorrelation function decays slowly (bottom left). SS parameter estimates, for example local Whittle estimates (bottom right), stay clearly away from 1/2 but less than 1 over a wide range of lower frequencies considered.

Here are the results from applying our proposed tests to S&P 500 squared daily log returns. Under 5% significance level, our sequential testing procedure found CM model with two breaks

$$\widehat{r}_t = .9080 - .4737 \times 1_{\{504<t\leq 1829\}} + .0443 \times 1_{\{1829<t\leq 2528\}},$$

where the break points correspond to Dec 30, 1991 and March 27, 1997, respectively, and p-values for CM model with two breaks are .1296 for local Whittle method and .7548 for GPH

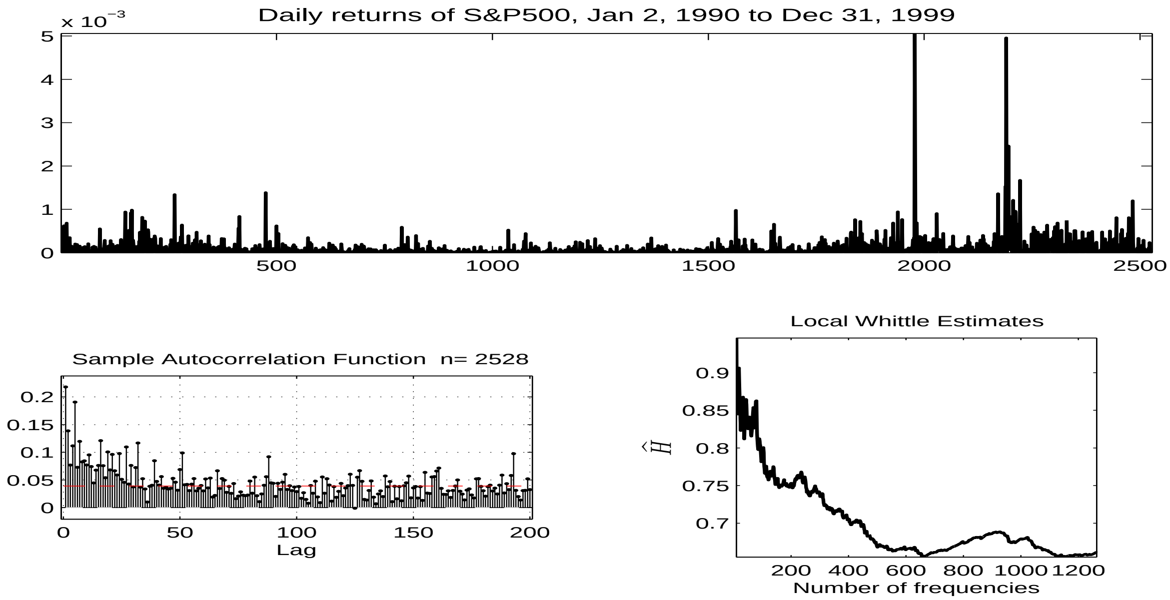

Figure 5.3: Time plots of S&P 500 squared returns from Jan 2, 1990 to Dec 31, 1999 (top) with sample autocorrelation plot (left bottom) and local Whittle estimates of SS parameter H (right bottom).

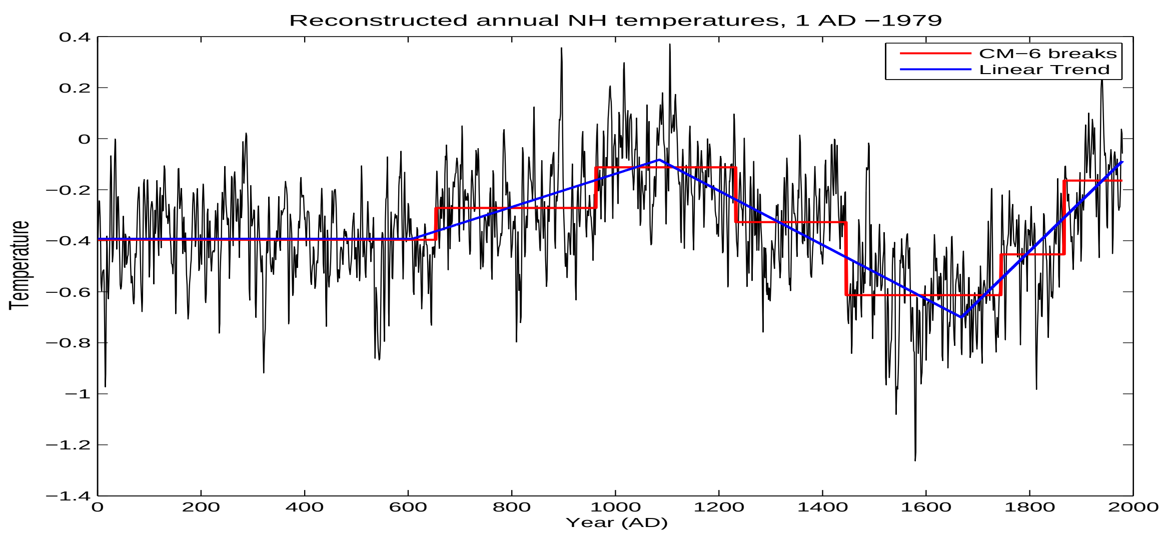

Figure 5.4: Northern Hemisphere temperature data with CM with 6 breaks (red) and piecewise linear model (blue) from Mills (2007).

method. For comparison when testing for CM model with two breaks, BHKS test statistic is $M_3 = 1.5073$ which corresponds to p-value of .0624. Based on our result, CM model with two breaks is preferred over LRD model for the squared daily log return of S&P 500 in 1990s.

We have also considered the historical Northern Hemisphere temperature data reconstructed by Moberg, Sonechkin, Holmgren, Datsenko and Karlen (2005), dating from 1 AD to 1979.

Applying our tests give CM model with 6 breaks with p-values 0.8235 and 0.8000 for local Whittle and GPH, respectively. In comparison, BHKS method finds CM model with 2 breaks with p-value .0819.

One important point to emphasize here is that our tests find a larger number of breaks in the temperature data. In fact, this result is more consistent with LRD series, rather than excludes such series. For example, we have generated FARIMA$(2, d, 2)$ series with the same parameters as used in Mills (2007) to fit a FARIMA model to the data. Out of 1,000 replications, the quartiles for the number of breaks found in the series were, respectively, $Q_1 = 5$, $Q_2 = 6$ and $Q_3 = 8$. Note also that the conclusion based on the BHKS test would be different - since only 2 breaks are found in the series, one would likely prefer to work with such simpler CM model. In case our tests also pointed to a small number of breaks, we would also recommend using such CM models.

On the other hand, our estimated CM model with 6 breaks is surprisingly similar to piecewise linear trend model with exogenously selected breaks at years 609, 1085 and 1668 (blue line in Figure 5.4) which is argued by Mills (2007) to eliminate LRD. The SS parameter estimates from the residuals obtained by subtracting the linear trend function are 0.4972 and 0.4167, and the corresponding p-values are 0.5147 and .8056 for local Whittle and GPH methods when testing for $H = 1/2$. This motivates extending our testing framework to piecewise linear trend model. This problem will be addressed in a future work.

5.7 Proofs

5.7.1 Proofs of Theorems 5.3.1 and 5.3.2

Proof of Theorem 5.3.1: We only consider the case $\widehat{k} \leq k^*$. Note that

$$R_j = \begin{cases} \mu + \epsilon_j - \overline{X}_1, & j \leq \widehat{k}, \\ \mu + \epsilon_j - \overline{X}_2, & \widehat{k} < j \leq k^* \\ \mu + \Delta + \epsilon_j - \overline{X}_2, & k^* < j, \end{cases}$$

where $\overline{X}_1 = 1/\widehat{k}\sum_{s=1}^{\widehat{k}} X_s$ and $\overline{X}_2 = 1/(n-\widehat{k})\sum_{s=\widehat{k}+1}^{n} X_s$. Then, since $\sum_{j=1}^{n} e^{-ij\omega_l} = 0$ if $\omega_l \neq 0$,

$$
\begin{aligned}
I_R(\omega_l) &= \frac{1}{2\pi n}\left|\sum_{j=1}^{\widehat{k}}(\mu+\epsilon_j-\overline{X}_1)e^{-ij\omega_l} + \sum_{j=\widehat{k}+1}^{k^*}(\mu+\epsilon_j-\overline{X}_2)e^{-ij\omega_l}\right.\\
&\qquad\left. + \sum_{j=k^*+1}^{n}(\mu+\Delta+\epsilon_j-\overline{X}_2)e^{-ij\omega_l}\right|^2\\
&= \frac{1}{2\pi n}\left|\sum_{j=1}^{n}\epsilon_j e^{-ij\omega_l} + (\overline{X}_2-\overline{X}_1-\Delta)\sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} - \sum_{j=\widehat{k}+1}^{k^*}\Delta e^{-ij\omega_l}\right|^2\\
&= \frac{1}{2\pi n}\left|\sum_{j=1}^{n}\epsilon_j e^{-ij\omega_l} + \left(\frac{1}{n-\widehat{k}}\sum_{j=1}^{n}\epsilon_j - \frac{n}{(n-\widehat{k})\widehat{k}}\sum_{j=1}^{\widehat{k}}\epsilon_j\right)\sum_{j=1}^{\widehat{k}} e^{-ij\omega_l}\right.\\
&\qquad\left. + \frac{\widehat{k}-k}{n-\widehat{k}}\Delta\sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} - \sum_{j=\widehat{k}+1}^{k^*}\Delta e^{-ij\omega_l}\right|^2 =: |x_1+x_2+x_3+x_4|^2,
\end{aligned}
\tag{5.64}
$$

where we used the fact that

$$
\overline{X}_2 - \overline{X}_1 - \Delta = \frac{\widehat{k}-k^*}{n-\widehat{k}}\Delta + \frac{1}{n-\widehat{k}}\sum_{j=1}^{n}\epsilon_j - \frac{n}{(n-\widehat{k})\widehat{k}}\sum_{j=1}^{\widehat{k}}\epsilon_j.
$$

For the term x_2, observe that, by the assumptions (5.20) and (A1)-(A3),

$$
\left(\frac{\widehat{k}}{n}, \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon_j, \frac{1}{\sqrt{n}}\sum_{j=1}^{\widehat{k}}\epsilon_j\right) \xrightarrow{d} (\theta, \sigma B(1), \sigma B(\theta)).
$$

Hence,

$$
x_2 = O_p(1)\frac{1}{n}\sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} = O_p(1)\frac{e^{-i(\widehat{k}+1)\omega_l} - e^{-i\omega_l}}{n(e^{-i\omega_l}-1)} = O_p\left(\frac{1}{l}\right).
$$

Similarly, by using assumptions (A1)-(A3),

$$
x_3 = O_p\left(\frac{1}{l\sqrt{n\Delta^2}}\right), \quad x_4 = O_p\left(\frac{(k^*-\widehat{k})\Delta}{\sqrt{n}}\right) = O_p\left(\frac{1}{\sqrt{n\Delta^2}}\right).
$$

It follows that

$$I_R(\omega_l) = I_\epsilon(\omega_l) + x_1 O_p\left(\frac{1}{l} + \frac{1}{\sqrt{n\Delta^2}}\right) + \overline{x}_1 O_p\left(\frac{1}{l} + \frac{1}{\sqrt{n\Delta^2}}\right) + O_p\left(\frac{1}{l^2} + \frac{1}{n\Delta^2}\right), \tag{5.65}$$

where

$$I_\epsilon(\omega_l) := |x_1|^2 = \left|\frac{1}{\sqrt{2\pi n}}\sum_{j=1}^{n} \epsilon_j e^{-ij\omega_l}\right|^2.$$

We shall rely next on the proofs of Theorems 1 and 2 in Robinson (1995a). An interested reader should therefore get familiar first with those proofs. As in Robinson (1995a), we first need to show consistency of $\widehat{H}_{lw}$. Since $H_0 = 1/2$ and by proceeding as in the proof of Theorem 1 in Robinson (1995a), it is enough to show that

$$\sup_{H\in\Theta_1} |A(H)| \xrightarrow{p} 0, \tag{5.66}$$

where

$$\Theta_1 = \{H : \Delta_1 < H < \Delta_2\}, \quad 0 < \Delta_1 < 1/2 < \Delta_2 < 1,$$
$$A(H) = \frac{2(H-1/2)+1}{m}\sum_{l=1}^{m}\left(\frac{l}{m}\right)^{2(H-1/2)}\left(\frac{I_R(\omega_l)}{G_0} - 1\right),$$

where $G_0 > 0$ is a constant such that the spectral density of ϵ satisfies $f_\epsilon(\omega) \sim G_0$ as $\omega \to 0$.

By using (5.65) and the fact that $\sup_{1\le l\le m} E|x_1| \le \infty$ (Theorem 2 in Robinson (1995b); to be more precise, as for Eq. (3.16) in Robinson (1995a), it is necessary to look into the proof of that Theorem 2),

$$\left(\frac{I_R(\omega_l)}{G_0} - 1\right) = \left(\frac{I_\epsilon(\omega_l)}{G_0} - 1\right) + O_p\left(\frac{1}{l} + \frac{1}{\sqrt{n\Delta^2}}\right) \tag{5.67}$$

Hence, (5.66) follows since

$$\sup_{H\in\Theta_1}\left|\frac{2(H-1/2)+1}{m}\sum_{l=1}^{m}\left(\frac{l}{m}\right)^{2(H-1/2)}\left(\frac{I_\epsilon(\omega_l)}{G_0} - 1\right)\right| \xrightarrow{p} 0$$

as proved in the proof of Theorem 1 in Robinson (1995a), and since

$$\frac{1}{m}\sum_{l=1}^{m}\left(\frac{l}{m}\right)^{2(H-1/2)} \to \int_0^1 x^{2(H-1/2)}dx$$

uniformly over Θ_1 by Lemma 1 in Robinson (1995a) and $1/l \to \infty$, as $l \to \infty$, and $n\Delta^2 \to \infty$.

For the asymptotic normality, the consistency of $\widehat{H}_{lw}$ implies that with probability approaching 1, as $n \to \infty$, $\widehat{H}_{lw}$ satisfies

$$0 = \frac{dR(\widehat{H}_{lw})}{dH} = \frac{dR(1/2)}{dH} + \frac{d^2R(\widetilde{H})}{dH^2}(\widehat{H}_{lw} - 1/2), \tag{5.68}$$

where $|\widetilde{H} - 1/2| \leq |\widehat{H}_{lw} - 1/2|$. Rewriting (5.68) gives

$$\sqrt{m}\left(\widehat{H}_{lw} - \frac{1}{2}\right) = -\sqrt{m}\frac{dR(1/2)/dH}{d^2R(\widetilde{H})/dH^2},$$

and the desired asymptotic normality follows if

$$\frac{d^2R(\widetilde{H})}{dH^2} \xrightarrow{p} 4, \qquad \sqrt{m}\frac{dR(1/2)}{dH} \xrightarrow{d} \mathcal{N}(0,4). \tag{5.69}$$

To show (5.69), we rely on the proof of Theorem 2 in Robinson (1995a). As in that proof (see Eq. (4.8) therein), we need to consider

$$\sum_{l=1}^{r}\left(\frac{I_R(\omega_l)}{G_0} - 2\pi I_\varepsilon(\omega_l)\right),$$

where ε appears in Assumption A3 of Robinson (1995a) through the representation $\epsilon_k = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{k-j}$. By using (5.67), this term is

$$\sum_{l=1}^{r}\left(\frac{I_\epsilon(\omega_l)}{G_0} - 2\pi I_\varepsilon(\omega_l)\right) + O_p\left(\log r + \frac{r}{\sqrt{n\Delta^2}}\right). \tag{5.70}$$

The first term in (5.70) is exactly that in Eq. (4.8) of Robinson (1995a), and is dealt in the proof of Theorem 2 of Robinson (1995a). We only need to make sure that the second term in (5.70) can be ignored in that proof. In this regard, it is enough to show that, for arbitrarily small $\delta > 0$,

$$\sum_{r=1}^{m} \left(\frac{r}{m}\right)^{1-2\delta} \frac{1}{r^2} \left(\log r + \frac{r}{\sqrt{n\Delta^2}}\right) + \frac{1}{m} \left(\log m + \frac{m}{\sqrt{n\Delta^2}}\right) = O(m^{-1/2})$$

(see the discussion around Eqs. (4.7)-(4.9) in Robinson (1995a)), and

$$m^{-1/2} \sum_{j=1}^{m} |\nu_j| \left(\frac{1}{j} + \frac{1}{\sqrt{n\Delta^2}}\right) = o(1),$$

where $\nu_j = \log j - m^{-1} \sum_{j=1}^{m} \log j$ (see the discussion around Eq. (4.11) in Robinson (1995a)). Both of these relations can be easily seen by using the assumption (5.45). $\square$

Proof of Theorem 5.3.2: Proceeding as in the proof of Theorem 1 in Robinson (1995a), it is enough to show that

$$\sup_{H \in \Theta_1} |A(H)| \xrightarrow{p} 0, \tag{5.71}$$

$$P\left(\left|\frac{1}{m} \sum_{l=1}^{m} (a_l - 1) \left(\frac{I_R(\omega_l)}{g_l} - 1\right)\right| \geq 1\right) \to 0. \tag{5.72}$$

Here,

$$A(H) = \frac{2(H - H_0) + 1}{m} \sum_{l=1}^{m} \left(\frac{l}{m}\right)^{2(H-H_0)} \left(\frac{I_R(\omega_l)}{g_l} - 1\right)$$

is defined in Eq. (3.10) of Robinson (1995a) and $g_l = G_0 \omega_l^{1-2H_0}$ with $G_0 > 0$ appearing in the spectral density of X, $f_X(\omega) \sim G_0 \omega^{1-2H_0}$ as $\omega \to 0$. The set Θ_1 in (5.71) is defined as $\Theta_1 = \{H : \Delta \leq H \leq \Delta_2\}$, where $\Delta = \Delta_1$ if $H_0 - 1/2 < \Delta_1$, and $H_0 - \frac{1}{2} < \Delta \leq H_0$ otherwise. In (5.72), the sequence $\{a_l\}$ is

$$a_l = \begin{cases} \left(\frac{l}{p}\right)^{2(\Delta - H_0)}, & \text{if } 1 \leq l \leq p := \exp(m^{-1} \sum_{j=1}^{m} \log j), \\ \left(\frac{l}{p}\right)^{2(\Delta_1 - H_0)}, & \text{if } p < l \leq m \end{cases}$$

(see the bottom of p. 1638 in Robinson (1995a)).

Observe now that

$$I_R(\omega_l) = \frac{1}{2\pi n} \left| \sum_{j=1}^{n} e^{-ij\omega_l} X_j + (\overline{X}_2 - \overline{X}_1) \sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} \right|^2.$$

$$= I_X(\omega_l) + w_l \frac{\overline{X}_2 - \overline{X}_1}{\sqrt{2\pi n}} \sum_{j=1}^{\widehat{k}} e^{ij\omega_l} + \overline{w}_l \frac{\overline{X}_2 - \overline{X}_1}{\sqrt{2\pi n}} \sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} + \frac{(\overline{X}_2 - \overline{X}_1)^2}{2\pi n} \left| \sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} \right|^2,$$

where $\overline{X}_1$, $\overline{X}_2$ are as in the proof of Theorem 5.3.1 above and

$$I_X(\omega_l) := |v(\omega_l)|^2 = \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X_j e^{-ij\omega_l} \right|^2.$$

Note that

$$\frac{1}{n} \sum_{l=1}^{\widehat{k}} e^{-ij\omega_l} = O\left(\frac{1}{l}\right). \tag{5.73}$$

Moreover, from (5.47), we know that

$$n^{1-H_0}(\overline{X}_2 - \overline{X}_1) \xrightarrow{d} \frac{1}{1-\xi} \left(B_{H_0}(1) - \frac{B_{H_0}(\xi)}{\xi} \right),$$

where ξ is defined by (5.29) with $H = H_0$, and hence that

$$\overline{X}_2 - \overline{X}_1 = O_p(n^{H_0 - 1}). \tag{5.74}$$

Furthermore, from the proof of Theorem 2 in Robinson (1995*b*),

$$\sup_{1 \le l \le m} \frac{E|v(\omega_l)|}{g_l^{1/2}} \le \sup_{1 \le l \le m} \left(\frac{E|v(\omega_l)|^2}{g_l} \right)^{1/2} \le C, \tag{5.75}$$

for some constant C.

To establish (5.71), and by using (5.73)-(5.75), it is enough to show that

$$\sup_{H \in \Theta_1} |A_i(H)| \xrightarrow{p} 0, \quad i = 1, 2, 3, \tag{5.76}$$

where

$$A_1(H) = \frac{2(H - H_0) + 1}{m} \sum_{l=1}^{m} \left(\frac{l}{m} \right)^{2(H-H_0)} \left(\frac{I_X(\omega_l)}{g_l} - 1 \right),$$

$$A_2(H) = \frac{1}{m} \sum_{l=1}^{m} \left(\frac{l}{m} \right)^{2(H-H_0)} \frac{1}{l^{3/2-H_0}}, \quad A_3(H) = \frac{1}{m} \sum_{l=1}^{m} \left(\frac{l}{m} \right)^{2(H-H_0)} \frac{1}{l^{3-2H_0}}.$$

The convergence (5.76) with $i = 1$ is shown in the proof of Theorem 1 in Robinson (1995*a*).

The convergence for $i = 3$ follows from that for $i = 2$ since $3/2 - H_0 > 0$. When $i = 2$, the convergence (5.76) follows from

$$\frac{1}{m}\sum_{l=1}^{m}\left(\frac{l}{m}\right)^{2(H-H_0)} \to \int_0^1 x^{2(H-H_0)}dx$$

uniformly over $H \in \Theta_1$ by Lemma 1 in Robinson (1995a), and $1/l^{3/2-H_0} \to 0$, as $l \to \infty$.

Arguing similarly as for (5.71), relation (5.72) is established by showing that

$$P\left(\left|\frac{1}{m}\sum_{l=1}^{m}(a_l - 1)\left(\frac{I_X(\omega_l)}{g_l} - 1\right)\right| \geq 1\right) \to 0 \tag{5.77}$$

and

$$\frac{1}{m}\sum_{l=1}^{m}|a_l - 1|\frac{1}{l^{3/2-H_0}} \to 0, \quad \frac{1}{m}\sum_{l=1}^{m}|a_l - 1|\frac{1}{l^{3-2H_0}} \to 0. \tag{5.78}$$

Convergence (5.77) is shown in the proof of Theorem 1 in Robinson (1995a). The relation (5.78) follows since

$$\frac{1}{m}\sum_{l=1}^{m} a_l = O(1)$$

from Eqs. (3.22) and (3.23) in Robinson (1995a) and $1/l^{3/2-H_0} \to 0$, as $l \to \infty$ and

$$\frac{1}{m}\sum_{l=1}^{m}\frac{1}{l^{3/2-H_0}} = O\left(\frac{m^{H_0-1/2}}{m}\right) = o(1). \qquad \square$$

5.7.2 Proof of Theorem 5.3.3

Proof of Theorem 5.3.3: We only consider the case $\widehat{k} \leq k^*$. As in the proof of Theorem 5.3.1 (see (5.64)), we have

$$I_R(\omega_l) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n}\epsilon_j e^{-ij\omega_l} + \left(\frac{1}{n-\widehat{k}}\sum_{j=1}^{n}\epsilon_j - \frac{n}{(n-\widehat{k})\widehat{k}}\sum_{j=1}^{\widehat{k}}\epsilon_j\right)\sum_{j=1}^{\widehat{k}} e^{-ij\omega_l}\right.$$

$$\left. + \frac{\widehat{k}-k}{n-\widehat{k}}\Delta\sum_{j=1}^{\widehat{k}} e^{-ij\omega_l} - \sum_{j=\widehat{k}+1}^{k^*}\Delta e^{-ij\omega_l}\right|^2 =: |x_1 + x_2 + x_3 + x_4|^2.$$

For x_1, by using assumption (5.51) and Theorem 3.2 of Phillips (2007) (which only uses our assumption (5.51) through Lemma 3.1 therein), note that

$$\sup_{1\leq l\leq m}\left|\frac{1}{\sqrt{2\pi n}}\sum_{j=1}^{n}\epsilon_j e^{-ij\omega_l}-\xi_l\right|=o_p\left(\frac{m}{n^{1/2-\delta}}\right), \tag{5.79}$$

where

$$\xi_l=\frac{1}{2\pi}\int_0^1 e^{2\pi ilr}dB(r) \tag{5.80}$$

are i.i.d. complex normal random variables. Therefore,

$$x_1=\frac{1}{\sqrt{2\pi n}}\sum_{j=1}^{n}\epsilon_j e^{-ij\omega_l}=\xi_l+o_p\left(\frac{m}{n^{1/2-\delta}}\right), \tag{5.81}$$

with the error $o_p(\cdot)$ being uniform in $l=1,\ldots,m$. For x_2, by using assumptions (5.51), (A1)-(A3) and $(1/2-\delta_0)-$Hölder continuity of Brownian motion, we get

$$x_2=\frac{1}{\sqrt{2\pi}}\left(\frac{B(1)}{1-\theta}-\frac{B(\theta)}{(1-\theta)\theta}\right)\frac{1}{n}\sum_{j=1}^{k^*}e^{-ij\omega_l}+O_p\left(\frac{1}{l(n\Delta^2)^{1/2-\delta_0}}\right)+o_p\left(\frac{1}{ln^{1/2-\delta}}\right) \tag{5.82}$$

for some arbitrarily small $\delta_0>0$, where the error terms $O_p(\cdot)$, $o_p(\cdot)$ are uniform in $l=1,\ldots,m$. Also, as in the proof of Theorem 5.3.1, by using assumptions (A1)-(A3),

$$x_3=O_p\left(\frac{1}{l\sqrt{n\Delta^2}}\right),\quad x_4=O_p\left(\frac{1}{\sqrt{n\Delta^2}}\right), \tag{5.83}$$

uniformly in $l=1,\ldots,m$.

Combining (5.81), (5.82) and (5.83) yields that

$$I_R(\omega_l)=\left|\xi_l+\frac{1}{\sqrt{2\pi}}\left(\frac{B(1)}{1-\theta}-\frac{B(\theta)}{(1-\theta)\theta}\right)\frac{1}{n}\sum_{j=1}^{k^*}e^{-ij\omega_l}+A_l\right|^2=:|\xi_l+\eta_l+A_l|^2, \tag{5.84}$$

where

$$\sup_{1\leq l\leq m}|A_l|=o_p\left(\frac{m}{n^{1/2-\delta}}\right)+O_p\left(\frac{1}{(n\Delta^2)^{1/2-\delta_0}}\right). \tag{5.85}$$

By using (5.84), our test statistic in (5.50) can be expressed as

$$M_{gph} = \sqrt{m}\left(\widehat{H}_{gph} - \frac{1}{2}\right) = \sqrt{m}\sum_{l=1}^{m} a_l \log|\xi_l + \eta_l + A_l|^2$$

$$= \sqrt{m}\sum_{l=1}^{m} a_l \log|\xi_l|^2 + \sqrt{m}\sum_{l=1}^{m} a_l \log\left|1 + \frac{\eta_l}{\xi_l}\right|^2 + \sqrt{m}\sum_{l=1}^{m} a_l \log\left|1 + \frac{A_l}{\xi_l + \eta_l}\right|^2 =: S_1 + S_2 + S_3. \tag{5.86}$$

The proof will follow by showing that $S_1 \xrightarrow{d} \mathcal{N}(0, \pi^2/24)$ and $S_k = o_p(1)$, $k = 2, 3$.

For the term S_1, observe first that

$$|\xi_l|^2 = \Re(\xi_l)^2 + \Im(\xi_1)^2$$

where $\Re(\xi_l)$ and $\Im(\xi_l)$ are independent $\mathcal{N}(0, \sigma^2/4\pi)$ variables. This implies that $|\xi_l|^2$ is distributed as $\sigma^2\chi^2(2)/4\pi$. Since $\sum_{l=1}^{m} a_l = 0$, rewrite S_1 as

$$\sqrt{m}\sum_{l=1}^{m} a_l \left(\log|\xi_l|^2 - E(\log\sigma^2\chi^2(2)/4\pi)\right).$$

Then, the asymptotic normality of S_1 is deduced from the Lindeberg-Feller Central Limit Theorem (see Robinson (1995*b*), p. 1070). The asymptotic variance is

$$\mathrm{Var}(\log|\xi_l|^2)\lim_{m\to\infty} m\sum_{l=1}^{m} a_l^2 = \frac{\pi^2}{24}$$

since

$$\mathrm{Var}(\log|\xi_l|^2) = \mathrm{Var}\left(\log\frac{\sigma^2}{4\pi}\chi^2(2)\right) = \mathrm{Var}(\log\chi^2(2)) = \frac{\pi^2}{6}$$

(see, for example, Johnson, Kotz and Balakrishnan (1995)) and

$$\lim_{m\to\infty} m\sum_{l=1}^{m} a_l^2 = \frac{1}{4}$$

from the approximations

$$(z_l - \bar{z}) = -2(\log(l/m) + 1 + o(1)), \quad \sum_{l=1}^{m}(z_l - \bar{z}_l)^2 = 4m(1 + o(1)) \tag{5.87}$$

(see Hurvich et al. (1998), p. 38).

For the term S_2 in (5.86), by using the inequality

$$|\log|1+a|| \le |a| + \frac{|a|}{|1+a|}, \tag{5.88}$$

we have

$$\begin{aligned}|S_2| &\le 2\sqrt{m}\sum_{l=1}^{m}|a_l|\left(\left|\frac{\eta_l}{\xi_l}\right| + \frac{\left|\frac{\eta_l}{\xi_l}\right|}{\left|1+\frac{\eta_l}{\xi_l}\right|}\right)\\ &\le 2\sqrt{m}\sum_{l=1}^{m}|a_l|\left|\frac{\eta_l}{\xi_l}\right| + 2\sqrt{m}\sum_{l=1}^{m}|a_l|\frac{|\eta_l|}{|\xi_l+\eta_l|} =: S_{2,1}+S_{2,2}.\end{aligned}$$

We will show that $S_{2,1} = o_p(1)$ and $S_{2,2} = o_p(1)$. Since $\left|1/n\sum_{j=1}^{k^*} e^{-ij\omega_l}\right| = O(1/l)$ and hence $|\eta_l| = O_p(1/l)$, and since $a_l = -(1+\log(l/m)+o(1))/2m(1+o(1))$ from (5.87), it is enough to show that $s_{2,1} = o_p(1)$ and $s_{2,2} = o_p(1)$, where

$$s_{2,1} = \sum_{l=1}^{m}\frac{|1+\log(l/m)|}{l\sqrt{m}}\frac{1}{|\xi_l|}, \quad s_{2,2} = \sum_{l=1}^{m}\frac{|1+\log(l/m)|}{l\sqrt{m}}\frac{1}{|\xi_l+\eta_l|}.$$

The result for $s_{2,1}$ follows from Lemma C in Phillips (2007), p. 113, and that for $s_{2,2}$ follows from Lemma 5.7.1 below.

Consider now the term S_3 in (5.86). Observe that

$$|S_3| \le 2\sqrt{m}\sum_{l=1}^{m}|a_l|\left|\log\left|1+\frac{A_l}{\xi_l+\eta_l}\right|\right| \le 2\sqrt{m}\sum_{l=1}^{m}|a_l|\frac{|A_l|}{|\xi_l+\eta_l|}B_l,$$

where

$$B_l = 1 + \frac{1}{\left|1+\frac{A_l}{\xi_l+\eta_l}\right|}.$$

First, we will show that

$$\sup_{1\le l\le m} B_l = O_p(1). \tag{5.89}$$

This follows from

$$\sup_{1\le l\le m}\frac{|A_l|}{|\xi_l+\eta_l|} \xrightarrow{p} 0.$$

To show the latter convergence, note that

$$\sup_{1\le l\le m}\frac{|A_l|}{|\xi_l+\eta_l|}\le \sup_{1\le l\le m}|A_l|\sum_{l=1}^{m}\frac{1}{|\xi_l+\eta_l|}$$

and

$$\sum_{l=1}^{m}\frac{1}{|\xi_l+\eta_l|}=O_p(m\log m)$$

by using Lemma 5.7.1 below. Therefore, together with (5.85),

$$\sup_{1\le l\le m}\frac{|A_l|}{|\xi_l+\eta_l|}=o_p\left(\frac{m^2\log m}{n^{1/2-\delta}}\right)+O_p\left(\frac{m\log m}{(n\Delta^2)^{1/2-\delta_0}}\right)\xrightarrow{p}0$$

by the assumption (5.52).

By using (5.89), (5.85) and (5.87), $S_3=o_p(1)$ if

$$\left(\frac{m}{n^{1/2-\delta}}+\frac{1}{(n\Delta^2)^{1/2-\delta_0}}\right)\frac{1}{\sqrt{m}}\sum_{l=1}^{m}\frac{|1+\log(l/m)|}{|\xi_l+\eta_l|}\xrightarrow{p}0.$$

Note that

$$\sum_{l=1}^{m}\frac{|1+\log(l/m)|}{|\xi_l+\eta_l|}=O_p(m\log m)$$

from Lemma 5.7.1 below. The convergence above then follows from

$$\left(\frac{m}{n^{1/2-\delta}}+\frac{1}{(n\Delta^2)^{1/2-\delta_0}}\right)\frac{m\log m}{\sqrt{m}}\to 0.$$ □

The following lemma was used several times in the proof above. It extends Lemma C of Phillips (2007) to a particular collection of *dependent* random variables.

Lemma 5.7.1. *Let ξ_l and η_l be random variables in (5.80) and (5.84), respectively. Suppose that a sequence of real numbers $\{y_{l,m}, m\ge 1, l=1,\dots,m\}$ satisfies*

$$\text{(C1)}\ \frac{1}{m}\sum_{l=1}^{m}|y_{l,m}|\to y<\infty,\quad \text{(C2)}\ \frac{1}{m\log m}\sum_{l=1}^{m}|y_{l,m}||\log|y_{l,m}||\to 0,\quad \text{(C3)}\ \sup_{l\le m}\frac{|y_{l,m}|}{m}\to 0.$$

Then,

$$\sum_{l=1}^{m}\frac{|y_{l,m}|}{|\xi_l+\eta_l|}=O_p(m\log m).$$

Proof: First note that
$$\sum_{l=1}^{m} \frac{|y_{l,m}|}{|\xi_l+\eta_l|} \leq \sum_{l=1}^{m} \frac{|y_{l,m}|}{|\Re(\xi_l)+\Re(\eta_l)|},$$
where
$$\Re(\xi_l) = \frac{1}{\sqrt{2\pi}} \int_0^1 \cos(2\pi l r) dB(r) =: \alpha_l,$$
$$\Re(\eta_l) = \frac{1}{\sqrt{2\pi}} \left(B(1) - \frac{B(\theta)}{\theta} \right) \frac{1}{1-\theta} \Re \left(\frac{1}{n} \sum_{j=1}^{k^*} e^{-ij\omega_l} \right) =: \beta k_l$$
with
$$\beta = \frac{1}{\sqrt{2\pi}} \left(B(1) - \frac{B(\theta)}{\theta} \right) = \frac{1}{\sqrt{2\pi}} \int_0^1 \left(1 - \frac{1_{[0,\theta]}(r)}{\theta} \right) dB(r).$$
Hence, it is sufficient to show that
$$\frac{1}{m \log m} \sum_{l=1}^{m} \frac{|y_{l,m}|}{|\alpha_l + \beta k_l|} \xrightarrow{p} b < \infty. \tag{5.90}$$

We follow the approach of the proof of Lemma C in Phillips (2007). Let
$$\gamma_l = \frac{|y_{l,m}|}{m \log m |\alpha_l + \beta k_l|},$$
$$\gamma_l' = \gamma_l 1_{\{\gamma_l < 1\}}, \quad \gamma_l'' = \gamma_l 1_{\{\gamma_1 \geq 1\}}, \quad b_l = E\gamma_l'.$$
Then, (5.90) follows by proving that
$$\text{(S1)} \ \sum_{l=1}^{m} b_l \to b < \infty, \quad \text{(S2)} \ \sum_{l=1}^{m} \gamma_l'' \xrightarrow{p} 0, \quad \text{(S3)} \ \sum_{l=1}^{m} (\gamma_l' - b_l) \xrightarrow{p} 0,$$
as $m \to \infty$. For later reference, note that
$$E\alpha_l^2 = \frac{\sigma^2}{2\pi} \int_0^1 \cos^2(2\pi l r) dr = \frac{\sigma^2}{4\pi}, \quad E\alpha_l \alpha_{l'} = 0, \quad l \neq l', \tag{5.91}$$
$$E\beta^2 = \frac{\sigma^2}{2\pi} \int_0^1 \left(1 - \frac{1_{[0,\theta]}(r)}{\theta} \right)^2 dr = \frac{\sigma^2}{2\pi} \left(1 - \frac{2\theta}{\theta} + \frac{\theta}{\theta^2} \right) = \frac{\sigma^2}{2\pi} \left(\frac{1}{\theta} - 1 \right), \tag{5.92}$$
$$E\beta\alpha_l = \frac{\sigma^2}{2\pi} \int_0^1 \left(\cos(2\pi l r) - \frac{1_{[0,\theta]}(r)}{\theta} \cos(2\pi l r) \right) dr = -\frac{\sigma^2}{2\pi} \frac{\sin 2\pi l \theta}{2\pi l \theta}. \tag{5.93}$$

For condition (S1), note that $\alpha_l + \beta k_l$ is a zero mean, Gaussian variable with standard deviation

$$\sqrt{E(\alpha_l + \beta k_l)^2} = \sqrt{\frac{\sigma^2}{4\pi} - 2k_l \frac{\sigma^2}{2\pi}\frac{\sin 2\pi l\theta}{2\pi l\theta} + \frac{\sigma^2}{2\pi}\left(\frac{1}{\theta} - 1\right)k_l^2} =: \sqrt{\frac{\sigma^2}{4\pi}\frac{1}{|x_{l,m}|}},$$

where we used (5.91)-(5.93). Then, (S1) follows as in (i) of the proof of Lemma C in Phillips (2007) as long as $1/m\sum_{l=1}^m |x_{l,m}y_{l,m}| \to a < \infty$, $1/(m\log m)\sum_{l=1}^m |x_{l,m}y_{l,m}|\,|\log|x_{l,m}y_{l,m}|| \to 0$ and $\sup_{l\leq m}|x_{l,m}y_{l,m}|/m \to 0$ are satisfied. Observe for $|x_{l,m}|$ that $|x_{l,m}| \to 1$ as $l \to \infty$ since $k_l = O(1/l)$. Hence, together with assumptions (C1)-(C3), the three conditions of Phillips (2007) above are satisfied. This shows (S1). Similarly, (S2) follows directly as in (ii) in the proof of Lemma C of Phillips (2007).

We now turn to condition (S3). Observe that

$$P\left(\sum_{l=1}^m (\gamma_l' - b_l) > \epsilon\right) \leq \epsilon^{-2}E\left(\sum_{l=1}^m (\gamma_l' - b_l)\right)^2$$

$$\leq \epsilon^{-2}\sum_{l=1}^m (E\gamma_l'^2 - b_l^2) + \epsilon^{-2}\sum_{l\neq l'}|E(\gamma_l' - b_l)(\gamma_{l'}' - b_{l'}')| =: \epsilon^{-2}(T_1 + T_2).$$

The first term T_1 converges to zero as in (iii) in the proof of Lemma C of Phillips (2007). It is then enough to show that

$$T_2 = o(1). \tag{5.94}$$

Denote $\gamma_l^* = (m\log m)\gamma_l'$ so that

$$T_2 = \frac{1}{m^2\log^2 m}\sum_{l\neq l'}|E\gamma_l^*\gamma_{l'}^* - E\gamma_l^* E\gamma_{l'}^*|\,.$$

Let also

$$\sigma_l^2 = E(\alpha_l + \beta k_l)^2, \quad \sigma_{l,l'} = E(\alpha_l + \beta k_l)(\alpha_{l'} + \beta k_{l'}), \quad \rho_{l,l'} = \frac{\sigma_{l,l'}}{\sigma_l\sigma_{l'}}$$

and write

$$\widetilde{\sigma}_l^2 = \frac{\sigma_l^2}{|y_{l,m}|^2}, \quad \widetilde{\sigma}_{l,l'} = \frac{\sigma_{l,l'}}{|y_{l,m}||y_{l',m}|}.$$

Note also from moment calculations (5.91)-(5.93) that for sufficiently large l,

$$\sigma_l^2 \sim \frac{\sigma^2}{4\pi}, \quad \sigma_{l,l'} = O\left(\frac{1}{ll'}\right). \tag{5.95}$$

Then,

$$E\gamma_l^* = 2\int_{|y_{l,m}|/(m\log m)}^{\infty} \frac{|y_{l,m}|}{z}\frac{1}{\sqrt{2\pi\sigma_l^2}}e^{-\frac{z^2}{2\sigma_l^2}}\,dz = 2\int_{1/(m\log m)}^{\infty}\frac{1}{z}\frac{1}{\sqrt{2\pi\widetilde{\sigma}_l^2}}e^{-\frac{z^2}{2\widetilde{\sigma}_l^2}}\,dz$$

$$= 2(\log(m\log m))\frac{1}{\sqrt{2\pi\widetilde{\sigma}_l^2}}e^{-\frac{1}{2\widetilde{\sigma}_l^2 m^2\log^2 m}} + 2\int_{1/(m\log m)}^{\infty}\frac{z\log z}{\sqrt{2\pi}\widetilde{\sigma}_l^3}e^{-\frac{z^2}{2\widetilde{\sigma}_l^2}}\,dz =: c_l + c_l^{(1)}.$$

Now, for $c_l^{(1)}$, observe that

$$\left|c_l^{(1)}\right| \le 2\int_0^{\infty}\frac{|z\log z|}{\sqrt{2\pi}\widetilde{\sigma}_l^3}e^{-\frac{z^2}{2\widetilde{\sigma}_l^2}}\,dz \le \frac{2}{\sqrt{2\pi}}\int_0^{\infty}\frac{|w(\log\widetilde{\sigma}_l+\log w)|}{\widetilde{\sigma}_l}e^{-\frac{w^2}{2}}\,dw$$

$$\le C\frac{1+|\log\widetilde{\sigma}_l|}{\widetilde{\sigma}_l} \le C'\left(1+\big|\log|y_{l,m}|\big|\right)|y_{l,m}|$$

for some constants C and C' by using (5.95). We can ignore the terms $c_l^{(1)}$ in further calculations. By using (5.95), for example, the cross term of c_l and $c_{l'}^{(1)}$ becomes

$$\frac{1}{m^2\log^2 m}\sum_{l\neq l'}|c_l c_{l'}^{(1)}| \le \frac{C}{m^2\log^2 m}\sum_{l\neq l'}\log(m\log m)|y_{l,m}|\left(1+\big|\log|y_{l',m}|\big|\right)|y_{l',m}|$$

$$\le \frac{C\log(m\log m)}{\log^2 m}\left(\frac{1}{m}\sum_{l=1}^{m}|y_{l,m}|\right)\left(\frac{1}{m}\sum_{l'=1}^{m}|y_{l',m}|\right)$$

$$+\frac{C\log(m\log m)}{\log m}\left(\frac{1}{m}\sum_{l=1}^{m}|y_{l,m}|\right)\left(\frac{1}{m\log m}\sum_{l'=1}^{m}\big|\log|y_{l',m}|\big||y_{l',m}|\right) \to 0$$

by the assumptions (C1)-(C2).

Similar expansion for $E\gamma_l^*\gamma_{l'}^*$ using integration by parts for two dimensions together with (5.95) and the assumptions (C1)-(C3) give the leading term as

$$c_{l,l'} := 2\frac{\log^2(m\log m)}{2\pi\widetilde{\sigma}_l\widetilde{\sigma}_{l'}\sqrt{1-\rho_{l,l'}^2}}\exp\left(\frac{1}{2(1-\rho_{l,l'}^2)}\left(\frac{1}{\widetilde{\sigma}_l^2 m^2\log^2 m}+\frac{1}{\widetilde{\sigma}_{l'}^2 m^2\log^2 m}\right)\right)$$

$$\times \left\{ \exp\left(\frac{1}{2(1-\rho_{l,l'}^2)} \frac{2\rho_{l,l'}}{\widetilde{\sigma}_l \widetilde{\sigma}_{l'} m^2 \log^2 m} \right) + \exp\left(-\frac{1}{2(1-\rho_{l,l'}^2)} \frac{2\rho_{l,l'}}{\widetilde{\sigma}_l \widetilde{\sigma}_{l'} m^2 \log^2 m} \right) \right\}.$$

Now, we bound

$$|c_{l,l'} - c_l c_{l'}| \le 4 \frac{\log^2(m \log m)}{2\pi \widetilde{\sigma}_l \widetilde{\sigma}_{l'}} \left(\frac{1}{\sqrt{1-\rho_{l,l'}^2}} - 1 \right) \exp\left(-\frac{1}{2m^2 \log^2 m} \left(\frac{1}{\widetilde{\sigma}_l^2} + \frac{1}{\widetilde{\sigma}_{l'}^2} \right) \right)$$

$$+4 \frac{\log^2(m \log m)}{2\pi \widetilde{\sigma}_l \widetilde{\sigma}_{l'} \sqrt{1-\rho_{l,l'}^2}} \left| \exp\left(-\frac{1}{2m^2 \log^2 m} \left(\frac{1}{\widetilde{\sigma}_l^2} + \frac{1}{\widetilde{\sigma}_{l'}^2} \right) \right) \right.$$

$$\left. - \exp\left(-\frac{1}{2m^2 \log^2 m (1-\rho_{l,l'}^2)} \left(\frac{1}{\widetilde{\sigma}_l^2} + \frac{1}{\widetilde{\sigma}_{l'}^2} \right) \right) \right|$$

$$+2 \frac{\log^2(m \log m)}{2\pi \widetilde{\sigma}_l \widetilde{\sigma}_{l'} \sqrt{1-\rho_{l,l'}^2}} \exp\left(-\frac{1}{2m^2 \log^2 m (1-\rho_{l,l'}^2)} \left(\frac{1}{\widetilde{\sigma}_l^2} + \frac{1}{\widetilde{\sigma}_{l'}^2} \right) \right) \times$$

$$\left| \exp\left(\frac{1}{2(1-\rho_{l,l'}^2)} \frac{2\rho_{l,l'}}{\widetilde{\sigma}_l \widetilde{\sigma}_{l'} m^2 \log^2 m} \right) \exp\left(-\frac{1}{2(1-\rho_{l,l'}^2)} \frac{2\rho_{l,l'}}{\widetilde{\sigma}_l \widetilde{\sigma}_{l'} m^2 \log^2 m} \right) - 2 \right| =: U_1 + U_2 + U_3.$$

Relation (5.94) will be established by showing that

$$\frac{1}{m^2 \log^2 m} \sum_{l \neq l'} (U_1 + U_2 + U_3) \to 0. \tag{5.96}$$

For U_1, observe first that

$$U_1 \le C \frac{\log^2 m}{\widetilde{\sigma}_l \widetilde{\sigma}_{l'}} \frac{|\rho_{l,l'}|}{\sqrt{1-\rho_{l,l'}^2}}$$

by using the inequality $|a^{1/2} - b^{1/2}| \le |a-b|^{1/2}$, $a, b > 0$. From (5.95) and since $k_l = O(1/l)$, this gives that

$$\frac{1}{m^2 \log^2 m} \sum_{l \neq l'} U_1 \le \frac{C}{m^2 \log^2 m} \sum_{l \neq l'} \frac{\log^2 m |y_{l,m}||y_{l',m}|}{ll'} \le C \left(\frac{1}{m} \sum_{l=1}^{m} \frac{|y_{l,m}|}{l} \right)^2 \to 0 \tag{5.97}$$

by the assumptions (C1) and (C3).

For U_2, by using the inequality $|e^{-a} - e^{-b}| \le |a-b|$, $a, b > 0$, and by arguing as for U_1

above,

$$\left|\exp\left(-\frac{1}{2m^2\log^2 m}\left(\frac{1}{\widetilde{\sigma}_l^2}+\frac{1}{\widetilde{\sigma}_{l'}^2}\right)\right)-\exp\left(-\frac{1}{2m^2\log^2 m(1-\rho_{l,l'}^2)}\left(\frac{1}{\widetilde{\sigma}_l^2}+\frac{1}{\widetilde{\sigma}_{l'}^2}\right)\right)\right|$$

$$\leq\left|\frac{1}{2m^2\log^2 m}\left(\frac{1}{\widetilde{\sigma}_l^2}+\frac{1}{\widetilde{\sigma}_{l'}^2}\right)\left(\frac{1}{1-\rho_{l,l'}^2}-1\right)\right|=O\left(\frac{|y_{l,m}|^2+|y_{l',m}|^2}{(m^2\log^2 m)l^2l'^2}\right).$$

Therefore,

$$\frac{1}{m^2\log^2 m}\sum_{l\neq l'}U_2\leq\frac{C}{m^4\log^2 m}\sum_{l\neq l'}\frac{|y_{l,m}||y_{l',m}|(|y_{l,m}|^2+|y_{l',m}|^2)}{l^2l'^2}$$

$$\leq\frac{C}{\log^2 m}\left(\frac{\sup_{l\leq m}|y_{l,m}|}{m}\right)^4\sum_{l\neq l'}\frac{1}{l^2l'^2}\to 0 \tag{5.98}$$

by the assumption (C3).

For U_3, observe that $|e^x+e^{-x}-2|=o(|x|)$, as $x\to 0$, implies similarly that

$$\left|\exp\left(\frac{1}{2(1-\rho_{l,l'}^2)}\frac{2\rho_{l,l'}}{\widetilde{\sigma}_l\widetilde{\sigma}_{l'}m^2\log^2 m}\right)+\exp\left(-\frac{1}{2(1-\rho_{l,l'}^2)}\frac{2\rho_{l,l'}}{\widetilde{\sigma}_l\widetilde{\sigma}_{l'}m^2\log^2 m}\right)-2\right|$$

$$=o\left(\frac{|y_{l,m}||y_{l',m}|}{(m^2\log^2 m)ll'}\right).$$

This leads to

$$\frac{1}{m^2\log^2 m}\sum_{l\neq l'}U_3\leq\frac{C}{m^2\log^2 m}\sum_{l\neq l'}\frac{\log^2 m|y_{l,m}|^2|y_{l',m}|^2}{(m^2\log^2 m)ll'}$$

$$\leq\frac{C}{\log^2 m}\left(\frac{\sup_{l\leq m}|y_{l,m}|}{m}\right)^2\frac{1}{m^2}\sum_{l\neq l'}\frac{|y_{l,m}||y_{l',m}|}{ll'}\to 0 \tag{5.99}$$

by the assumptions (C1) and (C3). Finally, (5.97), (5.98) and (5.99) imply (5.96). □

5.7.3 Proof of Theorem 5.4.1

Proof of Theorem 5.4.1: Note that

$$I_{R^{(1)}}(\omega_l)=\frac{1}{2\pi n}\left|\sum_{j=1}^{n}e^{-ij\omega_l}\epsilon_j+\Delta\sum_{j=1}^{k^*}e^{-ij\omega_l}\right|^2=\frac{n\Delta^2}{2\pi}\left(\left|\frac{1}{n}\sum_{j=1}^{k^*}e^{-ij\omega_l}\right|^2+\nu_l\right),$$

where

$$\nu_l = \frac{2\pi}{n\Delta^2} I_\epsilon(\omega_l) + \frac{\sqrt{2\pi}}{\sqrt{n\Delta^2}} 2\Re\left\{ \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} e^{-ij\omega_l}\epsilon_j \frac{1}{n}\sum_{j=1}^{k^*} e^{ij\omega_l} \right\}.$$

It follows that

$$\begin{aligned}
\widehat{H}_{lw} &= \underset{\Delta_1 \le H \le \Delta_2}{\operatorname{argmin}} \left\{ \log \frac{1}{m}\sum_{l=1}^{m} \omega_l^{2H-1} I_{R^{(1)}}(\omega_l) - (2H-1)\frac{1}{m}\sum_{l=1}^{m}\log\omega_l \right\} \\
&= \underset{\Delta_1 \le H \le \Delta_2}{\operatorname{argmin}} \left\{ \log \frac{1}{m}\sum_{l=1}^{m} l^{2H-1} I_{R^{(1)}}(\omega_l) - (2H-1)\frac{1}{m}\sum_{l=1}^{m}\log l \right\} \\
&= \underset{\Delta_1 \le H \le \Delta_2}{\operatorname{argmin}} \left\{ \log \sum_{l=1}^{m} l^{2H-1} \left(\left| \frac{1}{n}\sum_{j=1}^{k^*} e^{-ij\omega_l} \right|^2 + \nu_l \right) - (2H-1)\frac{1}{m}\sum_{l=1}^{m}\log l \right\}. \qquad (5.100)
\end{aligned}$$

We will show first that

$$\sup_{\Delta_1 \le H \le \Delta_2} \left| \sum_{l=1}^{m} l^{2H-1} \left(\left| \frac{1}{n}\sum_{j=1}^{k^*} e^{-ij\omega_l} \right|^2 + \nu_l \right) - \sum_{l=1}^{\infty} l^{2H-3} \frac{|1-e^{-i2\pi l\theta}|^2}{(2\pi)^2} \right| \xrightarrow{p} 0. \qquad (5.101)$$

Observe that

$$|l^2\nu_l| = O_p\left(\sqrt{\frac{m^2}{n\Delta^2}}\right)$$

uniformly in $l = 1, \ldots, m$ from $1/n\sum_{j=1}^{k^*} e^{ij\omega_l} = O(1/l)$ and $I_\epsilon(\omega_l) =: |v(\omega_l)|^2 = O_p(1)$ and $v(\omega_l) = O_p(1)$ from the proof of Theorem 2 in Robinson (1995*b*). Therefore,

$$\sup_{\Delta_1 \le H \le \Delta_2} \left| \sum_{l=1}^{m} l^{2H-1}\nu_l \right| = o_p(1)$$

because of the assumption (5.59) and absolute summability of $\sum_{j=1}^{\infty} j^{2H-3}$ at $H = \Delta_2$. Now, (5.101) follows from

$$\left| \frac{1}{n}\sum_{j=1}^{k^*} e^{-ij\omega_l} \right| = \frac{|1-e^{-ik^*\omega_l}|}{n|1-e^{-i\omega_l}|} \sim \frac{|1-e^{-i2\pi l\theta}|}{2\pi l}, \qquad (5.102)$$

as $l \to \infty$, by using (A1).

Since

$$0 < \sum_{l=1}^{\infty} l^{2H-3} \frac{|1-e^{-i2\pi l\theta}|^2}{2\pi^2} < \infty$$

and since $1/m \sum_{l=1}^{m} \log l \to \infty$, it follows from (5.100) that $\widehat{H}_{lw} \xrightarrow{p} \Delta_2$. $\square$

APPENDIX A

Geometric ergodicity of RDE

Let $\{X_n\}$ be given by RDE (2.1) where (A_n, B_n) are i.i.d. vectors. Then $\{X_n\}$ is a Markov chain and we recall here that it is geometrically ergodic. The following definitions will be used.

Markov chain $\{Y_n\}$ in a general state space S equipped with σ-field $\mathcal{S}$ is called $\mu-$irreducible for some non-degenerate measure μ on $(S, \mathcal{S})$, if $\mu(A) > 0$ implies

$$\sum_{N=1}^{\infty} p_N(y, A) > 0 \quad \text{for all } y \in S,$$

where $p_N(y, A)$ is the $N-$step transition probability of Markov chain starting from y to A. The Markov chain $\{Y_n\}$ is said to be geometrically ergodic if there is $\rho \in (0, 1)$ and constant C_y for each $y \in S$ such that,

$$\|p_N(y, \cdot) - \pi(\cdot)\| := \sup_{A \in \mathcal{S}} \{|p_N(x, A) - \pi(A)|\} \leq C_y \rho^n, \tag{A.1}$$

where $\pi(\cdot)$ denotes the invariant measure of the Markov chain.

Theorem A.0.1. *(Basrak et al. (2002b), Stelzer (2009)) Suppose there exists an $\epsilon > 0$ such that $E|A_1|^\epsilon < 1$ and $E|B_1|^\epsilon < \infty$. If the Markov chain $\{X_n\}$ is μ-irreducible, then it is geometrically ergodic.*

The condition of $\mu-$irreducibility is satisfied for most models of practical interest. For example, for the squares of ARCH(1) series $X_t = \xi_t^2 = \lambda \epsilon_t^2 \xi_{t-1}^2 + \beta \epsilon_t^2$, with $S = (0, \infty)$, $\mu =$ Lebesgue measure and $y > 0$, one obviously already has $p_1(y, A) = P((\lambda y + \beta)\epsilon_t^2 \in A) > 0$ whenever $P(A) > 0$ and ϵ_t^2 has a density on $(0, \infty)$. For the existence of ϵ such that $E|A_1|^\epsilon < 1$, consider $h(p) = E|A_1|^p$. Note that $h(0) = 1$ and $h'(0) = E \log |A_1| < 0$ assuming conditions of Theorem 2.1.1. Hence, there is ϵ such that $E|A_1|^\epsilon < 1$. Observe that

$$E|B_1|^\epsilon \leq (E|B_1|^\alpha)^{\epsilon/\alpha} < \infty,$$

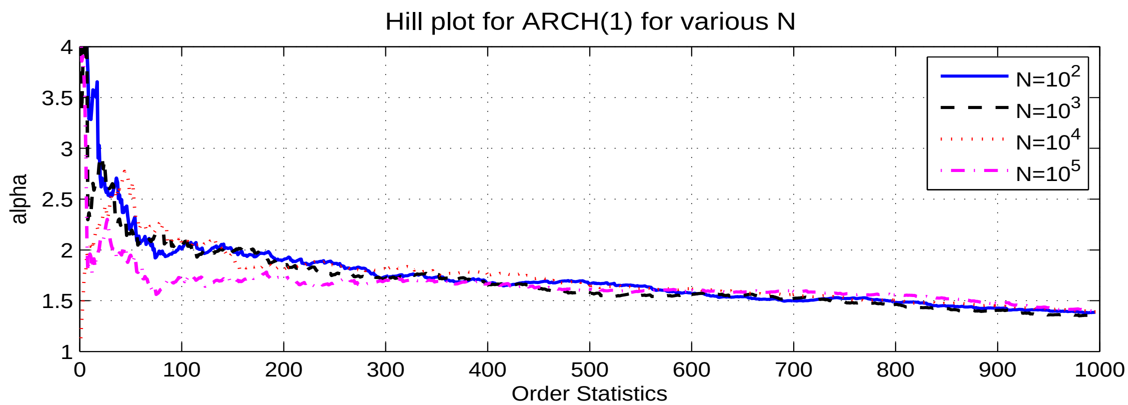

Figure A.1: Hill plot for ARCH(1) series with $\alpha = 10$ and $R = 5,000$ for various N.

from the assumption of Theorem 2.1.1. Therefore, as a result of Theorem A.0.1, P_N converges to P_∞ at an exponentially fast rate. Figure A.1 shows Hill plot based on 5,000 independent observations of ARCH(1) model with various choices of N. There is no difference in Hill plot, which supports the claim that the convergence is quite fast.

Tail exponent for multiplicative cascades with log-normal multipliers

For a log-normal multiplier $W = LN(-\sigma^2/2, \sigma^2)$,

$$\chi_2(h) = \log_2 EW^h - (h-1) = \log_2 \exp\left(-\frac{h\sigma^2}{2} + \frac{h^2\sigma^2}{2}\right) - (h-1)$$

$$= \frac{1}{\log 2}\left(-\frac{h\sigma^2}{2} + \frac{h^2\sigma^2}{2}\right) - (h-1).$$

To find the tail exponent, we need to set

$$\chi_2(h) = 0$$

and look for solution $h > 1$. This yields

$$(\sigma^2 h - 2\log 2)(h-1) = 0$$

or

$$\alpha = 2\log 2/\sigma^2 > 1. \tag{B.1}$$

BIBLIOGRAPHY

Aban, I. B. and Meerschaert, M. M. (2004), 'Generalized least-squares estimators for the thickness of heavy tails', *Journal of Statistical Planning and Inference* **119**(2), 341–352.

Abry, P., Flandrin, P., Taqqu, M. S. and Veitch, D. (2003), Self-similarity and long-range dependence through the wavelet lens, *in* 'Theory and Applications of Long-Range Dependence', Birkhäuser Boston, Boston, MA, pp. 527–556.

Altmann, E. G. and Kantz, H. (2005), 'Recurrence time analysis, long-term correlations, and extreme events', *Physical Review E. Statistical, Nonlinear, and Soft Matter Physics* **71**(5), 056106, 9.

Basrak, B., Davis, R. A. and Mikosch, T. (2002*a*), 'A characterization of multivariate regular variation', *The Annals of Applied Probability* **12**(3), 908–920.

Basrak, B., Davis, R. A. and Mikosch, T. (2002*b*), 'Regular variation of GARCH processes', *Stochastic Processes and their Applications* **99**(1), 95–115.

Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999), 'Tail index estimation and an exponential regression model', *Extremes* **2**(2), 177–200.

Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002), 'On exponential representations of log-spacings of extreme order statistics', *Extremes* **5**(2), 157–180.

Beirlant, J., Vynckier, P. and Teugels, J. L. (1996), 'Tail index estimation, Pareto quantile plots, and regression diagnostics', *Journal of the American Statistical Association* **91**(436), 1659–1667.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York.

Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.-M. (2006), 'On discriminating between long-range dependence and changes in mean', *The Annals of Statistics* **34**(3), 1140–1165.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989), *Regular Variation*, Vol. 27 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press.

Bisaglia, L. and Gerolimetto, M. (2009), 'An empirical strategy to detect spurious effects in long memory and occasional-break processes', *Communications in Statistics - Simulation and Computation* **38**, 172–189.

Boes, D. C. and Salas, J. D. (1978), 'Nonstationarity of the mean and the Hurst phenomenon', *Water Resourses Research* **14**(1), 135–143.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Bos, C., Franses, P. and Ooms, M. (1999), 'Long memory and level shifts: re-analyzing inflation rates', *Empirical Econometrics* **24**(3), 427–449.

Breiman, L. (1965), ‘On some limit theorems similar to the arc-sin law’, *Theory of Probability and its Applications* **10**, 351–360.

Chamayou, J.-F. and Letac, G. (1991), ‘Explicit stationary distributions for compositions of random functions and products of random matrices’, *Journal of Theoretical Probability* **4**(1), 3–36.

Charfeddine, L. and Guegan, D. M. (2009), Breaks or long memory behaviour: An empirical investigation, CES Working Papers, Université Panthéon-Sorbonne (Paris 1), Centre d’Economie de la Sorbonne.

Csörgő, S. and Viharos, L. (1997), ‘Asymptotic normality of least-squares estimators of tail indices’, *Bernoulli* **3**(3), 351–370.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory*, Springer Series in Operations Research and Financial Engineering, Springer, New York.

de Saporta, B., Guivarc’h, Y. and Le Page, E. (2004), ‘On the multidimensional stochastic equation $Y_{n+1} = A_n Y_n + B_n$’, *Comptes Rendus Mathématique. Académie des Sciences. Paris* **339**(7), 499–502.

Diebold, F. X. and Inoue, A. (2001), ‘Long memory and regime switching’, *Journal of Econometrics* **105**(1), 131–159.

Dolado, J. J., Gonzalo, J. and Mayoral, L. (2005), What is what?: A simple time-domain test of long-memory vs. structural breaks, Economics working papers, Department of Economics and Business, Universitat Pompeu Fabra.

Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds (2003), *Theory and Applications of Long-Range Dependence*, Birkhäuser Boston Inc., Boston, MA.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events*, Springer-Verlag, Berlin.

Embrechts, P. and Maejima, M. (2002), *Selfsimilar Processes*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ.

Feuerverger, A. and Hall, P. (1999), ‘Estimating a tail exponent by modelling departure from a Pareto distribution’, *The Annals of Statistics* **27**(2), 760–781.

Flandrin, P. (1992), ‘Wavelet analysis and synthesis of fractional Brownian motion’, *IEEE Transactions on Information Theory* **38**(2, part 2), 910–917.

Folland, G. B. (1999), *Real Analysis*, second edn, John Wiley & Sons Inc., New York.

Fryzlewicz, P. (2007), ‘Unbalanced Haar technique for nonparametric function estimation’, *Journal of the American Statistical Association* **102**(480), 1318–1327.

Gabaix, X. and Ibragimov, R. (2007), Rank-1/2: A simple way to improve the OLS estimation of tail exponents, Working Paper 342, National Bureau of Economic Research.

Geweke, J. and Porter-Hudak, S. (1983), ‘The estimation and application of long memory time series models’, *Journal of Time Series Analysis* **4**(4), 221–238.

Gil-Alana, L. A. (2008), 'Fractional integration and structural breaks at unknown periods of time', *Journal of Time Series Analysis* **29**(1), 163–185.

Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2003), 'Rescaled variance and related tests for long memory in volatility and levels', *Journal of Econometrics* **112**(2), 265–294.

Glynn, P. W. and Whitt, W. (1994), 'Large deviations behavior of counting processes and their inverses', *Queueing Systems* **17**(1-2), 107–128.

Goldie, C. M. (1991), 'Implicit renewal theory and tails of solutions of random equations', *The Annals of Applied Probability* **1**(1), 126–166.

Gomes, M. I., de Hann, L. and Rodrigues, L. H. (2008), 'Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses', *Journal of Royal Statitical Society series B* **70**(Part1), 31–52.

Gomes, M. I., Figueiredo, F. and Mendonça, S. (2005), 'Asymptotically best linear unbiased tail estimators under a second-order regular variation condition', *Journal of Statistical Planning and Inference* **134**(2), 409–433.

Gomes, M. I. and Martins, M. J. (2002), '"Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter', *Extremes* **5**(1), 5–31.

Gomes, M. I., Martins, M. J. and Neves, M. (2000), 'Alternatives to a semi-parametric estimator of parameters of rare events–the jackknife methodology', *Extremes* **3**(3), 207–229 (2001).

Granger, C. W. J. and Hyung, N. (1999), Occasional structural breaks and long memory, University of California at San Diego, Economics Working Paper Series 99-14, Department of Economics, UC San Diego.

Granger, C. W. J. and Hyung, N. (2004), 'Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns', *Journal of Empirical Finance* **11**(3), 399–421.

Granger, C. W. J. and Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**(1), 15–29.

Grintsyavichyus, A. (1981), 'On a random difference equation', *Lithuanian Mathematical Journal* **21**(4), 302–305.

Guivarc'h, Y. (1990), 'Sur une extension de la notion de loi semi-stable', *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques* **26**(2), 261–285.

Hall, P. and Welsh, A. H. (1985), 'Adaptive estimates of parameters of regular variation', *The Annals of Statistics* **13**(1), 331–341.

Hariz, S., Wylie, J. and Zhang, L. (2007), 'Optimal rate of convergence for nonparametric change-point estimators for non-stationary sequences', *The Annals of Statistics* **35**(4), 1802–1826.

Hastie, T., Tibshirani, R. and Friedman, J. (2001), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer-Verlag, New York. Data mining, inference, and prediction.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *The Annals of Statistics* **3**(5), 1163–1174.

Hosking, J. R. M. (1981), 'Fractional differencing', *Biometrika* **68**(1), 165–176.

Hurst, P. (1951), 'Long term storage capacity of reservoirs', *Transactions of the American Socieity of Civil Engineers* **116**, 770–808.

Hurvich, C. M., Deo, R. and Brodsky, J. (1998), 'The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series', *Journal of Time Series Analysis* **19**(1), 19–46.

Jach, A. and Kokoszka, P. (2008), 'Wavelet-domain test for long-range dependence in the presence of a trend', *Statistics. A Journal of Theoretical and Applied Statistics* **42**(2), 101–113.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995), *Distributions in Statistics. Continuous Univariate Distributions.*, John Wiley, New York.

Kahane, J.-P. and Peyrière, J. (1976), 'Sur certaines martingales de Benoit Mandelbrot', *Advances in Mathematics* **22**(2), 131–145.

Karagiannis, T., Molle, M. and Faloutsos, M. (2004), 'Long range dependence - ten years of internet traffic modeling', *IEEE Internet Computing* **8**(5), 57–64.

Kesten, H. (1973), 'Random difference equations and renewal theory for products of random matrices', *Acta Mathematica* **131**, 207–248.

Klemeš, V. (1974), 'The Hurst pheomenon: a puzzle?', *Water Resources Research* **10**(4), 675–688.

Klüppelberg, C. and Pergamenchtchikov, S. (2004), 'The tail of the stationary distribution of a random coefficient AR(q) model', *The Annals of Applied Probability* **14**(2), 971–1005.

Kratz, M. and Resnick, S. I. (1996), 'The QQ-estimator and heavy tails', *Communications in Statistics. Stochastic Models* **12**(4), 699–724.

Kuan, C.-M. and Hsu, C.-C. (1998), 'Change-point estimation of fractionally integrated processes', *Journal of Time Series Analysis* **19**(6), 693–708.

Kuswanto, H. (2009), A new simple test against spurious long memory using temporal aggregation, Preprint, Universität Hannover, Wirtschaftswissenschaftliche Fakultät.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983), *Extremes and related properties of random sequences and processes*, Springer Series in Statistics, Springer-Verlag, New York.

Liu, Q. (2000), 'On generalized multiplicative cascades', *Stochastic Processes and their Applications* **86**(2), 263–286.

Lo, A. W. (1991), 'Long-term memory in stock market prices', *Econometrica* **59**(5), 1279–313.

Mandelbort, B. (1974), 'Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier', *Journal of Fluid Mechanics* **62**(2), 331–358.

Mandelbrot, B. B. (1975), 'Limit theorems on the self-normalized range for weakly and strongly dependent processes', *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31**, 271–285.

Mandelbrot, B. B. and Van Ness, J. W. (1968), 'Fractional Brownian motions, fractional noises and applications', *SIAM Review* **10**, 422–437.

Mandelbrot, B. B. and Wallis, J. R. (1969), 'Computer experiments with fractional Gaussian noises. part 1: Sample graphs, averages and variances', *Water Resources Research* **5**, 228–241.

Martins, M. J., Gomes, M. I. and Neves, M. (1999), 'Some results on the behaviour of Hill's estimator', *Journal of Statistical Computation and Simulation* **63**(3), 283–297.

Mikosch, T. and Stărică, C. (2004), 'Changes of structure in financial time series and the GARCH model', *REVSTAT Statistical Journal* **2**(1), 41–73.

Mills, T. C. (2007), 'Time series modelling of two millennia of northern hemisphere temperatures: long memory or shifting trends?', *Journal of the Royal Statistical Society. Series A. Statistics in Society* **170**(1), 83–94.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. and Karlen, W. (2005), 'Highly variable northern hemisphere temperatures reconstructed from low- and high-resolution proxy data', *Nature* **433**, 613–617.

Müller, U. K. and Watson, M. W. (2008), 'Testing models of low-frequency variability', *Econometrica* **76**(5), 979–1016.

Ohanissian, A., Russell, J. R. and Tsay, R. S. (2008), 'True or spurious long memory? a new test', *Journal of Business & Economic Statistics* **26**, 161–175.

Ossiander, M. and Waymire, E. C. (2000), 'Statistical estimation for multiplicative cascades', *The Annals of Statistics* **28**(6), 1533–1560.

Palma, W. (2007), *Long-Memory Time Series*, John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Park, K. and Willinger, W. (2000), *Self-Similar Network Traffic and Performance Evaluation*, John Wiley & Sons, Inc., New York, NY, USA.

Peng, L. (1998), 'Asymptotically unbiased estimators for the extreme-value index', *Statistics & Probability Letters* **38**(2), 107–115.

Perron, P. and Qu, Z. (2006), An analytical evaluation of the log-periodogram estimate in the presence of level shifts and its implications for stock returns volatility, Boston University - Department of Economics - working papers series.

Phillips, P. C. B. (2007), 'Unit root log periodogram regression', *Journal of Econometrics* **138**(1), 104–124.

Pickands, III, J. (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* **3**, 119–131.

Puhalskii, A. A. and Whitt, W. (1997), 'Functional large deviation principles for first-passage-time processes', *The Annals of Applied Probability* **7**(2), 362–381.

Qu, Z. (2009), A test against spurious long memory, Preprint.

Reiss, R.-D. and Thomas, M. (2007), *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Verlag, Basel.

Resnick, S. (1992), *Adventures in Stochastic Processes*, Birkhäuser Boston Inc., Boston, MA.

Resnick, S. I. (1997), 'Heavy tail modeling and teletraffic data', *The Annals of Statistics* **25**(5), 1805–1869. With discussion and a rejoinder by the author.

Robinson, P. M. (1995*a*), 'Gaussian semiparametric estimation of long range dependence', *The Annals of Statistics* **23**(5), 1630–1661.

Robinson, P. M. (1995*b*), 'Log-periodogram regression of time series with long range dependence', *The Annals of Statistics* **23**(3), 1048–1072.

Robinson, P. M., ed. (2003), *Time Series with Long Memory*, Oxford University Press, Oxford.

Roughan, M. and Veitch, D. (1999), Measuring long-range dependence under changing traffic conditions, *in* 'Proceedings of IEEE INFOCOM', Vol. 3, pp. 1513–1521.

Samorodnitsky, G. (2006), 'Long range dependence', *Foundations and Trends® in Stochastic Systems* **1**(3), 163–257.

Shimotsu, K. (2006), Simple (but effective) tests of long memory versus structural breaks, Working Papers 1101, Queen's University, Department of Economics.

Smith, A. (2005), 'Level shifts and the illusion of long memory in economic time series', *Journal of Business & Economic Statistics* **23**(3), 321–335.

Smith, R. L. (1987), 'Estimating tails of probability distributions', *The Annals of Statistics* **15**(3), 1174–1207.

Stelzer, R. (2009), 'On Markov-switching ARMA processes – stationarity, existence of moments and geometric ergodicity', *Econometric Theory* **25**(01), 43–62.

Teverovsky, V., Taqqu, M. S. and Willinger, W. (1999), 'A critical look at Lo's modified R/S statistic', *Journal of Statistical Planning and Inference* **80**(1-2), 211–227.

Veitch, D. and Abry, P. (1999), 'A wavelet-based joint estimator of the parameters of long-range dependence', *IEEE Transactions on Information Theory* **45**(3), 878–897.

Veres, S. and Boda, M. (2000), The chaotic nature of TCP congestion control, *in* 'Proceedings of IEEE INFOCOM', Vol. 3, pp. 1715–1723.

Vostrikova, L. J. (1981), 'Detecting "disorder" in multidimensional random processes', *Soviet Math. Doklady* **24**, 55–59.

Wright, H. (1998), 'Testing for a structural break at unknown date with long-memory disturbances', *Journal of Time Series Analysis* **19**(3), 369–376.