University of Nottingham

School of Mathematical Sciences

Characterising the zero-crossing statistics of non-Gaussian stable processes

Daniel Slater

A thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

2018

For Mum, Dad and Jamie.

Abstract

Zero-crossing analysis is an old problem, with various attempts to tackle it not yet having led to a universally applicable solution. This is especially true for non-Gaussian stable processes, whose heavy-tails result in problematic statistical properties such as undefined autocorrelation. In this thesis we investigate the need for the coherence (as opposed to the ill-defined autocorrelation) and its applicability in predicting the zero-crossing statistics of non-Gaussian stable processes. This requires us to address the innate issues surrounding estimating the heavy-tailed index for empirical data sets within the stable regime. This includes a review of the ubiquitous Hill estimator, and a comparison of it with a proposed method based on characteristic function analysis by testing them on synthetic and real-world data. The theoretical framework of the thesis is then tested on synthetic data, using a novel method to generate non-Gaussian stable noises. From this it is concluded that coherence offers a means to predict the zero-crossing statistics of stable processes.

Acknowledgements

The work in this thesis is a result of the enormous support and encouragement I have received during my studies. I am very lucky to be able to count on so many people and will never forget that.

Principally, I would like to thank my supervisor Keith Hopcraft. I could not have wished for a better supervisor, one who met the challenging mathematical problems we faced with a positivity and excitement that inspired me. More importantly though, I am grateful to have had the experience and gained a friend. His support dragged me through the project in many situations and circumstances. I will miss our discussions, which I found often deviated rather quickly from zero-crossings to cricket. I wish you all the best in your much-deserved retirement.

I would also like to thank my co-supervisors Eric Jakeman and David Sirl. Both were enthusiastic to hear about how the project progressed and their input helped shape the project into what it was. Thank you also to my internal examiners Michael Tretyakov and Huiling Le, and my external examiner John McWhirter. Between them, they ensured my viva experience was transformed from something I was dreading to an enlightening and interesting discussion.

I would also like to thank the University of Nottingham. I have spent over 7 years here, and I leave with a great sense of affinity for the place I have called 'home'. Special thanks must go to University of Nottingham Futsal. The club has given me such great adventures during my studies, and it is these that have kept me sane.

I would like to thank all the friends and family that have supported me during this project. You do not realise how much you have helped. Special thanks must go to Peter Teakle for providing me with unwavering academic encouragement. Finally, I would like to thank Mike, Debbie and Jamie. Without you I wouldn't have been able to do this.

CONTENTS

1

INTRODUCTION AND BACKGROUND

1.1 BACKGROUND

The analysis of extreme fluctuations has long been of interest to mathematicians. Whilst extreme events, such as market crashes and natural disasters, are often rare and widely interspersed, their effects often account for huge damages when compared with commonly observed behaviour. Such a result makes their mathematical investigation of great importance.

The quantitative analyses of the financial world is dominated by the Gaussian distribution. Its self-similarity and analytical properties, together with its role in the Central Limit Theorem, has made its use prevalent. For example, the famous Black-Scholes model to calculate the European put and call options of stocks is reliant on it [1]. Extreme economic events display a particular vulnerability to such assumptions, however. The events of Black Monday in 1987 [2] saw the values of stock markets around the plummet at rates hitherto unseen. If the financial markets behaved according to Gaussian models, then such a crash would have been expected to occur just once in a time period 10 times that of the age of the universe. Whilst Black Monday is the most famous example of an extreme financial event, the regularity of which similar crashes (and booms) occur prompts a need to both recognise and introduce means of quantifying such behaviour.

A potential means to model extremal behaviour is through heavy-tails, where the asymptotic tail of the density function is a power-law of the form $\overset{a}{\sim} x^{-(1+\alpha)}$. These heavy-tails

often model large fluctuations more accurately than other common-place distributions, such as the Gaussian and exponential distributions. A particularly important set of heavy-tailed distributions is the set of non-Gaussian stable distributions, due to their self-similarity and role in the generalised Central Limit Theorem. Because of this, Mandelbrot proposed modelling cotton prices with them in 1963 [3]. They have subsequently been used to model a number of financial objects [4].

A key aspect of the probability distribution of the stable distributions (alongside the self-similarity property) is the exponent of the heavy-tail. For non-Gaussian stable distributions, the tail of the density function is $\overset{a}{\sim} x^{-(1+\alpha)}$ where $\alpha \in (0,2)$. This results in statistical properties that make their study particularly interesting. Specifically, the variance (and sometimes mean) of a non-Gaussian stable random variable is undefined. This means the covariance between stable random variables is also undefined, making quantifying their dependency problematic.

Despite the potential pitfalls, there is undeniable use and applicability of stable distributions. Aside from the financial world, they have been used in spectroscopy [5], and to model sandpiles [6] and solar flare waiting times [7]. More generally, heavy-tailed distributions are also widely used in a variety of contexts [8] [9], [10], [11], [12].

An unsolved problem in the study of stochastic processes is a complete mathematical description of zero-crossings and return-times. Zero-crossings occur when a signal crosses the x-axis, and the 'return-times' correspond to the intervals between subsequent crossings. Zero-crossings offer an alternative means of investigating extreme events by considering the time intervals between events and their subsequent distributions. The central aim of zero-crossing analysis is to determine the probability distribution(s) of the interval lengths between crossings. Since the initial work into zero-crossings by Kac [13] and Rice [14], this problem still remains (generally) unsolved. Zero-crossings are themselves applicable in a number of areas. There is extensive literature on such applications [15] including a review by Blake and Lindsey in [16].

Unlike non-Gaussian stable processes, the zero-crossings of Gaussian processes have previously been investigated in-depth. In 1945, Rice [14] showed that the mean zero-crossing rate for Gaussian processes is a function of the process' autocorrelation function by $\beta = \sqrt{-a''(0)}/\pi$ where a is the autocorrelation function. Rice was able to use this to approximate the zero-crossings pdf for small interval times. McFadden then used Rice's work to calculate the zero-crossings pdf under certain conditions [17], [18].

The result of McFadden's work was that zero-crossing events and the resulting statistical properties were directly related to the autocorrelation function of the underlying process. Such a conclusion made investigating the set of non-Gaussian stable processes problematic, as the heavy-tails resulted in an undefined autocorrelation. To analyse the zero-crossings of non-Gaussian stable processes, an alternative formulation was therefore required. This was done directly by Hopcraft and Jakeman in [19]. In their paper they proposed a new form of dependency structure termed the coherence. This was formulated and defined in terms of a joint characteristic function and calculated through the existence of fractional moments. The theoretical implications of this work were tested in [15], with a particular emphasis on the effect the coherence function's form had on zero-crossing statistics. The main aim of this thesis is to test the claims of [15] and [19] by testing the proposed methods on empirical stable processes.

Investigating non-Gaussian stable processes is itself problematic for several reasons. For a heavy-tailed process there is a need to be accurate in analysing the heavy-tail index, α, to determine whether or not a process is asymptotically stable. If not, then autocorrelation suffices as a means of quantifying the dependency structure, and the work into coherence need not be consulted. Additionally, for processes that are (asymptotically) stable, the proposals to estimate the underlying coherence require an accurate estimate of α.

The problem of calculating the heavy-tail index is itself an old one, with much literature and investigation into the topic. A review of a number of means of identifying the heavy-tail index can be found in [20]. The most obvious method of calculating the index, through graphical analysis, is error prone [21]. A number of alternatives, including Hill's estimator

[22] and quantile matching [23] have been put forward as potential means to estimate the index with accuracy. Despite a plethora of options, the most popular estimator remains Hill's estimator [25]. Parameter fitting methods specifically constructed for stable processes also exist, such as characteristic function analysis [26] which calculates (amongst other parameters) the heavy-tail index of stable processes.

Analysing the zero-crossing statistics of non-Gaussian stable processes therefore has the potential to be an extremely complicated task. In this thesis we attempt to do exactly that, through a number of steps. We first investigate tail estimation methods, with a specific focus on their applicability to non-Gaussian stable processes. This information is then used to motivate a need to generate synthetic stable processes with memory and a known distribution. The empirical crossing statistics of the synthetic data is then compared with the theoretical predictions calculated by using coherence. This allows us to comment on the applicability of coherence as a means of quantifying dependency in stable processes.

1.2 PROBABILITY THEORY

Before proceeding with the specifics of this thesis, it is useful to introduce basic probability theory, as well as the relevant nomenclature required for it. This is described in this chapter, and can be found in most introductory probability theory textbooks such as [27] and [28].

1.2.1 *Random variables*

Probability space

Fundamentally probability theory concerns itself with some experiment whose outcome is randomly decided. We therefore describe the mathematical framework in terms of such an experiment. To do so, we first introduce the probability space by defining three key properties: the sample space Ω, the set of events $\mathcal{F}$ and the probability function $\mathbb{P}$. The sample space Ω is defined as the set of all possible outcomes of the experiment. The set of events $\mathcal{F}$ is defined as the collection of all subsets of Ω (including the empty set and Ω itself). The probability function $\mathbb{P}$ is defined as some function $\mathbb{P} : \mathcal{F} \rightarrow [0,1]$ that

assigns every possible event with a probability of its occurence. Let E be an event such that $E \in \mathcal{F}$. Then the probability of E occuring is $\mathbb{P}(E)$ where $\mathbb{P}(E) = 0$ would signify that E had no chance of occuring and $\mathbb{P}(E) = 1$ would signify E would occur almost surely. We then say that the experiment is defined in terms of these three features, which we term the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

For example, consider flipping a fair coin and recording its outcome. By denoting $H =$ Heads and $T =$ Tails we can describe the relevant probability space. The sample space would be $\Omega = \{H, T\}$, the set of events $\mathcal{F} = \{\{\}, \{H\}, \{T\}, \{H, T\}\}$ and the probability function $\mathbb{P}(\{\}) = 0$ (as the coin flip must have an outcome), $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = 0.5$ (as the coin flip is fair), $\mathbb{P}(\{H, T\}) = 1$ (as the coin flip must have an outcome).

Random variables

In order to add mathematical rigour to the probability space we now introduce random variables. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a random variable (rv) is a measurable function $X : \Omega \to A$ where $A \subseteq \mathbb{R}$. Let $E \subseteq A$. Then the probability that $X \in E$ is $\mathbb{P}(\omega \in \Omega : X(\omega) \in E)$. For convenience, we can write this as $\mathbb{P}(X \in E)$.

Random variables can be split into two distinct categories: discrete random variables and continuous random variables. Let X be a random variable. Then X is a discrete random variable if and only if (iff) X can take on a countable number of possible outcomes. Conversely, X is continuous iff it can take on an uncountable number of values. For example, a coin toss would be represented by a discrete random variable, whilst a person's height a continuous random variable.

In this investigation it is important we discuss probabilistic concepts in terms of both discrete and continuous random variables. When considering empirical problems, it is often the case we use continuous random variables to model large discrete systems. Conversely, when investigating continuous random variables computationally we must use discrete samples.

Distribution functions

To quantify the relative frequency of outcomes of random variables we now introduce a series of distribution functions, defined in terms of the probability measure $\mathbb{P}$. Let X be a random variable. The cumulative distribution function (cdf) of X is defined as the function $F_X : \mathbb{R} \to [0,1]$ such that

$$F_X(x) := \mathbb{P}(X \leq x)$$

where $x \in \mathbb{R}$. The complementary cdf (ccdf) is defined as the function $\bar{F}_X(x) := 1 - F_X(x)$. It follows that $F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b)$ and $\bar{F}_X(x) = \mathbb{P}(X > x)$.

The cdf describes the cumulative probability of a random variable, but does not quantify the relative probability of individual outcomes or intervals. To do this, we introduce the probability mass function and probability density function. Let $X : \Omega \to A$ be a discrete random variable, where $A \subset \mathbb{R}$ is a countable set. Then the probability mass function (pmf) of X is defined as the function $p_X : A \to [0,1]$ such that

$$p_X(x) := \mathbb{P}(X = x)$$

for $x \in A$.

We now consider the quantity analagous to the pmf for continuous random variables, the probability density function (pdf). To do so, we must first introduce absolutely continuous random variables. Let X be a continuous random variable. Then X is absolutely continuous iff its cdf is absolutely continuous with respect to the Lebesque measure. To understand this definition, a thorough knowledge of measure theory beyond the scope of this thesis is required. For the purposes of our work it suffices to understand that to admit a pdf, a random variable must be absolutely continuous. This is true for the majority of work into continuous random variables, but there are notable examples where it is not; for example Cantor random variables. In this thesis, we shall work with continuous random variables under the assumption they are also absolutely continuous. Let X be a(n absolutely) continuous random variable. Then its probability density function (pdf) is defined as the function $f_X : \mathbb{R} \to [0, \infty)$ that satisfies

$$F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx$$

for all $a, b \in \mathbb{R}$ where F_X is its cdf.

Expected values and moments

Another means of quantifying a random variable's properties and the relative frequencies of its outcomes is by expected values and moments. Examples of these include mean and standard deviation which are used to quantify the average location and dispersion of a random variable.

Let $X : \Omega \to A$ be a random variable. Then we define the expected value of X by the Lebesque integral

$$\langle X \rangle := \int_{\omega \in \Omega} X(\omega)\, \mathbb{P}(d\omega).$$

We can generalise the idea of the expected value such that we can take the expected value of some function of a random variable. Let $g : A \to \mathbb{R}$ be a measurable function. Then the expected value of $g(X)$ is defined as the Lebesque integral

$$\langle g(X) \rangle := \int_{\omega \in \Omega} g(X(\omega))\, \mathbb{P}(d\omega).$$

The definitions for expected values can be used to produce expressions in terms of a rv's pmf or pdf. Let $X : \Omega \to A$ be a discrete random variable with pmf p_X. Then the expected value of X is

$$\langle X \rangle = \sum_{x \in A} x\, p_X(x).$$

Let $g : A \to \mathbb{R}$ be a measurable function. Then the expected value of $g(X)$ is

$$\langle g(X) \rangle = \sum_{x \in A} g(x)\, p_X(x).$$

Now, let $X : \Omega \to \mathbb{R}$ be an absolutely continuous random variable with pdf f_X. Then the expected value of X is

$$\langle X \rangle = \int_{\mathbb{R}} x\, f_X(x)\, dx.$$

Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function. Then the expected value of $g(X)$ is

$$\langle g(X) \rangle = \int_{\infty} g(x)\, f_X(x)\, dx.$$

An important family of expected values is the integer moments. Let X be a random variable. Then the nth raw moment of X is defined as

$$\mu_n := \langle X^n \rangle$$

for $n \in \mathbb{N}$. The mean μ of a random variable is defined as its 1st raw moment $\mu := \mu_1 = \langle X \rangle$. The nth central moment of X is defined as

$$\sigma_n := \langle (X - \mu)^n \rangle$$

for $n \in \mathbb{N}$. The variance σ^2 of a random variable is defined as its 2nd central moment $\sigma^2 := \sigma_2$. The nth standardised moment is defined as

$$\mu'_n := \frac{\sigma_n}{\sigma^n}$$

for $n \in \mathbb{N}$. Skewness and kurtosis are defined as the 3rd and 4th standardised moment respectively. Whilst the moments are traditionally considered just in terms of integer values, they may also be evaluated for fractional values. The nth fractional moment is defined as

$$\bar{\mu}_n := \langle |X|^n \rangle$$

for $n \in \mathbb{Q}$.

Characteristic functions

An alternative means of describing the probabilistic behaviour of a random variable is through its characteristic function. Let X be a random variable. Then the characteristic function (cf) of X is defined as the function $\phi : \mathbb{R} \to \mathbb{C}$ such that

$$\phi(t) := \langle \exp(itX) \rangle \tag{1.2.1}$$

where i is the imaginary unit and $t \in \mathbb{R}$. Whilst cf's are intuitively more complicated than cdf's, pmf's and pdf's, they have a number of properties that make their use advantageous in certain circumstances. Crucially, the characteristic function is well-defined for all real-valued random variables and can be calculated for any analytical pdf. Furthermore we can define random variables that do not have an analytical pdf in terms of an analytical cf (an example of this is the set of stable random variables). Additionally, the cf is bounded by $|\phi(t)| \leq 1$ and is unity at the origin. If a random variable's raw moments exist and are finite then its characteristic function can be expressed as a power series

$$\phi(t) = \sum_{n=0}^{\infty} \frac{1}{n!}\, i^n \mu_n t^n = 1 + i\mu_1 t - \frac{1}{2}\mu_2 t^2 - \frac{1}{6} i\mu_3 t^3 + \cdots,$$

by defining $\mu_0 = 1$ for convenience.

1.2.2 *Joint distributions*

Multivariate random variables

The ideas introduced to describe single random variables can be generalised to consider multivariate random variables (also called random vectors). These multivariate random variables can describe random experiments where more than one outcome is measured, or where multiple measurements are taken. We formalise multivariate random variables by again considering the probability space. Let $X_1, \ldots, X_n : \Omega \to A \subseteq \mathbb{R}$ be random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{R})$. Then we can define $\boldsymbol{X} = (X_1, \ldots, X_n)$ as an n-dimensional multivariate rv, or an n-dimensional random vector. We say a multivariate rv is discrete if its components are all discrete rv's, and continuous if its components are all continuous rv's. Additionally, we can say it is a mixed multivariate rv if some components are discrete and some are continuous.

Multivariate distribution functions

The definition of a multivariate rv can be used to introduce generalisations of the distribution functions. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an n-dimensional multivariate rv. Then the joint cumulative distribution function (joint cdf) of $\boldsymbol{X}$ is defined as the function $F_{\boldsymbol{X}} : \mathbb{R}^n \to [0,1]$ such that

$$F_{\boldsymbol{X}}(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

where $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Similarly, the joint complementary cdf (joint ccdf) is defined as $\bar{F}_{\boldsymbol{X}}(\boldsymbol{x}) := 1 - F_{\boldsymbol{X}}(\boldsymbol{x})$.

As was the case with univariate random variables, we may now describe the relative frequencies of the individual outcomes and intervals of multivariate random variables using joint pmf's and joint pdf's. Let $\boldsymbol{X}$ be an n-dimensional discrete multivariate rv such that each component $X_1, \ldots, X_n : \Omega \to A$, where $A \subset \mathbb{R}^n$ is a countable set. Then the joint probability mass function (joint pmf) of $\boldsymbol{X}$ is defined as the function $p_{\boldsymbol{X}} : A \to [0,1]$ such that

$$p_{\boldsymbol{X}}(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

where $\boldsymbol{x} = (x_1, \ldots, x_n) \in A$.

As per the case for univariate rv's; we first clarify that to admit a joint pdf a multivariate continuous rv's components must also be absolutely continuous. In the context of this investigation we shall assume that multivariate continuous rv's are made up of absolutely continuous rv's, unless stated otherwise. Let $\boldsymbol{X}$ be an n-dimensional multivariate continuous rv such that each component $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ is an absolutely continuous rv. Then the joint probability density function (joint pdf) of $\boldsymbol{X}$ is defined as the measurable function $f_{\boldsymbol{X}} : \mathbb{R}^n \to [0, \infty)$ that satisfies

$$\mathbb{P}(\boldsymbol{X} \in E) = \int_E f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}$$

for all $E \subseteq \mathbb{R}^n$.

It is worth noting we can define a combination of joint pmf's and joint pdf's for mixed multivariate random variables. Such a generalisation follows simply from the above definitions, and is not required in the context of this thesis.

A key concept when investigating multivariate rv's is dependence. Dependence refers to how random variables affect one another and how they are linked. This is described mathematically by the joint pmf or pdf. We say a set of random variables are independent if their behaviour does not affect one another. Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a multivariate discrete rv with joint pmf $p_{\boldsymbol{X}}$. Then $X_1, \ldots, X_n$ are independent iff

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{X_1}(x_1)\, p_{X_2}(x_2) \cdots p_{X_n}(x_n)$$

for all $\boldsymbol{x} \in A$ where p_{X_j} is the pmf of X_j. Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a multivariate absolutely continuous rv with joint pdf $f_{\boldsymbol{X}}$. Then $X_1, \ldots, X_n$ are independent iff

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ where f_{X_j} is the pdf of X_j. A set of rv's are dependent iff they are not independent.

Expected values

By introducing joint pmf's and joint pdf's we can now introduce expected values for multivariate rv's. Let $\boldsymbol{X}$ be a multivariate discrete random variable with joint pmf $p_{\boldsymbol{X}} : A \to [0,1]$ where $A \subset \mathbb{R}^n$ is countable. Let g be some function $g : A \to \mathbb{R}$. Then the expected value of $g(\boldsymbol{X})$ is

$$\langle g(\boldsymbol{X}) \rangle := \sum_{\boldsymbol{x} \in A} g(\boldsymbol{x})\, p_{\boldsymbol{X}}(\boldsymbol{x}).$$

Let $\boldsymbol{X}$ be a multivariate absolutely continuous rv with joint pdf $f_{\boldsymbol{X}} : \mathbb{R}^n \to [0, \infty)$. Let g be some function $g : \mathbb{R}^n \to \mathbb{R}$. Then the expected value of $g(\boldsymbol{X})$ is

$$\langle g(\boldsymbol{X}) \rangle := \int_{\mathbb{R}^n} g(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}.$$

Covariance and correlation

Expected values for multivariate rv's can be used to introduce the ideas of covariance and correlation. For a bivariate random variable (2-dimensional multivariate rv) covariance and correlation can quantify the relationship between those random variables. Let X and Y be random variables with means μ_X, μ_Y, and variances σ_X^2, σ_Y^2 respectively. Then the covariance between X and Y is defined as

$$\operatorname{cov}(X, Y) := \langle (X - \mu_X)(Y - \mu_Y) \rangle.$$

Correlation can broadly be used to describe any statistical relationship between random variables, but is usually formally defined by the Pearson correlation coefficient. The Pearson correlation coefficient ρ quantifies the linear relationship between two random random variables by the normalising the covariance, and is defined as

$$\rho(X, Y) := \frac{1}{\sigma_X \sigma_Y} \operatorname{cov}(X, Y).$$

It follows that ρ is bounded such that $-1 \leq \rho \leq 1$, where $\rho = 1$ signifies that X and Y have a perfectly positive linear relationship, $\rho = -1$ signifies a perfectly negative linear relationship and $\rho = 0$ signifies no linear relationship. Unless stated otherwise, 'correlation' henceforth refers to the Pearson correlation coefficient.

It is worth noting that $\rho = 0$ does not imply that two random variables are independent. For example, consider two random variables $X = \sin(Z)$ and $Y = \cos(Z)$, where $Z \sim \mathcal{U}[0, 2\pi)$, such that $\{X, Y\}$ correspond to the Cartesian coordinates drawn from random on the unit circle. Then there is no linear relationship between X and Y and $\rho(X, Y) = 0$, but they are clearly not independent.

Marginal distributions

For multivariate random variables, we are able to describe the behaviour of just one of its components via marginal distribution functions. For the purposes of this investigation it suffices to initially describe marginal distribution functions in terms of bivariate rv's. Let $\boldsymbol{X} = (X_1, X_2)$ be a discrete bivariate rv where $X_1,\ X_2 : \Omega \to A$ with joint pmf $p_{\boldsymbol{X}}$. Then the marginal pmf of X_1 is

$$p_{X_1}(x_1) = \sum_{x_2 \in A} p_{\boldsymbol{X}}(x_1, x_2).$$

Equivalently, let $\boldsymbol{X} = (X_1, X_2)$ be an absolutely continuous bivariate rv where $X_1,\ X_2 : \Omega \to \mathbb{R}$ with a joint pdf $f_{\boldsymbol{X}}$. Then the marginal pdf of X_1 is

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{\boldsymbol{X}}(x_1, x_2)\, dx_2.$$

These can be generalised to n-dimensional multivariate rv's, where the summation (for the joint pmf) or integration (for the joint pdf) is taken over all the random variable components apart from X_1. From this it is clear that for a multivariate random variable with a known joint pmf (or joint pdf) the marginal pmf's (or pdf's) of its individual components can all be calculated. The converse is not true, however. Specifically, a set of marginal pmf's (or marginal pdf's) do not imply a unique joint pmf (or joint pdf).

1.2.3 *Stochastic processes*

Discrete-time and continuous-time processes

The ideas of multivariate rv's are now used to introduce stochastic processes. Stochastic processes are used to describe random systems that evolve through time, and are therefore defined as time-ordered sets of random variables. From now on, we make no reference to

whether the random variables are themselves discrete or continuous, but only whether they are discrete or continuous in time.

A discrete-time process $\{X_t\}$ is defined as a time-ordered set of random variables defined at countably many points in time, such that

$$\{X_t\} = \{X_1,\ X_2,\ \ldots\}.$$

where $t \in T$ and $T \subseteq \mathbb{N}$. Similarly a continuous-time process $\{X_t\}$ is defined as a time-ordered set of random variables defined at uncountably many points on a given time interval, such that $t \in T$ where $T \subseteq \mathbb{R}$ is a continuous interval. An important class of continuous-time stochastic processes is the sample-continuous stochastic processes. Let $\{X_t\}$ be a continuous-time stochastic process defined on $t \in T$ where $T \subseteq \mathbb{R}$, and on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\{X_t\}$ is a sample-continuous process iff

$$\mathbb{P}\left(\omega \in \Omega : \lim_{s \to t} |X_t(\omega) - X_s(\omega)| = 0\right) = 1$$

almost surely for all $t \in T$. This ensures that its sample paths (discussed subsequently) are continuous. For convenience, we can refer to stochastic processes simply as 'processes' and sample-continuous stochastic processes simply as 'continuous stochastic processes' or 'continuous processes'.

Stationarity, autocovariance and autocorrelation

Stationarity is an important concept when investigating stochastic processes, as it allows us a framework in which past behaviour can be used to predict future behaviour. Let $\{X_t\}$ be a stochastic process. We denote its cdf at times $t = t_1,\ \ldots,\ t_n$ by $F_X(x_{t_1}, \cdots, x_{t_n})$. Then $\{X_t\}$ is stationary iff its cdf does not change in time, such that $F_X(x_{t_1}, \cdots, x_{t_n}) = F_X(x_{t_1+\tau}, \cdots, x_{t_n+\tau})$ for all $n,\ \tau,\ t_1,\ \ldots, t_n \in \mathbb{R}$. This results in ideal properties for the study of stochastic processes such that its statistical properties - such as mean, variance and marginal distributions - are consistent in time.

For a stochastic process $\{X_t\}$, the concept of covariance between a pair of rv's can be generalised to measure the covariance between all elements in a process. Because this

describes the covariance between elements in a single process, this is called the autocovariance. The autocovariance of a process at times t_1, t_2 is $\text{acov}(t_1, t_2) := \text{cov}(X_{t_1}, X_{t_2}) = \langle (X_{t_1} - \mu_1)(X_{t_2} - \mu_2) \rangle$ where $\mu_1 = \langle X_{t_1} \rangle$, $\mu_2 = \langle X_{t_2} \rangle$. Let $t_1 = t$, $t_2 = t + \tau$. If $\{X_t\}$ is stationary, it follows that $\mu_1 = \mu_2 = \mu$ and $\langle (X_t - \mu)(X_{t+\tau} - \mu) \rangle$ is independent of choice of t. It follows that the autocovariance of a stationary process can be expressed solely as a function of the time lag $\tau = |t_2 - t_1| \geq 0$ only, such that

$$\text{acov}(\tau) = \langle (X_t - \mu)(X_{t+\tau} - \mu) \rangle \tag{1.2.2}$$

for all $\tau \geq 0$. Similarly, correlation can be generalised for a stochastic process to consider the correlations between its various elements. For a stochastic process $\{X_t\}$, the autocorrelation function (acf) at times t_1, t_2 is $a(t_1, t_2) := \rho(X_{t_1}, X_{t_2})$. As per the case for autocovariance, the acf of a stationary process can be expressed solely as a function of the time lag $\tau \geq 0$ such that

$$a(\tau) = \frac{1}{\sigma^2} \text{acov}(\tau) \tag{1.2.3}$$

for all $\tau \geq 0$, where σ^2 is the variance of $\{X_t\}$. A weaker form of stationarity is wide-sense stationarity. Let $\{X_t\}$ be a stochastic process. Then $\{X_t\}$ is wide-sense stationary (wss) iff its mean, variance and autocorrelation are consistent in time, such that its autocovariance and autocorrelation are equal to Equation 1.2.2 and Equation 1.2.3 respectively.

In the same way that correlation quantifies the linear dependence between two random variables, the autocorrelation function quantifies the dependence in a process as a function of time lag. In this way, it is used to quantify the 'memory' inherent in a process. We shall see later that, despite its ubiquity in measuring memory in time-series data, the autocorrelation function has significant deficiencies in certain circumstances. A process with an acf $a(\tau) = \delta_{0,\tau}$, where $\delta_{i,j}$ is the Kronecker-delta function, is said to be δ-correlated.

Stochastic noise

We now introduce the concept of stochastic noise as a means to differentiate between a process whose elements are independent and a process whose elements are dependent. Let $\{X_t\}$ be a stationary stochastic process defined on $t \in T$. Then $\{X_t\}$ is a white noise iff $\langle X_t \rangle = 0$, $\langle X_t^2 \rangle < \infty$, and its elements are independent rv's for all $t \in T$. Because a white

noise is a sequence of independent rv's, its power-spectrum has equal intensity across all frequencies. It is from this that the name white noise is derived, as an analogy to white light which contains the entire spectrum of visible light.

Let $\{X_t\}$ be a stationary stochastic process defined on $t \in T$. Then $\{X_t\}$ is a coloured noise iff $\mu = \langle X_t \rangle = 0$, $\sigma^2 = \langle X_t^2 \rangle < \infty$ for all $t \in T$, and it is not a white noise. In other words, in contrast to white noise, coloured noise is a process whose elements have some form of dependency structure. This results in a power-spectrum that does not have equal intensity across all frequencies (analogous to coloured light).

For the purposes of this thesis, we shall assume that a white (or coloured) noise refers to a process whose elements are identically distributed. We also relax the assumption that the mean is necessarily $\mu = 0$.

Kolmogorov's extension theorem

In this thesis we often move between using discrete-time stochastic processes and continuous-time stochastic processes. The theory surrounding the zero-crossing statistics is based on continuous-time processes, whereas the business of investigating them numerically requires the use of discrete-time processes as representations. It is therefore important to introduce the framework that allows us to do this - the Kolmogorov extension theorem [29].

Let $T \subseteq \mathbb{R}$ be some time interval, and let $t_1, \ldots, t_k \in T$ be a discrete number of time points on T such that $k \in \mathbb{N}$. Let $p_{t_1,\cdots,t_k}$ be a probability measure on $\mathbb{R}^{nk}$. Suppose $p_{t_1,\cdots,t_k}$ satisfies the conditions:

1. for all permutations π on $\{1, \ldots, k\}$ and measurable sets $F_i \subseteq \mathbb{R}^n$,

$$p_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = p_{t_{\pi(1)},\cdots,t_{\pi(k)}}(F_{\pi(1)} \times \cdots \times F_{\pi(k)});$$

2. for all measurable sets $F_i \subseteq \mathbb{R}^n$, and $m \in \mathbb{N}$,

$$p_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = p_{t_1,\cdots,t_k,t_{k+1},\cdots,t_{k+m}}(F_1 \times \cdots \times F_k \times (\mathbb{R}^n)^m).$$

Then the Kolmogorov extension theorem states that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $\{X_t\}$ on Ω where $X_t : \Omega \to \mathbb{R}^n$ such that

$$p_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}(X_{t_1} \in F_1,\ \ldots,\ X_{t_k} \in F_k)$$

for all $t_i \in T$, $k \in \mathbb{N}$ and measurable sets $F_i \subseteq \mathbb{R}^n$.

Kolmogorov's extension theorem ensures that a stochastic process can be constructed in terms of finite-dimensional distributions, as long as those finite-dimensional distributions satisfy the two conditions. In the context of this thesis, it is sufficient to understand that Kolmogorov's extension theorem allows us to use the probability distributions of a process evaluated at discrete-time points to construct a continuous-time stochastic process. We shall henceforth assume that the processes we work with satisfy the two required conditions to apply Kolmogorov's extension theorem.

Sample statistics

Until now we have described the mathematical framework exclusively in terms of stochastics and randomness with no reference to observed outcomes. When investigating stochastic processes, we gauge the success of the theory by comparing its predictions with observed statistics. It is therefore important we introduce the nomenclature used to discuss the realisations of stochastic processes.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a sample point is an element of the sample space $\omega \in \Omega$. Let $X : \Omega \to A$ be a random variable. Then $X(\omega)$ is a sample random variable and a possible outcome of X. Let $\{X_t\}$ be a stochastic process such that $X_t : \Omega \to A$ for $t \in T$. Then $\{X_t(\omega)\}$ is a sample path for $t \in T$, and is a single possible path of $\{X_t\}$. When we observe a sample random variable (or a sample path) through computational generation or in empirical data, we call the resultant data a realisation of the random variable (or process).

We also make explicit the difference between the above concepts and sample statistics. The above use of 'sample' refers to a single possible outcome of stochastic objects. When working with realisations we may then calculate statistical information about the data. We

refer to this information as the sample statistics. For example, consider a set of data points $\{X_1, \ldots, X_n\}$. Then the value $\bar{X} = (X_1 + \cdots + X_n)/n$ is referred to as the sample mean. To test the theoretical framework of stochastic processes we can compare the theoretical statistics resulting from stochastic models to the sample statistics observed in the empirical data.

1.2.4 *Summary*

The work in section 1.2 offers a quick introduction to the probabilistic concepts required to understand the novel aspects of this thesis. Many of the concepts are only touched upon briefly in order to be able to use them later. For more information on the theoretical background, the referenced material offer a more in-depth discussion [27], [28], [29].

For the rest of the thesis, it suffices to understand the topics we have introduced, and to especially have an intuitive grasp of stochastic processes and the ability to move between continuous-time processes and discrete-time representations (and realisations).

We also note that after section 1.1 we discuss stochastic processes exclusively in terms of its evolution through time for a single sample. It is therefore notationally convenient to write the stochastic process as a function of time only, and not of the sample space. Therefore, we henceforth change the notation for a stochastic process from $\{X_t(\omega)\}$ to $\{X(t)\}$ for $\omega \in \Omega$, $t \in T$. We use $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ to instead denote a set of discrete realisatons or samples.

1.3 OUTLINE OF THESIS

In this chapter we have discussed the motivations and background information for the rest of this thesis. The basic probabilistic and statistical nomenclature and ideas have been discussed and introduced. This acts as the framework for the more complicated mathematical theory used subsequently.

In chapter 2 a series of key concepts required to investigate empirical stable processes are discussed. Hill's estimator is introduced as a tool for identifying the index of heavy-tailed distributions. The set of stable distributions is defined and the consequential failure of assigning an autocorrelation function to non-Gaussian stable processes is explained. Coherence is put forward as a non-trivial alternative measure of depency in a process to the autocorrelation for the set of stable distributions. Methods to generate non-stable processes with valid autocorrelation functions are introduced, with caveats of the methods explained. Established theory into the zero-crossings of Gaussian processes (a unique class of stable processes) is introduced, with McFadden's formula, Van-Vleck's theorem and Rice's formula all defined, and their role in calculating the zero-crossing statistics explained.

In chapter 3 we discuss the problematic issue of identifying the heavy-tail index, and why it is particularly important when investigating stable processes. We review Hill's estimator, and discuss the situations in which it is and isn't accurate. Specifically we investigate the difficulty in choosing the threshold value for heavy-tailed processes, the effect of autocorrelation on its error and its ineffectiveness for stable processes with indices close to the stable threshold. As an alternative, we introduce characteristic function methods. Using existing theory on stable parameter estimation, we consequently introduce a novel characteristic function method for estimating the heavy-tail index. All three methods (Hill's estimator, the stable characteristic function method and the general characteristic function method) are then tested on synthetic data and compared, allowing us to comment on optimum means for identifying the heavy-tail index. They are then used to investigate a real-world data set that is purpoted to be well modelled by a non-Gaussian stable distribution.

In chapter 4 a novel method for generating stable processes with memory is described. By extending a simple and ubiquitous method of generating a Cauchy random variable we are able to produce a Cauchy process with a non-trivial joint pdf. This is proposed as a candidate for generating Cauchy processes with memory. The empirical properties of the generated processes are investigated, and we show that coherence (as proposed in [19]) can quantify the resulting dependency.

In chapter 5 we directly approach the problem of investigating the zero-crossing statistics of stable processes. The theory used to investigate Gaussian stable processes is generalised to allow for the investigation of non-Gaussian stable processes. Generalised forms of Van-Vleck's theorem and Rice's formula are introduced, and defined in terms of coherence. This theory is then tested by comparing its predictions to empirical results produced by synthetic data, generated by the methods described in in chapter 4.

In chapter 6 we summarise the work of the thesis and highlight potential future avenues for research. This allows us to comment on the applicability of the coherence as a model for measuring dependency in stable processes and its use in predicting zero-crossing statistics, as proposed in [19].

2

EXISTING WORK

2.1 HILL'S ESTIMATOR

The work in this thesis is based around the tail-behaviour of random variables - its most extreme fluctuations. We are specifically interested in heavy-tailed random variables. The term 'heavy-tailed' can refer to a number of distributions and is often used interchangably with 'fat-tailed' or 'long-tailed'. In the context of this investigation, we define heavy-tailed as below. A random variable X is right heavy-tailed iff

$$\bar{F}_X(x) \overset{a}{\sim} |x|^{-\alpha}$$

as $x \to \infty$ where $\alpha > 0$ and $\overset{a}{\sim}$ denotes asymptotic equivalence [31] up to some positive constant. Similarly, a random variable is left heavy tailed iff $F_X(x) \overset{a}{\sim} |x|^{-\alpha}$ as $x \to -\infty$ where $\alpha > 0$. A random variable is (strictly) heavy-tailed iff it is both right and left heavy-tailed. The exponent α is henceforth called the heavy-tail index or α-tail index. It follows that a heavy-tailed random variable has pdf

$$f_X(x) \overset{a}{\sim} |x|^{-(1+\alpha)}$$

as $|x| \to \infty$. Heavy-tailed random variables only have a finite number of integer moments. This can be shown by noting that, for some $a > 0$,

$$\int_a^\infty x^{-k} dx < \infty$$

iff $k > 1$. Therefore, for a heavy-tailed process with heavy-tail index α, the nth integer moment μ_n is finite iff $\alpha > n$. For example, the third integer moment (skewness) is finite only when the heavy-tailed index $\alpha > 3$. This is true for central integer moments too.

Equivalently, a heavy-tailed random variable will have $\lceil \alpha \rceil - 1$ integer moments, where $\lceil \cdot \rceil$ is the ceiling function.

A heavy-tailed distribution can be easily identified by considering its complementary distribution function on a logarithmic scale. It is clear that

$$\log \bar{F}_X(x) \to -\alpha \log |x| + C \tag{2.1.1}$$

as $x \to \infty$ where $C > 0$ is some constant. Therefore a linear relationship would be observed with gradient $-\alpha$. It is therefore sufficient to say that we can propose a random variable is heavy-tailed if, for sufficiently large x, its cumulative distribution function is linear on a log-log plot. This also highlights a potential method to estimate the value of the heavy-tail index α. By fitting a linear model to Equation 2.1.1 (through least-squares regression or otherwise), the heavy-tail index α can be estimated from the gradient. This is referred to as the 'graphical method' using Log-Log Complementary Distribution (LLCD) plots [32]. The graphical method is problematic because for limited data sets volatility in the tail has an exaggerated effect on a logarithmic scale whilst errors in the central values are more difficult to scrutinise. It is sufficient for this method to be labelled too erronous for the purposes of this investigation.

In the context of this investigation, there is a need to be sufficiently accurate in estimating α. Firstly, it is important to clarify whether or not a heavy-tailed distribution is within the stable regime such that $\alpha \in (0,2)$. If it is in this range, then α is also required to quantify memory inherent in the process (in the form of the coherence function). Both of these points are expanded upon and subsequently explained in section 2.2. If $\alpha > 2$, an autocorrelation function exists and customary methods can be employed.

A popular method for accurately estimating the heavy-tail index was introduced by Hill in 1975 [22]. It assumes the tail of the distribution of the process is Pareto and fits a model to the data using Maximum Likelihood Estimation [8].

The derivation of Hill's estimator follows. Consider a set of n sample statistics $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ drawn from iid random variables with a Pareto pdf

$$f_X(x) = \begin{cases} kx^{-(1+\alpha)} & \text{if } x \geq x_0, \\ 0 & \text{if } x < x_0, \end{cases} \tag{2.1.2}$$

where $\alpha > 0$ is the heavy-tail index, $k > 0$ is a normalising constant and $x_0 > 0$ is some threshold value. We denote the order statistics of $\boldsymbol{X}$ by $\{X_{(1)}, \ldots, X_{(n)}\}$ such that $X_{(n)} \geq \ldots \geq X_{(1)}$. To fit the Pareto pdf in Equation 2.1.2 to the sample statistics, we set the threshold value to the first order statistic $x_0 = X_{(1)}$ resulting in the normalising constant $k = \alpha X_{(1)}^{\alpha}$. The log-likelihood function of α for $\{X_{(1)}, \ldots, X_{(n)}\}$ with pdf Equation 2.1.2 is

$$\begin{aligned} \mathcal{L}(\alpha) &= \log \prod_{j=1}^{n} \frac{\alpha}{X_{(1)}^{-\alpha}} X_{(j)}^{-(1+\alpha)} \\ &= \log \prod_{j=1}^{n} \frac{\alpha}{X_{(1)}} \left(\frac{X_{(j)}}{X_{(1)}} \right)^{-(1+\alpha)} \\ &= \sum_{j=1}^{n} \left[\log \left(\frac{\alpha}{X_{(1)}} \right) - (1+\alpha) \log \left(\frac{X_{(j)}}{X_{(1)}} \right) \right] \\ &= n \log \left(\frac{\alpha}{X_{(1)}} \right) - (1+\alpha) \sum_{j=1}^{n} \log \left(\frac{X_{(j)}}{X_{(1)}} \right). \end{aligned} \tag{2.1.3}$$

The optimal estimate for α is the Hill estimate $\hat{\alpha}_H$ for the heavy-tail index and occurs when Equation 2.1.3 is maximised. This occurs when $\mathcal{L}'(\hat{\alpha}_H) = 0$, such that

$$\begin{aligned} \mathcal{L}'(\hat{\alpha}_H) &= \frac{n}{\hat{\alpha}_H} - \sum_{j=1}^{n} \log \left(\frac{X_{(j)}}{X_{(1)}} \right) = 0 \\ \implies \hat{\alpha}_H &= \left[\frac{1}{n} \sum_{j=1}^{n} \log \left(\frac{X_{(j)}}{X_{(1)}} \right) \right]^{-1}. \end{aligned} \tag{2.1.4}$$

It can be shown that for a set of n realisations drawn from a set of iid rv's with a Pareto pdf $\hat{\alpha}_H \to \alpha$ as $n \to \infty$ almost surely [33]. Whilst the derivation of Hill's estimator is based upon a Pareto pdf, its use as an estimator for heavy-tailed processes is implemented by considering upper order statistics such that $\hat{\alpha}_H$ is based on sample statistics that behave according to the Pareto pdf. The methodology of selecting the upper order statistics is discussed later in section 3.1 and ensures Hill's estimator is therefore prevalent in the study of tail behaviour, such as in [32], [34].

2.2 STABLE PROCESSES AND THE COHERENCE FUNCTION

2.2.1 *The set of stable distributions*

The role of heavy-tailed distributions and the subsequent failure of integer moments prompts the need to introduce and define the set of stable distributions. A random variable is defined as stable iff the sum of independent copies of it has the same distribution as the random variable (up to scale and location parameters) [35]. Equivalently; a random variable X with independent and identical copies $X_1,\ X_2$ is stable iff $\forall\ a,b>0,\ \exists\ c>0,\ d\in\mathbb{R}$ such that

$$aX_1+bX_2\overset{d}{=}cX+d \tag{2.2.1}$$

where $\overset{d}{=}$ indicates equivalence in distribution. The set of distributions that satisfy Equation 2.2.1 can not be expressed analytically in terms of a distribution function or pdf (except for a limited number of examples). It can, however, be represented analytically in terms of the characteristic function

$$\phi(t;\alpha,\beta,\gamma,\mu)=\exp\left[it\mu-|\gamma t|^{\alpha}\left(1-i\beta\,\mathrm{sgn}(t)\psi(t)\right)\right] \tag{2.2.2}$$

where

$$\psi(t)=\begin{cases}\tan\frac{\pi\alpha}{2} & \text{if } \alpha\neq 1,\\ -\frac{2}{\pi}\log|t| & \text{if } \alpha=1,\end{cases}$$

i is the imaginary unit, $\alpha\in(0,2]$ is the stability parameter, $\beta\in[-1,1]$ is the skewness parameter, $\gamma>0$ is the scale parameter and $\mu\in\mathbb{R}$ is the location parameter. We denote a stable distribution by $\mathcal{S}(\alpha,\beta,\gamma,\mu)$ such that a random variable $X\sim\mathcal{S}(\alpha,\beta,\gamma,\mu)$ iff X has the characteristic function in Equation 2.2.2. In our investigation it suffices to investigate symmetric stable random variables which are defined in terms of the characteristic function

$$\phi(t)=\phi(t;\alpha,0,\gamma,0)=\exp\left(-|\gamma t|^{\alpha}\right). \tag{2.2.3}$$

We denote a symmetric stable distribution by $\mathcal{S}(\alpha,\gamma)$ such that a random variable $X\sim\mathcal{S}(\alpha,\gamma)=\mathcal{S}(\alpha,0,\gamma,0)$ iff X has the characteristic function in Equation 2.2.3.

There are a limited number of examples of stable random variables whose distribution functions can be expressed analytically. This includes the set of Gaussian random variables $\mathcal{N}(\mu,\sigma^2) = \mathcal{S}(2,0,\sigma/\sqrt{2},\mu)$, the set of Cauchy random variables $\mathcal{C}(\mu,\gamma) = \mathcal{S}(1,0,\gamma,\mu)$ and the set of Lévy random variables Lévy$(\mu,\gamma) = \mathcal{S}(1/2,1,\gamma,\mu)$ with pdf's

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$
$$f_X(x) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-\mu)^2+\gamma^2}\right],$$
$$f_X(x) = \sqrt{\frac{\gamma}{2\pi}}\,\frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\mu)}\right),$$

respectively. Despite stable distributions not necessarily having analytic expressions universally, their tail behaviour can be expressed analytically. For the set of non-Gaussian stable distributions (i.e. $\alpha \in (0,2)$), the pdf is

$$f_X(x) \overset{a}{\sim} \alpha\gamma^\alpha \sin\left(\frac{\pi\alpha}{2}\right)\Gamma(\alpha)\pi^{-1}(1+\beta)|x|^{-(1+\alpha)}$$

as $|x| \to \infty$ where Γ is the Gamma function [36]. Similarly, the complementary distribution is

$$\bar{F}_X(x) \overset{a}{\sim} \gamma^\alpha \sin\left(\frac{\pi\alpha}{2}\right)\Gamma(\alpha)\pi^{-1}(1+\beta)|x|^{-\alpha}$$

as $x \to \infty$. As discussed in section 2.1, this affects the integer moments. For all non-Gaussian stable random variables, the variance $\sigma^2 = \langle (X-\mu_1)^2\rangle$ is undefined. For stable random variables with $\alpha \leq 1$ (such as the Cauchy distribution), the mean $\mu_1 = \langle X\rangle$ is also undefined. By similar arguments, the autocovariance and autocorrelation of a non-Gaussian stable process are also undefined. This is problematic as it means a stable process' memory can not be expressed in terms of these traditional functions.

The prevalence and importance of the stable distributions arise from their self-similarity (as defined in Equation 2.2.1) and their role in the Central Limit Theorem (CLT), described subsequently. Let X_1, X_2, ... be iid random variables with finite mean μ and finite variance σ^2. Define

$$S_n := X_1 + X_2 + \cdots + X_n$$

and $A_n := S_n/n$. Then the (standard) Central Limit Theorem states that as $n \to \infty$, $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$ approximately and $A_n \sim \mathcal{N}(\mu, \sigma^2/n)$ approximately. This is equivalent to

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$
$$(A_n - \mu)\sqrt{n} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

as $n \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution. This form of the CLT breaks down however when considering heavy-tailed random variables within the stable regime - those whose heavy-tailed exponent is in the range $\alpha \in (0, 2)$. In these instances, the variance (and sometimes mean) is undefined. The Generalised Central Limit Theorem (GCLT) states that for some set of iid random variables with heavy-tails $f_X(x) \overset{a}{\sim} |x|^{-(1+\alpha)}$, S_n and A_n (as defined previously) converge such that

$$\frac{S_n - n\mu}{n^{1/\alpha}} \xrightarrow{d} \mathcal{S}(\alpha, 0, \gamma, 0),$$
$$(A_n - \mu)n^{(\alpha-1)/\alpha} \xrightarrow{d} \mathcal{S}(\alpha, 0, \gamma, 0)$$

as $n \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution and μ is the location parameter, which is equal to the mean iff $\alpha > 1$ [37].

2.2.2 *Coherence*

Due to their heavy-tails, the autocorrelation of non-Gaussian stable processes is undefined and therefore an inappropriate measure of their dependency structures. An alternative is therefore required, and such a candidate is proposed in [19]. It constructs a joint characteristic function for a simple symmetric stable process of the form

$$\phi(t_1, t_2) = \exp\left(\frac{-A}{1 + r^\alpha}\left(|t_1 + rt_2|^\alpha + |rt_1 + t_2|^\alpha\right)\right) \tag{2.2.4}$$

where $A > 0$ is some scaling constant (having the same role as γ in Equation 2.2.2), $\alpha \in (0, 2]$ is the stability index and $r \in [-1, 1]$ is some parameter that quantifies the dependence (in a role similar to the autocorrelation). The parameter r is termed the coherence and is proposed as a means to quantify memory for a stable process. To calculate the coherence for some time lag τ, it is first noted that although the integer moments of stable processes

do not exist (except the mean in some cases), a range of fractional moments do exist. Let $\{X(t)\}$ be a non-Gaussian stable process with the joint characteristic function in Equation 2.2.4 st $\alpha \in (0,2)$. Then the quantity $\langle |X(t)|^\gamma \rangle$ is finite for $-1 < \gamma < \alpha$. Furthermore, the structure function is defined by

$$S_\gamma(\tau) := \langle |X(t) - X(t+\tau)|^\gamma \rangle \tag{2.2.5}$$

and is again finite for $-1 < \gamma < \alpha$. It is shown in [19] that for a symmetric stable process with the joint characteristic function in Equation 2.2.4 the normalised structure function

$$s_\gamma(\tau) := \frac{S_\gamma(\tau)}{\langle |X(t)|^\gamma \rangle} \tag{2.2.6}$$

is related to the coherence r by

$$s_\gamma(\tau) = \left(\frac{2(1-r)^\alpha}{1+r^\alpha} \right)^{\gamma/\alpha}. \tag{2.2.7}$$

This means the coherence can be calculated for any time lag τ for some empirical process by calculating Equation 2.2.6 and inverting Equation 2.2.7. It is important to recognise how Equation 2.2.4 generalises for the Gaussian case to get an intuitive understanding of the coherence, especially in the context of autocorrelation. For $\alpha = 2$ the stable process is Gaussian, and Equation 2.2.4 reduces to

$$\phi(t_1, t_2) = \exp\left(-A \left\{ t_1^2 + t_2^2 + 2\left(\frac{2r}{1+r^2} \right) t_1 t_2 \right\} \right)$$

which is equal to the joint characteristic function of a Gaussian process with correlation

$$\rho = \frac{2r}{1+r^2}.$$

We therefore see that coherence quantifies the memory in a similar manner to autocorrelation, but is not equal to it. Let $\{X(t)\}$ be a stationary stable process st $X(t) \sim \mathcal{S}(\alpha, \gamma)$. Suppose that $\{X(t)\}$ has the joint characteristic function in Equation 2.2.4. Then $\{X(t)\}$ is a set of iid rv's iff $r = 0$ and a set of dependent rv's if $r \neq 0$. For Gaussian and non-stable processes, we could describe the dependency structures by 'white noise' or 'coloured noise'. Unlike those processes, non-Gaussian stable processes have an undefined autocorrelation which result in an undefined power-spectrum. Because of this, it becomes important to introduce a new term to distinguish non-Gaussian stable processes.

Let $\{X_t\}$ be a stationary non-Gaussian stable process with identically distributed random variables. Then $\{X(t)\}$ is a coherent noise iff its elements are not independent. For convenience, if the elements of $\{X(t)\}$ are independent, then it may be called a non-Gaussian stable iid noise.

2.3 FOURIER FILTERING TO GENERATE CORRELATED PROCESSES

2.3.1 *The Fourier Filtering Method (FFM)*

In order to properly investigate stochastic processes and test heavy-tail estimation methods, we introduce a means to generate stochastic processes with certain input properties. It is first shown how to generate a Gaussian coloured noise (set of correlated Gaussian rv's) using the Fourier Filtering Method (FFM) [38], [39]. Let

$$\boldsymbol{X} = \{X_1,\ X_2,\ \ldots,\ X_n\} \tag{2.3.1}$$

be a set of n iid Gaussian random variables. Let $a(\tau)$ be the input autocorrelation to be evaluated at non-negative integer time lags $\tau = 0,\ 1,\ \ldots,\ n-1$. The evaluation of the autocorrelations is expressed as an n-dimensional vector such that

$$\boldsymbol{A} = \{A_1,\ A_2,\ \ldots,\ A_n\} = \{a(0),\ a(1),\ \ldots,\ a(n-1)\}.$$

The Fourier-transforms of $\boldsymbol{X}$ and $\boldsymbol{A}$ are denoted $\boldsymbol{X}^* = \mathcal{F}[\boldsymbol{X}] = \{X_1^*,\ \ldots,\ X_n^*\}$ and $\boldsymbol{A}^* = \mathcal{F}[\boldsymbol{A}] = \{A_1^*,\ \ldots,\ A_n^*\}$. A new quantity $\boldsymbol{Y}^*$ is defined

$$\boldsymbol{Y}^* := \boldsymbol{X}^* \sqrt{|\boldsymbol{A}^*|} \tag{2.3.2}$$

such that $\boldsymbol{Y}^* = \{Y_1^*,\ \ldots,\ Y_n^*\}$ and $Y_j^* = X_j^* \sqrt{\left|A_j^*\right|}$. Through the inverse Fourier-transform of Equation 2.3.2, we define

$$\boldsymbol{Y} := \mathcal{F}^{-1}[\boldsymbol{Y}^*]. \tag{2.3.3}$$

Then $\boldsymbol{Y}$ is a process with distribution approximately Gaussian and autocorrelation approximately $a(\tau)$. If the sample autocorrelation function is denoted $a_Y(\tau)$ and the sample distribution function $F_Y(x)$, then it can be shown that $F_Y(x) \to F_X(x)$ and $a_Y(\tau) \to a(\tau)$ as $n \to \infty$. This is tested with an input autocorrelation function

$$a(\tau) = \exp(-k\tau^2)$$

with $k = 10^{-3}$. The distribution of the input white noise $\boldsymbol{X}$ is compared with the distribution of the output coloured noise $\boldsymbol{Y}$ for different sample sizes n in Figure 2.3.1. The sample autocorrelations $a_Y(\tau)$ are plotted alongside the theoretical autocorrelation input $a(\tau)$ for different sample sizes n in Figure 2.3.2. These plots successfully display evidence for the empirical convergence of both the distribution and autocorrelation as $n \to \infty$. It is noted that statistical tests (such as the Kolmogorov-Smirnov test) could also be used for affirmation.

It is accepted that the FFM is an effective means of generating a Gaussian coloured noise with a desired autocorrelation (provided the sample size is sufficiently large). The FFM does not work for non-Gaussian coloured noises, however. To show this, two non-Gaussian distribution functions are introduced

$$f_X(x) = \begin{cases} 1 & \text{if } x \in [0,1], \\ 0 & \text{otherwise;} \end{cases} \tag{2.3.4}$$

$$f_X(x) = \frac{1}{2(1+x^2)^2} \overset{a}{\sim} x^{-4}. \tag{2.3.5}$$

Equation 2.3.4 is the distribution of a uniform random variable $X \sim \mathcal{U}[0,1]$ and Equation 2.3.5 is an example of a heavy-tailed distribution outside the stable regime (i.e. one with a defined autocorrelation function). Sufficiently large realisations ($n = 2^{18}$) of iid random variables with the aforementioned distributions are used as the generating processes instead of Gaussian white noises in Equation 2.3.1. The FFM is then used to generate coloured noises. The resultant process' sample distribution function is plotted alongside the input distribution and a Gaussian fit in Figure 2.3.3. Whilst the FFM provides a means to generate a Gaussian coloured noise with some input autocorrelation function, it fails to produce satisfactory results when the input distribution function is non-Gaussian and, instead, produces another Gaussian coloured noise.

2.3.2 *The iterative Fourier Filtering Method (iFFM)*

To counter this, surrogate data methods and the consequential generation methods are introduced. Surrogate data testing is a means to test for non-linearity in empirical processes

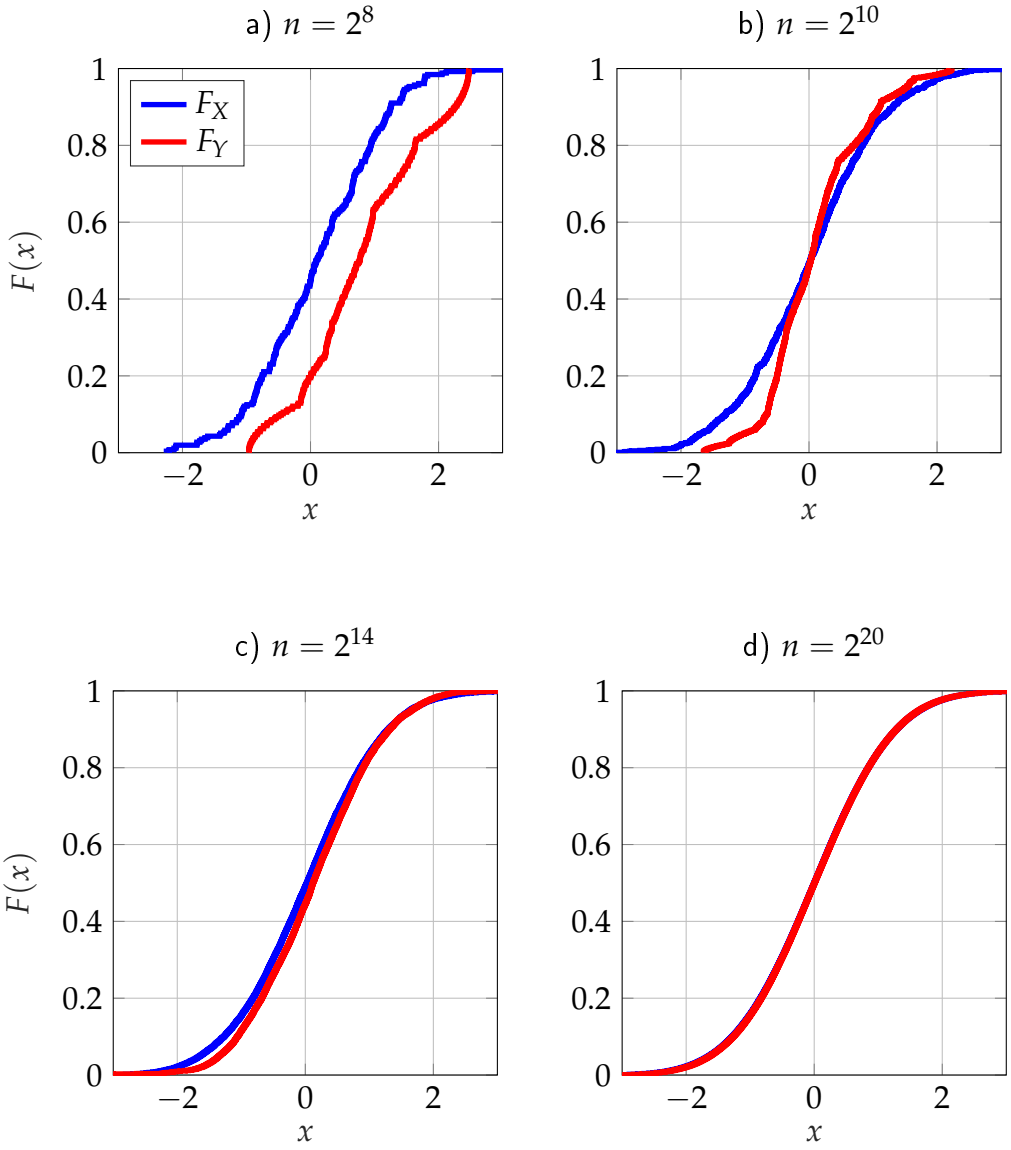

Figure 2.3.1.: Distribution functions of realisations of Gaussian coloured noise $\boldsymbol{Y}$ (red) compared with distribution of input of Gaussian white noise $\boldsymbol{X}$ (blue), for different sample sizes n. Generated with autocorrelation function $a(\tau) = \exp(-10^{-3}\tau^2)$.

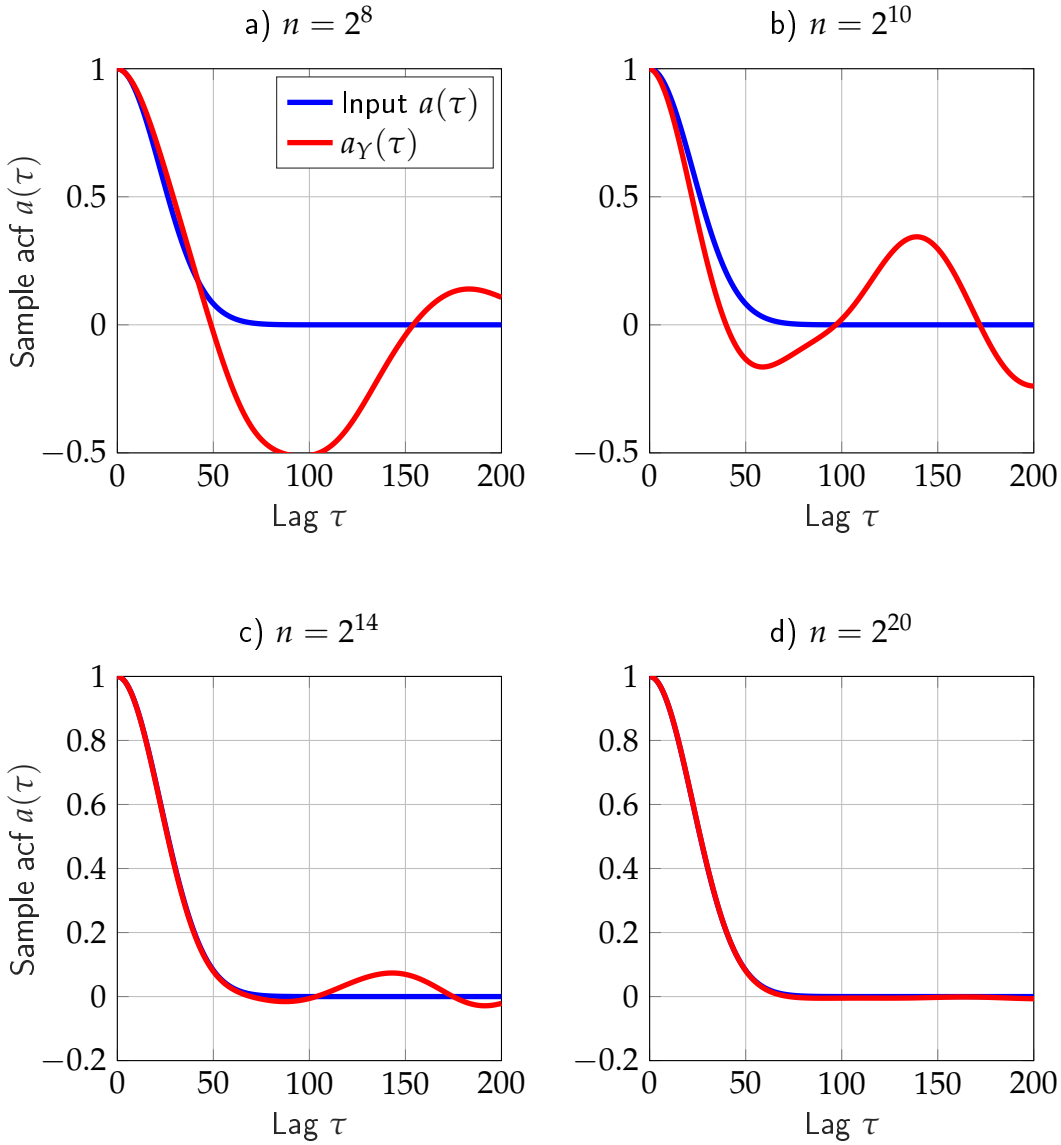

Figure 2.3.2.: Autocorrelation functions of realisations of Gaussian coloured noise $\boldsymbol{Y}$ (red) compared with input autocorrelation function $a(\tau) = \exp(-10^{-3}\tau^2)$, for different sample sizes n.

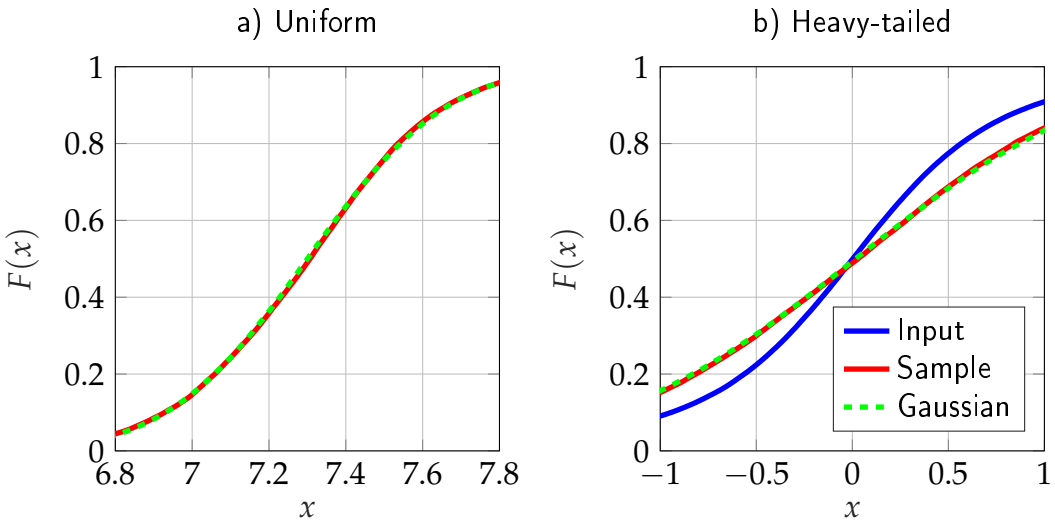

Figure 2.3.3.: Comparison between input and sample distributions for non-Gaussian processes using the FFM. Plotted for a) a uniform process (on $[0,1]$) with pdf Equation 2.3.4, and b) a heavy-tailed process with pdf Equation 2.3.5. In both cases, the sample distribution (red) is not equal to the input distribution (blue), and is instead Gaussian (green). For the uniform case, the input distribution is out of sight.

[40]. The key tool to testing for non-linearity is the generation of pseudorandom copies of the original process to be tested (called surrogate data). This surrogate data maintains the original distribution and an approximate of its memory (autocorrelation function) and is produced using the Amplitude Adjusted Fourier Transform (AAFT) method [40].

The AAFT method works as follows. Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be an empirical data set to be 'copied'. We first define the rank of $\boldsymbol{X}$ by $\mathcal{R}(\boldsymbol{X}) := \{\mathrm{rank}(X_1), \ldots, \mathrm{rank}(X_n)\}$ such that $\mathrm{rank}(X_j) = k \in \mathbb{N}$ denotes that X_j is the kth largest element in $\boldsymbol{X}$. A new data set $\boldsymbol{Y}$ is drawn from a set of n iid Gaussian random variables and ordered such that $\mathcal{R}(\boldsymbol{Y}) = \mathcal{R}(\boldsymbol{X})$. Then $\boldsymbol{Y}$ is a Gaussian process that 'follows' $\boldsymbol{X}$ in its evolution. This therefore means we can create a surrogate data set of $\boldsymbol{Y}$. To do so, we define its Fourier transform by $\boldsymbol{Y}^* := \mathcal{F}[\boldsymbol{Y}] = \{Y_1^*, \ldots, Y_n^*\}$. The phases of $\boldsymbol{Y}^*$ are randomised such that $\boldsymbol{Y}^* \to \boldsymbol{Y}^*_{\text{new}} = \{|Y_1^*| \exp(i\theta_1'), \ldots, |Y_n^*| \exp(i\theta_n')\}$ where $\theta_j' \sim \mathcal{U}[0, 2\pi]$. Then the resultant process $\boldsymbol{Y}_{\text{new}} = \mathcal{F}^{-1}[\boldsymbol{Y}^*_{\text{new}}]$ is a random copy of $\boldsymbol{Y}$ with matching distribution and power-spectrum. Finally the elements of $\boldsymbol{X}$ are rearranged to form some new process $\boldsymbol{X}_{\text{new}}$ such that $\mathcal{R}(\boldsymbol{X}_{\text{new}}) = \mathcal{R}(\boldsymbol{Y}_{\text{new}})$. Then the above procedure from $\boldsymbol{X} \to \boldsymbol{X}_{\text{new}}$ is the AAFT method for creating a surrogate data set.

The AAFT method works asymptotically as $n \to \infty$ but can experience problems for smaller data sets. Whilst the resultant process will preserve the distribution by construction, the resultant process can have a different power-spectrum due to the rank-replacement stage. This can render the resulting autocorrelation significantly different from the original. To remedy this, the iterative Amplitude Adjusted Fourier Transform (iAAFT) method is introduced [41].

For the iAAFT method we again let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be an empirical data set to be 'copied', and denote its Fourier transform $\boldsymbol{X}^* := \mathcal{F}[\boldsymbol{X}] = \{X_1^*, \ldots, X_n^*\}$. The power-spectrum of $\boldsymbol{X}$ is stored by $\boldsymbol{S} := \{|X_1^*|^2, \ldots, |X_n^*|^2\}$. We define $\boldsymbol{X}_{(0)}$ as a random shuffle of $\boldsymbol{X}$ and denote its respective Fourier-transform $\boldsymbol{X}^*_{(0)}$. The iAAFT is based on two main iterative steps:

1. The correct power-spectrum is enforced by $\boldsymbol{X}^*_{(j)} \to \boldsymbol{X}^*_{(j+1)} := \boldsymbol{X}^*_{(j)}\sqrt{\boldsymbol{S}}/|\boldsymbol{X}^*_{(j)}|$. Define $\boldsymbol{X}'_{(j+1)} := \mathcal{F}^{-1}[\boldsymbol{X}^*_{(j+1)}]$.
2. The correct distribution is enforced by rank-replacing the elements of $\boldsymbol{X}'_{(j+1)}$ with those of $\boldsymbol{X}$, and defining the resultant process $\boldsymbol{X}_{(j+1)}$. Equivalently, $\boldsymbol{X}_{(j+1)}$ is the rearrangement of $\boldsymbol{X}$ such that $\mathcal{R}(\boldsymbol{X}_{(j+1)}) = \mathcal{R}(\boldsymbol{X}'_{(j+1)})$.

By cycling through these steps from $\boldsymbol{X}_{(j)} \to \boldsymbol{X}_{(j+1)}$ a pseudorandom copy of $\boldsymbol{X}$ is created with matching distribution and convergent power-spectrum as $j \to \infty$.

Whilst the iAAFT method is constructed specifically to generate surrogate data sets from an empirical data set in order to investigate non-linearity, its methodology can be adapted to generate a realisation of a process with some target distribution and target autocorrelation, in the form of the iterative Fourier Filtering Method (iFFM) [42], which we describe below.

We first generate a realisation of a Gaussian coloured noise $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ with the desired autocorrelation $a(\tau)$ using the Fourier Filtering Method (or otherwise). Its Fourier transform is denoted $\boldsymbol{X}^* := \mathcal{F}[\boldsymbol{X}] = \{X^*_1, \ldots, X^*_n\}$ and its power-spectrum is stored by $\boldsymbol{S} := \{|X^*_1|^2, \ldots, |X^*_n|^2\}$. A set of realisations drawn from a set of iid random variables with the desired distribution is generated and denoted $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$. Let $\boldsymbol{Y}_{(0)} := \boldsymbol{Y}$. The iFFM is based on two main iterative steps:

1. Let $\boldsymbol{Y}^*_{(j)} := \mathcal{F}[\boldsymbol{Y}_{(j)}]$. Then $\boldsymbol{Y}_{(j)}$ is brought to the correct power-spectrum by $\boldsymbol{Y}^*_{(j)} \to \boldsymbol{Y}^*_{(j+1)} := \boldsymbol{Y}^*_{(j)}\sqrt{\boldsymbol{S}}/|\boldsymbol{Y}^*_{(j)}|$.
2. Let $\boldsymbol{Y}'_{(j+1)} := \mathcal{F}^{-1}[\boldsymbol{Y}^*_{(j+1)}]$. Then the correct distribution is enforced by rank-replacing the elements of $\boldsymbol{Y}'_{(j+1)}$ with those of $\boldsymbol{Y}$, and defining the resultant set $\boldsymbol{Y}_{(j+1)}$. Equivalently, $\boldsymbol{Y}_{(j+1)}$ is the rearrangement of $\boldsymbol{Y}$ such that $\mathcal{R}(\boldsymbol{Y}_{(j+1)}) = \mathcal{R}(\boldsymbol{Y}'_{(j+1)})$.

By repeating the above steps, the resultant data set $\boldsymbol{Y}_{(j)}$ has the distribution of $\boldsymbol{Y}$ whilst its autocorrelation converges to $a(\tau)$ as $j \to \infty$, as required.

The success of the iFFM can be determined by testing it empirically. To do so, the iFFM is used to generate a realisation with autocorrelation function $a(\tau) = \exp(-k_a\tau^2)$ with

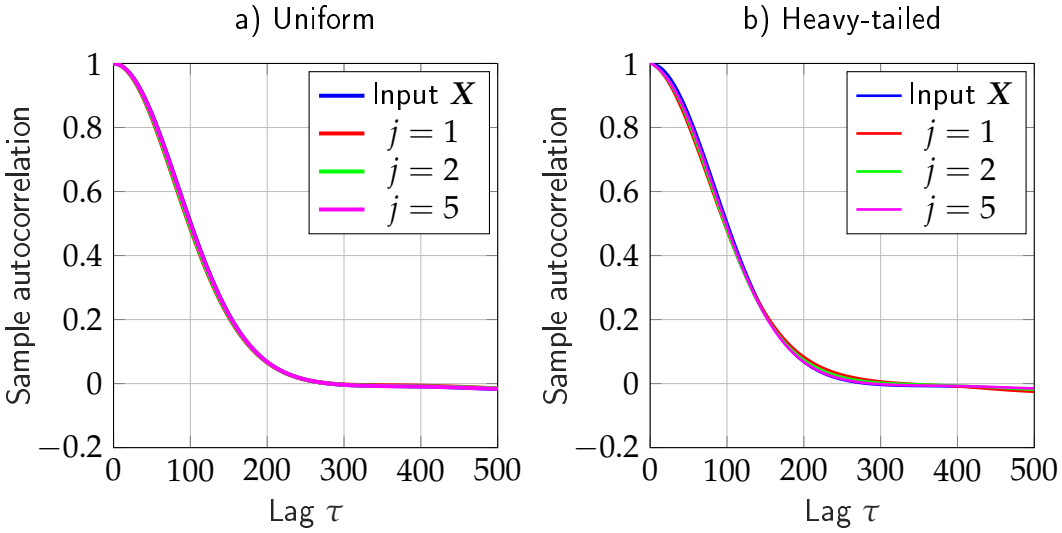

Figure 2.3.4.: Sample autocorrelations of non-Gaussian coloured noise produced by the iFFM. Several iterations plotted for realisations of processes with a) a uniform pdf (Equation 2.3.4), and b) a heavy-tailed pdf (Equation 2.3.5). In both cases, a single iteration is satisfactory, but repeated iterations do improve the results.

$k_a = \log(2)/100^2$ and a uniform distribution and heavy-tailed distribution as used previously (seen in Equation 2.3.4 and Equation 2.3.5 respectively). It is noted that during each loop of the iFFM, step 2 imprints the precise distribution function onto the output realisation through rank-replacement of the original generating realisation. The result is that the output realisation's distribution function is equal to the input distribution function, and its convergence is therefore a trivial measure of success. Instead, the determining factor of the success of the iFFM is the accuracy of the output's autocorrelation function. The method is tested on realisations with $n = 2^{18}$ elements and looped $j = 1,\ 2,\ 5$ times, with the results plotted in Figure 2.3.4.

From the results, it is clear that the iFFM is an effective method. In both example distributions, it is arguable that a single loop is enough to produce a data set with sufficient properties - although it is important to note that subsequent iterations do offer improvements. The requirement for the iteration is discussed in [42], where the iterations are particularly important (and effective) in converging the output sample path's autocorrela-

tion to the desired long-term correlations. It is important to note that there must exist some point of saturation that will eventually be reached. At this point, bringing the resultant realisation to the original power-spectrum will not perturb the system enough to enforce a change in rank of the variables. When this point is reached, there will be no more iterative progress by the algorithm. In the context of this investigation, it suffices to understand the iFFM and its viability for a limited number of iterations.

Before concluding, it is worth acknowledging a clear shortcoming in Fourier Filtering for non-Gaussian processes. Whilst the joint density function for correlated Gaussian random variables is fully encompassed by the marginal density functions and correlation, this is not true for non-Gaussian random variables. For non-Gaussian processes, an autocorrelation function and a marginal pdf do not uniquely characterise the joint density of the process. Hence the methods discussed in this chapter allow us to generate processes with a desired marginal density function and autocorrelation, but do not guarantee the form of the joint density function. Whilst this does not affect the validity of the testing in this investigation, it is still a point of consideration when using the iFFM.

2.4 ZERO-CROSSING STATISTICS OF GAUSSIAN PROCESSES

The central aim of this thesis is to investigate the zero-crossing statistics of coherent stable noises. Whilst this is a novel concept for non-Gaussian stable processes, there has been considerable work on the crossing statistics of Gaussian coloured noises. Here we describe the key work into these processes, which is later generalised to investigate the crossing statistics of non-Gaussian stable processes. Zero-crossings are an important mathematical concept, as they describe the excursions of a random signal from its origin. Special importance is attached to understanding the probability that a signal will return to zero within a certain time interval.

It is first important to define zero-crossings. Let $\{X(t)\}$ be a stationary continuous stochastic process. Then $X(t_0)$ is a zero-crossing with crossing time t_0 iff $\{X(t)\}$ changes sign at $t = t_0$. Let $X(t_1)$, $X(t_2)$ denote successive zero-crossings of $\{X(t)\}$ at times t_1,

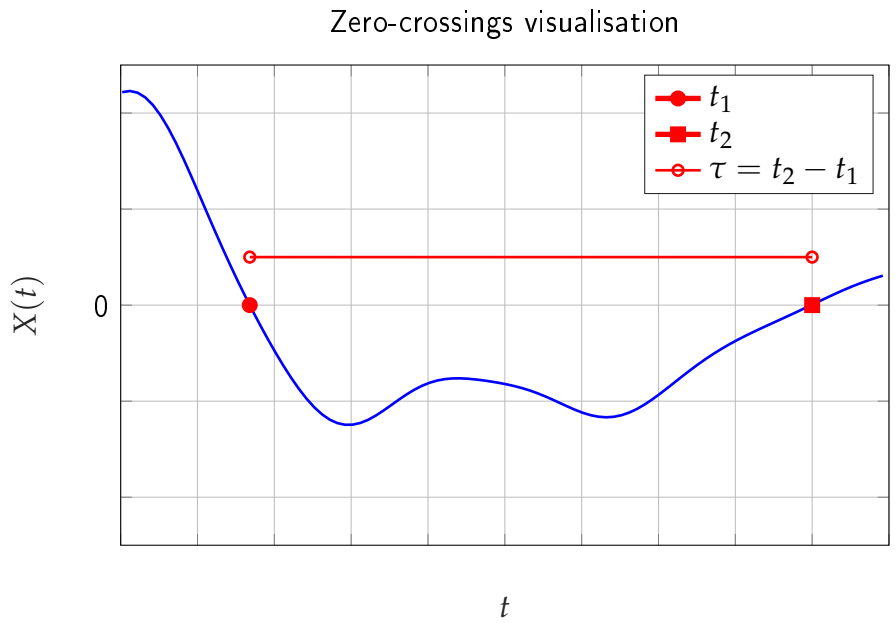

Figure 2.4.1.: A visualisation of a continuous stochastic process with its zero-crossing information highlighted.

t_2. Then the return time between these events is $\tau = t_2 - t_1$. The zero-crossings of a process are visualised in Figure 2.4.1. Although the mathematical framework regarding zero-crossings is based upon continuous stochastic processes, any empirical investigation into stochastic processes is limited to considering discrete data sets. We therefore also define zero-crossings in terms of discrete-time stochastic processes and their realisations. Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a discrete-time stochastic process. Then X_t is a zero-crossing with crossing-time t iff $\text{sgn}(X_t) \neq \text{sgn}(X_{t-1})$. Similarly; if X_{t_1}, X_{t_2} are successive zero-crossings then the return-time between these events is $\tau = t_2 - t_1$.

2.4.1 *McFadden's formula*

Let $\{X(t)\}$ be a differentiable stationary stochastic process with zero mean. In order to investigate the zero-crossings and return-times of $\{X(t)\}$, we introduce the clipped function $T(t)$, defined by

$$T(t) := \begin{cases} 1 & X(t) \geq 0, \\ -1 & X(t) < 0. \end{cases}$$

The clipped function is useful as, whilst it loses a lot of information about the original process $\{X(t)\}$, it has the same zero-crossings. The precise relevance of its use becomes clear in due course. Clearly, as $T(t)$ is bounded on $[-1,1]$, its autocorrelation is well-defined regardless of the underlying probability distribution of $X(t)$. We can therefore define the autocorrelation function of T as $R(\tau) := \langle T(t)T(t+\tau)\rangle$. This is referred to as the clipped autocorrelation of $\{X(t)\}$.

Let $p(n,\tau)$ describe the probability of finding exactly n zeros in the interval $(t,t+\tau)$. As $T(t) \in \{-1,1\}$, it follows that the product $T(t)T(t+\tau)$ is 1 if there are an even number of zero-crossings in $(t,t+\tau)$, and -1 if there are an odd number of crossings in that interval. It follows that the expected value of $T(t)T(t+\tau)$, and therefore the autocorrelation of the clipped function, can be expressed in terms of $p(n,\tau)$ by

$$R(\tau) = \langle T(t)T(t+\tau)\rangle = \sum_{n=0}^{\infty} (-1)^n \, p(n,\tau), \tag{2.4.1}$$

as first introduced by Rice [14]. The central aim of zero-crossing analysis is to describe the probability of a zero-crossing occuring in a prescribed epoch and the distribution of the intervals between these zero-crossings. To formalise this, let τ be an absolutely continuous random variable, denoting the time between two successive zero-crossings. We denote the probability density function of the zero-crossing of τ as $P_0(\tau)$, such that the probability that the return-time occurs between $[\tau_1,\tau_2)$ is

$$\mathbb{P}\left(\tau_1 \leq \tau < \tau_2\right) = \int_{\tau_1}^{\tau_2} P_0(\tau) d\tau.$$

Similarly $P_n(\tau)$, for $n = 1, 2, \ldots$, are defined as the probability density functions of crossing intervals between the mth and $(m+n+1)$th zeros for $m = 1, 2, \ldots$. Equivalently, $P_n(\tau)\, d\tau$ is the probability that the $(n+1)$th zero after time t occurs between $(t+\tau)$ and $(t+\tau+d\tau)$.

Let $\beta = 1/\langle\tau\rangle$ denote the mean crossing-rate. Then McFadden [17] showed that $p(n,\tau)$ could be related to $P_n(\tau)$ via the recursion relationships

$$p''(0,\tau) = \beta P_0(\tau), \tag{2.4.2}$$

$$p''(1,\tau) = \beta\left[P_1(\tau) - 2P_0(\tau)\right], \tag{2.4.3}$$

$$p''(n,\tau) = \beta\left[P_n(\tau) - 2P_{n-1}(\tau) + P_{n-2}(\tau)\right], \; n \geq 2, \tag{2.4.4}$$

where $p''(n,\tau)$ denotes the 2nd derivative of $p(n,\tau)$ with respect to τ. By taking the 2nd derivative of Equation 2.4.1 with respect to τ and substituting the recursion relationships in Equation 2.4.2, Equation 2.4.3 and Equation 2.4.4, the autocorrelation of the clipped function is related to $P_n(\tau)$ by

$$\frac{R''(\tau)}{4\beta} = \sum_{n=0}^{\infty} (-1)^n P_n(\tau). \tag{2.4.5}$$

This is referred to as McFadden's formula. From Equation 2.4.5, it is concluded that to accurately analyse the zero-crossings, and to analyse the zero-crossing density $P_0(\tau)$, both the mean crossing-rate β and the clipped autocorrelation function $R(\tau)$ are required, in addition to some model that connects $P_n(\tau)$ to $P_0(\tau)$, or some assumptions about the nature of $P_n(\tau)$ for $n \geq 1$.

In order to solve Equation 2.4.5 and calculate $P_0(\tau)$, McFadden used Laplace transforms. Let $h(t)$ be some function. Then the Laplace transform of $h(t)$ is defined

$$H(s) = \mathcal{L}\left[h(t)\right] := \int_0^{\infty} \exp(-st)\, h(t)\, dt \tag{2.4.6}$$

where s is the Laplace transform variable. Similarly, the inverse Laplace transform of $H(s)$ is some function $h(t) = \mathcal{L}^{-1}[H(s)]$ that satisfies Equation 2.4.6. By taking the Laplace transform of Equation 2.4.5, we introduce the quantities

$$g(s) := \mathcal{L}\left[\frac{R''(\tau)}{4\beta}\right],$$

$$p_n(s) := \mathcal{L}\left[P_n(\tau)\right],$$

so that Equation 2.4.5 is

$$g(s) = \sum_{n=0}^{\infty} (-1)^n p_n(s).$$

One advantage in using Laplace transforms comes when assuming that successive zero-crossing intervals are independent. By assuming independence, the probability density of the time between the mth and $m+n+1$th zero-crossing is equal to the probability density of $(n+1)\tau$. It follows that $P_n = P_0 \overset{n+1}{*} P_0$ where $\overset{n+1}{*}$ indicates the convolution operation repeated $n+1$ times. By the properties of Laplace transforms $\mathcal{L}[P_0 \overset{n+1}{*} P_0] = \mathcal{L}[P_0]^{n+1}$ and therefore the nth probability density in Laplace space can be expressed in terms of the 0th probability density by $p_n(s) = p_0^{n+1}(s)$. It follows that

$$g(s) = p_0(s) - p_0^2(s) + p_0^3(s) - \cdots = p_0(s)(1 - g(s)) \tag{2.4.7}$$

and therefore

$$p_0(s) = \frac{g(s)}{1 - g(s)}. \tag{2.4.8}$$

It follows that the probability density function of τ can be calculated via the inverse Laplace transform

$$P_0(\tau) = \mathcal{L}^{-1}\left[\frac{g(s)}{1 - g(s)}\right]. \tag{2.4.9}$$

Whilst it is based on the assumption of independence of succesive crossings, Equation 2.4.9 offers a means to calculate $P_0(\tau)$. In order to use Equation 2.4.9 to analyse the zero-crossings, we need to be able to calculate the clipped autocorrelation function $R(\tau)$ and the mean crossing rate β. How we do this is discussed subsequently.

2.4.2 *Van Vleck's theorem and Rice's formula*

In order to solve Equation 2.4.9 for a Gaussian coloured noise we introduce Van Vleck's theorem and Rice's formula. Let $\{X(t)\}$ be a stationary Gaussian process with autocorrelation function $a(\tau)$. Van Vleck's theorem [43] is derived in Appendix B and shows that the clipped autocorrelation function of $\{X(t)\}$ is

$$R(\tau) = \frac{2}{\pi} \sin^{-1} a(\tau). \tag{2.4.10}$$

This is also referred to as the arcsine law formula. Rice's formula [14] is derived in Appendix C and shows that the mean crossing-rate of $\{X(t)\}$ is

$$\beta = \frac{1}{\pi}\sqrt{-a''(0)}. \tag{2.4.11}$$

It follows that the pdf of the zero-crossings $P_0(\tau)$ is directly related to the autocorrelation $a(\tau)$ via McFadden's formula, Van Vleck's theorem and Rice's formula. At this point it suffices to just introduce the theoretical work into zero-crossings of Gaussian stable processes, with sufficient testing into empirical results covered in previous theses [15].

2.4.3 *The persistence parameter*

The final quantity defined to describe the zero-crossing properties is the persistence parameter. This quantifies the tail behaviour of the zero-crossings pdf. To do so, consider

Equation 2.4.8, where there exists a pole at $g(s) = 1$. Because $0 < g(s) \leq 1/2$ for $s \geq 0$, it follows that the pole can only exist for $s < 0$. For convenience, we denote $\theta > 0$ the persistence parameter, such that $g(-\theta) = 1$. The series expansion of $g(s)$ about $s = -\theta$ is

$$g(s) = g(-\theta) + g'(-\theta)(s+\theta) + O\left((s+\theta)^2\right)$$

such that

$$\frac{g(s)}{1-g(s)} \approx \frac{-1}{g'(-\theta)(s+\theta)} \tag{2.4.12}$$

for $s \approx \theta$. The Laplace transform of an exponential decay function is

$$\mathcal{L}\left[\exp(-k\tau)\right] = \frac{1}{k+\tau}$$

for $t \geq 0$. By inverting Equation 2.4.12 and by understanding the poles of the Laplace transform correspond to tail-behaviour, the asymptotic behaviour of the zero-crossing pdf is

$$P_0(\tau) \overset{a}{\sim} \exp(-\theta\tau)$$

where $\theta > 0$ as $\tau \to \infty$. This implies the zero-crossings pdf has an exponential tail if there exists a persistence parameter $\theta > 0$ that satisfies $g(-\theta) = 1$.

2.5 SUMMARY

In this chapter we have introduced a number of key concepts required to investigate the zero-crossings of stable processes. We initially looked at fundamental ideas to investigate empirical heavy-tailed and non-Gaussian stable random variables. Hill's estimator has been introduced as a fundamental tool in empirical heavy-tail analysis, and the set of stable distributions have been defined. We have seen that the set of non-Gaussian stable random variables have heavy-tails that result in an undefined autocorrelation. Because of this, coherence has been introduced as a potential means of quantifying the dependency between stable random variables. A means of calculating the coherence for empirical data sets using structure functions has been described. We then introduced established means of generating coloured noise with a target marginal distribution and autocorrelation. This was done through Fourier Filtering Methods. It was shown that the methods are successful

provided sufficiently large samples are used. This allowed us to generate non-stable heavy-tailed random processes with prescribed autocorrelation functions. Finally we introduced the established methods used to calculate the zero-crossing statistics of Gaussian coloured noises. This included McFadden's formula, Van-Vleck's theorem and Rice's formula. The work in this chapter gives us the framework to test the efficiency of Hill's estimator in chapter 3, as we investigate the effect of correlations on heavy-tail estimation methods.

3

HEAVY-TAIL INDEX ESTIMATORS

3.1 CRITICISM OF HILL'S ESTIMATOR

The prevalence of Hill's estimator as a means to estimate the heavy-tail index is testiment to its suitability in many scenarios [25]. Whilst there has been work into improving and optimising its use [44], [45], it suffices to consider the original formulation. In order to understand its strengths as well as possible weaknesses, we can test its behaviour in specific situations. It is therefore pertinent to be reminded of the definition of Hill's estimator (from section 2.1) as well as introducing certain other aspects and nomenclature.

Firstly, let us consider a set of realisations $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ sampled from n iid heavy-tailed random variables and denote $X_{(j)}$ its jth order statistic. The original formulation of Hill's estimator in Equation 2.1.4 is constructed for a sample that has a Pareto distribution, but this is not generally true for heavy-tailed processes. Because heavy-tails are only asymptotically Pareto, Hill's estimate is reformulated in terms of the k upper statistics, such that

$$\hat{\alpha}_H = \left[\frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{X_{(n-k+j)}}{X_{(n-k+1)}}\right)\right]^{-1}. \tag{3.1.1}$$

For a random sample drawn from a heavy-tailed process, the Hill estimate $\hat{\alpha}_H$ is an asymptotically correct estimate for the heavy-tail index α such that $\hat{\alpha}_H \to \alpha$ as $n \to \infty$ and $k/n \to 0$ [33]. For convenience $X_{(n-k+1)}$ is labelled the threshold value, k the threshold number and $\eta := k/n$ the threshold ratio.

In order to test the validity of Hill's estimator on empirical data, it is first tested on samples of iid random variables with the Pareto pdf

$$f_X(x) = \begin{cases} \alpha x^{-(\alpha+1)} & \text{if } x \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.2}$$

The heavy-tail index is chosen as $\alpha = 3$ and the convergence of Hill's estimate to the theoretical value is tested for different sample sizes $n = 2^6,\ 2^8,\ 2^{12}$. These samples represent a relatively large number of data points in the tail in the context of many practical applications. The results are plotted in Figure 3.1.1 as a function of threshold ratio η. It is clear that for each sample set, the accuracy of the estimate $\hat{\alpha}_H$ tends to increase as larger k (and η) is considered. It is also clear that for any chosen threshold ratio η, the accuracy of $\hat{\alpha}_H$ increases as sample size n increases. From this it is concluded that Hill's estimator is efficient for sets of iid random variables with a Pareto distribution, and its accuracy increases with sample size and threshold ratio. Nevertheless, large data sets are required to achieve good accuracy.

Hill's estimator (and the testing in Figure 3.1.1) is based on a Pareto distribution. The definition of a heavy-tail does not require this, however. Instead it is defined in terms of its asymptotic behaviour and, indeed, this is the reality for nearly all practical manifestations of heavy-tailed distributions. This inherently offers a problem with Hill's estimator in choosing the appropriate threshold value $X_{(n-k+1)}$. If a smaller threshold is chosen there is more data in the tail and the accuracy of Hill's estimator will increase for a process with a perfect power-law distribution - it will be less influenced by random fluctuations as seen by testing on processes with pdf Equation 3.1.2. For a heavy-tailed process with an asymptotic power-law distribution, however, choosing a smaller threshold will mean including values whose behaviour is not significantly influenced by the tail-behaviour, and not well modelled by it. To investigate this two heavy-tailed distributions with asymptotic power-law behaviour are introduced. The pdf's are defined as

$$f_X(x) := \frac{2}{\pi(1+x^2)^2} \overset{a}{\sim} x^{-4}, \tag{3.1.3}$$

$$g_X(x) := \frac{\sqrt{2}}{\pi(1+x^4)} \overset{a}{\sim} x^{-4}, \tag{3.1.4}$$

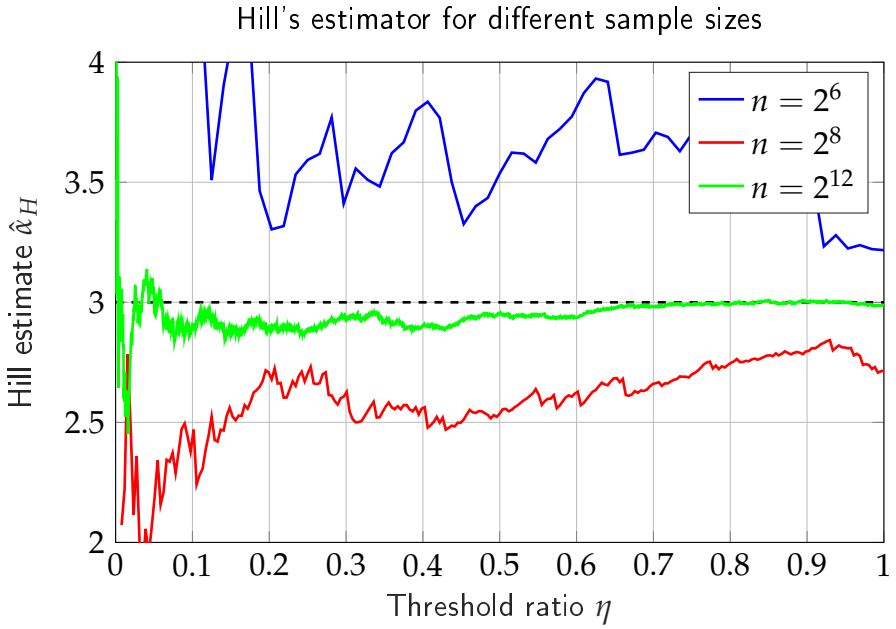

Figure 3.1.1.: Hill's estimator tested on 3 samples of white noise (of different sample size), each distributed according to the Pareto pdf in Equation 3.1.2 with heavy-tail exponent $\alpha = 3$. The Hill estimate is plotted as a function of the threshold ratio $\eta = k/n$, where k is the number of upper-order statistics used to generate the Hill estimate.

where $x \in \mathbb{R}$, such that their heavy-tailed exponents are both $\alpha = 3$. It is clear that $g_X \to x^{-4}$ faster than f_X, and that the functions behave differently around the origin. Hill's estimator is tested on samples of white noise with the pdf's f_X and g_X for different sample sizes $n = 2^{10}$, 2^{14}, 2^{18} up to a threshold ratio $\eta = 0.1$ (10%) (as discussed in [46]). The results are plotted as a function of the threshold ratio in Figure 3.1.2.

There are several conclusions to draw from these plots. Firstly; the difficulty in choosing the threshold number k can be seen. It can clearly be seen that choosing too small a value will result in an unreliable estimate. Conversely, choosing too large a value will see the estimate influenced by non-tail values resulting in a systematic under-estimate of α. Secondly; it is noted that the speed of convergence to the tail affects the accuracy. Whilst both f_X and g_X represent heavy-tailed distributions with the same heavy-tail index, there is a clear difference in the accuracy of Hill's estimator at different threshold ratios. Finally; the total sample size n has a clear effect on the accuracy when considering identical distributions and identical threshold ratios, as shown by comparisons in Figure 3.1.2. Through any combination of these considerations, Hill's estimator is ultimately unreliable if it does not measure enough values that are overwhelmingly influenced by the heavy-tail behaviour.

A key question arising from this is the optimal means of selectioning the threshold value $X_{(n-k+1)}$ (and consequently the optimum number of tail values k and threshold ratio η). An accepted means by which to select this is through analysing the Hill plot (where the Hill estimate is plotted as a function of threshold ratio). By considering a sample's Hill plot up to a certain maximum threshold ratio (0.1 is an often accepted upper limit [46]), the point at which the Hill estimate ceases to be too varied can be estimated. For sufficiently large sample sizes, there is a clear convergence around a value - such as the realisation with distribution g_X of size $n = 2^{18}$ in Figure 3.1.2 in the region $\eta \in (0.005, 0.025)$.

Whilst Hill's estimator has been shown to be relatively efficient on two general heavy-tailed random variables, it is pertinent to test it on stable processes. It is of particular interest to question how the efficiency of Hill's estimator is affected by the stablility-parameter α.

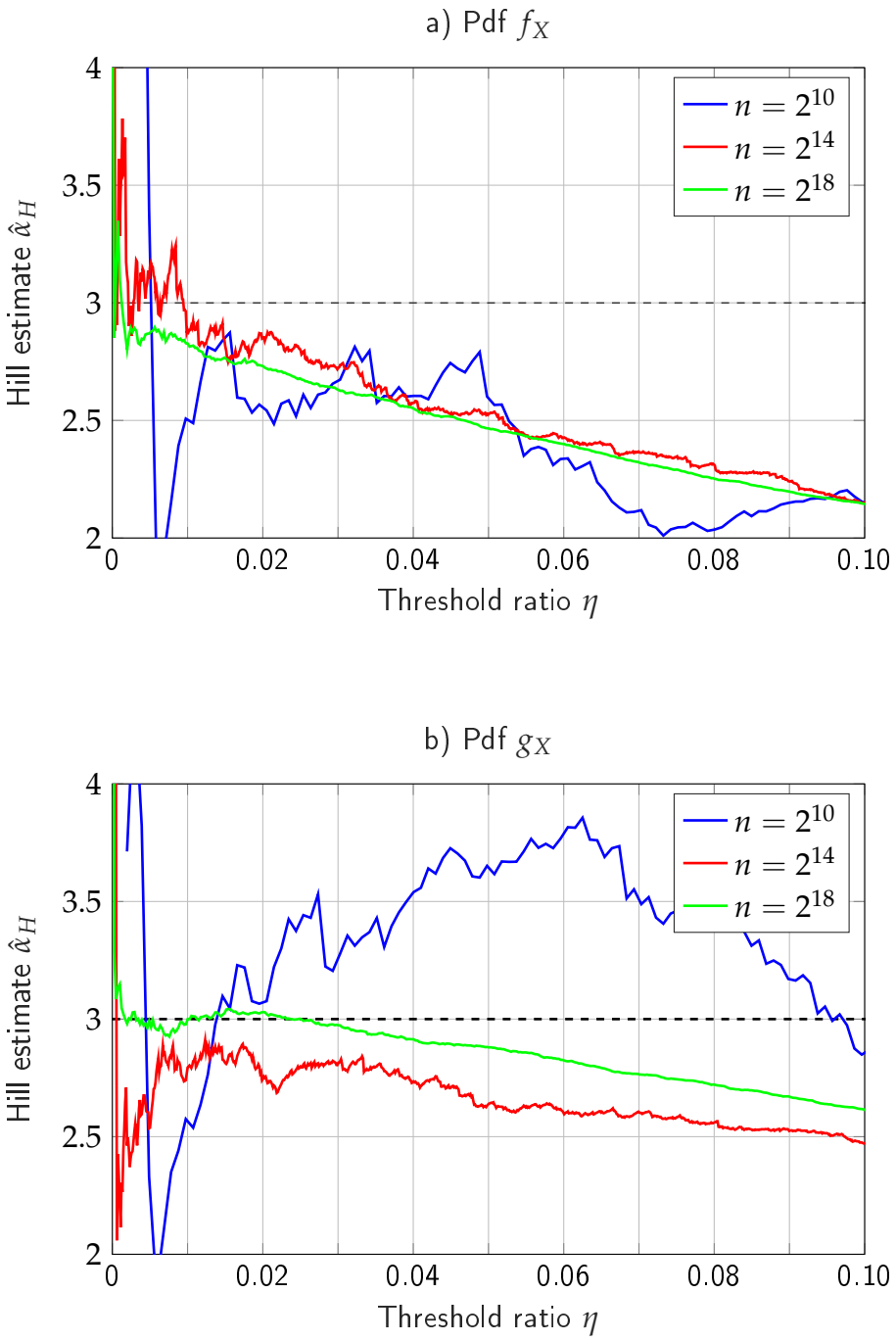

Figure 3.1.2.: Hill's estimator on samples with pdf's a) f_X (Equation 3.1.3) and b) g_X (Equation 3.1.4) and $n = 2^{10}$, 2^{14}, 2^{18} elements. Plotted against threshold ratio η.

To do this, the Hill plots of specific cases are investigated. The stability parameter is varied using $\alpha = 1.0,\ 1.5,\ 1.8,\ 1.9$. It is of no interest to consider the special case $\alpha = 2.0$ as the resulting process is Gaussian and will not exhibit heavy-tailed characteristics. For each case the sample sizes $n = 2^{10},\ 2^{14},\ 2^{18}$ are used. The results are plotted in Figure 3.1.3 as functions of the threshold ratio η. It is clear that Hill's estimator offers an accurate estimate under certain conditions. For smaller α and larger sample size n, the Hill plot is often non-volatile for a certain threshold ratio, which can allow us to make a relatively accurate estimate for α. As $\alpha \to 2$, however, it is clear that Hill's estimator struggles. Even for large sample sizes ($n = 2^{18}$) Hill's estimator often estimates that the heavy-tail index is well outside the stable regime.

3.1.1 *Selecting the optimum threshold value*

Whilst the effect of variability in the threshold value and the resultant Hill estimate has been shown, means to decide the optimum value have not been discussed. As shown previously in this chapter, optimum threshold values (and ratios) vary depending on sample size, distribution function and heavy-tail parameter. Several methods to choose the optimal threshold value are discussed in [47], including the graphical method [48], described below.

The graphical method involves creating a Hill plot as described previously, and identifying a horizontol region in the plot. This region represents a threshold that allows sufficient data to ensure the variance of the estimate is sufficiently small, but that the addition of smaller and more 'central' values do not significantly influence the estimate. Furthermore, it is useful to plot the threshold number logarithmically to identify the appropriate region [49]. When the region is identified, a number of methods may be used to decide which estimate to take (although it may be argued this aspect is trivial anyway as a relatively non-volatile region is considered). Let the Hill estimate using the threshold value $X_{(j)}$ be denoted as a function $\hat{\alpha}_H(X_{(j)})$. For an identified horizontal region with threshold numbers $k = k_1,\ \ldots,\ k_h$ and threshold values $X_{(n+1-k)} = X_{(n+1-k_1)},\ \ldots,\ X_{(n+1-k_h)}$, the median threshold is $X_{(n+1-\lceil h/2 \rceil)}$. The equivalent Hill estimate $\hat{\alpha}_H(X_{(n+1-\lceil h/2 \rceil)})$ is labelled the

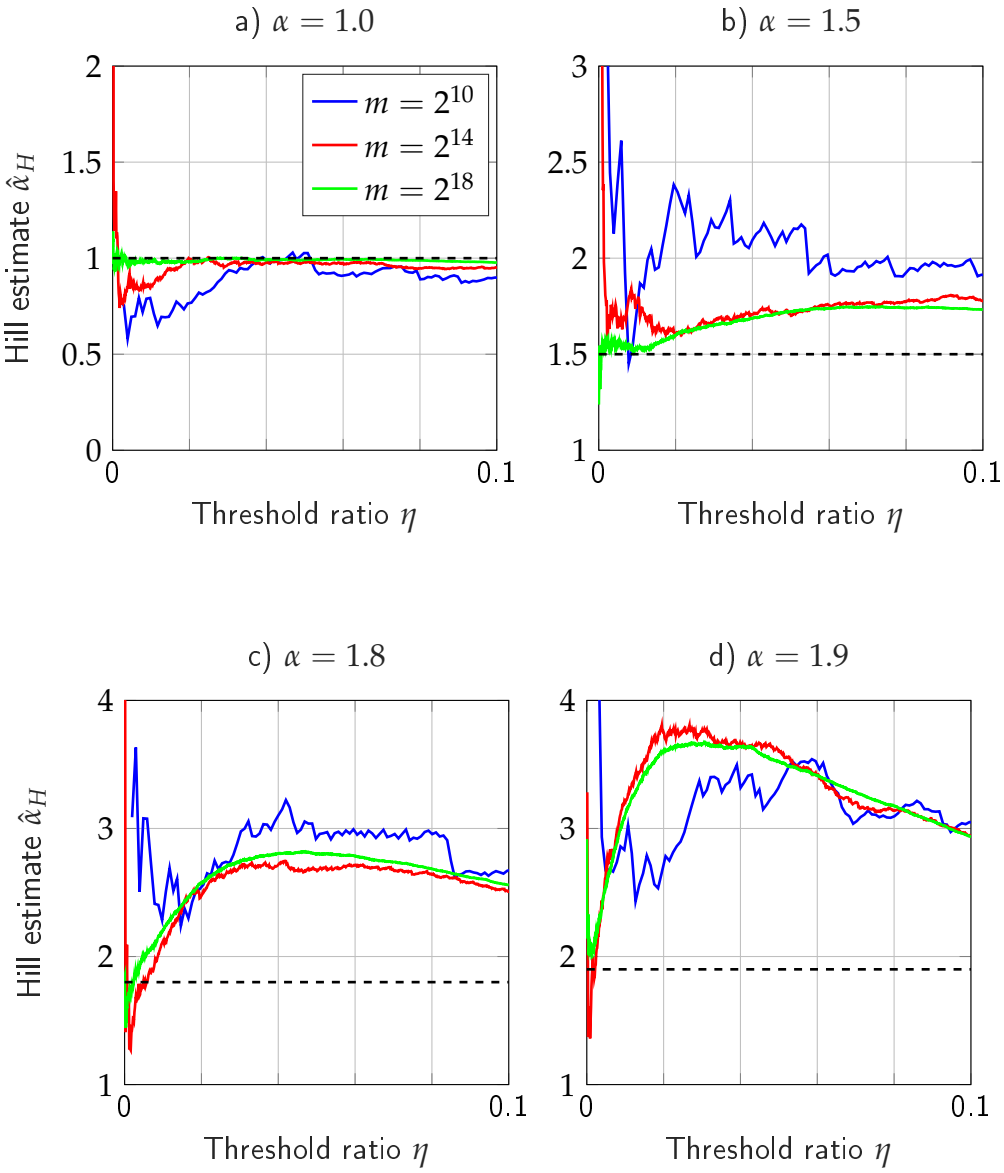

Figure 3.1.3.: Hill plots of sample stable processes with $n = 2^{10}$ (blue), 2^{14} (red) 2^{18} (green) elements. Plotted for heavy-tail indices a) $\alpha = 1.0$, b) $\alpha = 1.5$, c) $\alpha = 1.8$ and d) $\alpha = 1.9$.

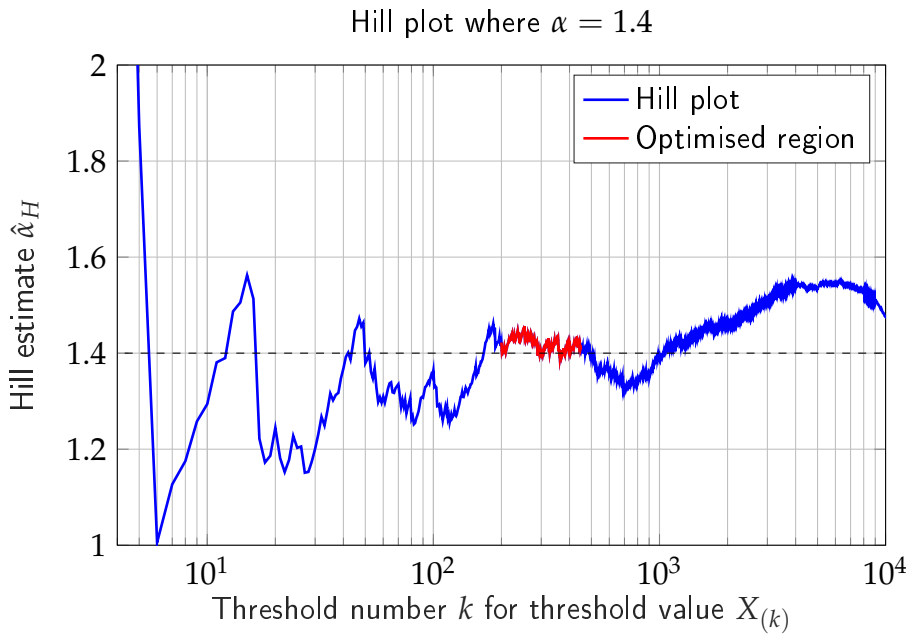

Figure 3.1.4.: Hill's estimator on stable process with $\alpha = 1.4$. Plotting on a logarithmic scale helps to identify the region to optimise (red). The resultant estimate is $\hat{\alpha}_H = 1.4176$.

median Hill estimate. Another possible measure is to take the mean average of all inputs on the identified non-volatile region $\langle \hat{\alpha}_H(X_{(j)}) \rangle_{j=n+1-k_1}^{n+1-k_h}$. This is labelled the mean Hill estimate.

To display this optimisation process, we can investigate a sample of a stable process with $\alpha = 1.4$ and $n = 2^{16}$. Using a Hill-plot (with logarithmic abscissa), a horizontal region is identified with threshold numbers $k = 200, \ldots, 450$ in Figure 3.1.4. From this region, the median Hill estimate is $\hat{\alpha}_H = 1.4143$ and the mean Hill estimate is $\hat{\alpha}_H = 1.4176$. In this case, both optimised methods produce an extremely accurate estimate for the underlying heavy-tail index $\alpha = 1.4$. It suffices to proceed using just the mean Hill estimate as the preferred optimisation method. Despite the ability to optimise it, it can be observed from Figure 3.1.3 that Hill's estimator still struggles under certain conditions.

3.2 STABLE CHARACTERISTIC FUNCTION METHOD

The criticisms of Hill's estimator for processes with a stability parameter near the stable threshold $\alpha = 2$ prompts a need to introduce and investigate an alternative approach for stable processes. Whilst stable random variables do not (in general) have analytical distribution functions, they can be expressed analytically in terms of their characteristic function. This allows us to introduce the Stable Characteristic Function (SCF) method [26]. The characteristic function $\phi(t)$ for a random variable is defined in Equation 1.2.1. Then for a set of realisations $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ the empirical characteristic function is

$$\tilde{\phi}(t) = \frac{1}{n}\sum_{j=1}^{n} \exp\left(itX_j\right).$$

It is shown in ?? that for a set of n realisations drawn from iid rv's with the characteristic function $\phi(t)$, the empirical characteristic function converges such that $\tilde{\phi}(t) \to \phi(t)$ almost surely as $n \to \infty$. As in section 2.2, the set of stable distributions may be defined in terms of the characteristic function

$$\phi(t) = \exp\left[it\mu - |\gamma t|^{\alpha}\left(1 - i\beta \text{sgn}(t)\psi(t)\right)\right].$$

Then the square of the modulus of the characteristic function is

$$|\phi(t)|^2 = \exp\left(-|\gamma t|^{\alpha}\right)^2 = \exp\left(-2|\gamma t|^{\alpha}\right)$$

where $\gamma > 0$ and $\alpha \in (0, 2]$. It follows that

$$\log\left(-\log|\phi(t)|^2\right) = \log\left(2\gamma^{\alpha}\right) + \alpha\log\left(|t|\right).$$

The stable characteristic function method proposes estimating the heavy-tail index α by using regression to fit a linear relationship between $\log(-\log|\tilde{\phi}(t)|^2)$ against $\log(|t|)$ using the empirical characteristic function. The gradient is estimated as the stablility parameter α. This method is advantageous for stable processes as the relationship is true generally as opposed to just in the tail of the density function. Characteristic function methods can be used to identify the other parameters (location parameter μ, scale parameter γ and skewness parameter β) to fit a complete stable model to an empirical data set if required.

Whilst there is a finite number of elements in the sample $\boldsymbol{X}$ that the empirical characteristic function $\tilde{\phi}(t)$ is calculated from, $\tilde{\phi}(t)$ itself can be evaluated for any choice t. It is shown in [26] that the mean-square error in the estimation of α varies very little with the exact choice of locations of t. What does have an effect is the size of region of t's. The proposed choices of t to evaluate the empirical characteristic function are taken from [26] as $t = t_k$ where

$$t_k = \frac{\pi k}{25}$$

and $k = 1, \ldots, K$. The optimal value of K is dependent on α and grows smaller as $\alpha \to 2$. Choosing the region upon which to regress is therefore discretionary, and should be chosen where there exists sufficient evidence for a linear region (when scaled logarithmically).

3.3 GENERAL CHARACTERISTIC FUNCTION METHOD

The Stable CF method takes advantage of the analytical form of the characteristic function of a stable random variable, but does not consider more general heavy-tailed random variables. To remedy this, the characteristic function of a general heavy-tailed random variable and a possible means of using it to calculate the heavy-tail index is investigated. This allows us to comment on the viability of characteristic function analysis as a general tool to estimate the heavy-tail index, especially in comparison to Hill's estimator. A more general heavy-tailed distribution is introduced, and defined by the pdf

$$f_X(x) := \frac{A}{\left(\delta^2 + x^2\right)^{(1+\alpha)/2}} \overset{a}{\sim} |x|^{-(1+\alpha)} \tag{3.3.1}$$

for $x \in \mathbb{R}$ where $A > 0$ is a normalising constant, $\delta > 0$ is a scale parameter and $\alpha > 0$ is the heavy-tail index (not necessarily in the stable regime). The characteristic function of a random variable with pdf Equation 3.3.1 can be calculated using the definition

$$\phi(t) = \langle \exp(itX) \rangle = \int_{-\infty}^{\infty} \frac{A \exp(itx)}{\left(\delta^2 + x^2\right)^{(1+\alpha)/2}}\, dx. \tag{3.3.2}$$

Using Euler's formula, and because the pdf Equation 3.3.1 is even, it follows that Equation 3.3.2 can be simplified to

$$\phi(t) = \int_{-\infty}^{\infty} \frac{A \cos(tx)}{\left(\delta^2 + x^2\right)^{(1+\alpha)/2}}\, dx. \tag{3.3.3}$$

By computing Equation 3.3.3 (and the normalising constant A), the characteristic function of a random variable with general heavy-tail pdf in Equation 3.3.1 is

$$\phi(t) = \frac{2^{1-\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)} |\delta t|^{\frac{\alpha}{2}} K_{\frac{\alpha}{2}}(|\delta t|) \tag{3.3.4}$$

where K_n is the modified Bessel K function of the second kind [50]. For non-integer values of n, K_n can be expressed as the power-series

$$K_n(t) = t^{-n}\left[t^{2n}\left\{2^{-(n+1)}\,\Gamma(-n) + \frac{2^{-(n+3)}}{n+1}\,\Gamma(-n)\;t^2 + O\left(t^4\right)\right\} + \left\{2^{n-1}\,\Gamma(n) - \frac{2^{n-3}}{n-1}\,\Gamma(n)\;t^2 + O\left(t^4\right)\right\}\right].$$

and therefore Equation 3.3.4 can be expressed as

$$\phi(t) = \left[\frac{2^{1-\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)}\right]\left[t^{\alpha}\left\{2^{-(\frac{\alpha}{2}+1)}\,\Gamma\left(-\frac{\alpha}{2}\right) + \frac{2^{-(\frac{\alpha}{2}+3)}}{\frac{\alpha}{2}+1}\,\Gamma\left(-\frac{\alpha}{2}\right)\;t^2 + O\left(t^4\right)\right\} + \left\{2^{\frac{\alpha}{2}-1}\,\Gamma\left(\frac{\alpha}{2}\right) - \frac{2^{\frac{\alpha}{2}-3}}{\frac{\alpha}{2}-1}\,\Gamma\left(\frac{\alpha}{2}\right)\;t^2 + O\left(t^4\right)\right\}\right]. \tag{3.3.5}$$

The characteristic function expressed in Equation 3.3.5 can be relabelled and simplified to emphasise the importance of the exponents, such that

$$\phi(t) = |t|^{\alpha}\sum_{j=0}^{\infty} l_j\,|t|^{2j} + \sum_{j=0}^{\infty} k_j\,|t|^{2j}$$

where l_j, k_j are coefficients of the series which may be calculated from Equation 3.3.5. The characteristic function can be considered up to $O(t^{\alpha})$ such that, for sufficiently small t,

$$\phi(t) \approx l_0\,|t|^{\alpha} + \sum_{j=0}^{m} k_m\,|t|^{2j} \tag{3.3.6}$$

where $m = \lceil \alpha/2 \rceil - 1$. The jth raw moment of a random variable $\mu_j = \langle X^j \rangle$ exists iff $j < \alpha$. When considering the expansion for the characteristic function in Equation 3.3.6, the coefficients of the integer power-series can be deduced by considering the raw moments. Specifically

$$k_j = \frac{\mu_{2j} i^{2j}}{(2j)!}$$

where i is the imaginary unit. It follows that Equation 3.3.6 can be rewritten

$$\phi(t) \approx l_0\,|t|^{\alpha} + \sum_{j=0}^{m} \frac{\mu_{2j}\,(it)^{2j}}{(2j)!} \tag{3.3.7}$$

for $t \approx 0$ where μ_j is the jth raw moment and $m = \lceil \alpha/2 \rceil - 1$. Whilst the above is derived from an even pdf (in Equation 3.3.1) such that the characteristic function is completely real, non-zero odd moments (such as mean and skewness) can be accounted for too. In these instances, Equation 3.3.7 can be adjusted to

$$\phi(t) \approx l_0 |t|^\alpha + \sum_{j=0}^{m} \frac{\mu_j (it)^j}{j!} \tag{3.3.8}$$

where $m = \lceil \alpha \rceil - 1$. The approximations therefore only include terms that can be calculated empirically through raw integer moments.

3.3.1 *Proposed characteristic function algorithm*

The approximation in Equation 3.3.8 suggests a possible method to estimate the heavy-tail index α for the general heavy-tailed pdf in Equation 3.3.1. Consider a process with empirical characteristic function $\tilde{\phi}(t)$. Define the kth truncated empirical characteristic function

$$\tilde{\phi}_k(t) = \tilde{\phi}(t) - \sum_{j=0}^{k} \left(\frac{\mu_j i^j}{j!} \right) t^j \tag{3.3.9}$$

where k is a non-negative integer and i is the imaginary unit. It follows that if $\alpha < k$ then

$$\tilde{\phi}_k(t) \overset{a}{\sim} |t|^\alpha$$

whilst, if $\alpha > k$, then

$$\tilde{\phi}_k(t) \overset{a}{\sim} |t|^k$$

and consequently

$$\log |\tilde{\phi}_k(t)| \approx A_0 + \min(k, \alpha) \log |t| \tag{3.3.10}$$

for sufficiently small t, where A_0 is some constant. With this relationship in mind, the following algorithm is proposed:

1. Let $k = 0$.
2. Using linear regression and Equation 3.3.10, calculate the slope of $\log |\tilde{\phi}_k(t)|$ against $\log |t|$, and denote this $\hat{\alpha}_G$.

α	$\phi(t)$
1.0	$1 + 0.50t^2 - 1.00\|t\|^{1.0}$
1.4	$1 + 0.83t^2 - 1.25\|t\|^{1.4}$
1.8	$1 + 2.50t^2 - 2.84\|t\|^{1.8}$
2.2	$1 - 2.50t^2 + 2.22\|t\|^{2.2}$
2.6	$1 - 0.83t^2 + 0.61\|t\|^{2.6}$

Table 3.3.1.: Table of expansions of characteristic function derived from general heavy-tail in Equation 3.3.4 with varying α and $\delta = 1$. Expanded up to third order.

3. If $\hat{\alpha}_G = k$, it follows that the leading term is k and $\alpha \geq k$. Therefore the kth moment $\mu_k = \langle X^k \rangle$ exists and the coefficient of the leading term $|t|^k$ can be calculated. Let $k = k + 1$ and repeat Step 2 and Step 3 until $\hat{\alpha}_G \neq k$.

4. If $\hat{\alpha}_G \neq k$, then $\hat{\alpha}_G$ is the estimate for the heavy-tail index α.

Henceforth, the above shall be referred to as the General Characteristic Function (GCF) method.

3.3.2 *GCF method near the stable threshold*

Whilst the GCF method does not have a limit on the value of the heavy-tail index it can be used to estimate, it is timely to recall its intended use. We are specifically interested in how its use might improve on estimating the index for heavy-tailed random variables that may or may not be asymptotically stable. We therefore consider the GCF method, and the behaviour of the characteristic function in general, as the tail index of a heavy-tailed random variable approaches, and passes through, the stable threshold $\alpha = 2$. This will allow us to comment on the method's viability as an alternative candidate to Hill's estimator, as well as the Stable CF method.

Consider the characteristic function of the general heavy-tailed random variable in Equation 3.3.4. For convenience, the scale parameter is set to $\delta = 1$. The resulting power-series expansion in Equation 3.3.5 is evaluated up to the third order (either $|t|^\alpha$ or t^2) such that

$$\begin{aligned}\phi(t) &= 1 + k_1 t^2 + l_0 |t|^\alpha \\ &= 1 + \frac{1}{2(2-\alpha)} t^2 + \frac{2^{-\alpha}\Gamma\left(-\frac{\alpha}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)} |t|^\alpha\end{aligned} \tag{3.3.11}$$

and tabulated for several choices of $\alpha \in [1,3)$ in Table 3.3.1. For $\alpha < 2$ and sufficiently small t the characteristic function will be primarily dominated by the $|t|^\alpha$ term, whilst for $\alpha > 2$ it will be primarily affected by the t^2 term. There must therefore exist some crossover as $\alpha \to 2$. This highlights a possible problem for the proposed method as the heavy-tail index approaches the stable threshold. The limiting behaviour as $\alpha \to 2$ can be calculated by

$$\begin{aligned}\phi(t) &= 1 + \lim_{\alpha\to 2} t^2 \left(k_1 + l_0 |t|^{\alpha-2}\right) \\ &= 1 + \frac{1}{4} t^2 \lim_{\alpha\to 2} \left(\frac{2}{2-\alpha} + \frac{\Gamma\left(-\frac{\alpha}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)} \left|\frac{t}{2}\right|^{\alpha-2}\right)\end{aligned} \tag{3.3.12}$$

The Taylor-series expansions of the terms in Equation 3.3.12 are

$$\begin{aligned}\Gamma\left(-\frac{\alpha}{2}\right) &= \frac{-2}{2-\alpha} + (\gamma_E - 1) + O(\alpha - 2), \\ \Gamma\left(\frac{\alpha}{2}\right) &= 1 - \frac{1}{2}\gamma_E (\alpha - 2) + O\left((\alpha-2)^2\right), \\ \left|\frac{t}{2}\right|^{\alpha-2} &= 1 + (\alpha-2)\log\left|\frac{t}{2}\right| + O\left((\alpha-2)^2\right),\end{aligned}$$

where $\gamma_E \approx 0.577$ is the Euler-Mascheroni constant. It also follows that, because $(1-x)^{-1} = 1 + x + O(x^2)$ about $x = 0$,

$$\Gamma\left(\frac{\alpha}{2}\right)^{-1} = 1 + \frac{1}{2}\gamma_E (\alpha - 2) + O\left((\alpha-2)^2\right).$$

Hence

$$\begin{aligned}\phi(t) = 1 + \frac{1}{4} t^2 \lim_{\alpha\to 2} \Bigg(\frac{2}{2-\alpha} + \left[\frac{-2}{2-\alpha} + (\gamma_E - 1)\right] \left[1 + \frac{1}{2}\gamma_E (\alpha-2)\right] \\ \left[1 + (\alpha-2)\log\left|\frac{t}{2}\right|\right] + O(\alpha-2)\Bigg)\end{aligned}$$

which, when evaluated gives the limit function

$$\phi(t) = 1 + \frac{1}{4} t^2 \left(2\gamma_E - 1 + \log\left|\frac{t}{2}\right|\right). \tag{3.3.13}$$

The interpretation of the limit (apart from the analytical result) is that the leading power term exhibits less 'dominance' over the 2nd power term as the heavy-tail index approaches the stable threshold and actually approaches a logarithmic relationship. Such departure from the single-term approximation behaviour will have an effect on its efficiency and viability. It is therefore important to clarify how quickly the single-term approximation becomes less accurate. To do this, the characteristic function is now expressed in terms of both t and the heavy-tail index α: $\phi(t,\alpha)$. This is compared to the 2-term approximation $\psi(t,\alpha) = 1 + l_0|t|^\alpha + k_1 t^2$ and the 1-term approximation $\zeta(t,\alpha) = 1 + l_0|t|^\alpha$ (for $\alpha < 2$), or $\zeta(t,\alpha) = 1 + k_1|t|^2$ (for $2 < \alpha < 3$), where l_0 and k_1 are the coefficients in Equation 3.3.11. The characteristic function and its approximations are compared for four specific values of the heavy-tail index $\alpha = 1.0,\ 1.5,\ 1.9,\ 2.5$. The results are plotted in Figure 3.3.1. Additionally, the complementary characteristic functions (and its approximations) are defined $\bar{\phi}(t,\alpha) = 1 - \phi(t,\alpha)$, $\bar{\psi}(t,\alpha) = 1 - \psi(t,\alpha)$ and $\bar{\zeta}(t,\alpha) = 1 - \zeta(t,\alpha)$. The complementary functions are plotted on a logarithmic scale in Figure 3.3.2.

From the plots on a linear scale, it is clear that the 1-term approximation becomes less accurate as $\alpha \to 2$, unlike the 2-term approximation which models the characteristic function closely. From the logarithmic plots, it is clear that (with the exception of $\alpha = 1.9$) the 1-term approximation is appropriate on the interval $[10^{-2}, 10^{-1}]$.

It is convenient to see how these models compare to empirically observed characteristic functions. To do so, realisations of the process with the general heavy-tail pdf in Equation 3.3.1 are generated. The empirical characteristic function is evaluated for certain t and plotted against the theoretical value and the 1-term approximation. The results are plotted for different sample sizes in Figure 3.3.3. From these plots, it is clear the empirical characteristic function converges to the theoretical characteristic function as the sample size increases (as expected). Similarly the empirical complementary characteristic functions are plotted for different sample sizes in Figure 3.3.4. From these plots, the variance and uncertainty for smaller sample sizes is clearer for small t.

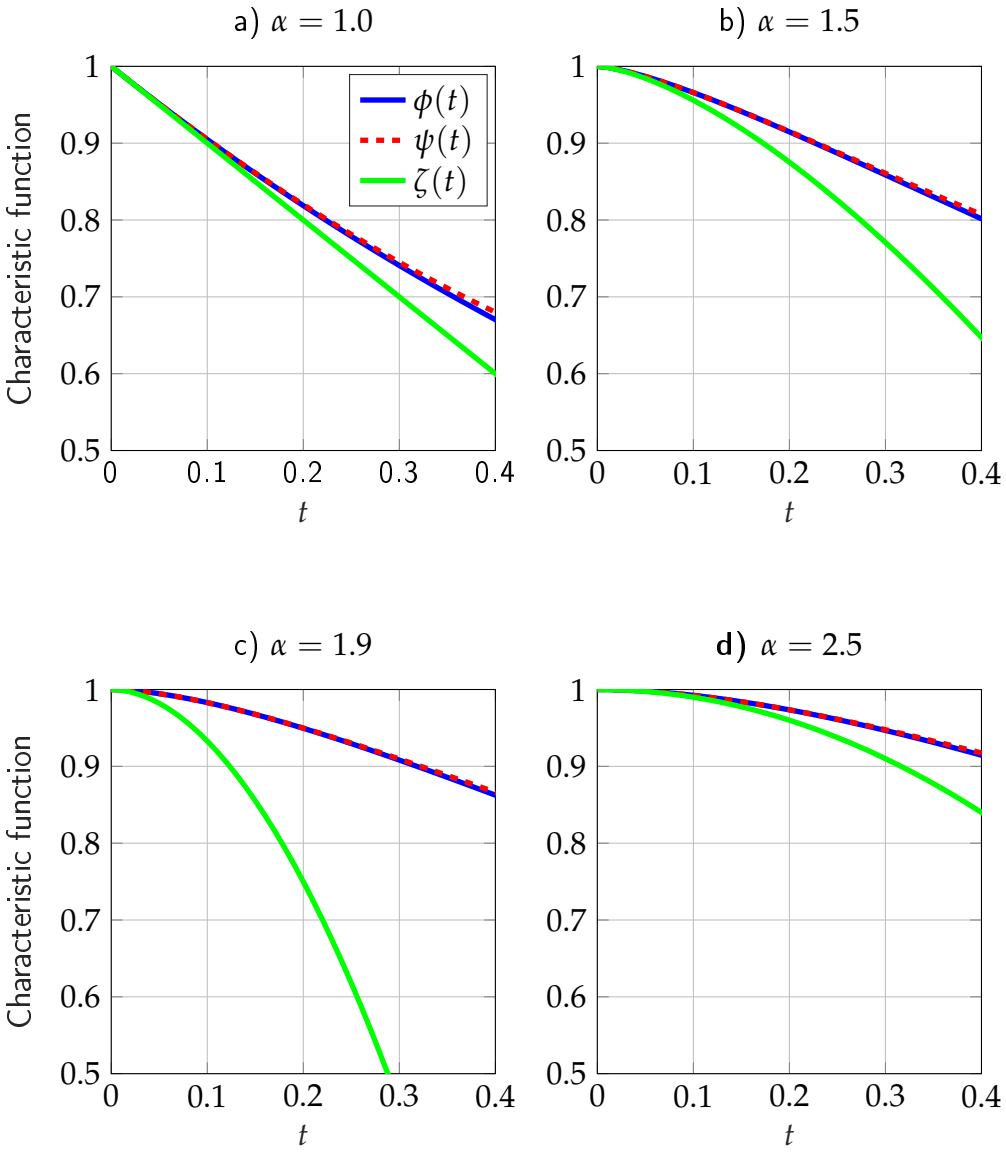

Figure 3.3.1.: Characteristic function of general heavy-tail random variable $\phi(t)$ (blue), 2-term approximation $\psi(t) = 1 + k_1 t^2 + l_0 |t|^\alpha$ (red) and 1-term approximation $\zeta(t) = 1 + l_0 |t|^\alpha$ (for $\alpha < 2$), or $\zeta(t) = 1 + k_1 t^2$ (for $\alpha > 2$) (green) for different heavy-tail indices $\alpha = 1.0,\ 1.5,\ 1.9,\ 2.5$.

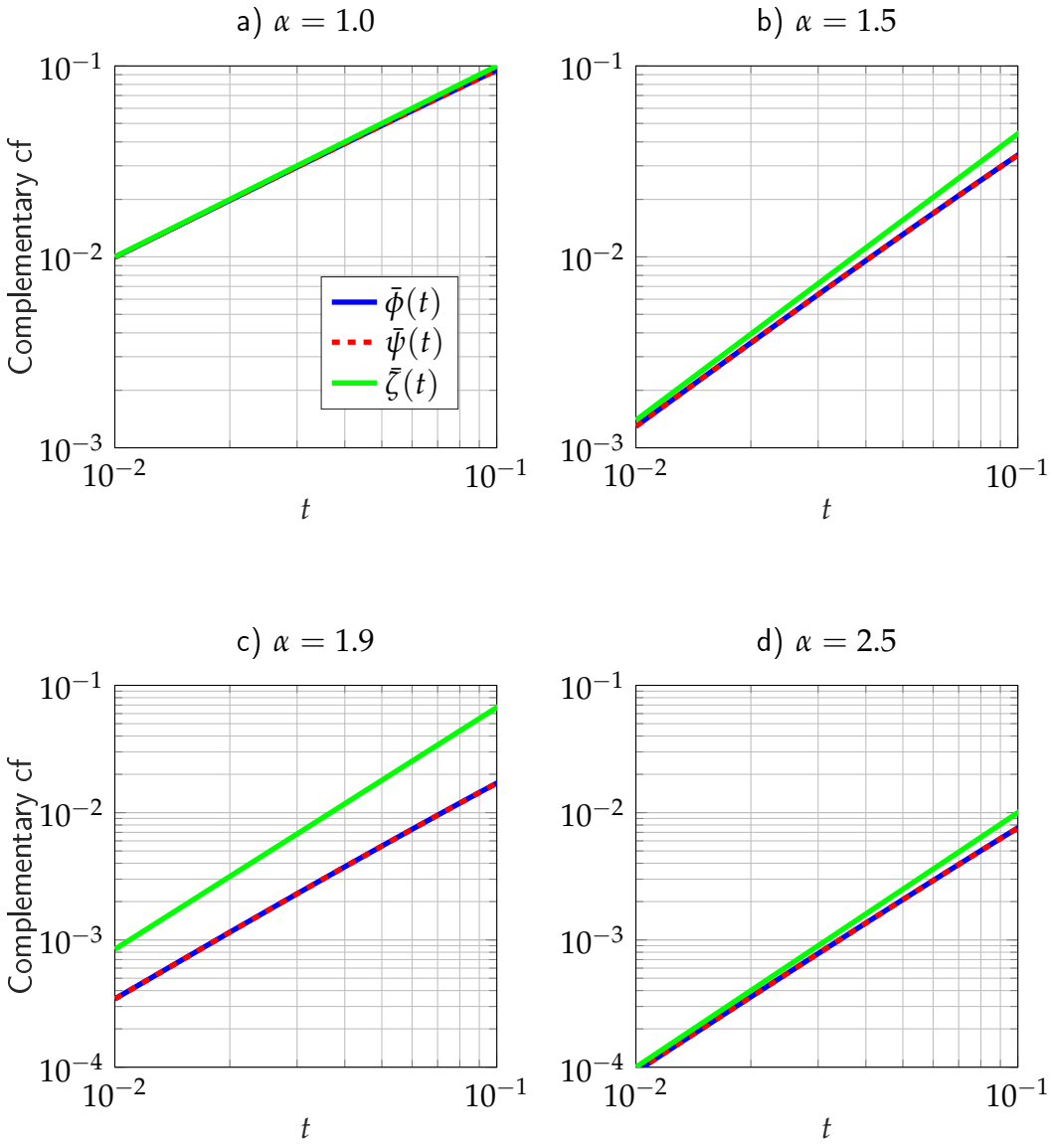

Figure 3.3.2.: Complementary characteristic function of general heavy-tail random variable $\bar{\phi}(t) = 1 - \phi(t)$ (blue), complementary 2-term approximation $\bar{\psi}(t) = 1 - \psi(t)$ (red) and complementary 1-term approximation $\bar{\zeta}(t) = 1 - \zeta(t)$ (green) for different heavy-tail indices $\alpha = 1.0,\ 1.5,\ 1.9,\ 2.5$.

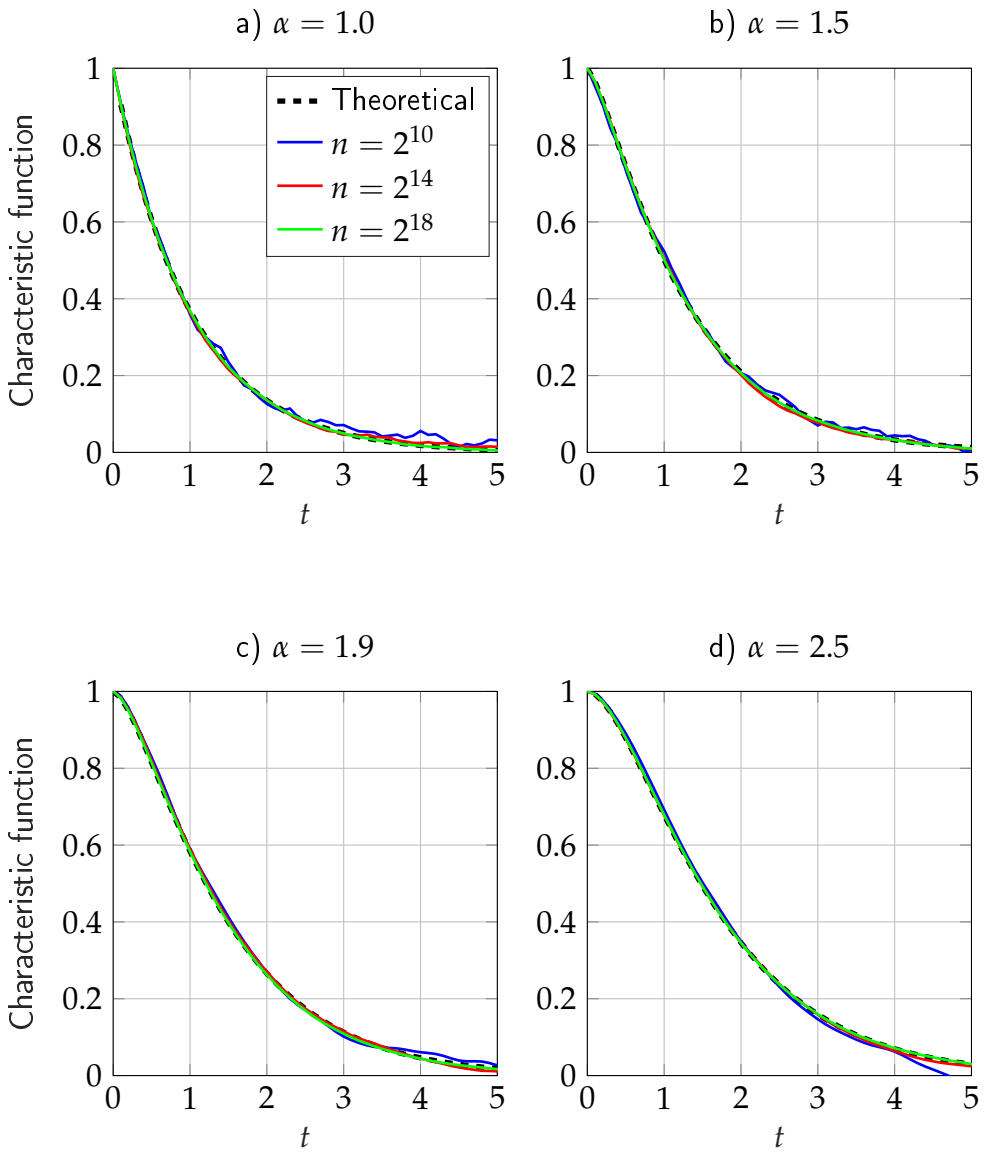

Figure 3.3.3.: Empirical characteristic functions of heavy-tailed processes for different sample sizes, alongside theoretical characteristic function. Plotted for sample sizes $n = 2^{10}$, 2^{14}, 2^{18} and heavy-tail indices $\alpha = 1.0,\ 1.5,\ 1.9,\ 2.5$.

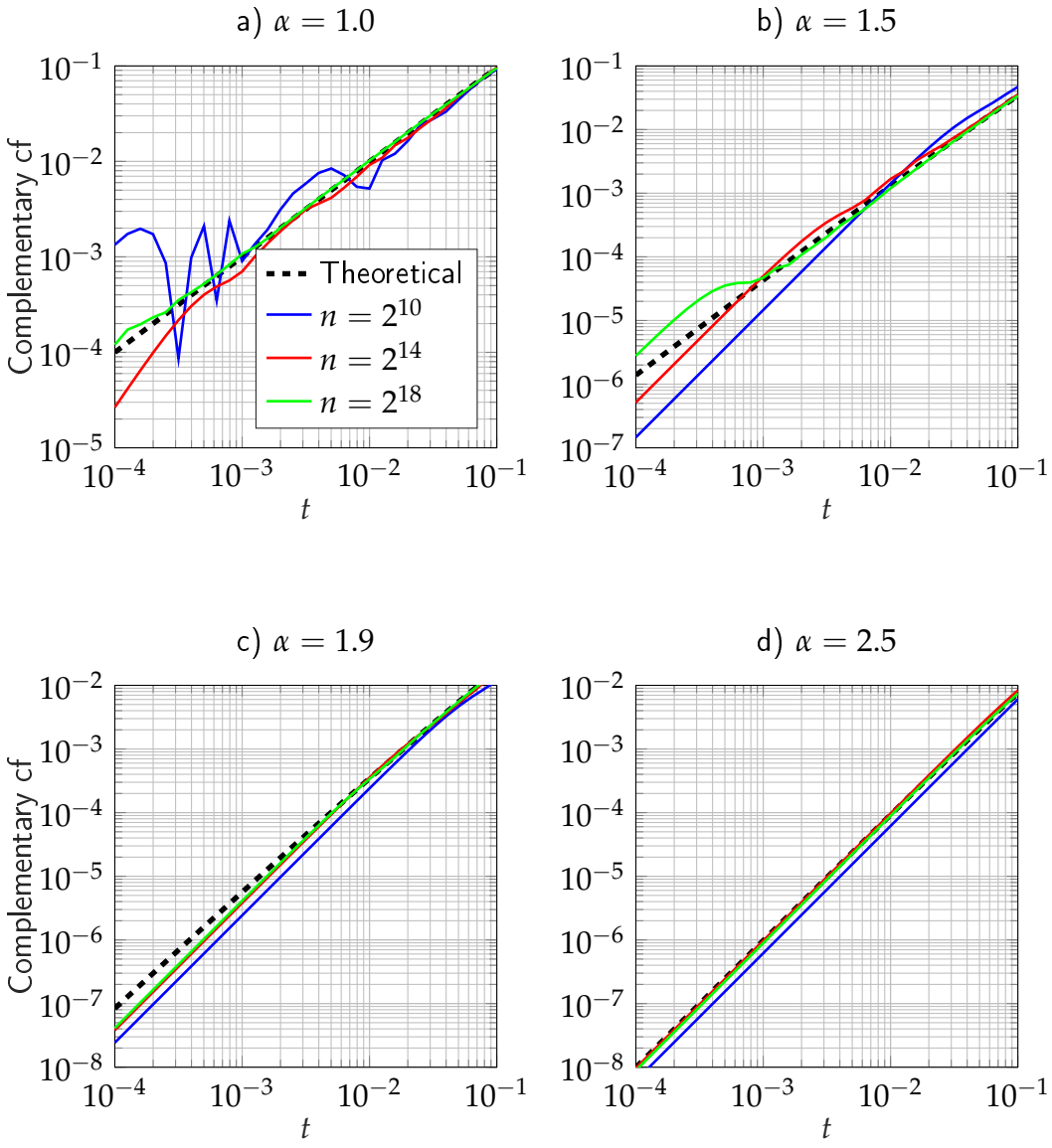

Figure 3.3.4.: Empirical complementary characteristic functions of heavy-tailed processes for different sample sizes, alongside theoretical characteristic function. Plotted for sample sizes $n = 2^{10}, 2^{14}, 2^{18}$ and heavy-tail indices $\alpha = 1.0, 1.5, 1.9, 2.5$. Plotted on a logarithmic scale.

From the plots in Figure 3.3.4, it is clear there is a need to be careful on selecting the region in which to use linear regression to estimate α. As discussed in section 3.2 and [26], it is not neccesarily important the the number of time points t (or the increments) upon which $\tilde{\phi}(t)$ is evaluated, but more important that the total size of the region is appropriate.

It is therefore proposed that in implementing the method, regularly spaced time intervals are used with $t > 10^{-2}$. For example, an appropriate spacing might be

$$t_k = \frac{\pi k}{100}$$

for $k = 1, \ldots, K$. For this spacing choice, $K = 30$ would be approximately equivalent to the interval $[10^{-2}, 10^{-1}]$ - a seemingly appropriate interval given the plotted empirical complementary characteristic functions. To optimise the general characteristic function method, it is therefore proposed that the empirical characteristic function is regressed on an interval chosen by the user, such that the region appears to be linear. Justification of linearity can be discretionary.

3.3.3 *SCF method on general heavy-tailed distributions*

Whilst both the Stable CF and General CF method take advantage of the behaviour of the characteristic function and its relationship with the heavy-tail index, the quantities that are compared (and subsequently upon which α is fitted and estimated) are different. Specifically, the GCF method uses the logarithm of the empirical characteristic function $\log|\tilde{\phi}(t) - 1|$ whilst the SCF method uses the quantity $\log(-\log|\tilde{\phi}(t)|^2)$. It is of interest to show that both methods should still estimate the heavy-tail parameter of both stable and general heavy-tailed random variables (for $\alpha < 2$). First, consider a stable random variable. Then the modulus of the characteristic function is

$$|\phi(t)| = \exp\left(-|\gamma t|^{\alpha}\right).$$

Using Taylor-series expansion about $t = 0$, the stable characteristic function is

$$\phi(t) \approx 1 - |\gamma t|^{\alpha}$$

and therefore the logarithm of the truncated characteristic function is

$$\log|\phi_0(t)| = \log|\phi(t) - 1| \approx \log\left|-|\gamma t|^{\alpha}\right| = \log\left(\gamma^{\alpha}\right) + \alpha \log|t|.$$

It follows that regression as per the GCF method will predict the heavy-tail index of a non-Gaussian stable random variable. Now consider a random variable with a general heavy-tailed characteristic function such that for $\phi(t) \approx 1 + l_0|t|^\alpha$. It follows that $|\phi(t)|^2 \approx 1 + 2l_0|t|^\alpha$. Using Taylor-series about $t = 0$, it follows that

$$\log|\phi(t)|^2 \approx \log\left|1 + 2l_0\,|t|^\alpha\right| \approx 2l_0\,|t|^\alpha\,.$$

For $\alpha < 2$, it is known that $l_0 < 0$ and therefore

$$\log\left(-\log|\phi(t)|^2\right) \approx \log|2l_0| + \alpha\log|t|\,.$$

It follows that the regression method proposed by the SCF method will work for the general heavy-tailed distribution (for $\alpha < 2$) for sufficiently small t.

3.4 TESTING ESTIMATION METHODS

We have introduced three methods that can be used to estimate the heavy-tail index: Hill's estimator, the Stable Characteristic Function (SCF) method and the General Characteristic Function (GCF) method. In order to comment on the viability of each of these methods, we now compare them directly by testing them on realisations of different processes. The methods are tested on three distinct types of processes: sets of iid stable random variables, heavy-tailed white noises and heavy-tailed coloured noises. For each sample process, the resulting estimates are labelled $\hat{\alpha}_H$ (Hill's estimate), $\hat{\alpha}_S$ (SCF method estimate) and $\hat{\alpha}_G$ (GCF method estimate). The chosen value for $\hat{\alpha}_H$ is selected using the optimised method described in subsection 3.1.1. The characteristic function methods are calculated using linear regression on the empirical characteristic function evaluated at points $t = k\pi/300$ for $k = 1, \ldots, 10$. This is approximately equal to the interval $[0.01, 0.10]$.

The results for samples of iid stable rv's are plotted in Figure 3.4.1 for sample sizes $n = 2^{10}, 2^{14}, 2^{18}$. The results clearly show that Hill's estimator provides a sufficient estimate of the heavy-tail index of stable processes if α is not near the stable threshold, and provided the sample size is large enough. When both conditions are not met however, the results can be misleading. For example, for $\alpha = 1.8$ and $n = 2^{14}$ (a relatively large data set), Hill's estimator predicts a value well outside the stable range, $\hat{\alpha}_H = 2.35$. Such a

prediction would lead us to erroneously believe that the autocorrelation was an appropriate measure of memory. These results indicate the relative success of characteristic function analysis and the relative failure of Hill's estimator. It is noteworthy to highlight that in many cases the General Characteristic Function method performs similarly to the Stable Characteristic Function method, and better than Hill's estimator, despite it being derived from a general heavy-tailed random distribution. This is in accordance with the analysis in subsection 3.3.3. This provides strong motivation and evidence for the use of characteristic function methods to estimate the heavy-tail index.

The methods are subsequently tested on samples of iid heavy-tailed random variables with the pdf

$$f_X(x) = \frac{A}{(1+x^2)^{(1+\alpha)/2}} \tag{3.4.1}$$

where A is a normalising factor and $\alpha > 0$. To test the methods both within and outside the stable regime, the heavy-tail index is varied on the interval $\alpha \in [1,3]$. Whilst previously there was only a requirement to consider the 1st-order General CF method, for $\alpha > 2$ the 2nd-order General CF method is considered. For the 2nd-order General CF method, the sample variance is subtracted from the empirical characteristic function in order to estimate the heavy-tail index. The resulting method is labelled $\hat{\alpha}'_G$. The 4 estimates ($\hat{\alpha}_H$, $\hat{\alpha}_S$, $\hat{\alpha}_G$, $\hat{\alpha}'_G$) are plotted for single realisations with varying input α with different samples $n = 2^{10}$, 2^{14}, 2^{18}. The results are plotted in Figure 3.4.2. Within the stable regime ($\alpha < 2$), both the characteristic function methods and Hill's estimator appear to be relatively successful. This is in contrast to the tests on stable processes in which Hill's estimator diverges as $\alpha \to 2$. Around the stable threshold $\alpha \approx 2$, the characteristic function methods underestimate the heavy-tail index. This is because the single-term approximation becomes less accurate, and the resultant empirical characteristic function behaves more logarithimically, as described in Equation 3.3.13.

Outside the stable regime, Hill's estimator appears sufficient providing the sample size is large enough. As expected, the Stable CF method and 1st-order General CF method saturate at $\alpha = 2$. The 2nd-order General CF method struggles and only outperforms Hill's

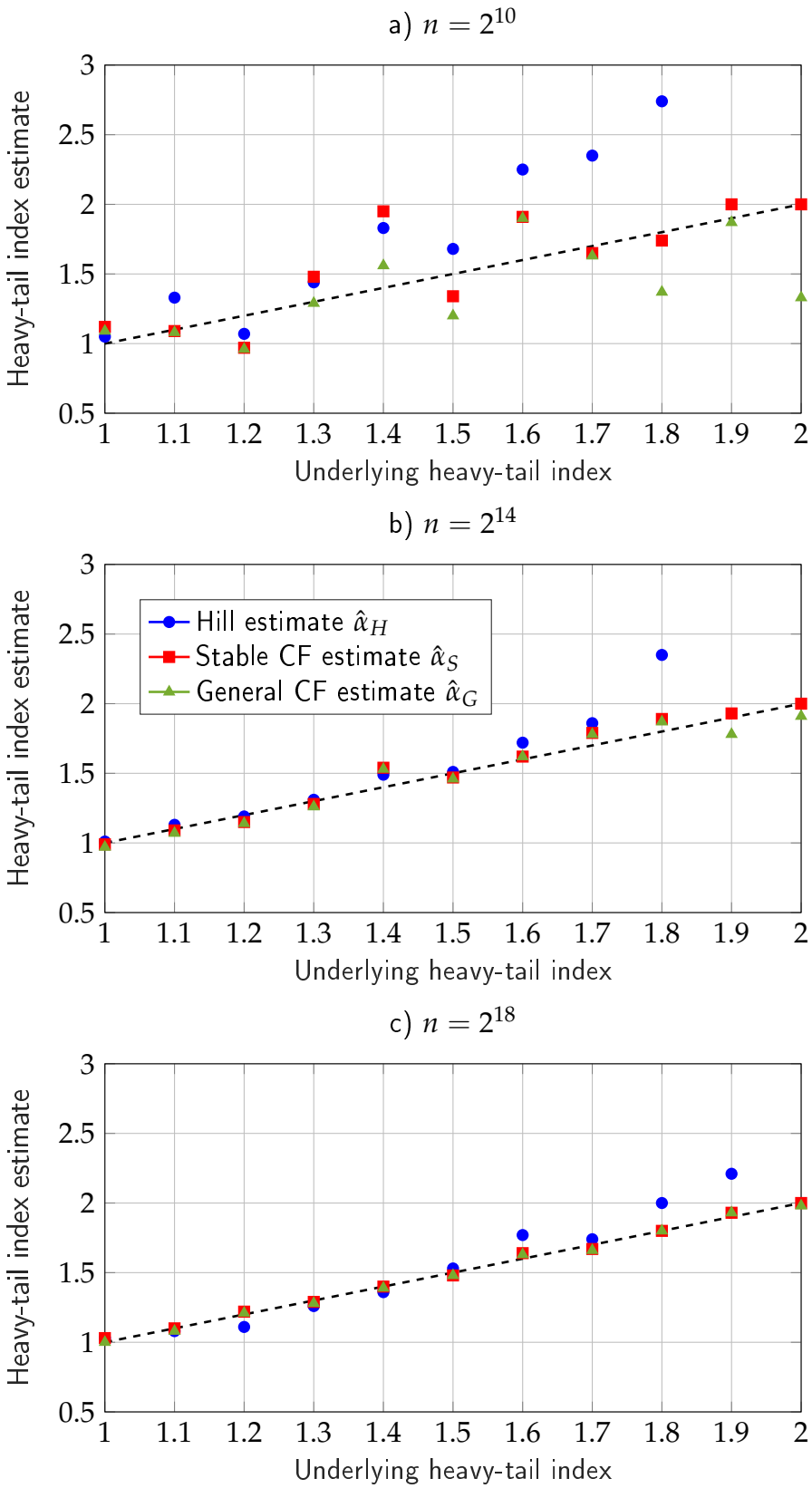

Figure 3.4.1.: Heavy-tail index estimators for sets of iid stable random variables with varying heavy-tail index α.

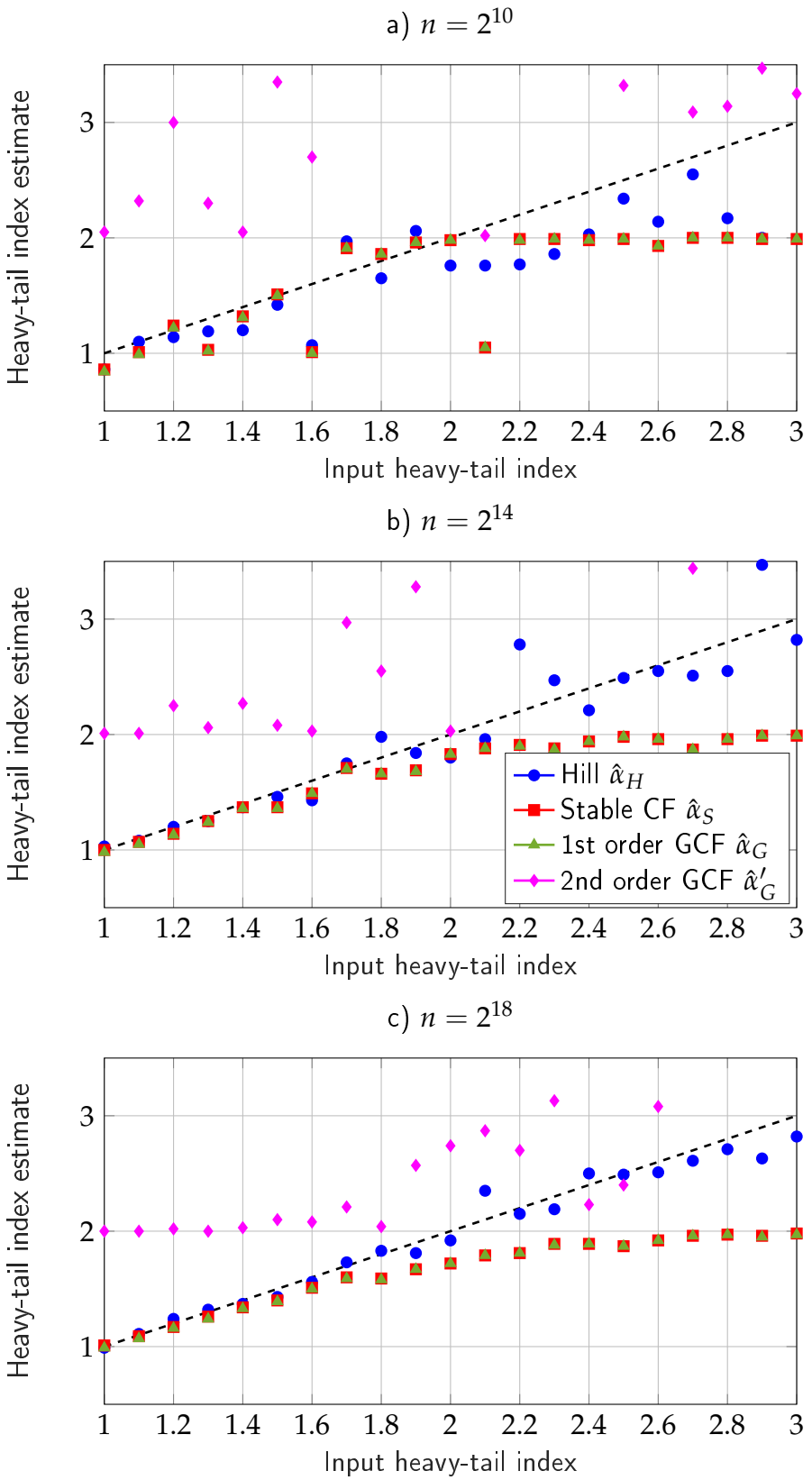

Figure 3.4.2.: Heavy-tail index estimates for samples of coloured noises with varying heavy-tail index α.

estimator on a few occasions. It is noticeably less consistent. Such results suggest a need to investigate the 2nd-order General CF method further.

Finally, the effect of memory on estimation methods is tested. As discussed in section 2.2, the autocorrelation for processes within the stable regime is an inappropriate measure of memory, and therefore only processes outside the stable regime can be considered at this point. A potential solution to consider stable processes is introduced and discussed later, in chapter 4. To test the effect of memory, we consider processes with pdf's

$$f_{X,1}(x) = \frac{A_1}{(1+x^2)^2}, \text{ for } x \in \mathbb{R}, \tag{3.4.2}$$

$$f_{X,2}(x) = \begin{cases} \dfrac{A_2}{x^4} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.3}$$

The iterative Fourier Filtering Method (iFFM) described in subsection 2.3.2 is used to generate realisations of heavy-tailed coloured noises with autocorrelation

$$a(\tau) = \exp\left(-k_a \tau^2\right)$$

where k_a is some scale factor. As described previously, the iFFM works by iteratively rearranging a set of iid random variables such that their order ensures the target autocorrelation function is replicated. Hill's estimator is independent of the order of elements however. It therefore follows that in order to test the effect of memory on a heavy-tailed coloured noise of size n, the iFFM must be used to generate a coloured noise of size m, such that $m \gg n$. The Hill estimator is then tested on an ordered subsequence of size n. This replicates the behaviour we wish to test: how memory exaggerates certain fluctuation sizes. To test this, the iFFM generates a process with $m = 2^{16}$ elements and a subsequence with $n = 2^{12}$ elements is used for testing purposes.

The iFFM is repeated to produce 500 samples. The average error between Hill's estimate and the underlying heavy-tail index is calculated over all the samples, such that $\epsilon = \langle |\hat{\alpha}_H - \alpha| / \alpha \rangle$. The errors are plotted as a function of threshold ratio η for different autocorrelation scale constants $k_a = \log(2)/\tau_{0.5}^2$ in Figure 3.4.3. From these plots, it is clear that autocorrelation and its scaling constant have a considerable impact on the

average errors. This is true for both the heavy-tailed distribution $f_{X,1}$ in Equation 3.4.2 and the Pareto distribution $f_{X,2}$ in Equation 3.4.3.

In order to display how these errors would manifest themselves in real-world data sets, we compare the Hill plots of single realisations of white noises and coloured noises. 5 realisations of white noises, and 5 realisations of coloured noises with $n = 2^{12}$ elements and the Pareto distribution in Equation 3.4.3 are produced. The coloured noises are simulated with a Gaussian autocorrelation with a scaling parameter $\tau_{0.5} = 8$. The resultant Hill plots are shown in Figure 3.4.4. Despite $\tau_{0.5} = 8$ representing a relatively modest value in the context of the average errors investigated, it has a noticeable effect on the spread of the Hill plots when compared to those of the white noises.

It is possible to mitigate the effects of short-term correlations by using larger sample sizes to ensure the long-term correlations are increasingly negligible. This can be shown by comparing the average error of Hill's estimator for coloured noises with identical autocorrelation but different sample sizes. Realisations of coloured noises are simulated with $n = 2^{12}$, 2^{13}, 2^{14} elements, the pdf $f_{X,2}$ in Equation 3.4.3 and a Gaussian autocorrelation with scaling parameter $\tau_{0.5} = 8,\ 16$. The average errors for each of these coloured noises is plotted in Figure 3.4.5 and confirms that considering larger sample sizes reduces the error of local non-negligible memory. It is noted that the inner scaling constant is effectively proportional to the sample size when it comes to considering average error. For example, the average error plot for $\tau_{0.5} = 8,\ n = 2^{12}$ is approximately equal to $\tau_{0.5} = 16,\ n = 2^{13}$.

3.5 IMPROVING THE 2ND-ORDER GCF METHOD

In section 3.3 the theoretical framework surrounding the General Characteristic Function method was introduced. It was seen in section 3.4 that the method was generally successful for $\alpha < 2$, but less so for $\alpha > 2$. Whilst some of the error can be attributed to the limiting logarithmic behaviour, the extent to which higher orders of the GCF method struggled requires addressing. The GCF method says that, for $2 < \alpha < 3$,

$$\tilde{\phi}(t) \approx 1 - \frac{1}{2}\langle X^2 \rangle t^2 + l_0 |t|^\alpha$$

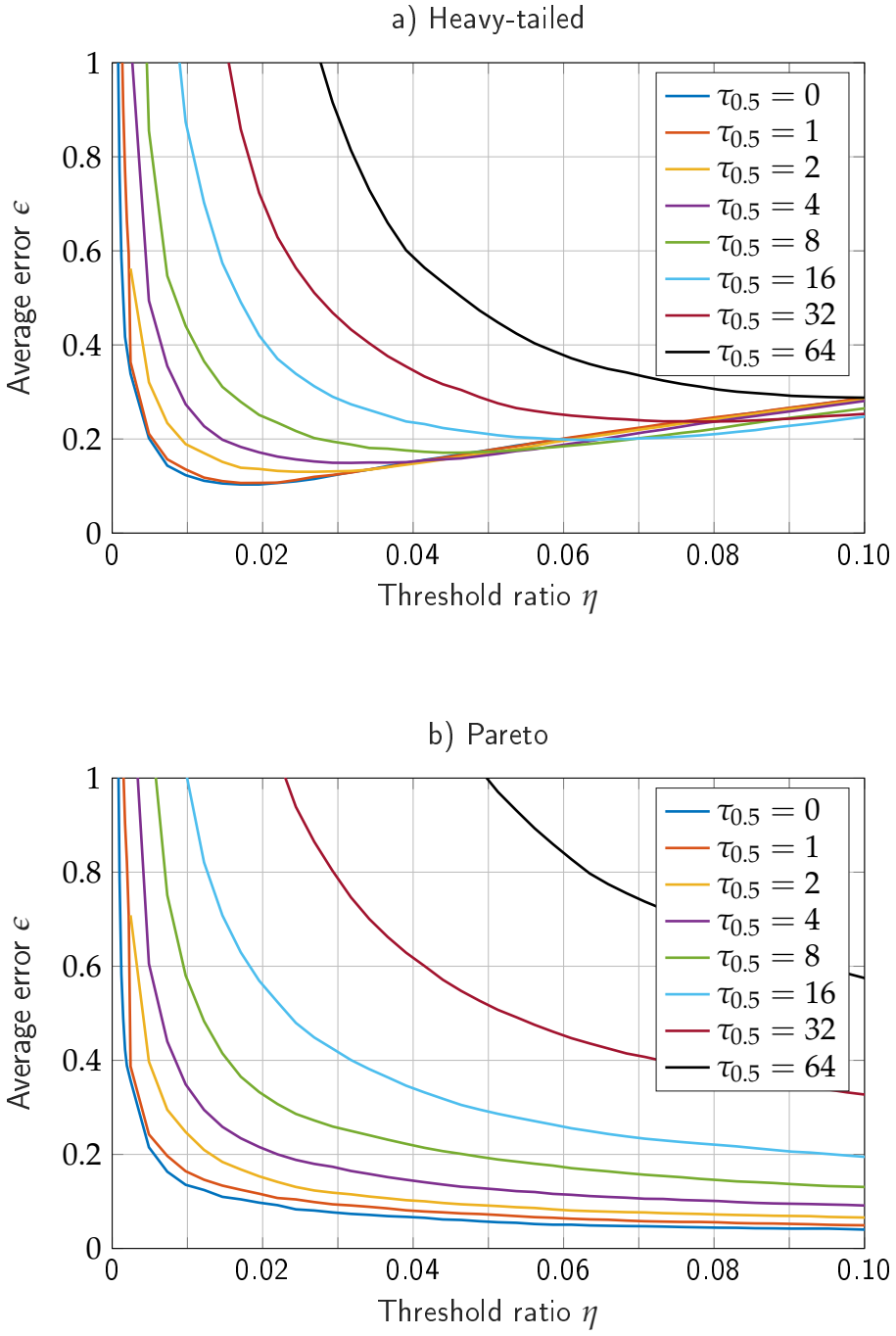

Figure 3.4.3.: Effect of memory on average error of Hill's estimator for coloured noises. Tested on processes with a) a heavy-tailed pdf $f_{X,1}$ (Equation 3.4.2) and b) a Pareto pdf $f_{X,2}$ (Equation 3.4.3) .

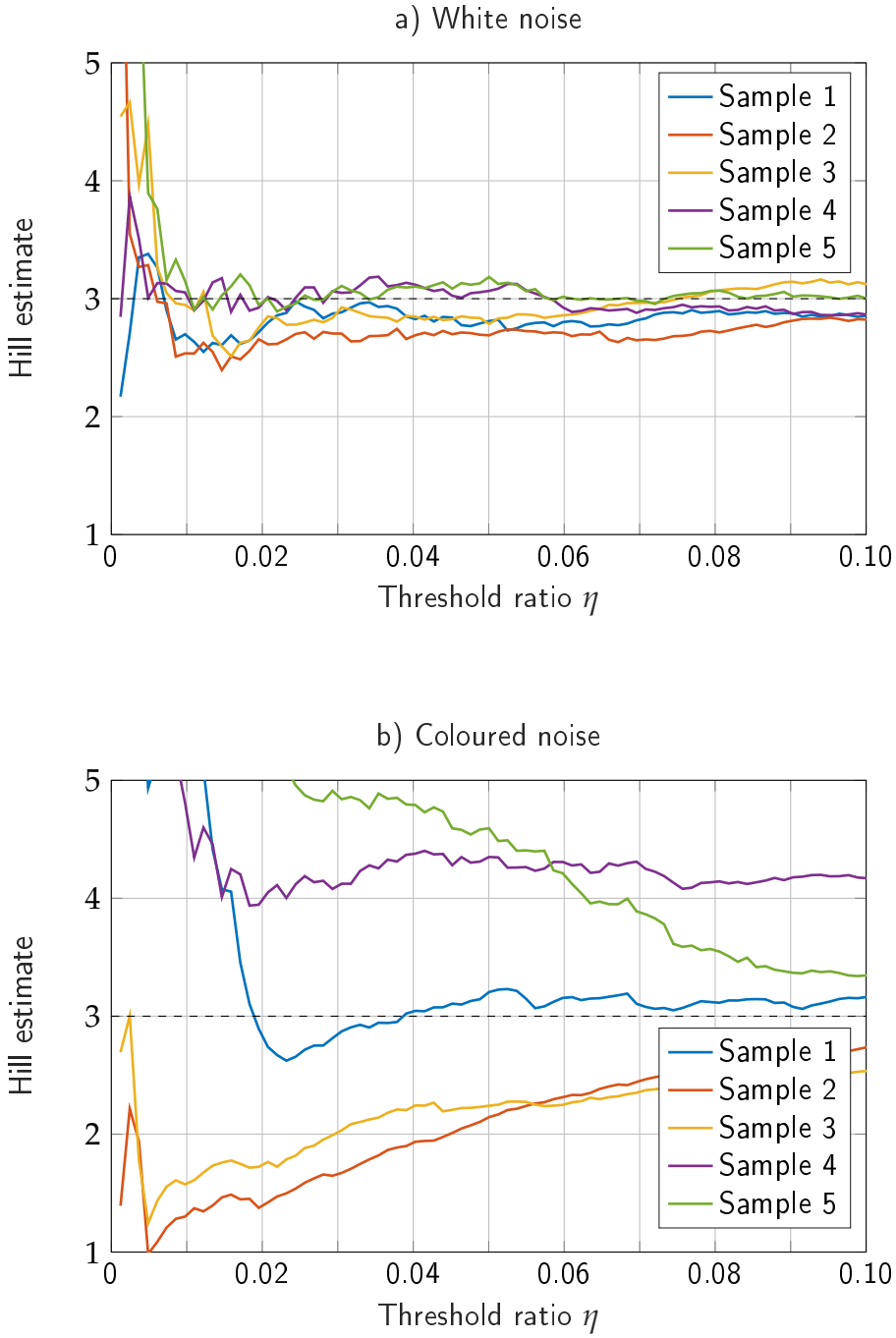

Figure 3.4.4.: Hill plots for 5 realisations of heavy-tailed processes. Tested on a) white noises and b) coloured noises with Gaussian autocorrelation with $\tau_{0.5} = 8$.

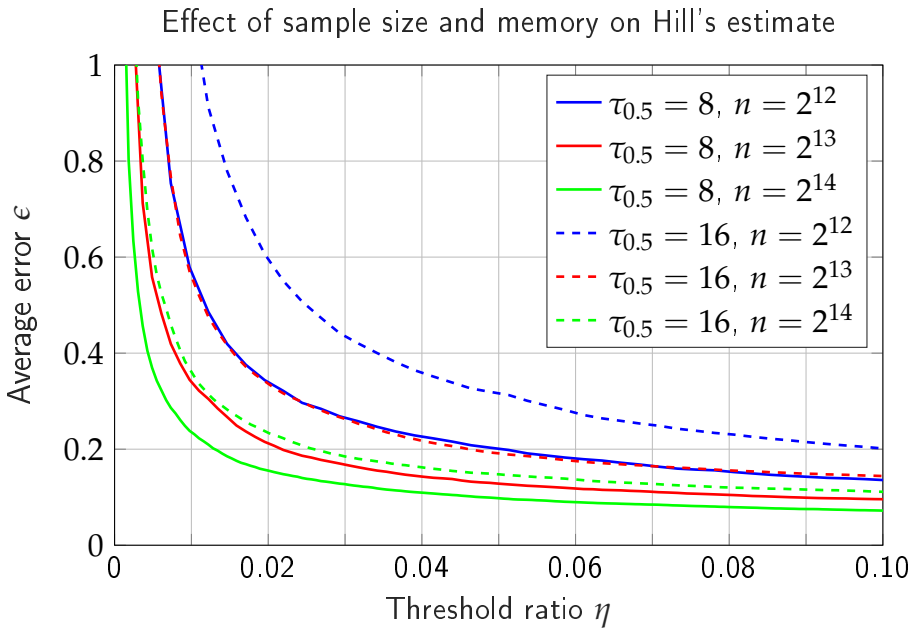

Figure 3.4.5.: Average error of Hill's estimator for coloured noises with scaling constants $\tau_{0.5} = 8,\ 16$ and sample sizes $n = 2^{12},\ 2^{13},\ 2^{14}$.

and therefore

$$\log\left(\tilde{\phi}(t) - \left(1 - \frac{1}{2}\langle X^2\rangle t^2\right)\right) \approx \log(l_0) + \alpha \log|t|\,. \tag{3.5.1}$$

In section 3.3 Equation 3.5.1 is computed for $t = t_k = \pi k/300$ for $k = 1,\ \ldots,\ 10$. This is approximately equal to the interval $[0.01, 0.10]$. This was chosen as previous tests suggested for $t < 10^{-2}$ the complementary characteristic function was noisy for smaller α and smaller data sets. This is not necessarily true, however, for all α. This can be observed in Figure 3.3.4 where, for $\alpha = 1.0,\ 1.5$ the empirical complementary characteristic functions are very noisy - unlike for $\alpha = 2.5$. However, because a consistent time interval was sought for all cases, $t > 10^{-2}$ was used. To investigate further, we look at a realisation with $n = 2^{18}$ elements drawn from iid random variables with the pdf

$$f_X(x) = \frac{A}{(1+x^2)^{(1+\alpha)/2}} \overset{a}{\sim} |x|^{-(1+\alpha)} \tag{3.5.2}$$

where A is a normalising constant and $\alpha = 2.5$. We then calculate Equation 3.5.1 for different values of t. For consistency, we look at a single realisation. We initially look at the the evaluation points $t = 0.01,\ \ldots,\ 0.10$ (as per the original testing). The resulting points are plotted in Figure 3.5.1 on both a linear and logarithmic scale. Using least-squares

regression on a logarithmic scale, the prediction for the heavy-tail index is $\hat{\alpha}_G = 3.33$. From the plot, the relative variations suggest the power-law fit might not be appropriate. To investigate further, we repeat the method but on a larger range of evaluation points, $t = 0.001, \ldots, 1.00$. A power-law function with the desired exponent $\alpha = 2.5$ is then fitted to the data to see whether this could describe the empirical behaviour. The behaviour is plotted in Figure 3.5.2 on both a linear and logarithmic scale. On the linear scale it is clear that the power-law function is accurate for a much larger range than previously postulated. Conversely, the logarithmic scale shows that there is a cusp around $t \approx 0.05$ where the empirical characteristic function deviates significantly from the fitted power-law function. This would explain the large error in the initial estimate of $\hat{\alpha}'_G$. We therefore test the linear regression on another region $t = 0.1, \ldots, 1.0$. The resulting estimate is $\hat{\alpha}'_G = 2.62$. This result is far more accurate than originally encountered. Whilst we have shown that considering too small value t will see the method fail, it is clear that we can't consider too large t. It is clear that $\tilde{\phi}_2(t) \overset{a}{\sim} t^2$ as $t \to \infty$ (where $\tilde{\phi}_2(t)$ is defined in Equation 3.3.9).

This new information prompts us to retest the 2nd-order general characteristic function method over the range $\alpha \in (1,3)$. To do so, we generate a realisation with $n = 2^{18}$ elements drawn from iid random variables with the pdf in Equation 3.5.2 and alter the heavy-tail index from $\alpha = 2.1, \ldots, 2.9$. The heavy-tail index is estimated on two regions: $\mathcal{R}_1 = [0.01, 0.10]$ (as done originally), and $\mathcal{R}_2 = [0.10, 1.00]$ (prompted by the work here). The results are plotted in Figure 3.5.3. From these plots, it is clear that this new region of consideration has considerably improved the estimates. It is worth noting that these regions were used regardless of any visual clues as to where the optimal region might be.

Whilst we have shown how improvements of the General Stable Characteristic Function method can be made, there is still work required to establish how this would be practically implemented. It appears likely that there is a unique optimal region to fit a power-law function for each heavy-tail index. Importantly, the trade off between the limiting behaviours has to be understood. The iterated characteristic function $\tilde{\phi}_2(t) \overset{a}{\sim} t^2$ as $t \to \infty$ whilst the estimate becomes noisy as $t \to 0$.

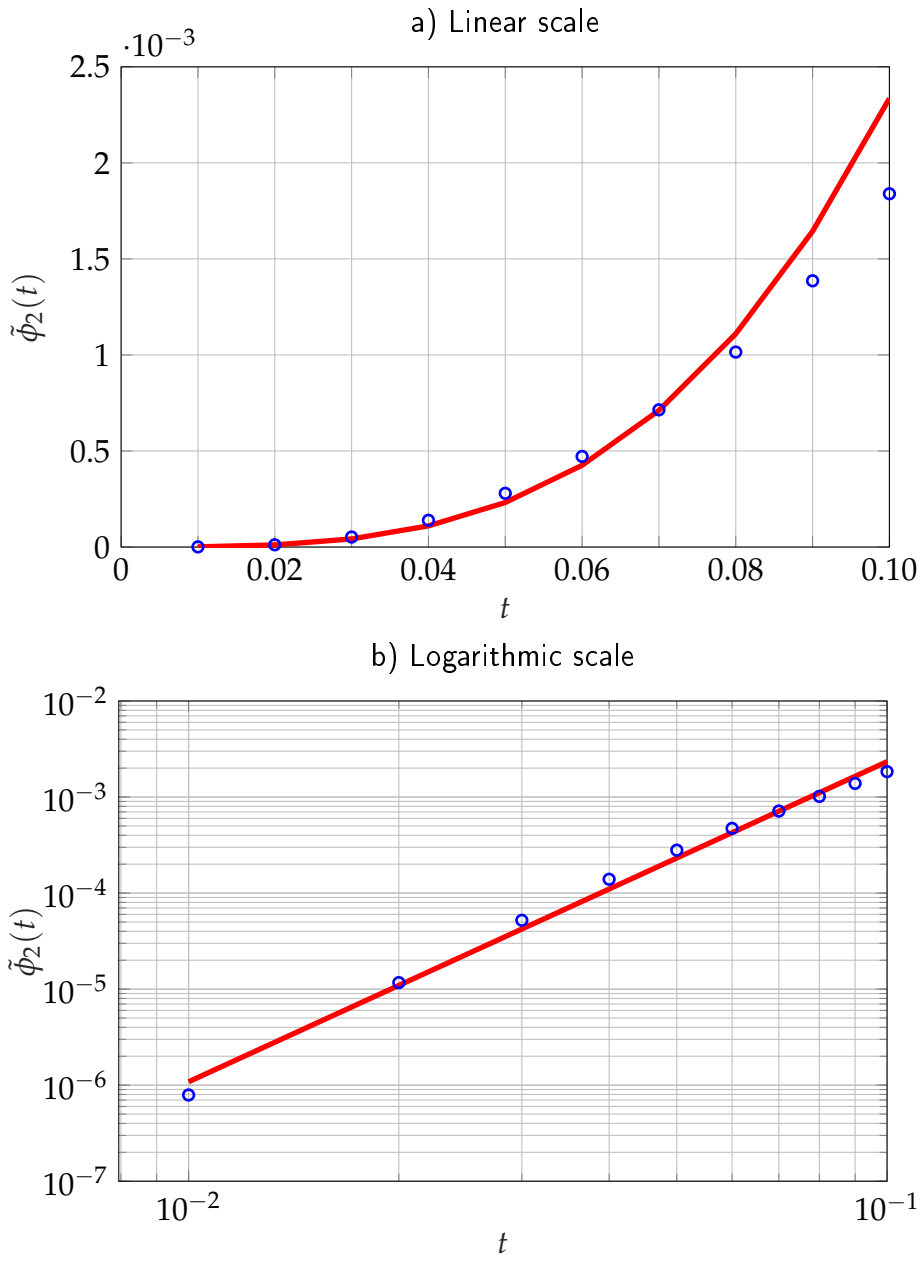

Figure 3.5.1.: Empirical characteristic function of heavy-tailed process with index $\alpha = 2.5$. The empirical characteristic function is computed for $t = 0.01, \ldots, 0.10$. A linear function is fitted on the logarithmic scale, resulting in a power-law fit $\tilde{\phi}_2(t) \approx k|t|^{3.33}$.

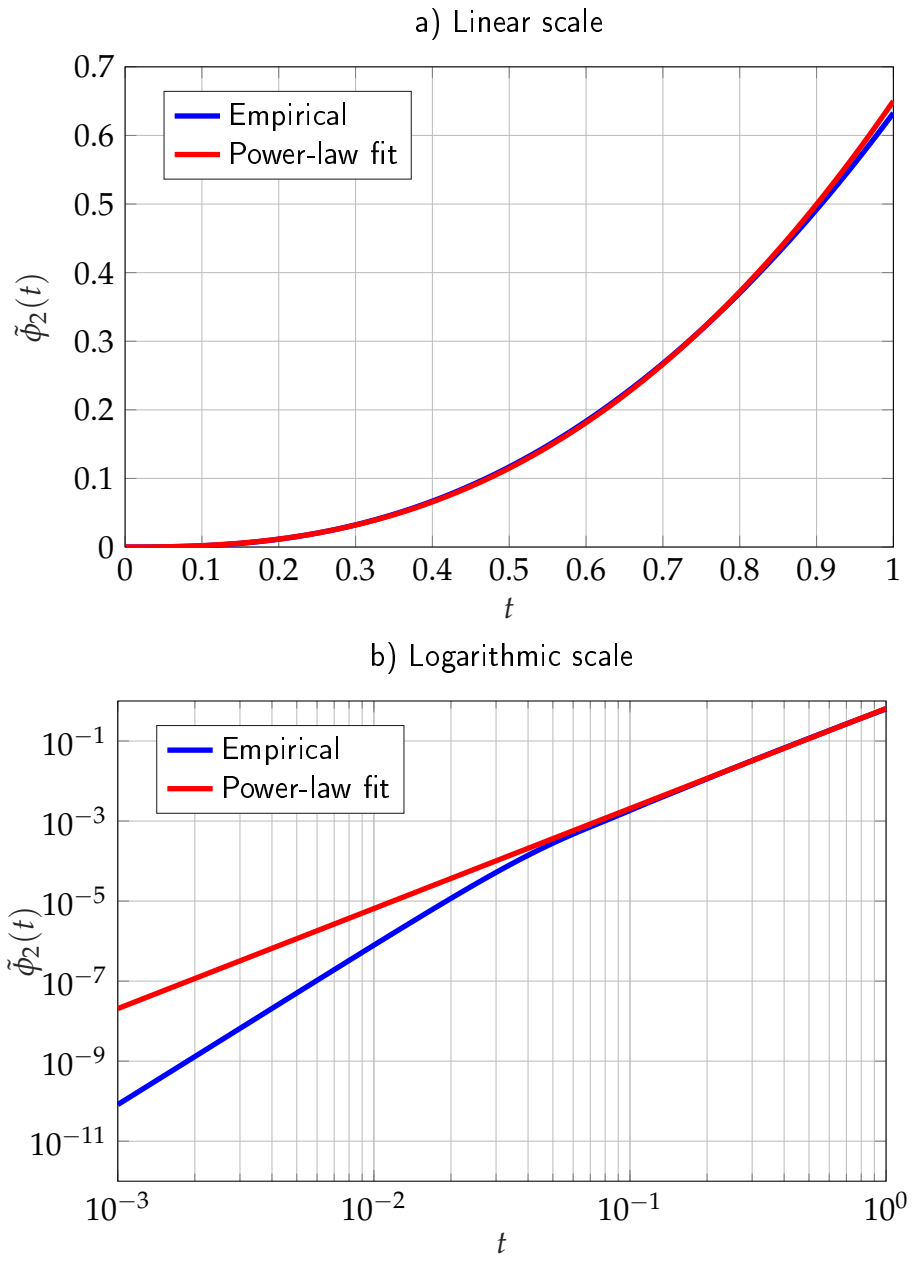

Figure 3.5.2.: Empirical characteristic function of heavy-tailed process with index $\alpha = 2.5$. The empirical characteristic function is computed for $t = 0.001, \ldots, 1.00$. A power-law function with the underlying heavy-tail index is fitted to the empirical process.

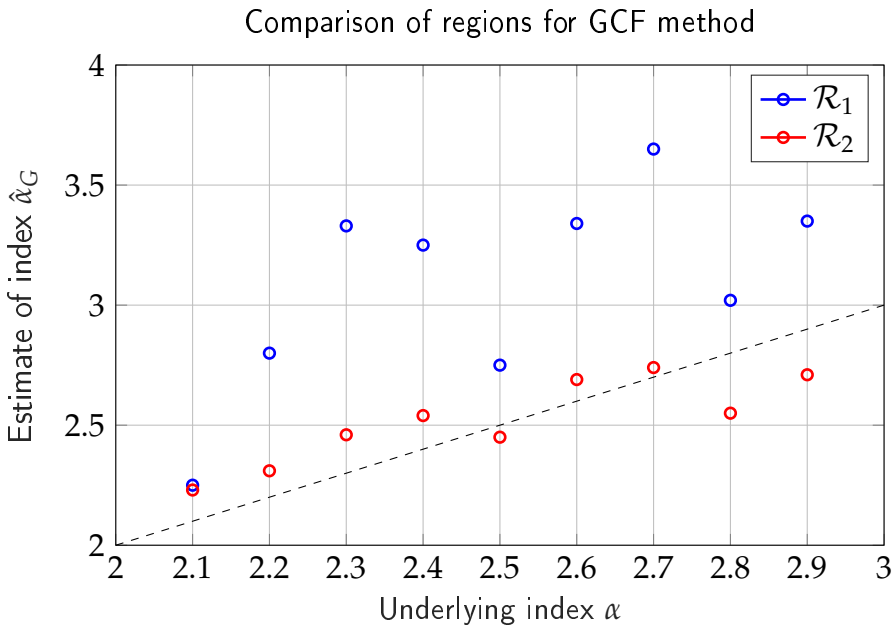

Figure 3.5.3.: Comparison of General Characteristic Function analysis on two regions, $\mathcal{R}_1 = [0.01, 0.10]$, $\mathcal{R}_2 = [0.1, 1.0]$. Fitting on $\mathcal{R}_2$ clearly gives an improved estimate.

3.6 THE IBM STOCK

We have so far only discussed the use of tail estimation methods on synthetic data. This has allowed us to compare empirical estimates of the heavy-tail index with the known underlying value, and comment on the efficiency of the estimators. It is useful, however, to give an example of its implementation on empirical data that is (purpoted to be) a non-Gaussian stable process. The use of stable distributions as a model for financial objects is not uncommon, having first been used as such by Mandelbrot when modelling cotton prices [3]. The motivation for doing so stems from the desire for self-similar models that could describe larger fluctuations more accurately than the Gaussian distribution.

An example of a stable model being used is Nolan's investigation of the IBM stock [51]. Nolan looked at the daily log-returns of the IBM stock over a 10-year period, from 01/01/2003 - 31/12/2012 [52]. We have taken the same data and investigated the tail

behaviour using the methods discussed in this chapter. The log-returns of the IBM stock are defined as

$$X_j := \log \frac{C_j}{C_{j-1}}$$

where $\{C_j\}$ is the set of adjusted daily close prices. The adjusted close prices and log-returns are plotted in Figure 3.6.1 and Figure 3.6.2 respectively. By using the Augmented Dickey-Fuller Test [53] in MATLAB, the set of log-returns is determined to be stationary to a 5% significance level. Nolan subsequently fitted a non-Gaussian stable distribution and a Gaussian distribution to the data using numerical likelihood estimation. The stable model had parameters $\alpha = 1.614873$, $\beta = 0.00000$, $\gamma = 0.007331$, $\delta = 0.000499$ and the Gaussian model had $\mu = 0.000400$, $\sigma^2 = 0.000192$. The sample pdf of the data is plotted alongside the model pdf's in Figure 3.6.3. From the plot it is clear that the stable distribution generally models the behaviour well, especially when compared to the Gaussian distribution. This is in itself is a useful conclusion, that if a self-similar model of the stock's behaviour is required then the non-Gaussian stable model offers a considerable improvement on the standard Gaussian.

It is not certain from this initial fitting, however, that the daily-returns are asymptotically stable. We therefore use the same data and apply our tail estimation methods to comment on the viability of the stable model in the tail. If the underlying behaviour is asymptotically stable, then the process will have those properties that we wish to investigate, such as undefined variance and autocovariance. Consequently, the IBM daily-returns (and similar data sets) could be used to test the coherence function as a means to quantify memory, as discussed in subsection 2.2.2. A Hill plot of the daily-returns is plotted in Figure 3.6.4. The heavy-tail index according to Nolan's stable model, $\alpha \approx 1.61$ is also plotted for comparison. The Hill plot shows no convergent 'horizontal' region from which the optimal Hill estimate might be taken, and any chosen threshold ratio (less than 10%) would give a heavy-tail estimate outside the stable regime. The data is also investigated using characteristic function analysis. The empirical characteristic function of the data, $\tilde{\phi}(t)$ is constructed for $t \in [-500, 500]$ and plotted in Figure 3.6.5, alongside $1 - \sigma^2 t^2/2$ where σ^2 is the sample variance. The complementary characteristic function $\bar{\phi}(t) = 1 - \tilde{\phi}(t)$ is plotted on a logarithmic scale in Figure 3.6.6 alongside $\sigma^2 t^2/2$. Both of the characteristic function plots

Figure 3.6.1.: Adjusted close price of the IBM stock from 2003-2012 from [52].

show that the small t behaviour is accurately described by $\overset{a}{\sim} t^2$. This corroborates Hill's estimator that the process is not asymptotically stable.

To further validate this conclusion, the complementary cdf is plotted on a logartihmic scale in Figure 3.6.7. It is clear that the tail is heavier than the Gaussian model, but not as heavy as the stable model. This confirms the predictions of both Hill's estimate and characteristic function analysis that both the stable and Gaussian models are erroneous in the tail. We now use both Hill's estimator and characteristic function analysis to determine if it is heavy-tailed in general and, if so, what the heavy-tail exponent is. To do so we first use the Hill plot and choose threshold ratios of $\eta = 0.01,\ 0.05,\ 0.08$. The resulting estimates for the heavy-tail index are $\alpha_H = 4.04,\ 2.77,\ 2.55$ respectively. The tail models are plotted against the empirical complementary cdf in Figure 3.6.8. There appears to be no convergence, with the latter two estimates clearly failing to adequately model the tail behaviour. The threshold ratio of $\eta = 0.01$ shows the best estimate for tail (as expected) but corresponds to only 26 data points and is therefore a relatively volatile estimate. We

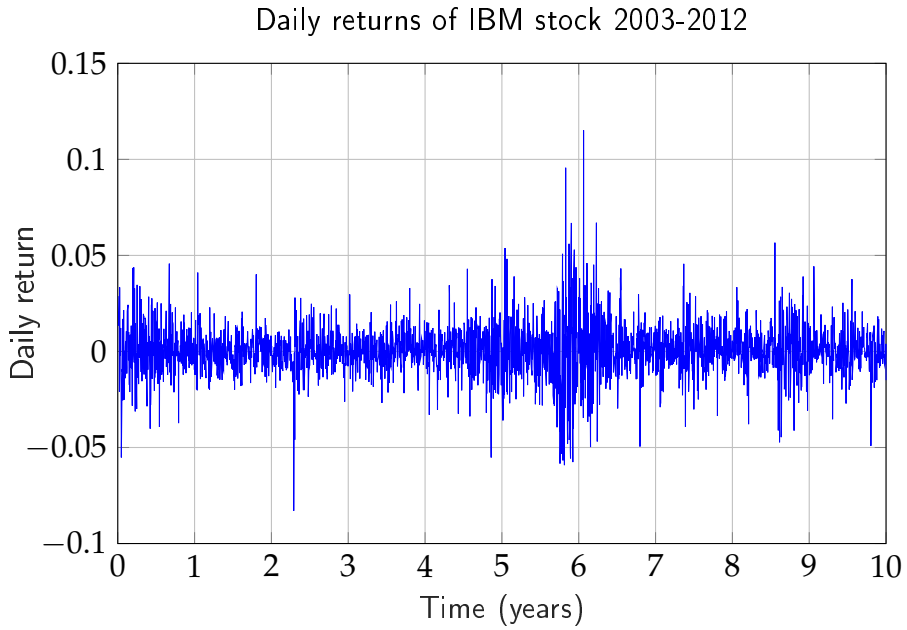

Figure 3.6.2.: The daily returns of the IBM stock from 2003-2012, calculated by $X_j = \log(C_j/C_{j-1})$ where $\{C_j\}$ is the adjusted close prices.

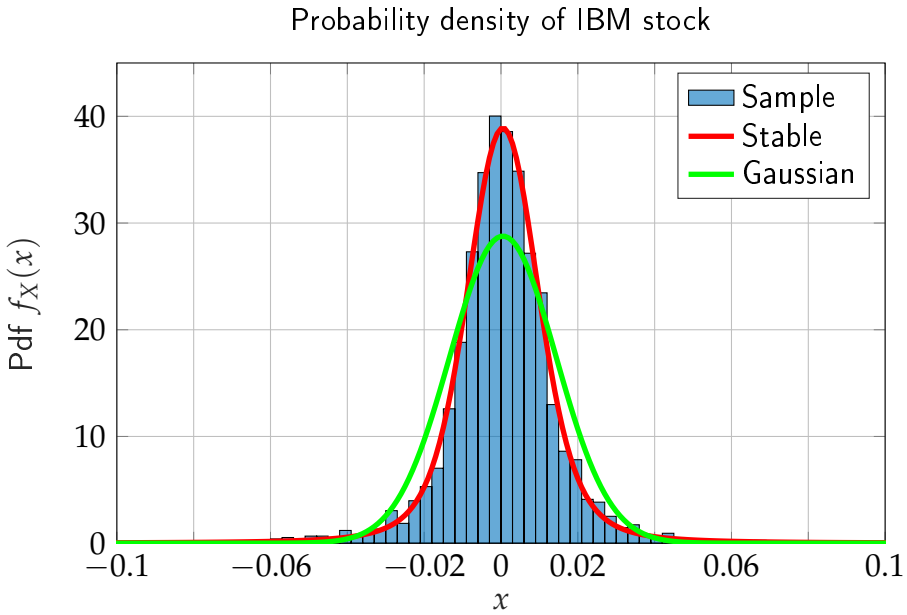

Figure 3.6.3.: Sample pdf of the daily returns, along with the stable and Gaussian models.

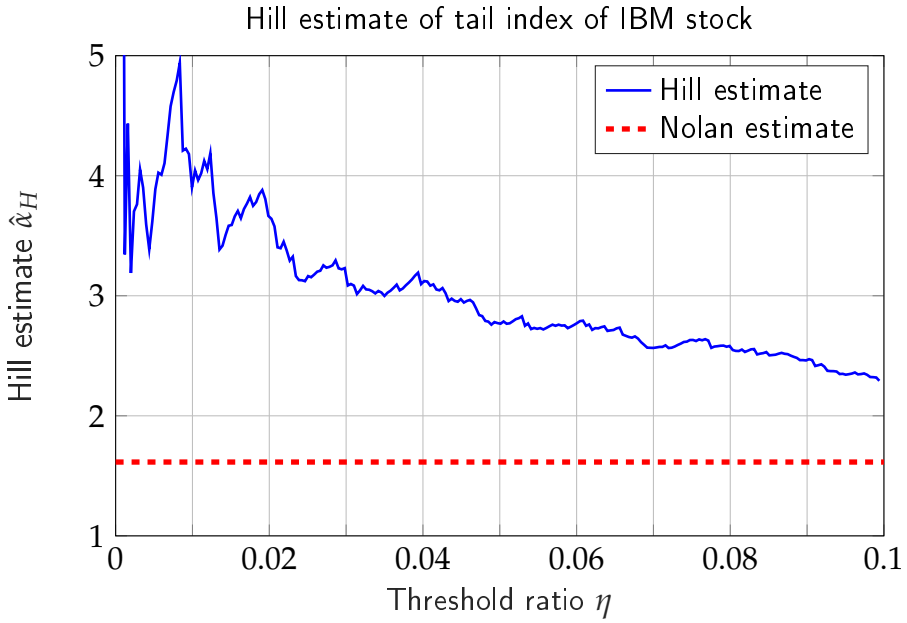

Figure 3.6.4.: Hill plot of the daily log-returns of the IBM daily returns up to a maximum threshold ratio of 10%.

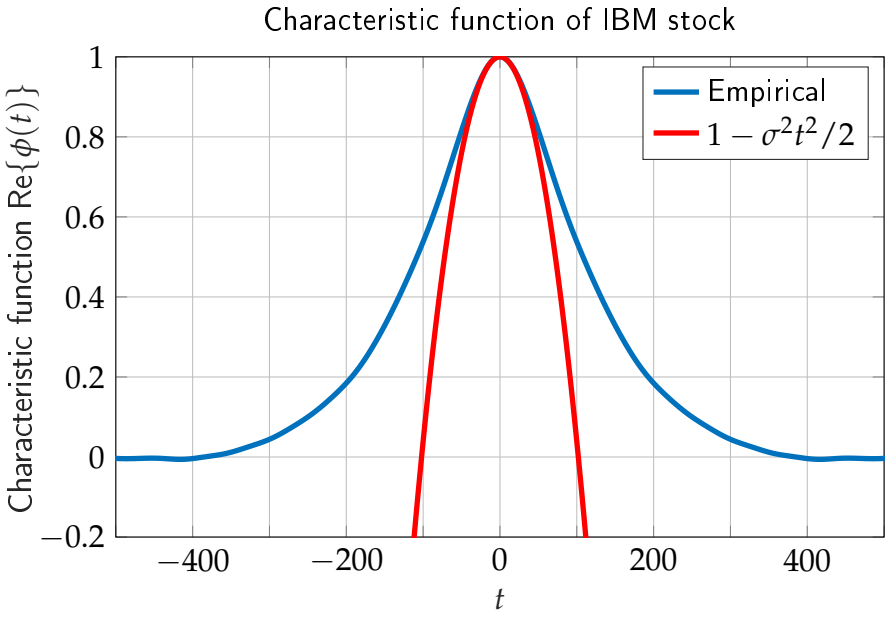

Figure 3.6.5.: Empirical characteristic function of daily returns of IBM stock. Also plotted the Taylor expansion using the sample variance. Imaginary components of the characteristic function not plotted.

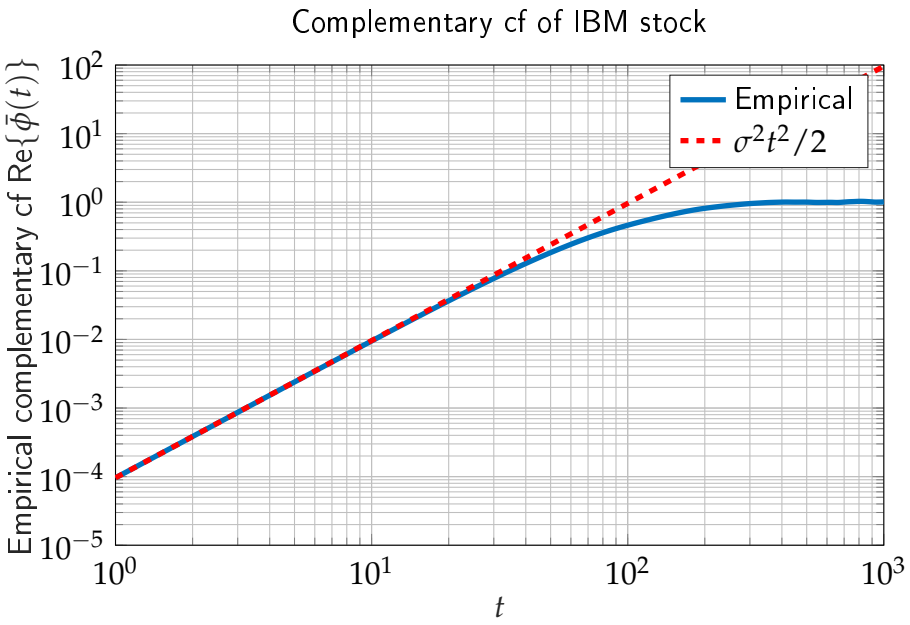

Figure 3.6.6.: Empirical complementary characteristic function of daily returns of IBM stock, plotted on a logarithmic scale. Displays that the behaviour for small t corresponds to the sample variance term with exponent $\overset{a}{\sim} t^2$.

now consider characteristic function analysis. The 2nd moment is subtracted from the empirical characteristic function to give the iterated empirical characteristic function

$$\tilde{\phi}_2(t) = \phi(t) - \left(1 - \frac{1}{2}\sigma^2 t^2\right).$$

$\tilde{\phi}_2(t)$ is plotted against $\mu_4 t^4/4!$ where $\mu_4 = 1.26 \times 10^{-8}$ is the sample raw kurtosis in Figure 3.6.9 and on a logarithmically scaled plot in Figure 3.6.10. From Figure 3.6.5 it is observed that the characteristic function has a relatively large spread. This is also reflected in the small values of the integer moments (including kurtosis). The result is that $\mu_4 t^4/4! < 1$ for $t < 94.4$ and small t behaviour for $\tilde{\phi}_2(t)$ is observed for a relatively large range of t. It is clear that for small t behaviour the truncated empirical characteristic function is behaving according to $\overset{a}{\sim} t^4$. The cusp around $t \approx 4$ is a result of the resolution limit of the empirical characteristic funtion, as discussed in section 3.5. This suggests that the 4th moment exists and, if the process is heavy-tailed, the heavy-tail index is greater than 4. By collaborating the results from Hill's estimator and the characteristic function method, it is concluded that there is insufficient evidence to classify the

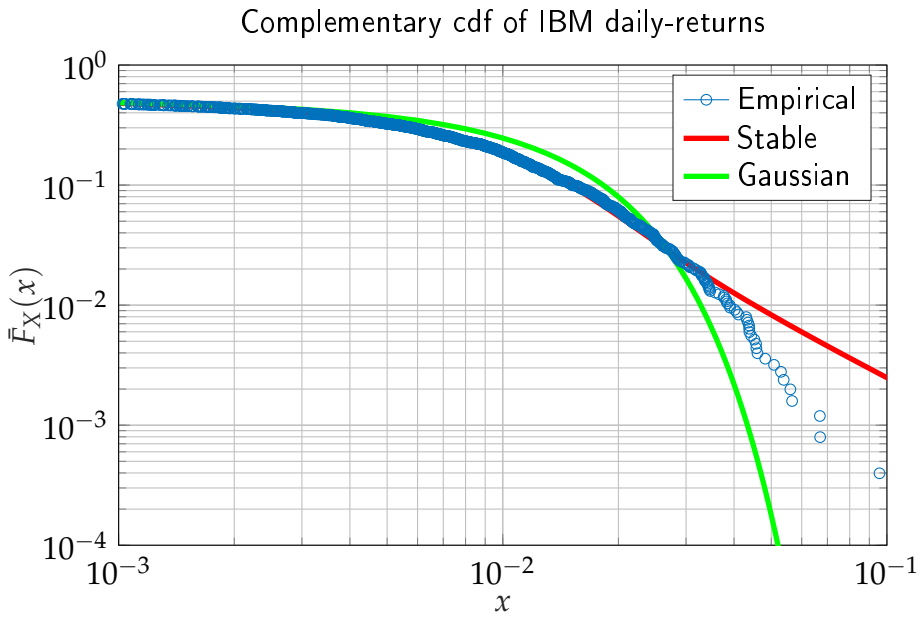

Figure 3.6.7.: Complementary cdf of IBM daily-returns, plotted against those predicted by Gaussian and Stable models.

process as heavy-tailed and, if it is, then the heavy-tail index is greater than (or equal to) 4.

The results again highlight the difficulty in correctly identifying and classifying heavy-tailed random variables. Whilst the stable distribution is a valid model for the IBM log-returns for central values, and is useful because of its self-similarity, it would be erroneous to investigate the tail properties of stable processes with them. It therefore remains an important facet of the investigation to identify empirical correlated stable data sets which can be investigated.

Although not applicable for investigating the coherence, we can calculate the autocorrelation function of the IBM stock which will be well-defined as $\{X_j\}$ is not asymptotically stable. Additionally, we calculate the autocorrelation of the absolute daily-returns, $\{X_{\text{abs},j}\}$ defined by $X_{\text{abs},j} := |X_j|$. The results are plotted in Figure 3.6.11. We see that whilst $\{X_j\}$ is δ-correlated, $\{X_{\text{abs},j}\}$ displays significant correlations. This phenomenon is referred to as 'volatility clustering' and suggests that large fluctuations are likely to be surrounded by more large fluctuations, and small fluctuations are likely to be surrounded by more small fluctuations. The lack of autocorrelation in the raw log-returns $\{X_j\}$, however, means that there is no correlation in the sign of the fluctuations. For example, a large positive

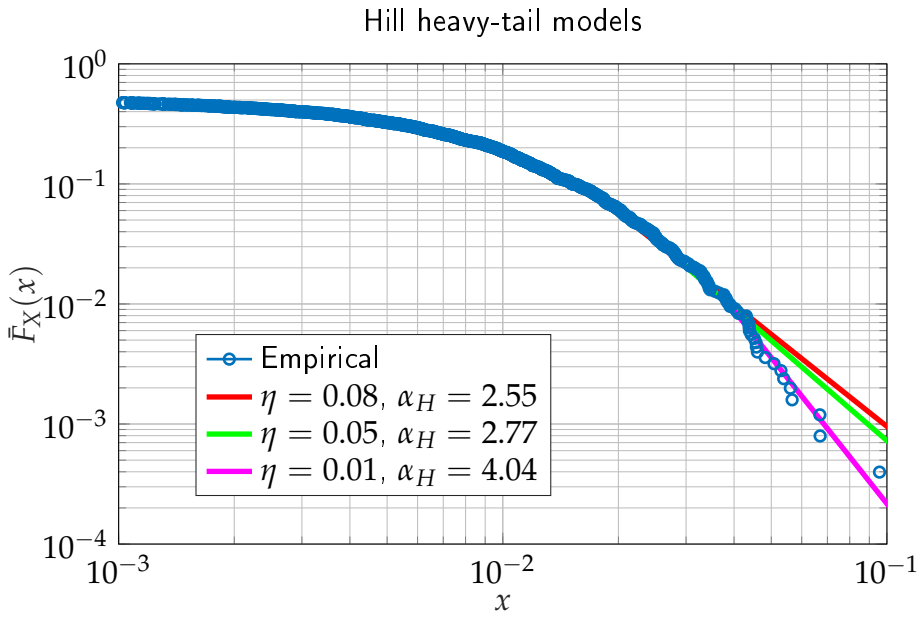

Figure 3.6.8.: Complementary cdf of IBM daily-returns, plotted with heavy-tail models calculated by Hill's estimator for 3 different threshold ratios.

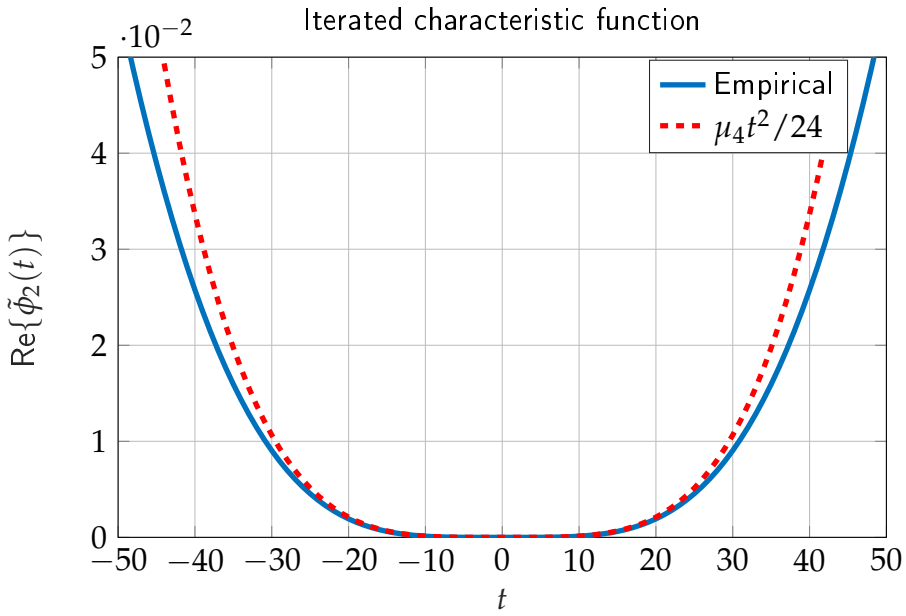

Figure 3.6.9.: 2nd iterated characteristic function of IBM daily-returns $\tilde{\phi}_2(t) = \tilde{\phi}(t) - (1 - \sigma^2 t^2/2)$. Plotted against $\mu_4 t^4/4!$ where μ_4 is the sample kurtosis.

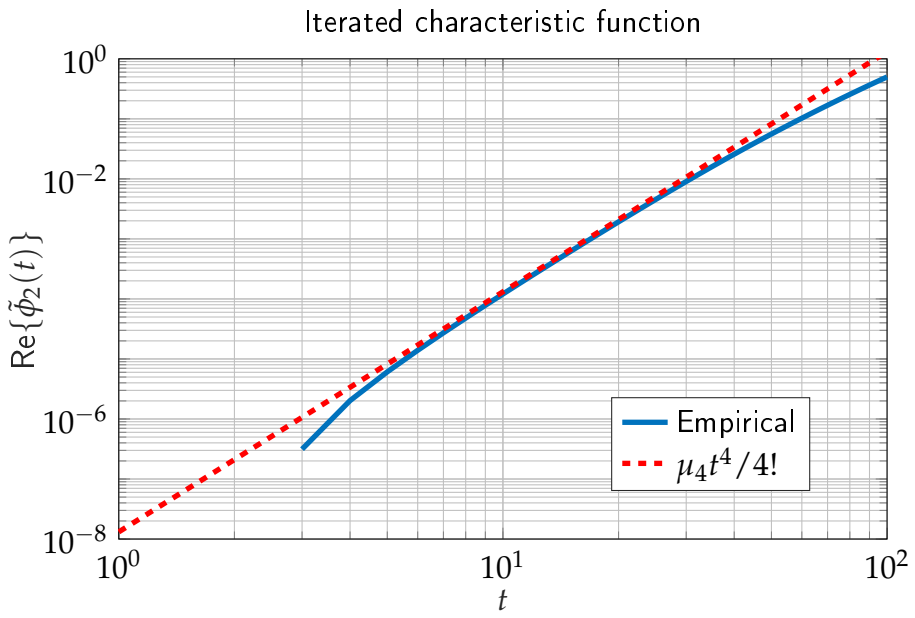

Figure 3.6.10.: 2nd iterated characteristic function of IBM daily-returns $\tilde{\phi}_2(t) = \tilde{\phi}(t) - (1 - \sigma^2 t^2/2)$. Plotted against $\mu_4 t^4/4!$ where μ_4 is the sample kurtosis on a logarithmic scale.

realisation is just as likely to be followed by a large negative realisation as a large positive realisation. This ensures there is no opportunity to predict if the IBM stock value is likely to increase or decrease (more than average) based on its previous realisation.

3.7 SUMMARY

Estimating the heavy-tail index of a heavy-tailed process is problematic. In this chapter two established identification methods (Hill's estimator and the Stable Characteristic Function method) have been discussed.

By extending the methodology of the Stable Characteristic Function method, a General Characteristic Function method for heavy-tailed random variables has been introduced. It was shown that there exist potential uses and applications for these ideas, with the ability to investigate heavy-tail indices outside the stable regime. It was shown that the method has considerable practical problems, however. As the heavy-tail index approaches integer

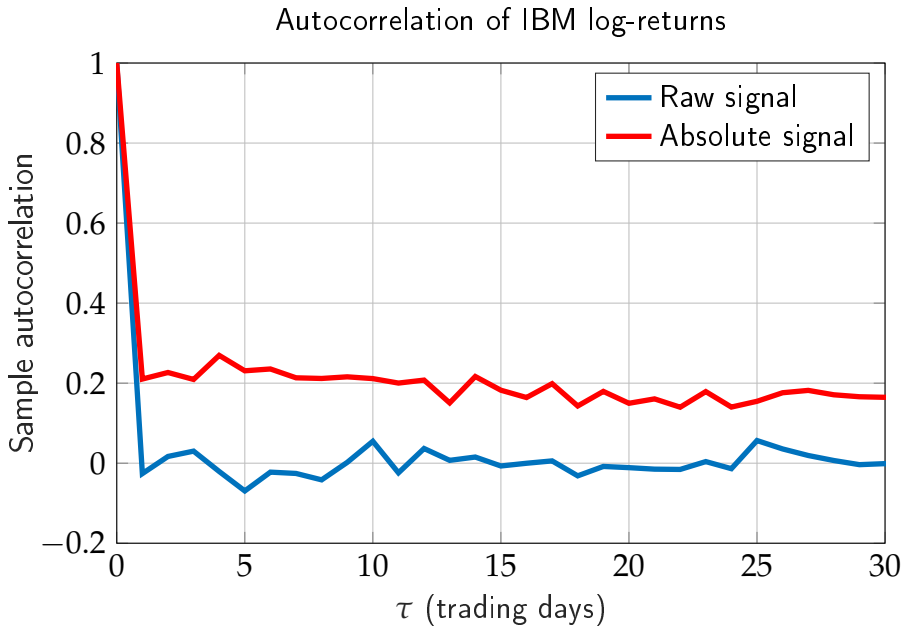

Figure 3.6.11.: Sample autocorrelation functions of the log-returns of the IBM stock and the absolute value of the log-returns. Plotted up to time lag $\tau = 30$ days $\approx$ 1 month

moments (we specifically investigated the stable threshold $\alpha = 2$), it was shown that the leading-term approximations become less appropriate, and the limiting behaviour is logarithmic. Such a result has an adverse effect on the accuracy of the heavy-tail estimate when testing. Furthermore, there were clear problems when considering higher-powers. Specifically, as integer moments are removed from the characteristic function, the leading-term approximation becomes more noisy. In general however, the adapted method showed value. It performed similarly to the Stable Characteristic Function method in our tests, and outperformed Hill's estimator when considering stable processes near the stable threshold $\alpha = 2$.

The effect of memory on extracting the heavy-tail index was also investigated. With no means (yet) introduced to generate coherent stable noise, the effect of the autocorrelation on Hill's estimator was tested. It was shown that the size of the autocorrelation function could vastly alter the success and reliability of Hill's estimator. The practical interpretation of this is that when considering empirical samples, certain fluctuations are emphasised

when there is significant autocorrelation (e.g. larger fluctuations will be clumped resulting in the tail appearing 'heavier'). The relationship between expected errors and sample size was shown, as well as the resultant expectations for individual realisations of both white noises and coloured noises.

Whilst the potential pitfalls and problems in attempting to estimate the heavy-tail index have been shown, it is useful to summarise what is known to reduce errors from certain sources. We also note that the investigation is restricted to stationary processes. When handling empirical data, it should be established whether or not the data is stationary. This can be done through a variety of methods, including the Augmented Dickey-Fuller Test [53]. By considering the below, it allows a methodology to optimise estimation techniques depending on what is known about the process.

1. Sample size: Regardless of identification method or underlying distribution function, larger sample sizes result in more accurate estimates. When looking at empirical data, knowledge of the sample size can be used in conjunction with the results from this chapter to ascertain how accurate the estimation methods might be.

2. Heavy-tail estimate method: Hill's estimator offers a good first estimate of the heavy-tail index. If it suggests the heavy-tail index is sufficiently below the stable threshold $\alpha = 2$, then the tail is sufficiently heavy that Hill's estimator is appropriate. Characteristic function methods can then be used to interrogate this initial estimate. If Hill's estimate is close to the stable threshold (or above it), then characteristic function methods should be used (especially if it is known that the underlying distribution is stable). If characteristic function methods suggest the variance exists, then Hill's estimator can be used as the most appropriate method. If characteristic function methods suggest the heavy-tail index is within the stable regime, then these can be taken as the heavy-tail index estimate. It should be noted that both methods can be used alongside each other. As noted in our tests, the success of methods is not consistent, and can be dependent on the underlying distribution function.

3. Memory: If the heavy-tail index is outside the stable regime, then the autocorrelation function represents an appropriate measure of memory for the process. This can then

be used to investigate the return-times and zero-crossing properties. If the heavy-tail index is within the stable regime, then the coherence function should be used. In these cases, an accurate estimate for α must be used to extract the underlying coherence function. When memory properties are known, this can be used to further comment on the apparent accuracy of the initial heavy-tail index estimate.

When dealing with heavy-tailed processes, there is an innate difficulty and uncertainty in estimating the heavy-tail index. For nearly all empirical and practical data sets, the amount of data in the tail is limited. These difficulties are amplified when considering processes with a dependency structure (e.g. coloured noises). Characteristic function analysis has clear advantages over Hill's estimator in certain circumstances - especially regarding heavy-tailed processes near the stable threshold. It has been shown that general characteristic function analysis has a theoretical grounding to investigate general heavy-tailed processes and not just stable processes. The result is that whilst characteristic function analysis can not be declared as an altogether better tool for estimating the heavy-tail index for all processes, there certainly exists enough evidence for it to be considered a useful addition. There also exists clear scope to continue investigations into how characteristic function methods can be optimised and improved.

Both Hill's estimator and the characteristic function methods were used to investigate the IBM daily-returns in section 3.6. We saw that whilst a stable model was generally successful in characterising the behaviour, both methods concluded that it was asymptotically non-stable. Such a result is a cautionary tale for using stable models. Even if the behaviour is stable in the central region of the distribution, the non-stable tail means many of the interesting properties such as undefined variance and autocorrelation do not hold.

The clear conclusion to draw from the work into the heavy-tail identification methods is the difficulty in accurately measuring the heavy-tail index for real-world empirical data sets. In these real-world data sets, we would expect to be working with limited set sizes with potentially significant memory. This is a clear problem in this investigation for several reasons. Firstly, the heavy-tail index is required to be able to categorise whether or not the process is (asymptotically) stable. This effects whether or not autocorrelation or coherence

is an appropriate tool for quantifying memory. Secondly, if a process is stable, the means to calculate the coherence from an empirical data requires α - both directly and indirectly in choosing an appropriate structure constant γ. The result is that if we want to test the validity of the coherence function as a means to predict zero-crossing statistics, it is strongly advisable that any initial investigation is based on processes whose heavy-tail index and underlying distribution is known.

4

GENERATING COHERENT STABLE NOISES

4.1 GENERATING A CAUCHY RANDOM VARIABLE

It was shown in section 2.3 that there are methods for generating heavy-tailed coloured noises whose exponent is outside the stable regime. In these cases, the memory inherent in the process is sufficiently characterised by the autocorrelation function. When considering heavy-tailed processes within the stable regime, however, the autocorrelation is undefined. An alternative means of describing the dependency between stable random variables is the coherence, as described in subsection 2.2.2. In order to test the validity of coherence and its ability to predict zero-crossing statistics we therefore introduce a means to generate stable random variables with a non-trivial joint pdf. To do so, we first describe a well-known method to generate a Cauchy random variable through the ratio of two Gaussian random variables [54]. Let $X, Y \sim \mathcal{N}(0,1)$ then $U := X/Y \sim \mathcal{C}(0,1)$ such that its pdf is

$$f_U(u) = \frac{1}{\pi(1+u^2)}.$$

To prove this, we introduce the Change of Variable theorem [55] (in 2 dimensions). Let the random variables $\{X, Y\}$ have joint pdf $f_{X,Y}(x,y)$, and let $\{U, V\}$ be some function of $\{X, Y\}$. Then the joint pdf of $\{U, V\}$ is

$$f_{U,V}(u,v) = |J| f_{X,Y}(x(u,v), y(u,v)) \tag{4.1.1}$$

where J is the Jacobian matrix

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

Let X, Y be independent Gaussian random variables with joint pdf $f_{X,Y}(x,y)$. Define the change of coordinate system from $(x,y) \to (u,v)$ by the functions $u = x/y$ and $v = y$ (and therefore $x(u,v) = uv$ and $y(u,v) = v$). Thus, the random variable U is the result of the quotient of X and Y. Consequently, the determinant of the Jacobian is

$$|J| = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = |v|. \tag{4.1.2}$$

The random variables X, Y are independent and therefore $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$. Using independence with Equation 4.1.1 and Equation 4.1.2 the joint pdf of U, V is

$$f_{U,V}(u,v) = |v| f_X(uv) f_Y(v).$$

Both $X, Y \sim \mathcal{N}(0,1)$ and therefore

$$\begin{aligned} f_{U,V}(u,v) &= |v| \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2v^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-v^2}{2}\right) \\ &= \frac{|v|}{2\pi} \exp\left(\frac{-v^2(1+u^2)}{2}\right). \end{aligned} \tag{4.1.3}$$

The pdf of U can be calculated by integrating the joint pdf in Equation 4.1.3 over v

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v) dv. \tag{4.1.4}$$

By noting that $f_{U,V}(u,v)$ is even about $v = 0$ and and defining $k = (1+u^2)/2$, Equation 4.1.4 becomes

$$\begin{aligned} f_U(u) &= \frac{1}{(-2k\pi)} \int_0^{\infty} (-2kv) \exp\left(-kv^2\right) dv \\ &= \frac{1}{\pi(1+u^2)}. \end{aligned} \tag{4.1.5}$$

Then U has a Cauchy pdf and is a Cauchy random variable. Therefore the ratio of two Gaussian random variables is a Cauchy random variable.

4.2 GENERATING A CAUCHY COHERENT NOISE

4.2.1 *The ratio method*

The methodology behind generating a Cauchy random variable can be extended to consider generating a Cauchy coherent noise. Let $\{X(t)\}$, $\{Y(t)\}$ be continuous-time Gaussian

processes for $t \geq 0$, both with the identical autocorrelation function $a(\tau)$. Let $\{X(t)\}$ and $\{Y(t)\}$ be independent of one another. A new process $\{U(t)\}$ is defined by

$$U(t) := \frac{X(t)}{Y(t)}$$

for $t \geq 0$. This new process $\{U(t)\}$ is considered at times t_1, t_2 such that $\tau = |t_1 - t_2|$ and $a(\tau) = \rho$. The sets of pairs of random variables at these times are denoted $\boldsymbol{X} = \{X(t_1), X(t_2)\}$, $\boldsymbol{Y} = \{Y(t_1), Y(t_2)\}$, $\boldsymbol{U} = \{U(t_1), U(t_2)\}$.

The general bivariate Gaussian distribution of two random variables X_1, X_2 with mean $\mu = 0$, variance $\sigma^2 = 1$ and correlation ρ is

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{x_1^2 + x_2^2 - 2\rho x_1 x_2\right\}\right]. \tag{4.2.1}$$

The two processes $\boldsymbol{X}, \boldsymbol{Y}$ are independent of one another. It follows that the joint pdf of the two processes at t_1, t_2 is

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y}) \tag{4.2.2}$$

where $\boldsymbol{x} = \{x_1, x_2\}$, $\boldsymbol{y} = \{y_1, y_2\}$ and $f_{\boldsymbol{X}}, f_{\boldsymbol{Y}}$ are the joint pdfs of $\boldsymbol{X}, \boldsymbol{Y}$ respectively, with bivariate Gaussian distributions as defined in Equation 4.2.1. Using Equation 4.2.1 and Equation 4.2.2, the joint pdf of $\{\boldsymbol{X}, \boldsymbol{Y}\}$ is

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{4\pi^2(1-\rho^2)} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{x_1^2 + x_2^2 + y_1^2 + y_2^2 - 2\rho(x_1 x_2 + y_1 y_2)\right\}\right]. \tag{4.2.3}$$

A change of coordinate system is introduced $(\boldsymbol{x}, \boldsymbol{y}) \to (\boldsymbol{u}, \boldsymbol{v})$, or, $(x_1, x_2, y_1, y_2) \to (u_1, u_2, v_1, v_2)$ via the equations $u_1 = x_1/y_1$, $u_2 = x_2/y_2$, $v_1 = y_1$, $v_2 = y_2$. The inverse relationships are therefore $x_j(u_1, u_2, v_1, v_2) = u_j v_j$ and $y_j(u_1, u_2, v_1, v_2) = v_j$, for $j = 1, 2$. This represents new random variables $U(t_1)$, $U(t_2)$ being generated through the ratio of two Gaussian random variables at times t_1, t_2, respectively. Then the joint pdf in the $(\boldsymbol{x}, \boldsymbol{y})$ coordinate system, $f_{\boldsymbol{X},\boldsymbol{Y}}$ (in Equation 4.2.3), is transformed to the joint pdf in the $(\boldsymbol{u}, \boldsymbol{v})$ coordinate system such that

$$f_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) = |J| f_{\boldsymbol{X},\boldsymbol{Y}}(u_1 v_1, u_2 v_2, v_1, v_2) \tag{4.2.4}$$

where the determinant of the Jacobian is

$$|J| = |v_1 v_2|.$$

Then the resultant joint pdf from Equation 4.2.4 is

$$f_{\mathbf{U},V}(u_1,u_2,v_1,v_2) = \frac{|v_1 v_2|}{4\pi^2(1-\rho^2)} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{v_1^2\left(1+u_1^2\right) + v_2^2\left(1+u_2^2\right) - 2\rho v_1 v_2(1+u_1u_2)\right\}\right]. \tag{4.2.5}$$

We can calculate the joint pdf of $f_{\mathbf{U}}$ by integrating Equation 4.2.5 over all v_1, v_2, such that

$$f_{\mathbf{U}}(u_1,u_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{U},V}(u_1,u_2,v_1,v_2)dv_1dv_2. \tag{4.2.6}$$

This will give the joint pdf of the random variables $U(t_1), U(t_2)$, that were generated by the ratio of correlated Gaussian random variables. It is important to note that both $U(t_1)$ and $U(t_2)$ are still Cauchy random variables from the proof for the single variable case. A breakdown of the methods to solve Equation 4.2.6 is in Appendix D, with result

$$f_{\mathbf{U}}(u_1,u_2) = \frac{(1-\rho^2)}{\pi^2 A}\left\{1 + \frac{\rho(1+u_1u_2)}{\sqrt{A}}\tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right\} \tag{4.2.7}$$

where

$$A = A(\rho,u_1,u_2) = \left(1+u_1^2\right)\left(1+u_2^2\right) - \rho^2(1+u_1u_2)^2. \tag{4.2.8}$$

The derivation for Equation 4.2.7 makes no assumption on the value of ρ (other than $\rho \in [-1,1]$) or lag τ. It follows that the lag may be evaluated for any $\tau \geq 0$ and Equation 4.2.7 is the joint pdf of $\{U(t)\}$, where the correlation ρ is now equal to the autocorrelation function of the Gaussian processes. Using this pdf, the special case of $\rho = 0 \;\forall\; \tau > 0$ (a set of iid Cauchy random variables produced by the two iid Gaussian processes) can be evaluated. The resultant joint pdf is

$$\begin{aligned} f_{\mathbf{U}}(u_1,u_2) &= \left\{\frac{1}{\pi(1+u_1^2)}\right\}\left\{\frac{1}{\pi(1+u_2^2)}\right\} \\ &= f_{U_1}(u_1)f_{U_2}(u_1) \end{aligned}$$

where U_1, U_2 are clearly Cauchy random variables. This checks the standard result - that the ratio of two Gaussian random variables is a Cauchy random variable - still holds. It is therefore concluded that a novel means of generating a Cauchy process with non-trivial

memory (in terms of an analytical joint pdf shown in Equation 4.2.7) has been produced. We call this the ratio method for generating a Cauchy coherent noise.

To comment on the validity of the ratio method as a viable candidate to investigate the zero-crossing statistics of stable processes, a single realisation is investigated. Whilst the ratio method is derived in terms of continuous-time stochastic processes, it is important to clarify that generating them computationally is done using discrete-time processes. Let $\boldsymbol{X} = \{X_1,\ X_2,\ \ldots,\ X_n\}$, $\boldsymbol{Y} = \{Y_1,\ Y_2,\ \ldots,\ Y_n\}$ be discrete-time Gaussian processes with autocorrelation function $a(\tau)$. Then the Cauchy process produced by the ratio method is

$$\boldsymbol{U} = \left\{ \frac{X_1}{Y_1},\ \frac{X_2}{Y_2},\ \ldots,\ \frac{X_n}{Y_n} \right\}.$$

This discrete-time process $\boldsymbol{U}$ is analagous to the continuous-time Cauchy process $\{U(t)\}$. A single realisation of $\boldsymbol{U}$ is generated with $n = 2^{18}$ elements by the ratio of two Gaussian realisations with identical autocorrelation functions $a(\tau) = \exp(-k_a^2\tau^2)$ where $k_a = \sqrt{\log(2)}/64$. The result is plotted in Figure 4.2.1, with a continuous plot fitted to the individual realisations. It is also plotted alongside the Gaussian realisations $\boldsymbol{X}$ and $\boldsymbol{Y}$ in Figure 4.2.2.

From these plots, it is clear that there is a discrepency between what might be expected from real-world data sets and the realisation produced. Specifically, there are relatively regular zero-crossings where the Cauchy realisation is significantly far away from 0 that it would be considered surprising in a real-world data set. The reason for this can be explained by inspecting Figure 4.2.2. Whilst the regular zero-crossings of $\boldsymbol{U}$ correspond to the zero-crossings of $\boldsymbol{X}$, there are also zero-crossings located at the zero-crossings of $\boldsymbol{Y}$ that cause discontinuities in $\boldsymbol{U}$. We can explain why this occurs by considering the original continuous-time processes. As the denominator process $Y(t) \to 0$ the Cauchy process $|U(t)| \to \infty$. Consequently, the zero-crossings of $\{Y(t)\}$ correspond to poles of $\{U(t)\}$ such that there is a sign change and a discontinuity. This is problematic as the resultant Cauchy process has zero-crossings that correspond to discontinuities whilst we want to investigate processes whose zero-crossings are not the result of such extreme switches. This motivates us to modify the ratio method to alleviate these issues.

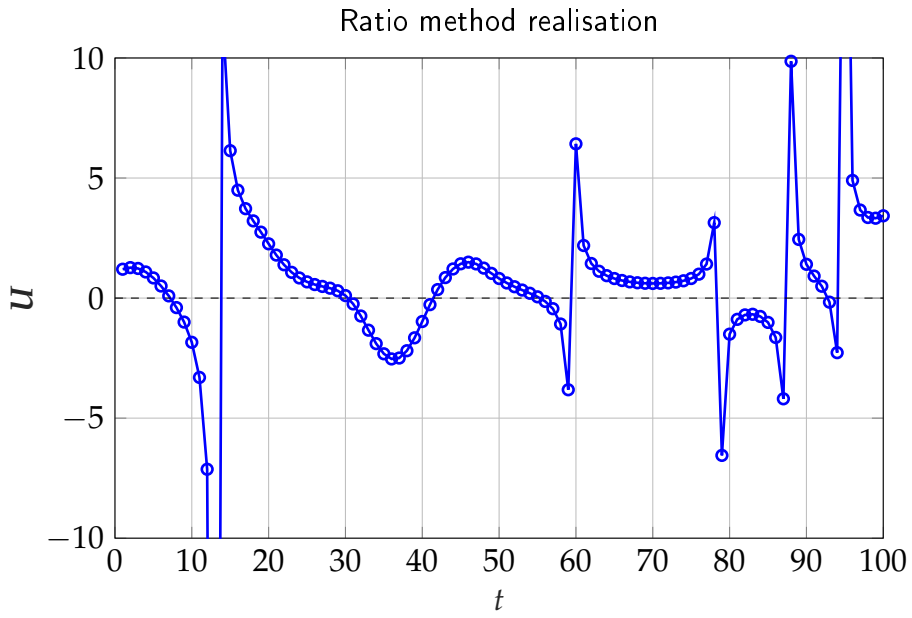

Figure 4.2.1.: Cauchy realisation generated by the ratio method. Although used to generate a discrete-time realisation (circled), a continuous model using interpolation is plotted to further illustrate its behaviour.

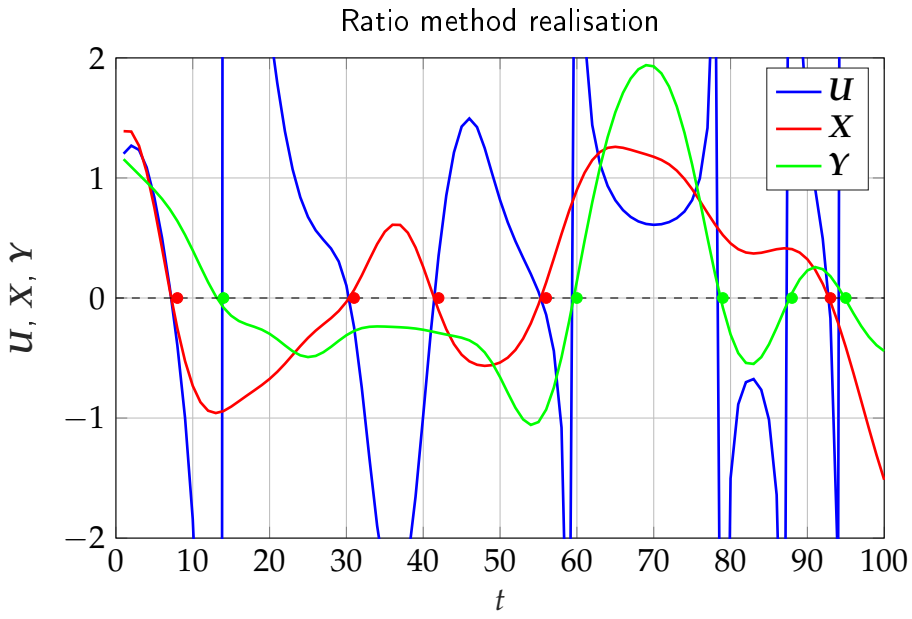

Figure 4.2.2.: Cauchy realisation $\boldsymbol{U}$ and its generating Gaussian realisations $\boldsymbol{X}$, $\boldsymbol{Y}$ using the ratio method. The zero-crossings of all 3 realisations are highlighted, where the crossings of $\boldsymbol{U}$ correspond to the crossings of both $\boldsymbol{X}$ and $\boldsymbol{Y}$.

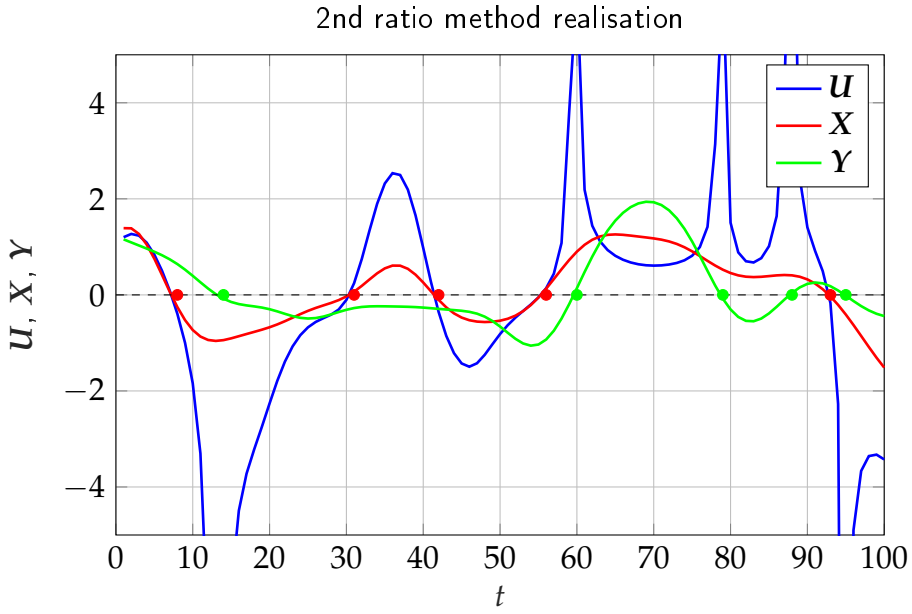

Figure 4.2.3.: Cauchy realisation generated by the 2nd ratio method. Plotted alongside the generating Gaussian realisations $\boldsymbol{X}$, $\boldsymbol{Y}$ where $\boldsymbol{U} = \boldsymbol{X}/|\boldsymbol{Y}|$. The zero-crossings of all 3 realisations are highlighted, where the zero-crossings of $\boldsymbol{U}$ correspond to the zero-crossings of $\boldsymbol{X}$ only.

4.2.2 *The 2nd ratio method*

A modification of the ratio method is to consider the modulus of $Y(t)$, such that $U(t) := X(t)/|Y(t)|$ and its discrete-time equivalent is

$$\boldsymbol{U} = \left\{ \frac{X_1}{|Y_1|}, \frac{X_2}{|Y_2|}, \ldots, \frac{X_n}{|Y_n|} \right\}. \tag{4.2.9}$$

This is designed to ensure the discontinuities in the process do not correspond to zero-crossings. Because X_j, Y_j are symmetric about 0 it follows that $\operatorname{sgn}(X_j) \overset{d}{=} \operatorname{sgn}(X_j)/\operatorname{sgn}(Y_j)$ where $\overset{d}{=}$ denotes equivalence in distribution. Therefore $X_j/|Y_j| = \operatorname{sgn}(X_j)\ |X_j/Y_j| \overset{d}{=} (\operatorname{sgn}(X_j)/\operatorname{sgn}(Y_j))\ |X_j/Y_j| = X_j/Y_j$. Therefore $U(t)$ is still Cauchy by the proof in section 4.1. An example of a realisation is plotted in Figure 4.2.3 alongside the Gaussian realisations that generate it. This alleviates a problem with the original ratio method, as the zero-crossings of $\{Y(t)\}$ no longer correspond to poles of $\{U(t)\}$ that register as zero-crossings, and instead just appear as discontinuities. To calculate the joint density function of the resultant process, we introduce the folded Gaussian distribution. Let $X \sim \mathcal{N}(\mu, \sigma^2)$

and define a new random variable $Y := |X|$. Then Y is a folded Gaussian random variable with pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right) + \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y+\mu)^2}{2\sigma^2}\right) \tag{4.2.10}$$

for $y \geq 0$, and 0 otherwise. We may denote this by $Y \sim \mathcal{N}_f(\mu, \sigma^2)$. An example of Equation 4.2.10 with $\mu = \sigma^2 = 1$ is plotted in Figure 4.2.4. In this investigation, it suffices to consider the simple case $Y \sim \mathcal{N}_f(0,1)$ such that the density function of Y is

$$f_Y(y) = \sqrt{\frac{2}{\pi}} \exp\left(\frac{-y^2}{2}\right) \tag{4.2.11}$$

for $y \geq 0$, and 0 otherwise. The mean and variance of Y are not equal to those of X and, using Equation 4.2.11, can be calculated as $\mu_Y = \sqrt{2/\pi}$ and $\sigma_Y^2 = 1 - 2/\pi$ respectively. The folded Gaussian pdf can be generalised to the bivariate folded Gaussian pdf, as described in [56]. Let $X_1, X_2 \sim \mathcal{N}(0,1)$ be random variables with correlation ρ_1 and define $Y_1 := |X_1|$, $Y_2 := |X_2|$ such that $Y_1, Y_2 \sim \mathcal{N}_f(0,1)$. Then the joint density function of Y_1, Y_2 is the (simple) folded bivariate Gaussian pdf

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\pi\sqrt{1-\rho_1^2}} \Bigg\{ &\exp\left[\frac{-1}{2(1-\rho_1^2)}\left(y_1^2 + y_2^2 - 2\rho_1 y_1 y_2\right)\right] \\ &+ \exp\left[\frac{-1}{2(1-\rho_1^2)}\left(y_1^2 + y_2^2 + 2\rho_1 y_1 y_2\right)\right] \Bigg\} \end{aligned} \tag{4.2.12}$$

for $y_1, y_2 \geq 0$, and 0 otherwise. The marginal pdf's of Equation 4.2.12 are equal to Equation 4.2.11, ensuring Equation 4.2.12 represents a joint pdf of the bivariate folded Gaussian random variables. It is noted that this is the folded form of Equation 4.2.1, and therefore the correlation between Y_1, Y_2 is not equal to ρ_1. Using the joint density in Equation 4.2.12 the correlation between Y_1, Y_2 is

$$\begin{aligned} \rho_Y &= \frac{1}{\sigma_Y^2} \int_0^\infty \int_0^\infty (y_1 - \mu_Y)(y_2 - \mu_Y) f_Y(y_1, y_2) dy_1 dy_2 \\ &= \frac{2}{\pi - 2} \left(\sqrt{1-\rho_1^2} - 1 + \rho_1 \tan^{-1}\left[\frac{\rho_1}{\sqrt{1-\rho_1^2}}\right] \right). \end{aligned} \tag{4.2.13}$$

In the context of this investigation, it suffices to use the relationship in this form, as we have not yet produced an analytical inverse solution of Equation 4.2.13 such that ρ_1 is expressed as a function of ρ_Y.

This information can be used to calculate the joint pdf of the Cauchy process described by Equation 4.2.9. Let $\{X(t)\}$, $\{\tilde{X}(t)\}$ be continuous-time Gaussian processes for $t \geq 0$, both with identical autocorrelation function $a(\tau)$. Let $\{X(t)\}$ and $\{\tilde{X}(t)\}$ be independent of one another. Define a new positive Gaussian process $\{Y(t)\}$ by $Y(t) := |\tilde{X}(t)|$. A new process $\{U(t)\}$ is defined by $U(t) := X(t)/Y(t)$ for $t \geq 0$. We consider the processes at times t_1, t_2, and denote the pairs of random variables $\boldsymbol{X} = \{X(t_1), X(t_2)\}$, $\boldsymbol{Y} = \{Y(t_1), Y(t_2)\}$, $\boldsymbol{U} = \{U(t_1), U(t_2)\}$. As $\boldsymbol{X}$, $\boldsymbol{Y}$ are independent of one another, their joint pdf is

$$g_{\boldsymbol{X},\boldsymbol{Y}}(x_1, x_2, y_1, y_2) = g_{\boldsymbol{X}}(x_1, x_2) g_{\boldsymbol{Y}}(y_1, y_2)$$

where $g_{\boldsymbol{X}}$ is the bivarate Gaussian density function in Equation 4.2.1, and $g_{\boldsymbol{Y}}$ is the bivariate positive Gaussian density function in Equation 4.2.12. For convenience we denote $\rho_X = \rho$. Therefore the joint pdf is

$$\begin{aligned} g_{\boldsymbol{X},\boldsymbol{Y}}(x_1, x_2, y_1, y_2) = & \frac{1}{2\pi^2 (1-\rho^2)} \left\{ \exp\left[\frac{-1}{2(1-\rho^2)} \left(x_1^2 + x_2^2 - 2\rho x_1 x_2 \right.\right.\right. \\ & \left. + y_1^2 + y_2^2 - 2\rho y_1 y_2 \right) \Big] + \exp\left[\frac{-1}{2(1-\rho^2)} \left(x_1^2 + x_2^2 \right.\right. \\ & \left.\left.\left. - 2\rho x_1 x_2 + y_1^2 + y_2^2 + 2\rho y_1 y_2 \right) \right] \right\} \end{aligned} \tag{4.2.14}$$

for $x_1, x_2 \in \mathbb{R}$, $y_1, y_2 \geq 0$, and 0 otherwise. We introduce a change of variables via $u_j = x_j / y_j$, $v_j = y_j$, such that $|J| = |v_1 v_2|$. Then the joint pdf in the new coordinate system is

$$\begin{aligned} g_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) = & \frac{|v_1 v_2|}{2\pi^2 (1-\rho^2)} \left\{ \exp\left[\frac{-1}{2(1-\rho^2)} \left(u_1^2 v_1^2 + u_2^2 v_1^2 - 2\rho u_1 u_2 v_1 v_2 \right.\right.\right. \\ & \left. + v_1^2 + v_2^2 - 2\rho v_1 v_2 \right) \Big] + \exp\left[\frac{-1}{2(1-\rho^2)} \left(u_1^2 v_1^2 + u_2^2 v_1^2 \right.\right. \\ & \left.\left.\left. - 2\rho u_1 u_2 v_1 v_2 + v_1^2 + v_2^2 + 2\rho v_1 v_2 \right) \right] \right\} \end{aligned} \tag{4.2.15}$$

for $u_1, u_2 \in \mathbb{R}$, $v_1, v_2 \geq 0$, and 0 otherwise. Then by integrating Equation 4.2.15 over v the joint pdf of $\boldsymbol{U}$ is

$$\begin{aligned} g_{\boldsymbol{U}}(u_1, u_2) = & \frac{1-\rho^2}{2\pi^2} \left\{ \frac{\rho(1+u_1 u_2)}{A^{3/2}} \left(\frac{\pi}{2} + \frac{\sqrt{A}}{\rho(1+u_1 u_2)} + \tan^{-1}\left[\frac{\rho(1+u_1 u_2)}{\sqrt{A}} \right] \right) \right. \\ & \left. - \frac{\rho(1-u_1 u_2)}{B^{3/2}} \left(\frac{\pi}{2} - \frac{\sqrt{B}}{\rho(1-u_1 u_2)} - \tan^{-1}\left[\frac{\rho(1-u_1 u_2)}{\sqrt{B}} \right] \right) \right\} \end{aligned} \tag{4.2.16}$$

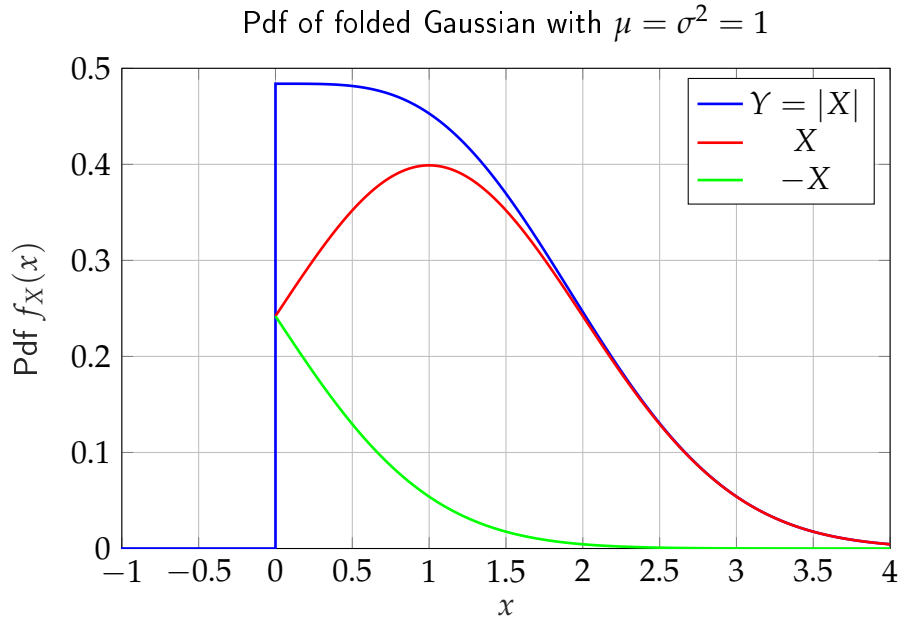

Figure 4.2.4.: The folded Gaussian pdf for a random variable $Y = |X|$ where $X \sim \mathcal{N}(1,1)$. Plotted alongside the pdfs for X and $-X$ (for $x \geq 0$ only). Clearly $f_Y = f_X + f_{-X}$.

where

$$A = A(\rho, u_1, u_2) = (1+u_1^2)(1+u_2^2) - \rho^2(1+u_1u_2)^2,$$
$$B = B(\rho, u_1, u_2) = (1+u_1^2)(1+u_2^2) - \rho^2(1-u_1u_2)^2$$

for $u_1, u_2 \in \mathbb{R}$. Then $g_{\boldsymbol{U}}$ in Equation 4.2.16 is the joint pdf of $\boldsymbol{U}$. We call this the 2nd ratio method for producing a Cauchy coherent noise.

4.2.3 *The 3rd ratio method*

For the 2nd ratio method described in subsection 4.2.2 the denominator process $\{Y(t)\}$ is produced by taking the element-wise modulus of a Gaussian coloured noise. Aside from the benefit of removing the zero-crossings corresponding to discontinuities, this method is useful as the joint density function of $\{Y(t)\}$ is then the folded Gaussian density function in Equation 4.2.12. Therefore the joint density function of the resulting Cauchy process $\{U(t)\}$ may be calculated. The empirical properties of the resulting realisations are not necessarily ideal, however, because the denominator process is not differentiable

as $Y(t) \to 0$. This results in poles that, whilst not corresponding to zero-crossings, are nevertheless present. It follows that realisations produced by the 2nd ratio method are not everywhere differentiable as might be expected for models of physical systems.

We therefore introduced a further adaptation which we dub the 3rd ratio method for generating Cauchy coherent noise. We first recall the iterative Fourier Filtering Method (iFFM) from subsection 2.3.2. Let $\boldsymbol{Y}_a$ be a realisation of a Gaussian coloured noise with autocorrelation function $a(\tau)$, and let $\boldsymbol{Y}_d$ be a realisation of a set of iid random variables with distribution d. Then recall the iFFM produces a new realisation $\boldsymbol{Z}$ with distribution d and autocorrelation $a(\tau)$. For convenience, we write this as $\boldsymbol{Z} = \text{iFFM}(\boldsymbol{Y}_a, \boldsymbol{Y}_d)$.

Let $\{X(t)\}$ be a Gaussian coloured noise with autocorrelation function $a(\tau)$, and let $\{Z(t)\}$ be a positive Gaussian coloured noise also with autocorrelation function $a(\tau)$ such that both $X(t)$ and $Z(t)$ are differentiable $\forall\ t \geq 0$. Then the 3rd ratio method defines a new Cauchy process $\{U(t)\}$ by $U(t) := X(t)/Z(t)$. To generate a discrete-time realisation representitive of this, let $\boldsymbol{X}$ be a discrete-time realisation of $\{X(t)\}$ as described previously. Let $\boldsymbol{Y}_a$ be a realisation of a Gaussian coloured noise with autocorrelation $a(\tau)$, $\boldsymbol{Y}_d$ be a set of realisations of a set of iid positive Gaussian random variables, and $\boldsymbol{Z} = \text{iFFM}(\boldsymbol{Y}_a, \boldsymbol{Y}_d)$ such that $\boldsymbol{Z}$ is the discrete-time analogy to $\{Z(t)\}$. Then

$$\boldsymbol{U} = \left\{ \frac{X_1}{Z_1},\ \ldots,\ \frac{X_n}{Z_n} \right\}$$

is the discrete-time realisation of the 3rd ratio method.

Whist the joint density function of the denominator process in the 2nd ratio method is known, the iFFM only enforces a certain marginal density function. It is therefore not certain the form of the joint density function of the denominator process for the 3rd ratio method. We therefore use the folded Gaussian distribution in Equation 4.2.12 to construct a model with the correct autocorrelation and marginal distribution. Let $\boldsymbol{X} = \{X(t_1), X(t_2)\}$ be a pair of correlated Gaussian random variables and $\boldsymbol{Y} = \{Y(t_1), Y(t_2)\}$ be a pair of correlated positive Gaussian random variables. Let $\boldsymbol{X}$, $\boldsymbol{Y}$ have matching correlation ρ. As $\boldsymbol{X}$, $\boldsymbol{Y}$ are independent of one another, their joint density function satisfies

$$h_{\boldsymbol{X},\boldsymbol{Y}}(x_1, x_2, y_1, y_2) = h_{\boldsymbol{X}}(x_1, x_2) h_{\boldsymbol{Y}}(y_1, y_2).$$

where $h_{\boldsymbol{X}}$ is equal to Equation 4.2.1 and $h_{\boldsymbol{Y}}$ is equal to Equation 4.2.12 with correlation coefficient ρ_1 satisfying Equation 4.2.13. As we have set $\boldsymbol{X}$ and $\boldsymbol{Y}$ to have identical autocorrelation functions, it follows that $\rho_X = \rho_Y = \rho$. Therefore

$$h_{\boldsymbol{X},\boldsymbol{Y}}(x_1, x_2, y_1, y_2) = \frac{1}{2\pi^2\sqrt{(1-\rho_1^2)(1-\rho^2)}} \exp\left[\frac{-1}{2(1-\rho^2)}\right.$$
$$\left.\left(x_1^2 + x_2^2 - 2\rho x_1 x_2\right)\right]\left\{\exp\left[\frac{-1}{2(1-\rho_1^2)}\left(y_1^2 + y_2^2 - 2\rho_1 y_1 y_2\right)\right]\right.$$
$$\left. + \exp\left[\frac{-1}{2(1-\rho_1^2)}\left(y_1^2 + y_2^2 + 2\rho_1 y_1 y_2\right)\right]\right\}. \tag{4.2.17}$$

for $x_1, x_2 \in \mathbb{R}$, $y_1, y_2 > 0$, and 0 otherwise. As we have not been able to produce an inverse of Equation 4.2.13, we are not able to simplify Equation 4.2.17 to a function of a single correlation parameter. We again introduce a change of coordinates using $u_j = x_j/y_j$, $v_j = y_j$. Then the joint pdf in the new coordinate system is

$$h_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) = \frac{|v_1 v_2|}{2\pi^2\sqrt{(1-\rho_1^2)(1-\rho^2)}} \exp\left[\frac{-1}{2(1-\rho^2)}\left(u_1^2 v_1^2\right.\right.$$
$$\left.\left. + u_2^2 v_2^2 - 2\rho u_1 u_2 v_1 v_2\right)\right]\left\{\exp\left[\frac{-1}{2(1-\rho_1^2)}\left(v_1^2 + v_2^2 - 2\rho_1 v_1 v_2\right)\right]\right.$$
$$\left. + \exp\left[\frac{-1}{2(1-\rho_1^2)}\left(v_1^2 + v_2^2 + 2\rho_1 v_1 v_2\right)\right]\right\} \tag{4.2.18}$$

for $u_1, u_2 \in \mathbb{R}$, $v_1, v_2 > 0$, and 0 otherwise. Then by integrating Equation 4.2.18 over v, the joint pdf of $\boldsymbol{U}$ is

$$\begin{aligned} h_{\boldsymbol{U}}(u_1, u_2) &= \int_0^\infty \int_0^\infty h_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) dv_1 dv_2 \\ &= \frac{m\sqrt{(1-\rho^2)(1-\rho_1^2)}}{2\pi^2} \\ &\quad \left\{\frac{\rho_1 + m\rho u_1 u_2}{C^{3/2}}\left(\frac{\pi}{2} + \frac{\sqrt{C}}{\rho_1 + m\rho u_1 u_2} + \tan^{-1}\left[\frac{\rho_1 + m\rho u_1 u_2}{\sqrt{C}}\right]\right)\right. \\ &\quad \left. - \frac{(\rho_1 - m\rho u_1 u_2)}{D^{3/2}}\left(\frac{\pi}{2} - \frac{\sqrt{D}}{\rho_1 - m\rho u_1 u_2} - \tan^{-1}\left[\frac{\rho_1 - m\rho u_1 u_2}{\sqrt{D}}\right]\right)\right\} \end{aligned} \tag{4.2.19}$$

for $u_1, u_2 \in \mathbb{R}$, where

$$C = C(\rho, \rho_1, u_1, u_2) = (1 + mu_1^2)(1 + mu_2^2) - (\rho_1 + m\rho u_1 u_2)^2,$$
$$D = D(\rho, \rho_1, u_1, u_2) = (1 + mu_1^2)(1 + mu_2^2) - (\rho_1 - m\rho u_1 u_2)^2$$

and

$$m = m(\rho_1, \rho) = \frac{1-\rho_1^2}{1-\rho^2}.$$

Then Equation 4.2.19 is a model for the joint density function of the 3rd ratio method for generating a Cauchy coherent noise. The steps for calculating $f_{\boldsymbol{U}}$, $g_{\boldsymbol{U}}$ and $h_{\boldsymbol{U}}$ are discussed in more detail in Appendix D.

4.2.4 *Comparison of the ratio methods*

We have proposed three ratio methods for generating a Cauchy coherent noise. It is useful for us to summarise all three methods, before comparing them. This will allow us to decide the best candidate to investigate the zero-crossing statistics of stable coherent noises. The original (or 1st) ratio method is introduced in subsection 4.2.1, the 2nd ratio method in subsection 4.2.2, and the 3rd ratio method in subsection 4.2.3.

Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ be realisations of Gaussian coloured noises with the identical autocorrelation function $a(\tau)$. Then

$$\boldsymbol{U}_1 = \left\{ \frac{X_1}{Y_1}, \ldots, \frac{X_n}{Y_n} \right\}$$

is a realisation of a Cauchy coherent noise produced via the 1st ratio method, with the resulting joint density function $f_{\boldsymbol{U}}$ in Equation 4.2.7.

Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ be realisations of Gaussian coloured noises with the identical autocorrelation function $a(\tau)$. Then

$$\boldsymbol{U}_2 = \left\{ \frac{X_1}{|Y_1|}, \ldots, \frac{X_n}{|Y_n|} \right\}$$

is a realisation of a Cauchy coherent noise produced via the 2nd ratio method, with the resulting joint density function $g_{\boldsymbol{U}}$ in Equation 4.2.16.

Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ be realisations of Gaussian coloured noises with the identical autocorrelation function $a(\tau)$, and let $\boldsymbol{Y}_d = \{Y_{d,1}, \ldots, Y_{d,n}\}$ be a realisaton of iid random variables with a positive Gaussian distribution. Using the

iFFM described in subsection 2.3.2 we can produce a realisation $\boldsymbol{Z} = \text{iFFM}(\boldsymbol{Y}, \boldsymbol{Y}_d) = \{Z_1, \ldots, Z_n\}$ of a coloured noise with a positive Gaussian distribution and autocorrelation $a(\tau)$, as well as being differentiable everywhere. Then

$$\boldsymbol{U}_3 = \left\{ \frac{X_1}{Z_1}, \ldots, \frac{X_n}{Z_n} \right\}$$

is a realisation of a Cauchy coherent noise produced via the 3rd ratio method, with a model joint density function $h_{\boldsymbol{U}}$ in Equation 4.2.19.

The first means of comparing the three ratio methods is by inspecting some example realisations directly. Using each method, three realisations are produced with $n = 2^{18}$ elements. For each method, the numerator process $\boldsymbol{X}$ has the autocorrelation function $a(\tau) = \exp(-k_a^2 \tau^2)$ where $k_a = \sqrt{\log(2)}/64$. The results of the first 1000 realisations are plotted in Figure 4.2.5. The plots show that the resulting realisations display the characteristics we expect from each one. Specifically the 1st method has discontinuities that result in the signal changing sign (and therefore registering as a zero-crossing), the 2nd method has discontinuities that do not change sign and the 3rd method is continuous everywhere. This highlights the advantage of the 3rd method as a suitable model for physical systems.

The other aspect of the ratio methods to examine is how well the analytical joint density functions predict their empirical behaviour. This can be done by comparing the sample structure functions calculated directly from the realisations to the computed structure functions predicted by the analytical joint density functions. Let $\boldsymbol{U} = \{U_j\}_{j=1}^n$ be the realisation produced by the ratio method. Then the sample structure function is calculated by

$$S_s(\tau) = \frac{1}{n-\tau} \sum_{j=1}^{n-\tau} \left| U_{j+\tau} - U_j \right|^\gamma$$

for $\tau = 0, 1, \ldots, n-1$, where γ is the structure constant such that $-1 < \gamma < 1$. Similarly, let $f_{\boldsymbol{U}}(\rho; u_1, u_2)$ be the joint density function of $\boldsymbol{U}$, where ρ is the correlation parameter in the pdf's. Then the computed structure function is

$$S_c(\tau) = \langle |U_{t+\tau} - U_t|^\gamma \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |u_1 - u_2|^\gamma \, f_{\boldsymbol{U}}(\rho; u_1, u_2) \, du_1 \, du_2$$

for $\tau = 0,\ 1,\ \ldots,\ n-1$, where $-1 < \gamma < 1$, and $\rho = a(\tau) = \exp(-k_a^2\tau^2)$ is the correlation parameter at lag τ. It is worth noting that for the 3rd ratio method, the joint pdf has two correlation parameters, ρ and ρ_1, and the latter is calculated from the former via Equation 4.2.13. For each method, both the sample structure function and computed structure function are calculated with a structure constant of $\gamma = 0.5$. The results are plotted in Figure 4.2.6. It is clear from these plots that the joint density functions describe the underlying behaviour of the Cauchy coherent noises generated via the 1st and 2nd ratio methods.

The joint density function is less successful, however, for the 3rd ratio method. This is a result of the iFFM perturbing the joint density function of the denominator process $\{Y(t)\}$. When we use the iFFM to enforce the desired properties onto the realisations (namely continuity and differentiability), the joint density function of $\{Y(t)\}$ changes such that it is not equal to Equation 4.2.12. This means that, whilst $\{Y(t)\}$ and $\{U(t)\}$ have known marginal density functions, we do not know their precise joint density functions. Equation 4.2.19 is therefore a (relatively accurate) model for the underlying behaviour, but is not precise.

From these results, we have seen that the different methods offer various advantages and disadvantages. Most notably, there is a trade-off between desirable empirical properties and knowledge of the joint density function. The 1st ratio method produces a process with a simplistic joint pdf and realisations with discontinuities that register as zero-crossings. The 2nd ratio method produces a process with a more complicated joint pdf (but one that is still correct), and realisations with discontinuities that do not affect the zero-crossings. The 3rd ratio method produces a process with a complicated joint pdf that can be used to model the underlying behaviour but is not precise, and realisations that are smooth and continuous everywhere. For the subsequent work into Cauchy coherent noises, we concern ourselves with the processes that most accurately model physical systems. It is therefore concluded that the 3rd ratio method offers us the best means of generating realisations of Cauchy coherent noises for these purposes.

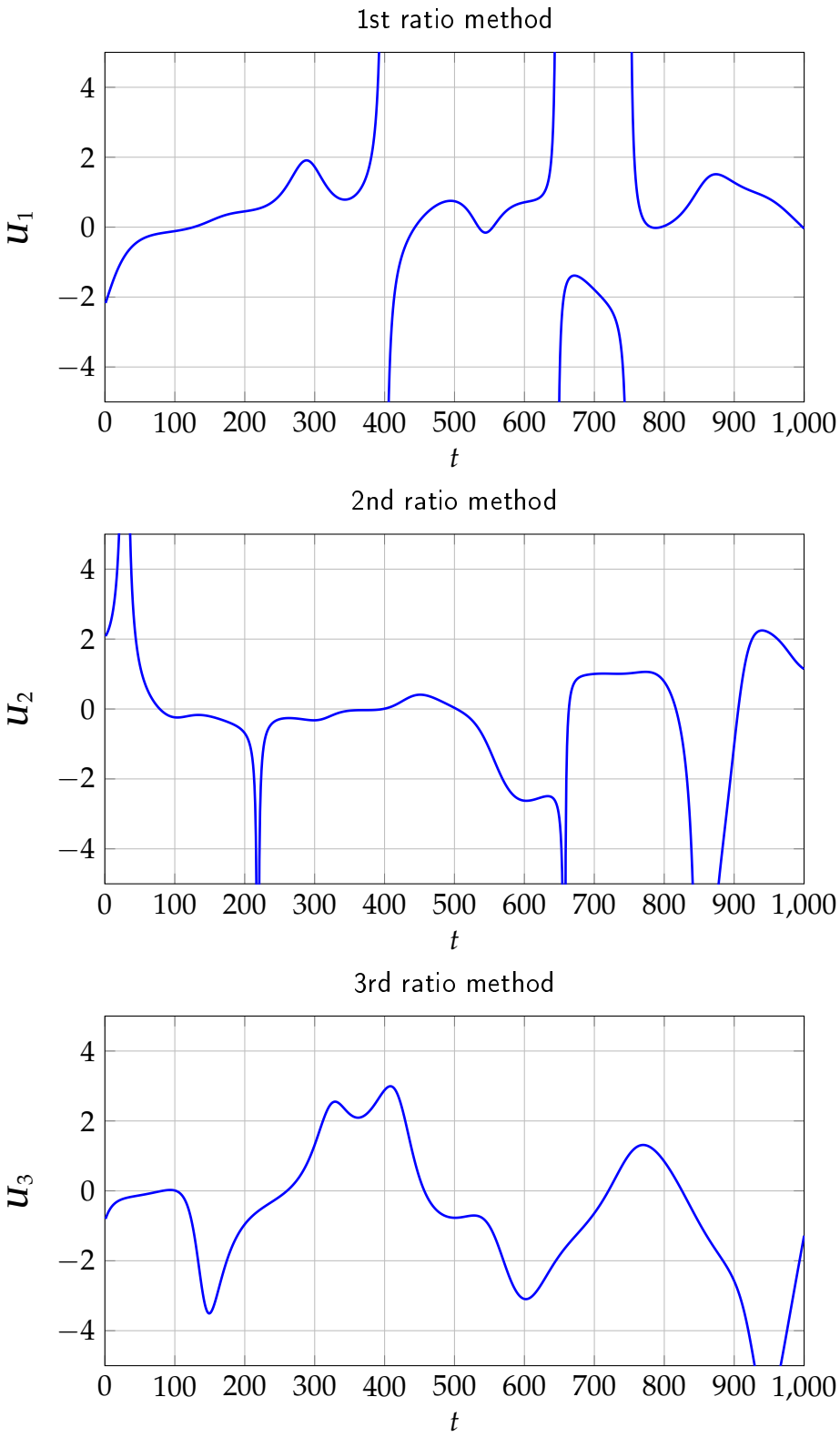

Figure 4.2.5.: Realisations of Cauchy coherent noises using the three ratio methods.

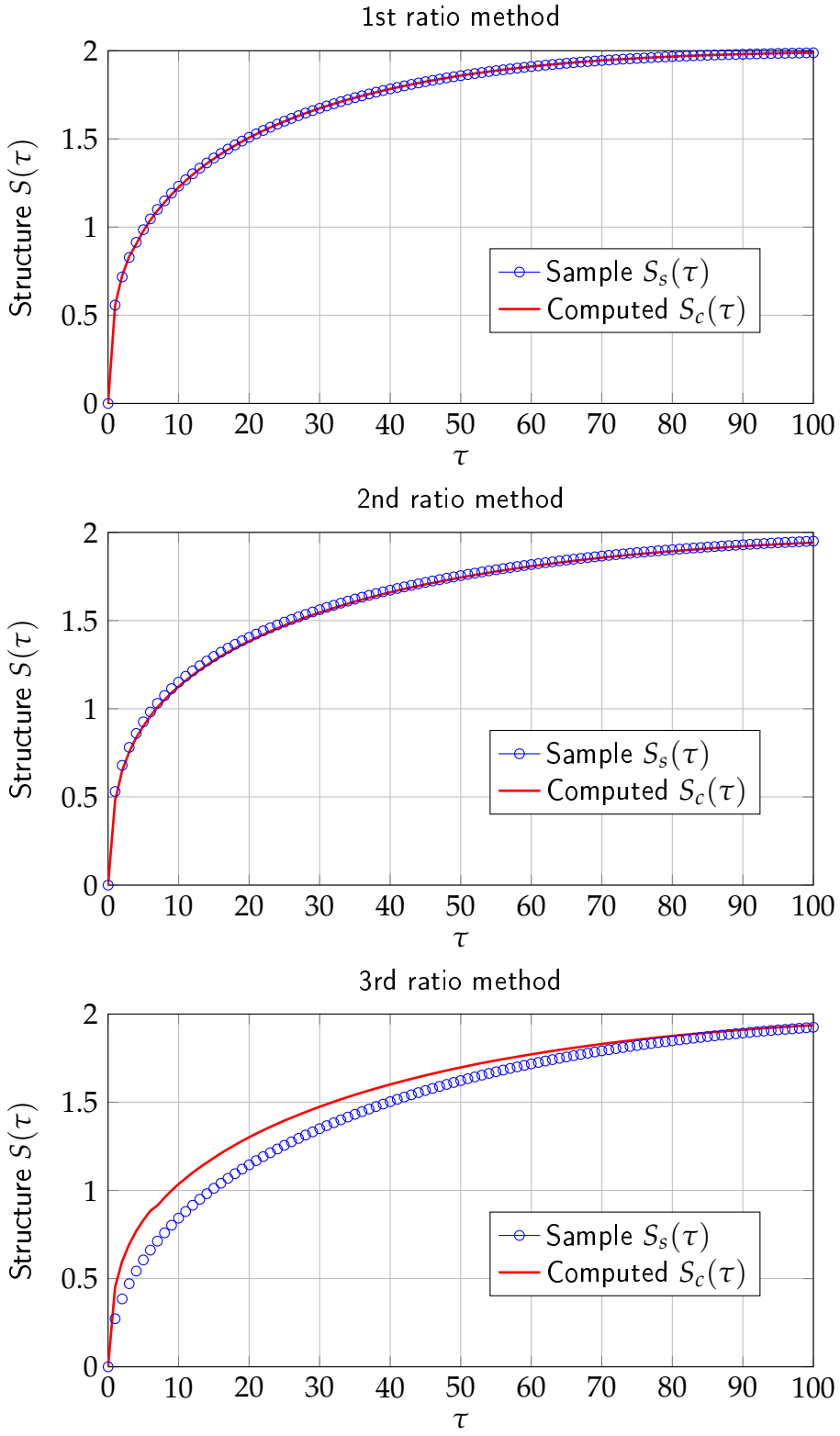

Figure 4.2.6.: Comparison of sample structure functions with those predicted computationally by the analytic joint pdf for the ratio methods.

4.3 EMPIRICAL PROPERTIES OF A CAUCHY COHERENT NOISE

It was concluded in subsection 4.2.4 that the best means of generating a Cauchy coherent noise is by the 3rd ratio method (henceforth referred to simply as the ratio method for convenience). It is of interest to briefly investigate the properties of the resultant realisations. Specifically, we look at its sample probability distribution (both centrally and in the tail), its sample autocorrelation function and its sample coherence function.

A realisation of a Cauchy process is produced by the ratio method such that the generating Gaussian process $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ and positive Gaussian process $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ have identical autocorrelation function $a(\tau) = \exp(-k_a^2\tau^2)$, where $k_a = \sqrt{\log(2)}/\tau_{0.5}$. We investigate a single realisation with $\tau_{0.5} = 64$ and $n = 2^{24}$ elements and denote it $\boldsymbol{U} = \{U_1, \ldots, U_n\}$. The generating realisations of $\boldsymbol{X}$, $\boldsymbol{Y}$ are plotted alongside the resultant Cauchy realisation $\boldsymbol{U}$ in Figure 4.3.1. As discussed in subsection 4.2.4, $\boldsymbol{U}$ is sufficiently smooth and displays large excursions away from 0 as expected.

We can ensure that $\boldsymbol{U}$ is indeed Cauchy by plotting its probability distribution. Its sample pdf is plotted alongside the theoretical Cauchy pdf in Figure 4.3.2. It is clear from this plot that the process is Cauchy for central (non-tail) values. In order to determine whether this is also true in the tail, the sample complementary cdf is plotted alongside the theoretical Cauchy complementary cdf on a logarithmic scale in Figure 4.3.3. From this plot it is similarly concluded that the resultant process is Cauchy in the tail too.

The sample autocorrelations of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{U}$ are plotted in Figure 4.3.4. From this plot it is clear that the sample autocorrelation of the resultant Cauchy process is approximately δ-correlated, especially in comparison to the autocorrelations of the input Gaussian processes.

We can now see if the memory inherent in $\boldsymbol{U}$ can be characterised by the coherence function, as introduced in subsection 2.2.2. The use of the sample structure function is important as it is theorised by [19] that it can be used to extract the coherence function, through the

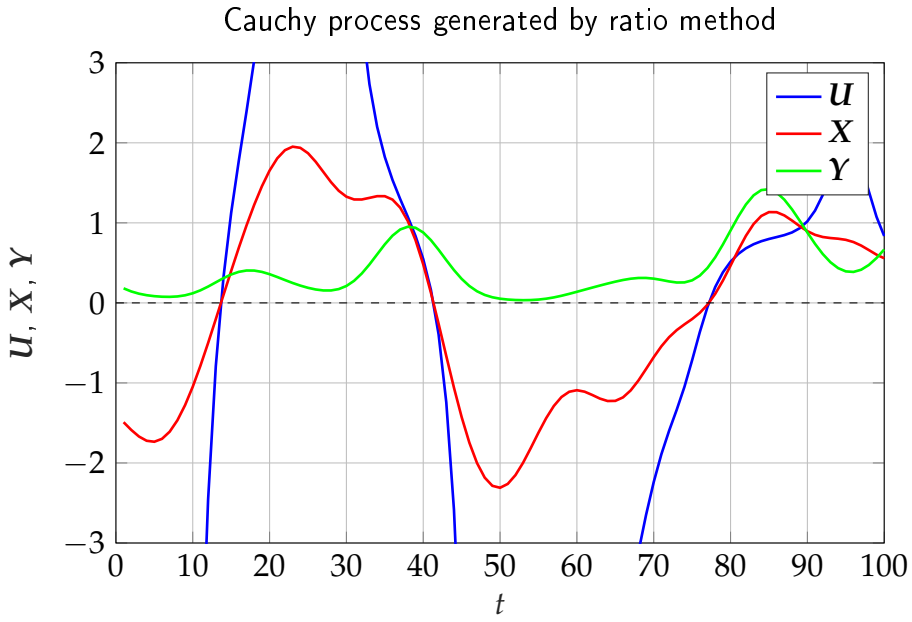

Figure 4.3.1.: Cauchy realisation $\boldsymbol{U}$ generated by the ratio method. Plotted alongside the generating Gaussian realisation $\boldsymbol{X}$ and positive Gaussian realisation $\boldsymbol{Y}$ where $\boldsymbol{U} = \boldsymbol{X}/\boldsymbol{Y}$.

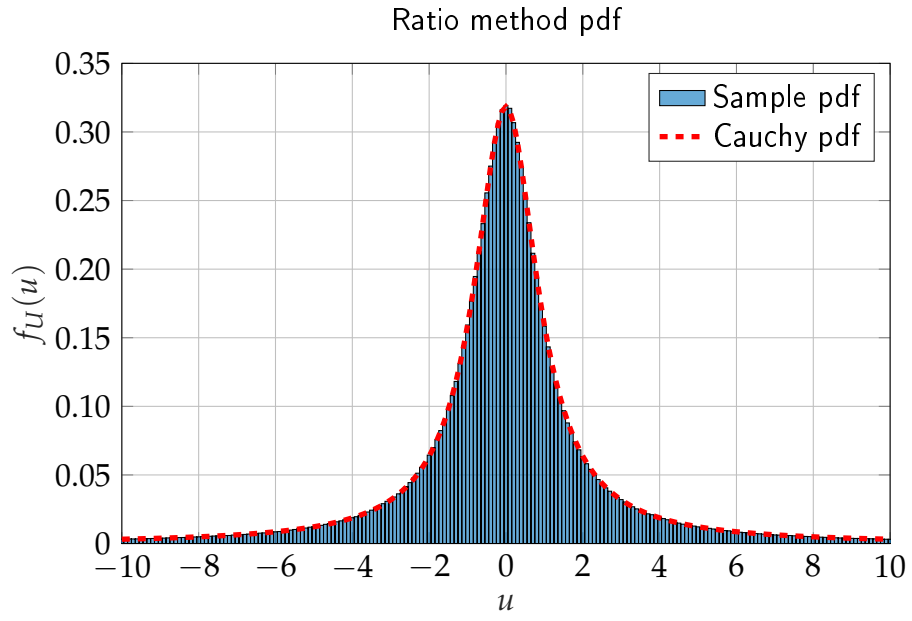

Figure 4.3.2.: Sample pdf of process $\boldsymbol{U}$, where $\boldsymbol{U}$ is generated by the ratio method. Shows that the process is Cauchy for central values.

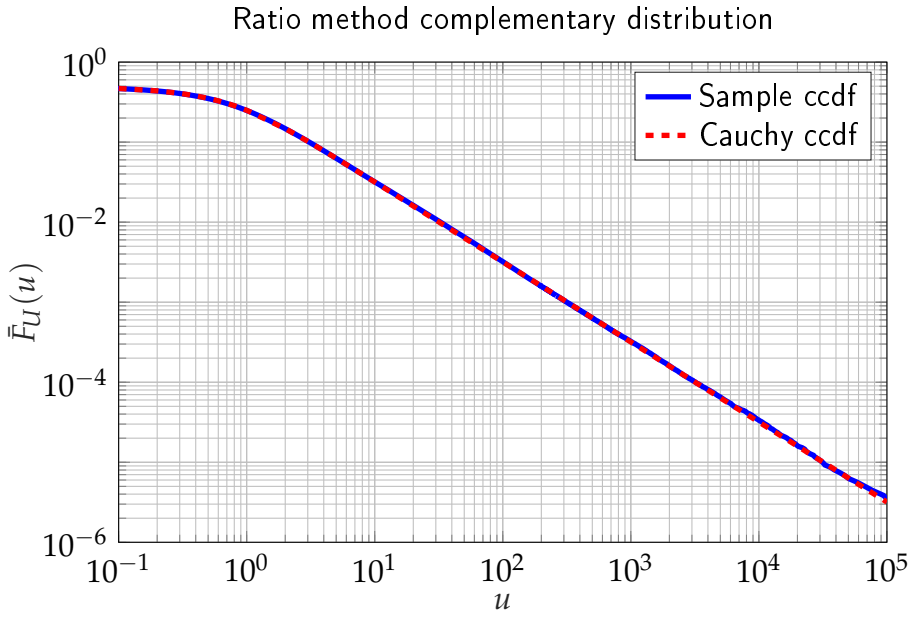

Figure 4.3.3.: Sample complementary cdf of process $\boldsymbol{U}$, where $\boldsymbol{U}$ is generated by the ratio method. Plotted on a logarithmic scale, displaying the process is Cauchy in the tail.

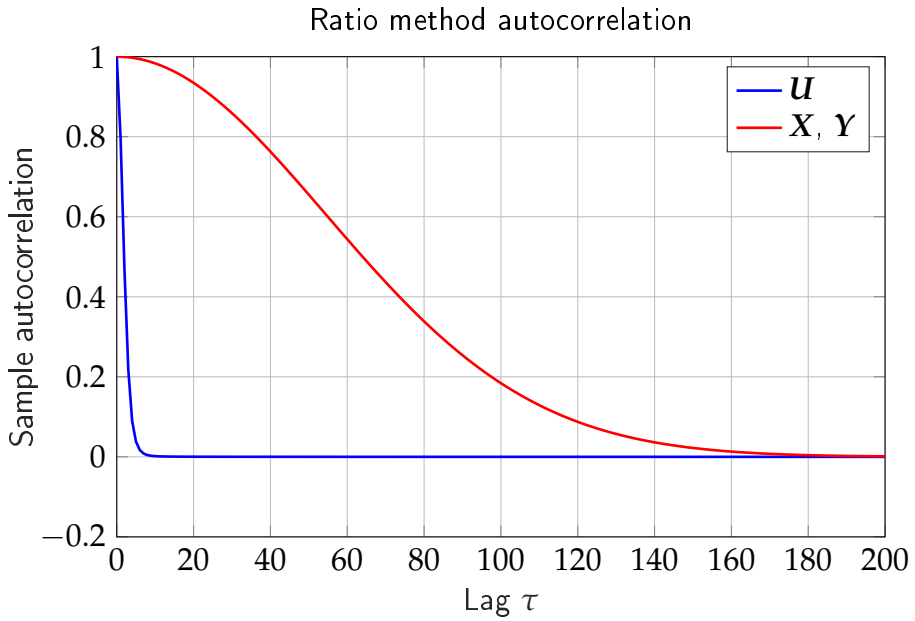

Figure 4.3.4.: Sample autocorrelation function of process $\boldsymbol{U}$, where $\boldsymbol{U}$ is generated by the ratio method. As expected, $\boldsymbol{U}$ is approximately δ-correlated, especially compared to the input autocorrelation of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

existence of fractional moments. Recall from Equation 2.2.5 that the structure function for $\boldsymbol{U}$ is

$$S_\gamma(\tau) = \langle |U_{t+\tau} - U_t|^\gamma \rangle \tag{4.3.1}$$

where $-1 < \gamma < \alpha$. We henceforth make explicit the difference between the sample coherence function and underlying coherence function (or just coherence function). The (underlying) coherence r is a function of time lag τ only, and appears in the joint characteristic function for the stable process according to

$$\phi(r; t_1, t_2) = \exp\left\{ \frac{-A}{1+r^\alpha} \left(|t_1 + rt_2|^\alpha + |t_2 + rt_1|^\alpha \right) \right\} \tag{4.3.2}$$

where A is some scale constant and α is the stability index. Conversely, the sample coherence function r_γ is a function of both the structure constant γ and time lag τ and is defined in terms of the structure function, such that it satisfies

$$\frac{S_\gamma}{\langle |U_t|^\gamma \rangle} = \left[\frac{2(1-r_\gamma)^\alpha}{1+r_\gamma^\alpha} \right]^{\gamma/\alpha} \tag{4.3.3}$$

where $-1 < \gamma < \alpha$. It is clear from the definition of the sample coherence in Equation 4.3.3 the need to distinguish it from the underlying coherence. As an example, if $\gamma = 0$, then Equation 4.3.3 can be satisfied by any coherence r_γ. It is posited in [19] that the coherence r satisfies Equation 4.3.3. To test this, Equation 4.3.3 can be inverted for the Cauchy case where $\alpha = 1$, such that

$$r_\gamma = \frac{2 - s_\gamma^{1/\gamma}}{2 + s_\gamma^{1/\gamma}} \tag{4.3.4}$$

where

$$s_\gamma = \frac{S_\gamma}{\langle |U_t|^\gamma \rangle} \ .$$

The sample coherence r_γ is plotted for different choices of γ in Figure 4.3.5. This is done for $\gamma = -0.5,\ 0.01,\ 0.5,\ 0.9$. From this plot, it is clear that the the sample coherence quantifies the memory in the process. As $\gamma \to -1$, the resultant coherence becomes more pronounced, whilst as $\gamma \to 1$ the coherence resembles a δ-correlation. The output sample coherence functions are relatively close, but there are non-negligible differences. It is important to emphasise that the range of r_γ outputs does not mean there is a range of underlying coherences r that satisfy the joint characteristic function. It is not yet clear which of the sample coherences correctly identify and model the underlying coherence in

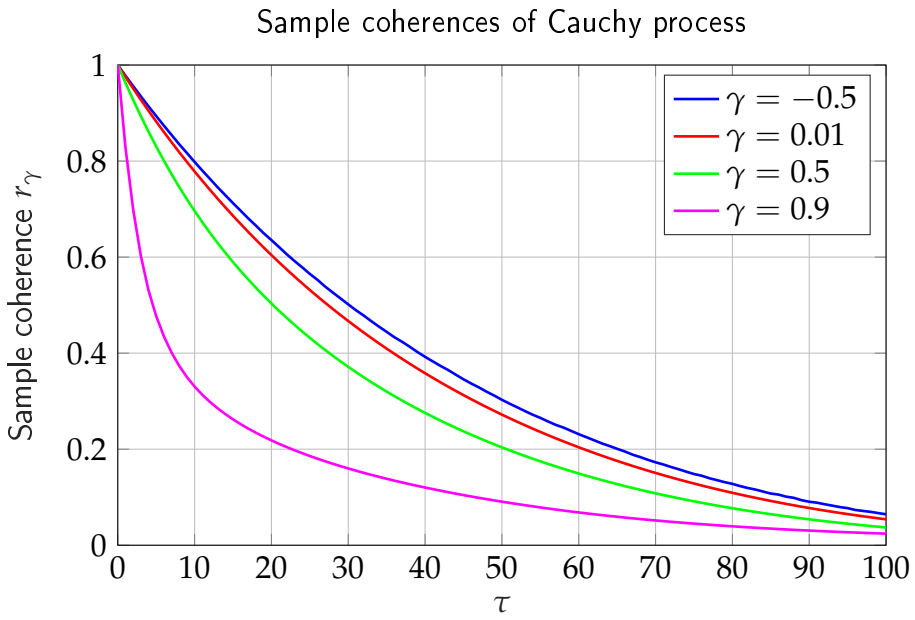

Figure 4.3.5.: Sample coherences r_γ plotted for $\gamma = -0.5,\ 0.01,\ 0.5,\ 0.9$ for a correlated Cauchy process.

Equation 4.3.2. For now, it suffices to accept that the sample coherence can describe the memory of a Cauchy coherent noise generated by the ratio method. For convenience, we henceforth choose to investigate the coherence function with the structure constant $\gamma = 0.5$, which is sufficently far from the limiting cases $|\gamma| \to 1$ and the trivial case $\gamma = 0$.

4.3.1 *An analytical model for the coherence function*

It has been shown that the memory inherent in a Cauchy coherent noise can be quantified by the sample coherence function. In this section we show that the sample coherence can be modelled with an exponential function for $\gamma = 0.5$. This serves to show how we might model coherence functions in general. The sample coherence function of the Cauchy coherent noise $\boldsymbol{U}$ is plotted in Figure 4.3.6 on both a linear and logarithmic scale. By observing that the logarithmic plot is approximately linear, it is proposed the coherence can be modelled by an exponential function. An exponential function is fitted by assuming a model of the form $r(\tau) = \exp(-k_r\tau)$ such that $r(0) = 1$ as required. The scaling

constant k_r is calculated such that the exponential model is equal to the sample coherence when $r(\tau) = 0.1$ (a point within the observed exponential range). For the initial case, the resulting scale constant is calculated as $k_r = 0.0317$. The exponential fits are also plotted alongside the sample coherence in Figure 4.3.6. It is concluded from this that the exponential fit is a satisfactory model. The discrepencies occur at large enough τ (and small enough r) that they are not concerning.

By adopting the exponential function as a model for the coherence, the scale constant of the coherence can be related to the scale constant of the autocorrelation function of the input Gaussian processes. This is done by equating the lag τ. We first remind ourselves that the autocorrelation function of the input Gaussian processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ is $a(\tau) = \exp(-k_a^2\tau^2)$. It follows that

$$\tau^2 = \frac{-1}{k_a^2}\log(a) > 0$$
$$\tau = \frac{-1}{k_r}\log(r) > 0.$$

By equating in terms of τ^2, then

$$\frac{-1}{k_a^2}\log(a) = \left(\frac{-1}{k_r}\log(r)\right)^2$$

and therefore

$$r = \exp\left[\frac{k_r}{k_a}\sqrt{-\log(a)}\right]. \tag{4.3.5}$$

Whilst Equation 4.3.5 might seem a trivial result, it is important to note that the derivation of the joint pdfs makes no assumption on the form of the input autocorrelation. The result is that Equation 4.3.5 is valid for any autocorrelation function $a(\tau)$ and structure constant $\gamma = 0.5$ and not just for Gaussian autocorrelations. Consequently, by altering the autocorrelation of the Gaussian processes used in the ratio method, the resulting Cauchy process' coherence will change according to Equation 4.3.5. It follows that the required autocorrelation for a desired output coherence function can be estimated by inverting Equation 4.3.5.

It also follows that the ratio k_r/k_a is constant and the input autocorrelation scale constant k_a and the output coherence scale constant k_r are related by some linear relationship $k_r = \nu k_a$. To verify this, the ratio method is repeated for different values of $k_a = 0.1, \ldots, 0.5$.

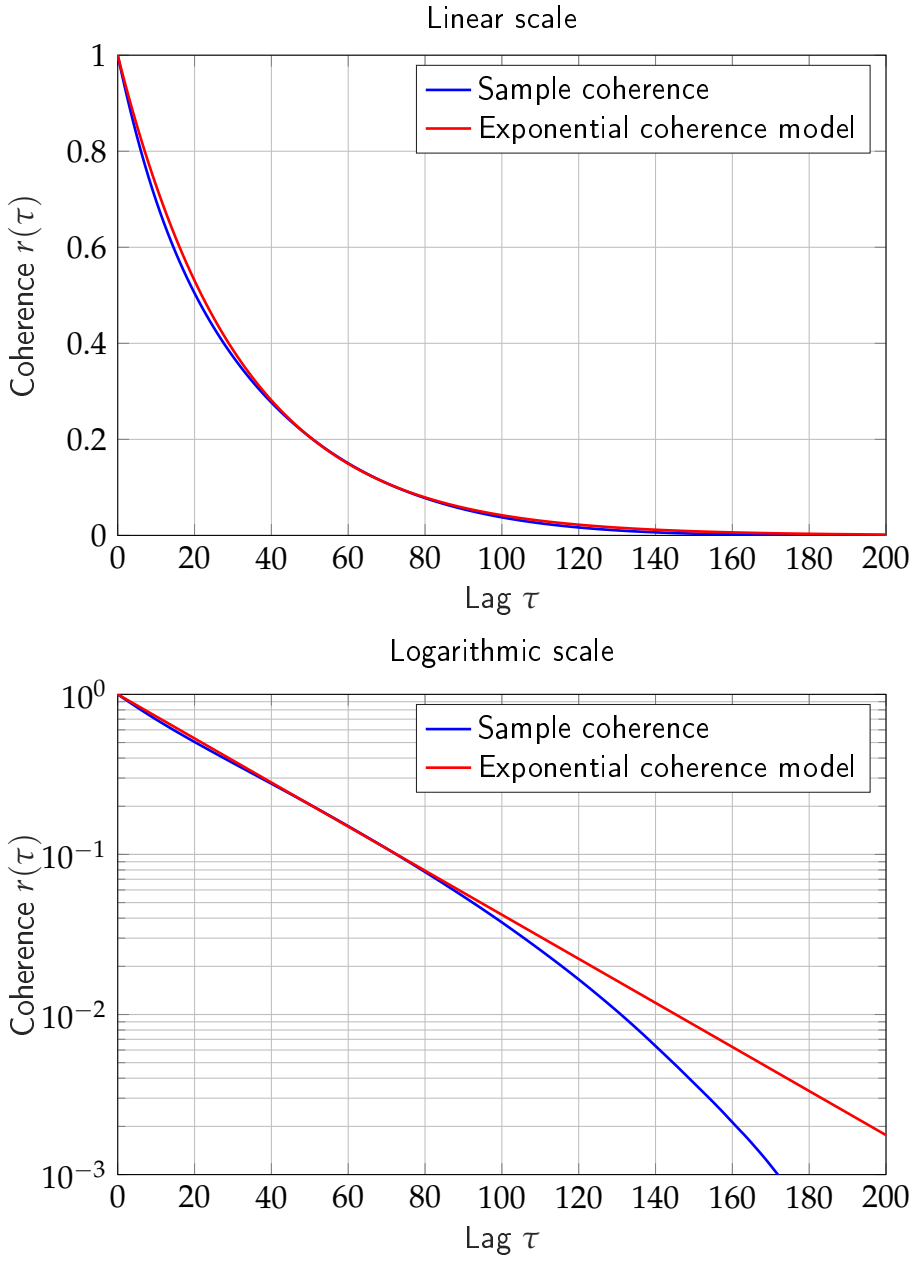

Figure 4.3.6.: Empirical coherence function of generated Cauchy process $\boldsymbol{U}$ plotted against an exponential fit $r(\tau) = \exp(-0.0317\tau)$. Plotted on linear and logarithmic scale.

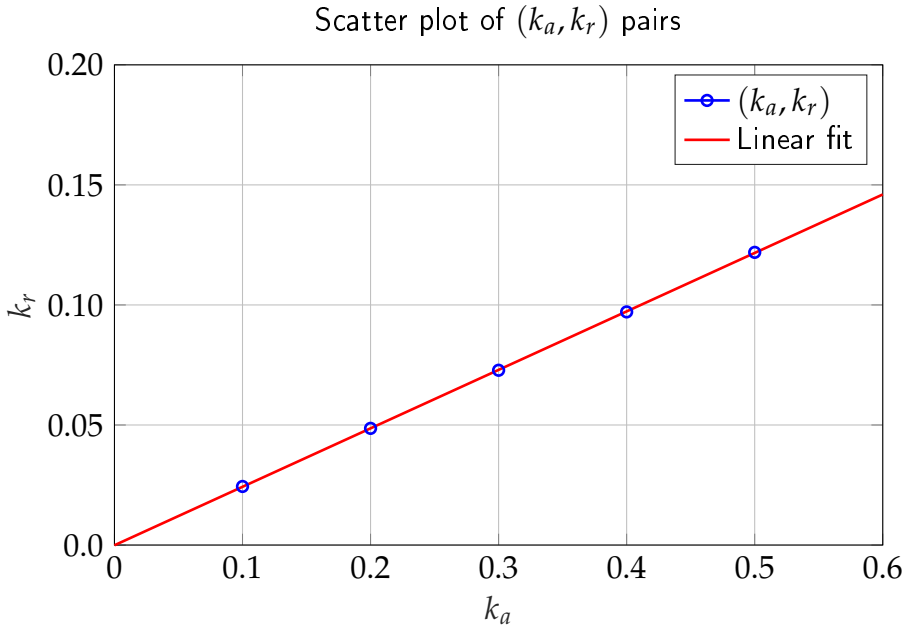

Figure 4.3.7.: Scatter plot of (k_a, k_r), where k_a is the scale constant of the input autocorrelation function, and k_r is the scale constant of the modelled exponential coherence. Fitted with a linear function $k_r = \nu k_a$ where $\nu = 0.244$.

By choosing k_r such that the sample coherence and exponential model have matching $r(\tau) = 0.1$ an estimate for k_r is calculated for each realisation. A scatter plot of the pairs of (k_a, k_r) is shown in Figure 4.3.7 and displays a strong linear relationship. Using linear regression, the scaling constant between k_a and k_r is calculated as $\nu = k_r/k_a = 0.244$.

4.4 THE RANK-REPLACEMENT METHOD

It is concluded that the testing validates the proposed means of generating a Cauchy coherent noise with a non-trivial coherence function. This only offers a specific (albeit very notable) example of a stable process. It is now shown that the methods introduced can be used to generate a realisation with any stable distribution and some non-trivial coherence. A realisation of a Cauchy coherent noise $\boldsymbol{U} = \{U_1, U_2, \ldots, U_n\}$ is generated via the ratio method. Additionally, a realisation of n iid stable random variables is generated and considered in terms of its order statistics $\boldsymbol{S} = \{S_{(1)}, S_{(2)}, \ldots, S_{(n)}\}$ where $S_{(1)} \leq S_{(2)} \leq \cdots \leq S_{(n)}$. The rank of each element of $\boldsymbol{U}$ is noted, such that $\text{rank}_{\boldsymbol{U}}(U_j) = k \in \mathbb{Z}$ says

that U_j is the kth largest element of $\boldsymbol{U}$. Then a new realisation $\boldsymbol{V} = \{V_1, V_2, \ldots, V_n\}$ is generated such that

$$V_j = S(\text{rank}_{\boldsymbol{U}}(U_j))$$

where $S(k) = S_{(k)}$ for notational convenience. This is equivalent to defining $\boldsymbol{V}$ as the shuffling of $\boldsymbol{S}$ such that its rank is equal to the rank of $\boldsymbol{U}$. This is described as the rank-replacement method for generating a stable coherent noise.

The initial realisation of the Cauchy coherent noise is generated by Gaussian processes with autocorrelation function $a(\tau) = \exp(-\log(2)\tau^2/64^2)$. The rank-replacement method is then used to generate realisations of stable coherent noises with stability parameters $\alpha = 0.2, 0.4, \ldots, 2.0$. The resulting sample coherence functions are calculated for structure constant $\gamma = \alpha/2$ and plotted in Figure 4.4.1.

As $\alpha \to 0$ (and the tail gets heavier), the memory (in the form of coherence function) becomes shorter, but retains the decay shape seen in the Cauchy generation method. It is concluded that this method allows us to generate a realisation of a process with any desired marginal stable distribution, that also that exhibits memory in the form of a non-trivial coherence function.

4.5 HEAVY-TAIL ESTIMATORS ON COHERENT NOISE

In section 3.4 the effect of memory on Hill's estimator was demonstrated by comparing Hill plots for different autocorrelation functions. At that point, commenting on whether this trend would be consistent when investigating non-Gaussian stable processes was difficult, as a means to generate stable coherent noises hadn't yet been introduced. With the work in this chapter, this is now possible. To do so, realisations of stable coherent noises are generated by the methods described in section 4.4. Estimates of the stability index can then be compared to the underlying value for both Hill's estimator and the characteristic function method estimator. To characterise the size of the memory and dependency inherent in the stable coherent noises, we use the 'half-life' of the coherence. This is the value

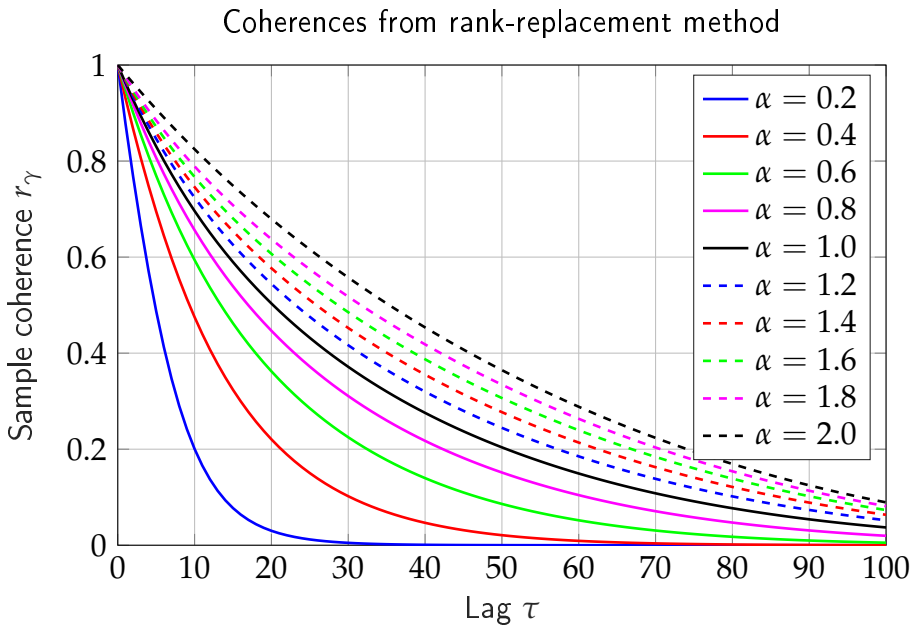

Figure 4.4.1.: Coherence functions of stable coherent noises generated by the rank-replacement method. Plotted for different stability parameters α with $\gamma = \alpha/2$.

$\tau_{0.5}$ such that the coherence $r(\tau_{0.5}) = 0.5$ (to the closest integer value).

To quantify the error in Hill's estimator, the average Hill estimator error $\epsilon = \langle|\hat{\alpha}_H - \alpha|\rangle/\alpha$ is calculated using 1000 samples, each with $n = 2^{12}$ elements, for stability indices $\alpha = 1.0,\ 1.5,\ 1.9$. The results are plotted for different threshold ratios in Figure 4.5.1, Figure 4.5.2 and Figure 4.5.3. These plots confirm expectations that memory continues to affect the viability of tail estimation using Hill's method, despite the autocorrelation being δ-correlated. The plots also quantify how Hill's estimator becomes more erroneous as $\alpha \to 2$ (regardless of coherence).

It has been shown previously that characteristic function methods are better tools for estimating the heavy-tail index α of stable processes, especially as $\alpha \to 2$. It is therefore important to clarify whether the characteristic function methods are anymore robust to memory in a stable process. To test this, the stable characteristic function method, as described in section 3.2, is used on points evaluated at $t_k = k\pi/300$ for $k = 1, \ldots, K$

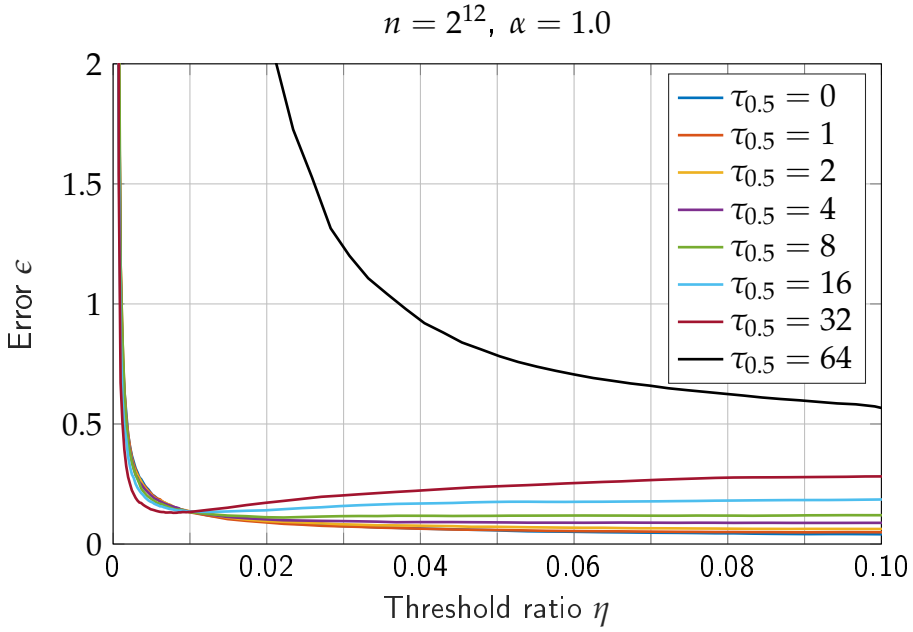

Figure 4.5.1.: Hill plots for stable coherent noises with heavy-tail index $\alpha = 1.0$. Plotted for different scales of coherence function.

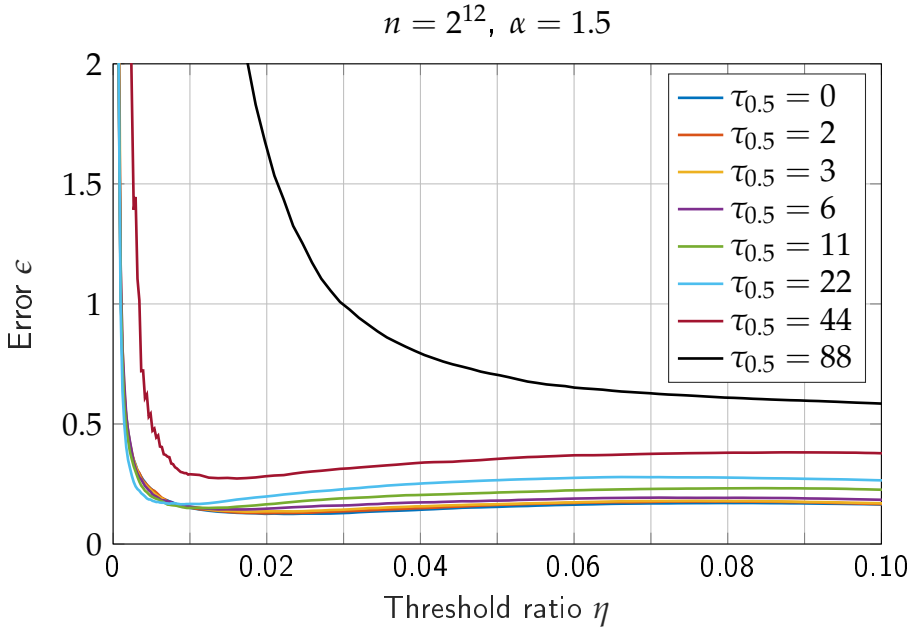

Figure 4.5.2.: Hill plots for stable coherent noises with heavy-tail index $\alpha = 1.5$. Plotted for different scales of coherence function.

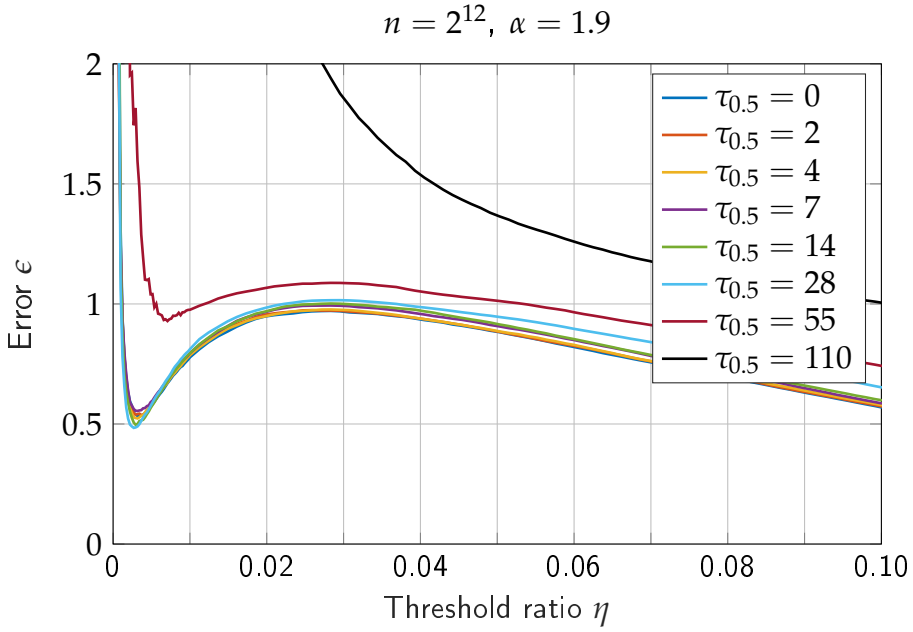

Figure 4.5.3.: Hill plots for stable coherent noises with heavy-tail index $\alpha = 1.9$. Plotted for different scales of coherence function.

for $K = 5,\ 10,\ 20$. This is done for realisations of stable coherent noise with $n = 2^{12}$ elements and the average error $\epsilon = \langle |\hat{\alpha}_S - \alpha| \rangle / \alpha$ is taken over 1000 samples. The results are plotted in Figure 4.5.4, Figure 4.5.5 and Figure 4.5.6. The results show that memory does have an effect on the accuracy of the estimation methods. For heavier-tails ($\alpha \to 0$) this becomes more pronounced. Near the upper limit $\alpha \approx 2$ the stable characteristic function method is more robust to the effect of memory. The results can be directly compared to the plots from Hill's estimator. For $\alpha = 1.0$, there is not a great deal of disparity between the validity of the estimators when memory is smaller. It is noted, however, that as memory becomes more sigificant, the stable characteristic function method is more robust. Similar conclusions can be drawn when comparing the results for $\alpha = 1.5$. It is clear that memory size has less of an effect on the average error. For $\alpha = 1.9$, the stable characteristic function method is significantly better regardless of memory. In fact, the stable characteristic function method error appears to be consistent, regardless of the scaling properties of the coherence function.

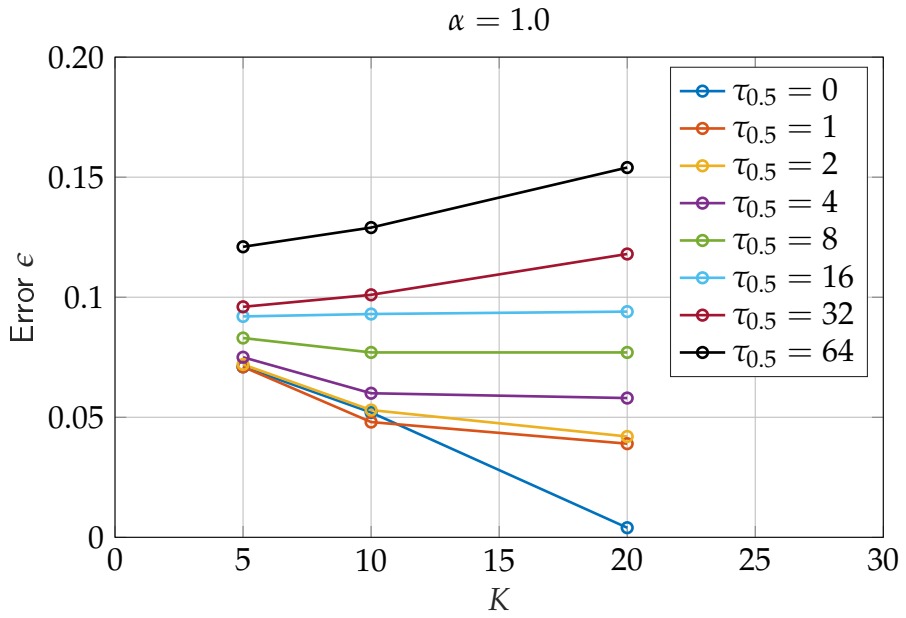

Figure 4.5.4.: Average errors $\epsilon = \langle|\hat{\alpha}_S - \alpha|\rangle/\alpha$ for stable characteristic function method estimate $\hat{\alpha}_S$. Plotted as a function of different sample intervals $t_k = k\pi/300$, $k = 1, \ldots, K$ with $\alpha = 1.0$.

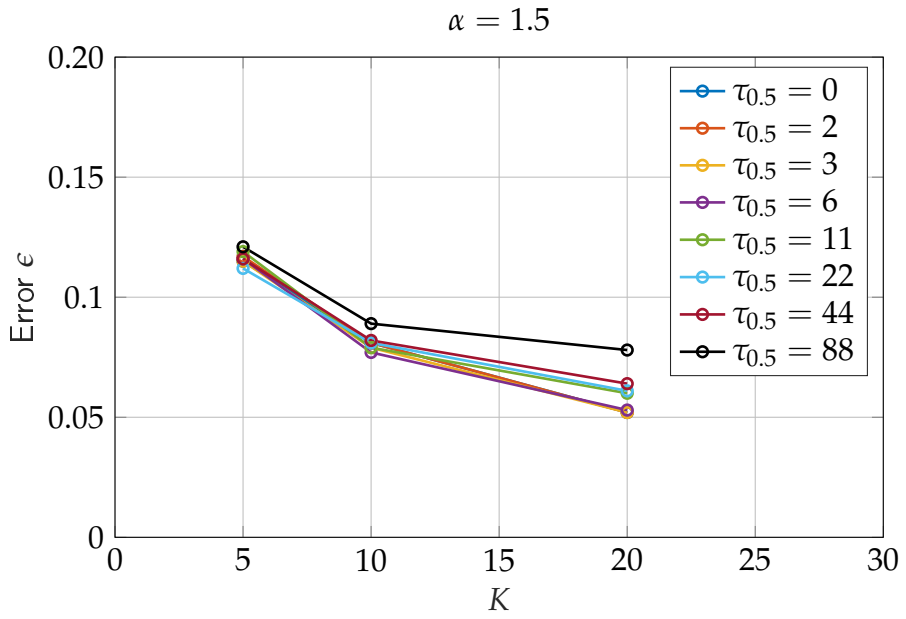

Figure 4.5.5.: Average errors $\epsilon = \langle|\hat{\alpha}_S - \alpha|\rangle/\alpha$ for stable characteristic function method estimate $\hat{\alpha}_S$. Plotted as a function of different sample intervals $t_k = k\pi/300$, $k = 1, \ldots, K$ with $\alpha = 1.5$.

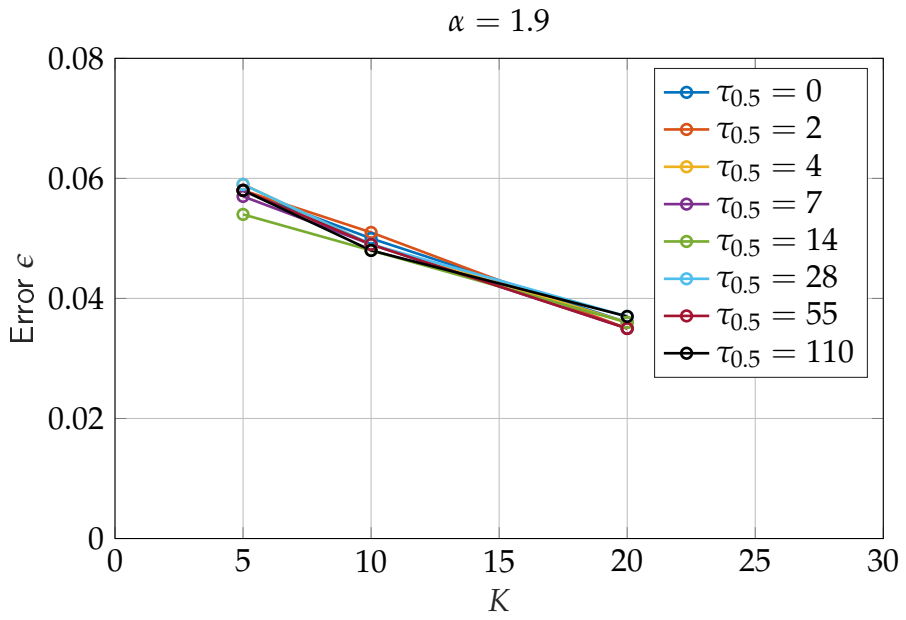

Figure 4.5.6.: Average errors $\epsilon = \langle|\hat{\alpha}_S - \alpha|\rangle/\alpha$ for stable characteristic function method estimate $\hat{\alpha}_S$. Plotted as a function of different sample intervals $t_k = k\pi/300$, $k = 1, \ldots, K$ with $\alpha = 1.9$.

The testing on stable coherent noises lends credence to the use of characteristic function methods over Hill's estimator. Whilst it is known that Hill's estimator is not accurate for stable processes as $\alpha \to 2$, it was not known the extent to which memory would further affect it. It has now been shown that it does make a significant difference. Conversely, the stable characteristic function method is not only more accurate for sets of iid stable random variables, but is more robust for stable coherent noise. Such a result further supports the use of characteristic function methods for identifying the stability index near the stable threshold $\alpha = 2$.

4.6 SUMMARY

In this chapter we have introduced a novel means of producing stable coherent noises. This has been done via the ratio methods, which initially generate realisations of Cauchy coherent noises. We have shown that the calculated joint pdf's of the 1st and 2nd ratio methods reflect the observed behaviour precisely and offer a model for the 3rd ratio

method. The empirical properties of the realisations produced by the 3rd ratio method have been investigated, by calculating the probability distribution, structure function and coherence function. As expected, the resultant realisations are δ-correlated, but do exhibit dependency that can be quantified by the coherence function. This lends credence to the use of the coherence function in general as a means to measure memory for stable processes.

The resulting sample coherence function has been modelled accurately by an exponential decay function when using the structure constant $\gamma = 0.5$. This also allowed us to relate the scale constant of the input autocorrelation function to the scale constant of the output coherence function. Such a result means Cauchy coherent noises can be generated with some target coherence function (not necessarily exponential).

We have shown that rank-replacing the elements of the realisation of a Cauchy coherent noise with those from a realisation of iid general stable random variables will conserve the memory whilst imprinting the desired distribution onto the process. The method is flawed in its ability to calculate the subsequent pdf, but still offers an adequate tool to generate a process with desired empirical properties.

The synthetic Cauchy coherent noises have then been used to investigate how memory affects tail-estimation methods for stable coherent noises. The methods described allowed us to complete the investigation from chapter 3 which was not possible without stable coherent noise. It was shown that memory does have an effect on the ability to measure the stability index using both Hill's estimator and characteristic function methods. It was observed, however, that characteristic function methods are more robust to the effect of memory.

The results from this chapter allow us to test the proposed theory to predict the zero-crossing statistics of stable coherent noise. Whilst we saw in chapter 3 that estimating the stability index accurately for stable processes is highly problematic, by generating the stable processes synthetically we can know the value innately. This is invaluable as it ensures testing of the theory in chapter 5 is optimal in choices for α and γ.

5

ZERO-CROSSINGS

Established work into the zero-crossing statistics of a Gaussian coloured noise was discussed in section 2.4. We saw McFadden's formula (Equation 2.4.5) could be solved for a Gaussian coloured noise by using Van Vleck's theorem (Equation 2.4.10) and Rice's formula (Equation 2.4.11). In this chapter, we will build on the precepts of Hopcraft and Jakeman in [19] and introduce generalisations of Van Vleck's theorem and Rice's formula for a stable coherent noise as functions of coherence. These will then be used to investigate the technically more problematic case of predicting the zero-crossing statistics of non-Gaussian stable processes.

We first restate McFadden's formula. Let $\{X(t)\}$ be a stationary continuous stochastic process with zero mean. Then the zero-crossing pdf's $P_n(\tau)$ are related to the clipped autocorrelation function $R(\tau)$ and the mean crossing-rate β by

$$\frac{R''(\tau)}{4\beta} = \sum_{n=0}^{\infty} (-1)^n P_n(\tau). \tag{5.0.1}$$

Here $P_n(\tau)$ is the pdf of the crossing intervals between the mth and $(m+n+1)$th zeros for $m = 1, 2, \ldots$. As discussed in subsection 2.4.1, by assuming independence of intervals, $P_0(\tau)$ can be solved via the inverse Laplace transform

$$P_0(\tau) = \mathcal{L}^{-1}\left[\frac{g(s)}{1-g(s)}\right] \tag{5.0.2}$$

where

$$g(s) = \mathcal{L}\left[\frac{R''(\tau)}{4\beta}\right]$$

We saw in subsection 2.4.2 that Equation 5.0.2 can be solved for a Gaussian process using Van Vleck's theorem and Rice's formula, both of which are functions of the Gaussian process' autocorrelation function.

5.1 GENERALISING VAN VLECK'S THEOREM AND RICE'S FORMULA

In order to solve McFadden's formula via Equation 5.0.2 for a stable coherent noise, the clipped autocorrelation $R(\tau)$ and the mean crossing-rate β must be determined. Here we introduce and describe the theoretical framework from Hopcraft and Jakeman in [19]. These equations describe the statistical properties of zero-crossings (and return-times) of stable coherent noises, taking advantage of the coherence function $r(\tau)$ introduced in subsection 2.2.2.

For the mean-crossing rate β, by noting the coherence function has expansion

$$r(\tau) = 1 - k\tau + O\left(\tau^2\right)$$

for $\tau \approx 0$, Hopcraft and Jakeman showed that

$$\beta = \frac{4k\alpha}{\pi^2} \int_0^1 \frac{q^{\alpha-1}}{(1+q^\alpha)^2} \log\left|\frac{1+q}{1-q}\right| dq. \tag{5.1.1}$$

This is a generalisation of Rice's formula. We can scrutinise the validity of Equation 5.1.1 by considering what happens for the Gaussian case where $\alpha = 2$. To do this, we are first reminded that, for the Gaussian case, the autocorrelation function and coherence function are related by

$$a(\tau) = \frac{2r(\tau)}{1+r^2(\tau)}. \tag{5.1.2}$$

By inserting the expansion of the autocorrelation function,

$$a(\tau) = 1 - \frac{a''(0)}{2}\tau^2 + O\left(\tau^3\right),$$

into Equation 5.1.2 we see that

$$r(\tau) \approx 1 - \sqrt{-a''(0)}\tau + O\left(\tau^2\right)$$

for $\tau \approx 0$ and therefore $k = \sqrt{-a''(0)}$. By inserting this (and $\alpha = 2$) into Equation 5.1.1, the mean-crossing rate is

$$\beta = \frac{1}{\pi}\sqrt{-a''(0)}$$

which corresponds to the Rice's formula, derived directly from the joint pdf in Appendix C. Similarly, Hopcraft and Jakeman showed that the clipped autocorrelation function for a symmetric stable process is

$$R(\tau) = \frac{4\,|r(\tau)|^{\alpha}}{\pi^2} \int_0^{\infty} \frac{q^{\alpha-1}}{1+|r(\tau)|^{\alpha} q^{\alpha}} \log\left|\frac{1+q}{1-q}\right| dq \tag{5.1.3}$$

where $r(\tau)$ is the coherence function and α is the stability index. As is the case with Equation 5.1.1, the integral in Equation 5.1.3 reduces to Van Vleck's theorem when we consider the special Gaussian case. Consequently, we have expressions for both the clipped autocorrelation function $R(\tau)$ and the mean-crossing rate β for stable coherent noises. Examples of the generalised forms of Rice's formula and Van Vleck's theorem being evaluated for several analytical coherence functions can be found in [15].

5.2 THEORETICAL CROSSING STATISTICS OF A CAUCHY PROCESS

5.2.1 *Identifying an exponential coherence*

To calculate the integral expressions for the generalisations of Rice's formula in Equation 5.1.1 and Van Vleck's theorem in Equation 5.1.3 an analytical coherence function is required. For the purposes of our investigation, we therefore need to identify an analytical model for the coherence function of the Cauchy coherent noise produced by the ratio method described in subsection 4.2.3.

To do this, a Cauchy realisation $\boldsymbol{X} = \{X_1, \ldots, X_n\} \sim \mathcal{C}(0,1)$ is generated using the ratio method. The input Gaussian processes have identical autocorrelation function

$$a(\tau) = \exp(-k_a^2 \tau^2)$$

where k_a is the scaling constant which is defined in terms of the parameter $\tau_{0.5}$ such that $k_a = \sqrt{\log(2)}/\tau_{0.5}$ and $a(\tau_{0.5}) = 0.5$. To ensure there is sufficient statistical information to make an informed decision regarding the coherence, a large $\tau_{0.5}$ is chosen so that there is significant memory in the process. Additionally, a large enough n is chosen such that there is a significant amount of total data and the dependency structures are small compared to the total sample size. Specifically, $\tau_{0.5} = 64$ and $n = 2^{24} \approx 16,800,000$ are selected.

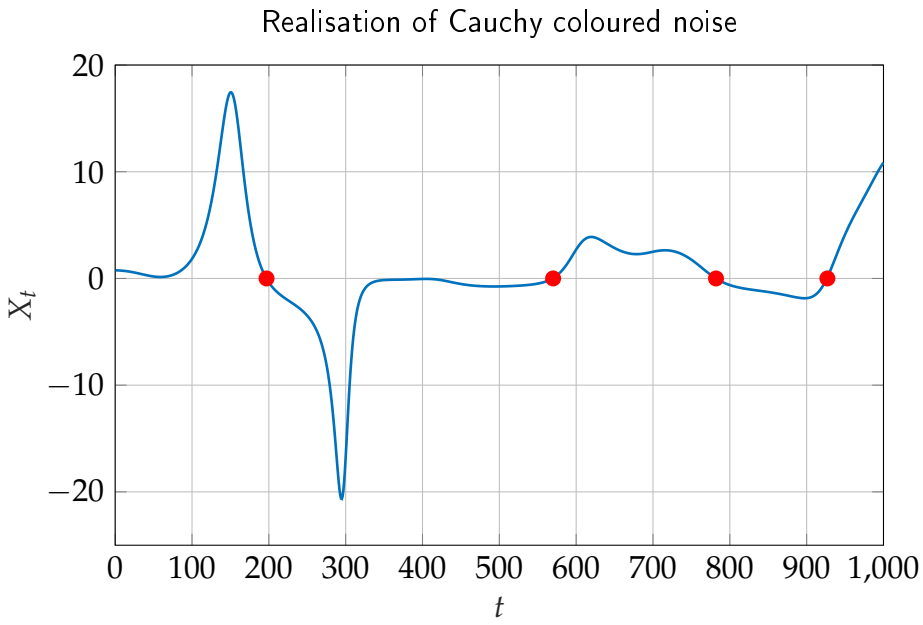

Figure 5.2.1.: Realisation of a Cauchy coherent noise $\boldsymbol{X} = \{X_t\}_{t=1}^{n}$ generated by ratio method, plotted for $t = 1, \ldots, 1000$. Zero-crossings highlighted in red circles. Note that this only displays a small portion of the entire process, which has $n = 2^{24}$ elements.

The realisation is plotted in Figure 5.2.1 for $t = 1, \ldots, 1000$. The zero-crossings in this interval are identified and circled.

To identify the coherence function, we remind ourselves that it was proposed in [19] that the coherence is related to the fractional moments by

$$\frac{\langle |X(t) - X(t+\tau)|^{\gamma} \rangle}{\langle |X(t)|^{\gamma} \rangle} = \left\{ \frac{2(1-r)^{\alpha}}{1+r^{\alpha}} \right\}^{\gamma/\alpha} \tag{5.2.1}$$

where r is the coherence, α is the stability index and $\gamma \in (-1, \alpha)$ is the structure constant. It was shown in section 4.3 that the choice of γ has an effect on the output coherence when inverting the sample structure function in Equation 5.2.1 to solve for r. We therefore express the sample coherence as a function of both τ and γ, such that $r = r_{\gamma}(\tau)$. This is displayed in Figure 5.2.2, where $r_{\gamma}(\tau)$ is plotted for $\gamma = -0.5,\ 0.1,\ 0.5,\ 0.9$. It is important to note that this isn't true for the underlying coherence, which is defined in terms of the joint characteristic function and makes no reference to γ. It is instead an artifact of its calculation through fractional moments. The result is that to properly investigate the

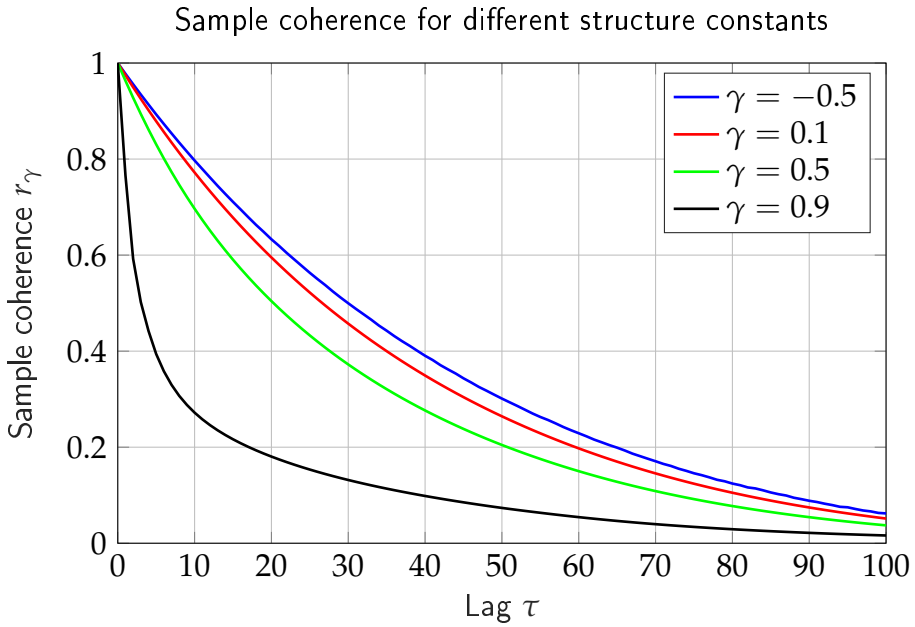

Figure 5.2.2.: Sample coherence function $r_\gamma(\tau)$ plotted for various structure constants γ. From this it is clear there is a need to choose an optimal γ that best models the empirical behaviour.

zero-crossing theories put forward, an optimal γ must be selected from the allowed interval $(-1, \alpha)$.

To do this, we assume the general form of Van Vleck's theorem in Equation 5.1.3 is correct. As the underlying distribution is Cauchy, its stability index is $\alpha = 1$ and the general form of Van-Vleck's theorem can be simplified to

$$R(\tau) = \frac{4r(\tau)}{\pi^2} \int_0^\infty \frac{1}{1 + q\,|r(\tau)|} \log \left| \frac{1+q}{1-q} \right| dq. \tag{5.2.2}$$

Whilst it is difficult to express an analytical inverse function of Equation 5.2.2 that would map the clipped autocorrelation R to the equivalent coherence r (to find an analytical function f^{-1} such that $r = f^{-1}(R)$), it is possible to approximate it numerically. Equation 5.2.2 can be computed for an arbitrary number of choices for input $r \in [0, 1]$. Then there is a discrete set of mappings from specific coherences to clipped autocorrelations, and therefore a discrete inverse mapping $\{r_1, \ldots, r_l\} \leftarrow \{R_1, \ldots, R_l\}$. Then for an arbitrary clipped autocorrelation $R(\tau)$ on the interval $R_j \leq R(\tau) < R_{j+1}$, it is possible to

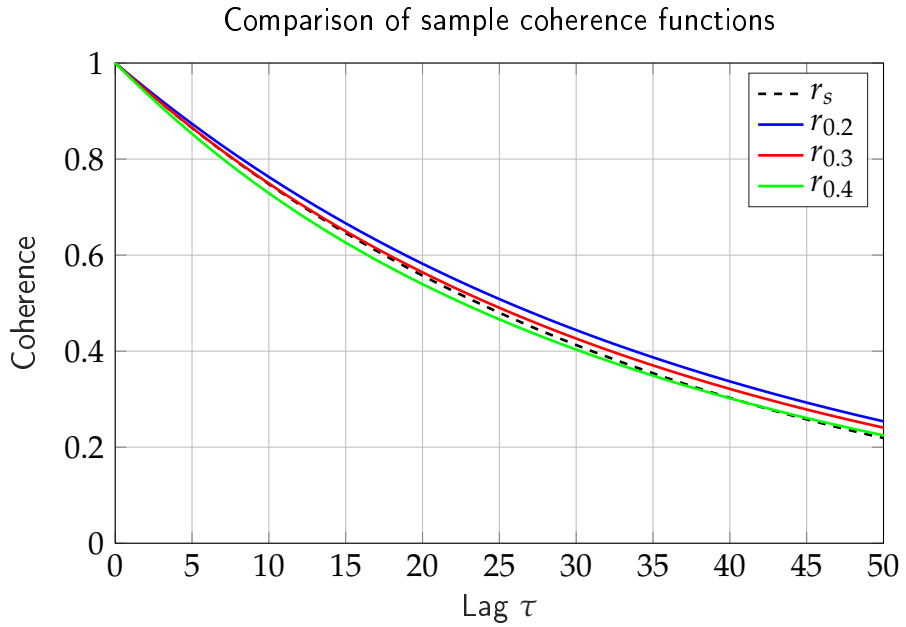

Figure 5.2.3.: Sample coherence functions for structure constants $\gamma = 0.2,\ 0.3,\ 0.4$, compared with the coherence r_s required to produce the empirical clipped autocorrelation.

approximate $r(\tau)$ by considering a linear interpolation from (R_j, r_j) to (R_{j+1}, r_{j+1}).

Using this linear interpolation, it is possible to therefore identify the coherence required to compute the sample clipped autocorrelation. Using the generated correlated Cauchy process $\boldsymbol{X} = \{X_1,\ \ldots,\ X_n\}$ and its subsequent clipped process $\boldsymbol{T} = \{T_1,\ \ldots,\ T_n\}$, the sample clipped autocorrelation is

$$R_s(\tau) = \langle T_t T_{t+\tau} \rangle = \frac{1}{n-\tau} \sum_{j=1}^{n-\tau} T_j T_{j+\tau}.$$

Using the linear interpolation methods with incremental coherence inputs of $\Delta r = 0.01$, an approximation of the coherence that would result in R_s is calculated, and denoted r_s. To determine the optimal γ to calculate the sample coherence r_γ, r_s is compared directly with r_γ for different choices of γ. Specifically we alter γ in increments of 0.1 from $\gamma = -0.9$ to $\gamma = 0.9$. We compare r_s to r_γ for several specific values of γ in Figure 5.2.3. From this it is concluded the optimal structure constant is $\gamma = 0.3$ (to 1 dp). In order to fit an analytical model to the optimal sample coherence $r_{0.3}$, it can be plotted on a logarithmic

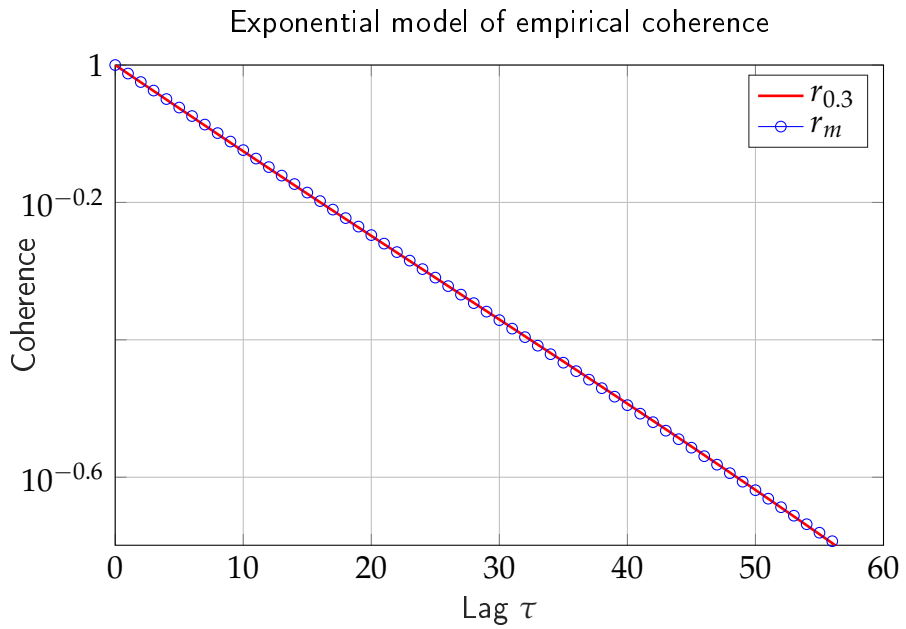

Figure 5.2.4.: Sample coherence $r_{0.3}$ and exponential fit r_m plotted on a logarithmic scale. This displays the coherence behaves exponentially.

scale, as seen in Figure 5.2.4. By doing so, it is apparent that an exponential decay model offers a good approximation. We denote the exponential model r_m, such that

$$r_m(\tau) = \exp(-k_1\tau)$$

where $k_1 > 0$ is some scaling constant. We choose k_1 such that $r_m(\tau) = r_{0.3}(\tau) \approx 0.5$ occurs at the same lag τ. We see that $r_{0.3}(24) = 0.5042$ and therefore choose $k_1 = 0.0285$ such that $r_m(24) = 0.5042$ also. The optimal sample coherence $r_{0.3}$, the desired sample coherence r_s and the exponential coherence model r_m are plotted alongside each other in Figure 5.2.5. It is concluded that the resultant Cauchy process produced by the ratio method has a coherence function that is accurately modelled by an exponential decay model r_m.

5.2.2 *Predicted crossing statistics of a Cauchy process*

In subsection 5.2.1 we saw that the coherence function of a Cauchy coherent noise produced by the ratio method is well modelled by an exponential function. Because of this, the theoretical crossing statistic measures introduced in section 5.1 can be calculated explicitly.

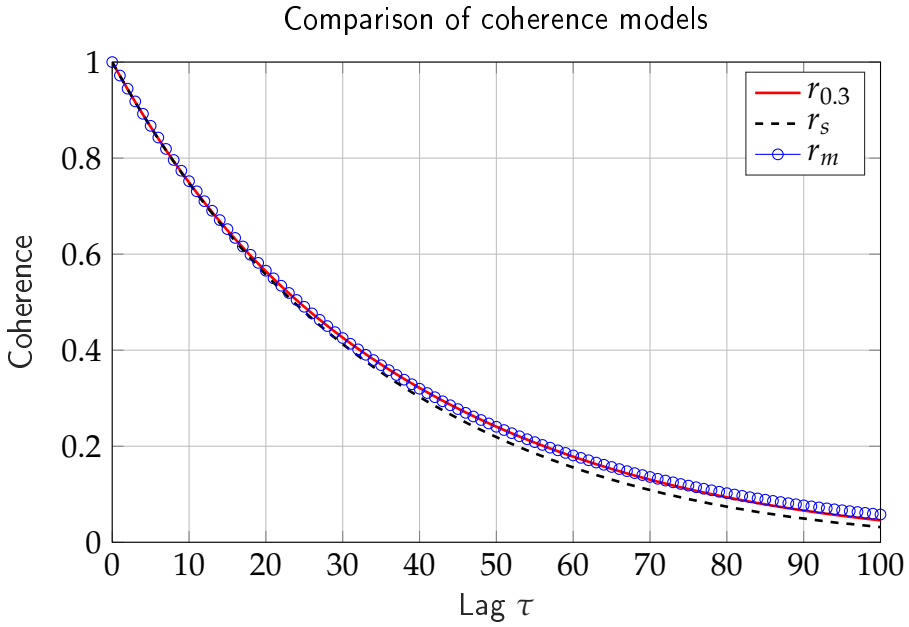

Figure 5.2.5.: Sample coherence $r_{0.3}$, an exponential model r_m and the sample coherence required to produce the empirical clipped autocorrelation r_s. The exponential model is $r_m(\tau) = \exp(-0.0285\tau)$. The consistency between all three shows that the exponential model is an appropriate model for the underlying coherence.

We shall henceforth assume $r(\tau) = \exp(-k_1\tau)$ where $k_1 > 0$ is the exponential scale constant.

The general form of Rice's formula for the mean crossing-rate of a stable coherent noise is expressed as an integral expression in Equation 5.1.1. For a Cauchy process, this can be simplified by inserting the stability index $\alpha = 1$ into Equation 5.1.1 such that

$$\begin{aligned}\beta &= \frac{4k}{\pi^2}\int_0^\infty \frac{1}{(1+q)^2}\log\left|\frac{1+q}{1-q}\right| dq \\ &= \frac{2k}{\pi^2}\end{aligned}$$

where k is a scale constant of the linear expansion of the coherence function about the origin, such that $r(\tau) = 1 - k\tau + O(\tau^2)$. Equivalently, $k = |r'(0)|$ (as we know $r \leq 1$). This is therefore equal to the exponential model scale constant k_1, because $r'(\tau) = -k_1\exp(-k_1\tau)$. It follows that the mean crossing-rate for a Cauchy process with an exponential coherence is

$$\beta = \frac{2k_1}{\pi^2} \tag{5.2.3}$$

where k_1 is the exponential scale constant.

The general form of Van Vleck's theorem for a stable coherent noise is expressed as an integral expression in Equation 5.1.3. For a Cauchy process where $\alpha = 1$, this simplifies to

$$R(\tau) = \frac{4\,|r(\tau)|}{\pi^2}\int_0^\infty \frac{1}{1+q\,|r(\tau)|}\log\left|\frac{1+q}{1-q}\right| dq.$$

By inserting an exponential coherence function, this further simplifies to

$$R(\tau) = \frac{4}{\pi^2}\int_0^\infty \frac{1}{q+\exp(k_1\tau)}\log\left|\frac{1+q}{1-q}\right| dq. \tag{5.2.4}$$

Using the clipped autocorrelation function and mean crossing-rate, we can investigate the pdf of the zero-crossings using McFadden's formula in Equation 2.4.5. By differentiating Equation 5.2.4 twice it follows that

$$R''(\tau) = \frac{4k_1^2}{\pi^2}\,\exp(k_1\tau)\int_0^\infty \frac{\exp(k_1\tau)-q}{(q+\exp(k_1\tau))^3}\log\left|\frac{1+q}{1-q}\right| dq$$

which, from [15], is equal to

$$R''(\tau) = \frac{4k_1^2}{\pi^2}\left(\frac{k_1\coth(k_1\tau)-1}{\sinh(k_1\tau)}\right). \tag{5.2.5}$$

Therefore (by Equation 5.2.5 and Equation 5.2.3)

$$\frac{R''(\tau)}{4\beta} = \frac{k_1}{2}\left(\frac{k_1\coth(k_1\tau)-1}{\sinh(k_1\tau)}\right)$$

and

$$\begin{aligned} g(s) &= \mathcal{L}\left[\frac{R''(\tau)}{4\beta}\right] \\ &= \frac{k_1}{2}\left[\frac{k_1^2+s^2}{k_1(k_1-s)^2} - \frac{s}{2k_1^2}\psi'\left(\frac{s-k_1}{2k_1}\right)\right] \end{aligned} \tag{5.2.6}$$

where ψ is the digamma function. The pdf of the zero-crossings can be found through the inverse Laplace transform

$$P_0(\tau) = \mathcal{L}^{-1}\left[\frac{g(s)}{1-g(s)}\right]. \tag{5.2.7}$$

The business of inverting the Laplace transform is a difficult operation to do analytically but can be done numerically via Talbot's method by assuming an analytical form of $g(s)/(1-g(s))$ (which we have done by approximating using exponential functions). A mathematical explanation of Talbot's method is discussed in Appendix E.

The final quantity to calculate is the persistence parameter θ, as described in subsection 2.4.3. We first define a normalised value of θ by

$$\tilde{\theta} := \frac{\theta}{k_1}$$

such that Equation 5.2.6 becomes

$$g(\theta) = \frac{1}{2}\left[\frac{1+\tilde{\theta}^2}{(1-\tilde{\theta})^2} - \frac{\tilde{\theta}}{2}\psi'\left(\frac{\tilde{\theta}-1}{2}\right)\right] \tag{5.2.8}$$

which is clearly no longer dependent on the scale factor k_1. In [15] the Newton-Raphson method [57] is used to solve Equation 5.2.8 for $g(\theta) = 1$, with result $\tilde{\theta} \approx 0.256$, such that the persistence parameter is

$$\theta \approx 0.256\, k_1 \tag{5.2.9}$$

for a Cauchy process. We have therefore presented the theoretical framework to calculate the key crossing-statistics of a Cauchy coherent noise with an exponential coherence function. These results can subsequently be checked against simulation results.

5.3 SAMPLE CROSSING STATISTICS OF A CAUCHY PROCESS

The theoretical framework surrounding the zero-crossing statistics of stable coherent noises is now compared to the sample statistics produced by synthetic Cauchy coherent noises. To generate the Cauchy processes to be tested, the 3rd ratio method from subsection 4.2.3 is used. We repeat the testing methods 6 times on realisations with $n = 2^{18}$ elements. As we wish to test the effect of memory on zero-crossings, the scale of the memory is altered for each realisation by adjusting the autocorrelation of the generating Gaussian processes. This is quantified by the value $\tau_{0.5}$, such that the autocorrelation function of the generating Gaussian processes is

$$a(\tau) = \exp(-k_a^2 \tau^2) \tag{5.3.1}$$

where $k_a = \sqrt{\log(2)}/\tau_{0.5}$. The testing methods are repeated for $\tau_{0.5} = 4$, 8, 16, 32, 64, 128, labelled $j = 1, \ldots, 6$ respectively for convenience. For each case, a model coherence function is fitted to the data. This is then used to calculate the theoretical crossing statistics using the methods from subsection 5.2.1. The raw data is then used to calculate the sample crossing statistics using the methods from subsection 5.3.1.

The success of the theory is judged on its accuracy in predicting the clipped autocorrelation $R(\tau)$, mean crossing-rate β, persistence factor θ and zero-crossing pdf $P_0(\tau)$. Additionally, the correlation between successive zero-crossing times is calculated as a measure of the accuracy of McFadden's assumption in Equation 2.4.7.

5.3.1 *Calculating sample crossing statistics*

To calculate the sample crossing statistics of a realisation, we now introduce simple means to compute the zero-crossings of a discrete-time signal. Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a discrete-time signal and $\boldsymbol{T} = \{T_1, \ldots, T_n\}$ be its equivalent clipped signal. Subsequently, define a new signal $\boldsymbol{\Pi} = \{\Pi_1, \ldots, \Pi_n\}$ by

$$\Pi_j := \begin{cases} 1 & j = 1, \\ T_j T_{j-1} & j = 2, \ldots, n. \end{cases}$$

As $T_j \in \{-1, 1\}$, it follows that $\Pi_j \in \{-1, 1\}$ and, more precisely, $\Pi_j = -1$ iff there is a sign change from $T_{j-1} \to T_j$ such that there is a zero-crossing at X_j.

The signal $\mathbf{\Pi}$ can be transformed to a string of 0's and 1's by a linear mapping $\mathbf{\Pi} \to \mathbf{\Lambda}$ where

$$\Lambda_j := \frac{1 - \Pi_j}{2}.$$

Then the resulting process $\mathbf{\Lambda}$ is the identifier signal of $\boldsymbol{X}$, such that $\Lambda_j = 1$ indicates a zero-crossing at X_j and $\Lambda_j = 0$ indicates no zero-crossing at X_j.

Subsequently, let F denote the find function, such that $F_{\mathbf{\Lambda}}(m)$ is the location of the mth $\Lambda_j = 1$ in the signal $\mathbf{\Lambda}$, and let $\boldsymbol{F} = \{F_{\mathbf{\Lambda}}(1),\ \ldots,\ F_{\mathbf{\Lambda}}(m)\}$ be the find signal, where m is the total number of 1's in $\mathbf{\Lambda}$. For the purposes of our investigation, $\boldsymbol{F}$ is calculated using the inbuilt MATLAB "Find" function [58].

The set of zero-crossing intervals (or return-times) $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_{m-1}\}$ is defined by

$$\tau_j := F_\Lambda(j+1) - F_\Lambda(j).$$

The pdf of the zero-crossings $P_0(\tau)$ is then calculated directly from $\boldsymbol{\tau}$. The mean return-time $\langle\tau\rangle$ is calculated by the expected value of $\boldsymbol{\tau}$, which is then used to calculate the mean crossing-rate $\beta = 1/\langle\tau\rangle$. The correlation between subsequent crossings can be calculated by $\rho_\tau = \langle[\tau_j - \langle\tau\rangle][\tau_{j+1} - \langle\tau\rangle]\rangle / \sigma_\tau^2$ where σ_τ^2 is the variance of $\boldsymbol{\tau}$, as a means to test McFadden's assumption.

To get an intuitive idea about how they are used and calculated, we compute each of the objects for a small example realisation

$$\boldsymbol{X} = \{0.4, 0.3, -0.2, -0.1, -0.1, 1.6, -0.4, -0.2\}.$$

Then the subsequent objects to investigate the zero-crossing statistics are

$$\boldsymbol{T} = \{1,1,-1,-1,-1,1,-1,-1\},$$
$$\boldsymbol{\Pi} = \{1,1,-1,1,1,-1,-1,1\},$$
$$\boldsymbol{\Lambda} = \{0,0,1,0,0,1,1,0\},$$
$$\boldsymbol{F} = \{3,6,7\},$$
$$\boldsymbol{\tau} = \{3,1\}.$$

It is these objects that are henceforth used to calculate the sample statistics of zero-crossings.

5.3.2 *The coherence function*

For each Cauchy realisation the coherence is calculated and modelled by an exponential function $r_m(\tau) = \exp(-k_1\tau)$ using the methods described in subsection 5.2.1. In each case, the optimal structure constant γ and resulting scale constant k_1 are recorded in Table 5.3.1. The sample coherences and exponential models for each sample are compared in Figure 5.3.1. A comparison between the exponential models on the same axes is plotted in Figure 5.3.2. We see that for every case the optimal structure constant is $\gamma = 0.3$ and this produces a coherence that is well modelled by an exponential function.

It is observed in Table 5.3.1 and Figure 5.3.2 that the scale of the coherence is approximately inversely proportional to $\tau_{0.5}$ of the input autocorrelation function. This can be understood further by producing a scatter plot of the pairs (k_a, k_1) for each realisation, where k_a is the autocorrelation scale constant in Equation 5.3.1. A linear plot $k_1 = m_a k_a + c_a$ is fitted and plotted in Figure 5.3.3. Using least-squares regression analysis, the linear parameters are determined to be $m_a = 2.200$, $c_a = 0.000$ to 3 dp. The relationship between k_1 and k_a is therefore approximately $k_1 = 2.200\ k_a$.

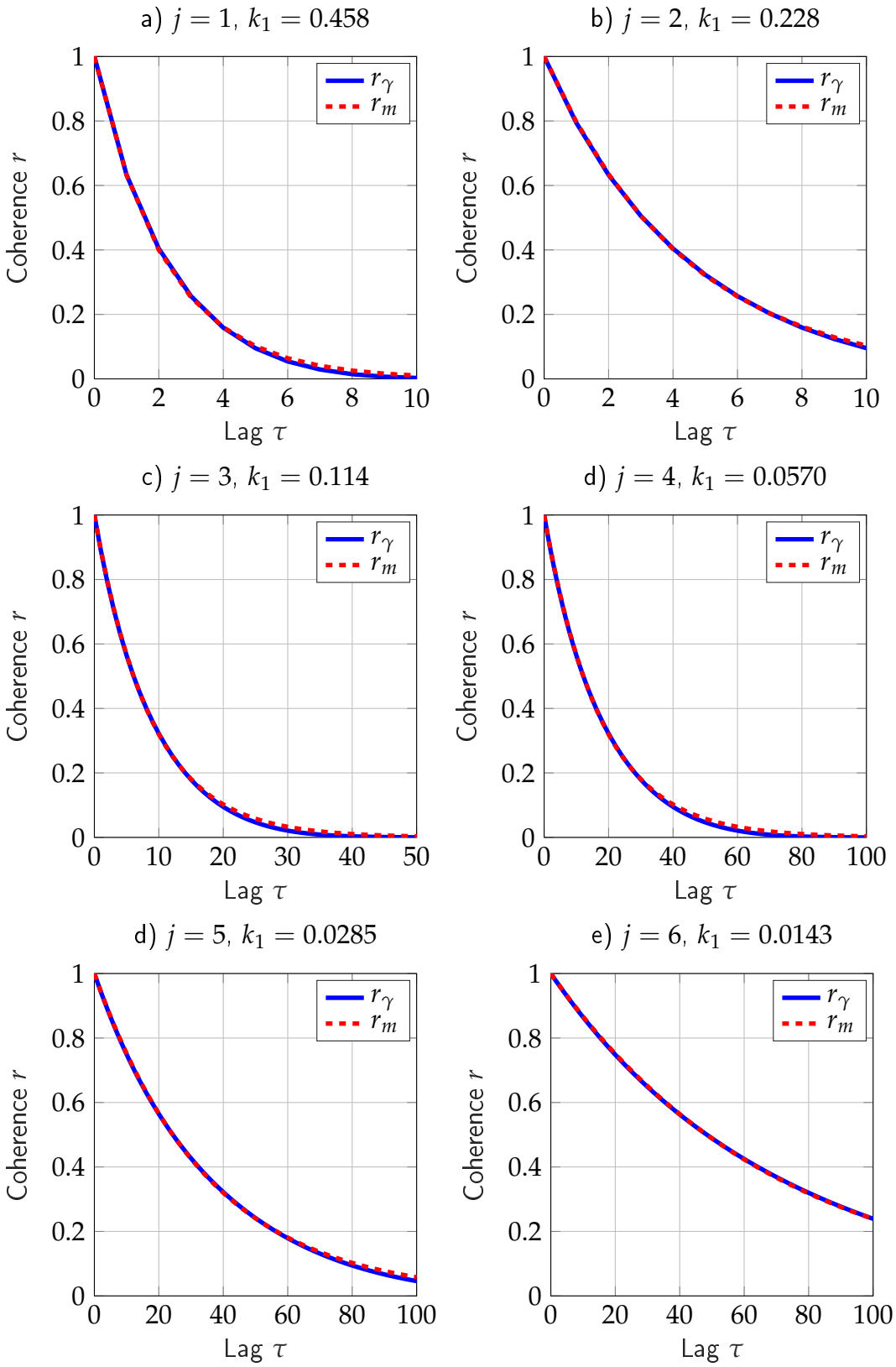

Figure 5.3.1.: Plots of empirical coherence r_γ calculated by optimal structure constant γ against the proposed exponential model r_m such that $r_m(\tau) = \exp(-k_1\tau)$.

j	$\tau_{0.5}$	γ	k_1
1	4	0.3	$4.58 \cdot 10^{-1}$
2	8	0.3	$2.28 \cdot 10^{-1}$
3	16	0.3	$1.14 \cdot 10^{-1}$
4	32	0.3	$5.70 \cdot 10^{-2}$
5	64	0.3	$2.85 \cdot 10^{-2}$
6	128	0.3	$1.43 \cdot 10^{-2}$

Table 5.3.1.: Record of results of Cauchy coherent noises and their exponential coherence model. For each realisation, we record the parameter $\tau_{0.5}$ of the Gaussian coloured noise, the optimal structure constant γ and the exponential model scale factor k_1.

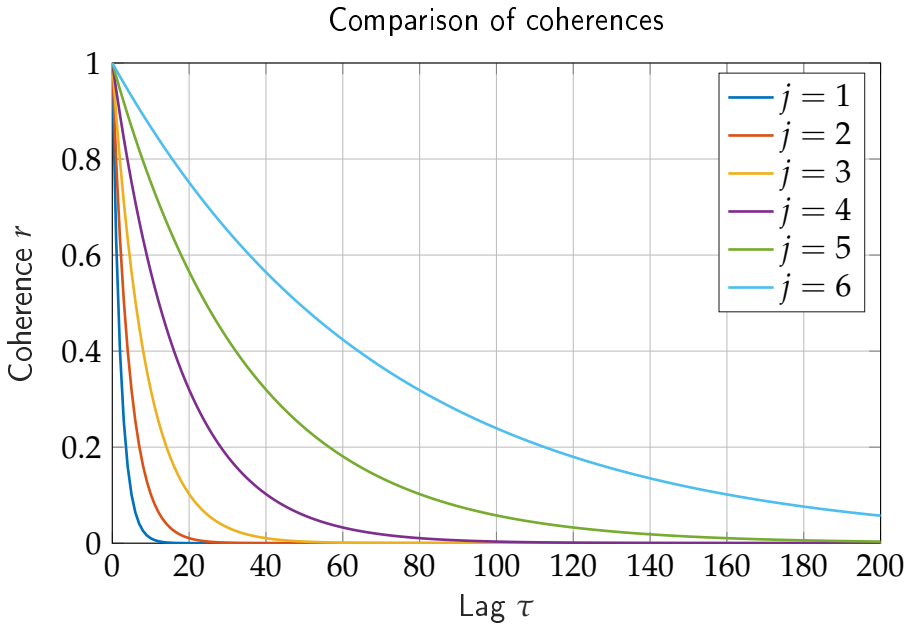

Figure 5.3.2.: Comparison of exponential models for coherence for $j = 1, 2, 3, 4, 5, 6$, with respective scale constants $k_1 = 0.458, 0.228, 0.114, 0.0570, 0.0285, 0.0143$.

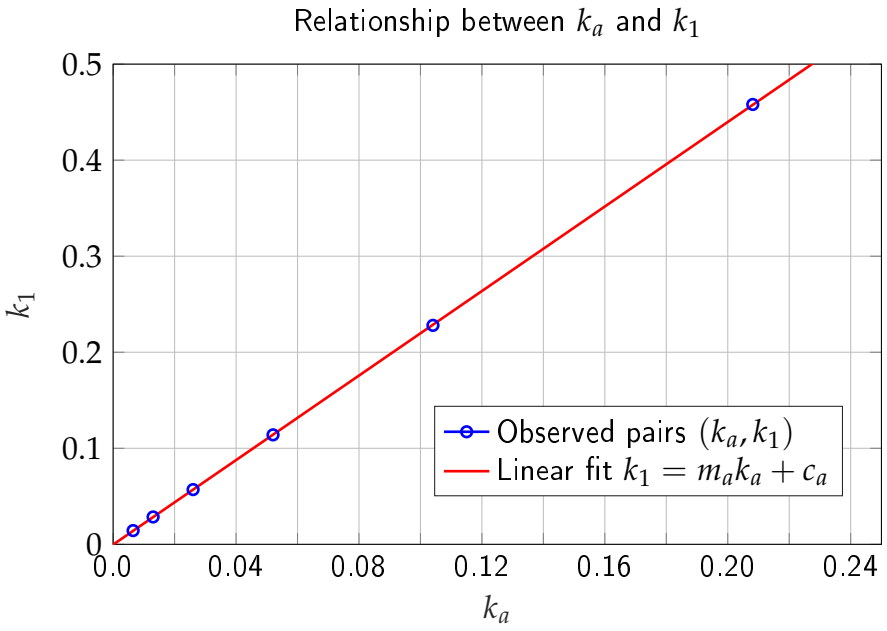

Figure 5.3.3.: Scatter plot of pairs (k_a, k_1) where k_a is the scale constant of the input autocorrelation function $a(\tau) = \exp(-k_a^2\tau^2)$ and k_1 is the scale constant of the resulting exponential model $r(\tau) = \exp(-k_1\tau)$. A linear relationship is fitted, such that $k_1 = m_a k_a + c_a$ where $m_a = 2.200$ and $c_a = 0.000$ to 3 dp.

j	k_1	β_s	β_m
1	$4.58 \cdot 10^{-1}$	$9.31 \cdot 10^{-2}$	$9.29 \cdot 10^{-2}$
2	$2.28 \cdot 10^{-1}$	$4.68 \cdot 10^{-2}$	$4.62 \cdot 10^{-2}$
3	$1.14 \cdot 10^{-1}$	$2.35 \cdot 10^{-2}$	$2.31 \cdot 10^{-2}$
4	$5.70 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$
5	$2.85 \cdot 10^{-2}$	$5.86 \cdot 10^{-3}$	$5.77 \cdot 10^{-3}$
6	$1.43 \cdot 10^{-2}$	$2.93 \cdot 10^{-3}$	$2.90 \cdot 10^{-3}$

Table 5.3.2.: Record of results of sample crossing-rate β_s and predicted crossing-rate β_m.

5.3.3 *Mean crossing-rate*

The first quantity to investigate is the mean crossing-rate β. For each realisation, the sample crossing-rate is calculated by $\beta_s = 1/\langle\tau\rangle$. Using the exponential model, the predicted crossing-rate is

$$\beta_m = \frac{2k_1}{\pi^2} \, .$$

The sample crossing-rate and predicted crossing-rate are recorded for each realisation in Table 5.3.2. It is seen that for every realisation the predicted mean crossing-rate is very close to the sample crossing-rate. The maximum error occurs in the $j = 3$rd realisation, which corresponds to an error of just 1.73%. It is therefore concluded that the theory is successful in predicting the mean crossing-rate.

5.3.4 *Clipped autocorrelation*

Next, the clipped autocorrelation $R(\tau)$ is investigated. For each realisation, the sample clipped autocorrelation function is

$$R_s(\tau) = \langle T_t T_{t+\tau} \rangle = \frac{1}{n-\tau} \sum_{j=1}^{n-\tau} T_j T_{j+\tau}$$

where $\boldsymbol{T} = \{T_1, \ldots, T_n\}$ is the clipped signal of the realisation $\boldsymbol{X}$. The predicted clipped autocorrelation for a Cauchy process with an exponential coherence function is $R_m(\tau)$, as defined in Equation 5.2.4. A comparison of the sample and predicted clipped

j	ρ_τ	θ_s	θ_m
1	$-1.51 \cdot 10^{-2}$	$1.24 \cdot 10^{-1}$	$1.13 \cdot 10^{-1}$
2	$-1.66 \cdot 10^{-2}$	$6.24 \cdot 10^{-2}$	$5.61 \cdot 10^{-2}$
3	$-1.72 \cdot 10^{-2}$	$3.14 \cdot 10^{-2}$	$2.80 \cdot 10^{-2}$
4	$-1.32 \cdot 10^{-2}$	$1.52 \cdot 10^{-2}$	$1.40 \cdot 10^{-2}$
5	$-1.86 \cdot 10^{-2}$	$7.85 \cdot 10^{-3}$	$7.01 \cdot 10^{-3}$
6	$-1.87 \cdot 10^{-2}$	$3.93 \cdot 10^{-3}$	$3.52 \cdot 10^{-3}$

Table 5.3.3.: Record of correlations between succesive return-times for each realisation, and sample and predicted persistence parameters.

autocorrelations are plotted in Figure 5.3.4 for each realisation. For every sample Cauchy process, the prediction is accurate for small τ where coherence is largest. This is significant as this region displays the most significant correlations and is chiefly responsible for the zero-crossing behaviour. There are slight errors in the tail of the clipped autocorrelation. This is a result of minor discrepencies between the exponential coherence model and the sample coherence in the tail. Such errors are not worrying in the context of this investigation.

5.3.5 *Probability density and persistence factor*

Using McFadden's formula, the clipped autocorrelation and mean crossing-rate can be used to calculate the zero-crossings pdf. McFadden's formula is constructed using the assumption that return-times are independent. This is used for the geometric progression equation in Equation 5.0.1. Therefore, before solving Equation 5.2.7 to predict the zero-crossings pdf, it is useful to calculate the correlations between successive return-times. For each realisation, the correlations between successive return-times are listed in Table 5.3.3. It is clear that the correlations are relatively insignificant, such that $|\rho_\tau| < 0.02$ for all cases. This ensures the approximation in McFadden's formula is appropriate for all the realisations.

By choosing appropriate bin sizes for the set of return-times $\boldsymbol{\tau}$ the sample zero-crossings pdf $P_{0,s}$ is produced. The means of deciding what bin sizes are appropriate is discretionary,

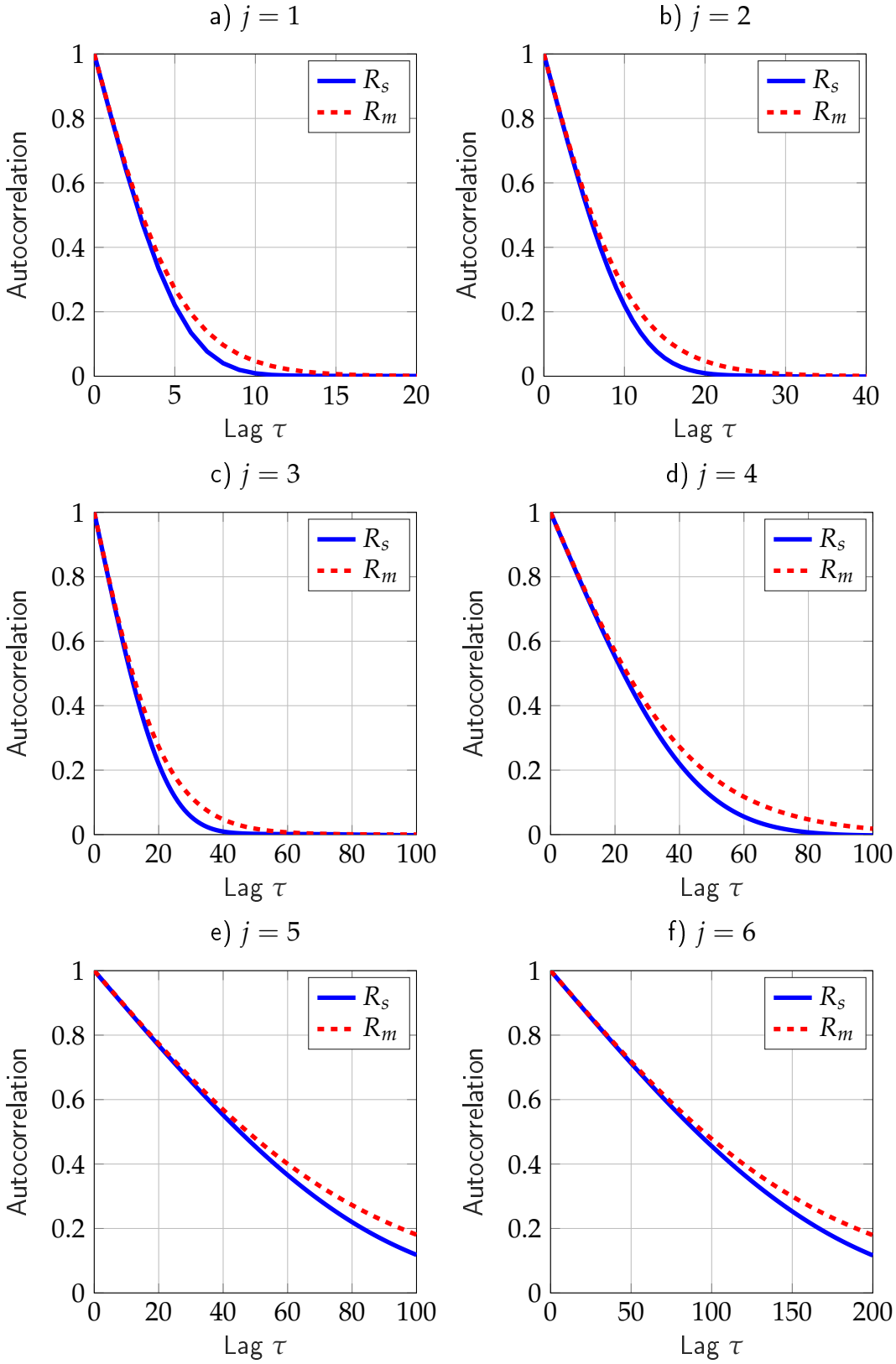

Figure 5.3.4.: Plots of sample clipped autocorrelation $R_s(\tau) = \langle T_t T_{t+\tau} \rangle$ against the predicted clipped autocorrelation function R_m calculated by Equation 5.2.4.

but we chose evenly spaced scaled intervals such that the central region is non-volatile for the linearly scaled axes, and the tail region is non-volatile for the logarithmically scaled axes. This allows for an accurate comparison between the predicted and observed behaviour. By plotting $P_{0,s}$ logarithmically and choosing an appropriate region upon which to fit an exponential tail, the sample persistence factor θ_s is calculated such that $P_{0,s}(\tau) \overset{a}{\sim} \exp(-\theta_s \tau)$ as $\tau \to \infty$. The pdf predicted by the exponential model, $P_{0,m}(\tau)$, is calculated by using Talbot's method (described in Appendix E) to solve Equation 5.2.7. The predicted persistence parameter θ_m is calculated by Equation 5.2.9. The sample and predicted pdf's, as well as the fitted exponential tail, are plotted on a linear scale in Figure 5.3.5 and on a logarithmic scale in Figure 5.3.6. The predicted and sample persistence parameters are tabulated in Table 5.3.3.

From Figure 5.3.5 is is clear that the model predicts the empirical behaviour very successfully for central values. There are very few discrepencies to address with any serious concern. In Figure 5.3.6 we see that the model is also accurate in the tail of the pdf, which is exponential as predicted in subsection 2.4.3. There are minor errors between the predicted and sample persistence factor, but only up to a maximum error of 12.1% for $j = 3$. These errors are noticeable, but not overwhelming when inspecting the logarithmically scaled plots in Figure 5.3.6.

It is concluded that the novel theory of coherence is successful in predicting the zero-crossing statistics of Cauchy coherent noises. By choosing an optimal coherence function, we have managed to produce accurate predictions for the mean crossing-rate β, the clipped autocorrelation function $R(\tau)$, the zero-crossings pdf $P_0(\tau)$ and the persistence factor θ. The results validate the generalisations of Rice's formula and Van-Vleck's theorem for symmetric stable processes, and the use of McFadden's formula.

5.4 SUMMARY

In this chapter we investigated the novel problem of predicting the zero-crossing statistics of a non-Gaussian stable coherent noise. To do this, generalised forms of Rice's formula

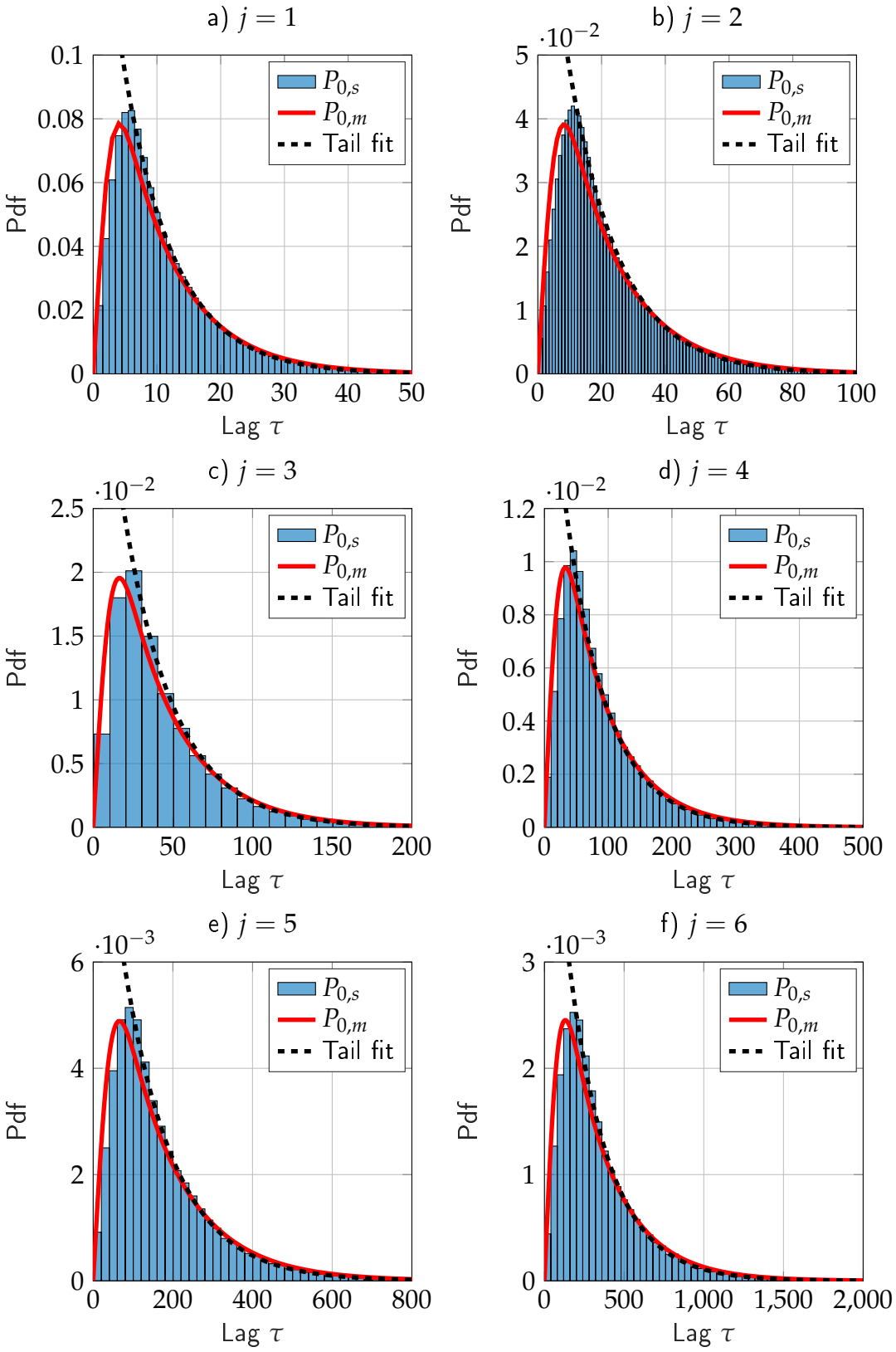

Figure 5.3.5.: Plots of sample zero-crossings $P_{0,s}$ against the predicted zero-crossings pdf $P_{0,m}$ calculated by using the exponential coherence model. Plotted on a linear scale.

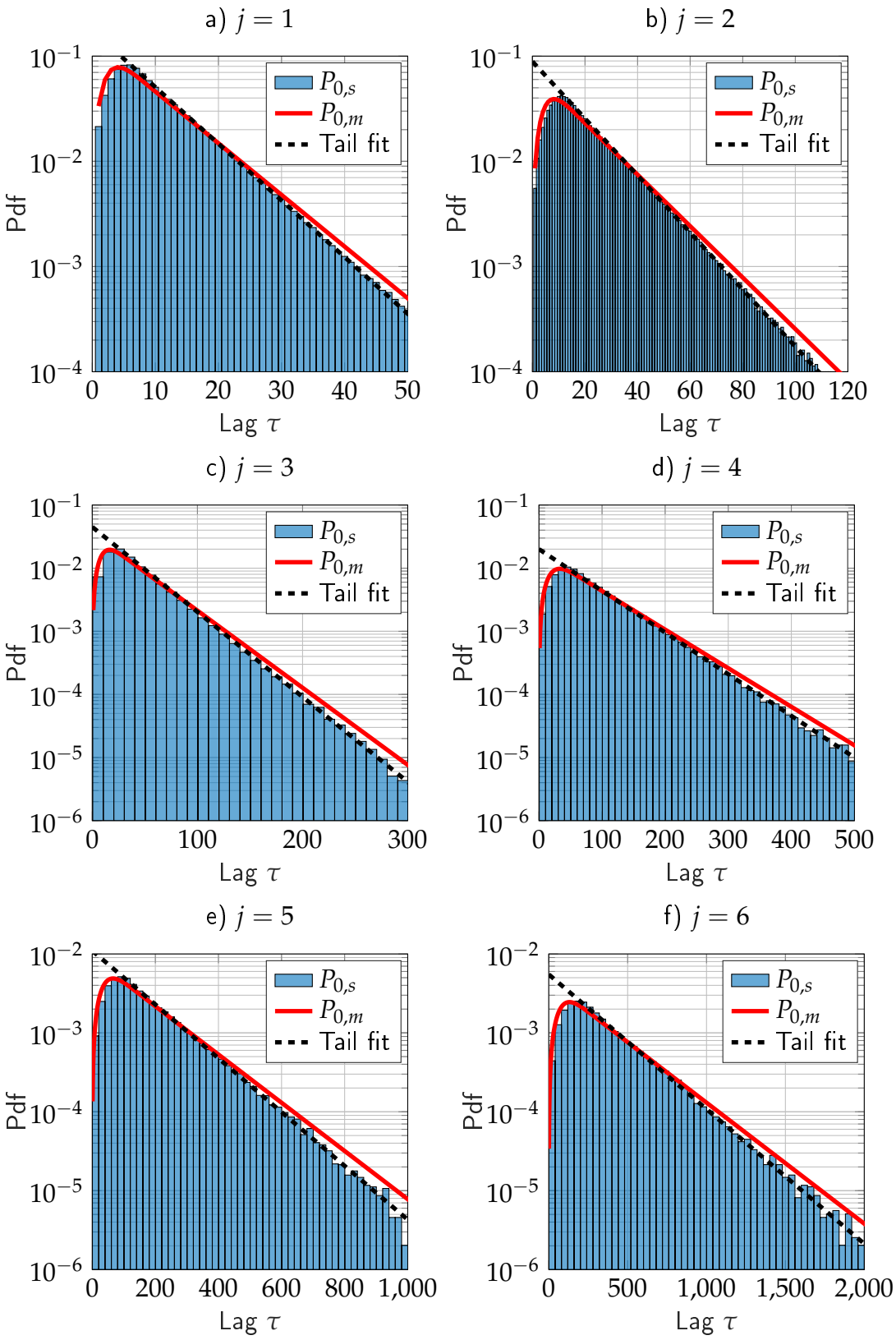

Figure 5.3.6.: Plots of sample zero-crossings pdf $P_{0,s}$ against the predicted zero-crossings pdf $P_{0,m}$ calculated by using the exponential coherence model. Plotted on a logarithmic scale.

and Van-Vleck's theorem were introduced for symmetric stable processes, as means of solving McFadden's formula. Both objects were expressed in terms of the coherence function.

These methods were then applied to Cauchy coherent noises produced by the ratio method from section 4.2. We were able to model the resultant processes' coherence function with an exponential function. By using this analytical model, we were then able to calculate Rice's formula, Van-Vleck's theorem and the theoretical persistence factor for a Cauchy coherent noise.

The predictions made by the generalisations of Rice's formula and Van-Vleck's theorem were compared directly with the sample crossing-rate and clipped autocorrelation in section 5.3. Additionally, they were used in conjunction with McFadden's formula to predict the zero-crossings pdf, which was then compared with the sample zero-crossings pdf. In all testing, the theory made accurate predictions of the zero-crossing properties. Such results validate the work in [19] as a means to encapsulate the memory in symmetric stable processes.

In summary, the results corroborate the few analytical properties that are known about non-Gaussian stable zero-crossings. We have shown that there exists a coherence function for a Cauchy coherent noise which, when combined with the theory of Hopcraft and Jakeman, and McFadden's formula, can produce an extremely accurate model of the zero-crossings. These validate the generalisations of Rice's formula and Van-Vleck's theorem for symmetric stable processes.

6

CONCLUSION

In this thesis we have investigated a proposed method to predict the zero-crossing statistics of stable coherent noises. Whilst work into the case of Gaussian stable processes is well established, we looked at the case of non-Gaussian stable processes, and the novel means of approaching it proposed by Hopcraft and Jakeman in [19]. To do this, we investigated several distinct areas.

We first introduced the existing work into zero-crossings of heavy-tailed stochastic processes, covering several distinct areas, in chapter 2. As a means to identify the heavy-tail index of heavy-tailed random variables (and processes) we introduced Hill's estimator based on upper-order statistics. The set of stable distributions were defined, and coherence was defined in terms of the joint characteristic function. Fourier Filtering was introduced as a means of generating (non-stable) heavy-tailed processes with a prescribed target autocorrelation function. The crossing statistics of Gaussian coloured noises were discussed, with specific reference to McFadden's formula, Van Vleck's theorem and Rice's formula. These topics were introduced as the fundamental framework required to investigate the novel case of non-Gaussian stable processes.

In chapter 3 we looked at the caveats in identifying the heavy-tail index. Hill's estimator was reviewed under different circumstances, allowing us to understand its flaws. We saw that whilst it performed well in idealistic settings (for large sets of iid non-stable random variables), its efficiency could be significantly affected if it wasn't. For stable processes, it struggled as the stability index approached the stable threshold (as $\alpha \to 2$). As an alternative we introduced the known characteristic function fitting method for stable processes.

Using this framework, we generalised the work to consider the characteristic function of non-stable heavy-tailed processes. The resulting work was used to propose an algorithm for calculating the heavy-tail index of a general heavy-tailed process. The expected behaviour and the potential pitfalls of the approach were investigated by considering the algorithm around the stable threshold $\alpha \approx 2$.

All three methods (Hill's estimator, stable characteristic function methods and general characteristic function methods) were tested on a range of synthetic data sets with varying (known) properties, in order to comment on each of the estimator's viability as a candidate to estimate real-world empirical data. It was shown that memory, data set size and proximity of the heavy-tail index to the stable threshold affected all of the methods with varying degrees of severity. It was concluded that characteristic function methods and Hill's estimator should be used in conjunction with each other to accurately estimate the heavy-tail index of a process, depending on the circumstances. The methods were then used to investigate the case of the IBM daily return price over a 10-year period, which was purpotedly well modelled by a stable distribution in [51]. In this case both Hill's estimator and the characteristic function methods agreed the data was asymptotically non-stable and that it would not be useful to consider such data sets to investigate zero-crossings of non-Gaussian stable processes.

The results from chapter 3 prompted a need to generate synthetic non-Gaussian stable processes. A means of doing this was introduced and discussed in chapter 4. We began by describing the well-known method of generating a Cauchy random variable via the ratio of two independent Gaussian random variables. This methodology was extended to consider the case of the ratio of two Gaussian coloured noises (that are themselves independent of one another). The resulting process was shown to be a Cauchy coherent noise, and the method for producing it was denoted the ratio method. The empirical properties of the Cauchy coherent noise were investigated, and it was shown that the presence of discontinuities made its use undesirable. We showed that by considering a positive denominator process these problems could be overcome.

The resulting optimal Cauchy process' empirical properties were then investigated, especially in the context of Hopcraft and Jakeman's proposed coherence function. We showed that the resulting process had the key properties expected of a Cauchy process - the correct sample distribution, a δ-correlated sample autocorrelation function and significant excursions from the 0-axis. It was shown that the joint probability density function of the process could be modelled and used to estimate the sample structure function with accuracy. The sample coherence function was calculated, and we showed that an exponential decay function offered an accurate model for it. This allowed us to accurately map the scale of the memory inherent in the generating Gaussian processes to the memory inherent in the resultant Cauchy process. We then proposed a means to generalise the ratio method to generate a stable coherent noise by rank-replacing the elements of a Cauchy coherent noise.

The Cauchy coherent noises were then used to investigate the tail estimation methods from chapter 3 for non-Gaussian stable coherent noises (as we hadn't been able to before). As was the case for non-stable processes, it was shown that memory had a significant effect on the ability of all of the estimators, as might be expected.

In chapter 5 we considered the problem of predicting the zero-crossing statistics of non-Gaussian stable processes. It was shown that McFadden's formula could still be used to estimate the zero-crossings pdf P_0, but Van-Vleck's theorem and Rice's formula could not. To alleviate this, generalisations of both Van-Vleck's theorem and Rice's formula, first introduced by Hopcraft and Jakeman, were defined. The theoretical framework was then tested on the synthetic Cauchy processes generated by the ratio method from chapter 4. It was shown that by choosing an optimal structure constant, the coherence could be modelled by an exponential decay function (as done previously), and such a function could be used to accurately predict both the mean crossing-rate and clipped autocorrelation. By using Talbot's method and McFadden's formula, the predicted zero-crossings pdf's were compared with the sample zero-crossings pdf's. Here we saw that the theory was successful in predicting the observed behaviour.

It was concluded from the investigation that the theoretical framework introduced by Hopcraft and Jakeman in [19] could be used to successfully predict the zero-crossing statistics of Cauchy coherent noises. Potential improvements on the investigation and scope for future work are discussed subsequently in section 6.1.

6.1 FURTHER WORK

6.1.1 *Improving the characteristic function method*

The characteristic function methods introduced in section 3.3 have been shown to have value in their ability to estimate the heavy-tail index of both stable and non-stable heavy-tailed processes. There remains scope to improve the methods however, especially when the heavy-tail index is outside the stable regime, such that $\alpha > 2$ and integer moments are subtracted from the empirical process.

Whilst in section 3.5 we have shown that choosing an optimal region improves the methods, we have not yet proposed a method to uniformally select this optimal region. If this can be done, then the characteristic function method can be investigated with an increased level of scrutiny in its ability to rival Hill's estimator for non-stable processes.

It is also noted that our investigation often deals with single realisations of the characteristic function method, and often does not look at the ensemble average of the error compared to Hill's estimator. By introducing a means to select an optimal region, this could be done and a more rigorous comparison with Hill's estimator could be achieved.

There are therefore several signficant areas in which the characteristic function method could be improved and investigated further. This is an avenue of potential future research, with a more complete picture of the characteristic function method desirable.

6.1.2 *Calculating the coherence and the role of optimal γ*

In subsection 2.2.2 the coherence between dependent stable random variables is defined as the quantity r that satisfies the joint characteristic function

$$\phi(t_1,t_2)=\exp\left(\frac{-1}{1+r^{\alpha}}\left(|t_1+rt_2|^{\alpha}+|rt_1+t_2|^{\alpha}\right)\right) \tag{6.1.1}$$

where $\alpha\in(0,2]$ is the stability index. For the Gaussian case with $\alpha=2$ the joint characteristic function reduces to

$$\phi(t_1,t_2)=\exp\left(-A\left\{t_1^2+t_2^2+2\left(\frac{2r}{1+r^2}\right)t_1t_2\right\}\right). \tag{6.1.2}$$

It was proposed in [19] that the coherence can be calculated by using fractional moments and the structure function, by solving

$$\left(\frac{2(1-r)^{\alpha}}{1+r^{\alpha}}\right)^{\gamma/\alpha}=\frac{\langle|X(t)-X(t+\tau)|^{\gamma}\rangle}{\langle|X(t)|^{\gamma}\rangle} \tag{6.1.3}$$

where $-1<\gamma<\alpha$. The value r that solves Equation 6.1.3 is labelled the sample coherence. Contrary to expectations, the sample coherence did not offer a single solution of the coherence, and instead gave a range of solutions depending on the chosen structure constant γ. Interestingly though, there was a certain choice of γ that allowed the sample coherence to successfully predict the zero-crossing statistics. There is therefore sufficient evidence that Equation 6.1.3 encompasses some of the properties of the underlying coherence, but is not sufficiently robust for all choices of γ. This therefore provides sufficient motivation and scope to investigate it further.

It is proposed that the reason for this inconsistency is due to the form of the joint characteristic function in Equation 6.1.1 used to calculate Equation 6.1.3. Whilst for the Gaussian case there is a unique joint characteristic function, as per Equation 6.1.2, this is not true for all stable random variables. For an n-dimensional general stable random variable, the joint characteristic function is

$$\phi(\boldsymbol{t})=\exp\left(i\boldsymbol{t}\cdot\boldsymbol{\delta}-\int_{\mathbb{S}}\omega_{\alpha}(\boldsymbol{t}\cdot\boldsymbol{s})\Lambda(d\boldsymbol{s})\right)$$

where $\boldsymbol{\delta}\in\mathbb{R}^n$ is a location vector, Λ is a finite measure on the unit sphere $\mathbb{S}=\{x\in\mathbb{R}^n:|\boldsymbol{x}|=1\}$, and

$$\omega_{\alpha}(t)=|t|^{\alpha}\left(1+i\,\mathrm{sgn}(t)\,\psi(t)\right)$$

where

$$\psi(t) = \begin{cases} \tan\frac{\pi\alpha}{2} & \text{if } \alpha \neq 1, \\ \frac{2}{\pi}\log|t| & \text{if } \alpha = 1. \end{cases}$$

This can be simplified to the joint characteristic function of an n-dimensional symmetric stable random variable

$$\phi(\boldsymbol{t}) = \exp\left(-\int_{\mathrm{S}} |\boldsymbol{t}\cdot\boldsymbol{s}|^{\alpha}\,\Lambda(d\boldsymbol{s})\right).$$

It is proposed that the ratio method used to generate the stable processes produces a process whose joint characteristic function is not exactly equal to the Equation 6.1.1, and hence Equation 6.1.3 does not hold for any choice of $-1 < \gamma < \alpha$. Due to the success of the coherence in estimating the zero-crossing statistics, it is clear that the joint characteristic function is (at least) very well modelled by Equation 6.1.1. Why a structure constant $\gamma = 0.3$ provides the best estimate of the coherence is therefore expected to be an artifact of the underlying joint characteristic function and the differences between it and Equation 6.1.1.

This provides several areas to investigate in the future. Firstly, it becomes important to clarify the joint characteristic function of the stable process produced by the ratio method. This will allow us to understand how accurately Equation 6.1.1 approximates the underlying joint characteristic function, and why $\gamma = 0.3$ allows us to extract the correct underlying coherence.

Secondly, given the success in this thesis in predicting the zero-crossing statistics using Equation 6.1.1, there exists an opportunity to generalise the coherence measure for all dependent stable random variables. By generalising the coherence and the theory in [19], it may be possible to correctly predict the zero-crossing statistics of all stable coherent noises.

6.1.3 *Generalising the ratio method*

In section 4.2 the ratio method for generating a Cauchy coherent noise was introduced. Specifically it was shown that for two Gaussian coloured noises $\{X(t)\}$, $\{Y(t)\}$, the ratio of the processes, defined by

$$U(t) := \frac{X(t)}{|Y(t)|}$$

will produce a Cauchy coherent noise $\{U(t)\}$. In section 4.4 we introduced a means to produce a realisation of a general stable coherent noise by rank-replacing the elements of a realisation of a Cauchy coherent noise. This method might be improved, such that a general stable realisation is produced directly from the ratio method. To do so, we consider the work into generating stable random variables in [59]. In this, a general stable random variable is generated by the ratio of two Gaussian random variables. First, let X, Y be Gaussian random variables with standard deviations σ_X, σ_Y respectively. Then

$$U_\alpha := \frac{X}{|Y|^{1/\alpha}}$$

has the pdf

$$f_{U_\alpha}(u) = \frac{1}{\pi\sigma_X\sigma_Y}\int_0^\infty y^{1/\alpha}\exp\left[-\left(\frac{y^2}{2\sigma_Y^2}+\frac{u^2y^{2/\alpha}}{2\sigma_X^2}\right)\right]dy$$

such that

$$f_{U_\alpha}(u) \overset{a}{\sim} |u|^{-(1+\alpha)}$$

as $|u| \to \infty$. Note that for $\alpha \in (0,2)$ this means the resulting process is asymptotically stable, (but is not stable in general). In [59] the standard deviations σ_X, σ_Y are adjusted such that f_{U_α} has the same tail as the stable distribution to be generated. Then n copies of the random variable are produced such that U_j is the jth copy, and a new random variable Z_n is defined by

$$Z_n := \frac{1}{n^{1/\alpha}}\sum_{j=1}^{n} U_j.$$

Then Z_n is a random variable whose distribution converges to the desired stable distribution as $n \to \infty$ (by the general CLT). It is proposed that this method can be adapted to produce a coherent noise that is (at least asymptotically) stable. To do this we would simply adjust

the current ratio method. If $\{X(t)\}$ is a Gaussian coloured noise and $\{Y(t)\}$ is a positive Gaussian coloured noise then it would follow that, by defining

$$U(t) := \frac{X(t)}{Y(t)^{1/\alpha}}$$

for $0 < \alpha < 2$, $\{U(t)\}$ is a(n asymptotically) non-Gaussian stable coherent noise. The exact nature of the resulting process would be of interest to investigate, especially regarding the ability to use the method to generate a general stable coherent noise.

6.1.4 *Testing on non-Cauchy stable processes*

The results in this thesis offer an initial investigation into the zero-crossings of a single type of stable coherent noises - specifically only Cauchy coherent noises have been tested empirically. With the ability to simulate general stable coherent noises, there is now an opportunity to investigate whether or not the theory continues to work for stable processes with stability parameters throughout the stable regime. It would be interesting to have a comprehensive review into this, and how a changing stability parameter might affect the results. Furthermore, it would then be interesting to find out if it could be applicable to real-world data sets that are purpoted to be non-Gaussian stable coherent noises.

A

GLOSSARY

A.1 OPERATORS

Below we list non-standard or potentially confusing mathematical objects and operators.

$f(x) \overset{a}{\sim} g(x)$ as $x \to \infty$: functions f and g are asymptotically equivalent up to some positive constant, such that $\lim_{x\to\infty} f(x)/g(x) = k$ where $k > 0$ (20).
$X \sim D$: random variable X has distribution D.
$X \overset{d}{=} Y$: random variable X is equal in distribution to random variable Y.
$\mathbb{P}(E)$: probability of event E.
$\langle X \rangle$: expected value of random variable X.
$X \xrightarrow{d} D$: random variable X converges in distribution to distribution D.

A.2 TERMS

Below we list uncommon, novel or especially important terms that may require referencing, along with the page number that first introduces or discusses them.

1st ratio method: see *Ratio method*.
2nd ratio method: (93).
3rd ratio method: (96).
2nd-order General Characteristic Function (GCF) method: (63).
α-tail index: see *Heavy-tail index*.

Amplitude Adjusted Fourier Transform (AAFT) method: (32).

Augmented Dickey-Fuller (ADF) test: (75).

Cauchy coherent noise: see *Coherent noise*.

Change of Variable theorem: (87)

Clipped function: (36)

Coherent noise: (27)

Coloured noise: (15).

Complementary characteristic function: (56).

Continuous process: see *Sample-continuous sample process*.

Continuous-time process: (13).

Discrete-time process: (13).

Empirical characteristic function: (50).

Fourier Filtering Method (FFM): (27)

General Characteristic Function (General CF, GCF) method: (54).

Heavy-tail index: (20).

Heavy-tailed: (20).

Hill's estimator: (21)

Iterative Amplitude Adjusted Fourier Transform (iAAFT) method: (32).

Iterative Fourier Filtering Method (iFFM): (33).

Log-Log Complementary Distribution (LLCD) plot: (21)

Noise: see stochastic noise.

Pareto distribution: (22)

Persistence parameter: (39)

Rank-replacement method: (111).

Ratio method: (88).

Rice's formula: (39).

Sample statistics: (16)

Sample path: (16)

Sample-continuous process: (13).

Stable (distribution, random variable, process etc.): (23).

Stable Characteristic Function (Stable CF, SCF) method: (50).

Stable coherent noise: see Coherent noise.

Stability parameter: (23).

Stationary: (13).

Stochastic noise: (14).

Threshold ratio: (42).

Truncated empirical characteristic function: (53).

Van Vleck's theorem: (39).

White noise: (14).

A.3 ACRONYMS

Below is a list of important acronyms that appear throughout the thesis. Their relevant page number (if not common and required) are listed above.

AAFT: Amplitude Adjusted Fourier Transform.

acf: autocorrelation function.

ccdf: complementary cumulative distribution function.

cdf: cumulative distribution function.

cf: characteristic function.

FFM: Fourier Filtering Method.

GCF method: General Characteristic Function method.

iAAFT: iterative Amplitude Adjusted Fourier Transform.

iff: if and only if.

iFFM: iterative Fourier Filtering Method.

iid: independent and identically distributed.

LLCD: Log-Log Complementary Distribution.

pdf: probability density function.

pmf: probability mass function.

rv: random variable.

SCF method: Stable Characteristic Function method.

B

VAN VLECK'S THEOREM

Here we derive the Van Vleck theorem for a Gaussian coloured noise [43]. Let $\{X(t)\}$ be a stationary Gaussian process with autocorrelation function $a(\tau)$, mean $\mu = 0$ and variance $\sigma^2 = 1$. Then the Van Vleck theorem (or arcsine law formula) states that the clipped autocorrelation function is

$$R(\tau) = \frac{2}{\pi} \sin^{-1} a(\tau).$$

To prove this, let $\chi(t)$ be the rectified clipped function such that

$$\chi(t) = \frac{1}{2}(1 + T(t)) = \begin{cases} 1 & T(t) \geq 0, \\ 0 & T(t) < 0. \end{cases} \tag{B.0.1}$$

where $T(t)$ is the clipped function of $X(t)$. The clipped autocorrelation function is defined $R(\tau) = \langle T(t)T(t+\tau)\rangle$. By stationarity of $T(t)$ it follows $R(\tau) = \langle T(0)T(\tau)\rangle$. Because the product $\chi(0)\chi(\tau) = 1$ iff $X(0), X(\tau) \geq 0$ (and is 0 otherwise), the autocorrelation function of $\chi(t)$ can be calculated by

$$\langle \chi(t)\chi(t+\tau)\rangle = \langle \chi(0)\chi(\tau)\rangle = \int_0^\infty \int_0^\infty m(x,y)\, dx\, dy \tag{B.0.2}$$

where $m(x,y)$ is the joint density function of $\{X(t)\}$. The joint density function of a bivariate Gaussian process with identical mean $\mu = 0$ and variance $\sigma^2 = 1$, and correlation ρ is

$$m(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{-(x^2 + y^2 - 2xy\rho)}{2(1-\rho^2)}\right). \tag{B.0.3}$$

Therefore, for a Gaussian process with autocorrelation $a(\tau)$, $\langle \chi(t)\chi(t+\tau)\rangle$ can be evaluated by using Equation B.0.3 in Equation B.0.2 with $\rho = a(\tau)$. The result is

$$\langle \chi(0)\chi(\tau)\rangle = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} a(\tau). \tag{B.0.4}$$

By taking the expectation of $\chi(0)\chi(\tau)$

$$\langle \chi(0)\chi(\tau) \rangle = \frac{1}{4} + \frac{1}{4}\langle T(0)T(\tau) \rangle = \frac{1}{4} + \frac{1}{4}R(\tau). \tag{B.0.5}$$

By equating Equation B.0.4 and Equation B.0.5 the clipped autocorrelation is

$$R(\tau) = \frac{2}{\pi} \sin^{-1} a(\tau) \tag{B.0.6}$$

as was to be demonstrated.

C

RICE'S FORMULA

Here we derive Rice's formula for a Gaussian coloured noise. Let $\{X(t)\}$ be a stationary differentiable Gaussian process with autocorrelation $a(\tau)$, mean $\mu = 1$ and variance $\sigma^2 = 1$. Then Rice's formula [14] states that the mean crossing-rate of $\{X(t)\}$ is

$$\beta = \frac{1}{\pi}\sqrt{-a''(0)}.$$

To prove this, consider the behaviour of $\{X(t)\}$ over the interval $(t_1, t_1 + \tau)$. We may choose τ to be small enough, such that the behaviour of $\{X(t)\}$ on this interval can be sufficiently modelled by a straight line, such that

$$X(t) = k\,(t - t_1) + X(t_1)$$

where k is the linear gradient, such that

$$k = \dot{X}(t) = \frac{dX(t)}{dt} = \left\{\frac{X(t_1 + \tau) - X(t_1)}{\tau}\right\}$$

on $t \in (t_1, t_1 + \tau)$. Suppose a zero-crossing occurs in this interval, such that $X(t) = 0$. Then

$$t = t_1 - \frac{X(t_1)}{\dot{X}(t)}.$$

As we consider $t \in (t_1, t_1 + \tau)$, it follows that

$$t_1 < t_1 - \frac{X(t_1)}{\dot{X}(t)} < t_1 + \tau. \tag{C.0.1}$$

Suppose $\dot{X}(t) > 0$ (a positive slope where the signal is crossing from a negative value to a positive value). Then, using Equation C.0.1, it follows that

$$-\dot{X}(t)dt < X(t_1) < 0. \tag{C.0.2}$$

Let $q(X, \dot{X})$ denote the joint pdf of $X = X(t_1)$, $\dot{X} = \dot{X}(t)$. Then the probability of X, $\dot{X}$ satisyfing Equation C.0.2 is

$$\int_0^\infty d\dot{X} \int_{-\dot{X}dt}^0 f(X, \dot{X}) dX = dt \int_0^\infty \dot{X} f(0, \dot{X}) d\dot{X} \tag{C.0.3}$$

by assuming sufficiently small dt sucht that $X \approx 0$. Now suppose $\dot{X}(t) < 0$ (a negative slope where the signal is crossing from a positive value to a negative one). Then using a similar methodology, the probability of X, $\dot{X}$ satisfying Equation C.0.2 is

$$-dt \int_{-\infty}^0 \dot{X} q(0, \dot{X}) d\dot{X}. \tag{C.0.4}$$

By combining Equation C.0.3, Equation C.0.4, the mean crossing rate is

$$\beta = \int_{-\infty}^\infty \left|\dot{X}\right| q(0, \dot{X}) d\dot{X}. \tag{C.0.5}$$

This is the general expression for the mean crossing-rate of a continuous and differentiable process. This formula can be directly applied to a Gaussian process by considering the joint pdf q. The joint pdf q can be calculated directly from the bivariate Gaussian distribution m defined in Equation B.0.3. This is done by considering a sufficiently small time interval. For convenience, let $X = X(t)$, $\dot{X} = \dot{X}(t)$ and $Y = X(t + dt)$. Then for sufficiently small dt, $Y = X + \dot{X}dt$. Through this, we consider the change of coordinate system $(X, Y) \to (X, \dot{X})$. Then the determinant of the Jacobian is

$$|J| = \left|\frac{\partial(X, Y)}{\partial(X, \dot{X})}\right| = t.$$

It follows that, if $m(X, Y)$ is the joint density function of X and Y, then the joint density function of X and $\dot{X}$ is

$$q(X, \dot{X}) = \lim_{t \to 0} t \; m(X, X + \dot{X}t). \tag{C.0.6}$$

Equation C.0.6 can be evaluated if the joint density function $m(X, Y)$ is known. For the Gaussian case, this can be done by considering the bivariate Gaussian pdf in Equation B.0.3. Rice showed that by inserting Equation C.0.6 into Equation B.0.3 that

$$q(X, \dot{X}) = \frac{1}{2\pi\sqrt{-a''(0)}} \exp\left(-\frac{1}{2}\left(X^2 - \frac{\dot{X}^2}{a''(0)}\right)\right)$$

and therefore

$$q(0, \dot{X}) = \frac{1}{2\pi\sqrt{-a''(0)}} \exp\left(\frac{\dot{X}^2}{2a''(0)}\right). \tag{C.0.7}$$

By substituting Equation C.0.7 into Equation C.0.5, the mean crossing-rate is

$$\begin{aligned}\beta &= \frac{1}{2\pi\sqrt{-a''(0)}}\, 2\int_0^\infty \dot{X}\exp\left(\frac{\dot{X}^2}{2a''(0)}\right) d\dot{X} \\ &= \frac{1}{\pi\sqrt{-a''(0)}}\left(-a''(0)\right) \\ &= \frac{1}{\pi}\sqrt{-a''(0)} \end{aligned} \tag{C.0.8}$$

for a Gaussian coloured noise, as was to be demonstrated.

D

COMPUTATION OF JOINT PDF OF CAUCHY PROCESS

In this we set out the methods used to produce the joint pdf of a Cauchy process, generated from the ratio of two Gaussian processes with non-trivial autocorrelation in section 4.2.

D.1 METHOD 1

Method 1 refers to the ratio of two Gaussian coloured noises, as per subsection 4.2.1. As per Equation 4.2.6, the joint pdf $f_{\boldsymbol{U}}$ of the resultant Cauchy process $\boldsymbol{U}$ is calculated by

$$f_{\boldsymbol{U}}(u_1, u_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2)dv_1dv_2 \tag{D.1.1}$$

where

$$f_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) = \frac{|v_1v_2|}{4\pi^2(1-\rho^2)}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{v_1^2\left(1+u_1^2\right)\right.\right.$$
$$\left.\left.+v_2^2\left(1+u_2^2\right) - 2\rho v_1v_2(1+u_1u_2)\right\}\right]. \tag{D.1.2}$$

To assist with the evaluation of the integral expressions *Mathematica* was used. To evaluate Equation D.1.1 the following simplifications are introduced. For convenience, the region of integration is denoted $\Omega = (-\infty,\infty)\times(-\infty,\infty)$. The following subregions, corresponding to the four quarters of the plane, are also introduced: $\Omega_1 = (0,\infty)\times(0,\infty)$, $\Omega_2 = (-\infty,0)\times(0,\infty)$, $\Omega_3 = (-\infty,0)\times(-\infty,0)$, $\Omega_4 = (0,\infty)\times(-\infty,0)$. For these subregions it is clear that

$$\int_{\Omega} f(v_1,v_2)\,dv_1\,dv_2 = \sum_{j=1}^{4}\int_{\Omega_j} f(v_1,v_2)\,dv_1\,dv_2$$

for any integrable function f on Ω. By inspecting Equation D.1.2, it is clear that $f_{\mathbf{U},\mathbf{V}}(u_1,u_2,v_1,v_2) = f_{\mathbf{U},\mathbf{V}}(u_1,u_2,-v_1,-v_2)$ and $f_{\mathbf{U},\mathbf{V}}(u_1,u_2,-v_1,v_2) = f_{\mathbf{U},\mathbf{V}}(u_1,u_2,v_1,-v_2)$ for all $v_1, v_2 > 0$. Therefore

$$\int_{\Omega_1} f_{\mathbf{U},\mathbf{V}}(u_1,u_2,v_1,v_2)dv_1dv_2 = \int_{\Omega_3} f_{\mathbf{U},\mathbf{V}}(u_1,u_2,v_1,v_2)dv_1dv_2 = I_1,$$
$$\int_{\Omega_2} f_{\mathbf{U},\mathbf{V}}(u_1,u_2,v_1,v_2)dv_1dv_2 = \int_{\Omega_4} f_{\mathbf{U},\mathbf{V}}(u_1,u_2,v_1,v_2)dv_1dv_2 = I_2.$$

It follows that

$$f_{\mathbf{U}}(u_1,u_2) = 2\left(I_1 + I_2\right). \tag{D.1.3}$$

As $v_1, v_2 > 0$ on Ω_1 it follows that $|v_1v_2| = v_1v_2$ and

$$I_1 = \frac{k}{2\pi^2}\int_0^\infty\int_0^\infty v_1v_2 \exp\Big[-k\Big\{v_1^2\left(1+u_1^2\right) + v_2^2\left(1+u_2^2\right) - 2\rho v_1v_2(1+u_1u_2)\Big\}\Big]dv_1dv_2 \tag{D.1.4}$$

where

$$k = \frac{1}{2(1-\rho^2)}.$$

Similarly, as $v_2 > 0 > v_1$ on Ω_2 it follows that $|v_1v_2| = -v_1v_2$ and

$$\begin{aligned} I_2 =& \frac{k}{2\pi^2}\int_0^\infty\int_{-\infty}^0 (-v_1v_2)\exp\Big[-k\Big\{v_1^2\left(1+u_1^2\right) + v_2^2\left(1+u_2^2\right) \\ & -2\rho v_1v_2(1+u_1u_2)\Big\}\Big]dv_1dv_2 \\ =& \frac{k}{2\pi^2}\int_0^\infty\int_0^\infty v_1v_2\ \exp\Big[-k\Big\{v_1^2\left(1+u_1^2\right) + v_2^2\left(1+u_2^2\right) \\ & +2\rho v_1v_2(1+u_1u_2)\Big\}\Big]dv_1dv_2. \end{aligned} \tag{D.1.5}$$

By *Mathematica*, Equation D.1.4 and Equation D.1.5 can be evaluated as

$$I_1 = \frac{1}{8kA\pi^2}\left\{1 + \frac{\rho(1+u_1u_2)}{\sqrt{A}}\left(\frac{\pi}{2} + \tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right)\right\},$$
$$I_2 = \frac{1}{8kA\pi^2}\left\{1 + \frac{\rho(1+u_1u_2)}{\sqrt{A}}\left(\frac{-\pi}{2} + \tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right)\right\}.$$

where $A = (1+u_1^2)(1+u_2^2) - \rho^2(1+u_1u_2)^2$. Then, from Equation D.1.3 it follows that

$$\begin{aligned} I &= \frac{1}{2kA\pi^2}\left\{1 + \frac{\rho(1+u_1u_2)}{\sqrt{A}}\tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right\} \\ &= \frac{(1-\rho^2)}{A\pi^2}\left\{1 + \frac{\rho(1+u_1u_2)}{\sqrt{A}}\tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right\}. \end{aligned}$$

Then it is concluded that the joint pdf of Method 1 is

$$f_{\boldsymbol{U}}(u_1,u_2)=\frac{(1-\rho^2)}{A\pi^2}\left\{1+\frac{\rho(1+u_1u_2)}{\sqrt{A}}\tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right\}$$

where

$$A=A(\rho,u_1,u_2)=\left(1+u_1^2\right)\left(1+u_2^2\right)-\rho^2(1+u_1u_2)^2.$$

D.2 METHOD 2

The 2nd ratio method refers to the ratio of a Gaussian coloured noise process with joint pdf

$$g_{\boldsymbol{X}}(x_1,x_2)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{x_1^2+x_2^2-2\rho x_1x_2\right\}\right] \tag{D.2.1}$$

for $x_1,x_2\in\mathbb{R}$ and a positive Gaussian coloured noise with joint pdf

$$g_{\boldsymbol{Y}}(y_1,y_2)=\frac{1}{\pi\sqrt{1-\rho^2}}\left\{\exp\left[\frac{-1}{2(1-\rho^2)}\left(y_1^2+y_2^2-2\rho y_1y_2\right)\right]+\exp\left[\frac{-1}{2(1-\rho^2)}\left(y_1^2+y_2^2+2\rho y_1y_2\right)\right]\right\}$$

for $y_1,y_2\geq 0$, and 0 otherwise, as per subsection 4.2.2. Then the joint pdf of the Cauchy coherent noise is

$$g_{\boldsymbol{U}}(u_1,u_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g_{\boldsymbol{U},\boldsymbol{V}}(u_1,u_2,v_1,v_2)dv_1dv_2$$

where

$$g_{\boldsymbol{U},\boldsymbol{V}}=\frac{|v_1v_2|}{2\pi^2\left(1-\rho^2\right)}\left\{\exp\left[\frac{-1}{2(1-\rho^2)}\left(v_1^2(1+u_1^2)+v_2^2(1+u_1^2)-2\rho v_1v_2(1+u_1u_2)\right)\right]+\exp\left[\frac{-1}{2(1-\rho^2)}\left(v_1^2(1+u_1^2)+v_2^2(1+u_1^2)+2\rho v_1v_2(1-u_1u_2)\right)\right]\right\} \tag{D.2.2}$$

for $u_1,u_2\in\mathbb{R}$, $v_1,v_2>0$, and 0 otherwise. It follows that

$$\begin{aligned}g_{\boldsymbol{U}}(u_1,u_2)&=\int_0^{\infty}\int_0^{\infty}g_{\boldsymbol{U},\boldsymbol{V}}(u_1,u_2,v_1,v_2)dv_1dv_2\\&=\frac{1}{2\pi^2(1-\rho^2)}\left(I_1+I_2\right)\end{aligned}$$

where

$$I_1 = \int_0^\infty \int_0^\infty v_1 v_2 \exp\left[\frac{-1}{2(1-\rho^2)}\left(v_1^2(1+u_1^2) + v_2^2(1+u_2^2) - 2\rho v_1 v_2(1+u_1u_2)\right)\right] dv_1 dv_2,$$

$$I_2 = \int_0^\infty \int_0^\infty v_1 v_2 \exp\left[\frac{-1}{2(1-\rho^2)}\left(v_1^2(1+u_1^2) + v_2^2(1+u_2^2) + 2\rho v_1 v_2(1-u_1u_2)\right)\right] dv_1 dv_2.$$

Using *Mathematica*, the integrals I_1, I_2 can be computed to

$$I_1 = \frac{\rho(1+u_1u_2)(1-\rho^2)^2}{A^{3/2}}\left[\frac{\pi}{2} + \frac{\sqrt{A}}{\rho(1+u_1u_2)} + \tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right]$$

$$I_2 = \frac{-\rho(1-u_1u_2)(1-\rho^2)^2}{B^{3/2}}\left[\frac{\pi}{2} - \frac{\sqrt{B}}{\rho(1-u_1u_2)} - \tan^{-1}\left[\frac{\rho(1-u_1u_2)}{\sqrt{B}}\right]\right]$$

where

$$A = A(\rho, u_1, u_2) = (1+u_1^2)(1+u_2^2) - \rho^2(1+u_1u_2)^2,$$

$$B = B(\rho, u_1, u_2) = (1+u_1^2)(1+u_2^2) - \rho^2(1-u_1u_2)^2.$$

By summing I_1 and I_2, it follows that

$$g_{\boldsymbol{U}}(u_1, u_2) = \frac{1-\rho^2}{2\pi^2}\left\{\frac{\rho(1+u_1u_2)}{A^{3/2}}\left(\frac{\pi}{2} + \frac{\sqrt{A}}{\rho(1+u_1u_2)} + \tan^{-1}\left[\frac{\rho(1+u_1u_2)}{\sqrt{A}}\right]\right) - \frac{\rho(1-u_1u_2)}{B^{3/2}}\left(\frac{\pi}{2} - \frac{\sqrt{B}}{\rho(1-u_1u_2)} - \tan^{-1}\left[\frac{\rho(1-u_1u_2)}{\sqrt{B}}\right]\right)\right\}. \quad \text{(D.2.3)}$$

Then Equation D.2.3 is the joint density function of the process produced by the 2nd ratio method.

D.3 METHOD 3

Method 3 refers to the ratio of a Gaussian coloured noise with joint pdf $h_{\boldsymbol{X}}(x_1, x_2) = g_{\boldsymbol{X}}(x_1, x_2)$ as per Equation D.2.1 and a positive Gaussian coloured noise with joint pdf

$$h_{\boldsymbol{Y}}(y_1, y_2) = \frac{1}{\pi\sqrt{1-\rho_1^2}} \left\{ \exp\left[\frac{-1}{2(1-\rho_1^2)} \left(y_1^2 + y_2^2 - 2\rho_1 y_1 y_2 \right) \right] + \exp\left[\frac{-1}{2(1-\rho_1^2)} \left(y_1^2 + y_2^2 + 2\rho_1 y_1 y_2 \right) \right] \right\}$$

where

$$\rho = \frac{2}{\pi - 2} \left(\sqrt{1-\rho_1^2} - 1 + \rho_1 \tan^{-1}\left[\frac{\rho_1}{\sqrt{1-\rho_1^2}} \right] \right). \tag{D.3.1}$$

This ensures that the $a_{\boldsymbol{Y}}(\tau) = a_{\boldsymbol{X}}(\tau)$. Then the joint pdf of the Cauchy coherent noise is

$$h_{\boldsymbol{U}}(u_1, u_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) dv_1 dv_2$$

where

$$h_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) = \frac{|v_1 v_2|}{2\pi^2\sqrt{(1-\rho_1^2)(1-\rho^2)}} \exp\left[\frac{-1}{2(1-\rho^2)} \left(u_1^2 v_1^2 + u_2^2 v_2^2 - 2\rho u_1 u_2 v_1 v_2 \right) \right] \left\{ \exp\left[\frac{-1}{2(1-\rho_1^2)} \left(v_1^2 + v_2^2 - 2\rho_1 v_1 v_2 \right) \right] + \exp\left[\frac{-1}{2(1-\rho_1^2)} \left(v_1^2 + v_2^2 + 2\rho_1 v_1 v_2 \right) \right] \right\}$$

for $u_1, u_2 \in \mathbb{R}$, $v_1, v_2 > 0$, and 0 otherwise. It follows that

$$\begin{aligned} h_{\boldsymbol{U}}(u_1, u_2) &= \int_0^{\infty} \int_0^{\infty} h_{\boldsymbol{U},\boldsymbol{V}}(u_1, u_2, v_1, v_2) dv_1 dv_2 \\ &= \frac{1}{2\pi^2\sqrt{(1-\rho^2)(1-\rho_1^2)}} \ (I_1 + I_2) \end{aligned}$$

where

$$I_1 = \int_0^\infty \int_0^\infty v_1 v_2 \ \exp\left[-\frac{1}{2}\left\{\frac{1}{1-\rho^2}\left(u_1^2 v_1^2 + u_2^2 v_2^2 - 2\rho u_1 u_2 v_1 v_2\right) + \frac{1}{1-\rho_1^2}\left(v_1^2 + v_2^2 - 2\rho_1 v_1 v_2\right)\right\}\right] dv_1 dv_2,$$

$$I_2 = \int_0^\infty \int_0^\infty v_1 v_2 \ \exp\left[-\frac{1}{2}\left\{\frac{1}{1-\rho^2}\left(u_1^2 v_1^2 + u_2^2 v_2^2 - 2\rho u_1 u_2 v_1 v_2\right) + \frac{1}{1-\rho_1^2}\left(v_1^2 + v_2^2 + 2\rho_1 v_1 v_2\right)\right\}\right] dv_1 dv_2.$$

Because Equation D.3.1 is non-invertible, it suffices to leave the integral forms for I_1, I_2 in the above forms. Using *Mathematica*, I_1, I_2 can be evaluated, with results

$$I_1 = \frac{m(1+\rho^2)(1+\rho_1^2)(\rho_1 + m\rho u_1 u_2)}{C^{3/2}}\left(\frac{\pi}{2} + \frac{\sqrt{C}}{\rho_1 + m\rho u_1 u_2} + \tan^{-1}\left[\frac{\rho_1 + m\rho u_1 u_2}{\sqrt{C}}\right]\right), \tag{D.3.2}$$

$$I_2 = \frac{-m(1+\rho^2)(1+\rho_1^2)(\rho_1 - m\rho u_1 u_2)}{D^{3/2}}\left(\frac{\pi}{2} - \frac{\sqrt{D}}{\rho_1 - m\rho u_1 u_2} - \tan^{-1}\left[\frac{\rho_1 - m\rho u_1 u_2}{\sqrt{D}}\right]\right), \tag{D.3.3}$$

where

$$C = C(\rho, \rho_1, u_1, u_2) = (1 + mu_1^2)(1 + mu_2^2) - (\rho_1 + m\rho u_1 u_2)^2,$$
$$D = D(\rho, \rho_1, u_1, u_2) = (1 + mu_1^2)(1 + mu_2^2) - (\rho_1 - m\rho u_1 u_2)^2,$$

and

$$m = m(\rho, \rho_1) = \frac{1-\rho_1^2}{1-\rho^2},$$

for $u_1, u_2 \in \mathbb{R}$.

E

TALBOT'S METHOD

Talbot's method is a means to solve the inverse Laplace transform and is described in [60]. Since its derivation, improvements have been made to it, including [61] and [62].

Let $f(t)$ be some function with Laplace transform $F(s) = \mathcal{L}[f(t)]$. Then the challenge of the inverse Laplace transform is to solve the Laplace transform relationship for an unknown $f(t)$ and known $F(s)$. We write this as $f(t) = \mathcal{L}^{-1}[F(s)]$. A standard means of solving the the inverse Laplace transform is via the integral expression

$$F(s) = \frac{1}{2\pi i} \int_B \exp(st)\ F(s)\ ds \tag{E.0.1}$$

where B is a vertical path from $\gamma_0 - i\infty$ to $\gamma_0 + i\infty$ and γ_0 is a real constant such that the real components of all singularities of $F(s)$ are less than γ_0. B is referred to as the Bromwich path, and Equation E.0.1 as the Bromwich integral [63]. Talbot's method improves upon this by deforming the Bromwich integral into a path described by

$$B(\theta) = \sigma_t + \mu_t\,(\theta \cot\theta + v_t i\theta)$$

where $\theta \in (-\pi, \pi)$, $\sigma_t \in \mathbb{R}$, $\mu_t, v_t > 0$ are chosen such that all singularities of $F(s)$ are to the left of the Bromwich path $B(\theta)$. Due to the exponential factor in the inverse Laplace transform, by defining the Bromwich integral as a contour such that $\mathrm{Re}(B(\theta)) \to -\infty$ as $|\theta| \to \pi$, the integrand decays rapidly on $B(\theta)$ and the integral converges rapidly [64].

Although there exists Talbot inversion software for MATLAB [65], [66], we computed the results in Mathematica using [67] based on the work in [61].

BIBLIOGRAPHY

[1] F. Black, M. Scholes "The Pricing of Options and Corporate Liabilities", *The Journal of Political Economy*, Vol. 81, No. 3, **637-654** (1973).

[2] D. Sanders "Extreme Events Part 2. Financial Catastrophes. The Overthrow of Modern Financial Theory" (2002). Available from: *https://www.actuaries.org.uk/documents/extreme-events-part-2-financial-catastrophes-overthrow-modern-financial-theory-0*.

[3] B. Mandelbrot "New methods in statistical economics", *The Journal of Political Economy*, Vol. 71, No. 5, **421-440** (1963).

[4] S. Rachev, S. Mittnik "Stable Paretian Models in Finance", Series in Financial Economics and Quantitative Analysis, *Wiley* (2000).

[5] G. Peach "Theory of the pressure broadening and shift of spectral lines", *Advances in Physics*, Vol. 30, No. 3, 367-474 (1981).

[6] P. Bak, C. Tang, K. Wiesenfeld, "Self-organized criticality", *Phys Rev A*, Vol. 38, No. 1, 364 (1988).

[7] D. Leddon "Statistical Study of Hard X-Ray Solar Flares", PhD Thesis, *University of North Texas* (2001).

[8] A. Clauset, C. R. Shalizi, M. E. J. Newman "Power-Law Distributions in Empirical Data", *SIAM Review*, Vol. 51, No. 4, **661-703** (2009).

[9] H. Aoyama, Y. Nagahara, M. P. Okazaki, W. Souma, H. Takayasu, M. Takayasu "Pareto's law for income of individuals and debt of bankrupt companies", *Fractals*, Vol. 8, No. 3, **293-300** (2000).

[10] M. Levy, S. Solomon "New evidence for the power-law distribution of wealth", *Physica A*, Vol. 242, **90-94** (1997).

[11] C. Furusawa, K. Kaneko "Zipf's Law in Gene Expression", *Phys. Rev. Lett*, Vol. 90, No. 8, **088102** (2003).

[12] T. Wilhelma, P. Hanggi "Power-law distributions resulting from finite resources", *Physica A*, Vol. 329, 499-508 (2003).

[13] M. Kac "On the average number of real roots of a random algebraic equation (ii)", *Proceedings of the London Mathematical Society*, Vol. 2, No. 1, **390-408** (1948).

[14] S. O. Rice "Mathematical analysis of random noise.", *Bell System Technical Journal*, Vol. 24, No. 1, **46-156** (1945).

[15] Y. Cao "Zero-crossing intervals of Gaussian and symmetric stable processes.", PhD thesis, University of Nottingham (2016).

[16] I. F. Blake, W. C. Lindsey "Level-crossing problems for random processes", *IEEE Trans. on. Info. Theory*, Vol. 19, No. 3, **295-315** (1973).

[17] J. A. McFadden "The axis-crossing intervals of random functions.", *Information Theory, IRE Transactions on,* Vol. 2, No. 4, **146-150** (1956).

[18] J. A. McFadden "The axis-crossing intervals of random functions-ii.", *Information Theory, IRE Transactions on,* Vol. 4, No. 1, **14-24** (1958).

[19] K. I. Hopcraft, E. Jakeman "On the joint statistics of stable random processes", *The Journal of Physics A: Mathematical and Theoretical*, Vol. 44, No. 43, **20pp** (2011).

[20] J. Panaretos, Z. Tsourti "Extreme Value Index Estimators and Smoothing Alternatives: A Critical Review", *Athens University of Economics and Business, Statistics Technical Report*, No. 149 (2001).

[21] J. Nair, A. Wierman, B. Zwart "The Fundamentals of Heavy Tails: Properties, Emergence, Identification", pre-print (2013).

[22] B. M. Hill "A Simple General Approach to Inference About the Tail of a Distribution", *The Annals of Statistics*, Vol. 3, No. 5 **1163-1174** (1975).

[23] E. Fama, R. Roll "Parameter estimates for symmetric stable distributions", *Journal of the American Statistical Association*, Vol. 66, **331-338** (1971).

[24] W. H. Du Mouchel "On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution", *The Annals of Statistics*, Vol. 1, **948-957** (1973).

[25] R.-D. Reiss, U. Cormann "An Example of Real-Life Data where the Hill Estimator is Correct", *International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo* (2006).

[26] I. A. Koutrouvelis "Regression-Type Estimation of the Parameters of Stable Laws", *Journal of the American Statistical Association*, Vol. 75, No. 372, **918-928** (1980).

[27] D. Williams *Weighing the Odds: A Course in Probability and Statistics*, Cambridge University Press (2001).

[28] R. F. Bass *Cambridge Series in Statistical and Probabilistic Mathematics: Stochastic Processes*, Cambridge University Press (2011).

[29] B. Øksendal *Stochastic Differential Equations: An Introduction with Applications* (Sixth ed.), Springer (2005).

[30] A. Feuerverge, R. A. Mureika "The Empirical Characteristic Function and Its Applications", *The Annals of Statistics*, Vol. 5, No. 1, **88-97** (1977).

[31] N. G. de Bruijn *Asymptotic Methods in Analysis*, Dover Publications (1981).

[32] M. E. Crovella "Explaining World Wide Web Traffic Self-Similarity", *Technical Report BUCS-1995-015*, Computer Science Department, Boston University, available from http://hdl.handle.net/2144/1574 (1995).

[33] B. Deheuvels, E. Haeusler, D. M. Mason "Almost sure convergence of the Hill estimator", *Math. Proc. Camb. Phil. Soc.*, Vol. 104, No. 2, **371-381** (1988).

[34] R. Reiss, U. Cormann "An Example of Real-Life Data where the Hill Estimator is Correct", *International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo* (2006).

[35] J. P. Nolan "Stable Distributions: Models for Heavy Tailed Data" (2009). http://academic2.american.edu/ jpnolan/stable/chap1.pdf

[36] M. Abramowitz, I. A. Stegun "Gamma (Factorial) Function" and "Incomplete Gamma Function", *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (9th printing)*. New York: Dover, **255-258** and **260-263** (1972).

[37] W. Feller "An Introduction to Probability Theory and its Applications", Vol. 2, 2nd edition, *New York: Wiley*, (1971).

[38] D. Saupe, H. O. Peitgen *The Science of Fractal Images*, Springer-Verlag, New York (1988).

[39] H. A. Makse, S. Havlin, M. Schwartz, H. E. Stanley "Method for generating long-range correlations for large systems", *Physical Review E*, Vol. 53, No. 5, **5445-5449** (1996).

[40] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J. D. Farmer "Testing for nonlinearity in time series: the method of surrogate data", *Physica D*, Vol. 58, **77-94** (1992).

[41] T. Schreiber, A. Schmitz "Improved Surrogate Data for Nonlinearity Tests", *Physical Review Letters*, Vol. 77, No 4, **635-638** (1996).

[42] J. F. Eichner, J. W. Kantelhardt, A. Bunde, S. Havlin "Statistics of return intervals in long-term correlated records", *Physical Review E* Vol. 75, No. 1, **011128(9)** (2007).

[43] D. Middleton, Institute of Electrical, and Electronics Engineers *An introduction to statistical communications theory* Vol. 960, McGraw-Hill New York (1960).

[44] I. B. Aban, M. M. Meerschaert "Shifted Hill's Estimator for Heavy Tails", *Commun. Statist. - Simula.* Vol. 53, No. 4, **949-962** (2001).

[45] S. Resnick, C. Starica "Smoothing the Hill Estimator", *Technical Report No. 1114, School of Operations Research and Industrial Engineering, College of Engineering, Cornell University* (1995).

[46] B. Podobnik, D. Horvatic, A. M. Petersen, H. E. Stanley "Cross-correlations between volume change and price change", *PNAS*, Vol. 106, No. 52, **22079-22084** (2009).

[47] F. Caeiro, M. I. Gomes "Threshold Selection in Extreme Value Analysis", *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman & Hall/CRC (2016).

[48] J. Beirlant, Y. Goegebeur, J. Segeers, J. L. Teugels, *Statistics of Extremes: Theory and Applications*, John Wiley & Sons (2004).

[49] B. de Sousa, G. Michailidis "A diagnostic plot for estimating the tail index of a distribution", *Journal of Computational and Graphical Statistics* Vol. 13 (2004).

[50] M. Abramowitz, I. A. Steugn "Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables", *Dover* (1964).

[51] J. P. Nolan "Financial modeling with heavy-tailed stable distributions", *WIREs Comput. Stat.* Vol. 6, **45-55** (2014).

[52] Yahoo! Finance "IBM Historical Data", Available from: *https://finance.yahoo.com/quote/IBM/history*.

[53] MATLAB, adftest "adftest", MATLAB Econometrics Toolbox Release 2018a, The Mathworks Inc., USA. *https://uk.mathworks.com/help/econ/adftest.html*.

[54] E. W. Weisstein "Normal Ratio Distribution", from MathWorld–A Wolfram Web Resource. *http://mathworld.wolfram.com/NormalRatioDistribution.html*.

[55] P. D. Lax "Change of Variables in Multiple Integrals.", *Amer. Math. Monthly* Vol. 106, **497-501** (1999).

[56] S. Psarakis, J. Panaretos "On Some Bivariate Extensions of the Folded Normal and the Folded-T Distributions", *Journal of Applied Statistical Science* Vol. 10, No. 22, **119-136** (2000).

[57] William H. Press. *Numerical recipes 3rd edition: The art of scientific computing.* Cambridge University Press (2007).

[58] MATLAB and Statistics Toolbox Release 2018a, The MathWorks, Inc., USA. *https://uk.mathworks.com/help/matlab/ref/find.html*.

[59] W. N. Mantegna "Fast, accurate algorithm for numerical sumulation of Levy stable stochastic processes", *Phys. Rev. E* Vol. 49, No. 5, **4677-4683** (1994).

[60] A. Talbot "The Accurate Numerical Inversion of Laplace Transform", *J. Inst. Maths. Applics.* Vol. 23, **97-120** (1979).

[61] J. Abate, P. P. Valko "Multi-precision Laplace transform inversion" *International Journal for Numerical Methods in Engineering* Vol. 60, No. 5-7, **979-993** (2004).

[62] B. Dingfelder, J. A. C. Weideman "An Improved Talbot Method For Numerical Laplace Transform Inversion", *Numerical Algorithms* Vol. 68, No.1, **167-182** (2015).

[63] G. Arfken "Inverse Laplace Transformation", *Mathematical Methods for Physicists, 3rd. ed.* Orlando, FL Academic Press, **853-861** (1985).

[64] J. A. C. Weideman "Optimizing Talbot's Contours For The Inversion Of The Laplace Transform", *SIAM J. Numer. Anal.* Vol. 44, No. 6, **2342-2362** (2006).

[65] T. McClure "`talbot_inversion`", *http://www.mathworks.com/matlabcentral/fileexchange/39035-numerical-inverse-laplace-transform/* (2013).

[66] B. Dingfelder, J. A. C. Weideman "`ModifiedTalbot`", *http://arxiv.org/e-print/1304.2505v3* (2014).

[67] J. Abate, P. P. Valko "`NumericalLaplaceInversion`", *http://library.wolfram.com/infocenter/MathSource/4738/* (2002).