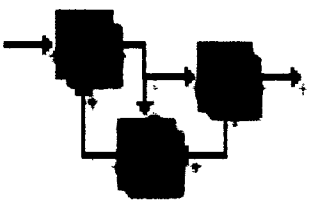
DISSERTATION

THE CODIFFERENCE FUNCTION FOR α–STABLE PROCESSES: ESTIMATION AND INFERENCE

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von



o. Univ. Prof. Dipl.-Ing. Dr. techn. Manfred Deistler

E 105
Institut für Wirtschaftsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dedi Rosadi
Matrikelnummer: 0126963
Lorenz Müllergasse 1A/5209, A1200 Wien

Wien, am 27.08.2004

This thesis is dedicated to My beloved Dad
died December 13, 2001
May God rest his soul in peace.

Contents

Abstract

In this thesis, we consider a strictly stationary univariate symmetric α-stable process $\{X_t, t \in \mathbb{Z}\}$ with $0 < \alpha \leq 2$. For Gaussian process ($\alpha = 2$), the covariance function completely describes the dependence structure over time for $\{X_t\}$. For $\alpha < 2$, the second moments of the process are infinite, and therefore the population covariance function is not defined. However, the (noncentral) sample autocovariance and the sample autocorrelation function (SACF) are well defined random variables. Davis and Resnick (1986) showed that for a class of linear models with infinite variance, the SACF will converge, and when it is properly normalized, converges weakly to a limiting distribution. Unfortunately, as a tool for modelling, the SACF of linear time series models with infinite variance has some drawbacks (see Section 1.2).

Due to the drawbacks of the SACF, for processes having no population second moments, it is desirable to have a dependence measure which does not depend on the existence of (second) moments. For this purpose, some generalizations of the autocovariance function as dependence measures of stationary process with infinite variance have been proposed in literature, e.g., the *autocovariation* (Cambanis and Miller, 1981), the *codifference* function (Kokoszka and Taqqu, 1994) and the *dynamical function* (Janicki and Weron, 1994*b*). In this thesis we study the codifference function and also consider the normalized codifference function of causal stable ARMA process with symmetric α-stable noise, $0 < \alpha \leq 2$. We consider estimators of the codifference and the normalized codifference function based on the empirical characteristic function. We further show the asymptotic properties of the proposed estimators. Finally, we discuss the application of the codifference function for identifying and estimating moving average models and doing Portmanteau-type test.

This thesis is organized as follows. In the first part of this thesis, comprising Chapters 1-3, we present some important concepts, which will be required for further understanding of the thesis. In the introductory chapter, the background and motivations to the problem that we consider are presented. Chapter 2 gives the reader an overview of stable distributions. Chapter 3 provides an overview of results related to the α-stable ARMA modelling.

The second part, consisting of Chapters 4 - 6, deals with the main results. Chapter 4 treats the codifference function. Here we provide the definition of the function, and give a summary of the asymptotic properties of the codifference function for causal stable ARMA case. Furthermore, based on the empirical characteristic function, we propose estimators of the codifference and the normalized codifference function. We show the consistency of the proposed estimators, where the underlying model is causal stable ARMA with symmetric α-stable noise, $0 < \alpha \leq 2$. In addition, we establish their limiting distributions.

Chapter 5 deals with the application of the results presented in Chapter 4. In the first section of this chapter, we address some practical issues for the calculation of the sample codifference and the normalized codifference function from time series data with finite size. In section two, we describe an order identification and estimation method for moving average models which uses the codifference function. In the last section, we discuss the application of the codifference function for Portmanteau-type test, i.e., testing for independence against serial dependence alternatives.

In the last chapter, we provide a brief summary, draw conclusions from the results presented in this thesis and outline possible further research directions.

Kurzfassung

In dieser Dissertation betrachten wir einen strikt stationären univariaten symmetrische α-stabilen Prozeß $\{X_t, t \in \mathbb{Z}\}$, wobei $0 < \alpha \leq 2$. Für einen Gaußschen Prozeß ($\alpha = 2$) beschreibt die Kovarianzfunktion in eineindeutiger Weise die Abhängigkeitsstruktur von $\{X_t\}$ über die Zeit hinweg. Für $\alpha < 2$ existieren die zweiten Momente des Prozesses hingegen nicht, und daher ist auch die Kovarianzfunktion nicht definiert. Die (nichtzentrierte) Sample-Autokovarianzfunktion und die Sample-Autokorrelationsfunktion (ACF) sind hingegen wohldefinierte Zufallsvariablen. Davis and Resnick (1986) haben gezeigt, daß für eine Klasse von linearen Modellen mit unbeschränkter Varianz die Sample-ACF konvergiert, und daß, bei geeigneter Normalisierung, schwache Konvergenz gegen eine Grenzverteilung vorliegt. Leider hat die Sample-ACF als Modellierung-Tool bei linearen Zeitreihenmodellen mit unbeschränkter Varianz einige Nachteile (siehe Abschnitt 1.2).

Wegen der oben angesprochenen Nachteile der Sample-ACF bei Prozessen mit unbeschränkten zweiten Momenten ist es wünschenswert ein Abhängigkeitsmaß zu betrachten, welches nicht von der Existenz (zweiter) Momente abhängt. Zu diesem Zweck wurden in der Literatur einige Verallgemeinerungen der Autokovarianzfunktion als Abhängigkeitsmaß für stationäre Prozesse mit unbeschränkter Varianz vorgeschlagen. Beispiele hierfür sind die *Autokovariation* (Cambanis and Miller, 1981), die *Kodifferenzfunktion* (Kokoszka and Taqqu, 1994) und die *dynamische Funktion* (Janicki and Weron, 1994*b*). In dieser Dissertation betrachten wir die Kodifferenzfunktion und die normalisierte Kodifferenzfunktion eines kausalen stabilen ARMA Prozeß mit symmetrischem α-stabilen Rauschen, $0 < \alpha \leq 2$. Die Schätzer, welche wir für diese Funktionen definieren, basieren auf der empirischen charakteristischen Funktion. Wir zeigen die asymptotischen Eigenschaften der Schätzer und diskutieren die Anwendung der Kodifferenzfunktion bei statistischen Tests, die dem Portmanteau Test sehr ähnlich sind und bei der Identifizierung und Schätzung von MA-Modellen.

Die Dissertation ist wie folgt gegliedert: Im ersten Teil, welcher aus den Kapiteln 1 - 3 besteht, präsentieren wir wichtige Konzepte, die für das Verständnis der Arbeit vonnöten sind. In der Einleitung wird der Hintergrund und die Motivation des betrachteten Problems beleuchtet. In Kapitel 2 erhält der Leser eine Übersicht über stabile Verteilungen. Kapitel 3 bietet eine Übersicht über bestehende Resultate zur α-stabilen ARMA-Modellierung.

Der zweite Teil der Arbeit besteht aus den Kapiteln 4 - 6 und beinhaltet die Hauptresultate. Kapitel 4 beschäftigt sich mit der Kodifferenzfunktion. Wir geben eine Definition der besagten Funktion und eine Zusammenfassung der asymptotischen Eigenschaften der Kodifferenzfunktion im kausalen stabilen ARMA Prozeß. Weiters schlagen wir Schätzer der Kodifferenz- und normalisierten Kodifferenzfunktion vor, die auf der empirischen charakteristischen Funktion basieren. Wir zeigen die Konsistenz der Schätzer, wobei das zugrundeliegende Modell ein kausalen stabilen ARMA-Modell mit symmetrischem α-stabilen Rauschen ist und $0 < \alpha \leq 2$. Zusätzlich leiten wir die Grenzverteilungen ab.

Kapitel 5 behandelt Anwendungen der Resultate aus Kapitel 4. Im ersten Abschnitt dieses Kapitels behandeln wir die praktische Berechnung der Sample-Kodifferenzfunktion und der normalisierten Sample-Kodifferenzfunktion aus endlich vielen Zeitreihendaten. In den folgenden Abschnitten beschreiben wir eine Methode zur Ordnungs- und Parameterschätzung für MA-Modelle, welche die Kodifferenzfunktion verwendet. Im letzten Abschnitt diskutieren wir die Anwendung

der Kodifferenzfunktion für Tests, die dem Portmanteau-Test ähnlich sind, d.h., wir testen Unabhängigkeit gegen serielle Abhängigkeit.

Das letzte Kapitel enthält eine kurze Zusammenfassung sowie Schlußfolgerungen aus den erzielten Resultaten. Einige mögliche Richtungen für weitere Forschungsaktivitäten werden abschließend genannt.

Acknowledgement

Here, I would like to thank all the people who supported me while I have been working on my thesis. First of all, I would like to express my deepest gratitude to my supervisor, Professor Manfred Deistler, for his academic guidance and all knowledge that he has given to me as well as giving me a complete liberty in what I liked to do, through all frustrating moments until the very end. Without his help, I would never have been able to accomplish this thesis.

I also would like to thank to Professor Kurt Hornik, from Computational Statistics group of the Department of Statistics and Mathematics, Vienna University of Economics and Business Administration, as the discussant and the second examiner in my thesis defence.

I gratefully acknowledge all colleagues in the group of Econometrics and System Theory for all kinds of help and creating a good working atmosphere in the office. Especially, thanks to Dietmar Bauer, Thomas Ribarits, and Eva Hamann, who have kindly shared the room during my stay in the institute. Thanks also to Wolfgang Scherrer and Emmanuel Gabreyohannes for any kind of assistance. Also many thanks to my Indonesian friends, some of them are: Joko, Dewi, Siswo, Khabib; they spent a lot of fun time with me in the past three years.

I acknowledge the Government of Austria, via ÖAD TSA project for the financial support during my study. In this regard, I want to acknowledge scholarship committee interviewer for ÖAD TSA project, who selected me as one of the scholarship holders for the project during periods 2001-2004. Furthermore, since this work gets a lot of benefits from the availability of R/GNU-S software, I gratefully acknowledge R-Development core team who has made R become an open source project.

Finally, I am indebted to my parents, brother and sisters for all support during my entire studies. Especially, I am deeply indebted to my lovely wife Heni and my wonderful daughter Putri, for their patience and love. They have supported me to overcome difficulties and nervousness throughout my Ph.D. study. Without them this thesis would not have been completed.

Chapter 1

Introduction

1.1 Motivation and background

In many cases, the assumption of normality of the observations seems to be reasonable. On the other hand, many large empirical data sets from diverse fields of studies, for instance from telecommunications and network traffics (e.g., Resnick, 1997), physics (see, e.g., Janicki and Weron, 1994*a*, section 4), finance (e.g., Fama, 1965; Mandelbrot, 1963; Rachev and Mittnik, 2000; Mittnik *et al.*, 1998), signal processing (e.g., Nikias and Shao, 1995) and others (e.g., Zolotarev, 1986, Chapter 1), are found to be leptokurtic (i.e., heavy-tailed and peaked around the center). It means that data with large value relative to the sample median will occur more frequently and can have a large spread compared to what the normal distribution can explain. Some distributions have been considered to replace the normal distribution in this case, e.g., the student-t distribution, the two-sided Weibull, the hyperbolic and the stable Paretian (stable non-Gaussian) distribution.

Among distributions that we mentioned above, an important and attractive class of distributions for modelling heavy-tailed data is the stable distribution (e.g., Rachev and Mittnik, 2000). In this respect, a crucial question arises: *Even if we acknowledge that large data sets are heavy-tailed, is it ever reasonable to use a stable model?*.

It has been argued that the class of stable distributions have many advantages compared to the numerous alternatives:

1. It contains the normal distribution as a special case, although the statistical analysis based on stable Paretian distribution in many cases is completely different from the Gaussian model, due to the nonexistence of the (second) moments and, except for a few cases, the nonexistence of the probability distribution function in a closed-form.

2. The importance of the stable distributions is theoretically supported by the *generalized central limit theorem*, which indicates that the stable distributions are the only possible non-trivial limiting distributions for the normed sum of independent and identically distributed random variables, either in the finite or in the infinite variance case.

3. In some applications, there are solid theoretical reasons for using non normal α-stable models, for instance (Feller, 1971; Zolotarev, 1986)

   - reflection of the radioactive emission on the screen yields the Cauchy distribution
   - hitting time for a Brownian motion yields the Levy distribution, and
   - the gravitational field of stars yields the Holtsmark distribution.

On the other hand, there is skepticism concerning the usage of the stable distributions in statistical modelling. For instance, from econometrics community, Campbell *et al.* (1997), p.19 wrote

> Although the stable distributions were popular in the 1960's and early 1970's, they are less commonly used today. They have fallen out of favor partly because they make modelling so difficult; standard finance theory almost always requires finite second moments of returns, and often finite higher moments as well. The stable distributions also have some counterfactual implications. First they imply that sample estimates of the variance and higher moments of returns will tend to increase as the sample size increases, whereas in practice these estimates seem to converge. Second, they imply that long horizon returns will be just as non-normal as short horizon returns since long horizon returns are sums of short horizon returns, and these distributions are stable under addition. In practice, the evidence for non-normality is much weaker for long horizon returns than for short horizon returns.

In particular, Campbell *et al.* (1997) presented an empirical kurtosis analysis of the daily and the monthly returns data to support their conclusion that the stable hypothesis should be rejected.

In our opinion, the arguments of Campbell *et al.* (1997) above are only partially correct and the analysis which is provided to support their conclusion has several weaknesses. In what follows, we discuss this in detail. Although it is not our main interest to argue against Campbell *et al.* (1997) or a similar opinion, we believe that it is important to discuss this here, to motivate the application of the stable distribution.

The first argument given in Campbell *et al.* (1997) seems to be one of the main reasons often used by practitioners against using stable models. In particular, it is often argued that since the stable non-Gaussian distribution has infinite variance, it is inappropriate for real data that have bounded range. However, note that it is a common procedure that we model (all) bounded data by some distributions which have infinite support (e.g., the popular normal distribution). As argued in Fofack and Nolan (1999), the reason behind this fitting procedure is mainly because the chosen distribution is able to describe the *shape* of the data, although it is not appropriate in the tails since the data always have bounded range [1]. Thus, the same justification can be used for stable non-Gaussian models. If the stable distribution can describe the shape and the other underlying properties of the data accurately and parsimoniously, then we may accept it as an alternative model for the data, although it (population) variance is infinite [2] [3]. Regarding the nonexistence of the variance of stable models, we refer to an argument given in Fofack and Nolan (1999)

> The variance is one measure of spread. The scale σ in a stable model is another measure can be used. Perhaps practitioners are so used to using the variance as the measure of spread. That they automatically retreat from a model without a variance[4]. The scale parameter σ can play a similar role for stable models[5]. Of course, all four parameters are necessary to completely describe the distribution.

Furthermore, to obtain a correct picture about the behavior of moments estimator of the process with infinite variance, we should deeply explore the next argument of Campbell *et al.* (1997). We recall it here

[1] In some sense, it has a similar interpretation to the well-known saying "All models are wrong, but some are useful"

[2] The modeler should remember the difference between the sample variance which is always finite for bounded data, and the population variance which can be infinite.

[3] Alternatively, one could also consider a new class of distributions as proposed in Menn and Rachev (2004), which has the "stability" property of the stable distributions and on the other side, has finite variance.

[4] In Gaussian case, the distribution can be well described by only two parameters, the mean and the variance, but for most of distributions, it does not

[5] For instance, the mean-variance analysis in finance theory can be replaced by mean-scale analysis ($1 < \alpha \leq 2$) or location-scale analysis ($0 < \alpha \leq 1$), see, e.g., Belkacem *et al.* (2000), Doganoglu and Mittnik (2003)

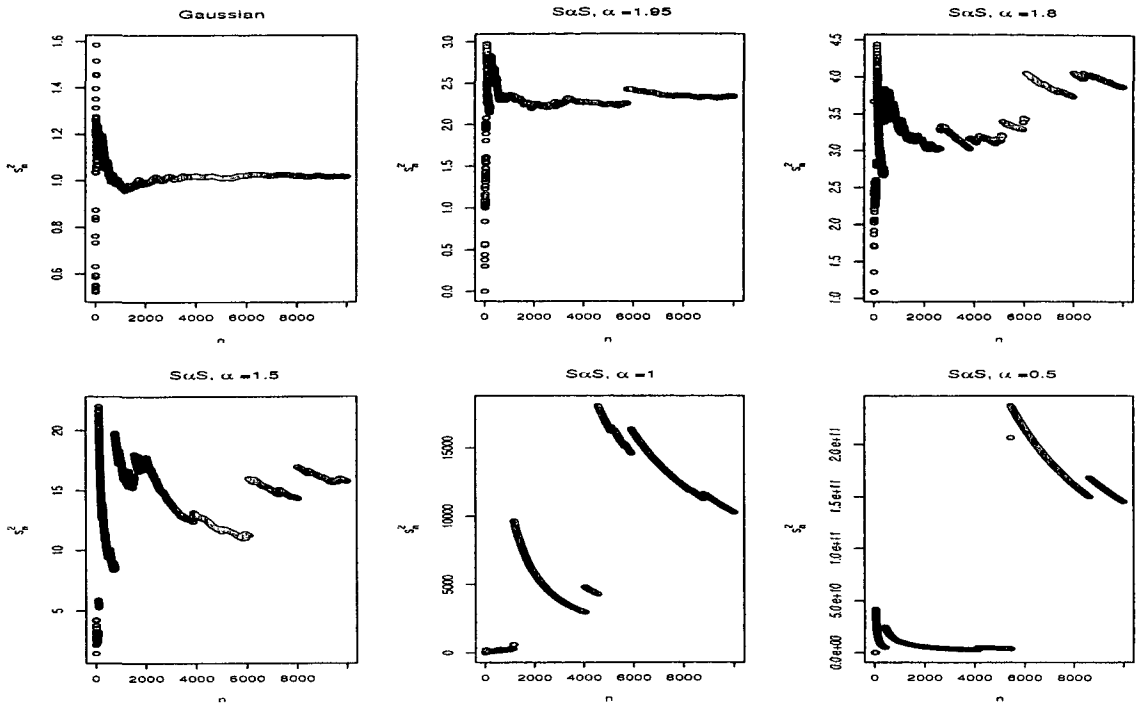


Figure 1.1: Cumulative sample variance plots of the simulated SαS series

> ...First they imply that sample estimates of the variance and higher moments of returns will tend to increase as the sample size increases, whereas in practice these estimates seem to converge ...

To investigate the claim of Campbell *et al.* (1997) above, let us consider the sample variance of the process, i.e., we consider the so-called "converging variance" test[6] (Granger and Orr, 1972). We plot the sample variance s_n^2 based on the first n observations, i.e., $s_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, with $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$, against n. If all X_t's are coming from the same distribution, when the population variance of X_t is finite, the sample variance should converge to a finite value for $n \to \infty$. Otherwise, the sample variance s_n^2 will diverge, and for finite sample size, typically there is a clear pattern of large jumps and irregularity in the graph. As an illustration for this fact, in Figure 1.1 we plot the sample variance of simulated standard α-stable (SαS) data for some index of stability α and for the spread measure $\sigma = 1$ against the sample size n. In this figure, when $\alpha < 2$, we can easily see the existence of large jumps when n is increased, however if the spread measure σ is small, the range of its values is not outrageous, especially for $1.5 < \alpha \leq 2$[7]. This simple counterexample shows that for moderate sample sizes, the argument of Campbell *et al.* (1997) above can be quite misleading.

Herewith we shortly discuss the stability testing methods for empirical data. In Campbell *et al.* (1997), it was argued that if the data were stable then the data of long horizon returns would be just as non-normal as short horizon returns since long horizon returns are sum of short horizon returns,

[6]Note that the "converging variance" test was originally designed for i.i.d. data, but in the presence of dependency in the data, Adler *et al.* (1998*a*) proposed to randomize the order of the series before applying the test.

[7]Indeed, we believe in many empirical applications, the data have finite mean. Therefore, in this respect, the range of $1 < \alpha \leq 2$ is empirically more important. Furthermore, in many empirical applications, it was often reported that the value of α is in the range of $[1.5, 2]$, see, e.g., Adler *et al.* (1998*b*), Rachev and Mittnik (2000), Weron (2001)

and their distributions are "stable" under addition. However, to support their conclusion, they provide an empirical example under which the long horizon returns have smaller sample kurtosis compared to the short horizon returns. Therefore, the stable hypothesis should be rejected, because the data do not show "stability". In our opinion, their argument does not provide a convincing approach for concluding non-"stability" of the data. Regarding this issue, we refer to Rachev and Mittnik (2000), p. 4, which note

> Kurtosis statistic reflects the peakedness of the center compared to that of the normal distribution, so that a value near three would be indicative of normality. Although formal tests could in principle be conducted, it should be kept in mind that under the Paretian stable hypothesis, second and higher moments do not exist, rendering such tests useless.

To check the stable (univariate) Paretian assumption (that is, α-stable distributed with index of stability $\alpha < 2$), in Section 2.3 we list several methods that have been suggested in literature, and notice that none of the proposed methods uses the sample kurtosis as the basis for concluding the summability of the data (i.e., the probability distribution of data is preserved under convolution). As can be found in literature, kurtosis statistic (and the sample kurtosis) higher than 3 is an indication that data are heavier than the normal distribution, and nothing can be said about the effect of summation of kurtosis (and sample kurtosis) when the moments of order higher than or equal to two is infinite (see a related discussion in Schmid and M.Trede, 2003).

It seems that the most prominent method for checking the summability of the data is done via the estimates of the index of stability α. Instead of checking the "stability" of the sample kurtosis, hence one can check the stability of $\hat{\alpha}$ for the data over different time horizons (e.g., over daily, weekly and monthly horizons) (e.g., Paolella, 2001). This verification can confirm that the (necessarily i.i.d.) data could have been generated from a stable law, because this property truly characterizes the stability of the sum of any finite number of i.i.d. stable random variables. If the true distribution of the data is a stable law, the estimated tail indexes of summed nonoverlapping j- length segments of the data can be expected roughly constant as a function of $j, j = 1, 2, \ldots$, while for non stable data, the estimates should tend to increase towards two as j increases.

Note that the stability test mentioned above is naturally a joint test of summability and i.i.d.-ness. As noted in Paolella (2001), the summability test could well reject the stable hypothesis if the data were generated from non-i.i.d. model driven by the stable Paretian innovations[8]. Moreover, this approach will work only if we can assume that the index α is non-dynamic over time horizon, since the invariability property of α under summation does not hold if we sum two α-stable distribution with different index α [9].

Let us further note that in the past, many empirical results have been reported to serve as evidences against the usage of the stable distributions. In some instances, the arguments provided are based on the inference of tail index α, obtained from Hill's estimator (see Section 2.3). It was argued that in certain stocks returns and foreign exchanges, the estimated indexes lie over two, and therefore the stable Paretian distribution is not an appropriate class of distributions for modelling the data (e.g., Gopikrishnan *et al.*, 1998; Gopikrishnan *et al.*, 1999; Plerou *et al.*, 1999). Regarding this matter, recently Paolella (2001) notices [10]

> Many authors use the popular Hill estimator to conduct inference on the tail index and, as in most studies, are uncertain as to the choice of the required cutoff value, k, which indicates at what threshold tail behavior begins. This is a serious problem as the choice of k can dramatically influence the estimate of the tail index α. Many suggestions for k have been proposed in literature, and are usually taken to be between

[8] In this respect, one can apply the considered stability testing to the fitted residual, after filtering data using appropriate non-i.i.d. models

[9] Note that the sum of two α-stable random variables with different index α is not α-stable (Weron, 2001)

[10] A similar argument can be found in Weron (2001)

one and ten percent of the sample size, T. We draw attention to the fact that with stable Paretian data, the optimal k is often around 40% of N, i.e., quite far from the tail, while use of k less than this value severely overestimates α.

Moreover, as noted by Rachev and Mittnik (2000), the analyses performed in many cases are focused merely on i.i.d. case, excluding the richer class of models driven by stable processes (as the innovation processes in the models). Thus, besides the heavy-tailedness, other common stylized facts often observed in the empirical data, for instance, the temporal dependence, the short or long range dependence, etc., can not be modelled by the considered model. Note that in this dissertation, we restrict ourselves to the study of the classical model of temporal dependence, i.e., the class of (univariate) linear homoscedastic models with heavy-tailed setting, in particular the Autoregressive Moving Average (ARMA) with symmetric α-stable distributed noise (see also Section 3.1).

Although the stable distributions provide an attractive class of distributions for data modelling, unfortunately it suffers from several mathematical difficulties, see Section 1.2. In particular here we focus on the consequence of nonexistence of second moments when $\alpha < 2$, and therefore, the population covariance function of stable stationary processes does not exist. However, the sample autocovariance and the sample autocorrelation function (SACF) are well defined random variables. Furthermore, for a class of linear models with infinite variance, the SACF will converge, and if properly normalized, it will have a limit distribution (Davis and Resnick, 1986). However, as a tool for modelling, the SACF of linear time series models with infinite variance has some drawbacks (see Section 1.2).

Due to the drawbacks of the SACF in the heavy-tailed setting, for processes having no population second moments, it is desirable to have a dependence measure which does not depend on the existence of (second) moments of the process. For this purpose, some generalizations of the autocovariance function as dependence measure of stationary processes with infinite variance have been proposed in literature, e.g., the *autocovariation* (Cambanis and Miller, 1981), the *codifference* function (Kokoszka and Taqqu, 1994) and the *dynamical function* (Janicki and Weron, 1994*b*). The dynamical function has a similar form to the codifference function and both of them do not require the existence of moments of any order. The concept of autocovariation is introduced in the probability literature (e.g., Samorodnitsky and Taqqu, 1994, Ch. 2) as a dependence measure of symmetric α-stable process with finite first moments (i.e., it is defined only for $1 < \alpha \leq 2$). Some properties of dependence measures mentioned above are summarized in Samorodnitsky and Taqqu (1994) and Janicki and Weron (1994*b*). Gallagher (1998) generalized the definition of the autocovariation function to be valid for all stationary processes with finite first moments. Furthermore, Gallagher (1998) defined the sample autocovariation function and used it as an empirical dependence measure of some ARMA models, to replace the role played by the SACF. In the same spirit as Gallagher (1998), in this dissertation we study the codifference function and also consider the normalized codifference function as generalization of the autocovariance and the autocorrelation function, respectively. We define estimators for the codifference and the normalized codifference function via the empirical characteristic function. In addition, we show the asymptotic properties of the proposed estimators for a class of symmetric α-stable linear models. The application of the codifference function for inference is also discussed.

1.2 Some problems in heavy-tail modelling

In the following, we list several problems of statistical modelling under the stable distribution framework.

1. Nonexistence of probability density function (*pdf*) in an explicit form except for a few cases. This problem has been hindering the usability of prominent estimation methods which rely on the existence of *pdf*, such as the Maximum Likelihood Estimation (MLE) procedure,

and was probably one of the major difficulties for the practical application of the stable distribution in the past. In this dissertation, we are not pursuing results in this direction, however we refer the interested reader to the results, which are reviewed in Rachev and Mittnik (2000), Section 3.7. and Calder and Davis (1998). Some recent results have been reported for the numerical approximation of the stable *pdf* (and likelihood function), e.g., for i.i.d. case in Doganoglu and Mittnik (1998), Mittnik *et al.* (1999) and for ARMA estimation procedure in Calder and Davis (1998). Those results are of practical importance; because of the availability of numerical routines for the approximation of stable Paretian densities, the MLE becomes numerically feasible. Therefore, a quick and accurate estimation procedure can be performed for the various statistical models, not only for i.i.d. model, but also for richer classes of models, driven by the stably distributed process.

2. Nonexistence of the second moments for α-stable random variables with $\alpha < 2$ implying that the population covariance and the population correlation function do not exist. However, as we already mentioned, the sample covariance and the sample correlation function are well defined random variables. For cases when the population covariance function does not exist, for strictly stationary process of the form

$$X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \sum_{j=0}^{\infty} |j|\,|c_j|^{\delta} < \infty \text{ for some } \delta \in (0,\alpha) \cap (0,1]$$

where $\{\epsilon_t\}$ is i.i.d. sequence of standard symmetric α stable (SαS) random variables, α denotes the index of stability, $0 < \alpha < 2$, with the identical characteristic function

$$E(\exp is\epsilon_1) = \exp(-\sigma^{\alpha}|s|^{\alpha}), s \in \mathbb{R}, \sigma > 0$$

Davis and Resnick (1986) show the following: the noncentral SACF at lag k, namely

$$\hat{\rho}(k) = \sum_{j=1}^{n-k} X_j X_{j+k} / \sum_{j=1}^{n} X_j^2$$

converges in probability to

$$\tilde{\rho}(k) = \sum_{j=0}^{\infty} c_j c_{j+k} / \sum_{j=0}^{\infty} c_j^2$$

Note that when $\alpha < 2$, $\tilde{\rho}(k)$ can not be interpreted as the usual population correlation function $\rho(k)$. Additionally, for each $m \geq 1$, the following central limit theorem holds

$$(n/\ln(n))^{1/\alpha}(\hat{\rho}(1) - \tilde{\rho}(1), \cdots, \hat{\rho}(m) - \tilde{\rho}(m))' \xrightarrow{d} (Y_1, \cdots, Y_m)'$$

Here $\xrightarrow{d}$ denotes convergence in distribution, and

$$Y_k = \sum_{j=1}^{\infty} [\tilde{\rho}(k+j) + \tilde{\rho}(k-j) - 2\tilde{\rho}(k)\tilde{\rho}(j)] \frac{G_j}{G_0}, k = 1, \ldots, m$$

where, $(G_j)_{j\geq 1}$ and G_0 are independent random variables, where G_0 is positive $\alpha/2$-stable with characteristic function

$$E\exp\{isG_0\} = \exp\{-\Gamma(1-\alpha/2)\cos(\pi\alpha/4)|s|^{\alpha/2}(1 - i\ sign(s)\tan(\pi\alpha/4))\}$$

and $(G_j)_{j\geq 1}$ are i.i.d. $S\alpha S$ random variables with characteristic function

$$E\exp\{isG_j\} = \begin{cases} \exp\{-\Gamma(2-\alpha)\cos(\pi\alpha/2)|s|^{\alpha} & \text{if } \alpha \neq 1 \\ \exp\{-\pi|s|/2\} & \text{if } \alpha = 1 \end{cases}$$

Here $\Gamma(\cdot)$ denotes the gamma function. For $\alpha > 1$, the result above is still valid if $\hat{\rho}(\cdot)$ is replaced by the centralized version.

As a dependence measure of linear stationary processes, the SACF has some drawbacks when the data are heavy-tailed (Gallagher, 2000),

(a) There is a discontinuity in the quantile of the limit distribution and in the normalization constant as $\alpha \to 2$. Here, the discontinuity of the quantile of limiting distribution is coming from the fact that the divisor G_0 is large with positive probability when $\alpha < 2$, but is unity when $\alpha = 2$. As noted in Gallagher (2000), this result creates a problem in applications. For instance, suppose that the unknown true α is equal to 2, but it is estimated a little bit smaller than the true value, e.g., $\alpha = 1.98$, then the constants and the quantiles of the distribution which are used for inference will change drastically (see Adler *et al.*, 1998*a*, table 3).

(b) In several simulation studies, it has been shown that a sample size of one million is required to obtain correct empirical levels of a statistical test (Adler *et al.*, 1998*a*). These simulation results indicate that the limiting distribution Y_k of $(\tilde{\rho}(\cdot) - \rho(\cdot))$ in (3.12) provides a very poor approximate distribution for moderate sample sizes.

These and related problems are investigated in this dissertation.

1.3 Thesis contribution

The overall objective of this dissertation is to study the class of (stable causal) Autoregressive Moving Average (ARMA) with symmetric α-stable ($S\alpha S$) distributed noise. In particular, we focus on the analysis of the dependence structure of this model using a quantity, called the codifference function. The main contribution of this dissertation can be summarized in the sequel:

1. In this dissertation, we study the asymptotic behavior of the codifference function of causal stable ARMA process with $S\alpha S$ noise. We consider three classes of the roots of the respective polynomial of the autoregressive part, namely real positive, real negative and complex, i.e. we slightly extend the results presented in Kokoszka and Taqqu (1994).

   See Section 4.1 and 4.2 for detailed discussion.

2. We consider the codifference and the normalized codifference function as dependence measures for stationary processes. Based on the empirical characteristic function, we propose estimators of the codifference and the normalized codifference function. We show the consistency of the proposed estimators, where the underlying model is stable causal ARMA with symmetric α-stable noise, $0 < \alpha \leq 2$. In addition, we derive their limiting distribution. Finally, we present a simulation study showing the dependence of the estimator on certain design parameter.

   These results are presented in Section 4.3 - 4.7 and 5.1.

3. In the traditional Box-Jenkins modelling, we use the sample autocorrelation function as a tool for fitting the plausible models for empirical data. In this dissertation, we consider the sample (normalized) codifference as a new tool for the preliminary order identification and estimation of pure univariate $S\alpha S$ moving average process with finite order. The performance of the proposed method is evaluated via simulation studies, and it is compared to the performance of a method which is based on the sample covariance/the sample autocorrelation function. Simulation results indicate that the method based on the codifference function works fairly well for identification and estimation purposes.

   See Section 5.2 for detailed discussion.

4. We consider a test for independence against serial dependence alternatives for symmetric α stable random variables with the exponent $0 < \alpha \leq 2$, using Box-Pierce Q-statistic which is defined using the codifference function. We obtain that unlike a similar test proposed in Runde (1997), the asymptotic distribution of the proposed statistic is similar to the classical case, that is asymptotically χ^2 distributed, both in the finite and the infinite variance cases.

For $1 < \alpha \leq 2$, simulation studies are performed to obtain the small sample performance of the proposed statistic. We found that the proposed statistic works fairly well for "small" sample, in the sense that in the infinite variance case, its empirical levels are much closer to the theoretical ones and its power is much better than Runde's statistic, and where in the finite variance case, its empirical levels and its power are approximately the same as that of Ljung Box's statistic (Ljung and Box, 1978). However, when the order of checked lag m is relatively large compared to the sample size n, the Portmanteau statistic, which is based on the codifference function, has been shown to have much closer empirical levels than the alternatives.

These results are presented in Section 5.3.

Chapter 2

The α-stable distribution family

2.1 A short historical overview

The study of stable distributions can be dated back to the early 18th century, when the first and most widely used Stable distribution, called the normal distribution, was discovered[1]. Although it is known that Abraham de Moivre in 1733 was the first mathematician who introduced the normal distribution in the context of approximating certain binomial distributions, the normal distribution is often devoted to Carl Friedrich Gauss who in 1809 emphasized the importance of the normal distribution in solving the least-squares regression problem within the field of astronomy[2]. Following the development of the theory of characteristic functions as the Fourier transforms of probability distributions, in 1810 Laplace used the characteristic function for the representation of the probability of a sum of large numbers of mutually independent random variables and showed that it is approximately normal distributed. Later on, Laplace's student, Augustin Louis Cauchy extended Laplace's analysis and discovered Cauchy law. Cauchy discovered in 1853 that the functions f_α which satisfies

$$\int_{-\infty}^{\infty} e^{i\theta x} f_\alpha(x)dx = e^{-\sigma^\alpha|\theta|^\alpha}, \alpha > 0, \theta \in \mathbb{R} \tag{2.1}$$

have the convolution property

$$(Af_\alpha(\cdot)) * (Bf_\alpha(\cdot)) = Cf_\alpha(\cdot)$$

for some $C = C(A, B)$ and all $A, B > 0$. However, Cauchy can show that $f_\alpha(x) \geq 0$ only in the cases $\alpha = 1$ and $\alpha = 2$.

In 1924, during his study regarding the limiting law of normed sums of independent and identically distributed (i.i.d.) random variables, Paul Pierre Lévy found that when the condition on the finiteness of the variance of i.i.d. random variables is relaxed, the limit distribution follows a stable law. Motivated by this result, Lévy studied the Fourier transform of the class of stable distributions. In this direction, in particular, Lévy extended the work of Cauchy by showing the positivity of f_α in (2.1) for $0 < \alpha \leq 2$. The work of Lévy originated the modern development of the theory of the stable distributions.

Despite its long history of research, the stable non-Gaussian distributions attracted almost no attention to applied researchers until the seminal work of Mandelbrot (1963) and Fama (1965). In their papers, they showed that many stocks price changes, and fluctuations in speculative prices

[1] Although the mathematical foundation of the normal distribution in general has a different history compared to the other members of the stable distributions, we mention this fact for the sake of completeness

[2] The terminology Gaussian distribution is widely used to refer to the normal distribution

and interest rates are poorly described by Gaussian distribution, but can be well described by a non-Gaussian stable distribution. Until recently, many applied works have been published within different fields of application such as signal processing, telecommunications, finance, physics and chemistry. For an application oriented discussion, we refer the reader to the monographs by Janicki and Weron (1994*b*), Nikias and Shao (1995), Embrechts *et al.* (1997), Adler *et al.* (1998*b*) and Rachev and Mittnik (2000).

In this chapter, we present some basic facts about the stable distributions which will be necessary for further understanding of the presentation. For discussion on the mathematical foundation of the stable distributions, we refer interested readers to the monographs by Gnedenko and Kolmogorov (1954), Feller (1971), Ibragimov and Linnik (1971), Zolotarev (1986), Samorodnitsky and Taqqu (1994), Janicki and Weron (1994*b*) and Nikias and Shao (1995).

2.2 Basic facts

In what follows, we recall some basic results related to the stable law, which will be used in this thesis. The subsequent discussion will follow closely the presentation in Samorodnitsky and Taqqu (1994).

2.2.1 Univariate stable distribution

We start by presenting some equivalent definitions of the univariate stable distribution.

Definition 2.2.1. *The random variable X is said to be stable if for any positive numbers A and B, there is a positive number C and a real number D such that*

$$AX_1 + BX_2 \stackrel{d}{=} CX + D \tag{2.2}$$

where X_1 and X_2 denote independent copies of X, and "$\stackrel{d}{=}$" denotes equality in distribution

This definition explains the "stability" property: the family of stable distributions is preserved under the convolutions [3]. Moreover, the stability definition above implies that there is a number $\alpha \in (0, 2]$ called *the index of stability* such that the number C in (2.2) satisfies (Samorodnitsky and Taqqu (1994), Theorem 1.1.2)

$$C^\alpha = A^\alpha + B^\alpha$$

Another equivalent definition to Definition 2.2.1 is given below.

Definition 2.2.2. *A random variable X is said to have a stable distribution if for any $n \geq 2$, there is a positive number C_n and a real number D_n such that*

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} C_n X + D_n \tag{2.3}$$

where $X_1, X_2, \ldots, X_n$ are independent copies of X.

It can be shown by induction that Definition 2.2.1 implies Definition 2.2.2. The reverse implication is also true (Feller, 1971, section VI.1). Here, we have necessarily $C_n = n^{1/\alpha}$ (Feller, 1971, Section VI.1).

The third definition constitutes that only stable random variables have a domain of attraction. This definition relates to *generalized central limit theorems*, i.e., it states that the stable distribution is the only distribution that can be obtained as a limit of normed sums of i.i.d. random variables.

[3] Note that various "stability" definitions are available in literature, not only restricted to summability notion, see, e.g., Rachev and Mittnik (2000)

Definition 2.2.3. *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e., there is a sequence of i.i.d. random variables $\{Y_t, t = 0, \pm 1, \pm 2, \ldots\}$ and sequences $a_n \geq 0$ and $b_n \in \mathbb{R}$ such that*

$$a_n^{-1} S_n + b_n \xrightarrow{d} X \tag{2.4}$$

where $S_n = Y_1 + Y_2 + \cdots + Y_n$

Definition 2.2.2 implies definition 2.2.3 , i.e., by taking Y_i's to be independent and distributed like X. The converse is shown in Gnedenko and Kolmogorov (1954), p.162. A random variable Y_i is said to be in the domain of *normal* attraction of stable law when $a_n = n^{1/\alpha}$. In general $a_n = n^{1/\alpha} L(n)$ for some slowly varying $L(n)$ [4] as $n \to \infty$.

The class of stable distributions is often characterized by its characteristic function. In literature, several parameterizations are available for this class of distributions (see, e.g., Zolotarev, 1986; Nolan, 1999, where some parameterizations are well suited for computational purposes). However, the most often used parameterization is the following

Definition 2.2.4. *A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$ and $\mu \in \mathbb{R}$, such that its characteristic function has the following form ($\theta \in \mathbb{R}$)*

$$\Phi(\theta) = E(\exp i\theta X) = \begin{cases} \exp(i\mu\theta - \sigma^\alpha |\theta|^\alpha (1 - i\beta (sign\ \theta) \tan \frac{\pi\alpha}{2}), & if\ \alpha \neq 1 \\ \exp(i\mu\theta - \sigma^\alpha |\theta|^\alpha (1 - i\frac{2}{\pi}\beta (sign\ \theta) \ln |\theta|), & if\ \alpha = 1 \end{cases} \tag{2.5}$$

Here

$$sign\ \theta = \begin{cases} 1 & if\ \theta > 0 \\ 0 & if\ \theta = 0 \\ -1 & if\ \theta < 0 \end{cases}$$

It is not difficult to see that Definition 2.2.4 will imply Definition 2.2.2. If $X_i, i = 1, \ldots, n$ are i.i.d. $S_\alpha(\sigma, \beta, \mu)$ then (2.3) holds, with $C_n = n^{1/\alpha}$ and $D_n = \mu(n - n^{1/\alpha})$ when $\alpha \neq 1$ and $D_n = \frac{2}{\pi}\sigma\beta n \ln n$ when $\alpha = 1$. The proof for the converse is rather involving, and for detail, we refer the interested reader to Gnedenko and Kolmogorov (1954), Section 34.

We often write $X \sim S_\alpha(\sigma, \beta, \mu)$ to indicate that X has (asymmetric) α-stable distribution $S_\alpha(\sigma, \beta, \mu)$[5]. The interpretation of the parameters α, σ, β, and μ can be explained from its properties. The parameter α is *the characteristic exponent* or *the index of stability* which defines *the fatness of the tails* of distributions (where smaller values of α correspond to heavier tailed distribution). For $a \in \mathbb{R}$, $X + a$ is $S_\alpha(\sigma, \beta, \mu + a)$, thus μ is a *shifting* parameter. For $a > 0$ and $\alpha \neq 1$ then aX is $S_\alpha(|a|\sigma, \text{sign}(a)\beta, a\mu)$, and when $\alpha = 1$, aX is $S_\alpha(|a|\sigma, \text{sign}(a)\beta, a\mu - \frac{2}{\pi}a(\ln |a| \sigma\beta))$, thus σ is similar to the *scale* parameter, although the name scale parameter for σ is a misnomer when $\alpha = 1$ and $\beta \neq 0$. Notice that when $\alpha = 1$, multiplication by a constant yields a nonlinear transformation of the shifting parameter μ. The nonstandard random variable $X \sim S_\alpha(\sigma, \beta, \mu)$ can be standardized by transformation

$$Z = \frac{X - \mu}{\sigma} \sim S_\alpha(\sigma = 1, \beta, \mu = 0) \tag{2.6}$$

If $\beta = 0$, the distribution of X is symmetric about μ, thus β is *the skewness parameter.* The distribution will be left or right skewed when $\beta > 0$ or $\beta < 0$, respectively. If $\beta = 0$ and $\mu = 0$, then the distribution is called *Symmetric α Stable* ($S\alpha S$) and has log characteristic function

$$\log(\Phi(\theta)) = -\sigma^\alpha |\theta|^\alpha, \theta \in \mathbb{R} \tag{2.7}$$

A $S\alpha S$ random variable is characterized by its scale parameter σ and characteristic exponent α. The value of σ is often considered as 1, which we call the *standard $S\alpha S$.*

[4] $L(x)$ is a slowly varying function for $x \to \infty$ if for every constant $c > 0, \lim_{x \to \infty} L(cx)/L(x) = 1$

[5] When $\alpha < 2$, it is often denoted as the stable Paretian distribution

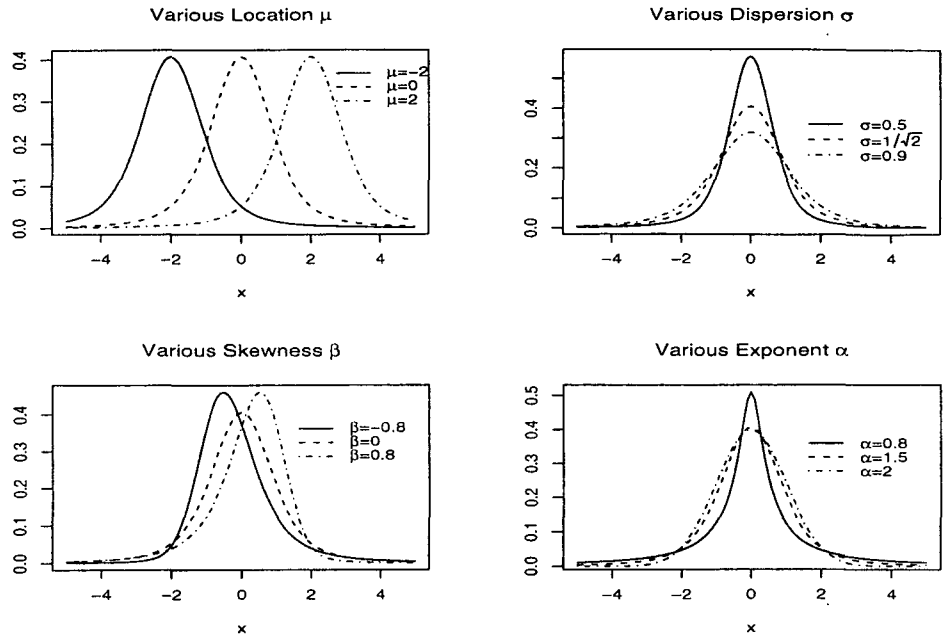


Figure 2.1: Plots of probability density function of $S_\alpha(\sigma, \beta, \mu)$ random variables for some values of the parameters. The default values of the parameters used in the plot are $\alpha = 1.5$, $\sigma = 1/\sqrt{2}$, $\beta = 0$, $\mu = 0$

As described in literature, the characteristic exponent α can be considered as the most important parameter of the stable distributions. Its importance comes from the fact that for every stable law with characteristic exponent $0 < \alpha < 2$, all absolute moments of the order $p \geq \alpha$ are infinite (see Appendix A.1). Therefore, the only stable distribution with finite variance is the Gaussian distribution.

Stable random variables have continuous probability density functions (Zolotarev, 1986), however, except for a few cases, it cannot be written in an explicit form. These exceptions are

1. $S_2(\sigma, 0, \mu)$ denotes the Gaussian distribution $N(\mu, 2\sigma^2)$, whose density is

$$(2\sigma\sqrt{\pi})^{-1} \exp((x-\mu)^2/4\sigma^2)$$

2. $S_1(\sigma, 0, \mu)$ denotes the symmetric (about μ) Cauchy distribution (it is sometimes called the Lorentz distribution), whose density is

$$f(x) = \frac{1}{\pi\sigma\left\{1 + \left(\frac{x-\mu}{\sigma}\right)^2\right\}}, x \in \mathbb{R}$$

3. $S_{0.5}(\sigma, 1, \mu)$ denotes the Lévy distribution, whose density is

$$f(x) = \frac{\sqrt{\sigma}}{\sqrt{2\pi}}(x-\mu)^{-3/2} \exp\left\{-\frac{\sigma}{2(x-\mu)}\right\} \text{ with } x > \mu$$

4. $S_\alpha(0, 0, \mu)$ has a degenerate distribution. However, usually degenerate distributions will be excluded from the analysis.

For the other choices of parameters, the density can only be generated numerically (see, e.g., Rachev and Mittnik, 2000, Section 3.7). Plots of probability density functions of stable random variables for several variations of the parameters are given in Figure 2.1.

Another important property of stable distributions is their unimodality (Yamazato, 1978). A distribution function $f(x)$ is called unimodal if there exists $a \in \mathbb{R}$, such that $f(x)$ is monotonically increasing for $x < a$ and monotonically decreasing for $x > a$. One usually says that a is the mode of $f(x)$. For numerical results about the problem of locating the mode of a stable distribution, see Fofack and Nolan (1999).

2.2.2 Multivariate stable distribution

In the following we present the definition of the multivariate stable distribution.

Definition 2.2.5. *A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is said to be a stable random vector in $\mathbb{R}^d$ [6] if for any positive numbers A and B, there is a positive number C and a vector $\mathbf{D} \in \mathbb{R}^d$ such that*

$$A\mathbf{X}^1 + B\mathbf{X}^2 \stackrel{d}{=} C\mathbf{X} + \mathbf{D} \tag{2.8}$$

where $\mathbf{X}^1$ and $\mathbf{X}^2$ are independent copies of $\mathbf{X}$.

Definition 2.2.6. *The vector $\mathbf{X}$ is called symmetric stable if it is stable and in addition satisfies the relation*

$$P\{\mathbf{X} \in A\} = P\{-\mathbf{X} \in A\}$$

for any Borel set A of $\mathbb{R}^d$.

The following properties define the relation between vector $\mathbf{X}$ and its components.

Proposition 2.2.7 (Samorodnitsky and Taqqu (1994), Theorem 2.1.2). *Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a stable (respectively, symmetric stable) d-dimensional random vector. Then there is a constant $\alpha \in (0, 2]$ such that in (2.8), $C = (A^\alpha + B^\alpha)^{1/\alpha}$. Moreover, any linear combination of the components of $\mathbf{X}$ of the type $Y = (\mathbf{b}, \mathbf{X}) = \sum_{k=1}^{d} b_k X_k$ is one dimensional α-stable (respectively, symmetric α-stable) random variable with the same index α for every $\mathbf{b} = (b_1, b_2, \ldots, b_k)$.*

Proposition 2.2.8 (Samorodnitsky and Taqqu (1994), Theorem 2.1.5). *Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$. Suppose that any linear combination $Y = (\mathbf{b}, \mathbf{X}), \mathbf{b} \in \mathbb{R}^d, \mathbf{b} \neq \mathbf{0}$ is one-dimensional stable, e.g., $(\mathbf{b}, \mathbf{X}) \sim S_{\alpha(\mathbf{b})}(\sigma(\mathbf{b}), \beta(\mathbf{b}), \mu(\mathbf{b}))$. Then there is one α that is the index of all linear combinations, i.e., $\alpha(\mathbf{b}) = \alpha$ is constant. If $\alpha \geq 1$, then $\mathbf{X}$ is a stable random vector in $\mathbb{R}^d$. If Y is symmetric stable, then $\mathbf{X}$ is a symmetric stable random vector in $\mathbb{R}^d$.*

As in the univariate case, the multivariate stable random variables are parameterized in the form of their characteristic function. Let $\mathbf{X} = (X_1, \ldots, X_d)$ be α-stable random vector in $\mathbb{R}^d$, $0 < \alpha < 2$ and let

$$\mathbf{\Phi}_\alpha(\theta) = \mathbf{\Phi}_\alpha(\theta_1, \theta_2, \ldots, \theta_d) = E\exp\{i(\theta, \mathbf{X})\} = E\exp\{i\sum\nolimits_{k=1}^{d} \theta_k X_k\}$$

denote its *joint characteristic function.*

Theorem 2.2.9 (Samorodnitsky and Taqqu (1994), Theorem 2.3.1). *Let $0 < \alpha < 2$. Then $\mathbf{X} = (X_1, \ldots, X_d)$ is an α-stable random vector in $\mathbb{R}^d$ if and only if there exist a finite measure Γ on the unit sphere $S_d = \{\mathbf{s} = \{s_1, \ldots, s_d\} : \|\mathbf{s}\| = 1\}$ of $\mathbb{R}^d$ and the vector μ^0 in $\mathbb{R}^d$ such that*

a. *If $\alpha \neq 1$*

$$\mathbf{\Phi}_\alpha(\theta) = \exp\left\{-\int_{S_d} |(\theta, \mathbf{s})|^\alpha \left(1 - i\, sign\,((\theta, \mathbf{s}))\tan\frac{\pi\alpha}{2}\right)\Gamma(d\mathbf{s}) + i(\theta, \mu^0)\right\} \tag{2.9}$$

[6] We often say "$X_1, X_2, \ldots, X_d$ are *jointly stable*", or "$\mathbf{X}$ has a stable distribution in $\mathbb{R}^d$", or "the distribution of $\mathbf{X}$ is multivariate stable" to denote the same terminology

b. *If* $\alpha = 1$

$$\Phi_\alpha(\theta) = \exp\left\{-\int_{S_d} |(\theta, \mathbf{s})| \left(1 + i\frac{2}{\pi}\, sign\,((\theta, \mathbf{s}))\ln|(\theta, \mathbf{s})|\right)\Gamma(d\mathbf{s}) + i(\theta, \mu^0)\right\} \tag{2.10}$$

The pair (Γ, μ^0) *is unique. The measure* Γ *is called the spectral measure of the* α*-stable random vector* $\mathbf{X}$.

To illustrate the relation between the measure Γ and the other parameters of stable distribution (σ, β, μ), we consider the following examples from Samorodnitsky and Taqqu (1994).

Example 2.2.10 (Samorodnitsky and Taqqu (1994), Example 2.3.3). *Let us consider the univariate case. Notice that for the univariate stable case, the sphere* S_d *only consists of two points* $\{-1, 1\}$. *Let* $\Gamma(-1)$ *and* $\Gamma(1)$ *denote the probability masses in those points, then for* $\alpha \neq 1$ *(a similar result applies for* $\alpha = 1$*), eq. (2.9) reduces to*

$$\Phi_\alpha(\theta) = \exp\left\{-|\theta|^\alpha \left(1 - i\, sign\,(\theta)\tan\frac{\pi\alpha}{2}\right)[\Gamma(1) - \Gamma(-1)] + i\theta\mu^0\right\}$$

which coincides to the eq. (2.5) with

$$\sigma = [\Gamma(1) + \Gamma(-1)]^{1/\alpha}, \quad \beta = \frac{\Gamma(1) - \Gamma(-1)}{\Gamma(1) + \Gamma(-1)}, \quad \mu = \mu^0$$

Here if the spectral measure Γ *is symmetric, i.e.,* $\Gamma(1) = \Gamma(-1)$, *then* X *is also symmetric with* $\beta = 0$.

Example 2.2.11 (Samorodnitsky and Taqqu (1994), Example 2.3.4). *Let* $\mathbf{X} = (X_1, \ldots, X_d)$ *be jointly stable random variable in* $\mathbb{R}^d$. *From Proposition 2.2.7 we know that any linear combination* $Y = (\mathbf{b}, \mathbf{X}) = \sum_{k=1}^d b_k X_k$ *is univariate stable, say*

$$(\mathbf{b}, \mathbf{X}) \sim S_\alpha(\sigma(\mathbf{b}), \beta(\mathbf{b}), \mu(\mathbf{b}))$$

Then from eq. (2.9) and (2.10), there are a finite measure Γ *on* $\mathbb{R}^d$ *and a location parameter* $\mu^0 \in \mathbb{R}^d$ *with*

$$\sigma(\mathbf{b}) = \left(\int_{S_d} |(\mathbf{b}, \mathbf{s})|^\alpha \Gamma(d\mathbf{s})\right)^{1/\alpha} \tag{2.11}$$

$$\beta(\mathbf{b}) = \frac{\int_{S_d} |(\mathbf{b}, \mathbf{s})|^\alpha sign(\mathbf{b}, \mathbf{s})\Gamma(d\mathbf{s})}{\int_{S_d} |(\mathbf{b}, \mathbf{s})|^\alpha \Gamma(d\mathbf{s})} \tag{2.12}$$

$$\mu(\mathbf{b}) = \begin{cases} (\mathbf{b}, \mu^0) & \text{if } \alpha \neq 1 \\ (\mathbf{b}, \mu^0) - \frac{2}{\pi}\int_{S_d} (\mathbf{b}, \mathbf{s}) \ln|(\mathbf{b}, \mathbf{s})|\Gamma(d\mathbf{s}) & \text{if } \alpha = 1 \end{cases} \tag{2.13}$$

Example 2.2.12 (Samorodnitsky and Taqqu (1994), Example 2.3.5). *An* α*-stable random vector* $\mathbf{X} = (X_1, \ldots, X_d)$ *has independent components if and only if its spectral measure* Γ *is discrete and concentrated on the intersection of the axes with the sphere* S_d. *For example, suppose* $d = 2$, *and consider independent random variables* $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i), i = 1, 2$. *Then* Γ *concentrates on the points* $(1, 0), (-1, 0), (0, 1), (0, -1)$ *that is*

$$\Gamma = a_1\delta((1, 0)) + a_2\delta((-1, 0)) + a_3\delta((0, 1)) + a_4\delta((0, -1))$$

where $\delta(\mathbf{s}_0)$ *assigns unit mass to the point* $\mathbf{s}_0$, *and* $a_i, i = 1, \ldots, 4$ *are given as*

$$a_1 = \sigma_1^\alpha \frac{1 + \beta_1}{2}, a_2 = \sigma_1^\alpha \frac{1 - \beta_1}{2}, a_3 = \sigma_2^\alpha \frac{1 + \beta_2}{2}, a_4 = \sigma_2^\alpha \frac{1 - \beta_2}{2}$$

An α-stable random vector in $\mathbb{R}^d$ is symmetric if and only if $\mu^0 = 0$ and Γ is a symmetric measure on S_d, i.e., $\Gamma(A) = \Gamma(-A)$ for any Borel set A of S_d. From Theorem 2.2.9, we obtain the following.

Theorem 2.2.13 (Samorodnitsky and Taqqu (1994), Theorem 2.4.3). *$\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is a $S\alpha S$ random vector in $\mathbb{R}^d$ with $0 < \alpha < 2$ if and only if there exists a unique symmetric positive finite measure Γ on the unit sphere S_d of $\mathbb{R}^d$ such that*

$$\Phi(\theta) = E\exp(i(\theta, \mathbf{X})) = E\exp(i\sum_{k=1}^{d}\theta_k X_k) = \exp\left\{-\int_{S_d}|(\theta, \mathrm{s})|^\alpha\,\Gamma(d\mathrm{s})\right\}$$

Γ is the spectral measure of the symmetric α-stable random vector $\mathbf{X}$.

Example 2.2.14. *If $d = 1$, we obtain $S_1 = \{-1, 1\}$, $\Gamma(\{1\}) = \Gamma(\{-1\})$, and a $S\alpha S$ random variable X is $S_\alpha(\sigma, 0, 0)$ distributed with*

$$\sigma = (\int_{S_1}|s|^\alpha\Gamma(ds))^{1/\alpha} = (2\Gamma(\{1\}))^{1/\alpha}$$

Example 2.2.15. *Let $\mathbf{X} = (X_1, X_2)$ be a symmetric α-stable random vector with independent components. Then from Example 2.2.12, we obtain that Γ_{X_1,X_2} is symmetric and concentrates on the points $(1,0), (-1,0), (0,1), (0,-1)$,*

$$\Gamma = a_1\delta((1,0)) + a_2\delta((-1,0)) + a_3\delta((0,1)) + a_4\delta((0,-1))$$

where $a_i, i = 1, \ldots, 4$ are given as

$$a_1 = a_2 = \frac{\sigma_1^\alpha}{2}, a_3 = a_4 = \frac{\sigma_2^\alpha}{2}$$

2.2.3 Stable stochastic processes

In this part, we present the definition of stable stochastic process $\{X_t, t \in T\}$ where T is an arbitrary set, although in this thesis we are only interested in $T = \mathbb{Z}$, that is the discrete time index.

Definition 2.2.16. *A stochastic process $\{X_t, t \in T\}$ is stable if all its finite dimensional distributions are stable. It is symmetric stable if all its finite-dimensional distributions are symmetric stable.*

The finite-dimensional distributions $\{X_t, t \in T\}$ are the distributions of the vectors

$$(X(t_1), X(t_2), \ldots, X(t_d)), t_1, t_2, \ldots, t_d \in T, d \geq 1$$

If the finite-dimensional distributions are stable then, by consistency they must all have the same index of stability α. Using Property 2.2.7, we obtain the following properties of the stable process.

Proposition 2.2.17 (Samorodnitsky and Taqqu (1994), Theorem 3.1.2). *Let $\{X_t, t \in T\}$ be a stochastic process*

(a) *$\{X_t, t \in T\}$ is symmetric stable if and only if all linear combinations*

$$\sum_{k=1}^{d} b_k X(t_k), t_1, t_2, \ldots, t_d \in T, b_1, b_2, \ldots, b_d \in \mathbb{R}, d \geq 1 \tag{2.14}$$

are symmetric stable

(b) *If $\alpha \geq 1$, then $\{X_t, t \in T\}$ is α-stable if and only if all linear combinations (2.14) are α-stable.*

The notion of (strictly) stationary α-stable process is often mentioned in this thesis. For the univariate case, we have the following definition.

Definition 2.2.18. *A stochastic process $\{X_t, t \in \mathbb{Z}\}$ is called a strictly stationary α-stable process if $\{X_t\}$ is stable and for any $d \geq 1$, $t_1, \ldots, t_d \in \mathbb{Z}$, the finite-dimensional distribution of the vector $(X_{t_1+h}, \ldots, X_{t_d+h})$ does not depend on the choice of $h \in \mathbb{Z}$*

2.3 Fitting data to the stable distributions

It has been shown in literature that many large empirical data sets from diverse fields of study are more leptokurtic than the normal distribution. As we already mentioned, a flexible class of distributions for modelling heavy-tailed data is a class of stable distributions. In this regard, when we model the time series data, one important question will arise: *how can we see from a given (univariate) time series data, whether or not it is generated from an α-stable time series process?.*

In literature, the methods that have been used for checking the plausibility of the (univariate) stable distributions [7] can be classified into two main approaches, the heuristic methods (e.g., by visual inspection via some graphical methods) and the inference procedures. Several graphical methods have been proposed in literature, for instance, by graphing the series, plotting the histogram and using the "converging variance" method (Adler *et al.*, 1998*a*; Rosenfeld, 1976); using kernel density plot and normal probability plot (i.e., pp and qq plots; see Nolan, 1999) and using the plot of ratios between the maximum and partial sum (Embrechts *et al.*, 1997). A popular method for detecting the heavy-tailedness of the data is by an inspection of the estimator $\hat{\alpha}$ of index α [8]. Here, it is important to note that making a statement about the tail is indeed quite different from making a statement about the entire distribution. The estimator of α only measures the "thickness" of the tail, but because most of the heavy-tailed distributions are not stable, we can not make a general statement regarding data. However, when $\hat{\alpha}$ is significantly greater than 2, we can discard the possibility of using the stable distributions. A test for stability based on $\hat{\alpha}$ over different time horizons is described in Paolella (2001).

The methods for estimating α consist of the graphical and the nongraphical methods. The statistical properties of the nongraphical methods have been studied rather extensively (see Section 2.4), and many methods will give consistent estimates for the true parameters. Using this consistency property, as a preliminary checking for the stability assumption, Nolan (1999) suggests to compare all the estimates of α obtained by the consistent methods. When the differences are substantial for a relatively large sample size, the stable distributions are less likely.

Apparently, the inference can be done not only for α, but also for the other parameters of the stable distributions (see Section 2.4). When all parameters have been estimated, one can use statistical criteria to measure the "goodness of fit" of the stable distributions, compared to the competing alternatives, e.g., by means of Kolmogorov distance (KD) statistic

$$KD = 100 \times \sup_{x \in \mathbb{R}} \left| F_s(x) - \hat{F}(x) \right|$$

Here for given sample $X_1, X_2, \ldots, X_n$, $\hat{F}(x)$ denotes the cumulative distribution function of the estimated (parametric) density and $F_s(x)$ denotes the empirical distribution, i.e., $F_s(x) = n^{-1} \sum_{t=1}^{n} \mathbb{1}_{(\infty, x]} \left[\frac{X_t - \hat{\mu}}{\hat{\sigma}} \right]$, where $\mathbb{1}(\cdot)$ denotes the indicator function. The other statistic is the

[7] For fitting the multivariate stable data, see, e.g., Nolan (1999)

[8] For instance, it is often noted that many economic data can be fitted by α-stable distribution on the range $1.5 \leq \alpha \leq 2$ (Fama, 1965; Weron, 2001). The conclusion is made merely on the base of $\hat{\alpha}$

Anderson-Dahling statistic

$$AD_0 = \sup_{x \in \mathbb{R}} \frac{\left| F_s(x) - \hat{F}(x) \right|}{\sqrt{\hat{F}(x)(1 - \hat{F}(x))}}$$

Furthermore, instead of looking at the maximum discrepancy, we also can look at the second and third largest discrepancies, and define AD_1 and AD_2, analogously as AD_0 (see Rachev and Mittnik (2000)).

Note that most of the methods that have been proposed in literature are designed to recognize the "heavy-tailedness" of i.i.d. data, i.e., they are not designed to see whether time series data are coming from an, e.g., α-stable ARMA process. In this regard, one possibility is to apply the considered method to the fitted residual, after filtering data using appropriate non i.i.d. models. Another possibility is to apply the methods directly to non i.i.d. stationary data. In this case, a common procedure is to firstly randomize the order of the data, in order to destroy serial dependence in the data (Adler *et al.*, 1998*a*). Note that this preprocessing procedure is not necessarily working well for the real data, e.g., when modelling financial time series, the dependency may not be removed after a random shuffle of the data, due to the persistence of strong GARCH effects (Paolella, 2001). However, the randomization should be working well in the ARMA (with small order) case.

To end this section, in the following, we briefly review Hill's estimator (Hill, 1975), one of the popular graphical estimators for the index α, which are based on the order statistics, within the class of heavy-tailed - not necessary stable - distributions [9]. Given the order statistics $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ from sample $X_1, X_2, \ldots, X_n$, Hill's estimator is defined as

$$\hat{\alpha}_{hill}^{-1} = k^{-1} \sum_{j=1}^{k} \ln(X_{n+1-j,n}) - \ln(X_{n-k,n})$$

with standard error $\widehat{SE}(\hat{\alpha}_{hill}) = k\hat{\alpha}_{hill}(k-1)^{-1}(k-2)^{-1/2}$ and $k = k(n) > 2$ denotes the threshold value where the tail area of the empirical distribution is "started". For $k = 1$, we only used the most extreme value for the estimation. The choice of k needs a trade-off, because it must be sufficiently small so $X_{n-k,n}$ is in the tail of the distribution, but if it is too small, the estimate has a precision problem. In literature, the value of k is commonly taken to be in between one and ten percent of the sample size. However, to get a reliable estimate $\hat{\alpha}$ of stable Paretian distributed data with $1.5 \leq \alpha \leq 1.95$, the optimal value for k is about $0.42n$ (Paolella, 2001). Despite its popularity in the stably distributed setting, Hill's estimator is severely biased even for a very large sample size. Several improvements for more accurate estimation of $\hat{\alpha}$ have been suggested, e.g., using Hill intercept method (Paolella, 2001) or smoothed Hill estimator (Resnick and Starica, 1997).

To illustrate this method, we simulate several independent data of length $n = 1000$ from $AR(2)$ process $X_t = 0.5X_{t-1} + 0.2X_{t-2} + \epsilon_t$ using the function `arima.sim` in R version 1.9.0 [10], where ϵ_t is $S\alpha S$ distributed, for several values of α. The unit symmetric stable variates are generated using the function `rstable` in the extension package `stable` [11]. For the comparison's sake, we also generate the process whose innovations are (0.5 Pareto(2,2)+ 0.5 (-Pareto(2,2)))[12], thus it is an infinite variance process but nonstably distributed. We also generate another $AR(2)$ process with standard Gaussian innovations, but it is randomly contaminated with the other $N(20, 10)$

[9] The other alternatives are Pickand's estimator, deHaan & Resnick's estimator (Mittnik and Rachev, 1996; Pictet *et al.*, 1998; Rachev and Mittnik, 2000) and "the log-tail test" (e.g., Mandelbrot, 1963; Weron, 2001)

[10] Available on the Web at `http://www.r-project.org` (Ihaka and Gentleman, 1996; R Development Core Team, 2004). Note that most of numerical results in this thesis are obtained using R. All R-codes that I used in this thesis are available upon request. See also `http://dedirosadi.staff.ugm.ac.id`

[11] Available on the Web at `http://alpha.luc.ac.be/~jlindsey/rcode.html`

[12] Recall that if X is $Pareto(\alpha, \sigma)$ distributed , $X \geq \sigma$, with scale parameter $\sigma > 0$ and shape parameter $\alpha > 0$, then its distribution function is given by $F(x) = 1 - (\sigma/x)^{\alpha}$. Its mean is given by $\frac{\alpha\sigma}{(\alpha-1)}$ and its variance equal to $\frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}$

data. The Hill's estimators for the data are plotted in Figure 2.2. Note that for Pareto i.i.d. data (Figure 2.2.a), the index obtained is approximately equal to the true index $\alpha = 2$ started from k is about 5%. For the α-stable series, the true value of α is obtained using about $15\% - 40\%$ of the extreme data, which follows the suggestion of Paolella (2001).

2.4 The estimation of the stable distribution parameters

In Section 2.3, we presented a quick and simple graphical method, called Hill's estimator, for estimating the index of stability α. For more general methods (i.e., not only for obtaining $\hat{\alpha}$, but also for the estimator of the other parameters of i.i.d. α-stable distributions), the estimation methods can be classified as follows: *the sample quantile* methods, i.e., Fama-Roll's method (Fama and Roll, 1971) and its extension by McCulloch (1986); *the empirical characteristic function* methods, i.e., methods of moments (Press, 1972), Koutrouvelis's regression method (Koutrouvelis, 1980*b*; Kogon and Williams, 1998), the minimum distance method (see Yu, 2004, for a review); and the *maximum likelihood estimator* (for a review, see Rachev and Mittnik, 2000, Section 3.7.). Some new methods have been recently proposed, such as *the extreme value method*, *the logarithmic moments method*, and *the fractional lower order moments* (Kuruoğlu, 2001). *The monte carlo* or *bootstrap* approach can also be used for estimating the parameters of heavy-tailed distributions, see, e.g., Pictet *et al.* (1998), Tsionas (1999). A systematic study of the performance of the estimation procedures based on Fama-Roll's method, McCulloch method, method of moments, and the regression method, can be found in Weron (1995). Among the methods considered, Weron (1995) suggests to implement the regression method as the most reliable estimator when the range of α and σ are unknown. It is important to know that although all of the methods mentioned above are designated for i.i.d. process, in some cases of non i.i.d. data, we still can use the quantile-based estimator for $\hat{\alpha}$ (e.g., McCulloch's method). For instance, Adler *et al.* (1998*a*) shows simulation examples that this method works well for small order MA and AR processes.

In what follows, we only present a short description of McCulloch's estimation method. This estimation procedure will be used in this thesis.

Sample quantile methods

Probably one of the oldest estimation methods for the parameters of stable Paretian distributions (i.e., the case with $\alpha < 2$) is based on sample quantiles. Here we only present the McCulloch's method. This method is computationally simple, and is shown to provide consistent estimates of all four parameters. Define

$$\nu_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}$$

$$\nu_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.50}}{x_{0.95} - x_{0.05}}$$

where x_p denotes the p-th quantile of x_i. ν_α and ν_β do not depend on σ and μ, and their values as the functions of α and β are tabulated in McCulloch (1986) (reproduced in Table 2.1 and Table 2.2). Moreover ν_α is a strictly decreasing function of α, where ν_β is a strictly increasing function of β. Let $\hat{\nu}_\alpha$ and $\hat{\nu}_\beta$ denote the sample values (which are the consistent estimates, since p-th sample quantile $\hat{x}_p$ is consistent estimates of x_p) of ν_α and ν_β, respectively. Since ν_α and ν_β are functions $\psi_1(\alpha, \beta)$ and $\psi_2(\alpha, \beta)$ of α and β, the parameters α and β are obtained as inversion of functions ψ_1 and ψ_2,

$$\alpha = \varphi_1(\nu_\alpha, \nu_\beta), \ \beta = \varphi_2(\nu_\alpha, \nu_\beta),$$

and the estimate of α and β can be obtained by substituting $\hat{\nu}_\alpha$ and $\hat{\nu}_\beta$ into Table 2.1 and Table 2.2. Here the interpolation, and sometimes the truncation (i.e., when $\hat{\alpha}$ or $\hat{\beta}$ out of its possible range) are necessary to obtain the estimates.

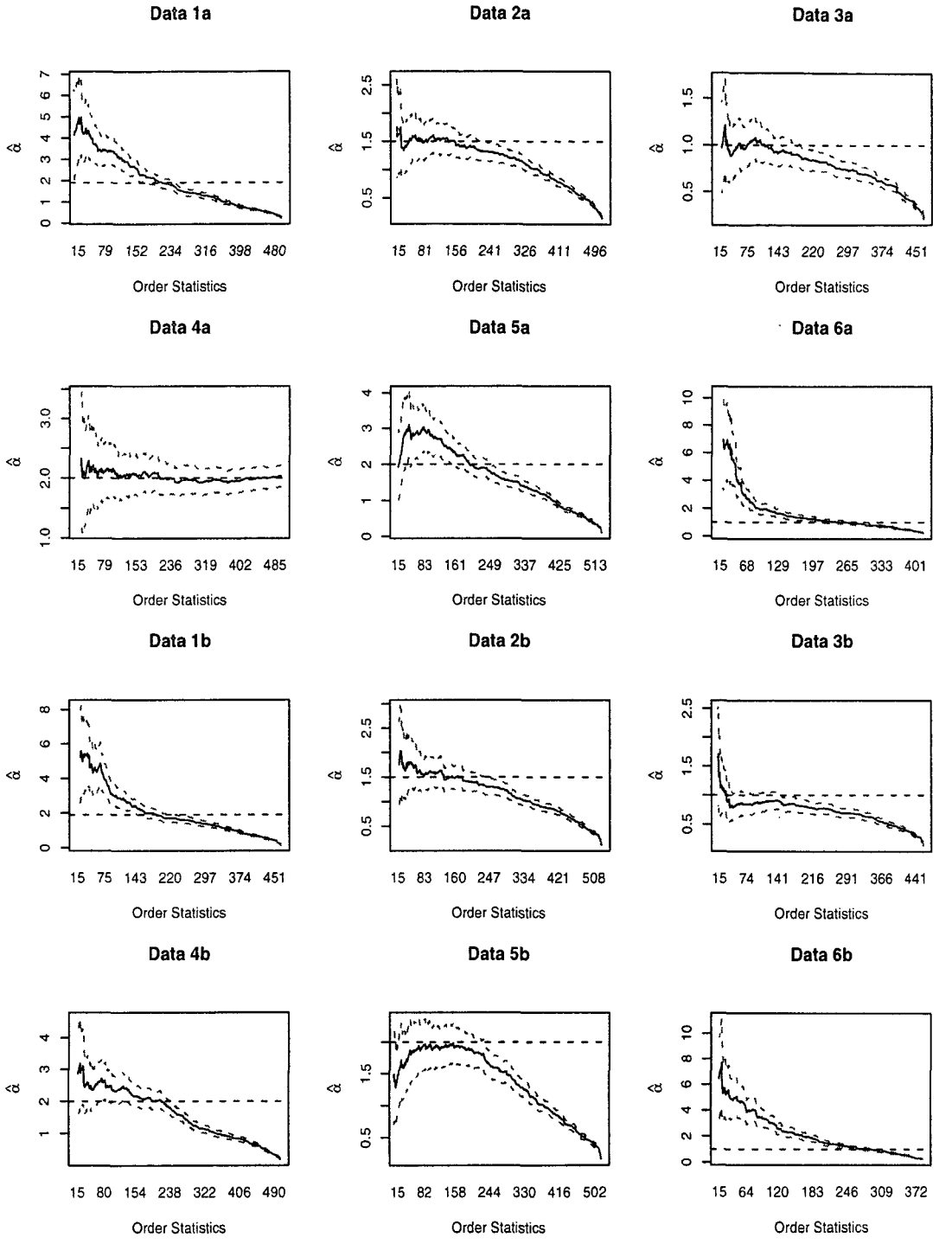


Figure 2.2: Hill's plots of AR(2) series (Figure 1b-3b) with innovations are i.i.d. SαS series, where (1a). $\alpha = 1.9$, (2a). $\alpha = 1.5$, (3a). Cauchy, respectively. For the comparison's sake, in Figure (4b-6b) we also present Hill's plots of AR(2) series with the innovations are assumed to be, respectively, (4a). Pareto distributed, (5a). Standard Gaussian randomly contaminated by the other Gaussian process with larger mean and larger variance, and also (6a). when the largest and the smallest 5 % of the Cauchy innovations are truncated.

v_α	v_β						
	0	0.1	0.2	0.3	0.5	0.7	1
2.439	2	2	2	2	2	2	2
2.5	1.916	1.924	1.924	1.924	1.924	1.924	1.924
2.6	1.808	1.813	1.829	1.829	1.829	1.829	1.829
2.7	1.729	1.73	1.737	1.745	1.745	1.745	1.745
2.8	1.664	1.663	1.663	1.668	1.676	1.676	1.676
3	1.563	1.56	1.553	1.548	1.547	1.547	1.547
3.2	1.484	1.48	1.471	1.46	1.448	1.438	1.438
3.5	1.391	1.386	1.378	1.364	1.337	1.318	1.318
4	1.279	1.273	1.266	1.25	1.21	1.184	1.15
5	1.128	1.121	1.114	1.101	1.067	1.027	0.973
6	1.029	1.021	1.014	1.004	0.974	0.935	0.874
8	0.896	0.892	0.887	0.883	0.855	0.823	0.769
10	0.818	0.812	0.806	0.801	0.78	0.756	0.691
15	0.698	0.695	0.692	0.689	0.676	0.656	0.595
25	0.593	0.59	0.588	0.586	0.579	0.563	0.513

Table 2.1: $\alpha = \varphi_1(\nu_\alpha, \nu_\beta) = \varphi_1(\nu_\alpha, -\nu_\beta)$

v_α	v_β						
	0	0.1	0.2	0.3	0.5	0.7	1
2.439	0	2.16	1	1	1	1	1
2.5	0	1.592	3.39	1	1	1	1
2.6	0	0.759	1.8	1	1	1	1
2.7	0	0.482	1.048	1.694	1	1	1
2.8	0	0.36	0.76	1.232	2.229	1	1
3	0	0.253	0.518	0.823	1.575	1	1
3.2	0	0.203	0.41	0.632	1.244	1.906	1
3.5	0	0.165	0.332	0.499	0.943	1.56	1
4	0	0.136	0.271	0.404	0.689	1.23	2.195
5	0	0.109	0.216	0.323	0.539	0.827	1.917
6	0	0.096	0.19	0.284	0.472	0.693	1.759
8	0	0.082	0.163	0.243	0.412	0.601	1.596
10	0	0.074	0.147	0.22	0.377	0.546	1.482
15	0	0.064	0.128	0.191	0.33	0.478	1.362
25	0	0.056	0.112	0.167	0.285	0.428	1.274

Table 2.2: $\beta = \varphi_2(\nu_\alpha, \nu_\beta) = -\varphi_2(\nu_\alpha, -\nu_\beta)$

To obtain the estimate of σ, McCulloch (1986) tabulated the behavior of function

$$\nu_\sigma = \frac{x_{0.75} - x_{0.25}}{\sigma}$$

as the function $\varphi_3(\alpha, \beta)$ (see Table 2.3). Then the consistent estimate of σ can be calculated using

$$\hat{\sigma} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{\varphi_3(\hat{\alpha}, \hat{\beta})}$$

In order to get a good estimate of the location parameter μ, we need to introduce a new parameter η defined as

$$\eta = \begin{cases} \mu + \beta\sigma \tan\frac{\pi\alpha}{2} & \text{for } \alpha \neq 1 \\ \mu & \text{for } \alpha = 1 \end{cases}$$

The parameter η does not have immediate interpretation, and it was introduced in Zolotarev (1986) as an alternative parameterization (which is referred to as Zolotarev's M parameterization) of the stable distribution. Define the index

$$\nu_\eta = \frac{\eta - x_{0.5}}{\sigma}$$

α	β				
	0	0.25	0.5	0.75	1
2	1.908	1.908	1.908	1.908	1.908
1.9	1.914	1.915	1.916	1.918	1.921
1.8	1.921	1.922	1.927	1.936	1.947
1.7	1.927	1.93	1.943	1.961	1.987
1.6	1.933	1.94	1.962	1.997	2.043
1.5	1.939	1.952	1.988	2.045	2.116
1.4	1.946	1.967	2.022	2.106	2.211
1.3	1.955	1.984	2.067	2.188	2.333
1.2	1.965	2.007	2.125	2.294	2.491
1.1	1.98	2.04	2.205	2.435	2.696
1	2	2.085	2.311	2.624	2.973
0.9	2.04	2.149	2.461	2.886	3.356
0.8	2.098	2.244	2.676	3.265	3.912
0.7	2.189	2.392	3.004	3.844	4.775
0.6	2.337	2.635	3.542	4.808	6.247
0.5	2.588	3.073	4.534	6.636	9.144

Table 2.3: $\nu_\sigma = \varphi_3(\alpha, \beta) = \varphi_3(\alpha, -\beta)$

which can be seen as the function $\psi_4(\alpha, \beta)$ (see Table 2.4). Then, the consistent estimate of η can be calculated from

$$\hat{\eta} = \hat{x}_{0.5} + \hat{\sigma}\psi_4(\hat{\alpha}, \hat{\beta})$$

and hence for $\alpha \neq 1$,

$$\hat{\mu} = \hat{\eta} - \hat{\beta}\hat{\sigma}\tan\frac{\pi\hat{\alpha}}{2}$$

α	β				
	0	0.25	0.5	0.75	1
2	0	0	0	0	0
1.9	0	-0.017	-0.032	-0.049	-0.064
1.8	0	-0.03	-0.061	-0.092	-0.123
1.7	0	-0.043	-0.088	-0.132	-0.179
1.6	0	-0.056	-0.111	-0.17	-0.232
1.5	0	-0.066	-0.134	-0.206	-0.283
1.4	0	-0.075	-0.154	-0.241	-0.335
1.3	0	-0.084	-0.173	-0.276	-0.39
1.2	0	-0.09	-0.192	-0.31	-0.447
1.1	0	-0.095	-0.208	-0.346	-0.508
1	0	-0.098	-0.223	-0.383	-0.576
0.9	0	-0.099	-0.237	-0.424	-0.652
0.8	0	-0.096	-0.25	-0.469	-0.742
0.7	0	-0.89	-0.262	-0.52	-0.853
0.6	0	-0.078	-0.272	-0.581	-0.997
0.5	0	-0.061	-0.279	-0.659	-1.198

Table 2.4: $\nu_\eta = \psi_4(\alpha, \beta) = -\psi_4(\alpha, -\beta)$

2.5 Generating the stable random variates

In order to do the Monte Carlo simulation of (univariate) stably distributed process[13], it is necessary to have the generator of the α-stable variates. For Cauchy, Gaussian and Lévy cases, the analytic expressions for the inverse probability distribution function are available. Therefore, the inverse transform method can be applied easily, and the generators are already implemented in

[13] A method for simulating bivariate stable vector is discussed in Modarres and Nolan (1994). It is implemented in `MVSTABLE.EXE`, available at `http://academic2.american.edu/~jpnolan/`

many standard statistical packages. However, as it is already known, the analytic expressions are not available for the other choices of parameters. In this case, the standard method for simulating stable random variates is based on the method presented in Chambers *et al.* (1976). We will present it below (for the proof, refer to Chambers *et al.*, 1976; Weron, 1996).

In the symmetric case ($\beta = 0$), to simulate α-stable random variable $X \sim S_\alpha(1, 0, 0)$ for $\alpha \in (0, 2]$, we use the following steps:

1. Generate a random variable U uniformly distributed on $(-\pi/2, \pi/2)$ and an exponential random variable V with mean 1, independent of U
2. Compute
$$X = \frac{\sin(\alpha U)}{\{\cos(U)\}^{1/\alpha}} \left\{ \frac{\cos(U - \alpha U)}{V} \right\}^{(1-\alpha)/\alpha} \tag{2.15}$$

From (2.6), using (2.15), we can generate $Y \sim S_\alpha(\sigma, 0, \mu)$ random variates as $\sigma X + \mu$.

To generate the skewed α-stable variates $X \sim S_\alpha(1, \beta, 0)$ with $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$, we can proceed as follows:

1. Generate a random variable U uniformly distributed on $(-\pi/2, \pi/2)$ and an exponential random variable V with mean 1, independent of U
2. If $\alpha = 1$, compute (see also Weron (1995))
$$X = \frac{2}{\pi} \left[(\tfrac{\pi}{2} + \beta U) \tan U - \beta \log \frac{V \cos U}{\frac{\pi}{2} + \beta U} \right]$$
If $\alpha \neq 1$, compute
$$X = S_{\alpha,\beta} \frac{\sin(\alpha(U + B_{\alpha,\beta}))}{(\cos(U))^{1/\alpha}} \left(\frac{\cos(U - \alpha(U + B_{\alpha,\beta}))}{V} \right)^{\frac{1-\alpha}{\alpha}} \tag{2.16}$$
where
$$B_{\alpha,\beta} = \tan^{-1}(t)/\alpha$$
$$S_{\alpha,\beta} = (1 + t^2)^{\frac{1}{2\alpha}}$$
and
$$t = \beta \tan(\tfrac{\pi\alpha}{2})$$

To generate $Y \sim S_\alpha(\sigma, \beta, \mu)$ random variates, for $\sigma \in \mathbb{R}^+, \mu \in \mathbb{R}$, we can use (2.16) by calculating

$$Y = \begin{cases} \sigma X + \mu & \text{for } \alpha \neq 1 \\ \sigma X + \frac{2}{\pi}\beta\sigma \log \sigma + \mu & \text{for } \alpha = 1 \end{cases}$$

The method presented above is implemented as `FORTRAN` routine `rstab` in Chambers *et al.* (1976) and its updated version in Samorodnitsky and Taqqu (1994). In the other programming languages, it is available online as function `rstab` in `S/S+`, in executable version of `STABLE.EXE` [14] or function `rstable` in the extension package `stable`, implemented in `R`.

[14] available on the Web at `http://academic2.american.edu/~jpnolan/`

Chapter 3

Review of heavy-tailed ARMA modelling

In this chapter we present a short summary of the results related to the heavy-tailed linear time series modelling, in particular, for the class of stable causal ARMA models with α-stable noise. The summary is presented by following the traditional systematic procedure for fitting time series models to data. This procedure is done in three stages: *order selection*, *parameter estimation* and *diagnostic checking*. All results are given without proofs, therefore we refer the readers who interested in mathematical details to literature. The purpose of this chapter is to present a unified summary for the results which are closely related to our topic. For the sake of brevity, the detailed presentation is given only for results which will be used in the subsequent discussion in this thesis. More details and systematic reviews for some parts of this chapter can be found in, e.g., Embrechts *et al.* (1997), Chapter 7 and Rachev and Mittnik (2000), Section 5.3. For practical issues, we refer the reader to, e.g., Adler *et al.* (1998*a*). In some cases, for comparison's sake and ease of reference, we start the presentation of results in the classical Gaussian model, and then turn to the heavy-tailed model.

3.1 Heavy-tailed time series

In this thesis, we are only concerned with **univariate**, **(strictly) stationary**, **causal-linear** processes, which are given as

$$X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, t \in \mathbb{Z} \tag{3.1}$$

with i.i.d. real-valued $\{\epsilon_t, t \in \mathbb{Z}\}$ and $c_0 = 1$. Furthermore, we assume the following

(C1). The coefficients c_j's are real-valued and satisfying $|c_j| < CQ^{-j}$ for some $C > 0$, $Q > 1$

(C2). ϵ_t is i.i.d. $S\alpha S$, $0 < \alpha \leq 2$ with $\sigma > 0$

Recall from Section 2.2 that ϵ_t has the characteristic function

$$E \exp(is\epsilon_1) = \exp(-\sigma^\alpha |s|^\alpha), s \in \mathbb{R} \tag{3.2}$$

where $0 < \alpha \leq 2$ and $\sigma > 0$. When $\alpha = 2$, ϵ_t is Gaussian with $\sigma_\epsilon^2 = 2\sigma^2$. When $\alpha < 2$, ϵ_t has infinite variance, thus only the notion of strictly stationary is available. In this thesis, we call this the stationary heavy-tailed case.

Note that classically we assume that $\{\epsilon_t\}$ has mean zero and finite variance σ_ϵ^2. In the classical setting, for fixed t, the series (3.1) converges a.s., provided that the real-valued coefficients c_j's satisfy the conditions

$$\sum_{j=0}^{\infty} |c_j| < \infty$$

and $\text{var}(X_t) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} c_j^2 < \infty$ (e.g., Brockwell and Davis (1987), Proposition 3.1.1). The process $\{X_t, t \in \mathbb{Z}\}$ is strictly stationary, the finite dimensional distributions of the process are invariant under time shift. In other words, the finite dimensional distributions for any instant of time $\{t_1, \cdots, t_d\} \in \mathbb{Z}$ are the same as for $\{t_1 + h, \cdots, t_d + h\} \in \mathbb{Z}$ for any $h \in \mathbb{Z}$. Every strictly stationary process with finite second moments is also stationary *in the wide sense*, i.e., there exists a constant μ such that $E(X_t) = \mu$ is a constant, and the covariance function EX_tX_s is only a function of time distance $|t - s|$.

Under conditions C1 and C2, (3.1) is well defined in the sense of a.s. convergence (by virtue of the 3-series theorem for a series of independent summands, see Appendix A.1). Moreover, under assumption C2, the process $\{X_t\}$ will be a (strictly) stationary $S\alpha S$ process with the same index of stability α but the scale parameter $\sigma_X = \sigma_\epsilon(\sum_{j=0}^{\infty} |c_j|^\alpha)^{1/\alpha}$. The latter property can be obtained by noting that

$$E\exp(isX_t) = \prod_{j=0}^{\infty} \exp(-\sigma_\epsilon^\alpha |c_j s|^\alpha) = \exp(-\sigma_\epsilon^\alpha \sum_{j=0}^{\infty} |c_j|^\alpha |s|^\alpha)$$

We write this "equality in law" as

$$X_t \stackrel{d}{=} \epsilon_t(\sum_{j=0}^{\infty} |c_j|^\alpha)^{1/\alpha}$$

To end this section, we deals with causal stable $ARMA(p, q)$ processes with $S\alpha S$ noise. We write the autoregressive moving average process of the order (p, q)

$$A(z)X_t = B(z)\epsilon_t \tag{3.3}$$

for $t \in \mathbb{Z}$, where z denotes backward-shift operator (here $z(X_t) = (X_{t-1})$) as well as the complex variable, and where the polynomials A and B are

$$A(z) = 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p = 1 - \sum_{i=1}^{p} a_i z^i \tag{3.4}$$

$$B(z) = b_0 + b_1 z + b_2 z^2 + \cdots + b_q z^q = \sum_{i=1}^{q} b_i z^i \tag{3.5}$$

for fixed order (p, q). The coefficients $a_1, a_2, \ldots, a_p$ and $b_1, \ldots, b_q$ are real-valued, and $b_0 = 1$. The polynomials A and B are assumed have no common roots, and $A(z)$ has no roots in the closed unit disk. The sequence ϵ_t fulfils the condition C2.

Theorem 3.1.1 (Kokoszka and Taqqu (1994), Proposition 2.1). *The system (3.3) has unique stationary solution of the form*

$$X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, t \in \mathbb{Z}, \text{ almost surely (a.s)} \tag{3.6}$$

where the c_j's are real-valued, satisfy $|c_j| < CQ^{-j}$ for some $C > 0$, $Q > 1$, if and only if $A(z)$ has no roots in the closed unit disk $\{z : |z| \leq 1\}$. The sequence $\{X_t, t \in \mathbb{Z}\}$ is then stationary and $S\alpha S$.

The c_j's are the coefficients in the series expansion of $B(z)/A(z), |z| < 1$

As a side remark here, note that in literature, many results are derived under more general condition on ϵ_t than the condition C2 (e.g., Davis and Resnick, 1986). To ensure the absolute convergence of series (3.1), it is required that

$$\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty, 0 < \delta < \alpha, \delta \leq 1 \tag{3.7}$$

For $\{\epsilon_t\}$ is i.i.d. sequence with Pareto tail distributed, i.e., it satisfies

$$\bar{F}(x) = P(|\epsilon_1| > x) \sim kx^{-\alpha}, x \to \infty$$

for some $\alpha > 0$ and $k > 0$, the absolute convergence has been shown in, e.g., Resnick (2001), p. 228, or Brockwell and Davis (1987). However, to make a unified summary for the results, note that throughout this thesis we only consider the causal model as in (3.1) with the condition C2 for $\{\epsilon_t\}$.

3.2 Order selection of ARMA models

Classically, the order selection stage of time series modelling is done via visual examination of plots of time series, the sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF). In this section, we will present the theoretical properties of the SACF of heavy-tailed linear time series models. In addition, we also present the properties of the autocovariation function and its sample analog for stationary processes with finite mean (Gallagher, 1998).

It is also important to mention here that alternative order selection can be done via information criterion. Under the finite variance assumption, a number of different criteria has been suggested in literature, such as Akaike's information criterion (AIC) and Bayesian information criterion (BIC). In the infinite variance setting, order determination of pure autoregressive or pure moving average processes has been treated in Bhansali (1988) and (Knight, 1989). Note that to the best of our knowledge, there is no theoretical result yet available for the heavy-tailed ARMA models. For Gaussian AR process, it is known that the AIC criterion does not give a consistent estimate for the true order, and in fact, it tends to overestimate the true order. However, for heavy-tailed AR process, under some regularity conditions, Knight (1989) shows that AIC criterion will give the consistent estimate of the true order of the process. This consistency property also obtained for FPE_{δ} criterion (Bhansali, 1988).

3.2.1 Estimation of the autocorrelation function

From i.i.d.-ness of ε_t, for the linear process (3.1), when $\sigma_{\epsilon}^2 < \infty$, we obtain its population covariance function as

$$\gamma(k) = \sigma_{\epsilon}^2 \sum_{j=0}^{\infty} c_j c_{j+|k|}$$

and therefore by definition, the ACF is

$$\rho(k) = \sum_{j=0}^{\infty} c_j c_{j+|k|} / \sum_{j=0}^{\infty} c_j^2$$

The estimators for $\gamma(k)$ and $\rho(k)$ are given by the sample covariance $\hat{\gamma}(k)$ and the SACF $\hat{\rho}(k)$

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} X_t X_{t+|k|}, k \in \mathbb{Z} \tag{3.8}$$

$$\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0) = \sum_{t=1}^{n-|k|} X_t X_{t+|k|} / \sum_{t=1}^{n} X_t^2, k \in \mathbb{Z} \tag{3.9}$$

and $\hat{\gamma}(k) = 0$ and $\hat{\rho}(k) = 0$ for $|k| > n$. In the classical case, the sample functions $\hat{\gamma}(k)$ and $\hat{\rho}(k)$ are consistent estimators of the population functions and asymptotically normal distributed. The limiting distribution property of the SACF in the classical case is summarized in the following theorem (see, e.g., Brockwell and Davis, 1987, Theorem 7.2.1. and 7.2.2.).

Theorem 3.2.1. *Let $\{X_t\}$ be the mean zero linear process (3.1). Suppose that either*

$$\sum_{j=0}^{\infty} |c_j| < \infty \text{ and } E\epsilon_t^4 < \infty$$

or

$$\sum_{j=0}^{\infty} |c_j| < \infty \ , \ \sum_{j=0}^{\infty} c_j^2 j < \infty \text{ and } \sigma_\epsilon^2 < \infty$$

Then for each $m \geq 1$,

$$\sqrt{n}(\hat{\rho}(k) - \rho(k))_{k=1,\ldots,h} \xrightarrow{d} (Y_k)_{k=1,\ldots,h}$$

where

$$Y_k = \sum_{j=1}^{\infty} [\rho(k+j) + \rho(k-j) - 2\rho(k)\rho(j)]\, G_j, k = 1, \ldots, h$$

and $(G_j)_{j\geq 1}$ are i.i.d. $N(0,1)$ random variables.

In particular, for each fixed $k \geq 1$, from this theorem we obtain Bartlett's formula

$$\sqrt{n}(\hat{\rho}(k) - \rho(k)) \xrightarrow{d} \left(\sum_{j=1}^{\infty} [\rho(k+j) + \rho(k-j) - 2\rho(k)\rho(j)] \right)^{1/2} G_1 \tag{3.10}$$

Now, under assumption C2, but $\alpha < 2$, then $\sigma_{\epsilon_t}^2 = \text{var}(\epsilon_t) = \infty$, and therefore the population autocovariance and autocorrelation do not exist. However, their sample version are well-defined random variables. The asymptotic properties of the SACF in this case are already summarized in Section 1.2, and for the ease of reference, we recall in the following theorem.

Theorem 3.2.2 (Davis and Resnick (1986)). *Let $\{X_t\}$ be the mean zero, strictly stationary linear process (3.1) where the coefficients c_j's satisfy the condition (3.7), and where $\{\epsilon_t\}$ is i.i.d. $S\alpha S$ process then when $\alpha < 2$,*

$$\hat{\rho}(k) \xrightarrow{p} \tilde{\rho}(k), k \geq 0 \tag{3.11}$$

where $\xrightarrow{p}$ denotes convergence in probability. In addition, for each $h \geq 1$,

$$(n/\ln(n))^{1/\alpha}(\hat{\rho}(1) - \tilde{\rho}(1), \ldots, \hat{\rho}(h) - \tilde{\rho}(h))' \xrightarrow{d} (Y_1, \ldots, Y_h)' \tag{3.12}$$

where $\xrightarrow{d}$ denotes convergence in distribution, $\tilde{\rho}(k) = \sum_{j=0}^{\infty} c_j c_{j+k} / \sum_{j=0}^{\infty} c_j^2$ and

$$Y_k = \sum_{j=1}^{\infty} [\tilde{\rho}(k+j) + \tilde{\rho}(k-j) - 2\tilde{\rho}(k)\tilde{\rho}(j)] \frac{G_j}{G_0}, k = 1, \ldots, h$$

Here, $(G_j)_{j\geq 1}$ and G_0 are independent random variables, where G_0 is positive $\alpha/2$-stable with characteristic function

$$E \exp\{isG_0\} = \exp\{-\Gamma(1 - \alpha/2)\cos(\pi\alpha/4)|s|^{\alpha/2}(1 - i\ \text{sign}(s)\tan(\pi\alpha/4))\} \tag{3.13}$$

and $(G_j)_{j\geq 1}$ *are i.i.d.* $S\alpha S$ *random variables with characteristic function*

$$E\exp\{isG_1\} = \begin{cases} \exp\{-\Gamma(2-\alpha)\cos(\pi\alpha/2)|s|^\alpha & \text{if } \alpha \neq 1 \\ \exp\{-\pi|s|/2\} & \text{if } \alpha = 1 \end{cases} \tag{3.14}$$

Here $\Gamma(\cdot)$ *denotes the gamma function. The marginal distribution of each* Y_k *somewhat simpler, and we have that*

$$Y_k = \left(\sum_{j=1}^{\infty} |\tilde{\rho}(k+j) + \tilde{\rho}(k-j) - 2\tilde{\rho}(k)\tilde{\rho}(j)|^\alpha\right)^{1/\alpha} \frac{G_1}{G_0} \tag{3.15}$$

When $\alpha > 1$, *(3.12) is still valid if* $\hat{\rho}(k), k = 1, \ldots, h$ *is replaced by the centralized version.*

Notice that the normalization constants and the limiting distributions are different in case of finite and infinite variance.

3.2.2 Estimation of the autocovariation function

Gallagher (1998) introduces a linear dependence measure which is called the autocovariation function and is designated as the generalization of the auto correlation function for the stationary processes with finite mean. We present a brief description of this measure as follows. In case of stationary process $\{X_t\}$ with finite second moments, we obtain the conditional expectation property:

$$E(X_t|X_{t-k}) = \rho(k)X_{t-k} \tag{3.16}$$

with $\rho(k)$ denote the autocorrelation of the process at lag k. By law of iterated expectation argument, we obtain

$$E(X_tX_{t-k}) = \rho(k)E|X_{t-k}|^2$$

Thus,

$$\rho(k) = \frac{E(X_tX_{t-k})}{E|X_{t-k}|^2}$$

which can be identified as the autocorrelation of the process at lag k. Now if we extend the linearity property (3.16) into the stationary process with finite first absolute moments, namely

$$E(X_t|X_{t-k}) = \lambda(k)X_{t-k} \tag{3.17}$$

then using the iterated expectation argument, Gallagher (1998) defines the autocovariation function at lag k as

$$\lambda(k) = \frac{E(X_tS_{t-k})}{E|X_{t-k}|}, k \in \mathbb{Z} \tag{3.18}$$

where $S_t = \text{sign}(X_t)$. Thus, for mean zero stationary processes with finite first absolute moments, $\lambda(k)$ is a measure of linear dependence between X_t and X_{t-k}. When X_t has finite variance, the autocovariation is an analog measure to the autocorrelation function.

The autocovariation function as in eq. (3.18) can be seen as the normalized version of the (non-normalized) autocovariation function (AcovF) at lag k

$$\delta(k) = E(X_tS_{t-k}) \text{ for } k = 0, \pm 1, \pm 2, \ldots$$

Notes that in general $\delta(k) \neq \delta(-k)$. The method of moments estimator of AcovF is defined in Gallagher (1998) as

$$\hat{\delta}(k) = n^{-1}\sum_{t=l}^{r} S_{t-k}X_t$$

where $l = \max(1, 1+k)$ and $r = \min(n, n+k)$. The sample normalized autocovariation function is defined as

$$\hat{\lambda}(k) = \frac{\sum_{t=l}^{r} S_{t-k} X_t}{\sum_{t=1}^{n} |X_t|}$$

The asymptotic properties of the sample autocovariation is studied in Gallagher (1998). We present a summary below.

Theorem 3.2.3. *If $\{X_t\}$ is a stationary ergodic sequence with $E|X_1| < \infty$, then*

$$\hat{\lambda}(k) \xrightarrow{a.s.} \lambda(k)$$

Applying this theorem, we obtain the following corollary.

Corollary 3.2.4. *If $\{X_t\}$ is a mean zero ARMA process where ϵ_t is i.i.d. $S\alpha S$ for some $\alpha > 1$ (thus $E|X_1| < \infty$) then*

$$\hat{\lambda}(k) \xrightarrow{a.s.} \lambda(k)$$

The weak limit properties of the sample autocovariation function are summarized in the following theorem

Theorem 3.2.5. *If $\{X_t\}$ satisfy*

$$X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}; \sum_{j=0}^{\infty} j\,|c_j| < \infty$$

where $\{\epsilon_t\}$ is a mean zero i.i.d. sequence with $E|\epsilon_t| < \infty$ and such that $P(X_1 = 0) = 0$. For any h, let

$$\lambda_h = (\lambda(-h), \ldots, \lambda(h)) \text{ and } \hat{\lambda}_h = (\hat{\lambda}(-h), \ldots, \hat{\lambda}(h))$$

(i). *If $E|\epsilon_1|^2 < \infty$ then*

$$n^{1/2}(\lambda_h - \hat{\lambda}_h) \xrightarrow{d} (E\,|X_1|)^{-1}\mathbf{X}$$

where $\mathbf{X}$ denotes multivariate normal distribution

(ii). *If $\{\epsilon_t\}$ is i.i.d. $S\alpha S$ then*

$$n^{1-\frac{1}{\alpha}}(\lambda_h - \hat{\lambda}_h) \xrightarrow{d} (E\,|X_1|)^{-1}\mathbf{Y}$$

where $\mathbf{Y}$ denotes multivariate stable distribution.

In particular, Gallagher (2000) shows that for $MA(q)$ process, $X_t = \sum_{j=0}^{q} c_j \epsilon_{t-j}$, $c_0 = 1$, where ϵ_t is i.i.d. $S\alpha S$ with $\alpha > 1$ and $|k| > q$,

$$n^{1-\frac{1}{\alpha}}\hat{\lambda}(k) \xrightarrow{d} \left(\frac{E|\mathbf{S}^T\mathbf{c}|^\alpha}{\sum_{j=0}^{q}|c_j|^\alpha}\right) X \tag{3.19}$$

where $X \sim S_\alpha(\frac{\pi}{2\Gamma(1-1/\alpha)}, 0, 0)$, $\mathbf{S} = (S_0, \ldots, S_q)^T$ and $\mathbf{c} = (c_0, \ldots, c_q)^T$. Notice that the constant in front of X reduces to 1 for i.i.d. process, since $q = 0$.

Note that for stationary symmetric α-stable processes, the autocovariation function is equivalent to a dependence measure, called the covariation coefficient (Nikias and Shao, 1995). A summary about the sample covariation coefficient can be found in Nikias and Shao (1995), Chapter 6 and 7.

3.2.3 Notes on the graphical order identification procedures

In this part, we discuss the application of the results presented in the previous subsections for identifying the order of $MA(q)$ processes. In the classical case, the correlation function $\rho(k)$ of $MA(q)$ process vanishes after lag q. Using this fact and the asymptotic distribution of the SACF $\hat{\rho}(\cdot)$, as the standard procedure in Box-Jenkins method, we attempt to identify the order q of moving average processes by plotting the SACF at lags $k = 1, 2, \ldots, K$ and looking for the lag q after which $\hat{\rho}(\cdot)$ is approximately zero. In the case of $MA(q)$ process where ϵ_t is i.i.d. mean zero and variance $\sigma^2 < \infty$, we obtain from Bartlett's formula (eq. 3.10) for $k > q$,

$$\sqrt{n}\hat{\rho}(k) \xrightarrow{d} (1 + 2\sum_{j=0}^{q} \rho(j)^2)^{1/2} Z \tag{3.20}$$

where Z is standard normal random variable. But in practice the values of $\rho(j), j = 0, 1, \ldots, q$ are not known. Therefore, to identify the order q, the practitioners will plot the SACF together with the bounds $\pm 1.96/\sqrt{n}$, n denotes the sample size, and consider the SACF is approximately zero if its values at certain lags are inside the interval. These bounds are obtained from the fact that for i.i.d. process, the asymptotic distribution of $\hat{\rho}(k), k \neq 0$ is a standard normal distribution.

This result motivates us to use similar order identification methods for moving average processes in the infinite variance case, e.g., as considered in Adler *et al.* (1998*a*). For $MA(q)$ process where ϵ_t is i.i.d. SαS with $\alpha < 2$ and $k > q$, from (3.15) we obtain

$$(n/\ln(n))^{1/\alpha}\hat{\rho}(k) \xrightarrow{d} (1 + 2\sum_{j=0}^{q} \tilde{\rho}(j)^{\alpha}))^{1/\alpha} G_1/G_0 \tag{3.21}$$

Thus, based on (3.19), (3.20), (3.21), we might consider the following identification strategies:

(i) The practitioner may be unaware that the data are coming from a process with heavy tails. In such case, he might plot $\hat{\rho}(k)$ at various lags and compare to the quantiles $1.96/\sqrt{n}$ of the normal distribution. This method is considered, e.g., in Rosenfeld (1976).

(ii) The practitioner might plot $\hat{\rho}^*(k) = (n/\log(n))^{1/\alpha}\hat{\rho}(k)$ at various lags and compare to the 2.5% and 97.5% quantiles of the distribution G_1/G_0, see Adler *et al.* (1998*a*).

(iii) The practitioner might plot $\hat{\lambda}(k)$ at various lags and compare to the 2.5% and 97.5% quantiles of $Xn^{(1/\alpha)-1}$ (Gallagher, 2000). Quantiles for the stable distributions are tabulated in, e.g., Samorodnitsky and Taqqu (1994), whereas a fast calculation of quantiles of the stable distributions can be performed using function `qstable` in the package `stable` of R.

3.3 Parameter estimation of ARMA models

First we recall the notion of $ARMA(p, q)$ process, which is a linear process given by a difference equation

$$X_t - a_1 X_{t-1} - \cdots - a_p X_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q}, t \in \mathbb{Z} \tag{3.22}$$

for fixed order (p, q). Define the polynomials

$$A(z) = 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p = 1 - \sum_{i=1}^{p} a_i z^i \tag{3.23}$$

$$B(z) = b_0 + b_1 z + b_2 z^2 + \cdots + b_q z^q = \sum_{i=1}^{q} b_i z^i \tag{3.24}$$

We write the vector of coefficients as

$$\beta = (a_1, \ldots, a_p, b_1, \ldots, b_q)^T$$

and use symbol β_0 for the true but unknown parameter vector. The observed time series $X_1, \ldots, X_n$ is supposed to come from the model (3.22) with $\beta = \beta_0$.

In the heavy-tailed time series literature, compared to the other stages, it seems that this stage of modelling has been studied rather extensively. In the following two subsections, we present some results related to the estimation of parameters β_0, which will be used in the subsequent discussion in this thesis. The other results are concisely summarized in the last subsection.

3.3.1 Yule-Walker and Generalized Yule-Walker Estimator

One of the well-known procedures for estimating the parameters of $AR(p)$ model is Yule-Walker (YW) method. YW estimator is defined as the solution of

$$\hat{\mathbf{a}}_{YW}(n) = \hat{\mathbf{V}}_p^{-1}\hat{\rho}(n) \tag{3.25}$$

where $\hat{\mathbf{a}}_{YW} = (\hat{a}_1, \ldots, \hat{a}_p)$ is the YW estimator of the coefficients of $AR(p)$ part in (3.22 , $\hat{\rho}_n = (\hat{\rho}(1), \ldots, \hat{\rho}(p))$ is the vector of the sample autocorrelation function where $\hat{\rho}(k)$ is given as (3.9) and $\hat{\mathbf{V}}_p = [\hat{\rho}(i-j)]_{i,j=1}^p$.

The asymptotic normality property of YW estimator in the classical case can be found in many time-series textbooks, e.g., in Brockwell and Davis (1987), Theorem 8.1.1 and Theorem 8.1.2. Because of its popularity, it is not surprising that the first method considered when estimating the autoregressive model in the infinite variance setting is the generalized Yule-Walker method. It is general in the sense it extends the relation (3.25) into the infinite variance model. The asymptotic properties of the estimates in α-stable case can be derived using the result of Theorem 3.2.2. In particular, since $\hat{\rho} \xrightarrow{p} \rho$ and $\hat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$ then using mean value theorem, we obtain

$$(n/\ln(n))^{1/\alpha}(\hat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} \mathbf{D}(Y_1, \ldots, Y_p)' \tag{3.26}$$

where $\mathbf{D}$ denotes the derivative of the vector function $\psi(\mathbf{z}) = \mathbf{V}_p(\mathbf{z})^{-1}\mathbf{z}$ and the distributions of the vector $Y_k, k = 1, \ldots, p$ are described in Theorem 3.2.2. Further discussion, and the simulation studies can be found in Adler *et al.* (1998*a*).

A similar generalized Yule-Walker estimator as (3.25) can be defined using the autocovariation function. A discussion for the estimation of $S\alpha S$ autoregressive process using the covariation function is presented in Gallagher (2001) and Nikias and Shao (1995), Section 7.2.1.

3.3.2 M-estimator

Suppose $X_1, \ldots, X_n$ are the data which are generated from ARMA model in (3.22) with true parameter vector $\beta_0 = (a_{01}, \ldots, a_{0p}, b_{01}, \ldots, b_{0q})'$. We define a M-estimate $\hat{\beta}_M$ of β_0 as any parameter vector which minimizes

$$\sum_{t=1}^{n} r(\epsilon_t(\beta)) \tag{3.27}$$

where $r(\cdot)$ is some suitably chosen loss function and $\{\epsilon_t(\beta)\}$ are estimates of the sequence $\{\epsilon_t\}$. Estimates of ϵ_t can be calculated for any particular vector of parameters $\beta = (a_1, \ldots, a_p, b_1, \ldots, b_q)'$ via

$$\begin{aligned}
&\epsilon_1(\beta) = X_1 \\
&\epsilon_2(\beta) = X_2 - a_1 X_1 - b_1\epsilon_1(\beta) \\
&\vdots \\
&\epsilon_n(\beta) = X_n - a_1 X_{n-1} - \cdots - a_p X_{n-p} - b_1\epsilon_{n-1}(\beta) - \cdots - b_q\epsilon_{n-q}(\beta)
\end{aligned} \tag{3.28}$$

Thus the objective is a function of the parameters through $\{\epsilon_t(\beta)\}$.

Several popular estimation procedures are the special cases of M-estimator. Least squares (LS) estimation has $r(x) = x^2$, and Least Absolute Deviation (LAD) has $r(x) = |x|$. Maximum likelihood Estimator (MLE) can be seen as a form of M-estimator with $r(x) = -\log p(x)$ where $p(\cdot)$ denotes the probability function of the underlying innovations. In the following theorem, we shortly present the asymptotic properties of M-estimator, as given in Davis (1996), following closely the presentation in Calder and Davis (1998).

First, we define a centered and rescaled parameter vector,

$$\mathbf{u} = a_n(\beta - \beta_0)$$

where the scaling constants $\{a_n\}$ are given by

$$a_n = \inf\left\{x : P(|\epsilon_1| > x) \leqslant n^{-1}\right\}$$

Then by substituting the vector $\mathbf{u}$ into (3.27) and centering, one obtains the process

$$W_n(\mathbf{u}) = \sum_{t=1}^{n} \left(r(\epsilon_t(\beta_0 + a_n^{-1}\mathbf{u})) - r(\epsilon_t(\beta_0))\right)$$

Minimizing $W_n(\mathbf{u})$ with respect to $\mathbf{u}$ is equivalent to minimize the original objective function (3.27) with respect to β. The asymptotic property of the estimator is summarized in the following theorem.

Theorem 3.3.1 (Davis (1996), Theorem 3.1.). *Suppose that $\{X_t\}$ is the ARMA(p,q) process (3.22) such that $\{\epsilon_t\}$ fulfils the condition C2. Assume that the loss function $r(x)$ is convex where its derivative $\psi(x)$ satisfying*

(a). *ψ is Lipschitz of order τ; i.e., $|\psi(x) - \psi(y)| \leqslant C\,|x - y|^{\tau}$ where $\tau > \max(\alpha - 1, 0)$ and C is a constant*

(b). *$E(\psi(\epsilon_1)) < \infty$ if $\alpha < 1$*

(c). *$E(\psi(\epsilon_1)) = 0$ and $\mathrm{var}(\psi(\epsilon_1)) < \infty$ if $\alpha \geqslant 1$*

then

$$n^{1/\alpha}(\beta_M - \beta) = \hat{\mathbf{u}}_n \xrightarrow{d} \hat{\mathbf{u}}$$

where $\hat{\beta}_M$ is the M-estimate of β. The limit random vector $\hat{\mathbf{u}}$ is the minimizer of a stochastic process, and typically has an intractable form.

In the following subsections, we present a short description of two special cases of the M-estimator.

Least square method

We start by presenting the result regarding the estimation of an autoregressive process. In pure autoregressive case $AR(p)$, the LS estimate is obtained by minimizing the objective function

$$\sum_{t=p}^{n}(X_t - (a_1X_{t-1} + \cdots + a_pX_{t-p}))^2$$

Then we might see that the LS estimate of $\mathbf{a} = (a_1, a_2, \ldots, a_p)^T$ is the solution of the following equation (assuming $\hat{\mathbf{C}}_{LS}$ is invertible)

$$\hat{\mathbf{C}}_{LS}\hat{\mathbf{a}}_{LS} = \hat{\mathbf{p}}_{LS} \tag{3.29}$$

where $\hat{\mathbf{C}}_{LS} = [\hat{\lambda}(i,j)]_{i,j=1,\ldots,p}$, $\hat{\mathbf{p}}_{LS} = [\hat{\lambda}(i)]_{i,j=1,\ldots,p}$ and

$$\hat{\lambda}(i,j) = \sum_{t=p}^{n} X_{t-j}X_{t-i} / \sum_{t=p}^{n} X_{t-i}^2$$

$$\hat{\lambda}(i) = \sum_{t=p}^{n} X_t X_{t-i} / \sum_{t=p}^{n} X_{t-i}^2$$

It can be shown that $\hat{\mathbf{a}}_{LS}$ is a strongly consistent estimator of the true coefficients. The order of convergence is given in the following theorem.

Theorem 3.3.2. *[Kanter and Steiger (1974), Yohai and Maronna (1977), Hannan and Kanter (1977)] Let n be the number of observations used in computing the least squares estimates as in (3.29). Then for any $\delta > \alpha$,*

$$\lim_{N\to\infty} n^{1/\delta}(\hat{a}_j - a_j) = 0 \text{ a.s. for } j = 1, 2, \ldots, p$$

For general ARMA model, one can expect to apply the result in Theorem 3.3.1 above. Unfortunately, under assumption C2 for $\{\epsilon_t\}$, the conditions of Theorem 3.3.1 are not fulfilled. However from Calder and Davis (1998), one obtains that

$$(\frac{n}{\ln n})^{1/\alpha}(\hat{\beta}_{LS} - \beta_0) \xrightarrow{d} \eta_{LS}$$

for some random vector η.

Least absolute deviation method

In the heavy-tailed autoregressive case, the least absolute deviation (LAD) estimator has been studied quite extensively, see, e.g., Bloomfield and Steiger (1983). In pure autoregressive case $AR(p)$, the LAD estimator of vector coefficients $\mathbf{a} = (a_1, a_2, \ldots, a_p)^T$ is obtained by minimizing the objective function

$$\sum_{t=p}^{N} |X_t - (a_1 X_{t-1} + \cdots + a_p X_{t-p})|$$

In the infinite variance case, it can be shown that the LAD estimator is a consistent estimator. Gross and Steiger (1979) show that if $\{\epsilon_t\}$ is i.i.d. α-stable and has unique median zero, then $\hat{\mathbf{a}}_{LAD}(n) \to \mathbf{a}$ almost surely as $n \to \infty$. Moreover, An and Chen (1982) show that in case of $1 < \alpha < 2$ and $\delta > \alpha$, $n^{1/\delta}(\hat{\mathbf{a}}_{LAD} - \mathbf{a}) \to 0$ in probability. Here the order of convergence of LAD estimates is comparable with the LS estimates. However, simulation studies in Bloomfield and Steiger (1983) and Gross and Steiger (1979) indicate that asymptotically LAD method will dominate LS. This conjecture is proved in Davis *et al.* (1992), where it is shown that

$$\frac{\|\hat{\mathbf{a}}_{LAD} - \mathbf{a}\|}{\|\hat{\mathbf{a}}_{LS} - \mathbf{a}\|} \xrightarrow{p} 0 \tag{3.30}$$

For general ARMA process, we expect to apply the result in Theorem 3.3.1 above. Unfortunately, under assumption C2 for $\{\epsilon_t\}$, the conditions of Theorem 3.3.1 are not fulfilled here. However, from Theorem 3.4. in Davis (1996), we obtain

$$n^{1/\alpha}(\beta_{LAD} - \beta) \xrightarrow{d} \eta_{LAD}$$

Comparing the rate of convergence of β_{LAD} to β_{LS}, in the ARMA case, the same conclusion as (3.30) can be obtained (Calder and Davis, 1998).

3.3.3 Summary of the other results

Besides the results presented in the previous subsections, there are some other important estimation methods for ARMA process with infinite variance that have been investigated in literature, such as the Whittle estimator (Mikosch *et al.*, 1995), the Gauss-Newton estimator (Davis, 1996), the empirical characteristic function (ECF) based estimator (Knight and Yu, 2002; Yu, 2004) and the Bayesian estimation method (Qiou and Ravishanker, 1998).

For estimation of pure autoregressive models, an approach based on Kalman filtering is considered in Stuck (1978). For $AR(1)$ with fixed parameter a where the noise follows $S\alpha S$ distribution with $1 < \alpha \leqslant 2$, using the minimum dispersion error criteria, Thavaneswaran and Abraham (1994) introduce a recursive type estimator for the parameter. This result is generalized by Thavaneswaran and Peiris (1999), in which they consider the penalized dispersion estimator, which is obtained by minimizing the absolute error between the estimator and the true values of the parameters. Their recursive estimator for time varying parameter of $AR(1)$ can be seen as a generalization of the result in Stuck and Kleiner (1974). Kuruoğlu *et al.* (1997) formulate L^p-norm minimization problem for estimating the parameters of $AR(p)$ model, and introduce reweighted least square model as the solution. They show that this estimator is an efficient method especially when short data length is available. Another related result for $AR(p)$ is presented in Jian and Pawitan (1994), in which they present the estimator for stable non causal $AR(p)$ process, and prove the consistency of the estimator in the heavy-tailed case. A linear programming approach for estimating the parameters of $AR(p)$ and also the asymptotic property of the estimator are discussed in, e.g., Feigin *et al.* (1994), where $\{\epsilon_t\}$ is assumed to follow a heavy-tailed distributed, but their values are always positive, which has different spirit with the process considered in this thesis. This type of process is often considered for modelling the teletraffic system.

Several estimation procedures have been developed for pure MA process with infinite variance. See for instance, Goryainova (1996) which presents a robust sign estimator type for parameters of MA model. Especially for non invertible MA(1) process with coefficient near or equal 1, Davis and Mikosch (1998) show that the rate of convergence of likelihood ratio estimator is $n^{1/2}$.

3.4 Diagnostic checking

The last stage in modelling is doing the diagnostic checking. This stage is done by examining the properties of the fitted residuals. When the estimated model fits the data, then the estimated residuals should resemble the true residuals, i.e., approximately an independent and identically symmetric α-stable distributed sequence. In the classical case, there are several diagnostic checking procedures available in literature, see, e.g., Section 9.4 of Brockwell and Davis (1987). Under the heavy-tailed setting, one can still apply the "heuristic" method to test the randomness of the fitted residuals. For instance, a simple graphical independency checking is possible by graphing the fitted residuals, although apparently as in the classical case, it is difficult to identify the correlation structure of the fitted residuals only from its graph. For the inference procedure in heavy-tailed setting, only limited results are available in literature, e.g., Krämer and Runde (2001). For "goodness of fit" of the stable distribution to the fitted residual, one can apply the methods presented in Section 2.3. In particular, here one may need to test the "stability" of $\hat{\alpha}$ obtained from the fitted residuals of different time horizons (e.g., Paolella, 2001).

Chapter 4

The codifference function

Let $\{X_t, t \in \mathbb{Z}\}$ be a strictly stationary symmetric α-stable process with $0 < \alpha \leq 2$. For Gaussian process, i.e., when $\alpha = 2$, the covariance function $\gamma(\cdot)$ completely describes the dependence structure over time for $\{X_t\}$. For $\alpha < 2$, the second moments of the process are infinite, and therefore, the population covariance function is not defined. In this chapter, we study the *codifference* function, and further consider its normalization, as the extension of the covariance and the correlation function, respectively. In Section 4.1, we present the definition of the codifference function, and discuss some of its basic properties. Section 4.2 describes the asymptotic properties of the codifference function of stable causal ARMA process. In Section 4.3 - 4.4, we propose estimators of the *codifference* and the *normalized codifference* function, which are defined via the empirical characteristic function. We show consistency of the proposed estimators, where the underlying model is stable causal ARMA with symmetric α-stable noise, $0 < \alpha \leq 2$. In addition, we show that the estimators are asymptotically normally distributed. The proofs of the asymptotic results will be presented in Section 4.5 - 4.7.

4.1 Definition of the codifference function

At least two different definitions of the codifference function have been proposed in literature. One way is by transforming the process $\{X_t\}$ into sequence $\{e^{isX_t}\}, s \in \mathbb{R}$ and defining a measure of dependence at lag k, $k \in \mathbb{Z}$ as $r(k) = \text{cov}(\exp(isX_{t+k}), \exp(isX_t))$, the covariance between $\exp(isX_{t+k})$ and $\exp(isX_t)$. Simple algebra gives

$$r(k) = r(s, -s; k) = \text{E}(\exp(is(X_{t+k} - X_t))) - \text{E}(\exp(isX_{t+k}))\text{E}(\exp(-isX_t)) \tag{4.1}$$

An alternative definition of the codifference function is considered in Yang *et al.* (2001)

$$\tau(k) = \tau(s, -s; k) = -\ln(\text{E}(\exp(is(X_{t+k} - X_t))))+\ln(\text{E}(\exp(isX_{t+k})))+\ln(\text{E}(\exp(-isX_t))) \tag{4.2}$$

where $s \in \mathbb{R}$ and $k \in \mathbb{Z}$. Both of the codifference function ((4.1) and (4.2)) require no moments conditions for the original process $\{X_t\}$; they are zero if the variables are independent and $k \neq 0$. Furthermore, note that if $\tau(k) \to 0$, when $k \to \infty$ then $|r(k) + K(s, -s)\tau(k)|$ for $k \to \infty$, i.e., $r(k)$ and $\tau(k)$ are asymptotically proportional. This relation can be obtain by noting that $r(k) = K(s, -s)(\exp(-\tau(k)) - 1)$ where $K(s, -s) = \text{E}(\exp(isX_{t+k}))\text{E}(\exp(-isX_t))$ is independent with k (Samorodnitsky and Taqqu, 1994).

In this thesis we only consider the codifference function as in (4.2). In the Gaussian case, the codifference function is proportional to the covariance function, i.e., $\tau(s, -s; k) = -s^2\gamma(k)$, where $\gamma(\cdot)$ denotes the covariance function of the stationary process $\{X_t\}$. Moreover, by defining the

normalized codifference function $I(k)$ as

$$I(k) = \frac{\tau(k)}{\tau(0)} = \frac{-\tau(k)}{-\tau(0)} \tag{4.3}$$

one directly obtains $I(k) = \rho(k)$ in the Gaussian case.

Note that in general $\tau(-k) = \tau(k)^*$, that is the conjugate of $\tau(k)$. However, for symmetric stationary process $\tau(-k) = \tau(k)$ holds. In particular, under assumptions C1 and C2 (see Section 3.1 in p. 23), we obtain that the codifference function $\tau(k)$ of linear process (3.1) is of the form (see Kokoszka and Taqqu (1994))

$$\begin{aligned}
\tau(k) &= -\ln \mathrm{E}e^{is(X_{t+k}-X_t)} + \ln \mathrm{E}e^{isX_{t+k}} + \ln \mathrm{E}e^{-isX_t} \\
&= -\ln \mathrm{E}e^{i(\sum_{j=0}^{\infty} sc_j\epsilon_{t+k-j} - \sum_{j=0}^{\infty} sc_j\epsilon_{t-j})} + \ln \mathrm{E}e^{i\sum_{j=0}^{\infty} sc_j\epsilon_{t+k-j}} + \ln \mathrm{E}e^{-i\sum_{j=0}^{\infty} sc_j\epsilon_{t-j}} \\
&= -\ln \mathrm{E}e^{i\left(\sum_{j=0}^{k-1} sc_j\epsilon_{t+k-j} + \sum_{j=0}^{\infty} s(c_{j+k}-c_j)\epsilon_{t-j}\right)} + \ln \mathrm{E}e^{i\sum_{j=0}^{\infty} sc_j\epsilon_{t+k-j}} + \ln \mathrm{E}e^{-i\sum_{j=0}^{\infty} sc_j\epsilon_{t-j}} \\
&= \sigma_\epsilon^\alpha \left[\sum_{j=0}^{k-1} |sc_j|^\alpha + \sum_{j=0}^{\infty} |s(c_{j+k}-c_j)|^\alpha - \sum_{j=0}^{\infty} |sc_j|^\alpha - \sum_{j=0}^{\infty} |-sc_j|^\alpha\right] \\
&= \sigma_\epsilon^\alpha \left[\sum_{j=0}^{\infty} (|s(c_{j+k}-c_j)|^\alpha - |sc_{j+k}|^\alpha - |-sc_j|^\alpha)\right] \\
&= \sigma_\epsilon^\alpha |s|^\alpha \left[\sum_{j=0}^{\infty} (|(c_{j+k}-c_j)|^\alpha - |c_{j+k}|^\alpha - |c_j|^\alpha)\right], k \geq 0
\end{aligned} \tag{4.4}$$

Remark 4.1.1. One could define more general codifference function than (4.1) and (4.2), as given in Hong (1999) and Kokoszka and Taqqu (1994), respectively. For $u, v \in \mathbb{R}$, a general version of (4.1) is

$$\tau_G(u, v; k) = \mathrm{E}(\exp(i(uX_{t+k} + vX_t))) - \mathrm{E}(\exp(iuX_{t+k}))\mathrm{E}(\exp(ivX_t)) \tag{4.5}$$

and where for (4.2) is

$$\tau_G(u, v; k) = -\ln(\mathrm{E}(\exp(i(uX_{t+k} + vX_t)))) + \ln(\mathrm{E}(\exp(iuX_{t+k}))) + \ln(\mathrm{E}(\exp(ivX_t))) \tag{4.6}$$

For our interest, definition (4.2) is more preferable. Note that under conditions C1 and C2, the normalized codifference function (4.3), using the codifference function as in (4.2), is independent of the choice of s (i.e., for given α, it is equal to $I(1, -1; k) = \tau(1, -1; k)/\tau(-1, 1; 0)$ for any choice of s), but a similar property can not be obtained using (4.6) except for $\alpha = 2$. It is important to note that for $S\alpha S$ process, $\tau(1, -1; k)$ coincides with the codifference function $u(k)$ as given in Samorodnitsky and Taqqu (1994), eq. (4.7.1). For given strictly stationary $S\alpha S$ process X_t, $u(k)$ is defined as

$$u(k) = 2(\sigma_{X_t})^\alpha - (\sigma_{X_{t+k}-X_t})^\alpha \tag{4.7}$$

where σ_Z and σ_{Y-Z} denote the scale parameters of Z and $Y - Z$, respectively. Expression for σ_{Y-Z} of jointly stable random vector (Y, Z) can be obtained from eq. (2.11). See also Theorem 2.2.13.

Remark 4.1.2. Notice that if the $S\alpha S$ stationary process X_t is independent, then for $k \neq 0$, $u(k) = 0$, or equivalently, $\tau(k)$ are zero for all s. Conversely, if $u(k) = 0, k \neq 0$ and $0 < \alpha < 1$, then X_t is independent. To obtain this property, we can use Example 2.2.15. From this example. we obtain that if X_t is independent, then $s_1 s_2 = 0, \Gamma_{X_1,X_2}$ *a.e.*, where Γ_{X_1,X_2} is the spectral

measure of (X_1, X_2). Hence

$$\begin{aligned}
&(\sigma_{X_{t+k}-X_t})^\alpha \\
&= \int_{S_2} |s_1 - s_2|^\alpha \Gamma_{X_{t+k},X_t}(ds) \\
&= \int_{S_2} |s_1|^\alpha \Gamma_{X_{t+k},X_t}(ds) + \int_{S_2} |s_2|^\alpha \Gamma_{X_{t+k},X_t}(ds) \quad (s_1 s_2 = 0, \Gamma_{X_{t+k},X_t} \text{ a.e.}) \\
&= (\sigma_{X_{t+k}})^\alpha + (\sigma_{X_t})^\alpha
\end{aligned}$$

i.e., $u(k) = 0$.

Further, if $0 < \alpha < 1$, then $|s_1 - s_2|^\alpha \leq |s_1|^\alpha + |s_2|^\alpha$ with equality only when $s_1 s_2 = 0$. This can be obtained as follows. Since $(|a + b|)^\alpha \leq (|a| + |b|)^\alpha$, we may suppose $a \geq 0$ and $b \geq 0$. If $0 < \alpha < 1$, then, for fixed $b > 0$ and for any $a \geq 0$, we have $g_b(a) = a^\alpha + b^\alpha - (a + b)^\alpha \geq 0$ with equality only at $a = 0$, because $g_b(0) = 0$ and $g_b'(a) > 0$ for $a > 0$.

From the preceding computation we obtain that if $u(k) = 0$, then $s_1 s_2 = 0, \Gamma_{X_{t+k},X_t}$ -a.e., and so X_{t+k} and X_t are independent.

When $1 \leq \alpha < 2$, $u(k) = 0$ does not imply that X_{t+k} and X_t are independent. For instance, choose $\Gamma_{X_{t+k},X_t} = \delta((1,0)) + \delta((-1,0))$, then $s_1 s_2 = 0, \Gamma_{X_{t+k},X_t}$ -a.e. but X_{t+k} and X_t are not independent according to Example 2.2.15.

Remark 4.1.3. Using Property 2.10.5 in Samorodnitsky and Taqqu (1994), we obtain that $u(k)$ of $S\alpha S$ process has the following properties:

$$u(k) \leq 2(\sigma_{X_t})^\alpha$$

and

$$u(k) \geq \begin{cases} 0 & \text{if } 0 < \alpha \leq 1 \\ 2(1 - 2^{\alpha-1})\sigma_{X_t}^\alpha & \text{if } 1 \leq \alpha \leq 2 \end{cases}$$

The upper bound is achieved at lag $k = 0$. For $0 < \alpha \leq 1$, the lower bound is achieved when X_t is an independent process, and when $X_{t+k} = -X_t$ if $1 \leq \alpha \leq 2$. From this relation, for the normalized codifference $I(k)$, we obtain

$$0 \leq I(k) \leq 1 \text{ if } 0 < \alpha \leq 1 \tag{4.8}$$

$$1 - 2^{\alpha-1} \leq I(k) \leq 1 \text{ if } 1 \leq \alpha \leq 2 \tag{4.9}$$

When $\alpha = 2$, (4.9) is equal to $-1 \leq \rho(k) \leq 1$, where $\rho(k)$ denotes the ACF.

Remark 4.1.4. Following Kokoszka and Taqqu (1994), we can consider the following interpretation of the codifference function (4.2). Consider two stationary symmetric α-stable sequences $\{X_t\}$ and $\{Y_t\}$ with identical scaling parameters equal to σ_X, and suppose that for some k

$$-\tau(k)^X \leq -\tau(k)^Y \tag{4.10}$$

If

$$\begin{aligned}
\xi_k &= -\ln(\mathrm{E}(\exp(is(X_{t+k} - X_t)))) \\
\nu_k &= -\ln(\mathrm{E}(\exp(is(Y_{t+k} - Y_t))))
\end{aligned}$$

then (4.10) becomes $2\sigma_X^\alpha - \xi_k \leq 2\sigma_X^\alpha - \nu_k$, or

$$\xi_k \geq \nu_k \tag{4.11}$$

Since $\xi_k^{-1}(X_{t+k} - X_t)$ and $\nu_k^{-1}(Y_{t+k} - Y_t)$ have the same distribution, we get, for any $c > 0$,

$$\begin{aligned}
\mathrm{P}(|X_{t+k} - X_t| > c) &= \mathrm{P}(\xi_k^{-1}|X_{t+k} - X_t| > \xi_k^{-1}c) \\
&= \mathrm{P}(\nu_k^{-1}|Y_{t+k} - Y_t| > \xi_k^{-1}c) \\
&= \mathrm{P}(|Y_{t+k} - Y_t| > \nu_k \xi_k^{-1}c) \\
&\geq \mathrm{P}(|Y_{t+k} - Y_t| > c)
\end{aligned}$$

The inequality $P(|X_{t+k} - X_t| > c) \geq P(|Y_{t+k} - Y_t| > c)$ means that Y_{t+k} and Y_t are less likely to differ than X_{t+k} and X_t and so are "more dependent". Thus, the larger τ, the "greater" the dependence.

Remark 4.1.5. Another (generalized) dependence measure proposed in literature is called the *dynamical function*. The dynamical function at lag k, DF(k), is defined as

$$\mathrm{DF}(k) = \mathrm{DF}(X_k, X_0) = \mathrm{E}(\exp(is(X_k - X_0))), s \in \mathbb{R} \tag{4.12}$$

One also can define more general dynamical function (c.f., Janicki and Weron (1994*b*)) as

$$\mathrm{DF}(n; \theta_1, \theta_2) = \mathrm{E}(\exp(i(\theta_1 X_n + \theta_2 X_0))), \theta_1, \theta_2 \in \mathbb{R} \tag{4.13}$$

Under assumptions C1 and C2, the dynamical function (4.12) is

$$\mathrm{DF}(k) = \mathrm{E}(\exp(is(X_{t+k} - X_t))) = \exp(-\sigma_\epsilon^\alpha \sum_{j=0}^{k-1} |sc_j|^\alpha - \sigma_\epsilon^\alpha \sum_{j=0}^{\infty} |s(c_{j+k} - c_j)|^\alpha)$$

The following relation between the dynamical function (4.12) and the codifference function (4.2) of the stationary process holds for every $k \in \mathbb{Z}$

$$\mathrm{DF}(k) = \exp(\tau(k) - \tau(0)) \tag{4.14}$$

The relation (4.14) also holds for (4.13) and (4.6).

4.2 The asymptotic behavior of the codifference function

First we recall from Theorem 3.1.1 that the $ARMA(p, q)$ process,

$$A(z)X_t = B(z)\epsilon_t \tag{4.15}$$

has unique stationary solution of the form

$$X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, t \in \mathbb{Z}, \text{ a.s.} \tag{4.16}$$

which fulfils conditions C1 and C2 if and only if the polynomial $A(z)$ has no roots in the closed unit disk $\{z : |z| \leq 1\}$. The c_j's in (4.16) are the coefficients in the series expansion of $B(z)/A(z), |z| < 1$. Here z denotes backward-shift operator (that is $z(X_t) = (X_{t-1})$), as well as the complex variable, and the polynomials A and B are

$$A(z) = 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p \tag{4.17}$$

$$B(z) = b_0 + b_1 z + b_2 z^2 + \cdots + b_q z^q \tag{4.18}$$

for fixed order (p, q).

In this section, the asymptotic behavior of the codifference function of ARMA processes (4.15) will be summarized based on the results in Kokoszka and Taqqu (1994). We first show in Theorem 4.2.1 that the codifference function of linear processes (4.16), which satisfy conditions C1 and C2, is bounded by an exponentially decaying function, like the covariance function. In other words, the sequence $\{\tau(k)\}$ of every stable causal ARMA process tends to zero exponentially fast. We further describe the exact asymptotic behavior of the codifference function by analyzing the coefficients c_j's in (4.16), based on the properties of roots of $A(z)$. Here we consider three classes of the roots: real positive, real negative and complex, i.e., we slightly extend the analysis, which is carried out in Kokoszka and Taqqu (1994). Note that Kokoszka and Taqqu (1994) consider the codifference function as (4.6), which is more general than (4.2), but contained (4.2) as the special case (when $u = s, v = -s, s \in \mathbb{R}$, see Remark 4.1.1).

Remark 4.2.1. Since the relation (4.14) holds for every $k \in \mathbb{Z}$, the information about the asymptotic behavior of the dynamical function can be obtained from the asymptotic properties of the codifference function.

Theorem 4.2.1 (Kokoszka and Taqqu (1994), Theorem 2.1.). *Consider the linear process (4.16), which fulfils the condition C2. If the coefficients c_j's of linear process (4.16) satisfy the condition C1, then*

$$\limsup_{k\to\infty} Q^{\alpha k} |\tau(k)| \leq K_1 |s|^\alpha \quad \text{for } 0 < \alpha \leq 1 \tag{4.19}$$

and

$$\limsup_{k\to\infty} Q^{k} |\tau(k)| \leq K_2 |s| \, |-s|^{\alpha-1} \quad \text{for } 1 < \alpha \leq 2 \tag{4.20}$$

where K_1 and K_2 are constants depending on α, Q and c_j's but not on s. Q is as given in p. 23.

The proof of this theorem depends on the following lemma

Lemma 4.2.2. *Consider the linear process (4.16), which fulfils the condition C2, and define*

$$\psi_k = (\sum_{j=k}^{\infty} |c_j|^\alpha)^{1/\alpha}$$

Then

$$\limsup_{k\to\infty} \frac{|\tau(k)|}{\psi_k^\alpha} \leq 2\sigma^\alpha |s|^\alpha \quad \text{for } 0 < \alpha \leq 1 \tag{4.21}$$

and

$$\limsup_{k\to\infty} \frac{|\tau(k)|}{\psi_k} \leq \alpha\sigma^\alpha \psi_0^{\alpha-1} |s| \, |-s|^{\alpha-1} \quad \text{for } 1 < \alpha \leq 2 \tag{4.22}$$

Proof. From (4.4), we obtain

$$\tau(k) = \sigma^\alpha \sum_{j=0}^{\infty} (|s(c_{j+k} - c_j)|^\alpha - |sc_{j+k}|^\alpha - |-sc_j|^\alpha) \tag{4.23}$$

Following Kokoszka and Taqqu (1994), we obtain the following inequalities which hold for any real numbers a and b

$$||a+b|^\alpha - |a|^\alpha - |b|^\alpha| \leq \begin{cases} 2|a|^\alpha & \text{for } 0 < \alpha \leq 1 \\ \alpha |a| \, |b|^{\alpha-1} + (\alpha+1)|a|^\alpha & \text{for } 1 < \alpha \leq 2 \end{cases} \tag{4.24}$$

If $0 < \alpha \leq 1$, then after applying (4.24) to each term in (4.23), with $a = sc_{j+k}$, we get

$$|\tau(k)| \leq 2\sigma^\alpha |s|^\alpha \psi_k^\alpha \tag{4.25}$$

If $1 < \alpha \leq 2$, then

$$\begin{aligned} |\tau(k)| &\leq \sigma^\alpha \sum_{j=0}^{\infty} \left\{ \alpha |sc_{j+k}| \, |-sc_j|^{\alpha-1} + (\alpha+1)|sc_{j+k}|^\alpha \right\} \\ &\leq \alpha\sigma^\alpha |s| \, |-s|^{\alpha-1} \left(\sum_{j=0}^{\infty} |c_{j+k}|^\alpha \right)^{1/\alpha} \left(\sum_{j=0}^{\infty} |c_j|^\alpha \right)^{(\alpha-1)/\alpha} + (\alpha+1)\sigma^\alpha |s|^\alpha \sum_{j=0}^{\infty} |c_{j+k}|^\alpha \end{aligned} \tag{4.26}$$

by Hölder Inequality. Relation (4.19) and (4.20) follow from (4.25) and (4.26) respectively. □

Proof of Theorem 4.2.1. Suppose k is such that $|c_j| < Q^{-j}$ for each $j > k$. Then

$$Q^k \phi_k = Q^k \left(\sum_{j=k}^{\infty} |c_j|^\alpha \right)^{1/\alpha} \leq Q^k \left(\sum_{j=k}^{\infty} Q^{-j\alpha} \right)^{1/\alpha} = (1 - Q^{-\alpha})^{-1/\alpha} \tag{4.27}$$

Applying Lemma 4.2.2, we obtain (4.20) and (4.19) with $K_1 = 2\sigma^\alpha (1 - Q^{-\alpha})^{-1}$ and $K_2 = \alpha \sigma^\alpha \psi_k^{\alpha-1} (1 - Q^{-\alpha})^{-1/\alpha}$ □

In Kokoszka and Taqqu (1994), it was shown that the asymptotic behavior of the sequence $\{\tau(k)\}$ of the ARMA process depends on α and the roots of the polynomials $A(z)$. In order to show this result, first we will determine the exact form of the coefficient c_j. As the polynomial $A(z)$ and $B(z)$ do not have common roots, i.e., the polynomial $\frac{B(z)}{A(z)}$ is reduced, we can write

$$\frac{B(z)}{A(z)} = P(z) + \frac{G(z)}{A(z)} \tag{4.28}$$

where $P(z) = p_0 + p_1 z + \cdots + p_{q-p} z^{q-p}$ (the p_i's real) if $q \geq p$ and $P \equiv 0$ if $q < p$. $G(z)$ is the polynomial of degree less than p. As in Kokoszka and Taqqu (1994), we assume here the different roots of $A(z)$ have different moduli. Since $A(z)$ has no roots in the closed unit circle $\{z : |z| \leq 1\}$, each root of $A(z)$ belongs to one of the following three classes (Nowicka, 1997):

1. The real positive roots:

   $$\exp(\kappa_1), \ldots, \exp(\kappa_s) \quad 0 < \kappa_1 < \cdots < \kappa_s$$

   with multiplicities $l_1, \ldots, l_s$

2. The real negative roots:

   $$-\exp(\nu_1), \ldots, -\exp(\nu_r) \quad 0 < \nu_1 < \cdots < \nu_r$$

   with multiplicities $g_1, \ldots, g_r$

3. The remaining complex roots:

   $$\exp(\lambda_1 \pm i\mu_1), \ldots, \exp(\lambda_t \pm i\mu_t), \quad 0 < \lambda_1 < \cdots < \lambda_t, \mu_j \in (0, 2\pi) \text{ for } j = 1, \ldots, t$$

   with multiplicities $m_1, \ldots, m_t$

Apparently, it is required that $l_1 + \cdots + l_s + g_1 + \cdots + g_r + 2(m_1 + \cdots + m_t) = p$. Since $G(z)/A(z)$ is a proper fraction, there are real numbers $d_{11}, \ldots, d_{1l_1}, \ldots, d_{s1}, \ldots, d_{sl_s}, f_{11}, \ldots, f_{1g_1}, \ldots, f_{r1}, \ldots, f_{rg_r}$ and complex numbers $e_{11}, \ldots, e_{1m_1}, \ldots, e_{t1}, \ldots, e_{tm_t}$ such that

$$\begin{aligned}
\frac{G(z)}{A(z)} = {} & \frac{d_{11}}{e^{\kappa_1} - z} + \cdots + \frac{d_{1l_1}}{(e^{\kappa_1} - z)^{l_1}} + \cdots + \frac{d_{s1}}{e^{\kappa_s} - z} + \cdots + \frac{d_{sl_s}}{(e^{\kappa_s} - z)^{l_s}} \\
& + \frac{f_{11}}{-e^{\nu_1} - z} + \cdots + \frac{f_{1g_1}}{(-e^{\nu_1} - z)^{g_1}} + \cdots + \frac{f_{r1}}{-e^{\nu_r} - z} + \cdots + \frac{f_{rg_r}}{(-e^{\nu_r} - z)^{g_r}} \\
& + \frac{e_{11}}{e^{\lambda_1 + i\mu_1} - z} + \cdots + \frac{e_{1m_1}}{(e^{\lambda_1 + i\mu_1} - z)^{m_1}} + \cdots + \frac{e_{t1}}{e^{\lambda_t + i\mu_t} - z} + \cdots + \frac{e_{tm_t}}{(e^{\lambda_t + i\mu_t} - z)^{m_t}} \\
& + \frac{\overline{e_{11}}}{e^{\lambda_1 - i\mu_1} - z} + \cdots + \frac{\overline{e_{1m_1}}}{(e^{\lambda_1 - i\mu_1} - z)^{m_1}} + \cdots + \frac{\overline{e_{t1}}}{e^{\lambda_t - i\mu_t} - z} + \cdots + \frac{\overline{e_{tm_t}}}{(e^{\lambda_t - i\mu_t} - z)^{m_t}}
\end{aligned} \tag{4.29}$$

For any natural number l and any complex number $a, |a| > 1$

$$\begin{aligned}\frac{1}{(a-z)^l} &= \left(\frac{a^{-1}}{1-a^{-1}z}\right)^l = \left(a^{-1}\sum_{j=0}^{\infty} a^{-j}z^j\right)^l \\ &= a^{-l}\sum_{j_1,\cdots,j_l=0}^{\infty} a^{-(j_1+\cdots+j_l)}z^{j_1+\cdots+j_l} \\ &= a^{-l}\sum_{j=0}^{\infty}\binom{j+l-1}{j}a^{-j}z^j \end{aligned} \tag{4.30}$$

Using (4.30), we obtain the coefficient $\tilde{c}_j$ in the series expansion $G(z)/A(z), |z| < 1$ which is given by

$$\begin{aligned}\tilde{c}_j &= D_1(j)e^{-\kappa_1 j} + \cdots + D_s(j)e^{-\kappa_s j} + F_1(j)(-e^{\nu_1})^{-j} + \cdots + F_r(j)(-e^{\nu_r})^{-j} \\ &\quad + E_1(j)e^{-(\lambda_1+i\mu_1)j} + \cdots + E_t(j)e^{-(\lambda_t+i\mu_t)j} + \overline{E_1(j)}e^{-(\lambda_1-i\mu_1)j} + \cdots + \overline{E_t(j)}e^{-(\lambda_t-i\mu_t)j} \\ &= D_1(j)e^{-\kappa_1 j} + \cdots + D_s(j)e^{-\kappa_s j} + F_1(j)(-1)^j e^{-\nu_1 j} + \cdots + F_r(j)(-1)^j e^{-\nu_r j} \\ &\quad + 2Re(E_1(j)e^{-i\mu_1 j})e^{-\lambda_1 j} + \cdots + 2Re(E_t(j)e^{-i\mu_t j})e^{-\lambda_t j}\end{aligned}$$

where

$$D_k(j) = \sum_{l=1}^{l_k} d_{kl}\exp(-\kappa_k l)\binom{j+l-1}{j}, k = 1,\ldots,s \tag{4.31}$$

$$F_k(j) = \sum_{g=1}^{g_k} f_{kg}(-1)^g\exp(-v_k g)\binom{j+g-1}{j}, k = 1,\ldots,r \tag{4.32}$$

$$E_k(j) = \sum_{m=1}^{m_k} e_{km}\exp(-(\lambda_k+i\mu_k)m)\binom{j+m-1}{j}, k = 1,\ldots,t \tag{4.33}$$

Note that here we use the relation

$$E_k(j)e^{-(\lambda_k+i\mu_k)j} + \overline{E_k(j)}e^{-(\lambda_k-i\mu_k)j} = 2e^{-\lambda_k j}\,\mathrm{Re}\{E_k(j)e^{-i\mu_k j}\}$$

Denote the numbers $\kappa_1,\ldots,\kappa_s,\nu_1,\ldots,\nu_r,\lambda_1,\ldots,\lambda_t$ as $\omega_1,\ldots,\omega_N$, $N = s+r+t$ in such a way that $0 < \omega_1 < \cdots < \omega_N$. Thus $\tilde{c}_j = \sum_{m=1}^{N} H_m(j)\exp(-\omega_m j)$, where for $m = 1,\ldots,N$

$$H_m(j) = \begin{cases} D_l(j) & \text{if } \omega_m \text{ is } l\text{-th element of the } \kappa\text{s} \\ (-1)^j F_l(j) & \text{if } \omega_m \text{ is } l\text{-th element of the } \nu\text{s} \\ 2Re(E_l(j)e^{-i\mu_m j}) & \text{if } \omega_m \text{ is } l\text{-th element of the } \lambda\text{s} \end{cases} \tag{4.34}$$

Therefore, from (4.28), we can write coefficient c_j as,

$$c_j = p_j + \tilde{c}_j = p_j + \sum_{m=1}^{N} H_m(j)\exp(-\omega_m j) \tag{4.35}$$

with understanding that $p_j = 0$ if $j > q-p$.

Below, we note some remarks. First, set $\gamma = l_1$ if $\omega_1 = \kappa_1$, $\gamma = g_1$ if $\omega_1 = \nu_1$ and $\gamma = m_1$ if $\omega_1 = \lambda_1$. From Kokoszka and Taqqu (1994), we have the following facts. For any natural numbers l and γ satisfying $1 \le l \le \gamma$, the following relation holds:

$$\forall j \ge 0 \lim_{k\to\infty}\binom{j+k+l-1}{j+k}\frac{1}{k^{\gamma-1}} = \begin{cases} [(l-1)!]^{-1} & \text{if } l = \gamma \\ 0 & \text{if } l < \gamma \end{cases} \tag{4.36}$$

$$\forall j \geq 0 \; \forall k \geq 4 \quad \binom{j+k+l-1}{j+k} \frac{1}{k^{\gamma-1}} < 1 \tag{4.37}$$

The following formulation will be used in subsequent proofs. First we recall from (4.4) that

$$\tau(k) = \sigma^\alpha \sum_{j=0}^{\infty} v_j(k)$$

where

$$v_j(k) = |s(c_{j+k} - c_j)|^\alpha - |sc_{j+k}|^\alpha - |-sc_j|^\alpha \tag{4.38}$$

We will rewrite $v_j(k)$ in more convenient form. Assume $k > p - q$ (and therefore $p_{j+k} = 0$ in (4.35)) and let $\sum_l$ stand for $\sum_{l=1}^N$, $N = s + r + t$. Set also $z^{<p>} = |z|^{p-1}\bar{z}$ for any complex number z and real number p. Then, as $n \to \infty$,

$$v_j(k) = \alpha(-s)^{<\alpha-1>} sc_j^{<\alpha-1>} c_{j+k} + o(c_{j+k}) - |sc_{j+k}|^\alpha \tag{4.39}$$

Indeed, from (4.38)

$$\begin{aligned} v_j(k) &= |-sc_j|^\alpha \left\{ \left| 1 + \frac{sc_{j+k}}{-sc_j} \right|^\alpha - 1 \right\} - |sc_{j+k}|^\alpha \\ &= |-sc_j|^\alpha \left\{ \frac{\alpha sc_{j+k}}{-sc_j} + o(c_{j+k}) \right\} - |sc_{j+k}|^\alpha \end{aligned}$$

yielding (4.39). By substituting (4.35) into (4.39), we obtain

$$\begin{aligned} v_j(k) = \alpha(-s)^{<\alpha-1>} s \left\{ p_j + \sum_l H_l(j) \exp(-\omega_l j) \right\}^{<\alpha-1>} \left[\sum_l H_l(j+k) \exp(-\omega_l(j+k)) \right] \\ + o\left[\sum_l H_l(j+k) \exp(-\omega_l(j+k)) \right] - \left| s \sum_l H_l(j+k) \exp(-\omega_l(j+k)) \right|^\alpha \end{aligned} \tag{4.40}$$

The following theorem presents the asymptotic behavior of the codifference function for $\alpha < 1$. Its form is determined by the "smallest" root of $A(z)$. If the "smallest" root is real and positive, $\omega_1 = \kappa_1$, then $\tau(k)$ is asymptotically proportional to $k^{(l_1-1)\alpha} \exp(-\kappa_1 \alpha k)$. When the "smallest" root is not positive, i.e., $\omega_1 = \nu_1$ or $\omega_1 = \lambda_1$, then $\tau(k)$ is oscillating, but $|\tau(k)|$ is asymptotically bounded from above by a function proportional to $k^{(g_1-1)\alpha} \exp(-\nu_1 \alpha k)$ or $k^{(m_1-1)\alpha} \exp(-\lambda_1 \alpha k)$, respectively.

Theorem 4.2.3. *Suppose $0 < \alpha < 1$ and consider the ARMA sequence (4.15), then*

(i). *If $\omega_1 = \kappa_1$ then*

$$\lim_{k\to\infty} k^{-(l_1-1)\alpha} \exp(\kappa_1 \alpha k) \tau(k) = \left| \frac{s d_{1l_1} \exp(-\kappa_1 l_1)}{(l_1-1)!} \right|^\alpha \frac{-\sigma^\alpha}{1 - \exp(-\kappa_1 \alpha)} \tag{4.41}$$

(ii). *If $\omega_1 = \nu_1$ then*

$$\limsup_{k\to\infty} k^{-(g_1-1)\alpha} \exp(\nu_1 \alpha k) \, |\tau(k)| \leq \left| \frac{s f_{1g_1} \exp(-\nu_1 g_1)}{(g_1-1)!} \right|^\alpha \frac{\sigma^\alpha}{1 - \exp(-\nu_1 \alpha)} \tag{4.42}$$

(iii). *If $\omega_1 = \lambda_1$ then*

$$\limsup_{k\to\infty} k^{-(m_1-1)\alpha} \exp(\lambda_1 \alpha k) \, |\tau(k)| \leq \left| \frac{s e_{1m_1} \exp(-\lambda_1 m_1)}{(m_1-1)!} \right|^\alpha \frac{2\sigma^\alpha}{1 - \exp(-\lambda_1 \alpha)} \tag{4.43}$$

Proof. First, set $\gamma = l_1$ if $\omega_1 = \kappa_1$, $\gamma = g_1$ if $\omega_1 = \nu_1$ and $\gamma = m_1$ if $\omega_1 = \lambda_1$. From (4.40)

$$\begin{aligned} k^{-(\gamma-1)\alpha}\exp(\omega_1\alpha k)v_j(k) &= \alpha(-s)^{<\alpha-1>}s\left\{p_j+\sum_l H_l(j)\exp(-\omega_l j)\right\}^{<\alpha-1>} \\ &\quad\times\left[\sum_l k^{-(\gamma-1)\alpha}H_l(j+k)\exp(-\omega_l j)\exp(-(\omega_l-\omega_1\alpha)k)\right] \\ &\quad+k^{-(\gamma-1)\alpha}\exp(\omega_1\alpha k)\ o\left[\sum_l H_l(j+k)\exp(-\omega_l(j+k))\right] \\ &\quad-\left|s\sum_l k^{-(\gamma-1)}H_l(j+k)\exp(-\omega_l j)\exp(-(\omega_l-\omega_1)k)\right|^{\alpha} \end{aligned} \tag{4.44}$$

Since $\alpha < 1$, $\omega_l - \alpha\omega_1 > 0$ and so the first two terms of (4.44) tend exponentially to zero as $k \to \infty$. For the last term, only the first term of the sum, namely $k^{-(\gamma-1)}H_1(j+k)\exp(-\omega_1 j)$ does not tend to zero. If $\omega_1 = \kappa_1$, then by (4.36),

$$\begin{aligned} \lim_{k\to\infty} k^{-(\gamma-1)}H_1(j+k) &= \lim_{k\to\infty} k^{-(l_1-1)}D_1(j+k) \\ &= d_{1l_1}\exp(-\kappa_1 l_1)\{(l_1-1)!\}^{-1} \end{aligned} \tag{4.45}$$

If $\omega_1 = \nu_1$, then $\lim_{k\to\infty} k^{-(g_1-1)}H_1(j+k)$ does not exist, because using (4.36), there are two subsequences that give two different limits, i.e.,

$$\lim_{k\to\infty} k^{-(g_1-1)}H_1(j+k) = -\frac{(-1)^{g_1+j}f_{1g_1}\exp(-\nu_1 g_1)}{(g_1-1)!} \text{ for odd } k \tag{4.46}$$

and

$$\lim_{k\to\infty} k^{-(g_1-1)}H_1(j+k) = \frac{(-1)^{g_1+j}f_{1g_1}\exp(-\nu_1 g_1)}{(g_1-1)!} \text{ for even } k \tag{4.47}$$

However,

$$\begin{aligned} \limsup_{k\to\infty} k^{-(\gamma-1)}|H_1(j+k)| &= \limsup_{k\to\infty} k^{-(g_1-1)}|F_1(j+k)| \\ &= |f_{1g_1}|\exp(-\nu_1 g_1)\{(g_1-1)!\}^{-1} \end{aligned} \tag{4.48}$$

If $\omega_1 = \lambda_1$ and $\mu_1 = \frac{x}{y}\pi$ (for any $x, y \in \mathbb{R}$ such that $x, y > 0, x \neq y$ and $x < 2y$), then $\lim_{k\to\infty} k^{-(m_1-1)}H_1(j+k)$ does not exist (Nowicka, 1997), because there are

$$M = \frac{2y}{\gcd(x, 2y)}$$

separate subsequences of k's that give different limits. However,

$$\begin{aligned} \limsup_{k\to\infty} k^{-(\gamma-1)}|H_1(j+k)| &\leq \limsup_{k\to\infty} 2k^{-(m_1-1)}|E_1(j+k)| \\ &= 2|e_{1m_1}|\exp(-\lambda_1 m_1)\{(m_1-1)!\}^{-1} \end{aligned} \tag{4.49}$$

Therefore, if $\omega_1 = \kappa_1$,

$$\lim_{k\to\infty} k^{-(\gamma-1)\alpha}\exp(\omega_1\alpha k)v_j(k) = \left|\frac{sd_{1l_1}\exp(-\kappa_1 l_1)}{(l_1-1)!}\right|^{\alpha}\exp(-\kappa_1\alpha j) \tag{4.50}$$

If $\omega_1 = \nu_1$, then

$$\limsup_{k\to\infty} k^{-(\gamma-1)\alpha}\exp(\omega_1\alpha k)|v_j(k)| = \left|\frac{sf_{1g_1}\exp(-\nu_1 g_1)}{(g_1-1)!}\right|^{\alpha}\exp(-\nu_1\alpha j) \tag{4.51}$$

And if $\omega_1 = \lambda_1$, then

$$\limsup_{k\to\infty} k^{-(\gamma-1)\alpha}\exp(\omega_1\alpha k)|v_j(k)| \le 2\left|\frac{se_{1m_1}\exp(-\lambda_1 m_1)}{(m_1-1)!}\right|^{\alpha}\exp(-\lambda_1\alpha j) \tag{4.52}$$

By (4.37), there is a constant K_1 depending only on the coefficients of polynomials $A(z)$ and $B(z)$ and is independent of k and j such that

$$\forall j,k \quad \left|s\sum_l k^{-(\gamma-1)}H_l(j+k)\exp(-(\omega_l-\omega_1)k)\exp(-(\omega_l-\omega_1)j)\right| \le K_1 \tag{4.53}$$

Therefore, using (4.24) in any case,

$$\begin{aligned}|k^{-(\gamma-1)\alpha}e^{\omega_1\alpha k}v_j(k)| &\le 2\left|s\sum_l k^{-(\gamma-1)}H_l(j+k)e^{-(\omega_l-\omega_1)j}e^{-(\omega_l-\omega_1)k}\right|^{\alpha}e^{-\omega_1\alpha j}\\ &\le |s|^{\alpha}2K_1^{\alpha}\exp(-\omega_1\alpha j)\end{aligned} \tag{4.54}$$

Thus, using the dominated convergence theorem when $\omega_1 = \kappa_1$ and a modification of Fatou's lemma if $\omega_1 = \nu_1$ and $\omega_1 = \lambda_1$, we obtain (4.41), (4.42) and (4.43). □

In the following theorem we focus on the case of $1 < \alpha \le 2$. For this range of α, if the "smallest" root is real and positive, $\omega_1 = \kappa_1$, then $\tau(k)$ is asymptotically proportional to $k^{(l_1-1)}\exp(-\kappa_1 k)$. When the "smallest" root is not positive, i.e., $\omega_1 = \nu_1$ or $\omega_1 = \lambda_1$, then $\tau(k)$ is oscillating, but $|\tau(k)|$ is asymptotically bounded from above by a function proportional to $k^{(g_1-1)}\exp(-\nu_1 k)$ or $k^{(m_1-1)}\exp(-\lambda_1 k)$, respectively.

Theorem 4.2.4. *Suppose $1 < \alpha \le 2$ and consider the ARMA sequence (4.15), then*

(i). *If $\omega_1 = \kappa_1$ then*

$$\lim_{k\to\infty} k^{-(l_1-1)}\exp(\kappa_1 k)\tau(k) = \frac{\alpha\sigma^{\alpha}(-s)^{\langle\alpha-1\rangle}sd_{1l_1}\exp(-\kappa_1 l_1)}{(l_1-1)!}\sum_{j=0}^{\infty}c_j^{\langle\alpha-1\rangle}\exp(-\kappa_1 j) \tag{4.55}$$

(ii). *If $\omega_1 = \nu_1$ then*

$$\limsup_{k\to\infty} k^{-(g_1-1)}\exp(\nu_1 k)|\tau(k)| \le \frac{\alpha\sigma^{\alpha}|-s|^{\alpha-1}|s||f_{1g_1}|\exp(-\nu_1 g_1)}{(g_1-1)!}\sum_{j=0}^{\infty}|c_j|^{\alpha-1}\exp(-\nu_1 j) \tag{4.56}$$

(iii). *If $\omega_1 = \lambda_1$ then*

$$\limsup_{k\to\infty} k^{-(m_1-1)}\exp(\lambda_1 k)|\tau(k)| \le \frac{2\alpha\sigma^{\alpha}|-s|^{\alpha-1}|s||e_{1m_1}|\exp(-\lambda_1 m_1)}{(m_1-1)!}\sum_{j=0}^{\infty}|c_j|^{\alpha-1}\exp(-\lambda_1 j) \tag{4.57}$$

Recall that for any complex number z and real number p, we denote $z^{\langle p\rangle} = |z|^{p-1}\bar{z}$.

Proof. For any fixed j and γ as in the proof of Theorem 4.2.3, write

$$k^{-(\gamma-1)}\exp(\omega_1 k)v_j(k) = \alpha(-s)^{<\alpha-1>}s\left\{p_j + \sum_l H_l(j)\exp(-\omega_l j)\right\}^{<\alpha-1>}$$
$$\times\left[\sum_l k^{-(\gamma-1)}H_l(j+k)\exp(-\omega_l j)\exp(-(\omega_l-\omega_1)k)\right]$$
$$+ k^{-(\gamma-1)}\exp(\omega_1 k)\ o\left[\sum_l H_l(j+k)\exp(-\omega_l(j+k))\right]$$
$$-\left|s\sum_l k^{-(\gamma-1)/\alpha}H_l(j+k)\exp(-\omega_l j)\exp(-(\omega_l-\omega_1/\alpha)k)\right|^{\alpha} \qquad (4.58)$$

The last two terms of (4.58) tend exponentially to zero as $k\to\infty$. In the first term, only the first term of the sum, namely

$$\alpha(-s)^{<\alpha-1>}s\left\{p_j + \sum_l H_l(j)\exp(-\omega_l j)\right\}^{<\alpha-1>} k^{-(\gamma-1)}H_1(j+k)\exp(-\omega_1 j)$$

does not tend to zero. Therefore, by (4.45), (4.48) and (4.49), if $\omega_1 = \kappa_1$,

$$\lim_{k\to\infty} k^{-(\gamma-1)}\exp(\kappa_1 k)v_j(k) = \alpha(-s)^{<\alpha-1>}s\left\{p_j + \sum_l H_l(j)\exp(-\omega_l j)\right\}^{<\alpha-1>}$$
$$\times d_{1l_1}\exp(-\kappa_1 l_1)\{(l_1-1)!\}^{-1}\exp(-\omega_1 j) \qquad (4.59)$$

and

$$\limsup_{k\to\infty} k^{-(\gamma-1)}\exp(\nu_1 k)v_j(k) \le \alpha|-s|^{\alpha-1}|s|\left|p_j + \sum_l H_l(j)\exp(-\omega_l j)\right|^{\alpha-1}$$
$$\times|f_{1g_1}|\exp(-\nu_1 g_1)\{(g_1-1)!\}^{-1}\exp(-\omega_1 j) \qquad (4.60)$$

if $\omega_1 = \nu_1$, and if $\omega_1 = \lambda_1$,

$$\limsup_{k\to\infty} k^{-(\gamma-1)}\exp(\lambda_1 k)v_j(k) \le \alpha|-s|^{\alpha-1}|s|\left|p_j + \sum_l H_l(j)\exp(-\omega_l j)\right|^{\alpha-1}$$
$$\times 2|e_{1m_1}|\exp(-\lambda_1 m_1)\{(m_1-1)!\}^{-1}\exp(-\omega_1 j) \qquad (4.61)$$

Formula (4.55), (4.56) and (4.57) will follow from (4.59), (4.60), (4.61), respectively, once we have shown that the sequence $\{v_j(k)\}_{k=0}^{\infty}$ is dominated by a summable sequence. To show this result, by (4.24) we write

$$|k^{-(\gamma-1)}e^{\omega_1 k}v_j(k)| \le \alpha\left|s\sum_l k^{-(\gamma-1)}H_l(j+k)e^{-\omega_l j}e^{-(\omega_l-\omega_1)k}\right|\left|(-s)\left\{p_j + \sum_l H_l(j)e^{-\omega_l j}\right\}\right|^{\alpha-1}$$
$$+(\alpha+1)\left|s\sum_l k^{-(\gamma-1)}H_l(j+k)k^{(1-1/\alpha)(\gamma-1)}e^{-(\omega_l-\omega_1/\alpha)k}e^{-(\omega_l-\omega_1)j}\right|^{\alpha}e^{-\omega_1\alpha j} \qquad (4.62)$$

Since $\lim_{j\to\infty}|c_j| = 0$, there is K_2 such that

$$\forall j \qquad \left|p_j + \sum_l H_l(j)\exp(-\omega_l j)\right| \le K_2$$

Therefore

$$|k^{-(\gamma-1)}\exp(\omega_1 k)v_j(k)| \leq \alpha|s||-s|^{\alpha-1}K_1K_2^{\alpha-1}\exp(-\omega_1 j)+(\alpha+1)|s|^{\alpha}K_1^{\alpha}K_3\exp(-\omega_1\alpha j) \quad (4.63)$$

where K_1 is as given in (4.53) and

$$K_3 = \sup_k \left\{ k^{(1-1/\alpha)(\gamma-1)} \exp\left(-\left(1-\frac{1}{\alpha}\right)\omega_1 k\right)\right\}$$

□

It remains to give the result for $\alpha = 1$, which is discussed in the following theorem

Theorem 4.2.5. *Suppose $\alpha = 1$ and consider the ARMA sequence (4.15), then*

(i). *If $\omega_1 = \kappa_1$ then*

$$\lim_{k\to\infty} k^{-(l_1-1)}\exp(\kappa_1 k)\tau(k) = \frac{\sigma^{\alpha}\exp(-\kappa_1 l_1)}{(l_1-1)!}\sum_{j=0}^{\infty}((-s)\text{sign}(c_j)d_{1l_1} + |sd_{1l_1}|)\exp(-\kappa_1 j) \quad (4.64)$$

(ii). *If $\omega_1 = \nu_1$ then*

$$\limsup_{k\to\infty} k^{-(g_1-1)}\exp(\nu_1 k)\,|\tau(k)| \leq \frac{2\sigma^{\alpha}\,|sf_{1g_1}|\exp(-\nu_1 g_1)}{(g_1-1)!\{1-\exp(-\nu_1)\}} \quad (4.65)$$

(iii). *If $\omega_1 = \lambda_1$ then*

$$\limsup_{k\to\infty} k^{-(m_1-1)}\exp(\lambda_1 k)\,|\tau(k)| \leq \frac{4\sigma^{\alpha}\,|se_{1m_1}|\exp(-\lambda_1 m_1)}{(m_1-1)!\{1-\exp(-\lambda_1)\}} \quad (4.66)$$

Proof. The proof of Theorem 4.2.5 is similar to the proof of Theorem 4.2.4 above, the only difference is the last term of (4.58) does not tend to zero. Furthermore, instead of (4.62), we can use simpler relation (4.54) to obtain a summable sequence which dominated the sequence $\{v_j(k)\}_{k=0}^{\infty}$. □

4.3 The estimator of the codifference function

As the codifference function is defined via characteristic functions (*cf*), it can be estimated by empirical characteristic functions (*ecf*)(see, e.g., Yu, 2004, for a review on *ecf*). Given a sample $X_1, X_2, \ldots, X_n$, an estimator for the codifference function at lag $k \in \mathbb{Z}$ can be defined as ($s \in \mathbb{R}$)

$$\hat{\tau}(s,-s;k) = \sqrt{(n-k)/n} \times [-\ln(\phi(s,-s;k)) + \ln(\phi(s,0;k)) + \ln(\phi(0,-s;k))] \quad (4.67)$$

where for $u, v \in \mathbb{R}$

$$\phi(u,v;k) = \begin{cases} (n-k)^{-1}\sum_{t=1}^{n-k}\exp(i(uX_{t+k}+vX_t)) \text{ when } k \geq 0 \\ (n+k)^{-1}\sum_{t=1}^{n+k}\exp(i(uX_{t-k}+vX_t)) \text{ when } k < 0 \end{cases} \quad (4.68)$$

Accordingly, $\hat{I}(s,-s;k) = \frac{\hat{\tau}(s,-s;k)}{\hat{\tau}(s,-s;0)}$ can be used as the estimator of the normalized codifference $I(k)$. Here we consider a discrete estimation procedure, i.e., we evaluate the codifference function at r points $s_1 < s_2 < \cdots < s_r$, for $s_i \in \mathbb{R}, s_i \neq 0, i = 1, \ldots, r$. In what follows, we denote the vectors $\mathbf{s} = \{s_1, \ldots, s_r\}$,

$$\hat{\tau}(\mathbf{s},k) = [\hat{\tau}(s_1,-s_1;k), \hat{\tau}(s_2,-s_2;k), \ldots, \hat{\tau}(s_r,-s_r;k)]^T$$

and

$$\hat{I}(\mathbf{s},k) = [\hat{I}(s_1,-s_1;k),\hat{I}(s_2,-s_2;k),\ldots,\hat{I}(s_r,-s_r;k)]^T$$

Note that one can replace the factor $\sqrt{(n-k)/n}$ in (4.67) by unity, and also the divisor $(n-k)$ in (4.68) by n without changing the asymptotic properties of the estimator, however, the choices in (4.67) and (4.68) will give a better finite sample performance than the alternative. Under assumptions C1 and C2, $\tau(-k) = \tau(k)$ such one can restrict the analysis to the case of $k \geq 0$. Notice that two similar estimators for the codifference function have recently been proposed in Yang *et al.* (2001) and in Hong (1999).

4.4 The asymptotic properties of the estimator

The asymptotic properties of the estimator is summarized in the following theorems.

Theorem 4.4.1. *Let $X_t, t \in \mathbb{Z}$ be the stationary linear process (3.1) satisfying conditions C1 and C2. For $s \in \mathbb{R}, s \neq 0$, its codifference estimator $\hat{\tau}(s,-s;k)$ and the sample normalized codifference $\hat{I}(s,-s;k)$ are (weakly) consistent estimators for $\tau(k), k \in \{0,1,2,\ldots\}$ and $I(k)$, respectively.*

The proof is given in Section 4.5.

The asymptotic distribution of the sample codifference function (and the sample normalized codifference function) of the linear process (3.1) can be derived using the central limit theorem for empirical characteristic function (Theorem A.4.1). For convenience, we split $\hat{\tau}$ into its real and imaginary parts. We write

$$\mathrm{Re}(\hat{\tau}(\mathbf{s},k)) = [\mathrm{Re}(\hat{\tau}(s_1,-s_1;k)),\mathrm{Re}(\hat{\tau}(s_2,-s_2;k)),\ldots,\mathrm{Re}(\hat{\tau}(s_r,-s_r;k))]^T$$

and

$$\mathrm{Im}(\hat{\tau}(\mathbf{s},k)) = [\mathrm{Im}(\hat{\tau}(s_1,-s_1;k)),\mathrm{Im}(\hat{\tau}(s_2,-s_2;k)),\ldots,\mathrm{Im}(\hat{\tau}(s_r,-s_r;k))]^T$$

Here, $\mathrm{Re}(z)$ and $\mathrm{Im}(z), z \in \mathbb{C}$ denote the real and imaginary parts of z. As $\hat{\tau}(s,-s,0)$ by definition is a real function, we therefore obtain

$$\mathrm{Re}(\hat{I}(\mathbf{s},k)) = \begin{bmatrix} \mathrm{Re}(\hat{\tau}(s_1,-s_1;k))/\hat{\tau}(s_1,-s_1;0) \\ \mathrm{Re}(\hat{\tau}(s_2,-s_2;k))/\hat{\tau}(s_2,-s_2;0) \\ \vdots \\ \mathrm{Re}(\hat{\tau}(s_r,-s_r;k))/\hat{\tau}(s_r,-s_r;0) \end{bmatrix}, \quad \mathrm{Im}(\hat{I}(\mathbf{s},k)) = \begin{bmatrix} \mathrm{Im}(\hat{\tau}(s_1,-s_1;k))/\hat{\tau}(s_1,-s_1;0) \\ \mathrm{Im}(\hat{\tau}(s_2,-s_2;k))/\hat{\tau}(s_2,-s_2;0) \\ \vdots \\ \mathrm{Im}(\hat{\tau}(s_r,-s_r;k))/\hat{\tau}(s_r,-s_r;0) \end{bmatrix} \tag{4.69}$$

In the following theorem, a result regarding the asymptotic distribution of the sample normalized codifference is given. The proof is given in Section 4.6.

Theorem 4.4.2. *Let $X_t, t \in \mathbb{Z}$ be the stationary linear process (3.1), satisfying conditions C1 and C2. Then for $h \in \{1,2,\ldots\}$,*

$$\left[\begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},1)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},1)) \end{pmatrix}, \begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},2)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},2)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},h)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},h)) \end{pmatrix}\right]^T$$

is

$$AN\left(\left[\begin{pmatrix} I(1) \\ 0 \end{pmatrix}, \begin{pmatrix} I(2) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} I(h) \\ 0 \end{pmatrix}\right]^T, n^{-1}\mathbf{W}\right) \tag{4.70}$$

The matrix variance-covariance $\mathbf{W}$ is given in (4.96).

Applying this theorem, we obtain the following corollary. The proof is given in Section 4.7.

Corollary 4.4.3. *Let $X_t, t \in \mathbb{Z}$ be an i.i.d. sequence satisfying the condition C2. Then for $k \in \{1, 2, \ldots\}$,*

$$\mathrm{Re}(\hat{I}(\mathbf{s}, k)) \text{ is } AN(0, n^{-1}\mathbf{W}_1) \tag{4.71}$$

and

$$\mathrm{Im}(\hat{I}(\mathbf{s}, k)) \text{ is } AN(0, n^{-1}\mathbf{W}_2) \tag{4.72}$$

where the (i,j)-th elements of matrix $\mathbf{W}_1$ and $\mathbf{W}_2$ are,

$$W_1(i,j) = \frac{f_{ij}}{g_{ij}} \text{ and } W_2(i,j) = \frac{h_{ij}}{g_{ij}} \quad , i,j = 1, \ldots, r \tag{4.73}$$

with

$$f_{ij} = e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i - s_j|^\alpha)} \left\{ \tfrac{1}{2} e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i - s_j|^\alpha)} - 1 \right\} + e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i + s_j|^\alpha)} \left\{ \tfrac{1}{2} e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i + s_j|^\alpha)} - 1 \right\} + 1$$

$$h_{ij} = e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i - s_j|^\alpha)} \left\{ \tfrac{1}{2} e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i - s_j|^\alpha)} - 1 \right\} + e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i + s_j|^\alpha)} \left\{ 1 - \tfrac{1}{2} e^{\sigma^\alpha(|s_i|^\alpha + |s_j|^\alpha - |s_i + s_j|^\alpha)} \right\}$$

and

$$g_{ij} = 4\sigma^{2\alpha} |s_i|^\alpha |s_j|^\alpha$$

4.5 Proof of Theorem 4.4.1

To show consistency of the codifference estimator, the following two lemmas are necessary.

Lemma 4.5.1. *Let $X_t, t \in \mathbb{Z}$ be the stationary linear process (3.1), satisfying conditions C1 and C2, and let $\Phi(s) = \mathrm{E}(\exp(isX_t))$ denote its first order characteristic function. For $k \in \{0, 1, 2 \ldots\}$ and $s \in \mathbb{R}, s \neq 0$*

$$\ln(\phi(s,k)) = \ln\left((n-k)^{-1} \sum_{t=1}^{n-k} \exp(isX_t) \right)$$

is a consistent estimator of $\ln(\Phi(s))$.

Proof. Let $y_t = \exp(isX_t)$. Apparently, the magnitude of y_t is equal to one, and therefore it is a second order stationary process. For the sake of simplicity, instead of working with $\phi(s,k)$, we first show consistency of $\phi^*(s,k) = n^{-1}\sum_{t=1}^{n} \exp(isX_t)$. Here, $\phi^*(s,k)$ is an unbiased estimator for $\Phi(s) = E(y_t)$. To show the weak consistency of this estimator, we show that y_t is a mean ergodic process. A sufficient condition for y_t to be mean ergodic, i.e., $\phi^*(s,k) \to \mathrm{E}(y_t)$ in the mean square sense, is that its covariance function tends to zero as time lags tends to ∞ (see Theorem A.3.7). The covariance function of y_t at lag k can be expressed as

$$c(k) = |\Phi(s)|^2 \left(\frac{\mathrm{E}(\exp(is(X_{t+k} - X_t)))}{\mathrm{E}(\exp(isX_{t+k}))\mathrm{E}(\exp(-isX_t))} - 1 \right) = |\Phi(s)|^2 \left(\exp(-\tau(k)) - 1\right) \tag{4.74}$$

From Theorem 4.2.1, we see that $c(k) \to 0$ when $k \to \infty$ exponentially fast. As mean square convergence entails convergence in probability, $\phi^*(s,k) \xrightarrow{p} \Phi(s)$. Moreover, under assumptions C1 and C2, we have $\Phi(s) = \exp(-\sum_{j=0}^{\infty} \sigma^\alpha |sc_j|^\alpha)$, a real-valued function. Therefore we can conclude $\mathrm{Re}(\phi^*(s,k)) \xrightarrow{p} \mathrm{Re}(\Phi(s)) = \Phi(s)$ and $\mathrm{Im}(\phi^*(s,k) \xrightarrow{p} \mathrm{Im}(\Phi(s)) = 0$.

By taking the principal value of $\ln(\cdot)$ function in the complex domain, we obtain that $\ln(\cdot)$ is a continuous and well-defined function on $\mathbb{C}$ minus the negative real line. Because $|c_j| < CQ^{-j}$ for

some $C > 0, Q > 1$, we conclude $\mathrm{Re}(\Phi(s))$ always strictly greater than 0, which implies with the probability converging to 0, $\mathrm{Re}(\phi^*(s,k))$ will be less than or equal to 0. Therefore, without loss of generality, we can restrict the definition of the real and imaginary parts of $\ln(\phi^*(s,k))$ only on the right half plane where $\mathrm{Re}(\phi^*(s,k)) > 0$, and equal to 0 on the other case. From this consideration, we obtain $\mathrm{Re}(\ln(\phi^*(s,k))) = \frac{1}{2}\ln((\mathrm{Re}(\phi^*(s,k)))^2 + (\mathrm{Im}(\phi^*(s,k)))^2)$ and $\mathrm{Im}(\ln(\phi^*(s,k))) = \arctan(\frac{\mathrm{Im}(\phi^*(s,k))}{\mathrm{Re}(\phi^*(s,k))})$. From the continuity of the logarithm function in the considered domain, we can deduce that $\mathrm{Re}(\ln(\phi^*(s,k))) \xrightarrow{p} \mathrm{Re}(\ln(\Phi(s))) = \ln(\Phi(s))$ and $\mathrm{Im}(\ln(\phi^*(s,k))) = \arg(\phi^*(s,k)) \xrightarrow{p} 0$, when $n \to \infty$. In other words, we obtain $\ln(\phi^*(s,k)) \xrightarrow{p} \ln(\Phi(s))$. To complete our proof, it is sufficient to show $\phi^*(s,k) - \phi(s,k) \xrightarrow{p} 0$. By assumption of the model, $\mathrm{Re}(\Phi(s)) > 0$, thus $\mathrm{E}\,|\mathrm{Re}(\phi^*(s,k)) - \mathrm{Re}(\phi(s,k))| < 2\frac{k}{n-k}$ and $\mathrm{E}\,|\mathrm{Im}(\phi^*(s,k)) - \mathrm{Im}(\phi(s,k))| < 2\frac{k}{n-k}$, and therefore we can conclude $\phi^*(s,k) - \phi(s,k) = o_p(1)$. □

Lemma 4.5.2. *Let $X_t, t \in \mathbb{Z}$ be the stationary linear process (3.1), satisfying conditions C1 and C2, and for $k \in \{0,1,2,\ldots\}$ and $s \in \mathbb{R}, s \neq 0$, let $\Phi(s,-s;k) = \mathrm{E}(\exp(is(X_{t+k} - X_t)))$ be its second-order characteristic function evaluated at $(s,-s)$. Then as $n \to \infty$*

$$\ln(\phi(s,-s;k)) \xrightarrow{p} \ln(\Phi(s,-s;k))$$

where $\phi(s,-s;k)$ is as given in (4.68).

Proof. For the proof, we proceed in a similar way as the previous lemma. For simplicity, instead of working with $\phi(s,-s;k)$, we first show the consistency of $\phi^*(s,-s;k) = n^{-1}\sum_{t=1}^{n}\exp(is(X_{t+k} - X_t))$. A sufficient condition for y_t to be autocovariance ergodic (Proakis and Manolakis, 1996, p.A10), i.e., $\phi^*(s,-s,k) \to \Phi(s,-s;k)$, in the mean square sense is that

$$\mathrm{E}(\exp(is(X_t - X_{t+k} - X_{t+l} + X_{t+l+k}))) \to |\Phi(s,-s;k)|^2$$

as $l \to \infty$ where the index l denotes the lag of covariance among the sample autocovariance function. Hence, we have

$$\begin{aligned}&\mathrm{E}(\exp(is(X_t - X_{t+k} - X_{t+l} + X_{t+l+k})))\\ &= |\Phi(s,-s;k)|^2 \frac{\mathrm{E}(\exp(is((X_t - X_{t+k}) - (X_{t+l} - X_{t+l+k}))))}{\mathrm{E}(\exp(is(X_t - X_{t+k})))\mathrm{E}(\exp(is(X_{t+l+k} - X_{t+l})))} = |\Phi(s,-s;k)|^2\exp(-C_l)\end{aligned}$$

where

$$\begin{aligned}C_l = &-\ln(\mathrm{E}(\exp(is((X_t - X_{t+k}) - (X_{t+l} - X_{t+l+k})))))\\ &+ \ln(\mathrm{E}(\exp(is(X_t - X_{t+k})))) + \ln(\mathrm{E}(\exp(is(X_{t+l+k} - X_{t+l}))))\end{aligned}$$

To obtain a simple expression for C_l, we can write the elements of C_l as follows

$$\begin{aligned}&-\ln \mathrm{E}\exp is((X_t - X_{t+k}) - (X_{t+l} - X_{t+l+k}))\\ &= -\ln \mathrm{E}\exp is\left(\sum_{j=0}^{\infty} c_j\epsilon_{t-j} - \sum_{j=0}^{\infty} c_j\epsilon_{t+k-j} - \sum_{j=0}^{\infty} c_j\epsilon_{t+l-j} + \sum_{j=0}^{\infty} c_j\epsilon_{t+l+k-j}\right)\\ &= -\ln \mathrm{E}\exp is\left(\sum_{j=0}^{\infty}(c_j - c_{j+k} - c_{j+l} + c_{j+l+k})\epsilon_{t-j}\right.\\ &\qquad\left. - \sum_{j=0}^{k-1} c_j\epsilon_{t+k-j} - \sum_{j=0}^{l-1} c_j\epsilon_{t+l-j} + \sum_{j=0}^{l+k-1} c_j\epsilon_{t+l+k-j}\right)\\ &= \sigma^\alpha\left[\sum_{j=0}^{\infty}|s(c_j - c_{j+k} - c_{j+l} + c_{j+l+k})|^\alpha + \sum_{j=0}^{k-1}|-sc_j|^\alpha + \sum_{j=0}^{l-1}|-sc_j|^\alpha + \sum_{j=0}^{l+k-1}|sc_j|^\alpha\right]\end{aligned}$$

$$\ln \mathrm{E} \exp is(X_t - X_{t+k}) = \ln \mathrm{E} \exp is \left(\sum_{j=0}^{\infty} c_j \epsilon_{t-j} - \sum_{j=0}^{\infty} c_j \epsilon_{t+k-j} \right)$$
$$= \ln \mathrm{E} \exp is \left(\sum_{j=0}^{\infty} (c_j - c_{j+k}) \epsilon_{t-j} - \sum_{j=0}^{k-1} c_j \epsilon_{t+k-j} \right)$$
$$= \sigma^\alpha \left[- \sum_{j=0}^{\infty} |s(c_j - c_{j+k})|^\alpha - \sum_{j=0}^{k-1} |-sc_j|^\alpha \right]$$

and

$$\ln \mathrm{E} \exp is(X_{t+l+k} - X_{t+l}) = \ln \mathrm{E} \exp is \left(\sum_{j=0}^{\infty} c_j \epsilon_{t+l+k-j} - \sum_{j=0}^{\infty} c_j \epsilon_{t+l-j} \right)$$
$$= \ln \mathrm{E} \exp is \left(\sum_{j=0}^{l+k-1} c_j \epsilon_{t+l+k-j} - \sum_{j=0}^{l-1} c_j \epsilon_{t+l-j} + \sum_{j=0}^{\infty} (c_{j+l+k} - c_{j+l}) \epsilon_{t-j} \right)$$
$$= \sigma^\alpha \left[- \sum_{j=0}^{l+k-1} |sc_j|^\alpha - \sum_{j=0}^{l-1} |-sc_j|^\alpha - \sum_{j=0}^{\infty} |s(c_{j+l+k} - c_{j+l})|^\alpha \right]$$

Therefore C_l can be written as

$$C_l = \sigma^\alpha \left[\sum_{j=0}^{\infty} |s(c_j - c_{j+k} - c_{j+l} + c_{j+l+k})|^\alpha - |s(c_{j+l+k} - c_{j+l})|^\alpha - |s(c_j - c_{j+k})|^\alpha \right] \quad (4.75)$$
$$= \sigma^\alpha \left[\sum_{j=0}^{\infty} |s(k_j - k_{j+l})|^\alpha - |-sk_{j+l}|^\alpha - |sk_j|^\alpha \right] \quad (4.76)$$

where $k_j = c_j - c_{j+k}$. This expression is the codifference function $\tau_G(u, v; l)$ for coefficients k_j's and parameters $u = -s, v = s$. Because $|c_j| < CQ^{-j}$ for some $C > 0, Q > 1$, then $|k_j| < C_1 Q^{-j}$ for some $C_1 = 2C > 0, Q > 1$. As Theorem 4.2.1 also applies for $\tau_G(u, v; \cdot)$, by (4.22) and (4.21), we can conclude that $\exp(-C_l)$ will converge to 1 exponentially fast. In other words, $\mathrm{E}(\exp(is(X_t - X_{t+k} - X_{t+l} + X_{t+l+k}))) \to |\Phi(s, -s; k)|^2$ for $n \to \infty$, and we obtain the mean square convergence of $\phi^*(s, -s; k)$ to $\Phi(s, -s; k)$ and therefore $\phi^*(s, -s; k) \xrightarrow{p} \Phi(s, -s; k)$. For the rest of the proof, we can proceed similarly to the proof of previous lemma, as we have $\Phi(s, -s; k) = \exp(-\sigma^\alpha(\sum_{j=0}^{k-1} |sc_j|^\alpha - \sum_{j=0}^{\infty} |s(c_{j+k} - c_j)|^\alpha))$ also a real-valued function, strictly greater than 0. □

Proof of Theorem 4.4.1. As for finite k and $n \to \infty$ we obtain $\sqrt{1 - k/n} \to 1$, then using the results in lemma 4.5.1 and lemma 4.5.2, we have as $n \to \infty$, for $i = 1, \ldots, r$

$$\hat{\tau}(s_i, -s_i; k) \xrightarrow{p} - \ln(\Phi(s_i, -s_i; k)) + \ln(\Phi(s_i)) + \ln(\Phi(-s_i)) = \tau(s_i, -s_i; k) \quad (4.77)$$

□

4.6 Proof of Theorem 4.4.2

In this part, we will derive the asymptotic distribution of the sample codifference function of linear processes. The proof will be given as a series of propositions, where the main results are presented

in Theorem 4.6.4 and also the proof of Theorem 4.4.2 at the end of this part. The proof will follow closely an approach for obtaining the limiting distribution of the SACF in the classical case, e.g., Theorem 7.2.1 in Brockwell and Davis (1987).

For notational simplicity, instead of working with $\hat{\tau}(s_i, -s_i; k)$, $i = 1, \ldots, r$, in the following first we will consider the similar estimator $\hat{\tau}^*(s_i, -s_i; k)$,

$$\hat{\tau}^*(s_i, -s_i; k) = -\ln(\phi^*(s_i, -s_i; k)) + \ln(\phi^*(s_i, 0; k)) + \ln(\phi^*(0, -s_i; k)) \tag{4.78}$$

where $\phi^*(u, v; k) = n^{-1} \sum_{t=1}^{n} \exp(i(uX_{t+k} + vX_t))$, $u, v \in \mathbb{R}$. The required result will be presented in Theorem 4.6.4.

Proposition 4.6.1. *Let $X_t, t \in \mathbb{Z}$ be the stationary linear process (3.1), satisfying conditions C1 and C2. Then if $p \geq 0$ and $q \geq 0$,*

$$\lim_{n\to\infty} n \operatorname{cov}\left(\begin{pmatrix} \operatorname{Re}(\hat{\tau}^*(\mathbf{s}, p)) \\ \operatorname{Im}(\hat{\tau}^*(\mathbf{s}, p)) \end{pmatrix}, \begin{pmatrix} \operatorname{Re}(\hat{\tau}^*(\mathbf{s}, q)) \\ \operatorname{Im}(\hat{\tau}^*(\mathbf{s}, q)) \end{pmatrix}\right) = \lambda \mathbf{L}_2^p \mathbf{V}_{pq} \mathbf{L}_2^q \lambda^T$$

where the matrices $\lambda, \mathbf{L}_2^k, k = p, q$ and $\mathbf{V}_{pq}$ are given in (4.80), (4.87) and (4.89) below. Here $\operatorname{cov}(X, Y)$ denotes the covariance between X and Y.

Proof. To obtain a complete variance-covariance structure of the estimator, we consider the following representation of $\hat{\tau}^*(\mathbf{s}, k)$

$$\begin{pmatrix} \operatorname{Re}(\hat{\tau}^*(\mathbf{s}, k)) \\ \operatorname{Im}(\hat{\tau}^*(\mathbf{s}, k)) \end{pmatrix} = \begin{pmatrix} \operatorname{Re}(\hat{\tau}^*(s_1, -s_1, k)) \\ \operatorname{Re}(\hat{\tau}^*(s_2, -s_2, k)) \\ \vdots \\ \operatorname{Re}(\hat{\tau}^*(s_r, -s_r, k)) \\ \operatorname{Im}(\hat{\tau}^*(s_1, -s_1, k)) \\ \operatorname{Im}(\hat{\tau}^*(s_2, -s_2, k)) \\ \vdots \\ \operatorname{Im}(\hat{\tau}^*(s_r, -s_r, k)) \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} \tag{4.79}$$

where

$$\lambda = \begin{pmatrix} \mathbf{I}_r \otimes \lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r \otimes \lambda_1 \end{pmatrix} \tag{4.80}$$

$$\lambda_1 = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$$

and

$$\mathbf{Y} = \begin{pmatrix} \operatorname{Re}(\ln(Y_1^k)) \\ \operatorname{Re}(\ln(Y_2^k)) \\ \vdots \\ \operatorname{Re}(\ln(Y_r^k)) \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \operatorname{Im}(\ln(Y_1^k)) \\ \operatorname{Im}(\ln(Y_2^k)) \\ \vdots \\ \operatorname{Im}(\ln(Y_r^k)) \end{pmatrix}$$

Here $\mathbf{I}_r$ denotes the matrix identity of size r, where we denote

$$Y_i^k = \begin{pmatrix} \phi^*(0, -s_i; k) \\ \phi^*(s_i, 0; k) \\ \phi^*(s_i, -s_i; k) \end{pmatrix} = \begin{pmatrix} \phi_1(s_i, k) \\ \phi_2(s_i, k) \\ \phi_3(s_i, k) \end{pmatrix}$$

and the logarithm function is defined componentwise, i.e., we have

$$\operatorname{Re} \ln Y_i^k = \begin{pmatrix} \operatorname{Re}(\ln(\phi_1(s_i, k))) \\ \operatorname{Re}(\ln(\phi_2(s_i, k))) \\ \operatorname{Re}(\ln(\phi_3(s_i, k))) \end{pmatrix}$$

and similarly for the imaginary part. Let us denote

$$\mathrm{E}(Y_i^k) = \begin{pmatrix} \mathrm{E}(\phi_1(s_i,k)) \\ \mathrm{E}(\phi_2(s_i,k)) \\ \mathrm{E}(\phi_3(s_i,k)) \end{pmatrix} = \begin{pmatrix} \Phi_1(s_i,k) \\ \Phi_2(s_i,k) \\ \Phi_3(s_i,k) \end{pmatrix}$$

Notice that $\Phi(u,v;k) = \mathrm{E}(\exp i(uX_{t+k} + vX_t))$, $u,v \in \mathbb{R}$. Using mean value theorem, we can expand the codifference function into

$$\begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(\mathbf{s},k)) \\ \mathrm{Im}(\hat{\tau}^*(\mathbf{s},k)) \end{pmatrix} = \lambda \left\{ \mathbf{L}_1^k + \bar{\mathbf{L}}_2^k \mathbf{Z}_n^k \right\} \tag{4.81}$$

where

$$\mathbf{L}_1^k = \begin{pmatrix} \mathrm{Re}(\mathbf{L}_1^k) \\ \mathrm{Im}(\mathbf{L}_1^k) \end{pmatrix}, \ \mathbf{Z}_n^k = \begin{pmatrix} \mathrm{Re}(\mathbf{Z}_n^k) \\ \mathrm{Im}(\mathbf{Z}_n^k) \end{pmatrix} = \begin{pmatrix} \mathrm{Re}(\varphi_n^k) - \mathrm{Re}(\psi_n^k) \\ \mathrm{Im}(\varphi_n^k) - \mathrm{Re}(\psi_n^k) \end{pmatrix}$$

with

$$\mathrm{Re}(\mathbf{L}_1^k) = \begin{pmatrix} \mathrm{Re}(\ln(\mathrm{E}(Y_1^k))) \\ \mathrm{Re}(\ln(\mathrm{E}(Y_2^k))) \\ \vdots \\ \mathrm{Re}(\ln(\mathrm{E}(Y_r^k))) \end{pmatrix}, \mathrm{Re}(\varphi_n^k) = \begin{pmatrix} \mathrm{Re}(Y_1^k) \\ \mathrm{Re}(Y_2^k) \\ \vdots \\ \mathrm{Re}(Y_r^k) \end{pmatrix}, \mathrm{Re}(\psi_n^k) = \begin{pmatrix} \mathrm{Re}(\mathrm{E}(Y_1^k)) \\ \mathrm{Re}(\mathrm{E}(Y_2^k)) \\ \vdots \\ \mathrm{Re}(\mathrm{E}(Y_r^k)) \end{pmatrix}$$

and similarly for the imaginary parts, and where and $\bar{\mathbf{L}}_2^k = (\bar{d}_{ij}^k)_{i,j=1,\ldots,6}$ denotes Jacobian of (4.79), which is evaluated at $\mathbf{c}$ ($\left\|\mathbf{c} - \psi_n^k\right\| < \left\|\varphi_n^k - \psi_n^k\right\|$). From assumption C2, we obtain

$$\Phi_3(s_i,k) = \Phi(s_i,-s_i;k) = \exp(-\sum_{j=0}^{k-1} \sigma^\alpha \left|s_i c_j\right|^\alpha - \sum_{j=0}^{\infty} \sigma^\alpha \left|s_i(c_{j+k} - c_j)\right|^\alpha) \tag{4.82}$$

and $\Phi_1(s_i,k) = \Phi_2(s_i,k)$, i.e.,

$$\Phi(s_i,0;k) = \Phi(0,-s_i;k) = \exp(-\sum_{j=0}^{\infty} \sigma^\alpha \left|s_i c_j\right|^\alpha) \tag{4.83}$$

From identities (4.82)-(4.83) and further applying assumption C1, we obtain that the elements of $\mathrm{Re}(\psi_n^k)$ are always strictly greater than 0. Therefore, with a probability convergent to 0, the elements of $\mathrm{Re}(\varphi_n^k)$ will be less than or equal to 0. Hence, without changing the limiting distribution of the estimator, we can restrict the definition of the real and the imaginary components of $\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix}$ in (4.79) only in the right half plane where the elements of $\mathrm{Re}(\varphi_n^k) > 0$, and equal to 0 in the other case. Thus, we can conclude that the Jacobian matrix $\bar{\mathbf{L}}_2^k$ is well defined here. By Theorem 4.4.1, $\bar{\mathbf{L}}_2^k$ will converge in probability to $\mathbf{L}_2^k$, where

$$\mathbf{L}_2^k = \nabla \mathbf{L}_1^k$$

Here ∇g denotes the Jacobian of g. From (4.82), (4.83), we have the following identities

$$\mathrm{Re}(\Phi(s_i,-s_i;k)) = \mathrm{E}(\cos(s_i(X_{t+k} - X_t))) = \Phi(s_i,-s_i,k) \tag{4.84}$$

$$\mathrm{Re}(\Phi(s_i,0;k)) = \mathrm{E}(\cos(s_i X_{t+k})) = \Phi(s_i,0,k) \tag{4.85}$$

$$\mathrm{Re}(\Phi(0,-s_i;k)) = \mathrm{E}(\cos(-s_i X_t)) = \Phi(0,-s_i,k) \tag{4.86}$$

and $\mathrm{Im}(\Phi(s_i,-s_i;k)) = \mathrm{E}(\sin(s_i(X_{t+k}-X_t))) = 0$, $\mathrm{Im}(\Phi(s_i,0;k)) = \mathrm{E}(\sin(s_iX_{t+k})) = 0$ and $\mathrm{Im}(\Phi(0,-s_i;k)) = \mathrm{E}(\sin(-s_iX_t)) = 0$. Using these identities, after some algebra we directly obtain

$$\mathbf{L}_2^k = \begin{pmatrix} \mathbf{I}_r\mathbf{d}^k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r\mathbf{d}^k \end{pmatrix} \tag{4.87}$$

where $(\mathbf{d}^k)^T = [\mathbf{d}_1^k, \mathbf{d}_2^k, \ldots, \mathbf{d}_r^k]$, and the elements of $\mathbf{d}_i^k, i = 1,\ldots,r$ are

$$\begin{aligned} d_i^k(1,1) &= (\mathrm{Re}(\Phi(0,-s_i;k)))^{-1} \\ d_i^k(2,2) &= (\mathrm{Re}(\Phi(s_i,0;k)))^{-1} \\ d_i^k(3,3) &= (\mathrm{Re}(\Phi(s_i,-s_i;k)))^{-1}) \end{aligned}$$

and equal to 0, otherwise. The asymptotic variance-covariance matrix is obtained from (4.81) as

$$\lim_{n\to\infty} n \ \mathrm{cov}\left(\begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(s,-s;p)) \\ \mathrm{Im}(\hat{\tau}^*(s,-s;p)) \end{pmatrix}, \begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(s,-s;q)) \\ \mathrm{Im}(\hat{\tau}^*(s,-s;q)) \end{pmatrix}\right) = \lambda\mathbf{L}_2^p\mathbf{V}_{pq}\mathbf{L}_2^q\lambda^T \tag{4.88}$$

where

$$\mathbf{V}_{pq} = \begin{pmatrix} \mathbf{V}_{pq}^{RR} & \mathbf{V}_{pq}^{RI} \\ \mathbf{V}_{pq}^{IR} & \mathbf{V}_{pq}^{II} \end{pmatrix} = \lim_{n\to\infty} n \begin{pmatrix} \mathrm{cov}(\mathrm{Re}(\mathbf{Z}_n^p),\mathrm{Re}(\mathbf{Z}_n^q)) & \mathrm{cov}(\mathrm{Re}(\mathbf{Z}_n^p),\mathrm{Im}(\mathbf{Z}_n^q)) \\ \mathrm{cov}(\mathrm{Im}(\mathbf{Z}_n^p),\mathrm{Re}(\mathbf{Z}_n^q)) & \mathrm{cov}(\mathrm{Im}(\mathbf{Z}_n^p),\mathrm{Im}(\mathbf{Z}_n^q)) \end{pmatrix} \tag{4.89}$$

The matrix $\mathbf{V}_{pq}$ can be obtained by applying Theorem A.4.1. Its elements can be derived in a similar way as obtaining variance-covariance matrix in Theorem A.4.1. This is possible, because it can be shown that all elements of $\mathbf{V}_{pq}$ (in the form of sum of the absolute components) are finite. Therefore, one can apply the property of the sample mean of ergodic processes (Theorem A.3.7). Notice that here in particular, we obtain all elements of $\mathbf{V}_{pq}$ with respect to $\mathrm{cov}(\mathrm{Re}(\mathbf{Z}_n^p),\mathrm{Im}(\mathbf{Z}_n^q))$ and $\mathrm{cov}(\mathrm{Im}(\mathbf{Z}_n^p),\mathrm{Re}(\mathbf{Z}_n^q))$ are zeros. The elements of $\mathbf{V}_{pq}$ with respect to $\mathrm{cov}(\mathrm{Re}(\mathbf{Z}_n^p),\mathrm{Re}(\mathbf{Z}_n^q))$ and $\mathrm{cov}(\mathrm{Im}(\mathbf{Z}_n^p),\mathrm{Im}(\mathbf{Z}_n^q))$ can be shown to be finite using identities (4.82)-(4.83) and applying a similar approach as obtaining eq. (4.74) and (4.76), and further applying Theorem 4.2.1, or sometimes, eq.(4.24) together with the similar steps as the proof of Theorem 4.2.1. However, we omit details. □

Proposition 4.6.2. *Let $X_t, t \in \mathbb{Z}$ be the moving average process of order m, $X_t = \sum_{j=0}^m c_j\epsilon_{t-j}$, satisfying conditions C1 and C2. Then for $h \in \{1,2,\ldots\}$, $s \in \mathbb{R}, s \neq 0$*

$$\left[\begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(\mathbf{s},0)) \\ \mathrm{Im}(\hat{\tau}^*(\mathbf{s},0)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(\mathbf{s},h)) \\ \mathrm{Im}(\hat{\tau}^*(\mathbf{s},h)) \end{pmatrix}\right] \text{ is } AN\left(\left[\begin{pmatrix} \tau(\mathbf{s},0) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \tau(\mathbf{s},h) \\ 0 \end{pmatrix}\right], n^{-1}M\right)$$

where $\mathbf{M}$ is the covariance matrix

$$\mathbf{M} = \left[\lambda\mathbf{L}_2^p\mathbf{V}_{pq}\mathbf{L}_2^q\lambda^T\right]_{p,q=0,\ldots,h}$$

and the matrices $\lambda, \mathbf{L}_2^k, k = p,q$ and $\mathbf{V}_{pq}$ are as given in Proposition 4.6.1 above.

Proof. To show this relation, define vectors $\{\mathbf{Y}_t\}$ by

$$\mathbf{Y}_t^T = (\mathbf{Z}_t, \mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_{t+h})$$

where

$$\mathbf{Z}_{t+k} = \begin{pmatrix} \mathbf{X}_1^k \\ \mathbf{X}_2^k \\ \vdots \\ \mathbf{X}_r^k \end{pmatrix}$$

where for $j = 1,\ldots,r$

$$\mathbf{X}_j^k = \begin{pmatrix} \exp(-is_jX_t) \\ \exp(is_jX_{t+k}) \\ \exp(is_j(X_{t+k}-X_t)) \end{pmatrix}$$

By definition, $\{\mathbf{Z}_{t+k}\}$ is $m+k$-dependent sequence and therefore $\{\mathbf{Y}_t\}$ is $m+h$-dependent sequence. Next define

$$\zeta_t^T = (\xi_t, \xi_{t+1}, \ldots, \xi_{t+h})$$

where

$$\xi_{t+j} = \begin{pmatrix} \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n Z_{t+j})) \\ \mathrm{Im}(\ln(n^{-1}\sum_{t=1}^n Z_{t+j})) \end{pmatrix}$$

and

$$\mathrm{Re}(\ln(n^{-1}\sum\nolimits_{t=1}^n Z_{t+j})) = \begin{pmatrix} \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \mathbf{X}_1^j)) \\ \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \mathbf{X}_2^j)) \\ \vdots \\ \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \mathbf{X}_r^j)) \end{pmatrix}$$

where $l = 1, \ldots, r$

$$\mathrm{Re}(\ln(n^{-1}\sum\nolimits_{t_1=1}^n \mathbf{X}_l^j)) = \begin{pmatrix} \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \exp(-is_l X_t))) \\ \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \exp(is_l X_{t+j}))) \\ \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \exp(is_l(X_{t+j} - X_t)))) \end{pmatrix}$$

(similarly for the imaginary part. Note that the summation and the principal value of $\ln(\cdot)$ are defined componentwise), then we have

$$\lambda \begin{pmatrix} \mathrm{Re}(\ln(n^{-1}\sum_{t=1}^n \mathbf{Y}_t^T)) \\ \mathrm{Im}(\ln(n^{-1}\sum_{t=1}^n \mathbf{Y}_t^T)) \end{pmatrix} = \lambda\zeta_t^T = \left[\begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(\mathbf{s},0)) \\ \mathrm{Im}(\hat{\tau}^*(\mathbf{s},0)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{\tau}^*(\mathbf{s},h)) \\ \mathrm{Im}(\hat{\tau}^*(\mathbf{s},h)) \end{pmatrix} \right]$$

where λ is as given in (4.80). We therefore need to show that when $n \to \infty$

$$\mathbf{a}^T(\lambda[\xi_t, \xi_{t+1}, \ldots, \xi_{t+h}])^T \text{ is } AN\left(\mathbf{a}^T\left(\begin{pmatrix} \mathrm{Re}(\tau(0)) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\tau(h)) \\ 0 \end{pmatrix}\right)^T, n^{-1}\mathbf{a}^T\mathbf{M}\mathbf{a}\right) \quad (4.90)$$

for all vectors $\mathbf{a} = (a_0, \ldots, a_h)^T \in \mathbb{R}^{h+1}$ such that $\mathbf{a}^T\mathbf{M}\mathbf{a} > 0$. For any such $\mathbf{a}$, the sequence $\{\mathbf{a}^T(\lambda\zeta_t^T)^T\}$ is $(m+h)$-dependent and since by Proposition 4.6.1

$$\lim_{n\to\infty} n \ \mathrm{var}(\mathbf{a}^T(\lambda[\xi_t, \xi_{t+1}, \ldots, \xi_{t+h}])^T) = \mathbf{a}^T\mathbf{M}\mathbf{a} > 0$$

where $\mathbf{M}$ is the covariance matrix

$$\mathbf{M} = \left[\lambda\mathbf{L}_2^p\mathbf{V}_{pq}\mathbf{L}_2^q\lambda^T\right]_{p,q=0,\ldots,h}$$

and the vectors $\lambda, \mathbf{L}_2^p, \mathbf{L}_2^q$, matrix $\mathbf{V}_{pq}$ are as given in Proposition 4.6.1 above. We can conclude that $\{\mathbf{a}^T(\lambda\zeta_t^T)^T\}$ satisfies the conditions of central limit theorems for m-dependent processes (Theorem A.2.15), and therefore by this theorem, for $n \to \infty$, we obtain the required result (4.90). The relation $\mathrm{Im}(\tau(\mathbf{s}, j)) = 0, j = 0, 1, \ldots, h$ can be obtained directly from identities (4.82)-(4.83). □

Proposition 4.6.3. *Proposition 4.6.2 remains true for $X_t, t \in \mathbb{Z}$ being a stationary linear process (3.1), satisfying conditions C1 and C2.*

Proof. For the proof, we will apply the result of Proposition 4.6.2 to the truncated sequence $X_{tm} = \sum_{j=0}^m c_j\epsilon_{t-j}$ and then derive the result for X_t by letting $m \to \infty$. For $0 \le p \le h$, we define

$$\hat{\tau}_m^*(s, -s; p) = -\ln(\phi_m^*(s, -s; p)) + \ln(\phi_m^*(s, 0; p)) + \ln(\phi_m^*(0, -s; p)) \quad (4.91)$$

where $\phi_m^*(u, v; p) = n^{-1}\sum_{t=1}^n \exp(i(uX_{(t+p)m} + vX_{tm}))$. Then by Proposition 4.6.2

$$n^{1/2}\left[\begin{pmatrix} \mathrm{Re}(\hat{\tau}_m^*(\mathbf{s},0)) - \mathrm{Re}(\tau_m(\mathbf{s},0)) \\ \mathrm{Im}(\hat{\tau}_m^*(\mathbf{s},0)) - \mathrm{Im}(\tau_m(\mathbf{s},0)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{\tau}_m^*(\mathbf{s},h)) - \mathrm{Re}(\tau_m(\mathbf{s},h)) \\ \mathrm{Im}(\hat{\tau}_m^*(\mathbf{s},h)) - \mathrm{Im}(\tau_m(\mathbf{s},h)) \end{pmatrix}\right] \Rightarrow Y_m$$

where $Y_m \sim N(0, M_m)$. Here $\mathbf{M}_m$ is the covariance matrix

$$\begin{aligned}\mathbf{M}_m &= \begin{pmatrix} \text{cov}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)), \text{Re}(\hat{\tau}_m^*(\mathbf{s},q))) & \text{cov}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)), \text{Im}(\hat{\tau}_m^*(\mathbf{s},q))) \\ \text{cov}(\text{Im}(\hat{\tau}_m^*(\mathbf{s},p)), \text{Re}(\hat{\tau}_m^*(\mathbf{s},q))) & \text{cov}(\text{Im}(\hat{\tau}_m^*(\mathbf{s},p)), \text{Im}(\hat{\tau}_m^*(\mathbf{s},q))) \end{pmatrix}_{p,q=0,\ldots,h} \\ &= \left[\lambda \mathbf{L}_2^p(m) \mathbf{V}_{pq}^m \mathbf{L}_2^q(m) \lambda^T\right]_{p,q=0,\ldots,h}\end{aligned}$$

where λ is defined as (4.80) and the Jacobian matrix $\mathbf{L}_2^k(m)$ and matrix $\mathbf{V}_{pq}^m$ are defined for X_{tm} as in (4.87) and (4.89), respectively. Now, as $m \to \infty$,

$$\mathbf{M}_m \to \mathbf{M}$$

where $\mathbf{M}$ is defined like $\mathbf{M}_m$ by replacing X_{tm} by X_t. Hence

$$Y_m \Rightarrow Y \text{ where } Y \sim N(0, \mathbf{M})$$

The proof now can be completed by applying Theorem A.2.7 provided we can show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \text{P}(n^{1/2} |\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)) - \text{Re}(\tau_m(\mathbf{s},p)) - \text{Re}(\hat{\tau}^*(\mathbf{s},p)) + \text{Re}(\tau(\mathbf{s},p))| > \epsilon) = 0 \qquad (4.92)$$

for $p = 0, 1, \ldots, h$ (and similarly for the imaginary part). The probability in (4.92) is bounded by

$$\begin{aligned}&\epsilon^{-2} n \text{ var}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)) - \text{Re}(\hat{\tau}^*(\mathbf{s},p))) \\ &= \epsilon^{-2} [n \text{ var}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p))) + n \text{ var}(\text{Re}(\hat{\tau}^*(\mathbf{s},p))) - 2n\text{cov}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)), \text{Re}(\hat{\tau}^*(\mathbf{s},p)))]\end{aligned}$$

From the calculation of Proposition 4.6.1 and further noting that Theorem A.4.1 can be applied for the finite moving average process by setting some of the coefficients c_j's to be zero, we obtain

$$\lim_{m\to\infty} \lim_{n\to\infty} n \text{ var}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p))) = \lim_{n\to\infty} n \text{ var}(\text{Re}(\hat{\tau}^*(\mathbf{s},p))) = m_{pp}^{RR}$$

where m_{pq}^{RR} denotes the covariance between the real elements in (p,q)- block of covariance matrix $\mathbf{M}$. Moreover, using the same steps to that given in the proof of Proposition 4.6.1, it can be shown that

$$\lim_{m\to\infty} \lim_{n\to\infty} n \text{cov}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)), \text{Re}(\hat{\tau}^*(\mathbf{s},p))) = m_{pp}^{RR}$$

Thus

$$\lim_{m\to\infty} \limsup_{n\to\infty} \epsilon^{-2} n \text{ var}(\text{Re}(\hat{\tau}_m^*(\mathbf{s},p)) - \text{Re}(\hat{\tau}^*(\mathbf{s},p))) = 0$$

Similar results can be obtained for the imaginary part. This established (4.92). □

Theorem 4.6.4. *Let $X_t, t \in \mathbb{Z}$ be the stationary linear process (3.1), satisfying conditions C1 and C2. Then for $h \in \{1, 2, \ldots\}$, $s \in \mathbb{R}, s \neq 0$*

$$\left[\begin{pmatrix} \text{Re}(\hat{\tau}(\mathbf{s},0)) \\ \text{Im}(\hat{\tau}(\mathbf{s},0)) \end{pmatrix}, \ldots, \begin{pmatrix} \text{Re}(\hat{\tau}(\mathbf{s},h)) \\ \text{Im}(\hat{\tau}(\mathbf{s},h)) \end{pmatrix}\right] \text{ is } AN\left(\left[\begin{pmatrix} \tau(\mathbf{s},0) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \tau(\mathbf{s},h) \\ 0 \end{pmatrix}\right], n^{-1}M\right)$$

where $\mathbf{M}$ is as given in Proposition 4.6.2 above.

Proof. To show the convergence of the estimator $\text{Re}(\hat{\tau}(\mathbf{s},j))$ and $\text{Im}(\hat{\tau}(\mathbf{s},j))$ to the same limit as $\text{Re}(\hat{\tau}^*(\mathbf{s},j))$ and $\text{Im}(\hat{\tau}^*(\mathbf{s},j))$, respectively, with $0 \leq j \leq h$, it suffices to show that as $n \to \infty$

$$n^{1/2} \left\{ \lambda_2 \begin{pmatrix} \text{Re}(\phi^*(s_k, -s_k; j)) \\ \text{Re}(\phi^*(s_k, 0; j)) \\ \text{Re}(\phi^*(0, -s_k; j)) \end{pmatrix} - \lambda_2 \begin{pmatrix} \text{Re}(\phi(s_k, -s_k; j)) \\ \text{Re}(\phi(s_k, 0; j)) \\ \text{Re}(\phi(0, -s_k; j)) \end{pmatrix} \right\} = o_p(1)$$

(and similarly for the imaginary part), where $\phi^*(u,v;j) = n^{-1} \sum_{t=1}^n \exp(i(uX_{t+j} + vX_t))$, $\phi(u,v;j) = (n-j)^{-1} \sum_{t=1}^{n-j} \exp(i(uX_{t+j} + vX_t))$ and $\lambda_2 = \begin{bmatrix} -1 & 1 & 1 \end{bmatrix}$. The required result then follows

from Slutzky's theorem (see p. 92).
Simple algebra gives, for $0 \leq j \leq h$,

$$
\begin{aligned}
& n^{1/2}\mathrm{E}\left|\lambda_2\begin{pmatrix} \mathrm{Re}(\phi^*(s_k,-s_k;j)) \\ \mathrm{Re}(\phi^*(s_k,0;j)) \\ \mathrm{Re}(\phi^*(0,-s_k;j)) \end{pmatrix} - \lambda_2\begin{pmatrix} \mathrm{Re}(\phi(s_k,-s_k;j)) \\ \mathrm{Re}(\phi(s_k,0;j)) \\ \mathrm{Re}(\phi(0,-s_k;j)) \end{pmatrix}\right| \\
& = n^{1/2}\mathrm{E}\left|\lambda_2\begin{pmatrix} \frac{j}{(n-j)}\frac{1}{n}\sum_{t=1}^{n}\cos(is_k(X_{t+j}-X_t)) - \frac{1}{n-j}\sum_{t=n-j+1}^{n}\cos(is(X_{t+j}-X_t)) \\ \frac{j}{(n-j)}\frac{1}{n}\sum_{t=1}^{n}\cos(is_kX_{t+j}) - \frac{1}{n-j}\sum_{t=n-j+1}^{n}\cos(is_kX_{t+j}) \\ \frac{j}{(n-j)}\frac{1}{n}\sum_{t=1}^{n}\cos(-is_kX_t) - \frac{1}{n-j}\sum_{t=n-j+1}^{n}\cos(-is_kX_t) \end{pmatrix}\right| \\
& \leq 6j(n-j)^{-1/2}(\tfrac{n}{n-j})^{1/2}
\end{aligned}
$$

The required result is obtained from $3j(n-j)^{-1/2} \to 0$ and $n/(n-j) \to 1$ as $n \to \infty$. Using the same arguments, similar results can be obtained for the imaginary part. The conclusion of the theorem then follows from Proposition 4.6.3 above. □

Proof of Theorem 4.4.2. Let $\mathbf{g}(\cdot)$ be the function from $\mathbb{R}^{2r\times(h+1)}$ to $\mathbb{R}^{2r\times h}$ defined by

$$
\begin{aligned}
& \mathbf{g}\left(\left[\begin{pmatrix} \hat{\tau}(\mathbf{s},0) \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Re}(\hat{\tau}(\mathbf{s},1)) \\ \mathrm{Im}(\hat{\tau}(\mathbf{s},1)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{\tau}(\mathbf{s},h)) \\ \mathrm{Im}(\hat{\tau}(\mathbf{s},h)) \end{pmatrix}\right]^T\right) \\
& = \left[\begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},1)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},1)) \end{pmatrix}, \begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},2)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},2)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},h)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},h)) \end{pmatrix}\right]^T
\end{aligned}
$$

where for $0 < j \leq h$ and $\hat{\tau}(0) \neq 0$, we have $\mathrm{Re}(\hat{I}(s_i,-s_i;j)) = \frac{\mathrm{Re}(\hat{\tau}(s_i,-s_i;j))}{\hat{\tau}(s_i,-s_i;0)}$ and $\mathrm{Im}(\hat{I}(s_i,-s_i;j)) = \frac{\mathrm{Im}(\hat{\tau}(s_i,-s_i;j))}{\hat{\tau}(s_i,-s_i;0)}$, for $i = 1,\ldots,r$. By applying delta method and Theorem 4.6.4 above, we can show that

$$
\left[\begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},1)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},1)) \end{pmatrix}, \begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},2)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},2)) \end{pmatrix}, \ldots, \begin{pmatrix} \mathrm{Re}(\hat{I}(\mathbf{s},h)) \\ \mathrm{Im}(\hat{I}(\mathbf{s},h)) \end{pmatrix}\right]^T
$$

is asymptotically normal distributed with mean

$$
\begin{aligned}
& \mathbf{g}\left(\left[\begin{pmatrix} \tau(\mathbf{s},0) \\ 0 \end{pmatrix}, \begin{pmatrix} \tau(\mathbf{s},1) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \tau(\mathbf{s},h) \\ 0 \end{pmatrix}\right]^T\right) \\
& = \left[\begin{pmatrix} I(1) \\ 0 \end{pmatrix}, \begin{pmatrix} I(2) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} I(h) \\ 0 \end{pmatrix}\right]^T
\end{aligned}
$$

and variance $n^{-1}\mathbf{DMD^T}$. Here the matrix $\mathbf{M}$ is as given in Proposition 4.6.2 and $\mathbf{D}$ is the Jacobian matrix of $\mathbf{g}(\cdot)$. To obtain the elements of matrix $\mathbf{D}$, we proceed as follows. First, note that the codifference function at lag 0 is a real-valued function. Therefore, for $0 \leq j \leq h$, and $\tau(0) \neq 0$, we obtain $\mathrm{Re}(I(j)) = \frac{\mathrm{Re}(\tau(j))}{\tau(0)} = I(j)$ and $\mathrm{Im}(I(j)) = \frac{\mathrm{Im}(\tau(j))}{\tau(0)} = 0$. It is straightforward to obtain the Jacobian matrix $\mathbf{D}$ as

$$
\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} & 0 & \ldots & 0 \\ D_{21} & 0 & D_{23} & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_{h1} & 0 & 0 & \ldots & D_{h(h+1)} \end{bmatrix} \tag{4.93}
$$

where

$$
D_{l1} = \begin{bmatrix} D_{l1}^{11} & \mathbf{0}_r \\ \mathbf{0}_r & \mathbf{0}_r \end{bmatrix} \tag{4.94}
$$

and

$$D_{l(l+1)} = \begin{bmatrix} D^{11}_{l(l+1)} & \mathbf{0}_r \\ \mathbf{0}_r & D^{11}_{l(l+1)} \end{bmatrix} \tag{4.95}$$

for $l = 1, \ldots, h$, where

$$D^{11}_{l1} = \mathbf{I}_r \left[\frac{-I(l)}{\tau(s_1, -s_1; 0)}, \frac{-I(l)}{\tau(s_2, -s_2; 0)}, \ldots, \frac{-I(l)}{\tau(s_r, -s_r; 0)} \right]^T$$

and

$$D^{11}_{l(l+1)} = \mathbf{I}_r \left[\frac{1}{\tau(s_1, -s_1; 0)}, \frac{1}{\tau(s_2, -s_2; 0)}, \ldots, \frac{1}{\tau(s_r, -s_r; 0)} \right]^T$$

Let's denote w_{ij}, for $i, j = 1, \ldots, h$, the (i,j)-th block element of $\mathbf{DMD}^T$ and m_{ij}, for $i, j = 0, 1, \ldots, h$, the (i,j)-th block element of $\mathbf{M}$. We find that

$$\begin{aligned} w_{ij} &= \begin{bmatrix} \operatorname{cov}(\operatorname{Re}(\hat{I}(\mathbf{s}, i)), \operatorname{Re}(\hat{I}(\mathbf{s}, j))) & \operatorname{cov}(\operatorname{Re}(\hat{I}(\mathbf{s}, i)), \operatorname{Im}(\hat{I}(\mathbf{s}, j))) \\ \operatorname{cov}(\operatorname{Im}(\hat{I}(\mathbf{s}, i)), \operatorname{Re}(\hat{I}(\mathbf{s}, j))) & \operatorname{cov}(\operatorname{Im}(\hat{I}(\mathbf{s}, i)), \operatorname{Im}(\hat{I}(\mathbf{s}, j))) \end{bmatrix} \\ &= D_{i1} m_{00} D_{j1} + D_{i(i+1)} m_{i0} D_{j1} + D_{i1} m_{0j} D_{j(j+1)} + D_{i(i+1)} m_{ij} D_{j(j+1)} \\ &= \begin{bmatrix} D^{11}_{i1} m^{RR}_{00} D^{11}_{j1} + D^{11}_{i(i+1)} m^{RR}_{i0} D^{11}_{j1} + D^{11}_{i1} m^{RR}_{0j} D^{11}_{j(j+1)} + D^{11}_{i(i+1)} m^{RR}_{ij} D^{11}_{j(j+1)} & \mathbf{0}_r \\ \mathbf{0}_r & D^{11}_{i(i+1)} m^{II}_{ij} D^{11}_{j(j+1)} \end{bmatrix} \end{aligned} \tag{4.96}$$

Here m^{RR}_{ij} and m^{II}_{ij} denote the partitions of m_{ij} which correspond to the real and the imaginary components, respectively. □

4.7 Proof of Corollary 4.4.3

Proof of Corollary 4.4.3. As MA(0) is a special case of the linear process (3.1), by applying Theorem 4.4.2, one can conclude the asymptotic normality of

$$\left[\begin{pmatrix} \operatorname{Re}(\hat{I}(\mathbf{s}, 1)) \\ \operatorname{Im}(\hat{I}(\mathbf{s}, 1)) \end{pmatrix}, \begin{pmatrix} \operatorname{Re}(\hat{I}(\mathbf{s}, 2)) \\ \operatorname{Im}(\hat{I}(\mathbf{s}, 2)) \end{pmatrix}, \ldots, \begin{pmatrix} \operatorname{Re}(\hat{I}(\mathbf{s}, h)) \\ \operatorname{Im}(\hat{I}(\mathbf{s}, h)) \end{pmatrix} \right]^T$$

for $h \in \{1, 2, \ldots\}$. The true codifference function of i.i.d. process X_t is

$$\begin{aligned} \tau(s, -s; k) &= -\ln(\mathrm{E}(\exp(is(X_{t+k} - X_t)))) + \ln(\mathrm{E}(\exp(isX_{t+k}))) + \ln(\mathrm{E}(\exp(-isX_t)) \\ &= \begin{cases} -2\sigma^\alpha |s|^\alpha & \text{for } k = 0 \\ 0 & \text{for } k > 0 \end{cases} \end{aligned}$$

which enables us to conclude that the real and the imaginary parts of $I(k) = 0$ whenever $k > 0$.

From (4.96), we obtain that $w_{pq}, p, q > 0$ for $p, q = 1, \ldots, h$, the (p,q)-th block element of $\mathbf{DMD}^T$ is reduced to

$$w_{pq} = D_{p(p+1)} m_{pq} D_{q(q+1)} \tag{4.97}$$

where matrix $D_{k(k+1)}, k = p, q$ is as given in (4.95), with

$$D^{11}_{l(l+1)} = \mathbf{I}_r \left[\frac{1}{-2\sigma^\alpha |s_1|^\alpha}, \frac{1}{-2\sigma^\alpha |s_2|^\alpha}, \ldots, \frac{1}{-2\sigma^\alpha |s_r|^\alpha} \right]^T$$

and where

$$m_{pq} = \lambda \mathbf{L}^p_2 \mathbf{V}_{pq} \mathbf{L}^q_2 \lambda^T \tag{4.98}$$

with λ is as given in (4.80), and the elements of the matrix $\mathbf{L}_2^k, k = p, q$ and the covariance matrix $\mathbf{V}_{pq}$ will be given below.

Let us denote

$$\mathbf{V}_{pq}^{RR}(i,j) = [\mathrm{cov}(\mathrm{Re}(\phi_l(s_i,p)), \mathrm{Re}(\phi_m(s_j,q)))]_{l,m=1,2,3}$$

and

$$\mathbf{V}_{pq}^{II}(i,j) = [\mathrm{cov}(\mathrm{Im}(\phi_l(s_i,k)), \mathrm{Im}(\phi_m(s_j,k)))]_{l,m=1,2,3}$$

as the (i,j)-th block elements of $\mathbf{V}_{pq}^{RR}$ and $\mathbf{V}_{pq}^{II}$, respectively. Using identities (4.84)-(4.86) (and the identities for the imaginary part afterwards) in p.52, we can obtain their components, as follows

$$\begin{aligned}\mathrm{cov}(\mathrm{Re}(\phi_1(s_i,p)), \mathrm{Re}(\phi_1(s_j,q))) &= \mathrm{cov}(\cos(-s_iX_t), \cos(-s_jX_t)) \\ &= \tfrac{1}{2}\{e^{-\sigma^\alpha|s_i+s_j|^\alpha} + e^{-\sigma^\alpha|s_i-s_j|^\alpha}\} - e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha)}\end{aligned}$$

$$\mathrm{cov}(\mathrm{Re}(\phi_1(s_i,p)), \mathrm{Re}(\phi_2(s_j,q))) = \mathrm{cov}(\mathrm{Re}(\phi_2(s_i,p)), \mathrm{Re}(\phi_1(s_j,q))) = \mathrm{cov}(\mathrm{Re}(\phi_1(s_i,p)), \mathrm{Re}(\phi_1(s_j,q)))$$

$$\mathrm{cov}(\mathrm{Re}(\phi_2(s_i,p)), \mathrm{Re}(\phi_2(s_j,q))) = \mathrm{cov}(\mathrm{Re}(\phi_1(s_i,p)), \mathrm{Re}(\phi_1(s_j,q)))$$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Re}(\phi_1(s_i,p)), \mathrm{Re}(\phi_3(s_j,q))) \\ &= \mathrm{cov}(\cos(-s_iX_t), \cos(s_j(X_{t+q}-X_t))) + \mathrm{cov}(\cos(-s_iX_{t+q}), \cos(s_j(X_{t+q}-X_t))) \\ &= e^{-\sigma^\alpha(|s_j|^\alpha+|s_i-s_j|^\alpha)} + e^{-\sigma^\alpha(|s_j|^\alpha+|s_i+s_j|^\alpha)} - 2e^{-\sigma^\alpha(|s_i|^\alpha+|2s_j|^\alpha)}\end{aligned}$$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Re}(\phi_3(s_i,p)), \mathrm{Re}(\phi_1(s_j,q))) \\ &= \mathrm{cov}(\cos(-s_jX_t), \cos(s_i(X_{t+p}-X_t))) + \mathrm{cov}(\cos(-s_jX_{t+p}), \cos(s_i(X_{t+p}-X_t))) \\ &= e^{-\sigma^\alpha(|s_i|^\alpha+|s_i-s_j|^\alpha)} + e^{-\sigma^\alpha(|s_i|^\alpha+|s_i+s_j|^\alpha)} - 2e^{-\sigma^\alpha(|s_j|^\alpha+|2s_i|^\alpha)}\end{aligned}$$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Re}(\phi_2(s_i,p)), \mathrm{Re}(\phi_3(s_j,q))) \\ &= \mathrm{cov}(\cos(s_iX_{t+q}), \cos(s_j(X_{t+q}-X_t))) + \mathrm{cov}(\cos(s_iX_{t+p}), \cos(s_j(X_{t+p+q}-X_{t+p}))) \\ &= e^{-\sigma^\alpha(|s_j|^\alpha+|s_i-s_j|^\alpha)} + e^{-\sigma^\alpha(|s_j|^\alpha+|s_i+s_j|^\alpha)} - 2e^{-\sigma^\alpha(|s_i|^\alpha+|2s_j|^\alpha)}\end{aligned}$$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Re}(\phi_3(s_i,k)), \mathrm{Re}(\phi_2(s_j,k))) \\ &= \mathrm{cov}(\cos(s_jX_{t+k}), \cos(s_i(X_{t+k}-X_t))) + \mathrm{cov}(\cos(s_jX_{t+k}), \cos(s_i(X_{t+2k}-X_{t+k}))) \\ &= e^{-\sigma^\alpha(|s_i|^\alpha+|s_i-s_j|^\alpha)} + e^{-\sigma^\alpha(|s_i|^\alpha+|s_i+s_j|^\alpha)} - 2e^{-\sigma^\alpha(|s_j|^\alpha+|2s_i|^\alpha)}\end{aligned}$$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Re}(\phi_3(s_i,p)), \mathrm{Re}(\phi_3(s_j,q))) \\ &= \mathrm{cov}(\cos(s_i(X_{t+p}-X_t)), \cos(s_j(X_{t+q}-X_t))) + \mathrm{cov}(\cos(s_i(X_{t+p+q}-X_{t+q})), \cos(s_j(X_{t+q}-X_t))) \\ &\quad + \mathrm{cov}(\cos(s_i(X_{t+p}-X_t)), \cos(s_j(X_{t++p+q}-X_{t+p}))) + c_{\mathrm{Re}}^{pq}\end{aligned}$$

where

$$c_{\mathrm{Re}}^{pq} = \begin{cases} 0 & \text{if } p = q \\ \mathrm{cov}(\cos(\,s_i(X_{t+q}-X_{t+q-p})), \cos(\,s_j(X_{t+q}-X_t))) & \text{if } q > p \\ \mathrm{cov}(\cos(\,s_i(X_{t+p}-X_t)), \cos(\,s_j(X_{t+p}-X_{t+p-q}))) & \text{if } p > q \end{cases}$$

yielding for $p = q$

$$\begin{aligned}\mathrm{cov}(\mathrm{Re}(\phi_3(s_i,p)), \mathrm{Re}(\phi_3(s_j,q))) &= \tfrac{1}{2}e^{-2\sigma^\alpha|s_i+s_j|^\alpha} + \tfrac{1}{2}e^{-2\sigma^\alpha|s_i-s_j|^\alpha} - 3e^{-\sigma^\alpha(2|s_i|^\alpha+|2s_j|^\alpha)} \\ &\quad + e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha+|s_i-s_j|^\alpha)} + e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha+|s_i+s_j|^\alpha)}\end{aligned}$$

and for $p \neq q$

$$\begin{aligned}\mathrm{cov}(\mathrm{Re}(\phi_3(s_i,p)),\mathrm{Re}(\phi_3(s_j,q))) &= 2e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha+|s_i-s_j|^\alpha)} \\ &\quad + 2e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha+|s_i+s_j|^\alpha)} - 4e^{-\sigma^\alpha(2|s_i|^\alpha+|2s_j|^\alpha)}\end{aligned}$$

$$\mathrm{cov}(\mathrm{Im}(\phi_1(s_i,p)),\mathrm{Im}(\phi_1(s_j,q))) = \mathrm{cov}(\sin(-s_iX_t),\sin(-s_jX_t)) = \tfrac{1}{2}\{e^{-\sigma^\alpha|s_i-s_j|^\alpha} - e^{-\sigma^\alpha|s_i+s_j|^\alpha}\}$$

$$\mathrm{cov}(\mathrm{Im}(\phi_1(s_i,p)),\mathrm{Im}(\phi_2(s_j,q))) = \mathrm{cov}(\mathrm{Im}(\phi_2(s_i,p)),\mathrm{Im}(\phi_1(s_j,q))) \quad = -\mathrm{cov}(\mathrm{Im}(\phi_1(s_i,p)),\mathrm{Im}(\phi_1(s_j,q)))$$

$$\mathrm{cov}(\mathrm{Im}(\phi_2(s_i,p)),\mathrm{Im}(\phi_2(s_j,q))) = \mathrm{cov}(\mathrm{Im}(\phi_1(s_i,p)),\mathrm{Im}(\phi_1(s_j,q)))$$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Im}(\phi_3(s_i,p)),\mathrm{Im}(\phi_3(s_j,q))) \\ &= \mathrm{cov}(\sin(s_i(X_{t+p}-X_t)),\sin(s_j(X_{t+q}-X_t))) + \mathrm{cov}(\sin(s_i(X_{t+p+q}-X_{t+q})),\sin(s_j(X_{t+q}-X_t))) \\ &\quad + \mathrm{cov}(\sin(s_i(X_{t+p+q}-X_{t+p})),\sin(s_j(X_{t+p}-X_t))) + c^{pq}_{\mathrm{Im}}\end{aligned}$$

where

$$c^{pq}_{\mathrm{Im}} = \begin{cases} 0 & \text{if } p = q \\ \mathrm{cov}(\sin(s_i(X_{t+q}-X_{t+q-p})),\sin(s_j(X_{t+q}-X_t))) & \text{if } q > p \\ \mathrm{cov}(\sin(s_i(X_{t+p}-X_t)),\sin(s_j(X_{t+p}-X_{t+p-q}))) & \text{if } p > q \end{cases}$$

yielding for $p = q$

$$\begin{aligned}&\mathrm{cov}(\mathrm{Im}(\phi_3(s_i,k)),\mathrm{Im}(\phi_3(s_j,k))) \\ &= \tfrac{1}{2}e^{-2\sigma^\alpha|s_i-s_j|^\alpha} - \tfrac{1}{2}e^{-2\sigma^\alpha|s_i+s_j|^\alpha} + e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha+|s_i+s_j|^\alpha)} - e^{-\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha+|s_i-s_j|^\alpha)}\end{aligned}$$

and $\mathrm{cov}(\mathrm{Im}(\phi_3(s_i,k)),\mathrm{Im}(\phi_3(s_j,k))) = 0$ for $p \neq q$. The other elements are all 0. The elements of $\mathbf{L}_2^k, k = p,q$ are as given in (4.87), where the elements of $\mathbf{d}_i^k, i = 1,\ldots,r$ are

$$\begin{aligned}d_i^k(1,1) &= (\mathrm{Re}(\Phi(0,-s_i;k)))^{-1} = e^{\sigma^\alpha|s_i|^\alpha} \\ d_i^k(2,2) &= (\mathrm{Re}(\Phi(s_i,0;k)))^{-1} = e^{\sigma^\alpha|s_i|^\alpha} \\ d_i^k(3,3) &= (\mathrm{Re}(\Phi(s_i,-s_i;k)))^{-1} = e^{2\sigma^\alpha|s_i|^\alpha})\end{aligned}$$

As from (4.98) we obtain

$$m^{RR}_{pq} = \mathrm{cov}(\mathrm{Re}(\hat{\tau}(\mathbf{s},p)),\mathrm{Re}(\hat{\tau}(\mathbf{s},q))) = (\mathbf{I}_r \otimes \lambda_1)d^pV^{RR}_{pq}d^q(\mathbf{I}_r \otimes \lambda_1^T)$$

and

$$m^{II}_{pq} = \mathrm{cov}(\mathrm{Im}(\hat{\tau}(\mathbf{s},p)),\mathrm{Im}(\hat{\tau}(\mathbf{s},q))) = (\mathbf{I}_r \otimes \lambda_1)d^pV^{RR}_{pq}d^q(\mathbf{I}_r \otimes \lambda_1^T)$$

then the (i,j)-th element of m^{RR}_{pq} and m^{II}_{pq} is obtained from

$$m^{RR}_{pq}(i,j) = \lambda_1 d_i^pV^{RR}_{pq}(i,j)d_j^q\lambda_1^T$$

and

$$m^{II}_{pq}(i,j) = \lambda_1 d_i^pV^{II}_{pq}(i,j)d_j^q\lambda_1^T$$

which therefore after a simple algebra, we obtain for $p = q = k$

$$\begin{aligned}m^{RR}_{kk}(i,j) &= e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i-s_j|^\alpha)}\left\{\tfrac{1}{2}e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i-s_j|^\alpha)} - 1\right\} \\ &\quad + e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i+s_j|^\alpha)}\left\{\tfrac{1}{2}e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i+s_j|^\alpha)} - 1\right\} + 1\end{aligned}$$

$$\begin{aligned}m^{II}_{kk}(i,j) &= e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i-s_j|^\alpha)}\left\{\tfrac{1}{2}e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i-s_j|^\alpha)} - 1\right\} \\ &\quad + e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i+s_j|^\alpha)}\left\{1 - \tfrac{1}{2}e^{\sigma^\alpha(|s_i|^\alpha+|s_j|^\alpha-|s_i+s_j|^\alpha)}\right\}\end{aligned}$$

and for $p \neq q$, $m^{RR}_{pq}(i,j) = 0$ and $m^{II}_{pq}(i,j) = 0$. The required result follows directly from (4.97). □

Chapter 5

Application

In the previous chapter, we considered estimators of the codifference and the normalized codifference function and established the asymptotic properties of the proposed estimators, where the underlying process was strictly stationary heavy-tailed linear process. This chapter deals with more practical problems related to the results presented in Section 4.4. In the first section of this chapter, we address some problems for the calculation of the sample codifference and the sample normalized codifference function for time series data with finite size. In Section 5.2, we present simulation studies for the application of the sample normalized codifference for identifying the order and estimating the coefficients of pure moving average processes. In the last section, we discuss the application of the codifference function for Portmanteau-type test of randomness, i.e., testing for independence against serial dependence alternatives.

5.1 Practical considerations

5.1.1 The choice of s

Before we proceed, we make a remark about the sample and the population codifference function. From (4.4) and the fact that all c_j's are real, we have that the codifference function of the models we consider here is a real-valued function but the estimator (4.67) is complex. Therefore, one possibility is to use only the real part of the estimator. Because in practice we are working with a finite sample, the imaginary part of the estimator is still present, but will vanish asymptotically.

In what follows, we are only working with the estimator of the normalized codifference $\hat{I}(k)$. Hence, we note that unlike the true normalized codifference (4.3), from (4.67) and Corollary 4.4.3, one can see that the sample normalized codifference function and its limiting variance depend on $\mathbf{s} = \{s_1, \ldots, s_r\}$. Apparently $\hat{I}(\cdot)$ is defined for all $\mathbf{s} > \mathbf{0}$, and from Theorem 4.4.1, we know that it is a consistent estimator. However, in a finite sample, the accuracy of the estimates to the population values depends on the choice of $\mathbf{s}$. Therefore, for estimation, $\mathbf{s}$ is a design parameter which has to be chosen appropriately. In other words, $\hat{I}(\cdot)$ should be calculated from those values of $\mathbf{s}$ which gives the most accurate estimates of the true function $I(\cdot)$.

To be more precise, in practice the number of grid points r and more importantly, the location of $s_1, \ldots, s_r$, have to be chosen. As the normalized codifference function is defined based on the *ecf*, we can apply here the known results for *ecf*. For a fixed r, Koutrouvelis (1980*a*) and Kogon and Williams (1998) show that for calculating *ecf*, the location of the grid points $s_1, \ldots, s_r$, should be chosen close, but not equal to zero. It has been shown in this case, the *ecf* will most accurately estimate the characteristic function. This particular choice of grid points seems to be reasonable for calculating $\hat{I}(\cdot)$, as for instance, can be shown in Figure 5.2. To determine the location of s_i's.

we suggest to plot $\mathrm{Re}\,\hat{I}(k)$ within the interval $0.01 \leq s \leq 2$, for some values of lag $k > 0$. These graphs will show the interval of s around zero which has relatively small bias, as the best location for evaluating the estimator. The best choice for the interval of s clearly depends on the data itself and in general also on the lag k. However, we suggest to use $s_a = 0.01$ as the left bound of the interval, where the best choice for the right bound can be determined from the graphs, i.e., as the threshold point s_b where the graphs $\mathrm{Re}\,\hat{I}(k)$, for some lag k, are still relatively flat. The individual choices for s_i's can be chosen in one of two ways:

1. If we wish to use equal spacing of s_i's, we can set $\mathbf{s} = \{s_1 = 0.01, 0.01 + i\frac{s_b - s_a}{r-1}, s_b\}, i = 1, \ldots, r-2$. To obtain r, we can minimize the determinant of covariance matrix in (4.70). Unfortunately, the covariance matrix in (4.70) depends on the unknown parameters c_j's, α, σ and the distance between s_i's. One possibility is to replace (4.70) with its consistent estimate. However, here we consider a different approach for choosing the s_i's, i.e., with the help of the trajectories of the estimator, for instance, as given in Figure 5.2. Figure 5.2 indicates that these trajectories depend on the sample size n and, more strongly, on α. From our numerical studies, we observe that for given α, the behavior of these trajectories is typical for arbitrary lag k. For $\alpha = 2$, $\hat{I}(k)$ is relatively smooth, where for $\alpha < 2$, $\hat{I}(k)$ has an erratic behavior, and this behavior will be stronger when α goes farther away from 2. This result suggests that when $\alpha = 2$, there is no benefit by choosing the s_i's very close to each other, and conversely for $\alpha < 2$. This conjecture can be checked in i.i.d. case using the determinant of the covariance matrix (4.71). Here, we propose to use s_i's with distance $d = 0.01$ for $\alpha \leq 1$, $0.01 < d \leq 0.05$ for $1 < \alpha \leq 1.5$, $0.05 < d \leq 0.1$ for $1.5 < \alpha < 2$ and $d = 0.1$ or larger for $\alpha = 2$. Especially in i.i.d. case we can show that these choices are sufficient, in the sense that for given α, choosing a smaller distance between grid points will not significantly decrease the determinant of the covariance matrix (4.71). Notice that in practice it is not necessary to know α. As the erratic behavior of the estimator is typical for given α, we can observe this property using the plot of $\mathrm{Re}\,\hat{I}(k)$ within an interval near zero, for some values of lag $k > 0$.

2. If equal spacing is not considered, when r has been fixed, the s_i's can be chosen using the determinant of covariance matrix in (4.70). Here, we can use a similar consideration as above, i.e., we choose the s_i's sufficiently close, depending on the erratic behavior of the estimates $\mathrm{Re}\,\hat{I}(\cdot)$.

The last thing to consider is the number of points r. For $r \geq 1$, the final estimate $\hat{I}(k)$ can be defined as the weighted sum of the estimates at the grid points $s_1, \ldots, s_r$, i.e., we define $\hat{I}(k) = \sum_{i=1}^{r} w_i \hat{I}(s_i, -s_i; k)$ with $\sum_{i=1}^{r} w_i = 1$. For instance, we can use a simple average with $w_i = 1/r$ or a negative exponentially weighted average with $w_i = \exp(-s_i^2)/\sum_{j=1}^{r}\exp(-s_j^2)$. In i.i.d. case, we obtain that by averaging the estimates at different points, the asymptotic variance of the estimator will be smaller or equal to the variance of the estimator obtained at single point. which can be seen directly from Figure 5.1. Figure 5.1 shows that for $\alpha = 2$, there is no difference in terms of the asymptotic variance between estimating $\hat{I}(\cdot)$ either at a single point or at more points, whereas for $\alpha < 2$, the difference is significant, especially when α is small. Furthermore, Figure 5.1 also shows that the smaller α is, the smaller the covariance between the grid points. In the finite sample case, this fact agrees with the typical erratic behavior of the plot of $\mathrm{Re}\,\hat{I}(\cdot)$, both for the i.i.d. and non i.i.d samples, for instance, as shown in Figure 5.2 and Figure 5.3. Based on these results and from our numerical studies, in the finite sample case we suggest the choice of the number of grid points r as follows. For $\alpha = 2$ (i.e., for smooth graphs of $\mathrm{Re}\,\hat{I}(k)$, $k > 0$), we observe that $r = 1$ is sufficient, whereas for $\alpha < 2$ (i.e., for erratic graphs of $\mathrm{Re}\,\hat{I}(k)$, $k > 0$), at least two points should be chosen, and more points are required when α is farther away from 2 (i.e., for more erratic graphs of $\mathrm{Re}\,\hat{I}(k)$, $k > 0$). It is important to note that from our simulation experience, the accuracy of the estimator is more sensitive to the location of grid points than to the number of points r.

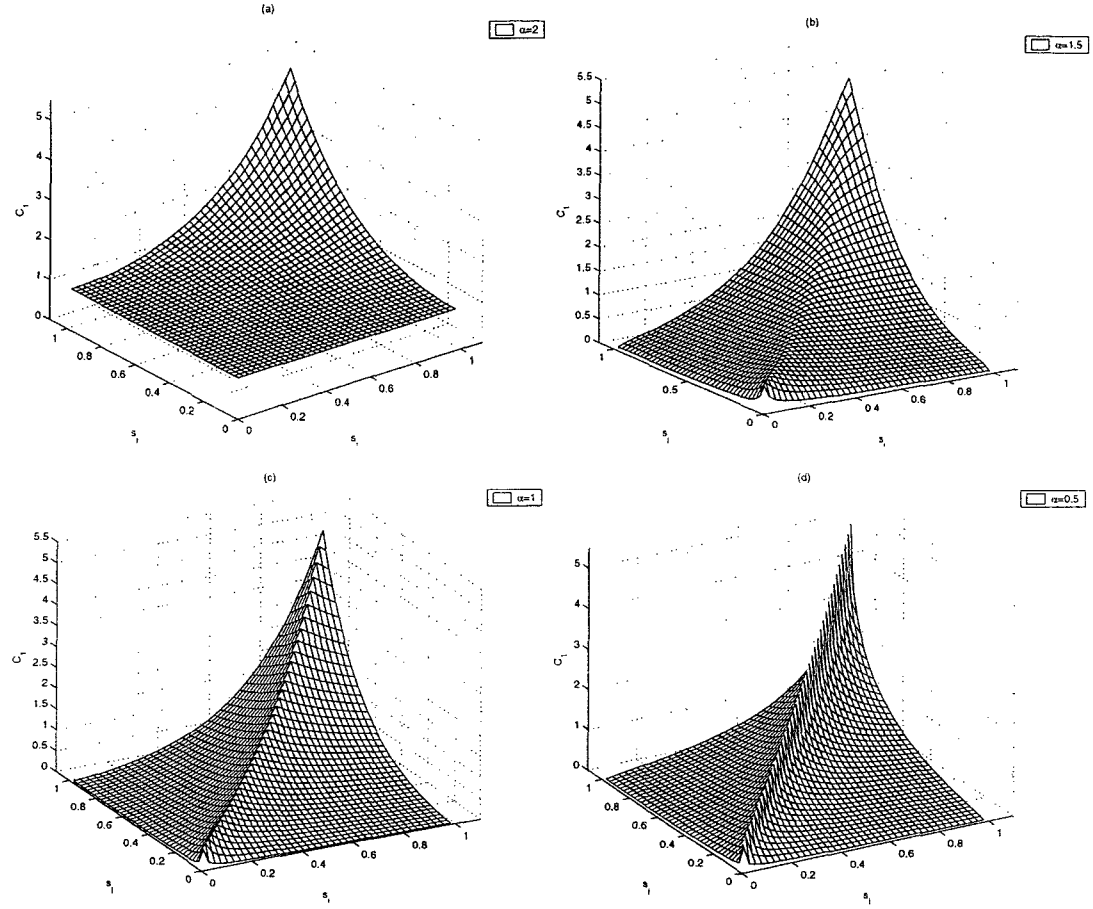


Figure 5.1: Plots of $W_1(i,j)$ as given in (4.73), for $s_i, s_j \in [0.01, 1]$, and some α's.

In the following subsection, we will investigate the choice of grids points through monte carlo simulations.

5.1.2 Simulation results

To investigate the proposed choice of s and also the finite sample behavior of the estimator, we run several monte-carlo simulations using `R` version 1.9.0, where we use function `rstable` in the extension package `stable`, to generate the unit symmetric α stable innovations and function `arima.sim` in the package `stats` to generate $X_t = \epsilon_t + c_1\epsilon_{t-1} + c_2\epsilon_{t-2}$ processes where (c_1, c_2) are

$$I.\,(2, 1.111)\quad II.\,(-1, 0.5)\quad III.\,(0.55, 0.05)\quad IV.\,(-0.4, 0.7) \tag{5.1}$$

and from now on we refer to these as experiment I - experiment IV, respectively.

In Gaussian framework, models in experiments I, II and III were examined in Bhansali (1983). The roots of the polynomial $1 + c_1 z + c_2 z^2 = 0$ are as follows. In experiment I, the roots are $-0.9 \pm 0.3i$, close to the invertibility region. In experiment II and IV, the roots are $0.5 \pm 0.5i$, and $-0.2857 \pm 0.247i$, so the absolute values of the roots are 0.71, and 0.378, respectively. In these experiments. the models have similar roots properties which are neither too close to 1 nor to 0.

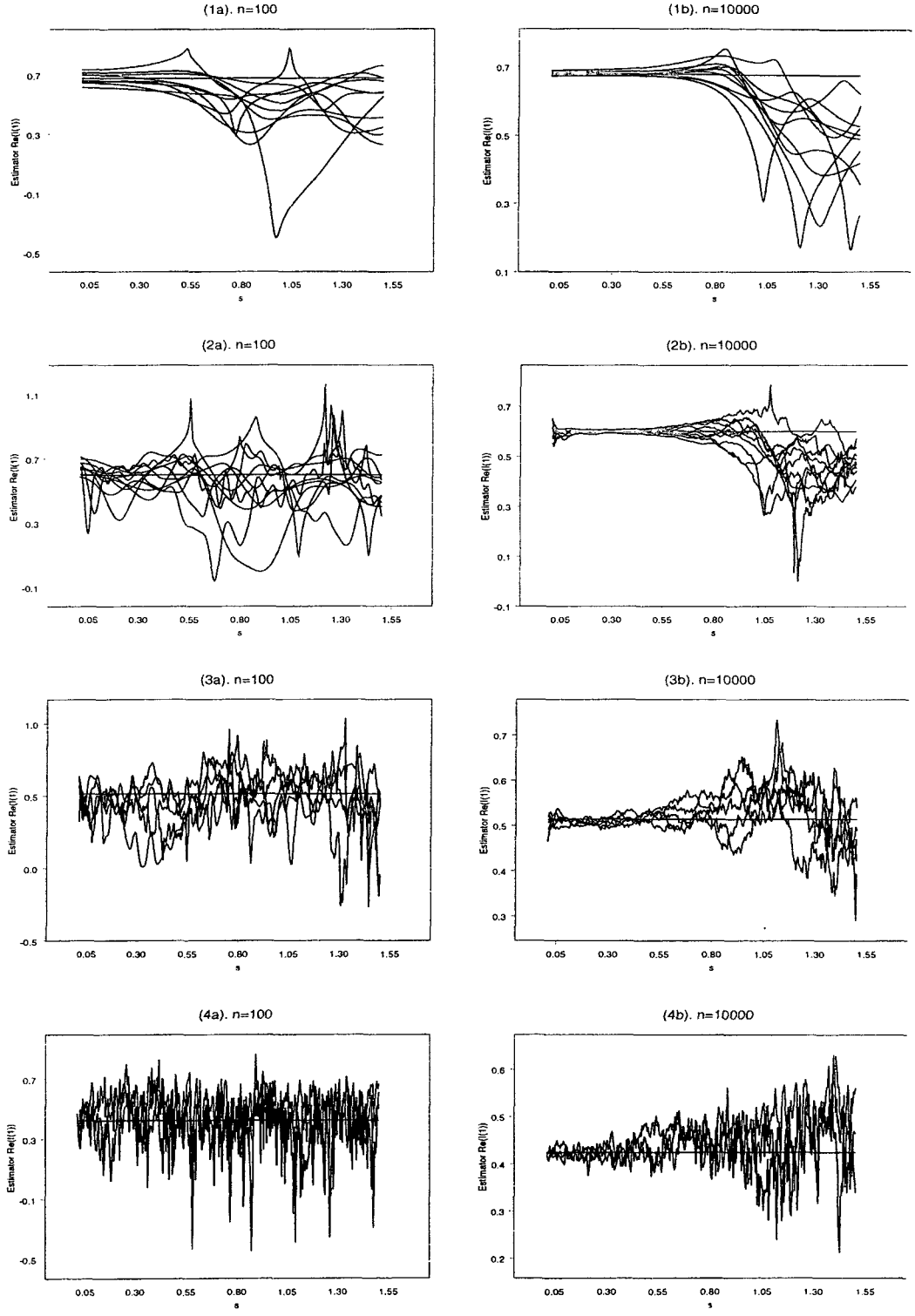


Figure 5.2: Plots of $\mathrm{Re}\,\hat{I}(1)$ for several simulation runs where $(1a-1b).\alpha = 2$, $(2a-2b).\alpha = 1.5$, $(3a-3b).\alpha = 1$ and $(4a-4b).\alpha = 0.5$ and $\sigma = 1$, $s \in [0.01, 1.55]$. The data are generated from experiment I, that is MA(2) process with $c_0 = 1$, $c_1 = 2$ and $c_2 = 1.111$. The straight lines denote the true values of $I(1)$.

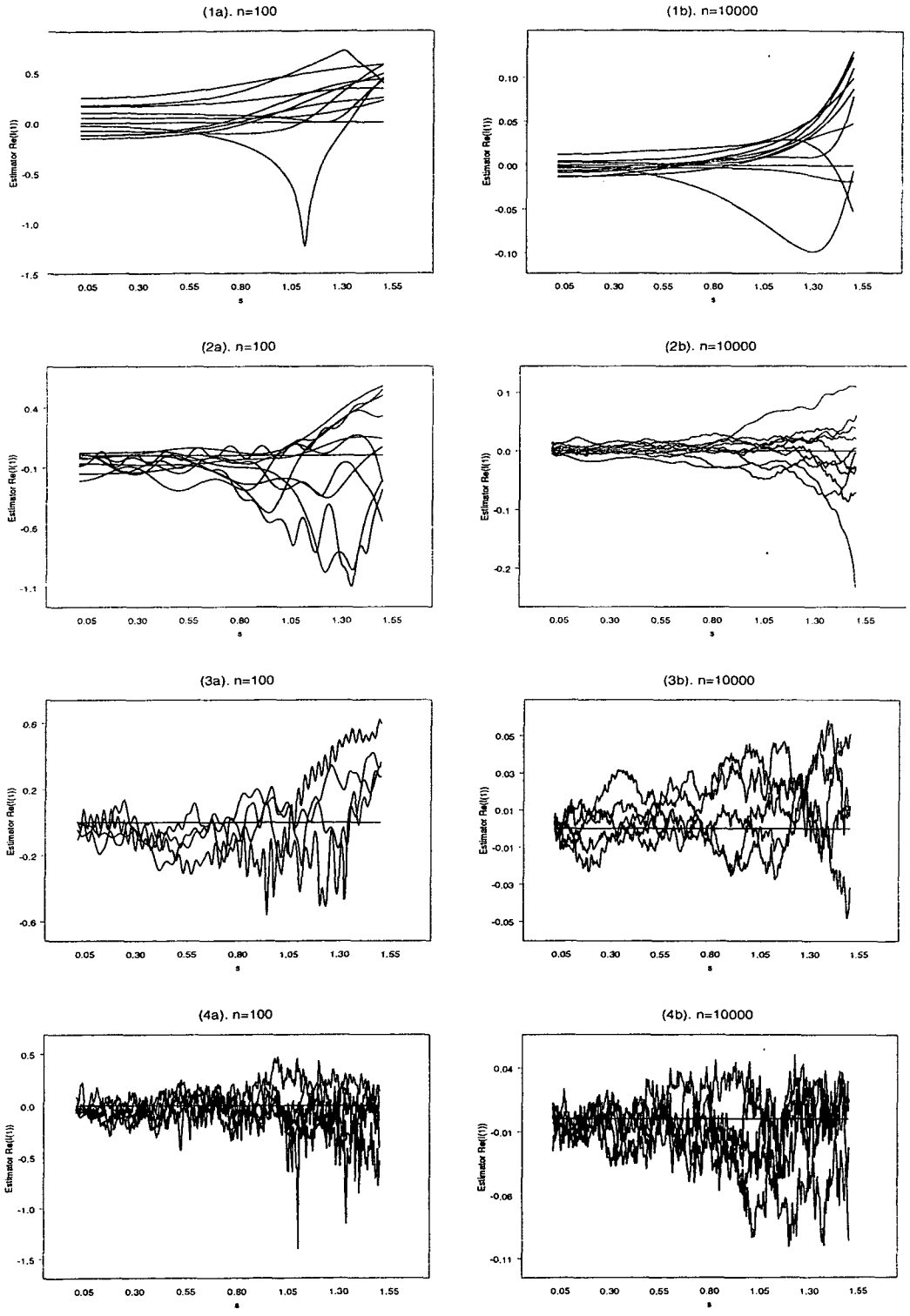


Figure 5.3: Plots of $\mathrm{Re}\,\hat{I}(1)$ for several simulation runs where $(1a-1b).\alpha = 2$, $(2a-2b).\alpha = 1.5$, $(3a-3b).\alpha = 1$ and $(4a-4b).\alpha = 0.5$ and $\sigma = 1$, $s \in [0.01, 1.55]$. The data are generated from i.i.d. $S\alpha S$ processes. The straight lines denote the true values of $I(1) = 0$.

In experiment III, the roots are real-valued, equal to -0.435 and -0.115, one close to 0.5 and the other close to 0.

For $\alpha = 2$, the true values of the normalized codifference (equal to the correlation function) at lag k , $(I(1), I(2)) = (\rho(1), \rho(2))$ in experiment I-IV are:

$$(I).(0.677, 0.178),\quad (II).(-0.667, 0.222),\quad (III).(0.443, 0.038),\quad (IV).(-0.412, 0.424)$$

In experiment I and II, the values of I(k) are closer to 1 at lag 1 and not too close to 0 at lag 2 while for experiment III, at lag 1 close to 0.5 but almost 0 at lag 2. For the last experiment, at lag 1 it is negative but it is positive at lag 2, with absolute values near 0.5.

In order to see the performance of the estimator, we simulate the time series in experiment I - IV for several different values of α with $\sigma = 1$ and two sample sizes, the "small" one is $n = 100$ and the "large" one is $n = 1000$. All experiments are replicated $T = 1000$ times. The estimates of the normalized codifference are calculated for lags 1 till lag 10. Figure 5.2 suggests that in the interval $0.01 \leq s \leq 0.5$, $\mathrm{Re}\,\hat{I}(\cdot)$ is relatively less biased, although the best interval of s depends on the index α. For checking the best location of s and also the choice of grid points, we choose several different sets of $\mathbf{s}_i = \{s_1, \ldots, s_r\}, i = 1, 2, \ldots, 28$. Here we consider the equidistant and non equidistant grid points. The complete listing of the choices is as follows : $\mathbf{s}_1 = \{0.01\}$, $\mathbf{s}_2 = \{0.1\}$, $\mathbf{s}_3 = \{0.2\}$, $\mathbf{s}_4 = \{0.3\}$, $\mathbf{s}_5 = \{0.5\}$, $\mathbf{s}_6 = \{1\}$, $\mathbf{s}_7 = \{0.01, 0.1\}$, $\mathbf{s}_8 = \{0.01, 0.2\}$, $\mathbf{s}_9 = \{0.01, 0.5\}$, $\mathbf{s}_{10} = \{0.01, 1\}$, $\mathbf{s}_{11} = \{0.1, 0.2\}$, $\mathbf{s}_{12} = \{0.1, 0.5\}$, $\mathbf{s}_{13} = \{0.1, 1\}$, $\mathbf{s}_{14} = \{0.5, 1\}$, $\mathbf{s}_{15} = \{0.01, 0.1, 0.2\}$, $\mathbf{s}_{16} = \{0.01, 0.1, 0.5\}$, $\mathbf{s}_{17} = \{0.01, 0.1, 1\}$, $\mathbf{s}_{18} = \{0.1, 0.2, 0.3\}$, $\mathbf{s}_{19} = \{0.1, 0.3, 0.5\}$, $\mathbf{s}_{20} = \{0.01, 0.5, 1\}$, $\mathbf{s}_{21} = \{0.1, 0.5, 1\}$, $\mathbf{s}_{22} = \{0.1, 0.2, 0.3, 0.4, 0.5\}$, $\mathbf{s}_{23} = \{0.1, 0.2, \ldots, 1\}$, $\mathbf{s}_{24} = \{0.01, 0.06, 0.11, 0.16, 0.21\}$, $\mathbf{s}_{25} = \{0.01, 0.02, \ldots, 0.2\}$, $\mathbf{s}_{26} = \{0.01, 0.02, \ldots, 0.1\}$, $\mathbf{s}_{27} = \{0.11, 0.12, \ldots, 0.2\}$ and $\mathbf{s}_{28} = \{0.5, 0.55, \ldots, 1\}$. For each choice of $\mathbf{s}_i$ in run h, the final estimate is calculated as the weighted average of estimates among the choices of grid points $s_{ij}, j = 1, \ldots, r_i$, denoted by $\mathrm{Re}\,\hat{I}(\cdot)_{ih} = \sum_{j=1}^{r_i} w_{ij}\,\mathrm{Re}\,\hat{I}(s_{ij}, -s_{ij}, \cdot)_h$, where $\mathrm{Re}\,\hat{I}(s_{ij}, -s_{ij}, \cdot)_h$ denotes the real part of the estimates of the normalized codifference in run h at certain lags, calculated at $s_{ij}, j = 1, \ldots, r_i$. Here we consider two methods for weighting the estimates, first we use a simple average of the estimates (which is denoted by method Avg.) and the second, we use a negative exponential weighted average (which is denoted by method Exp.). Results of simulation studies are summarized in Table 5.1.1 - 5.1.4. In the tables, we also record the best choices of s, which are defined as the values of grid points, which minimize the sum of mean absolute deviation (MAD) of estimates at lag 1 and lag 2, among all considered choices of grid points above. Here, MAD at lag k and for grid $\mathbf{s}_i$ is defined as $MAD_{ik} = \frac{1}{T}\sum_{h=1}^{T} |\,\mathrm{Re}\,\hat{I}(k)_{ih} - I(k)|$, $k = 1, 2$. For the comparison's sake, when $\alpha = 2$ we also record the estimates of sample (central) ACF.

From the simulation results, as expected, we observe that the estimation accuracy will be improved when the sample size is increased. Furthermore, throughout the simulation studies, the results indicate that the accuracy of the estimates of the normalized codifference function depends on the choice of grid points s, where the optimal choice of grids depends on the index α and the sample size n. When $\alpha = 2$, surprisingly that under suitable choices of grid points, we find in some cases, the sample normalized codifference can provide a better estimate (in terms of total MAD for the first two lags) to the true values (of the normalized codifference, which is equal to ACF) than the estimates given by the SACF. When $\alpha < 2$, it seems that there is a great benefit by evaluating $\mathrm{Re}\,\hat{I}(\cdot)$ at several points of s, that is $r > 1$, where under the appropriate choice of grid points s, the performance of the weighting methods (the simple average and the exponential weight) are approximately the same. For MA(2) models, in all cases we consider here, we also find that the estimation accuracies are significantly better if we choose the grids point $s < 0.5$ than $s \geq 0.5$. In general, there is a benefit in terms of the estimation accuracy to include a point close to zero. We further observe that when $\alpha < 1.5$, the choice of equidistant grids with a distance between 0.01 and 0.05 seems to be adequate. For $\alpha \geq 1.5$, a distance 0.1 seems to be adequate, because a smaller grid distance does not really improve the accuracy of the estimates. As a general conclusion, from this simulation studies, we may conclude that the optimal choice of grid points s will follow the lines of our proposed choice of grids points s as in Section 5.1.1.

Using the findings in Section 5.1.1 and simulation studies in this section, in the practical situation, to obtain the estimates of the normalized codifference with a good accuracy for the real data, first we suggest to plot $\mathrm{Re}\,\hat{I}(k)$ within the interval $s \in [0.01, 2]$, for some values of lag $k > 0$. These graphs will show two important things. First, these graphs suggest the interval of s near zero which has relatively small bias (i.e., the region $0.01 < s < s_b$, s_b denotes the threshold point of s where the graphs $\mathrm{Re}\,\hat{I}(k)$, for some k, are still relatively flat), as the best location for evaluating $\hat{I}(\cdot)$. Secondly, it reveals the erratic behavior of the estimates. When the graphs are smooth, the choice of one point $s = 0.01$ is sufficient for estimating $\mathrm{Re}\,\hat{I}(k)$. If the graphs are erratic, at least two points are required and more points are better for more erratic graphs. If the equidistant points $s_1, \ldots, s_r$ are considered, when the graphs are highly erratic, we can use a small distance between points, e.g., $d = 0.01$, where for less erratic graphs, we can use a bigger distance, such as $d = 0.05$ or $d = 0.1$. If the non equidistant points are used, we should include one point close to zero in the choice of s_i's and the chosen points are sufficiently close to each other. Finally, we define the final estimate as the weighted average of the estimates on $s_1, \ldots, s_r$.

Remark 5.1.1. Notice that in this thesis we have considered a method for calculating the codifference and the normalized codifference function directly from the data. As an obvious alternative, once one knows the estimated parameters and the orders of the estimated models, one could estimate the codifference and the normalized codifference using equation (4.4). The methods for estimating the parameters of stable ARMA models have been reviewed in Section 3.3. Notice that for small order MA and AR processes, the tail index α can be well estimated using a quantile based estimator (i.e., McCulloch's method, see Section 2.4), see, e.g., Adler *et al.* (1998*a*) for simulation evidences. In our opinion, for inference purpose, such as a hypothesis testing of independence (e.g. Section 5.3), the "direct" estimation method is more preferable than estimating the codifference function via estimated parameters.

To illustrate this method, we consider $AR(2)$ processes with the roots: $(r_1, r_2) = (2, 5)$ and also $(r_1, r_2) = \frac{1}{3}(1 \pm i\sqrt{3})$, for some αs. The $\hat{\alpha}$ is estimated using McCulloch method, where the coefficients of AR(2) processes are estimated using the following methods (see Section 3.3 for description of each methods): Least Absolute Deviation (LAD), Least Square (LS), Generalized Yule Walker via Autocorrelation function (ACF) and Gallagher's autocovariation function (Acov). The estimator of the normalized codifference function is calculated using (4.4), approximated until the order $j = 200$. For the direct estimation procedure, the normalized codifference is evaluated at $\mathbf{s} = \{0.01, 0.02, \ldots, 0.05\}$, and the final estimate is defined as the average of the estimates at these points. The sample size used is $n = 1000$ and 1000 independent replication of the data were generated. The results of estimation are summarized in Table 5.2. The findings indicate that the estimation result obtained using the "direct" and the other methods approximately have the same order of efficiency when α is closer to 2. However, for small α, the differences are significant. Clearly the order of efficiency depends on the choice of the methods used for estimating the parameters of the models.

Table 5.1.1 Experiment I:$c_0 = 1$, $c_1 = 2$ and $c_2 = 1.111$

N	α	Method	s	$I(1)$	Avg. $\hat{I}(1)$	MAD_1	$I(2)$	Avg. $\hat{I}(2)$	MAD_2
100	2	Avg.	$\{0.01\}$		0.66180	0.04600		0.14647	0.10186
		Exp.	$\{0.01, 0.1, 1\}$	0.67722	0.63195	0.05732	0.17821	0.19387	0.09025
		ACF	-		0.65848	0.04644		0.14500	0.10115
	1.8	Avg.	$\{0.01, 0.2\}$	0.64700	0.64938	0.04415	0.19237	0.15860	0.09297
		Exp.	$\{0.01, 0.1, 1\}$		0.62509	0.04760		0.20409	0.08011
	1.5	Avg.	$\{0.01, 0.1, 0.2\}$	0.59903	0.60826	0.04839	0.21343	0.17158	0.07927
		Exp.	$\{0.01, 0.1, 1\}$		0.59728	0.04888		0.21062	0.06729
	1.3	Avg.	$\{0.01, 0.06, \ldots, 0.21\}$	0.56554	0.57235	0.05032	0.22665	0.19017	0.06974
		Exp.	$\{0.01, 0.1, 1\}$			0.05548		0.22380	0.06105
	1	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.51350	0.50708	0.04745	0.24325	0.21576	0.06049
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.50717	0.04739		0.21577	0.06036
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.47792	0.47274	0.04357	0.25014	0.22644	0.05599
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.47276	0.04355		0.22643	0.05597
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.42379	0.40713	0.04463	0.24809	0.22776	0.06033
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.40715	0.04459		0.22776	0.06028
1000	2	Avg.	$\{0.01\}$		0.67577	0.01335		0.17495	0.03096
		Exp.	$\{0.01\}$	0.67722	0.67577	0.01335	0.17821	0.17495	0.03096
		ACF	-		0.67544	0.01336		0.17477	0.03094
	1.8	Avg.	$\{0.01, 0.06, \ldots, 0.21\}$	0.64700	0.65059	0.01645	0.19237	0.18859	0.02708
		Exp.	$\{0.01, 0.06, \ldots, 0.21\}$		0.65067	0.01649		0.18854	0.02708
	1.5	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.59903	0.60204	0.01815	0.21343	0.20879	0.02155
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.60209	0.01822		0.20878	0.02154
	1.3	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.56554	0.56751	0.01640	0.22665	0.22312	0.01996
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.56754	0.01644		0.22310	0.01992
	1	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.51350	0.51396	0.01521	0.24325	0.24105	0.01577
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.51396	0.01521		0.24105	0.01576
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.47792	0.47742	0.01383	0.25014	0.24811	0.01511
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.47742	0.01382		0.24811	0.01510
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.42379	0.42239	0.01296	0.24809	0.24548	0.01852
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.42239	0.01295		0.24548	0.01851

Table 5.1.2 Experiment II:$c_0 = 1$, $c_1 = 1$ and $c_2 = 0.5$

N	α	Method	s	$I(1)$	Avg. $\hat{I}(1)$	MAD_1	$I(2)$	Avg. $\hat{I}(2)$	MAD_2
100	2	Avg.	$\{0.01\}$		0.64691	0.04695		0.18651	0.09944
		Exp.	$\{0.01, 0.1, 1\}$	0.66667	0.64133	0.04873	0.22222	0.21072	0.09010
		ACF	-		0.64366	0.04758		0.18463	0.09894
	1.8	Avg.	$\{0.01, 0.2\}$	0.65583	0.64535	0.04312	0.21861	0.19124	0.08987
		Exp.	$\{0.01, 0.1, 1\}$		0.63840	0.04468		0.21643	0.08259
	1.5	Avg.	$\{0.01, 0.2\}$	0.63733	0.63720	0.03920	0.21244	0.19092	0.07168
		Exp.	$\{0.01, 0.1, 1\}$		0.62975	0.03935		0.21746	0.06720
	1.3	Avg.	$\{0.01, 0.1, 0.2\}$	0.62341	0.62514	0.03750	0.20780	0.18555	0.06558
		Exp.	$\{0.01, 0.1, 1\}$		0.61851	0.04002		0.21564	0.06279
	1	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.60000	0.59356	0.03491	0.20000	0.17702	0.05580
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.59360	0.03488		0.17704	0.05566
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.58267	0.57785	0.03498	0.19422	0.17540	0.05315
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.57787	0.03497		0.17540	0.05312
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.55410	0.54519	0.03263	0.18470	0.15929	0.06053
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.54519	0.03262		0.15929	0.06048
1000	2	Avg.	$\{0.1, 0.2\}$		0.66604	0.01361		0.22077	0.03038
		Exp.	$\{0.1, 0.2\}$	0.66667	0.66604	0.01361	0.22222	0.22078	0.03038
		ACF	-		0.66572	0.01361		0.22064	0.03042
	1.8	Avg.	$\{0.01, 0.1, 0.5\}$	0.65583	0.65722	0.01294	0.21861	0.21590	0.02585
		Exp.	$\{0.01, 0.1, 0.5\}$		0.65756	0.01307		0.21614	0.02572
	1.5	Avg.	$\{0.01, 0.06, \ldots, 0.21\}$	0.63733	0.64032	0.01386	0.21244	0.21069	0.02150
		Exp.	$\{0.01, 0.06, \ldots, 0.21\}$		0.64037	0.01391		0.21071	0.02147
	1.3	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.62341	0.62333	0.01201	0.20780	0.20601	0.01802
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.62335	0.01203		0.20601	0.01800
	1	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.60000	0.60056	0.01132	0.20000	0.19934	0.01535
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.60056	0.01132		0.19934	0.01534
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.58267	0.58252	0.00975	0.19422	0.19249	0.01592
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.58252	0.00975		0.19249	0.01591
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.55410	0.55377	0.00934	0.18470	0.18384	0.01769
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.55377	0.00933		0.18384	0.01767

Table 5.1.3 Experiment III: $c_0 = 1$, $c_1 = 0.55$ and $c_2 = 0.05$

N	α	Method	s	$I(1)$	Avg. $\hat{I}(1)$	MAD_1	$I(2)$	Avg. $\hat{I}(2)$	MAD_2
100	2	Avg.	$\{0.01\}$		0.42368	0.06336		0.01561	0.09345
		Exp.	$\{0.01\}$	0.44253	0.42368	0.06336	0.03831	0.01561	0.09345
		ACF			0.42155	0.06344		0.01545	0.09258
	1.8	Avg.	$\{0.01, 0.1, 0.5\}$	0.43169	0.42068	0.05635	0.03447	0.01666	0.08108
		Exp.	$\{0.01, 0.1, 1\}$		0.42031	0.05682		0.03082	0.08003
	1.5	Avg.	$\{0.01, 0.1, 0.5\}$	0.41278	0.41010	0.05133	0.03003	0.01142	0.07324
		Exp.	$\{0.01, 0.1, 1\}$		0.41229	0.05317		0.03426	0.06992
	1.3	Avg.	$\{0.01, 0.06, \ldots, 0.21\}$	0.39846	0.39796	0.04819	0.02867	0.01133	0.06354
		Exp.	$\{0.01, 0.1, 1\}$		0.40088	0.05031		0.03278	0.06038
	1	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.37500	0.36673	0.04107	0.03125	0.00991	0.05910
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.36678	0.04103		0.00994	0.05900
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.35902	0.35297	0.03780	0.03835	0.01411	0.05902
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.35298	0.03779		0.01411	0.05900
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.33811	0.32857	0.03734	0.06333	0.04037	0.06664
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.32858	0.03731		0.04037	0.06660
1000	2	Avg.	$\{0.1\}$		0.44077	0.01913		0.03536	0.02851
		Exp.	$\{0.1\}$	0.44253	0.44077	0.01913	0.03831	0.03536	0.02851
		ACF	-		0.44055	0.01913		0.03533	0.02849
	1.8	Avg.	$\{0.01, 0.1, 0.5\}$	0.43169	0.43345	0.01781	0.03447	0.03410	0.02470
		Exp.	$\{0.01, 0.1, 0.5\}$		0.43367	0.01800		0.03410	0.02471
	1.5	Avg.	$\{0.01, 0.06, \ldots, 0.21\}$	0.41278	0.41579	0.01720	0.03003	0.02911	0.02030
		Exp.	$\{0.01, 0.06, \ldots, 0.21\}$		0.41585	0.01724		0.02912	0.02028
	1.3	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.39846	0.39945	0.01594	0.02867	0.02699	0.01911
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.39946	0.01596		0.02700	0.01909
	1	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.37500	0.37510	0.01208	0.03125	0.03062	0.01726
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.37510	0.01206		0.03062	0. 01722
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.35902	0.35776	0.01188	0.03835	0.03579	0.01772
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.35776	0.01187		0.03579	0.01772
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.33811	0.33742	0.01128	0.06333	0.06230	0.01995
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.33742	0.01127		0.06230	0.01994

Table 5.1.4 Experiment IV: $c_0 = 1$, $c_1 = -0.4$ and $c_2 = 0.7$

N	α	Method	s	$I(1)$	Avg. $\hat{I}(1)$	MAD_1	$I(2)$	Avg. $\hat{I}(2)$	MAD_2
100	2	Avg.	$\{0.01\}$		-0.40620	0.09564		0.40349	0.07117
		Exp.	$\{0.01\}$	-0.41212	-0.40620	0.09564	0.42424	0.40349	0.07117
		ACF	-		-0.40415	0.09516		0.39943	0.07129
	1.8	Avg.	$\{0.01, 0.2\}$	-0.32268	-0.39110	0.10727	0.41077	0.40732	0.06488
		Exp.	$\{0.01, 0.1, 1\}$		-0.30549	0.08134		0.40839	0.06508
	1.5	Avg.	$\{0.01, 0.1, 1\}$	-0.19541	-0.13719	0.10087	0.38652	0.39666	0.06615
		Exp.	$\{0.01, 1\}$		-0.22875	0.07796		0.40537	0.06349
	1.3	Avg.	$\{0.01, 0.1, 1\}$	-0.11483	-0.09366	0.09876	0.36731	0.38476	0.06611
		Exp.	$\{0.1, 0.2, \ldots, 1\}$		-0.07723	0.10007		0.36050	0.06427
	1	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.00000	-0.06593	0.10727	0.33333	0.33612	0.05201
		Exp.	$\{0.1, 0.2, \ldots, 1\}$		0.04434	0.08885		0.32922	0.06823
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.07269	0.02966	0.08416	0.30689	0.30179	0.04934
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.02952	0.08411		0.30183	0.04924
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.17608	0.14997	0.06301	0.26101	0.24732	0.05198
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.14998	0.06297		0.24733	0.05195
1000	2	Avg.	$\{0.01\}$		-0.41207	0.02860		0.42400	0.02344
		Exp.	$\{0.01\}$	-0.41212	-0.41207	0.02860	0.42424	0.42400	0.02344
		ACF			-0.41187	0.02859		0.42358	0.02342
	1.8	Avg.	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$	-0.32268	-0.32889	0.04062	0.41077	0.40963	0.02131
		Exp.	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$		-0.32893	0.04054		0.40969	0.02130
	1.5	Avg.	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$	-0.19541	-0.20161	0.03743	0.38652	0.38506	0.01929
		Exp.	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$		-0.20160	0.03708		0.38514	0.01922
	1.3	Avg.	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$	-0.11483	-0.12088	0.03527	0.36731	0.36613	0.01859
		Exp.	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$		-0.12074	0.03462		0.36617	0.01842
	1	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.00000	-0.00656	0.02912	0.33333	0.33350	0.01588
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		-0.00658	0.02913		0.33351	0.01587
	0.8	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	0.07269	0.06921	0.02193	0.30689	0.30576	0.01499
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$		0.06920	0.02188		0.30576	0.01496
	0.5	Avg.	$\{0.01, 0.02, \ldots, 0.1\}$	0.17608	0.17218	0.01904	0.26101	0.26032	0.01669
		Exp.	$\{0.01, 0.02, \ldots, 0.1\}$		0.17218	0.01902		0.26032	0.01668

Table 5.1: The true values $I(k)$ and the estimates $\hat{I}(k)$, $k = 1, 2$, from the experiment I, II, III and IV. Here, $Avg.\hat{I}(i) = \frac{1}{T}\sum_{j=1}^{T} \mathrm{Re}\,\hat{I}(i)_j$ and $MAD_i = \frac{1}{T}\sum_{j=1}^{T} |\mathrm{Re}\,\hat{I}(i)_j - I(i)|$. $i = 1, 2$, where $\mathrm{Re}\,\hat{I}(i)_j$ denotes the estimates at lag i in run j.

Table 5.2.1 $(r_1, r_2) = (2, 5)$

		Lag					Lag			
α	Method	1	2	3	4	α	1	2	3	4
2	True Value	0.636364	0.345455	0.178182	0.090182	1.5	0.627956	0.326052	0.159521	0.076885
	Direct	0.633064	0.340991	0.174688	0.088034		0.627273	0.326535	0.160179	0.075819
		(0.023159)	(0.036088)	(0.042819)	(0.045835)		(0.028948)	(0.032112)	(0.037686)	(0.037622)
	LAD	0.634643	0.341823	0.174841	0.088739		0.628016	0.32678	0.160402	0.077703
		(0.02917)	(0.047376)	(0.049192)	(0.04004)		(0.012123)	(0.018715)	(0.01809)	(0.013881)
	ACF	0.632878	0.339119	0.172188	0.086296		0.627453	0.326115	0.160148	0.078028
		(0.023372)	(0.036808)	(0.038036)	(0.030742)		(0.022968)	(0.036906)	(0.034713)	(0.025964)
	Acov	0.633642	0.341632	0.174671	0.088436		0.622925	0.331062	0.164549	0.088052
		(0.028739)	(0.044088)	(0.047058)	(0.039068)		(0.084035)	(0.114694)	(0.131799)	(0.147884)
	LS	0.634886	0.342267	0.175165	0.088565		0.627979	0.326619	0.160491	0.078234
		(0.023104)	(0.036403)	(0.037789)	(0.030739)		(0.022655)	(0.036861)	(0.034729)	(0.025998)
1	True Value	0.6	0.32	0.164	0.0828	0.8	0.581821	0.325878	0.179734	0.098848
	Direct	0.583056	0.301945	0.15055	0.072238		0.539047	0.286876	0.152552	0.081048
		(0.028827)	(0.028664)	(0.033636)	(0.033552)		(0.047267)	(0.043538)	(0.037518)	(0.033493)
	LAD	0.599262	0.320318	0.164786	0.083639		0.580963	0.326673	0.181262	0.100512
		(0.006503)	(0.003539)	(0.005316)	(0.004993)		(0.007152)	(0.003882)	(0.008368)	(0.008752)
	ACF	0.598132	0.319343	0.164344	0.0838		0.580139	0.326082	0.181208	0.101083
		(0.018823)	(0.026193)	(0.025184)	(0.019365)		(0.022024)	(0.029021)	(0.029181)	(0.024535)
	LS	0.598409	0.319634	0.164579	0.083966		0.580397	0.326403	0.181502	0.101314
		(0.018614)	(0.026097)	(0.025118)	(0.019323)		(0.021878)	(0.028892)	(0.029088)	(0.024477)

Table 5.2.2 $(r_1, r_2) = \frac{1}{3}(1 \pm i\sqrt{3})$

		Lag					Lag			
α	Method	1	2	3	4	α	1	2	3	4
2	True Value	0.48	-0.2025	-0.42188	-0.2025	1.5	0.46211	-0.10459	-0.21073	-0.10076
	Direct	0.478959	-0.20191	-0.42173	-0.20423		0.465392	-0.13468	-0.2721	-0.13153
		(0.014609)	(0.028524)	(0.03338)	(0.037368)		(0.023329)	(0.072476)	(0.141504)	(0.071686)
	LAD	0.478643	-0.19238	-0.40061	-0.1917		0.4635	-0.10488	-0.21572	-0.10203
		(0.018696)	(0.038095)	(0.049369)	(0.032947)		(0.008206)	(0.018354)	(0.03527)	(0.01751)
	ACF	0.478025	-0.1929	-0.39998	-0.1908		0.464044	-0.10205	-0.2146	-0.09979
		(0.014821)	(0.031146)	(0.044035)	(0.027903)		(0.013339)	(0.022223)	(0.035282)	(0.019445)
	Acov	0.478188	-0.19151	-0.39868	-0.19078		0.467123	-0.09135	-0.21576	-0.09291
		(0.018864)	(0.037538)	(0.05252)	(0.034997)		(0.038756)	(0.046058)	(0.048479)	(0.042662)
	LS	0.478904	-0.19238	-0.40088	-0.19229		0.464223	-0.10216	-0.21494	-0.1
		(0.014742)	(0.03139	(0.043642)	(0.027378)		(0.013387)	(0.022243)	(0.035401)	(0.019505)
1	True Value	0.428571	0	0	0	0.8	0.408082	0.044719	0.088067	0.042438
	Direct	0.406268	-0.00759	-0.01066	-0.00551		0.358947	0.030405	0.063201	0.028815
		(0.032393)	(0.03944)	(0.06732)	(0.036994)		(0.053965)	(0.031479)	(0.048142)	(0.030758)
	LAD	0.429389	0.000658	-0.00155	0.000454		0.409662	0.043983	0.084085	0.041647
		(0.006875)	(0.016476)	(0.032407)	(0.015618)		(0.008139)	(0.015958)	(0.031378)	(0.015179)
	ACF	0.433901	0.010279	-0.0014	0.00806		0.417289	0.057225	0.084156	0.052389
		(0.015944)	(0.02523)	(0.032321)	(0.021746)		(0.018461)	(0.027143)	(0.031495)	(0.023165)
	LS	0.434339	0.010017	-0.00152	0.007851		0.417371	0.057034	0.084186	0.052307
		(0.014675)	(0.025108)	(0.03257)	(0.021786)		(0.018528)	(0.027048)	(0.031485)	(0.023169)

Table 5.2: Comparison between the "direct" estimation method and the estimation via estimated parameters using eq. (4.4) for AR(2) process with the roots (r_1, r_2). We generated $T = 1000$ independent samples with the size of each sample $n = 1000$. The direct estimation method here is denoted by "Direct" in the column "Method", and it is evaluated at $\mathbf{s} = \{0.01, 0.02, \ldots, 0.05\}$. The other methods considered here are named via the method used for estimating the coefficients of AR(2) process, where $\hat{\alpha}$ is estimated using McCulloch quantile estimation method. Here, we record the average and the root of mean square error (in the brackets) of $\hat{I}(k)$ for lag $k = 1, \ldots, 4$.

5.2 Order identification and estimation of α-stable moving averages

5.2.1 Introduction

In Section 3.2.3, we presented a guide for graphical order identification methods of pure moving average processes of finite order using some dependence measures. In this part, we present the application of the normalized codifference estimator to attempt to identify the order of pure moving average process (of finite order). If the order of the process is already known, we also discuss how to estimate the coefficients of the model using the method of moments estimator.

Notice that for $MA(q)$ process, $X_t = \sum_{j=0}^{q} c_j \epsilon_{t-j}$, where ϵ_t fulfils the condition C2, by a similar method as for obtaining (4.4), we obtain for $k \leq q$,

$$\tau(k) = \sigma^\alpha |s|^\alpha \left[\sum_{j=0}^{q-k} (|(c_{j+k} - c_j)|^\alpha - |c_{j+k}|^\alpha - |-c_j|^\alpha) \right] \tag{5.2}$$

and $\tau(k)$ vanishes after lag $k > q$.

5.2.2 Fitting moving average process

Using the fact that $\tau(k)$ vanishes after lag $k > q$, we can use $\tau(\cdot)$ for a preliminary attempt of identifying the order of the underlying moving average process. By mimicking the traditional identification procedure using the SACF, we will use the results in Section 4.4 to identify the order of $MA(q)$ process.

Applying Theorem 4.4.2, we can obtain for $MA(q)$ process (similarly as the proof of Corollary 4.4.3), and $k > q$,

$$\sqrt{n}(\mathrm{Re}\, I(k)) \Rightarrow (W_1)^{1/2} Z, \quad \sqrt{n}(\mathrm{Im}\, I(k)) \Rightarrow (W_2)^{1/2} Z \tag{5.3}$$

where Z denotes standard normal random variable, and

$$W_1 = (\tau(0))^{-2} m_{11}^{kk}, \quad W_2 = (\tau(0))^{-2} m_{22}^{kk}$$

Here we have $(\tau(0))^2 = \left(\sum_{j=0}^{q} 2 |sc_j|^\alpha \right)^2$ and m_{11}^{kk}, m_{22}^{kk} can be obtained from (4.98) in a similar way as derivation of results in Section 4.7 (set $r = 1$). Unfortunately, from the form of $\tau(0)$, m_{11}^{kk} and m_{22}^{kk}, one can see that W_1 and W_2 depend on the unknown parameters $c_j, j = 1, \ldots, q$. Therefore, we propose to use the same strategy as when using the correlation function, an order identification method using the bounds of i.i.d. sequence. For this simple model, from Corollary 4.4.3 (consider $r = 1$), we obtain for $k > 0$, W_1 and W_2 are

$$W_1 = \{4\sigma^{2\alpha} |s|^{2\alpha}\}^{-1} \left[e^{2\sigma^\alpha |s|^\alpha} \left\{ \tfrac{1}{2} e^{2\sigma^\alpha |s|^\alpha} - 1 \right\} + e^{\sigma^\alpha (2|s|^\alpha - |2s|^\alpha)} \left\{ \tfrac{1}{2} e^{\sigma^\alpha (2|s|^\alpha - |2s|^\alpha)} - 1 \right\} + 1 \right] \tag{5.4}$$

and

$$W_2 = \{4\sigma^{2\alpha} |s|^{2\alpha}\}^{-1} \left[e^{2\sigma^\alpha |s|^\alpha} \left\{ \tfrac{1}{2} e^{2\sigma^\alpha |s|^\alpha} - 1 \right\} + e^{\sigma^\alpha (2|s|^\alpha - |2s|^\alpha)} \left\{ 1 - \tfrac{1}{2} e^{\sigma^\alpha (2|s|^\alpha - |2s|^\alpha)} \right\} \right] \tag{5.5}$$

which are independent with c_j's. Like the SACF, we plot the real and imaginary parts of $\hat{I}(k)$ together with the bounds $\pm 1.96\sqrt{W_1}/\sqrt{n}$, and $\pm 1.96\sqrt{W_2}/\sqrt{n}$, respectively, where n denotes the sample size. We consider that the real and imaginary parts of $\hat{I}(k)$ are approximately zero if their values at certain lags k are inside the bounds. However, here we only use $\mathrm{Re}\, \hat{I}(k)$ for the order identification purpose.

If the order q of the process is already known, given the estimates of the codifference function $\hat{\tau}(k)$ for some lags $k \geq 0$, one can use (5.2) to find the estimates of the coefficients $\hat{c}_j$'s. For given order q, the closed form expressions for $\hat{c}_j$'s, $j = 1, \ldots, q$ in the form of $\hat{\tau}(k), k = 0, \ldots, q$ are obtained only for $\alpha = 2$. For the process with infinite second moments, the estimate of the coefficients can be found by finding the solution of a set of appropriate nonlinear equations (considering only the equations for lag $k \neq 0$). Note that from consistency of $\hat{\tau}(\cdot)$(Theorem 4.4.1), this method should also be consistent.

Remark 5.2.1. Note that in case of Gaussian (or generally for the process with finite second moments), the identification of stationary process using the ACF has a solid theoretical reason. i.e., the characterization theorem of the autocovariance function (see Theorem A.3.6). Unfortunately, for the covariation and the codifference function, or the ACF in case of infinite variance process (where the population ACF indeed does not exist), similar results are not available. Therefore, the methods of order identification presented in Section 3.2.3 and the method proposed in this section should not be interpreted in the same way as the result obtained in the classical case.

5.2.3 Simulation results

Identification of moving average models

In this part, we present simulation studies for the order identification of several $MA(2)$ processes that we consider in Section 5.1.2, which we call as the experiments I-IV. For $\alpha = 2$, the true values of the normalized codifference and the correlation at lag k , $(I(1), I(2)) = (\rho(1), \rho(2))$ in experiment I-IV are:

$$(I).(0.677, 0.178), \quad (II).0.667, 0.222), \quad (III).(0.443, 0.038), \quad (IV).(-0.412, 0.424)$$

In experiment I and II, the values of I(k) are close to 1 at lag 1 and not too close to 0 at lag 2 while for experiment III, at lag 1, it is close to 0.5 but almost 0 at lag 2. Hence, it can be expected that the probability of underidentified will be high in the experiment III. For the last experiment, it is negative at lag 1 but positive at lag 2, with the absolute values near 0.5.

To investigate the performance of the proposed method, we simulate 1000 time series in experiment I - IV for several different values of $\alpha \in (0, 2]$, each series is having sample size $n = 1000$. From the experience of simulation studies in Section 5.1.2, we obtain that the optimal choice for grids s lies on an interval around zero. Thus for the comparison's sake, we consider the following sets of $\{s_1, \ldots, s_r\}$: $\mathbf{s}_1 = \{0.001\}$, $\mathbf{s}_2 = \{0.01\}$, $\mathbf{s}_3 = \{0.1\}$, $\mathbf{s}_4 = \{0.2\}$, $\mathbf{s}_5 = \{0.3\}$, $\mathbf{s}_6 = \{0.5\}$, $\mathbf{s}_7 = \{1\}$, $\mathbf{s}_8 = \{0.001, 0.01\}$, $\mathbf{s}_9 = \{0.001, 0.1\}$, $\mathbf{s}_{10} = \{0.001, 0.2\}$, $\mathbf{s}_{11} = \{0.001, 0.5\}$, $\mathbf{s}_{12} = \{0.001, 1\}$, $\mathbf{s}_{13} = \{0.01, 0.1\}$, $\mathbf{s}_{14} = \{0.01, 0.2\}$, $\mathbf{s}_{15} = \{0.01, 0.3\}$, $\mathbf{s}_{16} = \{0.01, 0.5\}$, $\mathbf{s}_{17} = \{0.01, 1\}$, $\mathbf{s}_{18} = \{0.01, 0.1, 0.2\}$, $\mathbf{s}_{19} = \{0.01, 0.1, 0.5\}$, $\mathbf{s}_{20} = \{0.01, 0.1, 1\}$, $\mathbf{s}_{21} = \{0.1, 0.2, 0.5\}$, $\mathbf{s}_{22} = \{0.1, 0.2, 1\}$, $\mathbf{s}_{23} = \{0.01, 0.06, \ldots, 0.21\}$, $\mathbf{s}_{24} = \{0.01, 0.02, \ldots, 0.1\}$, $\mathbf{s}_{25} = \{0.1, 0.11, \ldots, 0.2\}$, $\mathbf{s}_{26} = \{0.01, 0.02, \ldots, 0.2\}$. When $r > 1$, the estimator $\hat{I}(\cdot)$ and the bound (5.3) will be calculated as their weighted average of the estimates in the considered grid points. We consider two weighting methods, similarly as the calculation in Section 5.1.2, namely the simple average and the negative exponential weighted average. Here we calculate the sample normalized codifference $\hat{I}(k)$ and the SACF (the centralized ACF when $\alpha = 2$, and the non-centralized ACF when $\alpha < 2$), and consider the identification strategy as presented in Section 3.2.3, and Section 5.2.2 above. We count the number of times the order of the model is successfully identified, i.e., the case when statistic $\mathrm{Re}\,\hat{I}(\cdot)$ is outside the bounds at lag 2, and within the bounds for each lag $3, \ldots, 10$. The results are presented in Table 5.3 - 5.4.

Notice that throughout the simulation studies in this part, we consider only standard $S\alpha S$ for ϵ_t, where we assume that the value of α is known. For practical application, we suggest to use McCulloch's estimator for obtaining $\hat{\alpha}$ (see Section 2.4). For small order MA and AR processes, Adler *et al.* (1998*a*) give simulation evidences that McCulloch's method performs well.

From the simulation results, we obtain that when $1 < \alpha < 2$, the performance of the identification procedure, which uses the sample normalized codifference $\hat{I}(\cdot)$, in most cases will be similar to the performance of the procedure, which uses the SACF and normal bounds (i.e., as if we are unaware that the data are coming from the stable process). When α is getting closer to zero, the performance of $\hat{I}(\cdot)$ will be the worst compare to the other considered methods. When $\alpha = 2$, the performance of the procedure which uses $\hat{I}(\cdot)$ is very much similar to the performance of the standard identification procedure using the SACF. If the sample size is increased, in any case we find that the performance of $\hat{I}(\cdot)$ is improved. Furthermore, it is quite interesting to see the result of experiment III, where with a high probability the correct order of the model is underidentified, the performance of the other methods is outperformed by identification procedure which uses $\hat{I}(\cdot)$ in many choices of α.

In the table, for identification method using $\hat{I}(\cdot)$, we record the best choices of s for each α, which are defined as the sets of grid points which maximize the number of successful identification (i.e., the true order is correctly identified) among all considered choices of sets of grid points above. Throughout the simulation studies, the results indicate that the number of successful identification depends on the choice of grid points $\mathbf{s}$, where the optimal choice of grids depends on the index α and the sample size n. However, it seems that there is a benefit by evaluating $\hat{I}(\cdot)$ at several points of s, that is $r > 1$, where under the appropriate choices of $\mathbf{s}$, the performance of the weighting methods (the simple average and the exponential weighted average) are approximately the same. We can see that throughout the simulation studies, the performance of the identification procedure which uses $\hat{I}(\cdot)$ will be improved by including a point very close to 0. However, it seems that the choice $\mathbf{s} = \{0.01\}$ is adequate for the order identification purpose, since the performance of the identification procedure will be tapered off at this point.

Furthermore, the findings in Table 5.3 - 5.4 suggest that the boundary of independent sequence (MA(0)) is possibly less appropriate for the identification of $MA(q)$ model, $q > 0$, using any of graphical method. In identification using the SACF, instead of using stringent boundaries $\pm 1.96/\sqrt{n}$, using the estimate of the right hand side of (3.20), we often use bounds $\pm 1.96(1 + 2\sum_{j=0}^{q} \hat{\rho}(j)^2)^{1/2}/\sqrt{n}$ to identify the order of $MA(q)$ process (by first fixing the order q). However, for the sample normalized codifference, it is not clear how to attack the problem using a similar approach, as the limiting distribution of the sample normalized codifference depends on the unknown parameters $c_j, j = 1, \ldots, q$.

Estimation of moving averages models

In this part, we present a simulation study to illustrate the usage of the sample codifference function for estimating the coefficients of moving average models with finite order. Here we only consider MA(2) process as in the experiment I - IV. Note that the codifference of MA(2) process is

$$\tau(0) = -2\sigma^\alpha |s|^\alpha (1 + |c_1|^\alpha + |c_2|^\alpha) \tag{5.6}$$

$$\tau(1) = \sigma^\alpha |s|^\alpha (|(c_1 - 1)|^\alpha + |(c_2 - c_1)|^\alpha - 2|c_1|^\alpha - |c_2|^\alpha - 1) \tag{5.7}$$

$$\tau(2) = \sigma^\alpha |s|^\alpha (|(c_2 - 1)|^\alpha - |c_2|^\alpha - 1) \tag{5.8}$$

and equal to 0 for the other lags.

We generate 1000 independent series using function `arima.sim` in the package `stats` of `R` version 1.9.0 and function `rstable` in the extension package `stable`, to generate the unit symmetric α stable innovations. In the simulation, we consider two sample sizes, the small one is $n = 100$, and the large one is $n = 1000$. We use the following sets of $\{s_1, \ldots, s_r\}$: $\mathbf{s}_1 = \{0.0001\}$, $\mathbf{s}_2 = \{0.001\}$, $\mathbf{s}_3 = \{0.01\}$, $\mathbf{s}_4 = \{0.1\}$, $\mathbf{s}_5 = \{0.2\}$, $\mathbf{s}_6 = \{0.0001, 0.1\}$, $\mathbf{s}_7 = \{0.0001, 0.2\}$, $\mathbf{s}_8 = \{0.0001, 0.3\}$, $\mathbf{s}_9 = \{0.01, 0.1\}$, $\mathbf{s}_{10} = \{0.01, 0.2\}$, $\mathbf{s}_{11} = \{0.01, 0.3\}$, $\mathbf{s}_{12} = \{0.01, 0.1, 0.2\}$, $\mathbf{s}_{13} = \{0.01, 0.02, \ldots, 0.1\}$, $\mathbf{s}_{14} = \{0.1, 0.11, \ldots, 0.2\}$, $\mathbf{s}_{15} = \{0.01, 0.02, \ldots, 0.2\}$. When $r > 1$, the estimator $\hat{I}(\cdot)$ will be calculated as the weighted average of the estimates in the considered grid points. We consider two

weighting methods as in Section 5.1.2, namely the simple average (which we call as method Avg.) and the negative exponential weighted average (which we call as method Exp.).

For calculating the estimators $\hat{c}_1$ and $\hat{c}_2$, we plug-in the values of $\hat{\tau}(1)$ and $\hat{\tau}(2)$ into the equation (5.7) and (5.8). Notice that (5.7) and (5.8) have a closed-form solution in terms of c_1 and c_2 only when $\alpha = 2$. When $\alpha < 2$, these equations are nonlinear in c_1 and c_2 and only numerical solution can be obtained, i.e., by finding the zeros of (5.7) and (5.8). These solutions can be searched for instance, using function `uniroot` in R. Unfortunately, here we have numerical difficulties when working with the function `uniroot`, especially when $\alpha \leq 1$. For this reason, we only consider two values of α in simulation, which are equal to 1.5 and 2. Note that in the calculation of estimate $\hat{c}_1$ and $\hat{c}_2$ using equation (5.7) and (5.8), we assume that the values of α are known. The simulation results in this part are summarized in Table 5.5. In the table, when $\alpha = 2$, we compare the performance of the estimation procedures using the sample codifference function and using the sample covariance function $\hat{\gamma}(\cdot)$. When we use the sample codifference function, we record the sets of grid points s, which minimize the sum of mean absolute deviation (MAD) of the estimators $\hat{c}_1$ and $\hat{c}_2$. In the table we denote this minimization method as method Avg. (when $\hat{I}(\cdot)$'s are obtained using the simple average of the estimates on the considered points) and method Exp. (when the negative exponential weighted average is used for obtaining $\hat{I}(\cdot)$'s).

Throughout of simulation studies, the findings show that the performance of the estimation procedure which uses the sample codifference function is fairly well. As it can be expected from consistency of the codifference estimator, the performance of the estimators $\hat{c}_1$ and $\hat{c}_2$ improves when the sample size is increased. The optimum performance of weighting method (the average and the negative exponential weight) seems to be approximately the same here. The choice of optimal grids point also agrees to the procedure that we propose in Section 5.1.1. Furthermore from the table, we found that when $\alpha = 2$, the performance of the estimation procedure which uses the sample codifference function is approximately the same as the similar estimation procedure, which is based on the sample covariance $\hat{\gamma}(\cdot)$.

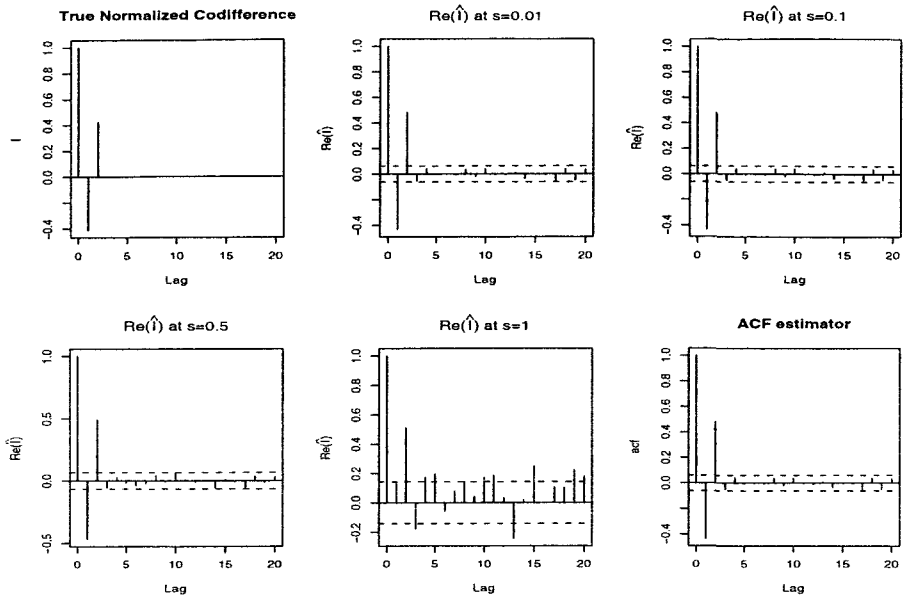


Figure 5.4: Plot of $I(k)$ and the real part of its estimator, $\mathrm{Re}\,\hat{I}(k)$, at lag $0, \ldots, 20$ for one simulation run of experiment IV with $c_0 = 1$, $c_1 = -0.4$, $c_2 = 0.7$ at some choices of s, where $\alpha = 2$, and the sample size $n = 1000$. We also plot the SACF to graphically compare the performance of $\mathrm{Re}\,\hat{I}(k)$.

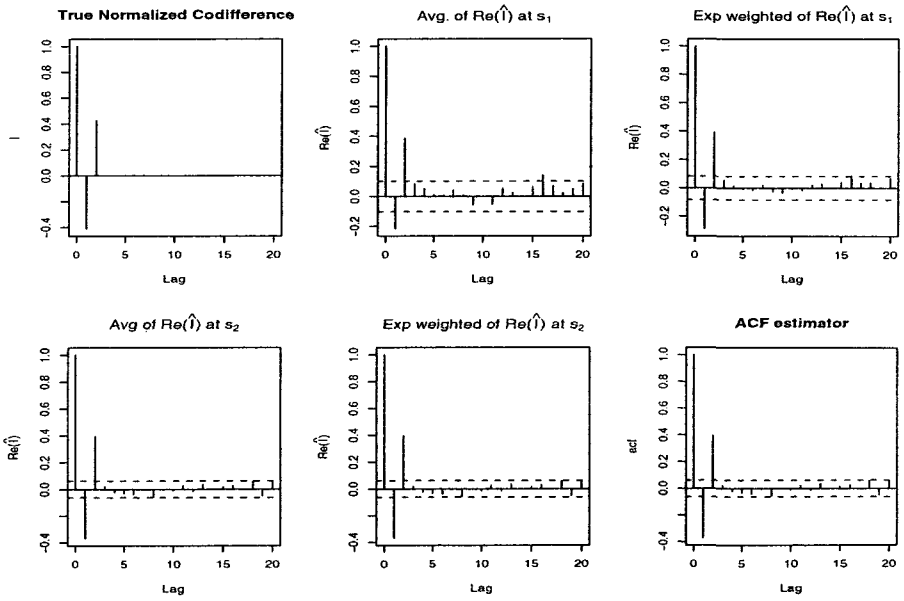


Figure 5.5: Similar to Figure 5.4, but different sets of grid points. Here $s_1 = \{0.01, 1\}$ and $s_2 = \{0.01, 0.02, \ldots, 0.2\}$. The estimates and the bounds are calculated either as the average or as the negative exponential weighted average of the estimates and the bounds on the considered grids.

Table 5.3.1

α	method	s	$\hat{I}(k)$	$\hat{\rho}(k)$	$\hat{\rho}(k)_\alpha$
0.3	Avg.	s_8	318	970	881
	Exp.	s_8	318	970	881
0.5	Avg.	s_8	652	932	893
	Exp.	s_8	652	932	893
0.7	Avg.	s_8	828	915	917
	Exp.	s_8	828	915	917
0.9	Avg.	s_8	850	851	901
	Exp.	s_8	850	851	901
1.1	Avg.	s_8	748	796	885
	Exp.	s_8	748	796	885
1.3	Avg.	s_{13}	654	721	840
	Exp.	s_{13}	654	721	840
1.5	Avg.	s_{13}	575	629	764
	Exp.	s_{13}	575	629	764
1.7	Avg.	s_{13}	510	555	516
	Exp.	s_{13}	508	555	516
1.9	Avg.	s_9	432	453	86
	Exp.	s_9	433	453	86

Table 5.3.2

α	method	s	$\hat{I}(k)$	ACF	ACF_α
0.3	Avg.	s_8	416	966	884
	Exp.	s_8	416	966	884
0.5	Avg.	s_8	677	942	900
	Exp.	s_8	677	942	900
0.7	Avg.	s_8	832	888	889
	Exp.	s_8	832	888	889
0.9	Avg.	s_8	809	830	879
	Exp.	s_9	809	830	879
1.1	Avg.	s_9	757	807	895
	Exp.	s_9	758	807	895
1.3	Avg.	s_{13}	697	716	830
	Exp.	s_{13}	698	716	830
1.5	Avg.	s_{14}	603	632	760
	Exp.	s_{14}	602	632	760
1.7	Avg.	s_{14}	559	581	540
	Exp.	s_{19}	561	581	540
1.9	Avg.	s_{10}, s_{14}	447	454	111
	Exp.	s_{14}	446	454	111

Table 5.3.3

α	method	s	$\hat{I}(k)$	ACF	ACF_α
0.3	Avg.	s_8	531	5	894
	Exp.	s_8	531	5	894
0.5	Avg.	s_8	548	8	860
	Exp.	s_8	548	8	860
0.7	Avg.	s_8	303	16	16
	Exp.	s_8	303	16	16
0.9	Avg.	s_{13}	204	36	14
	Exp.	s_{13}	204	36	14
1.1	Avg.	s_{24}	182	49	8
	Exp.	s_{24}	183	49	8
1.3	Avg.	s_1	177	82	29
	Exp.	s_1	177	82	29
1.5	Avg.	s_9	166	117	70
	Exp.	s_9	167	117	70
1.7	Avg.	s_{15}	151	146	146
	Exp.	s_{15}	154	146	146
1.9	Avg.	s_{10}, s_{14}	134	136	73
	Exp.	s_2,s_8, s_{10}	133	136	73

Table 5.3.4.

α	method	s	$\hat{I}(k)$	ACF	ACF_α
0.3	Avg.	s_8	445	972	904
	Exp.	s_8	445	972	904
0.5	Avg.	s_8	744	945	903
	Exp.	s_8	744	945	903
0.7	Avg.	s_8	828	905	906
	Exp.	s_8	828	905	906
0.9	Avg.	s_8	834	857	895
	Exp.	s_8	836	857	895
1.1	Avg.	s_8	799	814	891
	Exp.	s_8	799	814	891
1.3	Avg.	s_{14}	728	756	865
	Exp.	s_{14}	726	756	865
1.5	Avg.	s_{19}	662	690	800
	Exp.	s_{19}	676	690	800
1.7	Avg.	s_{19}	567	583	530
	Exp.	s_{19}	569	583	530
1.9	Avg.	s_{15}	514	538	101
	Exp.	s_{15}	513	538	101

Table 5.3: Number of simulated data sets out of 1000 for experiments I-IV that were correctly identified as coming from MA(2) processes, for some $\alpha < 2$ and sample size $n = 1000$. The parameters of models are : $c_1 = 2$, $c_2 = 1.111$ (Table 5.3.1), $c_1 = 1$, $c_2 = 0.5$ (Table 5.3.2), $c_1 = 0.55$, $c_2 = 0.05$ (Table 5.3.3) and $c_1 = -0.4$, $c_2 = 0.7$ (Table 5.3.4), respectively. Here, we only listed the grids s, which give the best performance, where $s_1 = \{0.001\}$, $s_2 = \{0.01\}$,$s_8 = \{0.001, 0.01\}$, $s_9 = \{0.001, 0.1\}$, $s_{10} = \{0.001. 0.2\}$, $s_{13} = \{0.01, 0.1\}$, $s_{14} = \{0.01, 0.2\}$, $s_{15} = \{0.01, 0.3\}$, $s_{19} = \{0.01, 0.1, 0.5\}$, $s_{24} = \{0.01, 0.02, \ldots, 0.1\}$.

Table 5.4.1

N	method	s	$\hat{I}(k)$	$\hat{\rho}(k)$
100	Avg.	s_{10},s_{14},s_{23}	164	175
	Exp.	s_{20}	233	175
500	Avg.	s_{18},s_{23}	438	432
	Exp.	s_{18},s_{23}	438	432
1000	Avg.	s_{25}	429	427
	Exp.	s_{25}	429	427
5000	Avg.	s_{23}	430	424
	Exp.	s_{18},s_{23}	429	424

Table 5.4.2

N	method	s	$\hat{I}(k)$	$\hat{\rho}(k)$
100	Avg.	s_1,s_2,s_4,s_8	223	231
	Exp.	s_{20}	245	231
500	Avg.	s_8,s_{13},s_{24}	436	439
	Exp.	s_8,s_{13},s_{24}	436	439
1000	Avg.	s_{11},s_{16}	413	410
	Exp.	s_{11},s_{16}	413	410
5000	Avg.	s_1,s_2,s_8	456	457
	Exp.	s_1,s_2,s_8	456	457

Table 5.4.3

N	method	s	$\hat{I}(k)$	$\hat{\rho}(k)$
100	Avg.	s_{25}	37	37
	Exp.	s_{15},s_{25}	37	37
500	Avg.	s_4	89	85
	Exp.	s_4	89	85
1000	Avg.	s_{25}	132	131
	Exp.	s_{25}	132	131
5000	Avg.	s_{20}	400	396
	Exp.	s_{20}	408	396

Table 5.4.4

N	method	s	$\hat{I}(k)$	$\hat{\rho}(k)$
100	Avg.	s_1,s_2,s_8	504	541
	Exp.	s_{20}	529	541
500	Avg.	s_{10},s_{14},s_{18},s_{23}	498	502
	Exp.	s_{20}	508	502
1000	Avg.	s_{25}	507	506
	Exp.	s_{25}	507	506
5000	Avg.	s_{20}	468	465
	Exp.	s_9,s_{17}	486	465

Table 5.4: Number of simulated data sets out of 1000 for experiment I-IV that were correctly identified as coming from MA(2) processes, for $\alpha = 2$ and some sample size n. The parameters of models are : $c_1 = 2$, $c_2 = 1.111$ (Table 5.4.1), $c_1 = 1$, $c_2 = 0.5$ (Table 5.4.2), $c_1 = 0.55$, $c_2 = 0.05$ (Table 5.4.3) and $c_1 = -0.4$, $c_2 = 0.7$ (Table 5.4.4), respectively. Here, we only listed the grid s, which gives the best performance, where $s_1 = \{0.001\}$, $s_2 = \{0.01\}$,$s_4 = \{0.2\}$, $s_8 = \{0.001, 0.01\}$, $s_9 = \{0.001, 0.1\}$, $s_{10} = \{0.001, 0.2\}$, $s_{11} = \{0.001, 0.5\}$, $s_{13} = \{0.01, 0.1\}$, $s_{14} = \{0.01, 0.2\}$, $s_{15} = \{0.01, 0.3\}$, $s_{16} = \{0.01, 0.5\}$, $s_{17} = \{0.01, 1\}$, $s_{18} = \{0.01, 0.1, 0.2\}$, $s_{20} = \{0.01, 0.1, 1\}$, $s_{23} = \{0.01, 0.06, \ldots, 0.21\}$, $s_{24} = \{0.01, 0.02, \ldots, 0.1\}$, and $s_{25} = \{0.1, 0.11, \ldots, 0.2\}$

Table 5.5.1: $c_1 = 2$, $c_2 = 1.111$

α	n	Method	s	med $\hat{c}_1$	MAD_1	med $\hat{c}_2$	MAD_2
2	100	Avg.	$\{1e-04, 0.1\}$	2.106618	0.979001	0.819056	0.675340
		Exp.	$\{1e-04, 0.1\}$	2.107266	0.979825	0.819090	0.675329
		$\hat{\gamma}(\cdot)$	—	2.103775	1.051022	0.804442	0.670264
	1000	Avg.	$\{0.0001\}$	2.005643	0.145549	1.104687	0.220927
		Exp.	$\{0.0001\}$	2.005643	0.145549	1.104687	0.220927
		$\hat{\gamma}(\cdot)$	—	2.005664	0.145463	1.103581	0.220717
1.5	100	Avg.	$\{0.1\}$	1.856800	1.353239	0.864860	1.100492
		Exp.	$\{0.1\}$	1.856800	1.353239	0.864860	1.100492
	1000	Avg.	$\{0.01, 0.02, \ldots, 0.2\}$	1.989153	0.280088	1.166471	0.369550
		Exp.	$\{0.01, 0.02, \ldots, 0.2\}$	1.990065	0.279532	1.166951	0.371309

Table 5.5.2 : $c_1 = 1$, $c_2 = 0.5$

α	n	Method	s	med $\hat{c}_1$	MAD_1	med $\hat{c}_2$	MAD_2
2	100	Avg.	$\{0.0001\}$	0.987103	0.138540	0.413601	0.254906
		Exp.	$\{0.0001\}$	0.987103	0.138540	0.413601	0.254906
		$\hat{\gamma}(\cdot)$	—	0.985827	0.138118	0.409585	0.253441
	1000	Avg.	$\{0.0001\}$	1.000157	0.037844	0.494991	0.084040
		Exp.	$\{0.0001\}$	1.000157	0.037844	0.494991	0.084040
		$\hat{\gamma}(\cdot)$	—	1.000040	0.037835	0.494495	0.083988
1.5	100	Avg.	$\{0.2\}$	0.982443	0.353930	0.427296	0.271203
		Exp.	$\{0.2\}$	0.982443	0.353930	0.427296	0.271203
	1000	Avg.	$\{0.2\}$	0.994395	0.090274	0.487849	0.070800
		Exp.	$\{0.2\}$	0.994395	0.090274	0.487849	0.070800

Table 5.5.1: $c_1 = 0.55$, $c_2 = 0.05$

α	n	Method	s	med $\hat{c}_1$	MAD_1	med $\hat{c}_2$	MAD_2
2	100	Avg.	$\{0.0001\}$	0.525610	0.105586	0.018763	0.121484
		Exp.	$\{0.0001\}$	0.525610	0.105586	0.018763	0.121484
		$\hat{\gamma}(\cdot)$	—	0.522891	0.105573	0.018502	0.120367
	1000	Avg.	$\{0.001\}$	0.548176	0.031573	0.045342	0.039548
		Exp.	$\{0.001\}$	0.548176	0.031573	0.045342	0.039548
		$\hat{\gamma}(\cdot)$	—	0.547908	0.031569	0.045296	0.039513
1.5	100	Avg.	$\{0.2\}$	0.518537	0.359696	0.004775	0.157516
		Exp.	$\{0.2\}$	0.518537	0.359696	0.004775	0.157516
	1000	Avg.	$\{0.2\}$	0.549811	0.045314	0.041396	0.041955
		Exp.	$\{0.2\}$	0.549811	0.045314	0.041396	0.041955

Table 5.5.2: $c_1 = -0.4$, $c_2 = 0.7$

α	n	Method	s	med $\hat{c}_1$	MAD_1	med $\hat{c}_2$	MAD_2
2	100	Avg.	$\{0.001\}$	-0.412268	0.105113	0.653730	0.183803
		Exp.	$\{0.001\}$	-0.412268	0.105113	0.653730	0.183803
		$\hat{\gamma}(\cdot)$	—	-0.411937	0.105076	0.647112	0.183078
	1000	Avg.	$\{0.001\}$	-0.401349	0.033203	0.696237	0.055984
		Exp.	$\{0.001\}$	-0.401349	0.033203	0.696237	0.055984
		$\hat{\gamma}(\cdot)$	—	-0.401311	0.033202	0.695540	0.055971
1.5	100	Avg.	$\{0.2\}$	-0.479764	0.319835	0.670749	0.249772
		Exp.	$\{0.2\}$	-0.479764	0.319835	0.670749	0.249772
	1000	Avg.	$\{0.2\}$	-0.422774	0.102640	0.690173	0.068098
		Exp.	$\{0.2\}$	-0.422774	0.102640	0.690173	0.068098

Table 5.5: Median and maximum absolute deviation of the estimates of the coefficients MA(2) processes in the experiments I-IV, for some α and sample sizes n.

5.3 Testing for independence in heavy-tailed time series

5.3.1 Introduction

Let $X_1, X_2, \ldots$ be a sequence of identically distributed random variables and consider the problem of testing the hypothesis,

$$H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0, m < n \tag{5.9}$$

against serial dependence alternatives,

$$H_1 : \rho(k) \neq 0 \text{ at least for one } k, \text{with } k = 1, \ldots, m$$

where $\rho(\cdot)$ denotes the autocorrelation function.

If X_i's are i.i.d. with $E(X_i^2) < \infty$, then this hypothesis can be tested with a Portmanteau statistic, such as Box-Pierce's Q-statistic (Box and Pierce, 1970)

$$Q_{n,m} = n \sum_{k=1}^{m} (\hat{\rho}(k))^2 \tag{5.10}$$

or its finite sample correction, Ljung-Box's Q-statistic (Ljung and Box, 1978)

$$\tilde{Q}_{n,m} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} (\hat{\rho}(k))^2 \tag{5.11}$$

where

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^{n} (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}, k = 1, \ldots, m \tag{5.12}$$

The statistic $Q_{n,m}$ and $\tilde{Q}_{n,m}$ are asymptotically χ^2_m distributed when $n \to \infty$.

Let's now assume that X_i's are i.i.d. SαS random variables with $1 < \alpha < 2$ (thus the variance is infinite but $E(X_i) < \infty$). Under this condition, for testing the hypothesis (5.9), Runde (1997) considers the statistic

$$Q^{\alpha}_{n,m} = (\frac{n}{\ln n})^{2/\alpha} \sum_{k=1}^{m} (\hat{\rho}(k))^2 \tag{5.13}$$

where the autocorrelation function $\rho(\cdot)$ in the hypothesis (5.9) is replaced by the pseudo-correlogram $\tilde{\rho}(k)$, see Theorem 3.2.2. It was shown that under H_0

$$Q^{\alpha}_{n,m} \xrightarrow{d} Q_m(\alpha) \text{ as } n \to \infty$$

where $Q_m(\alpha)$ is distributed as

$$\frac{1}{G_0^2}(G_1^2 + \cdots + G_m^2) \tag{5.14}$$

and $G_0, G_1, \ldots$ are as given in Theorem 3.2.2. The critical values of $Q_m(\alpha)$ and the empirical levels of the test for selected α and m can be found in Runde (1997), Table 1 and 2, respectively.

Gallagher (2000) lists several drawbacks of the SACF for heavy-tailed modelling. Since the Runde's statistic is based on the SACF, it has similar shortcomings, as follows:

1. Here, it is unclear what is exactly tested in the hypothesis proposed in Runde (1997), because $\tilde{\rho}(k)$ (when $\alpha < 2$) is not as easy interpreted as the population correlation function $\rho(k)$ (when $\alpha = 2$).

2. There is discontinuity in the quantiles of the limiting distribution and in the constant of statistics used for the test as $\alpha \to 2$. Note that when $\alpha < 2$, for $Q^{\alpha}_{n,m}$ we use constant $(n/\ln n)^{2/\alpha}$ but when $\alpha = 2$, for $Q_{n,m}$ we use constant n. Here, the quantile of $Q_m(\alpha)$ is discontinuous, as the divisor G_0 can be large with positive probability when $\alpha < 2$, but it becomes unity when $\alpha = 2$. As analogously noted in Gallagher (2000), these discontinuities cause problems in application, as for instance, if the true $\alpha = 2$, but is estimated a little bit smaller than the true value, e.g., $\alpha = 1.98$, then the constant and the quantiles of the distribution, which are used for the test will change drastically.

3. Due to the nonexistence of the limiting distribution $Q_m(\alpha)$ in a closed form, the critical values of the limiting distribution are relatively difficult to obtain, although it can be found using computer-intensive numerical approximations (see Runde (1997)). Moreover, simulation results in Runde (1997), Table 2, indicate that $Q_m(\alpha)$ provides a poor approximation distribution for small sample sizes. It was confirmed as from simulation studies (see Section 5.3.3), using the same parameters as in Table 2 in Runde (1997), we find that most of the empirical rejection probabilities are significantly different from the theoretical ones even when the sample size n is taken to be 1000 (it seems that the simulation results in Table 2 of Runde (1997) are incorrect). This flaw is probably due to the slow convergence of the sample correlation function to its asymptotic distribution in the infinite variance case (Adler *et al.*, 1998*a*), where it was reported in some cases, the sample size of one million is necessary to get the correct empirical level test 5%.

5.3.2 A Portmanteau-type test using the codifference function

Motivated by the above problems, let us consider the hypothesis of zero codifference up to order m, $m < n$, i.e.,

$$H_0 : I(1) = I(2) = \cdots = I(m) = 0 \tag{5.15}$$

against

$$H_1 : I(k) \neq 0 \text{ at least for one } k, k = 1, \ldots, m$$

Note that for Gaussian process, the hypothesis (5.15) is equivalent to the hypothesis (5.9).

To test the hypothesis (5.15), we propose to use the statistic

$$Q_{GC} = n \sum_{j=1}^{m} \left(\operatorname{Re} \tilde{I}(j) \right)^2 \tag{5.16}$$

where

$$\operatorname{Re} \tilde{I}(j) = \frac{\operatorname{Re} \hat{I}(j)}{\hat{S}} \tag{5.17}$$

Here $\operatorname{Re} \hat{I}(j)$ denotes the real part of the sample normalized codifference $\hat{I}(k)$, and $\hat{S}^2$ is given in the Theorem 5.3.1 below. The asymptotic property of Q_{GC} under the null hypothesis is summarized below.

Theorem 5.3.1. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. symmetric α-stable with the characteristic function $E \exp(isX_1) = \exp(-\sigma^{\alpha}|s|^{\alpha}), s \in \mathbb{R}$, where $\alpha \in (0, 2]$ and $\sigma > 0$, then as $n \to \infty$ ($m \in \mathbb{N}$)*

$$Q_{GC} = n \sum_{j=1}^{m} \left(\operatorname{Re} \tilde{I}(j) \right)^2 \xrightarrow{d} \chi^2_m$$

where $\operatorname{Re} \tilde{I}(j)$ is as given in (5.17) and χ^2_m denotes a chi-square random variable with m degrees of freedom and where $\hat{S}^2$ denotes a consistent estimate of the asymptotic variance of $\operatorname{Re} \hat{I}(k)$, $k > 0$, i.e.,

$$S^2 = (2\sigma^{\alpha}|s|^{\alpha})^{-2} \left[e^{2\sigma^{\alpha}|s|^{\alpha}} \left\{ \tfrac{1}{2} e^{2\sigma^{\alpha}|s|^{\alpha}} - 1 \right\} + e^{\sigma^{\alpha}(2|s|^{\alpha} - |2s|^{\alpha})} \left\{ \tfrac{1}{2} e^{\sigma^{\alpha}(2|s|^{\alpha} - |2s|^{\alpha})} - 1 \right\} + 1 \right] \tag{5.18}$$

Remark 5.3.1. Note that under the alternative hypothesis that the process is a stationary linear process $X_t = \sum_{j=0}^{\infty} c_j \epsilon_t$ where conditions C1 and C2 hold, a similar statistic to Q_{GC} can be shown to be asymptotically chi-square distributed using Theorem 4.4.2.

Theorem 5.3.1 is obtained directly via continuous mapping theorem (see Section A.3 in the Appendix) using the result of the following proposition.

Proposition 5.3.2. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $S\alpha S$ with the characteristic function $E\exp(isX_1) = \exp(-\sigma^\alpha |s|^\alpha), s \in \mathbb{R}$, where $\alpha \in (0,2]$ and $\sigma > 0$. Then for $h \in \mathbb{N}$*

$$\left[\operatorname{Re}\hat{I}(1), \operatorname{Re}\hat{I}(2), \ldots, \operatorname{Re}\hat{I}(h)\right]^T \text{ is } AN\left(\mathbf{0}, n^{-1}S^2 I_h\right)$$

where I_h denotes the identity matrix with size h and S^2 is as given in (5.18).

Proof. This result follows directly using the result in Section 4.7, by setting $r = 1$. □

5.3.3 Simulation results

The choice of s

Note that from (4.68) and (5.16), one can see that the sample normalized codifference function $\operatorname{Re}\hat{I}(\cdot)$ and therefore, statistic Q_{GC}, depend on s. For our purpose, the statistic Q_{GC} should be calculated from the values of s, which give the best empirical levels for the test. This can be achieved by choosing the grids s, which will give the most accurate estimates for $\operatorname{Re}\hat{I}(\cdot)$. For each lag $k = 1, \ldots, m$, here we consider a discrete estimation procedure for $\operatorname{Re}\hat{I}(k)$, i.e., we evaluate $\operatorname{Re}\hat{I}(k)$ at some points $s_1, \ldots, s_r$, $r \geq 1$. As the asymptotic variances in (5.18) has different values for different choices of grid points $s_i, i = 1, \ldots, r$, here we calculate $\operatorname{Re}\tilde{I}(s_i, k)$ for all considered points $s_i, i = 1, \ldots, r$. The final estimate of $\operatorname{Re}\tilde{I}(k)$ is calculated as the weighted value of the estimates among the choices of grid points, i.e., $\operatorname{Re}\tilde{I}(k) = \sum_{j=1}^{r} w_j \operatorname{Re}\tilde{I}(s_j, k)$, $\sum_{j=1}^{r} w_j = 1$. In this part, we consider the weights as the simple average of the estimates with $w_j = 1/r_i$ (we called method *avg.*) or the negative exponential weighted average, with $w_j = \exp(-s_j^2)/\sum_{i=1}^{r}\exp(-s_i^2)$ (we called method *exp.*). Using this approach, we do not take into account the covariance between $\operatorname{Re}\hat{I}(s_i; k)$ and $\operatorname{Re}\hat{I}(s_j; k)$, $i \neq j, i, j = 1, \ldots, r$, $k = 1, \ldots, m$ in the definition of Q_{GC}.

The empirical levels of Q_{GC}

Although the asymptotic distribution of the Portmanteau statistic that we consider in (5.16) is chi-square, the usage of quantiles of the chi-square distribution in the finite-sample case is still an open question. In this regard, we will investigate whether the finite-sample distribution of the statistic (5.16) is well approximated by the chi-square distribution, and furthermore how it compares to the performance of the other "standard" statistic.

The simulation study in this part consists of two parts, the case with the infinite variance and with the finite variance. For the first part, we generate independent samples with "small" sample size $n = 20$ and 50 and "large" sample size $n = 1000$ using `R`, where we use function `rstable` in the extension package `stable` to generate the unit symmetric α stable process. To be able to compare it with the result in Runde (1997), for the infinite variance case, we only consider the case of $\alpha > 1$. Here we consider some different values of α to generate the data. For calculating the statistic $Q_{GC} = Q_{GC}(\alpha, \sigma)$, we use both the true values and the estimators, where $\hat{\alpha}$ and $\hat{\sigma}$ are obtained using McCulloch's method (see Section 2.4). For "small" sample size, i.e., when $n = 20$ and 50, we generate $T = 10000$ independent realizations of the process (and for $n = 1000$ we take $T = 4000$), where we calculated the statistic (5.16) for several values of m. To compare the performance of Q_{GC} with $Q^{\alpha}_{n,m}$ (see (5.13)), here we only consider lag $m = 2$ and $m = 5$, since the

	$m=2$			$m=5$		
	$1-\theta$			$1-\theta$		
α	0.9	0.95	0.99	0.9	0.95	0.99
1.9	15.6	17.94	22.07	21.75	24.87	26.31
1.8	14.29	16.74	20.82	20.57	23.16	25.14
1.7	13.41	15.83	19.68	19.42	21.73	24.08
1.6	12.09	14.47	18.54	18.23	20.43	22.95
1.5	11.12	13.53	17.138	17.2	19.17	21.89
1.4	10.27	12.37	16.2	16.11	18.24	21.06
1.3	9.31	11.52	15.03	15.09	17.14	19.97
1.2	8.87	10.87	13.92	14.28	15.92	19.04
1.1	8.11	10.05	13.04	13.7	15.08	18.18

Table 5.6: Critical values of the limiting distribution $Q_m(\alpha)$

limiting distribution $Q_m(\alpha)$ of the statistic $Q^{\alpha}_{n,m}$ as in (5.14) are available only for limited values of m (i.e., only for $m = 2, 3, 4, 5$, see Table 1 in Runde (1997)). The critical values of $Q_m(\alpha)$ for considered values of the parameters used in the test above are reproduced in Table 5.6 above). To appreciate the accuracy of χ^2 approximation, the nominal levels of test are chosen as $\theta = 0.01$ and 0.05.

From the experience of simulation studies in Section 5.1, we know that the location of grid points $s_1, \ldots, s_r$ is of critical importance for the estimation accuracy of $\hat{I}(\cdot)$, and therefore also for Q_{GC}. In Section 5.1.1, it was discussed that for a fixed choice of r, the $s_j, j = 1, \ldots, r$ have to be chosen within an interval near zero, since at this interval, we can accurately estimate $I(\cdot)$ and at the other choices of grid points, the value of $\hat{I}(\cdot)$ will be highly biased. The sets of grid points $\mathbf{s} = \{s_1, \ldots, s_r\}$ which are considered in the simulation study in this part are listed as follows: $\mathbf{s}_1 = \{0.01, 0.2\}$, $\mathbf{s}_2 = \{0.01, 0.3\}$, $\mathbf{s}_3 = \{0.01, 0.5\}$, $\mathbf{s}_4 = \{0.01, 0.1, 0.2, 0.3\}$, $\mathbf{s}_5 = \{0.1, 0.2, 0.3\}$, $\mathbf{s}_6 = \{0.01, 0.06, \ldots, 0.26\}$, $\mathbf{s}_7 = \{0.01, 0.02, \ldots, 0.3\}$, $\mathbf{s}_8 = \{0.1, 0.11, \ldots, 0.3\}$. When $\alpha = 1.1$ and $\alpha = 1.9$, in addition to these sets of grid points, we also consider the following sets: $\mathbf{s}_9 = \{0.01, 0.05\}$, $\mathbf{s}_{10} = \{0.01, 0.1\}$, $\mathbf{s}_{11} = \{0.1, 0.2, \ldots, 0.5\}$, $\mathbf{s}_{12} = \{0.1, 0.2, \ldots, 1\}$. For $n = 1000$, we additionally consider the other sets of grid points, as follows: $\mathbf{s}_{13} = \{0.01\}$, $\mathbf{s}_{14} = \{0.05\}$, $\mathbf{s}_{15} = \{0.1\}$, $\mathbf{s}_{16} = \{0.2\}$, $\mathbf{s}_{17} = \{0.3\}$,$\mathbf{s}_{18} = \{0.4\}$, $\mathbf{s}_{19} = \{0.5\}$, $\mathbf{s}_{20} = \{0.01, 0.05\}$,$\mathbf{s}_{21} = \{0.01, 1\}$.

Results of the simulation in the first part are summarized in Table 5.7 and 5.8. In Table 5.7, we also record two sets of grid points $\mathbf{s}$ among all considered grid points above, which give the closest empirical levels to the theoretical ones using both of the weighting methods, based on the performance of the statistic $Q_{GC}(\hat{\alpha}, \hat{\sigma})$. The findings show that the optimal choices of grid $\mathbf{s}$ depend on sample size n, the maximum lag m, the level θ, and also the index α. The results in this table indicate that throughout the whole experiments, the small sample performance of the statistic Q_{GC} is relatively poor, as its empirical levels are in many cases higher than the theoretical ones, although its performance clearly depends on the choice of $\mathbf{s}$. However, it is very interesting to note that the small sample performance of Q_{GC} in terms of its empirical levels is much better than the performance of statistic $Q^{\alpha}_{n,m}$ (which turns out to be completely different from the simulation results reported in Runde (1997)). Note that our results regarding statistic $Q^{\alpha}_{n,m}$ in Table 5.8 are similar to the finite sample property of a similar statistic reported in Krämer and Runde (2001), namely the limiting distribution for $Q^{\alpha}_{n,m}$ provides a very poor approximation for finite sample distribution. Indeed, as similarly reported in Krämer and Runde (2001), we find that the discrepancies are still persistent even for the sample size as large as $n = 1000$.

In the second part, to asses the small sample properties of the proposed statistic in case of the finite variance, we conducted a simulation study where it was generated three "small" samples of N(0,1) sequences with sample size $n = 20, 50$ and 100. We use function `rnorm` in `R` to generate $T = 10000$ independent realization of standard Gaussian white noise. Here we use $\alpha = 2$ and σ in (5.18) is estimated as $\hat{\sigma} = \sqrt{s_n^2}/\sqrt{2}$, where s_n^2 denotes the sample variance. We calculated the statistic (5.16) for several values of m. To appreciate the accuracy of χ^2 approximation, the nominal levels of test are chosen as $\theta = 1, 5, 10$ and 20 percent. In simulation, the following sets

Level	α	Weighting Method	s	$n=20$ $Q_{GC}(\alpha,\sigma)$ $m=2$	$m=5$	$Q_{GC}(\hat{\alpha},\hat{\sigma})$ $m=2$	$m=5$	s	$n=50$ $Q_{GC}(\alpha,\sigma)$ $m=2$	$m=5$	$Q_{GC}(\hat{\alpha},\hat{\sigma})$ $m=2$	$m=5$	s	$n=1000$ $Q_{GC}(\alpha,\sigma)$ $m=2$	$m=5$	$Q_{GC}(\hat{\alpha},\hat{\sigma})$ $m=2$	$m=5$
0.01	1.1	Avg.	s_2	0.73	1.56	0.58	1.07	s_1	0.8	1.13	0.77	0.95	s_{10}	0.925	1.375	0.925	1.375
			s_{10}	1.69	3.57	1.23	2.36	s_{10}	1.67	2.2	1.47	1.92	s_3	0.85	1.075	0.825	1.05
		Exp.	s_2	0.74	1.61	0.59	1.04	s_1	0.81	1.16	0.77	1	s_2	0.925	1.275	0.875	1.25
			s_{10}	1.69	3.59	1.24	2.36	s_{10}	1.67	2.2	1.49	1.92	s_3	1.025	1.55	1.025	1.575
	1.3	Avg.	s_1	1.75	3.44	1.08	2.19	s_6	1.34	1.58	1.04	1.25	s_{10}	1.15	1.45	1.15	1.475
			s_8	1.25	2.39	0.78	1.58	s_7	1.08	1.29	0.83	1.01	s_{21}	0.85	1.075	0.825	1.025
		Exp.	s_1	1.78	3.45	1.08	2.21	s_6	1.37	1.61	1.06	1.31	s_1	0.975	1.275	1	1.275
			s_8	1.24	2.39	0.79	1.59	s_7	1.1	1.32	0.87	1.04	s_3	1	1.425	1	1.425
	1.5	Avg.	s_5	1.47	3.32	1.01	2.1	s_4	1.34	2.18	1.07	1.62	s_{10}	1.175	1.475	1.075	1.475
			s_8	1.48	3.33	1.01	2.13	s_5	1.17	1.71	0.88	1.36	s_1	0.825	1.025	0.825	1.025
		Exp.	s_5	1.5	3.33	1.02	2.11	s_4	1.37	2.26	1.1	1.67	s_{10}	1.175	1.5	1.15	1.475
			s_8	1.49	3.35	1.01	2.13	s_5	1.18	1.77	0.88	1.37	s_2	0.875	1.025	0.85	1.025
	1.7	Avg.	s_5	1.48	3.01	1.09	2.05	s_3	0.86	1.51	0.74	1.2	s_{16}	1.125	0.75	1	0.75
			s_8	1.49	3.04	1.08	2.02	s_8	1.24	2.48	1.01	1.92	s_1	0.975	1.125	0.85	1
		Exp.	s_3	1.43	2.99	1.05	2.01	s_5	1.22	2.45	1.02	1.92	s_{16}	1.125	0.75	1	0.75
			s_8	1.49	3.05	1.08	2.02	s_8	1.25	2.49	1.01	1.97	s_1	1.025	1.125	0.9	1.05
	1.9	Avg.	s_8	0.97	2.19	1.46	2.88	s_3	0.94	1.39	1.15	1.88	s_1	1	1.075	1	1.2
			s_{12}	0.26	0.38	0.65	0.96	s_{12}	0.68	0.74	0.74	1.14	s_2	0.85	0.9	0.875	1.075
		Exp.	s_3	1.24	2.45	1.48	2.77	s_3	0.96	1.34	1.17	1.91	s_{16}	1.05	1.1	1.1	1.225
			s_5	0.96	2.18	1.48	2.86	s_{12}	0.72	1.04	0.82	1.27	s_1	0.975	1.075	1	1.2
0.05	1.1	Avg.	s_{10}	6.91	9.16	5.14	6.77	s_9	7.6	10.65	6.95	9.73	s_{14}	5.4	5.675	5.425	5.65
			s_1	4.77	6.39	3.65	4.94	s_{10}	5.04	6.78	4.72	6.07	s_{15}	5.325	5.2	5.3	5.225
		Exp.	s_{10}	6.93	9.21	5.15	6.77	s_9	7.6	10.65	6.95	9.74	s_{14}	5.4	5.675	5.425	5.65
			s_1	4.84	6.45	3.69	4.96	s_{10}	5.07	6.78	4.72	6.08	s_{15}	5.325	5.2	5.3	5.225
	1.3	Avg.	s_1	6.27	9.66	4.01	6.11	s_1	4.92	6.43	4.11	5.3	s_{15}	5.125	5.425	5.15	5.45
			s_7	5.09	7.92	3.39	5.17	s_6	4.19	5.35	3.45	4.51	s_{16}	5.475	5.525	5.525	5.475
		Exp.	s_1	6.32	9.71	4.05	6.12	s_1	4.99	6.49	4.19	5.37	s_{15}	5.125	5.425	5.15	5.45
			s_7	5.13	8	3.4	5.21	s_6	4.21	5.42	3.48	4.53	s_{16}	5.475	5.525	5.525	5.475
	1.5	Avg.	s_1	7.74	11.49	5.26	7.36	s_1	6.35	8.46	4.96	6.29	s_{14}	5.55	7.6	5.35	7.65
			s_6	7.33	10.77	4.94	6.96	s_6	5.77	7.58	4.42	5.53	s_{20}	4.825	7.2	4.725	6.925
		Exp.	s_1	7.76	11.55	5.31	7.39	s_1	6.43	8.51	5.06	6.35	s_{14}	5.55	7.6	5.35	7.65
			s_3	5.14	7.46	3.64	5.16	s_7	5.35	6.85	4.15	5.04	s_{20}	4.825	7.2	4.725	6.925
	1.7	Avg.	s_2	6.26	9.34	5.14	7.23	s_5	5.42	6.43	4.33	5.3	s_{14}	5.975	7.55	5.85	7.15
			s_5	6.24	9.32	5.09	7.08	s_6	6.16	7.52	5.1	6.31	s_{10}	4.775	5.75	4.65	5.575
		Exp.	s_5	6.26	9.33	5.09	7.1	s_6	6.16	7.56	5.1	6.35	s_{14}	5.975	7.55	5.85	7.15
			s_8	6.34	9.45	5.15	7.17	s_7	5.92	7.17	4.83	5.88	s_{10}	4.8	5.775	4.65	5.6
	1.9	Avg.	s_7	5.14	7	6.16	8.82	s_3	4.51	5.23	5.14	6.02	s_{15}	5.35	5.275	5.175	5.225
			s_8	5.23	7.09	6.05	8.67	s_{11}	4.68	5.53	5.24	6.23	s_{10}	5.65	5.625	5.475	5.85
		Exp.	s_3	5.1	7.19	6.04	8.59	s_3	4.64	5.33	5.31	6.14	s_{15}	5.35	5.275	5.175	5.225
			s_{12}	2.83	3.17	3.62	4.48	s_{11}	4.73	5.5	5.29	6.31	s_{10}	5.65	5.625	5.475	5.875

Table 5.7: Empirical levels (in %) of the statistic Q_{GC} for some sample size n. Under H_0, the process is assumed $S\alpha S$ distributed.

		$n = 20$		$n = 50$		$n = 1000$	
		m		m		m	
level	α	2	5	2	5	2	5
0.01	1.1	0.13	0.26	2.29	4.29	10.3	22.05
	1.3	0	0.02	0.04	0.07	2.9	6.8
	1.5	0	0	0	0	0.675	1.1
	1.7	0	0	0	0	0.05	0.075
	1.9	0	0	0	0	0	0
0.05	1.1	0.69	1.21	4.5	7.63	12.475	25.5
	1.3	0	0.01	0.18	0.26	3.75	7.95
	1.5	0	0	0	0	0.65	1.2
	1.7	0	0	0	0	0.075	0.125
	1.9	0	0	0	0	0	0

Table 5.8: Empirical levels (in %) of the statistic $Q^{\alpha}_{n,m}$, for some sample sizes. Under H_0, the process is assumed $S\alpha S$ distributed.

of grid points $\mathbf{s}_i = \{s_1, \ldots, s_{r_i}\}, i = 1, 2, \ldots, 4$ are considered: $\mathbf{s}_1 = \{0.01\}$, $\mathbf{s}_2 = \{0.01, 0.5\}$, $\mathbf{s}_3 = \{0.01, 0.1, 1\}$, $\mathbf{s}_4 = \{0.1, 0.2, \ldots, 1\}$. Furthermore, to compare the accuracy of the statistic Q_{GC}, we also recorded the nominal levels of the statistic $\tilde{Q}_{n,m}$ as in (5.11). It was shown in Ljung and Box (1978) that this statistic has better empirical sizes than the alternative statistic $Q_{n.m}$ as in (5.10).

Results of the experiments are summarized in Table 5.9. The findings indicate that in many cases, under the suitable choices of grid points $\mathbf{s}$, the empirical levels of Q_{GC} are closer to the theoretical ones than Ljung-Box statistic $Q_{n,m}$. This result will be more apparent when the lag m is large with respect to the sample size n, in which case the difference between the empirical and the theoretical levels of Ljung-Box statistic $Q_{n,m}$ can be significantly large. The optimal choice of grids $\mathbf{s}$ in general depends on the maximum lag m, the nominal level θ and the sample size n. However, for its practical application, the choice of $\mathbf{s}$ as above seems to be sufficient, at least for the considered sample size.

The empirical power of Q_{GC}

To examine the power of the test, we consider simulation studies using the alternative hypothesis that the process is autoregressive process of order 1, AR(1). For each sample size, the function `arima.sim` in R is used to generate AR(1) processes $X_t = \phi X_{t-1} + \epsilon_t$, with the parameters $\phi = 0.3$ and $\phi = 0.8$. The sample sizes considered are $n = 50$ and 200.

In this part, we consider two simulation sets. For the case of infinite variance, we consider ϵ_t is i.i.d. $S\alpha S$ for some different values of α and σ is assumed to be 1, where we generated 10000 independent realization of the process. In the finite variance case, we generated 4000 independent realization of the process with the normal standard (N(0,1)) innovations, thus $\sigma = 1/\sqrt{2}$ in (5.18). The statistic Q_{GC} is evaluated using the true values of α and σ, using the following sets of grid points: $\mathbf{s}_1 = \{0.01, 0.1\}$, $\mathbf{s}_2 = \{0.01, 0.2\}$, $\mathbf{s}_3 = \{0.01, 0.3\}$, $\mathbf{s}_4 = \{0.01, 0.5\}$, $\mathbf{s}_5 = \{0.01, 1\}$, $\mathbf{s}_6 = \{0.01, 0.1, 0.3\}$,$\mathbf{s}_7 = \{0.01, 0.1, 0.5\}$, $\mathbf{s}_8 = \{0.1, 0.2, 0.3\}$, $\mathbf{s}_9 = \{0.1, 0.2, 0.5\}$, $\mathbf{s}_{10} = \{0.1, 0.2, 1\}$, $\mathbf{s}_{11} = \{0.01, 0.02, \ldots, 0.1\}$, $\mathbf{s}_{12} = \{0.1, 0.11, \ldots, 0.2\}$, $\mathbf{s}_{13} = \{0.2, 0.21, \ldots, 0.3\}$, $\mathbf{s}_{14} = \{0.01, 0.02, \ldots, 0.3\}$. Here we only use the simple average as the weighting methods. In case of infinite variance process, we compare the power of the statistic Q_{GC} with the statistic $Q^{\alpha}_{n,m}$ as in (5.13), whereas for the finite variance we compare the power of the statistic Q_{GC} with the statistic $\tilde{Q}_{n.m}$ as in (5.11). Under the same consideration as in the previous subsection, here we consider lags $m = 2$ and $m = 5$ in the infinite variance case. On the other hand, for finite variance case, the choice of m is more flexible since both of Q_{GC} and $\tilde{Q}_{n,m}$ are asymptotically χ^2_m distributed and therefore, the quantiles of the limiting distribution can be obtained easily for any choice of m. For the simulation, the level is chosen to be 5%. The results of simulation are summarized in table 5.10 and table 5.11 for the infinite and the finite variance cases, respectively.

			$n = 20$				$n = 50$						$n = 100$						
			m				m						m						
% level	method	grids	1	3	5	10	1	3	5	10	15	25	1	3	5	10	15	25	50
1	$Q_{GCAvg.}$	s_1	0.5	0.75	1.24	1.73	0.88	0.86	1.18	1.8	2.48	2.96	0.78	0.98	1.3	1.62	2.05	2.81	3.88
		s_2	0.7	0.99	1.48	2.29	0.88	0.96	1.22	2.01	2.62	3.33	0.76	1.02	1.29	1.57	2.04	2.92	4.04
		s_3	1.04	1.22	1.66	2.17	0.94	1.01	1	1.74	2.33	2.83	0.68	0.76	0.95	1.02	1.3	1.8	2.3
		s_4	1.24	1.49	2.19	2.93	1.1	1.19	1.32	2.15	2.95	3.75	0.78	0.99	1.19	1.37	1.74	2.45	3.39
	$Q_{GCExp.}$	s_1	0.5	0.75	1.24	1.73	0.88	0.86	1.18	1.8	2.48	2.96	0.78	0.98	1.3	1.62	2.05	2.81	3.88
		s_2	0.67	0.91	1.45	2.18	0.89	0.95	1.22	1.97	2.61	3.24	0.76	1.01	1.28	1.55	2.04	2.9	3.98
		s_3	0.64	0.81	1.2	1.66	0.77	0.78	0.94	1.69	2.07	2.64	0.65	0.77	1.03	1.06	1.49	1.95	2.73
		s_4	1.09	1.25	1.88	2.68	0.97	1.06	1.13	2.03	2.64	3.5	0.8	0.97	1.18	1.35	1.77	2.53	3.43
	$\tilde{Q}_{n,m}$	-	0.83	1.2	1.75	2.74	1.03	1.04	1.39	2.23	2.91	3.74	0.83	1.06	1.45	1.79	2.25	3.14	4.56
5	$Q_{GCAvg.}$	s_1	4.47	3.77	4.57	5.41	4.61	4.17	4.49	5.78	6.48	7.51	4.74	4.96	5.12	5.54	6.35	7.45	8.95
		s_2	4.97	4.55	5.33	6.49	4.73	4.32	4.57	6.12	6.93	7.94	4.79	4.86	5.03	5.68	6.39	7.57	9.11
		s_3	4.73	4.56	5.03	5.62	4.17	3.82	4	4.98	5.66	6.37	3.93	3.85	3.63	3.95	4.38	5.09	5.71
		s_4	5.46	5.69	6.55	7.37	4.72	4.51	4.77	6.42	7.16	8.08	4.53	4.73	4.56	5.11	5.68	7.02	7.97
	$Q_{GCExp.}$	s_1	4.47	3.77	4.57	5.41	4.61	4.17	4.49	5.78	6.48	7.51	4.74	4.96	5.12	5.54	6.35	7.45	8.95
		s_2	4.96	4.4	5.22	6.35	4.77	4.28	4.47	6	6.93	7.84	4.79	4.87	5.04	5.65	6.35	7.54	9.07
		s_3	4.21	3.77	4.44	4.95	4.24	3.72	3.88	5.03	5.65	6.21	4.24	4.16	4.12	4.44	5.02	5.88	6.51
		s_4	5.08	5.2	5.97	7.01	4.78	4.31	4.53	6.11	6.85	7.79	4.57	4.77	4.68	5.27	5.82	6.92	7.92
	$\tilde{Q}_{n,m}$	-	5.67	5.5	6.3	8.12	5.03	4.77	5.26	6.73	7.95	9.3	5.01	5.25	5.61	6.03	6.95	8.42	10.17
10	$Q_{GCAvg.}$	s_1	9.06	7.52	8.41	9.21	9.25	9.26	9.36	10.25	10.98	11.99	9.63	9.73	9.87	10.14	10.73	12.13	13.33
		s_2	9.69	8.51	9.38	10.57	9.41	9.59	9.82	10.78	11.65	12.67	9.76	9.58	9.82	10.32	10.93	12.15	13.71
		s_3	8.59	7.89	8.51	8.8	8.57	8.33	8.62	9.28	9.57	9.95	8.58	7.89	7.79	7.51	7.93	8.48	8.66
		s_4	9.93	9.77	10.66	11.67	9.44	9.65	9.95	10.93	11.86	12.65	9.52	9.26	9.18	9.44	9.97	10.98	12.27
	$Q_{GCExp.}$	s_1	9.06	7.52	8.41	9.21	9.25	9.26	9.36	10.25	10.98	11.99	9.63	9.73	9.87	10.14	10.73	12.13	13.33
		s_2	9.58	8.35	9.3	10.3	9.39	9.58	9.7	10.65	11.57	12.51	9.72	9.57	9.8	10.28	10.9	12.16	13.63
		s_3	8.85	7.03	7.86	8.25	8.65	8.56	8.49	9.1	9.55	10.23	8.95	8.37	8.42	8.28	8.58	9.47	10.3
		s_4	9.98	9.17	10.11	11.06	9.29	9.39	9.71	10.46	11.62	12.38	9.47	9.42	9.33	9.39	9.92	11.02	12.22
	$\tilde{Q}_{n,m}$	-	10.68	10.04	11.21	13.36	9.93	10.27	10.61	11.78	12.86	14.61	9.92	10.19	10.47	11.16	11.7	13.38	15.21
20	$Q_{GCAvg.}$	s_1	18.98	16.65	16.42	16.9	19.58	18.54	18.34	18.41	19.11	19.29	20.38	19.63	19.59	19.35	19.66	20.5	21.97
		s_2	19.54	18.01	18.17	18.49	19.74	19	19.01	19.3	19.96	20.19	20.39	19.5	19.67	19.24	19.66	20.95	22.07
		s_3	17.45	15.89	15.93	15.58	18.11	16.77	16.21	16.48	16.66	15.89	18.03	16.7	15.86	14.54	14.82	14.96	15.03
		s_4	19.31	18.84	19.45	19.2	19.45	19.05	18.87	19.55	19.85	19.63	19.46	18.51	18.08	17.76	18.08	18.96	19.85
	$Q_{GCExp.}$	s_1	18.98	16.65	16.42	16.9	19.58	18.54	18.34	18.41	19.11	19.29	20.38	19.63	19.59	19.35	19.66	20.5	21.97
		s_2	19.49	17.75	17.87	18.34	19.65	18.85	18.85	19.13	19.71	19.96	20.38	19.48	19.55	19.26	19.49	20.71	22.03
		s_3	18.14	15.89	15.52	15.25	18.68	17.09	16.76	16.58	16.9	16.18	19.16	17.49	17.19	16.28	16.32	16.6	17.16
		s_4	19.47	18.32	18.74	18.72	19.37	18.83	18.44	19.1	19.57	19.17	19.48	18.69	18.36	17.92	18.3	19.22	20.12
	$\tilde{Q}_{n,m}$	-	21.02	20.68	20.65	22.64	20.59	20.39	20.27	21.05	21.88	22.85	20.86	20.38	20.38	20.68	21.15	22.44	24.47

Table 5.9: Empirical levels (in %) of test based on Q_{GC} and $\tilde{Q}_{n,m}$ at some nominal significance level (in %). Under hypothesis H_0, process is assumed $N(0,1)$ distributed. Here $\mathbf{s}_1 = \{0.01\}$, $\mathbf{s}_2 = \{0.01, 0.5\}$, $\mathbf{s}_3 = \{0.01, 0.1, 1\}$ and $\mathbf{s}_4 = \{0.1, 0.2, \ldots, 1\}$.

				Q_{GC}		$Q^{\alpha}_{n,m}$	
				m		m	
ϕ	n	α	s	2	5	2	5
0.3	50	1.1	s_1	62.875	49.125	33.27	29.48
		1.3	s_1	59.3	48.375	4.98	3.51
		1.5	s_1	57	46.925	0.15	0.06
		1.7	s_1	48.1	40.375	0	0.01
		1.9	s_1	40.3	33.2	0	0
	200	1.1	s_2	99.925	99.825	99.1	98.99
		1.3	s_2	99.875	99.725	95.12	90.7
		1.5	s_1	99.825	99.7	33.11	15.7
		1.7	s_1	99.625	98.975	1.26	0.34
		1.9	s_1	97.8	94.375	0	0
0.8	50	1.1	s_2	99.99	99.97	99.18	99.16
		1.3	s_2	99.97	99.95	97.95	96.61
		1.5	s_2	99.97	99.91	88.49	80.06
		1.7	s_2	99.92	99.77	50.2	42.85
		1.9	s_2	99.79	99.48	5.82	12.13
	200	1.1	all	100	100	99.94	99.94
		1.3	all	100	100	99.97	99.97
		1.5	all	100	100	99.94	99.94
		1.7	all	100	100	99.94	99.94
		1.9	all	100	100	99.94	99.94

Table 5.10: Empirical power (in %) of test based on Q_{GC} and $Q^{\alpha}_{n,m}$ at nominal significance level 5% where the assumed model is i.i.d. SαS, $\alpha < 2$, but the true model is AR(1), $X_t = \phi X_{t-1} + \epsilon_t$, ϵ_t is SαS process, $\alpha < 2$. Here, we only record the grid points which maximize the power, however, when $\phi = 0.3$ and $n = 200$ or $\phi = 0.8$ and $n = 50$, the powers of all considered grid points are approximately the same.

Throughout the experiments, we found that in infinite variance case, the test based on Q_{GC} has better power than the alternative $Q^{\alpha}_{n,m}$ (which turns out have a poor power). However, in case of finite variance, the performance of Q_{GC} is slightly lower than the alternative statistic. Thus in general, we may conclude that the Pormanteau statistic, which is based on the codifference function, is a powerful statistic for detecting serial dependence in AR(1) form.

			Q_{GC}			$\tilde{Q}_{n,m}$		
			m			m		
ϕ	n	s	5	15	30	5	15	30
0.3	50	s_1	25.475	23.65	23.725	27.6	25.825	28.175
	200	s_1	91.8	77.475	66.725	91.95	77.975	67.775
0.8	50	s_1	99.1	97.575	96.45	99.225	97.875	97.4
	200	all	100	100	100	100	100	100

Table 5.11: Empirical power (in %) of test based on Q_{GC} and $\tilde{Q}_{n,m}$ at nominal significance level 5% where the assumed model is i.i.d. N(0,1), but the true model is AR(1), $X_t = \phi X_{t-1} + \epsilon_t$, ϵ_t is N(0,1). Here, we only record the grid points which maximize the power, however, when $\phi = 0.3$ and $n = 200$ or $\phi = 0.8$ and $n = 50$, the powers of all considered grid points are approximately the same.

Chapter 6

Conclusions

The central discussion throughout this thesis is on a dependence measure of univariate strictly stationary time series processes, which is called the codifference function. This measure is designated for generalization of the covariance function, and it does not require the existence of moments of the process of any order. We in particular focus our study on the properties of the population codifference function and its estimator for the particular class of heavy-tailed linear time series model, which we call causal stable ARMA with symmetric α-stable noise, $0 < \alpha \leq 2$. Some necessary information on the α-stable distributions were reviewed in Chapter 2, where a quick overview of the available results in literature about the α-stable ARMA modelling was summarized in Chapter 3.

The main contribution of this thesis is given in Chapter 4. We proposed estimators for the codifference and the normalized codifference function, defined via the *ecf*, where for linear processes with geometrically bounded coefficients and $S\alpha S$ noise, we established the asymptotic properties of the proposed estimators. The results in Chapter 4 solve some problems of the SACF that we posed in Section 1.2. Here we see that unlike the ACF estimator, there is no discontinuity in either the normalization or the limiting distribution of the proposed estimators of the codifference and the normalized codifference when $\alpha \to 2$. Moreover, when $\alpha < 2$ we note that unlike the SACF, which has an unfamiliar limiting distribution and is relatively difficult to obtain the quantiles of the limiting distribution, the estimators of the codifference and the normalized codifference is asymptotically normally distributed at the same rate as the SACF in the classical case (that is $n^{-1/2}$), although the asymptotic variance is different.

In Chapter 5, we consider some inference problems using the results, which are presented in Chapter 4. In Section 5.1, we presented simulation studies for investigating the small sample properties of the normalized codifference estimator. In particular, we address some practical issues in the calculation of the codifference estimator. A short guideline for calculating the codifference function is discussed in Section 5.1.1.

In Section 5.2, we discussed a new method for fitting pure moving averages models to dependent heavy-tailed data, which was based on the codifference function. The performance of the proposed method is evaluated via simulation studies. Simulation results indicate that the method based on the codifference function works fairly well for identification and estimation purposes.

Further application of the sample codifference function was discussed in Section 5.3. Here we consider a Portmanteau-type test of randomness, i.e., the test for independence against serial dependence alternatives, for symmetric α stable random variables with the exponent $0 < \alpha \leq 2$, using Box-Pierce Q-statistic which is defined using the codifference function. We obtain that, unlike a similar test proposed in Runde (1997), the asymptotic distribution of the proposed statistic is similar to the classical case, that is asymptotically χ^2 distributed, both in the finite and the infinite variance cases. Simulation studies are performed to obtain the small sample performance

of the proposed statistic. We found that the proposed statistic works fairly well for small sample, although it introduces the problem of choosing the appropriate sets of grid points s. In the infinite variance case, its empirical levels are much closer to the theoretical ones and its power is much better than Runde's statistic. In the finite variance case, its empirical level and its power are approximately the same as that of Ljung Box's statistic (Ljung and Box, 1978). This result will be more apparent if the order of checked lag is relatively large with respect to the sample size. However, when the order of checked lag m is relatively large compared to the sample size n, the Portmanteau statistic, which is based on the codifference function, has been shown to have much closer empirical levels than the alternatives.

Notice that throughout this thesis, we have restricted ourselves with ARMA models. To the best of our knowledge, in literature the theoretical properties of the codifference function have been already studied for several time series models, for instance, for the class of Fractional Autoregressive Integrated Moving Averages (FARIMA) models (Kokoszka and Taqqu, 1995) and Randomized Generalized Autoregressive Conditionally Heteroscedastic (R-GARCH) (Nowicka and Weron, 2001). It is an interesting topic for further research to investigate the properties of the sample codifference function for those model classes.

Appendix A

Appendix

A.1 Some concepts related to the stable law

Theorem A.1.1 (Samorodnitsky and Taqqu (1994), Theorem 1.2.15). *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then,*

$$\lim_{x \to \infty} x^\alpha P(X > x) = C_\alpha \frac{1+\beta}{2} \sigma^\alpha$$

and

$$\lim_{x \to \infty} x^\alpha P(X < -x) = C_\alpha \frac{1-\beta}{2} \sigma^\alpha$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \text{if } \alpha \neq 1 \\ 2/\pi & \text{if } \alpha = 1 \end{cases}$$

Property A.1.2 (Fractional Lower Order Moment). *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then*

$$E|X|^p < \infty \text{ for any } 0 < p < \alpha$$

and

$$E|X|^p = \infty \text{ for any } p \geqslant \alpha$$

Proof. Let f denote the density function of a nonnegative random variable X. By changing the order of integration, according to Fubini's theorem, we obtain

$$\begin{aligned} \int_0^\infty P(X > \lambda) d\lambda &= \int_0^\infty \left(\int_\lambda^\infty f(x) dx \right) d\lambda \\ &= \int_0^\infty \left(\int_0^x f(x) d\lambda \right) dx \\ &= \int_0^\infty f(x) \left(\int_0^x d\lambda \right) dx \\ &= \int_0^\infty x f(x) dx = E(X) \end{aligned}$$

Therefore,

$$\begin{aligned} E(|X|^p) &= \int_0^\infty P(|X|^p > t) dt \\ &= \int_0^\infty p s^{p-1} P(|X| > s) ds \end{aligned} \tag{A.1}$$

Applying Theorem A.1.1 above, one can see that

$$s^{p-1}P(|X| > s) = O(s^{p-1}) \text{ as } s \to 0$$

and

$$s^{p-1}P(|X| > s) = O(s^{p-\alpha-1}) \text{ as } s \to \infty$$

Thus, one can conclude (A.1) is finite if and only if $0 \leq p < \alpha$ □

Theorem A.1.3 (Kolmogorov's three series theorem). *(See, e.g., Resnick, 2001, Theorem 7.6.1 and Remark 7.6.1.) Let $\{X_n, n \geq 1\}$ be an independent sequence of random variables. In order for $\sum_n X_n$ to converge a.s., it is necessary and sufficient that there exist $c > 0$ such that*

i. $\sum_n P(|X_n| > c) < \infty$

ii. $\sum_n \text{var}(X_n \mathbb{1}_{(|X_n| \leq c)}) < \infty$

iii. $\sum_n E(X_n \mathbb{1}_{(|X_n| \leq c)})$ *converges*

If $\sum_n X_n$ converges a.s., then (i), (ii), (iii) hold for any $c > 0$. Thus if the three series converge for one value of $c > 0$, they converge for all $c > 0$. If $\{X_n, n \geq 1\}$ are independent and nonnegative, it is only necessary to check convergence of (i) and (iii).

Proposition A.1.4. *Suppose we have time series, defined by*

$$X_n = \sum_{j=0}^{\infty} c_j \epsilon_{n-j}, n = 0, 1, \ldots$$

Under the conditions

C1. $|c_j| < CQ^{-j}$, *for some* $C > 0$ *and* $Q > 1$

C2. ϵ_t *is iid* $S\alpha S$

then by virtue of Kolmogorov three series theorem

$$\sum_{j=1}^{\infty} |c_j \epsilon_j| < \infty, \text{ almost surely}$$

Proof. By Kolmogorov's three series theorem, we need to verify that conditions C1 and C2 imply convergence of the two series

$$\sum_{j=1}^{\infty} \mathrm{P}(|c_j \epsilon_j| > 1) = \sum_{j=1}^{\infty} \bar{F}\left(\frac{1}{|c_j|}\right) < \infty \tag{A.2}$$

$$\sum_{j=1}^{\infty} \mathrm{E}(|c_j \epsilon_j| \mathbb{1}_{(|c_j \epsilon_j| \leq 1)}) = \sum_{j=1}^{\infty} |c_j| m\left(\frac{1}{|c_j|}\right) < \infty \tag{A.3}$$

where

$$m(t) = E(|\epsilon_1| \mathbb{1}_{(|\epsilon_1| \leq t)})$$

Note that from Theorem A.1.1, we obtain

$$P(|\epsilon_1| > x) = \bar{F}(x) \sim \tfrac{C_\alpha}{2} \sigma^\alpha x^{-\alpha}, x \to \infty \tag{A.4}$$

where C_α denotes constant which depends on α. Thus, we have here as $j \to \infty$

$$P(|c_j \epsilon_j| > 1) \sim \tfrac{C_\alpha}{2} \sigma^\alpha |c_j|^\alpha$$

Then from assumption C1

$$\sum_{j=1}^{\infty} P(|c_j\epsilon_j| > 1) \sim \sum_{j=1}^{\infty} \frac{C_\alpha}{2}\sigma^\alpha |c_j|^\alpha < \infty$$

which verifies (A.2). To verify (A.3), we observe that by Fubini's theorem

$$\begin{aligned} m(t) &= \int_0^t x dF(x) = \int_{x=0}^t \left(\int_{u=0}^x du \right) dF(x) \\ &= \int_{u=0}^t \left(\int_{x=u}^t dF(x) \right) du = \int_0^t \bar{F}(u)du - t\bar{F}(t) \\ &\leq \int_0^t \bar{F}(u)du \end{aligned} \tag{A.5}$$

From (A.4), given $\theta > 0$, there exists x_0 such that $x > x_0$ implies

$$\bar{F}(x) \leq (1+\theta)\frac{C_\alpha}{2}\sigma^\alpha x^{-\alpha} := k_1 x^{-\alpha}$$

thus from (A.5)

$$m(t) \leq \int_0^{x_0} + \int_{x_0}^t \leq c + k_1 \int_{x_0}^t u^{-\alpha} du, t \geq x_0 \tag{A.6}$$

For $\alpha > 1$, $E(|\epsilon_t|) < \infty$, so that

$$\sum\nolimits_j E(|c_j\epsilon_j| \mathbb{1}_{(|c_j\epsilon_j| \leq 1)}) \leq \sum\nolimits_j |c_j| E(|\epsilon_1|) < \infty$$

For $\alpha = 1$, from (A.6) we find that

$$m(t) \leq c' + k_2 \log(t), t \geq x_0$$

for positive constant c', k_2. From (A.3) we obtain for another constant c''

$$\begin{aligned} \sum\nolimits_j |c_j| m(\frac{1}{|c_j|}) &\leq c'' \sum\nolimits_j |c_j| + k_2 \sum\nolimits_j |c_j| \log(\frac{1}{|c_j|}) \\ &\leq c'' \sum\nolimits_j |c_j| + k_3 \sum_j |c_j| < \infty \end{aligned}$$

Finally, for $\alpha < 1$, $t > x_0$

$$m(t) \leq c_1 + k_1 t^{1-\alpha}$$

Thus

$$\sum\nolimits_j |c_j| m(\frac{1}{|c_j|}) \leq c2 \sum\nolimits_j |c_j| + k_4 \sum\nolimits_j |c_j|^{1+\alpha-1} < \infty$$

□

A.2 Stochastic processes

In this part, we recall a few basic concepts from probability theory

Definition A.2.1 (Random Variable). *A scalar random variable X is a mapping attaching each element ω in some probability space $(\Omega, \mathcal{A}, P)$ to a real number $X(\omega)$, which is $\mathcal{B}$ measurable. Here $\mathcal{A}$ denotes a σ-algebra on Ω and P a probability measure on $(\Omega, \mathcal{A})$, and $\mathcal{B}$ denotes the Borel σ-algebra on $\mathbb{R}$.*

Here we often deal with the sequence of random variables which often denoted in our context as the stochastic process. We will always assume that all random variables are defined on the same probability space.

Definition A.2.2. *A stochastic process is a family of random variables* $\{X_t, t \in T\}$ *all defined on the same probability space* $(\Omega, \mathcal{A}, P)$.

Note that in time series analysis, we interested with the index set T as a set of time points. In this thesis, we only consider $T = \mathbb{Z} = 0, \pm 1, \pm 2, \ldots$.

In the following, we recall several concepts from the asymptotic theory.

Definition A.2.3 (Convergence in probability). *A sequence* X_n *of scalar random variables is said to converge in probability to a random variable* X_0 *(in symbols* $X_n \xrightarrow{p} X_0$ *or* $X_n = o_p(1)$*), if for every* $\varepsilon > 0$, $P(|X_n - X_0| > \varepsilon) \to 0$ *for* $n \to \infty$.

Definition A.2.4 (Convergence in r-th mean, $r > 0$). *A sequence* X_n *of scalar random variables is said to converge in* r*-th mean to* X_0 *, written* $X_n \xrightarrow{r} X_0$ *if* $E(|X_n - X_0|^r) \to 0$ *for* $n \to \infty$. *An important order here is* $r = 2$*, that is the mean-square convergence, written* $l.i.m.X_n = X_0$.

Definition A.2.5 (Almost sure convergence). *A sequence* X_n *of scalar random variables is said to converge almost sure to a random variable* X_0 *(in symbols* $\lim X_n = X_0$ *a.s.), if* $P(\omega \in \Omega : X_n(\omega) - X_0(\omega)|) = 1$ *for* $n \to \infty$

Definition A.2.6 (Convergence in distribution). *A sequence* X_n *of scalar random variables is said to converge almost sure to a random variable* X_0 *(in symbols* $X_n \xrightarrow{d} X_0$*), if for every continuity point* x *of cdf* F_0 *of* X_0*, it holds that the real sequence of values* $F_n(x)$*, where* F_n *denote the cdf of* X_n*, converges to* $F_0(x)$*, i.e.* $F_n(x) \to F_0(x)$

The following laws are useful to analyze the behavior of a given random sequence.

- $l.i.m.X_n = X_0 \Rightarrow X_n \xrightarrow{p} X_0$
- (Slutzky's Theorem) Let $X_n \in \mathbb{R}^m$ for some integer m and let $\mathbf{X}$ be a vector constant. Further let the function $g : \mathbb{R}^m \to \mathbb{R}^k$ be continuous mapping. Then $X_n \xrightarrow{p} \mathbf{X} \Rightarrow g(X_n) \xrightarrow{p} g(\mathbf{X})$. The statement holds with a.s. convergence replacing convergence in probability
- (Continuous mapping theorem) Let g be as above. Then $X_n \xrightarrow{d} X_0 \Rightarrow g(X_n) \xrightarrow{d} g(pX_0)$.
- (Cramer's theorem) Let $X_n \in \mathbb{R}^m$ and $Y_n \in \mathbb{R}^m$. If $X_n \xrightarrow{d} X_0$ and $Y_n \xrightarrow{p} a$ where a constant vector, then (i) $(X_n + Y_n) \xrightarrow{d} X_0 + a$, (ii) $Y_n' X_n \xrightarrow{d} a' X_0$
- (The Cramer-Wold Device) Let $X_n \in \mathbb{R}^m$ be a sequence of random m-vectors. Then $X_n \xrightarrow{d} X$ if and only if $\lambda^T X_n \xrightarrow{d} \lambda^T X$ for all $\lambda = (\lambda_1, \ldots, \lambda_m)^T \in \mathbb{R}^m$.

The following theorem will be very useful for establishing the asymptotic distribution of some function of random variables.

Theorem A.2.7 (Brockwell and Davis (1987), Theorem 6.3.9). *Let* $\{\mathbf{X}_n\}, n = 1, 2, \ldots$ *and* $\{\mathbf{Y}_{nj}\}, n = 1, 2, \ldots, j = 1, 2, \ldots, n = 1, 2, \ldots$ *be a sequence of random* k*-vectors such that*

1. $\mathbf{Y}_{nj} \xrightarrow{d} \mathbf{Y}_j$ *as* $n \to \infty$ *for each* $j = 1, 2, \ldots$
2. $\mathbf{Y}_j \xrightarrow{d} \mathbf{Y}$ *as* $j \to \infty$*, and*

3. $\lim_{j\to\infty} \limsup_{n\to\infty} P(|\mathbf{X}_n - \mathbf{Y}_{nj}| > \epsilon) = 0$ *for every* $\epsilon > 0$

then

$$\mathbf{X} \xrightarrow{d} \mathbf{Y} \text{ as } n \to \infty$$

Theorem A.2.8 (Delta Method). *Let* $y_t \in \mathbb{R}^p$ *be a (vector valued) random sequence, such that* $\sqrt{N}(y_t - y_0) \xrightarrow{d} Z$*, where* y_0 *is a constant and* Z *is a multivariate normal random variable with mean zero and variance* V *. Let* $g : \mathbb{R}^p \to \mathbb{R}^m$ *denote a mapping, which is continuously differentiable at* y_0*. Then*

$$\sqrt{N}(g(y_t) - g(y_0)) \xrightarrow{d} \nabla(g)Z$$

where $\nabla(g)$ *denotes the Jacobian of* g *evaluated at* y_0*.*

The following results shows the situation when we can safely interchange expectation and limits. The proofs can be found in, e.g., Resnick (2001)

Theorem A.2.9 (Monotone Convergence Theorem (MCT)). *If* $0 \leq X_n \uparrow X$ *then* $0 \leq E(X_n) \uparrow E(X)$

Theorem A.2.10 (Series Version of MCT). *If* $X_n \geq 0$ *are non negative random variables for* $n \geq 1$ *then*

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n)$$

Theorem A.2.11 (Fatou Lemma). *If* $X_n \geq 0$ *then*

$$E(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} E(X_n)$$

More generally, if there exists $Z \in L_1$ *and* $X_n \geq Z$ *then*

$$E(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} E(X_n)$$

Theorem A.2.12 (Modified Fatou Lemma). *If* $X_n \leq Z$ *where* $Z \in L_1$*, then*

$$E(\limsup_{n\to\infty} X_n) \geq \limsup_{n\to\infty} E(X_n)$$

Theorem A.2.13 (Dominated Convergence Theorem). *If* $X_n \to X$ *and there exists a dominating random variable* $Z \in L_1$ *such that* $|X_n| \leq Z$

$$E(X_n) \to E(X)$$

In this thesis, we need for a central limit theorem which applies to sums of dependent random variables. It will be sufficient here to have a theorem which applies to m-dependent strictly stationary sequences, defined as follow

Definition A.2.14 (m-Dependence). *A strictly stationary sequence of random variables* $\{X_t\}$ *is said to be* m*-dependent (where* m *is a non-negative integer) if for each* t*, the two sets of random variables* $\{X_j, j \leq t\}$ *and* $\{X_j, j \geq t + m + 1\}$ *are independent.*

The property of m-dependence generalizes that of independence in a natural way. Observations of an m-dependent process are independent provided they are separated in time by more than m time units. In the special case when $m = 0$, m-dependence reduces to independence.

The following central limit theorem extend the classical central limit theorem of *iid* sequence to m-dependent sequences. See, e.g, Theorem 6.4.2 in Brockwell and Davis (1987) for the proof.

Theorem A.2.15 (The Central Limit Theorem for Strictly Stationary m-Dependent Sequences). *If* $\{X_t\}$ *is a strictly stationary* m*-dependent sequence of random variables with mean zero and autocovariance function* $\gamma(\cdot)$*, and if* $v_m = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j) \neq 0$*, then*

1. $\lim_{N\to\infty} N\text{var}(\bar{X}_N) = v_m$
2. $\bar{X}_N$ *is* $AN(0, v_m/N)$

A.3 The autocovariance function

Definition A.3.1 (Autocovariance function). *If $\{X_t, t \in T\}$ is a process such that $\mathrm{var}(X_t) < \infty$ for each $t \in T$, then the autocovariance function $\gamma_X(\cdot,\cdot)$ of $\{X_t\}$ is defined by*

$$\gamma_X(r,s) = \mathrm{cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)],\ r,s \in T$$

Definition A.3.2 (Weak Stationarity). *The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be weak stationarity if*

(i) $E|X_t|^2 < \infty$ *for all* $t \in \mathbb{Z}$

(ii) $EX_t = m$ *for all* $t \in \mathbb{Z}$ *and*

(iii) $\gamma_X(h) = \gamma_X(h,0) = \mathrm{cov}(X_{t+h}, X_t)$

Definition A.3.3 (Strict Stationarity). *The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strict stationarity if the joint distribution of $(X_{t_1}, \ldots, X_{t_k})^T$ and $(X_{t_{1+h}}, \ldots, X_{t_{k+h}})^T$ are the same for all positive integers k and for all $t_1, \ldots, t_k, h \in \mathbb{Z}$.*

In fact, the requirement of finite second moments in the definition weak stationary process exclude the stable Paretian process (for $\alpha < 2$). Therefore in that case, the term stationary is always referred to the strict stationary definition.

The following properties makes the importance of the autocovariance function in time series analysis (see, e.g.,, Brockwell and Davis (1987), Section 1.5.)

Property A.3.4. *If $\gamma(\cdot)$ is the autocovariance function of a (weak) stationary process $\{X_t, t \in \mathbb{Z}\}$, then*

(i) $\gamma(0) \geq 0$

(ii) $\gamma(h) \leq \gamma(0)$ *for all* $h \in \mathbb{Z}$ *and*

(iii) $\gamma(h) = \gamma(-h)$ *for all* $h \in \mathbb{Z}$

Definition A.3.5 (Non-negative Definiteness). *A real-valued function on the integers, $f : \mathbb{Z} \to \mathbb{R}$ is said to be non-negative definite if and only if*

$$\sum_{i,j=1}^{n} a_i f(t_i - t_j) a_j \geqslant 0$$

for all positive integers n and for all vectors $\mathbf{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$ and $\mathbf{t} = (t_1, \ldots, t_n)^T \in \mathbb{Z}^n$.

Theorem A.3.6 (Characterization of Autocovariance function). *A real-valued even function defined on the set $\mathbb{Z}$ of all integers is non-negative definite if and only if it is the autocovariance function of a stationary time series*

The following theorem defines the mean ergodicity property of stationary process.

Theorem A.3.7 (Mean Ergodicity). *Let $\{\mathbf{X}_t\}$ is stationary with mean μ and autocovariance function $\gamma(\cdot)$, then as $n \to \infty$,*

$$\mathrm{var}(\bar{X}_n) = E(\bar{X}_n - \mu)^2 \to 0 \ \textit{if}\ \gamma(n) \to 0$$

and

$$nE(\bar{X}_n - \mu)^2 \to \sum_{h=-\infty}^{\infty} \gamma(h) \ \textit{if} \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

See, e.g., Theorem 7.1.1. in Brockwell and Davis (1987) for the proof.

A.4 The central limit theorem for ECF of linear processes

Suppose that the distribution function of X is $F(x)$. The characteristic function of random variable X is defined as,

$$\Phi(s) = E\exp(isX) = \int \exp(isX)dF(x), s \in \mathbb{R} \tag{A.7}$$

For a given *iid* random sample $X_1, \ldots, X_n$, the empirical characteristic function (ECF) is defined as

$$\hat{\Phi}_n(s) = \frac{1}{n}\sum_{j=1}^{n}\exp(isX_j) = \int \exp(isX)dF_n(x), s \in \mathbb{R} \tag{A.8}$$

where $F_n(x)$ is the empirical distribution function (EDF).

Theorem A.4.1 (Hesse (1990), Theorem 1 and Remark 2.6). *Let $X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ be a linear process satisfying the conditions*

(C1). *The coefficients c_j satisfied $|c_j| < c\rho^j, 0 < \rho < 1, c > 0$*

(C2). *The distribution F of X_t has bounded density*

(C3). *ϵ_t are independent and identically distributed random variables*

in addition, let the sums in (A.9), (A.10), (A.11) be finite. Then the real and imaginary parts of

$$\phi_n(s) = n^{1/2}(\hat{\Phi}_n(s) - \Phi(s))$$

$s \in \mathbb{R}$, converge to a normal distribution with zero mean and variance-covariance structure given by

$$\lim_{n\to\infty} \mathrm{var}(\mathrm{Re}\,\phi_n(s)) = \frac{1}{2}(1 + \mathrm{Re}\,\Phi(2s)) - (\mathrm{Re}\,\Phi(s))^2 + 2\sum_{j=1}^{\infty}\mathrm{cov}(\cos sX_1, \cos sX_{1+j}) \tag{A.9}$$

$$\lim_{n\to\infty} \mathrm{var}(\mathrm{Im}\,\phi_n(s)) = \frac{1}{2}(1 - \mathrm{Re}\,\Phi(2s)) - (\mathrm{Im}\,\Phi(s))^2 + 2\sum_{j=1}^{\infty}\mathrm{cov}(\sin sX_1, \sin sX_{1+j}) \tag{A.10}$$

$$\begin{aligned}&\lim_{n\to\infty} \mathrm{cov}(\mathrm{Re}\,\phi_n(s), \mathrm{Im}\,\phi_N(s))\\ &= \frac{1}{2}\,\mathrm{Im}\,\Phi(2s) - \mathrm{Re}\,\Phi(s)\,\mathrm{Im}\,\Phi(s) + \sum_{j=1}^{\infty}(\mathrm{cov}(\cos sX_1, \sin sX_{1+j}) + \mathrm{cov}(\sin sX_1, \cos sX_{1+j}))\end{aligned} \tag{A.11}$$

Here $\mathrm{Re}(x)$ *and* $\mathrm{Im}(x)$ *denote, respectively, the real and imaginary parts of* x, *and* $\mathrm{cov}(X, Y)$ *denotes the covariance between* X *and* Y.

A.5 The complex logarithmic function

If z is a complex number then any solution w of $e^w = z (z \neq 0)$ is called a logarithmic of z. Notice that unlike real variable, the complex exponential is not $1 - 1$, it is periodic with period $2\pi i$, and so does not have a well-defined inverse. Because of periodicity all solutions will differ from one another by an integer multiple of $2\pi i$.

If we write the complex number $z = |z|\exp(i\mathrm{arg}z)$, then we have $w = \ln z = \ln|z| + i\ \arg z$. Now there is a choice of arguments, reflecting the periodicity discussed above. So unless we specify an interval in which the arguments must lie we have not defined w as a function of z.

Definition A.5.1. *If we choose the principal argument, satisfying* $-\pi < arg\ z \leq \pi$, *this defines the principal value of the logarithm, denoted by Log z*

Definition A.5.2. *If we choose the interval* $\alpha - \pi < arg_\alpha z \leq \alpha + \pi$, *this defines the branch of logarithmic function, given by* $\log_\alpha z = ln|z| + i arg_\alpha z$

Now $\ln|z|$ is continuous for all non-zero z, but $\arg_\alpha z$ is discontinuous for $\arg_\alpha z = \alpha + \pi$.

Bibliography

Adler, R. J., R. E. Feldman and C. M. Gallagher (1998*a*). Analyzing stable time series. In: *A Practical Guide to Heavy Tails : Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman and M. S. Taqqu, Eds.). pp. 133–158. Birkhäuser. Boston.

Adler, R. J., R. E. Feldman and M. S. Taqqu (1998*b*). *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser. Boston.

An, H. Z. and Z. G. Chen (1982). On converge of LAD estimates in autoregression with infinite variances. *Journal of Multivariate Analysis* **12**, 335–345.

Belkacem, L., J. L. Véhel and C. Walter (2000). CAPM, risk and portfolio selection in "stable" markets. *Fractals* **8**(1), 99–115.

Bhansali, R. J. (1983). Estimation for the order of MA model from autoregressive and windows estimates of the inverse correlation function. *Journal of Time Series Analysis* **4**, 137–162.

Bhansali, R. J (1988). Consistent order determination for processes with infinite variance. *Journal of the Royal Statistical Society B* **50**(1), 46–60.

Bloomfield, P. and W. L. Steiger (1983). *Least Absolute Deviations : Theory, Applications, and Algorithms*. Birkhäuser. Boston.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual correlations in Autoregressive-Integrated Moving Average time series models. *Journal of the American Statistical Association* **65**, 1509–1526.

Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods*. Springer-Verlag. New York.

Calder, M. and R. A. Davis (1998). Inference for linear processes with stable noise. In: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman and M. S. Taqqu, Eds.). pp. 159–176. Birkhäuser. Boston.

Cambanis, S. and G. Miller (1981). Some path properties of p-th order and symmetric stable processes. *Annals of Probability* **8**, 1148–1156.

Campbell, J. Y., A. W. Lo and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton Univ. Press.

Chambers, J. M., C. L. Mallows and B. W. Stuck (1976). A method for simulating stable random variables. *Jounal of the American Statistical Association* **71**, 340–344.

Davis, R. A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Processes and their Applications* **63**, 75–95.

Davis, R. A. and S. I. Resnick (1986). Limit theory for the sample correlation function of moving averages. *Annals of Statistics* **14**, 533–558.

Davis, R. A. and T. Mikosch (1998). Gaussian likelihood-based inference for non-invertible MA(1) processes with $S\alpha S$ noise. *Stochastic Processes and their Applications* **77**, 99–122.

Davis, R. A., K. Knight and J. Liu (1992). M-estimation for autoregressions with infinite variance.. *Stochastic Processes and Their Applications* **40**(1), 145–180.

Doganoglu, T. and S. Mittnik (1998). An approximation procedure for asymmetric stable Paretian densities. *Computational Statistics* **13**, 463–475.

Doganoglu, T. and S. Mittnik (2003). Portfolio selection, risk assessment and heavy tails. Technical report. Ludwig-Maximilians-Universität München.

Embrechts, P., C. Klüppelberg and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer Verlag. Berlin.

Fama, E. (1965). The behavior of stock market prices.. *Journal of Business* **38**, 34–105.

Fama, E. and R. Roll (1971). Parameter estimates for symmetric stable distributions.. *Journal of the American Statistical Association* **66**, 331–338.

Feigin, P., M. Kratz and S. Resnick (1994). Parameter estimation for Moving Averages with positive innovations. Technical report. School of Operations Research and Industrial Engineering, Cornell University. Available as `TR1111.ps.Z` at `http://www.orie.cornell.edu/trlist/trlist.html`.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*. Volume 2, 2nd edition, Wiley, New York.

Fofack, H. and J. P. Nolan (1999). Tail behavior, modes and other characteristics of stable distributions. *Extremes* **2**(1), 39–58.

Gallagher, C. (1998). Fitting ARMA models to heavy tailed data. PhD thesis. University of California, Santa Barbara.

Gallagher, C. (2000). Estimating the autocovariation functions with application to heavy-tailed ARMA modeling. Technical report. Department of Mathematical Sciences, Clemson University.

Gallagher, C. (2001). A method for fitting stable autoregressive models using the autocovariation function. *Statistics and Probability Letters* **53**, 381–390.

Gnedenko, B. V. and A. N. Kolmogorov (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley. Reading, MA.

Gopikrishnan, P., M. Meyer, L. A. N. Amaral and H. E. Stanley (1998). Inverse cubic law for the probability distribution of stock price variations. *European Journal of Physics B: Rapid Communications* **3**, 139–140.

Gopikrishnan, P., V. Plerou, L. A. N. Amaral, M. Meyer and H. E. Stanley (1999). Scaling of the distributions of fluctuations of financial market indices. *Physical Review E* **60**, 5305–5316.

Goryainova, E. P. (1996). Sign estimates in moving average models with infinite variance. *Avtomatika i Telemekhanika (Automation and Remote Control)* **57**(11), 29–38.

Granger, C. W. J. and D. Orr (1972). Infinite variance and research strategy in time series analysis.. *Jounal of the American Statistical Association* **67**(338), 275–285.

Gross, S. and W. L. Steiger (1979). Least absolute deviation estimates in autoregression with infinite variance. *Journal of Applied Probability* **167**, 104–116.

Hannan, E. J. and M. Kanter (1977). Autoregressive processes with infinite variance. *Journal of Applied Probability* **14**, 411–415.

Hesse, C. H. (1990). Rates of convergence for the empirical distribution function and the empirical characteristic function of a broad class of linear process. *Journal of Multivariate Analysis* **35**, 186–202.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**, 1163–1174.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association* **94**(448), 1201–1214.

Ibragimov, I. A. and Yu. V. Linnik (1971). *Independent and Stationary Sequences of Variables Random*. Wolters-Noordhoff. Groningen.

Ihaka, R. and R. Gentleman (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**(3), 299–314.

Janicki, A. and A. Weron (1994*a*). Can one see α-stable variables and processes?. *Statistical Science* **9**(1), 109–126.

Janicki, A. and A. Weron (1994*b*). *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*. Marcel Dekker. New York.

Jian, H. and Y. Pawitan (1994). Consistent estimator for non-Gaussian non-causal autoregressive process. *Journal of Time Series Analysis* **20**, 417–123.

Kanter, M. and W. L. Steiger (1974). Regression and autoregression with infinite variance. *Advances in Applied Probability* **6**, 768–83.

Knight, J. L. and J. Yu (2002). Empirical characteristic function in time series estimation. *Econometric Theory* **18**, 691–721.

Knight, K. (1989). Consistency of Akaike's information criterion for infinite variance autoregressive processes. *Annals of Statistics* **17**(2), 824–840.

Kogon, S. M. and D. B. Williams (1998). Characteristic function based estimation of stable distribution parameters. In: *Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman and M. S. Taqqu, Eds.). pp. 311–335. Birkhäuser. Boston.

Kokoszka, P. S. and M. S. Taqqu (1994). Infinite variance stable ARMA processes. *Journal of Time Series Analysis* **15**, 203–220.

Kokoszka, P. S. and M. S. Taqqu (1995). Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications* **60**, 19–47.

Koutrouvelis, I. A. (1980*a*). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika* **67**, 238–240.

Koutrouvelis, I. A. (1980*b*). Regression type estimation of the parameters of stable laws. *Journal of the American Statistical Association* **75**, 918–928.

Krämer, W. and R. Runde (2001). Diagnostic checking in linear process with infinite variance. *Mathematical and Computer Modelling* **34**, 1123–1131.

Kuruoğlu, E. E. (2001). Density parameter estimation of skewed α-stable distributions. *IEEE Transactions on Signal Processing* **49**(10), 2192–2201.

Kuruoğlu, E.E., P. J. W. Rayner and W. J. Fitzgerald (1997). Least L^p -norm estimation of autoregressive model coefficients of Symmetric α -Stable processes. *IEEE Signal Processing Letters* **4**(7), 201–203.

Ljung, G. M. and G. E. P Box (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics Computation and Simulation* **15**, 1109–1136.

Menn, C. and S. T. Rachev (2004). A new class of probability distributions and its application to finance. Technical report. Department of Statistics and Applied Probability, University of California, Santa Barbara.

Mikosch, T., T. Gadrich, C. Klüppelberg and R. Adler (1995). Parameter estimation for ARMA models with infinite variance innovations. *Annals of Statistics* **23**, 305–326.

Mittnik, S. and S. T. Rachev (1996). Tail estimation of the stable index α. *Applied Mathematics Letters* **9**(3), 53–56.

Mittnik, S., S. T. Rachev and M. S. Paolella (1998). Stable Paretian modeling in finance: Some empirical and theoretical aspects. In: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman and M. S. Taqqu, Eds.). pp. 79–110. Birkhäuser. Boston.

Mittnik, S., S. T. Rachev, T. Doganoglu and D. Chenyao (1999). Maximum likelihood estimation of stable Paretian models. *Mathematical and Computer Modelling.* **29**(10-12), 275–293.

Modarres, R. and J. P. Nolan (1994). A method for simulating stable random vectors. *Computational Statistics* **1**(9), 11–19.

Nikias, C. L. and M. Shao (1995). *Signal Processing with α-Stable Distributions and Applications.* Wiley. New York.

Nolan, J. P. (1999). Fitting data and assessing goodness-of-fit with stable distributions. Technical report. Departement of Mathematics and Statistics, American University.

Nowicka, J. (1997). Asymptotic behavior of the covariation and the codifference for ARMA models with stable innovations. *Stochastic Models* **13**, 673–686.

Nowicka, J. and A. Weron (2001). Dependence structure of stable R-GARCH processes. *Probability and Mathematical Statistics* **21**, 371–380.

Paolella, M. S. (2001). Testing the stable Paretian assumption. *Mathematical and Computer Modelling* **34**, 1095–1112.

Pictet, O. V., M. M. Dacorogna and U. A. Müller (1998). Hill, bootstrap and jackknife estimators for heavy tails. In: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman and M. S. Taqqu, Eds.). pp. 283–310. Birkhäuser.

Plerou, V., P. Gopikrishnan, L. A. N. Amaral, M. Meyer and H. E. Stanley (1999). Scaling of the distribution of price fluctuations of individual companies. *Physical Review E* **60**, 6519–6529.

Press, S. J. (1972). Estimation in univariate and multivariate stable distribution. *Journal of the American Statistical Association* **67**, 842–846.

Proakis, J. G. and D. G. Manolakis (1996). *Digital Signal Processing.* Prentice Hall. London.

Qiou, Z. and N. Ravishanker (1998). Bayesian inference for time series with infinite variance stable innovations. *Journal of Time Series Analysis* **19**(2), 235–249.

R Development Core Team (2004). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-00-3.

Rachev, S. T. and S. Mittnik (2000). *Stable Paretian Models in Finance.* Wiley. New York.

Resnick, S. I. (1997). Heavy tailed modeling and teletraffic data. *Annals of Statistics* **25**, 1805–1869.

Resnick, S. I. (2001). *A Probability Path.* Birkhäuser.

Resnick, S. I. and C. Starica (1997). Smoothing the Hill estimator. *Advances in Applied Probability* **29**, 271–293.

Rosenfeld (1976). Identification of time series with infinite variance. *Applied Statistics* **25**, 147–153.

Runde, R. (1997). The asymptotic null distribution of the Box-Pierce Q statistic for random variables with infinite variance - with an application to German stock returns. *Journal of Econometrics* **78**, 205 – 216.

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable NonGaussian Processes: Stochastic Models with Infinite Variance.* Chapman and Hall. New York.

Schmid, F. and M.Trede (2003). Simple tests for peakedness, fat tails and leptokurtosis based on quantiles. *Computational Statistics and Data Analysis* **43**, 1–12.

Stuck, B. W. (1978). Minimum error dispersion linear filtering of scaler symmetric stable processes. *IEEE Transactions on Automatic Control* **23**, 507–509.

Stuck, B. W. and B. Kleiner (1974). A statistical analysis of telephone noise. *The Bell Technical Journal* **53**, 1263–1320.

Thavaneswaran, A. and B. Abraham (1994). A note on model reference adaptive system (MRAS) estimate with infinite variance. *Statistica Neerlandica* **48**, 253–257.

Thavaneswaran, A. and S. M Peiris (1999). Estimation for regression with infinite variance errors. *Mathematical and Computer Modelling* **29**(10-12), 177–180.

Tsionas, E. G. (1999). Monte carlo inference in econometric models with symmetric stable disturbances. *Journal of Econometrics* **88**, 365–401.

Weron, R. (1995). Performance of the estimators of stable law parameters. Technical Report HSC/95/1. Wroclaw University of Technology.

Weron, R. (1996). On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statistics & Probability Letters* **28**, 165–171.

Weron, R. (2001). Levy-stable distributions revisited: Tail index $\alpha > 2$ does not exclude the Lévy-stable regime. *International Journal of Modern Physics C* **12**(2), 209–223.

Yamazato, M. (1978). Unimodality of infinitely divisible distributions of class. *Annals of Probability* **6**, 523–531.

Yang, Q., A. Petropulu and J. C. Pesquet (2001). Estimating long-range dependence in impulsive traffic flows. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* pp. 3413 –3416.

Yohai, V. and R. A. Maronna (1977). Asymptotic behavior of least squares estimates for autoregressive processes with infinite variances. *Annals of Statistics* **5**, 554–560.

Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews* **23**, 93–123.

Zolotarev, V. M. (1986). *One Dimensional Stable Distributions.* American Mathematical Society. Providence, RI.

CURRICULUM VITAE

Name	: Dedi Rosadi
Place and Date of Birth	: Belitung - Indonesia, December 20, 1974
Sex	: Male
Marital Status	: Married, one child

EDUCATION BACKGROUND

Elementary School 1979-1985 : SDN I Gantung, Belitung, Indonesia

Junior High School 1985-1988 : SMPN Gantung, Belitung, Indonesia

Senior High School 1988-1991 : SMAN Manggar, Belitung, Indonesia

University

- August 1991 - January 1996: Sarjana S1 (a degree in Indonesia somewhat higher than B.Sc.)
  Specialization: Statistics
  Faculty of Mathematics and Natural Sciences
  Gadjah Mada University, Yogyakarta, Indonesia

- August 1997 - June 1999: Master of Science in Engineering Mathematics
  Specialization: Stochastic and Operation Research
  Faculty of Applied Mathematics
  Twente University, The Netherlands
  Title of Thesis : *Performance analysis of link sharing policies*

- October 2001 - September 2004 :
  Ph.D. Study at Institute for Mathematical Methods in Economics, Research Unit: Econometrics and System Theory, Vienna University of Technology

WORKING EXPERIENCE

April 1996 until present: Lecturer in Department of Statistics, Gadjah Mada University, Indonesia.

December 1998 - June 1999: Research student at KPN Research, Leidschendam, The Netherlands