Goodness of Fit Tests for Non-Causal Time Series with Infinite Variance

by

Mohamad Fasih

Supervisor: Dr. Saeid Rezakhah
Advisor: Dr. Mina Amin Ghafari

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

Mathematical Statistics

(Statistics)

AMIRKABIR UNIVERSITY OF TECHNOLOGY

(Tehran)

January 2016

Abstract

This thesis examines the use of goodness-of-fit tests for non-causal autoregressive time series with stable, non-Gaussian errors. It employs a class of Portmanteau tests to assess the model's fit. These tests are conducted by analyzing the sample autocorrelations of the model's residuals, specifically focusing on a sequence of lags. Since time series with infinite variance lack second-order moments, the analysis utilizes trimmed residuals. This approach leverages the asymptotic normality of the trimmed residuals' sample autocorrelations, enabling the application of general goodness-of-fit tests even for such time series. The performance of tests was investigated by simulation and the result showed that almost all of the tests demonstrated strong performance and tend to be conservative. Furthermore the lag value had low impact on general performance of tests.
Keywords: Non-causal Autoregressive Time Series, Trimmed Residuals, Portmanteau Tests, α-Stable Distributions

Table of Contents

List of Tables

List of Figures

Chapter 1

Basic Concepts and Definitions

1.1 Introduction

Time series analysis has a wide range of applications. for instance, it can be a great aid in the field of economics. As a case in point, we could look at the water level of a lake as a time series. In addition to these examples, other phenomena that exhibit a consistent pattern over time, such as air temperature, population growth, biological growth rates, and product sales figures, can also be considered time series. In general, a series of observations of a variable measured over time is referred to as a time series. In time series analysis, firstly, a suitable model that aligns with the data must be selected. Subsequently, the parameters of the chosen model are estimated. After determining the parameter values, the adequacy of the model is assessed. This step involves determining whether the selected model accurately represents the data. In fact, one of the tasks of statisticians is to know which probability model can describe the data well. Hence the question that often arises is that which model provides a better fit to the data. For the first step, that is the specification of the model for autoregressive moving average time series models, it is common to use sample moments that include the sample autocorrelation function and the sample partial autocorrelation function. For autoregressive moving average models, specifying the model includes determining the order of the autoregressive and determining the order of the moving average of the model. In the second stage of analysis, which is estimating the specified model parameters, the maximum likelihood method can usually be used to estimate the parameters. For a well-specified model, the problem of parameter estimation is usually not very complicated. For a pure autoregressive process, there are at least two common methods for parameter estimation: Estimation of parameters using least squares method and estimation of parameters using Yule-Walker equations. And the third stage, which is the stage of checking the adequacy of the model, and in this

thesis, this problem has been investigated for autoregressive time series. These aims are investigated using goodness of fit tests and techniques such as the overfitting problem, the residuals diagram, and most importantly, whether the residuals of the fitted model are uncorrelated or not. A good model should be able to describe the data dependency structure well. An important way to measure the dependency is checked through the autocorrelation function. In other words, a good model should be able to produce residuals that are almost uncorrelated. Goodness of fit tests are performed on the residuals of the fitted model. in such a way that, the statistics are based on the autocorrelations of the residuals. For individual lags, since the autocorrelations are approximately normally distributed, the tests can be based on this fact. On the other hand, in addition to looking at the autocorrelation of residuals in individual lags, it can be useful to examine them as a group, so tests can be performed more generally for a group of correlations. These general tests are often called portmanteau tests. In this thesis, several portmanteau goodness-of-fit tests for selecting the most appropriate model for an autoregressive time series have been investigated. In the first chapter, basic concepts and definitions that are needed in the next chapters of the thesis are presented. Since identification tests for autoregressive models with non-Gaussian stable errors have been investigated, the second chapter is devoted to the introduction of stable distributions. In the third chapter, goodness-of-fit tests are presented to identify non-causal autoregressive models with stable non-Gaussian errors, and simulation studies are also performed in third chapter. In this chapter, we express the definitions and basic concepts that will be mentioned in different parts of the thesis. The contents and proofs of this chapter can be found in reference 9.

1.2 Time Series

In this chapter, we describe the basic definitions and concepts that will be mentioned in different parts of the thesis. The contents and proofs of this chapter can be found in [1].

A time series $\{X_t, t \in T\}$ is an ordered sequence of observations, usually ordered by time, especially at equal time intervals. [**?**]It is the inherent nature of time series that its observations are dependent or correlated, so the order of observations is important and statistical methods based on the assumption of independence are no longer applicable and other methods are required. Time series analysis is used in various fields [**?**]. In agriculture, the amount of the product and its annual prices, in commerce and economics, the

prices available at the end of the day, and in engineering, we see electrical signals and voltage, which are examples of time series. A time series is called discrete-time if the set T is a discrete set, and continuous-time if the set T is a continuous time interval. A stochastic process is a set of random variables that time series is its observations. In this thesis we may call time series model or even time series instead of stochastic process. The stochastic process $X_t \sim IID\left(\mu, \sigma^2\right)$ which comes from an independent and identical random variable with mean μ and variance σ^2, may be the simplest example of time series and it is called *i.i.d.* noise. According to the definition of such a sequence, for any positive integer n and real numbers $x_1, \cdots, x_n$, we have:

$$P\left(X_1 \leq x_1, \cdots X_n \leq x_n\right) = P\left(X_1 \leq x_1\right) \cdots P\left(X_n \leq x_n\right) = F\left(x_1\right) \cdots F\left(x_n\right)$$

where F is the cumulative distribution function of each of the random variables $X_1 \cdots X_n$. In this model, there is no dependency between observation so for every $h \geq 0$ and $x_1, \cdots, x_n$:

$$P\left(X_{n+h} \leq x_1 \middle| X_1 = x_1, \cdots, X_n = x_n\right) = P\left(X_{n+h} \leq x\right)$$

This shows that knowing about $X_1 \cdots X_n$ provides no information about the behavior of X_{n+h}. In fact the *i.i.d.* noise process is not predictable.

Example 1.1. The sequence of random variables $X_t, T = 1, 2, \cdots$ with $P\left(X_t = -1\right) = 1 - p$ and $P\left(X_t = 1\right) = p$ where $p = \frac{1}{2}$, is an example of pure model. The time series obtained from the sequence of tossing of coins, which assigns the number 1 to the head and -1 to the tail, is an example of this model.

1.3 Definitions

Definition 1.2. Let X_t be a time series with $EX_t^2 < \infty$. Mean and auto covariance functions of X_t are defined as follows respectively:

$$\mu_X\left(t\right) = E\left(X_t\right)$$

$$\gamma_x\left(r, s\right) = cov\left(X_r, X_s\right) = E\left[\left(X_r - \mu_X\left(r\right)\right)\left(X_s - \mu_X\left(s\right)\right)\right]$$

Definition 1.3. Time series $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is strictly stationary if for any integer h and $n > 0$, $X_1, \cdots X_n$ and $X_{1+h}, \cdots X_{n+h}$ have the same joint distributions.

Definition 1.4. Time series $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is weakly stationary if for any integer h $\mu_X t$ and $\gamma_X(t, t+h)$ be independent of t.

It is obvious that if X_t is strictly stationary and for every t $E\left(X_t^2\right) < \infty$, then X_t is weakly stationary as well and Weak stationarity does not imply strict stationarity. A sequence of independent and identical random variables is a strictly stationary process. Wherever stationary is mentioned, weak stationary is meant unless specifically stated. If X_t is a sequence of uncorrelated random variables with mean zero and variance σ^2 it is obviously stationary. This sequence is called white noise with mean zero and variance σ^2 and written as $\{X_t\} \sim WN\left(0, \sigma^2\right)$.

Definition 1.5. Process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is called moving average process with order q or $MA(q)$ if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $\{Z_t\} \sim WN\left(0, \sigma^2\right)$ and $\theta_1, \theta_2, \cdots, \theta_q$ are constants. In the above definition if Z_t is a sequence of $i.i.d.$ random variables then X_t will be strictly stationary.

Definition 1.6. Process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is called autoregressive process with order p or $AR(p)$ if

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

where $\{Z_t\} \sim WN\left(0, \sigma^2\right)$ and $\phi_1, \phi_2, \cdots, \phi_p$ are constants.

The class of linear time series models provides a wide context for investigation of stationary process.

Definition 1.7. Process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is a linear process if for every integer t

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

where $\{Z_t\} \sim WN\left(0, \sigma^2\right)$ and ψ_j is a sequence of constants that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

A linear process is called infinite moving average or $MA(\infty)$ if for every $j < 0$, $\psi_j = 0$ or

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

using this definition $AR(p)$ and $MA(q)$ are stationary and linear process.

Definition 1.8. Process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is a $ARMA(p,q)$ process, if X_t is stationary and for every t:

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

where $\{Z_t\} \sim WN\left(0, \sigma^2\right)$ and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common zeroes.

A brief form of definition (1.8) can be written as $\phi(B) X_t = \theta(B) Z_t$ where $\phi(\cdot)$, $\theta(\cdot)$ are following polynomials:

$$(1 - \phi_1 z - \cdots - \phi_p z^p), \qquad (1 + \theta_1 z + \cdots + \theta_q z^q)$$

and B is the backshift operator which is defined as follows:

$$B^j X_t = X_{t-j}, \qquad B^j Z_t = Z_{t-j}, \quad j = 0, \pm 1, \cdots$$

Time series X_t is a $AR(p)$ if $\theta(z) \equiv 1$ and is a $MA(q)$ process if $\phi(z) \equiv 1$.

Definition 1.9. Process $ARMA(p,q)$ is a causal process or a causal function of Z_t, if there are ψ_j so that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and for every t, $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$. Causality is equivalent to following condition:

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \qquad \forall |z| \leq 1$$

Definition 1.10. Process $ARMA(p,q)$ is an invertible process if there are π_j so that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and for every t, $Z_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}$. Invertibility is equivalent to following condition:

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0 \qquad \forall |z| \leq 1$$

1.4 ACVF, ACF, and PACF functions

Definition 1.11. Let X_t be a process that for every t, $E\left(X_t^2\right) < \infty$. Then the auto covariance function $(ACVF)$ $\gamma(\cdot, \cdot)$ is defined as following:

$$\gamma_x(r,s) = cov(X_r, X_s) = E\left[\left(X_r - \mu_X(r)\right)\left(X_s - \mu_X(s)\right)\right] \qquad r, s \in T$$

If X_t is a stationary process, for every $r, s \in T$ $\gamma_X(r,s) = \gamma_X(r-s, 0)$. Hence for stationary processes it is better to use only one variable for its $ACVF$:

$$\gamma_X(h) = \gamma_X(h,0) = cov(X_{t+h}, X_t) \qquad t, h \in T$$

As an example, consider $MA(1)$, $X_t = Z_t + \theta Z_{t-1}$, $t = 0, \pm 1, \pm 2, \cdots$ with $Z_t \sim WN\left(0, \sigma^2\right)$ and $\theta \in \mathbb{R}$. For this model we have:

$$E(X_t) = 0$$

$$E\left(X_t^2\right) = \sigma^2\left(1 + \theta^2\right) < \infty$$

$$\gamma_X(h) = \begin{cases} \sigma^2\left(1+\theta^2\right) & h = 0 \\ \sigma^2\theta & h = \pm 1 \\ 0 & |h| > 1 \end{cases}$$

and in general for a $MA(q)$ process we have:

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & x\,|h| < q \\ 0 & |h| > q \end{cases}$$

Definition 1.12. The auto correlation function (ACF) for X_t at lag h is given as follows:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = corr\left(X_{t+h}, X_t\right) \qquad t, h \in T$$

$\rho(0) = 1$ and $|\rho(h)| = 1$ are easily derived from the definition (1.12). For $MA(1)$ the auto correlation function is defined as :

$$\rho_X(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2} & h = \pm 1 \\ 0 & |h| > 1 \end{cases}$$

Proposition 1.13. *If $\gamma(.)$ be the autocovariance function of stationary process $\{X_t, t \in T\}$, then:*

- $\gamma(.) > 0$
- *for every $h \in \mathbb{Z}$, $|\gamma(h)| \leq |\gamma(0)|$*
- $\gamma(.)$ *is a even function, that is for every $h \in \mathbb{Z}$ $\gamma(h) = \gamma(-h)$.*

1.4.1 Estimating the ACF and $ACVF$

Let $x_1, x_2, \cdots x_n$ be a realization of time series X_t, we have the following definitions
sample mean:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

sample autocovariance function:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(x_{t+|h|} - \bar{x}\right)(x_t - \bar{x}) \qquad |h| < n$$

sample autocorrelation function:

$$\hat{\rho}(h) = \frac{\hat{\gamma(h)}}{\hat{\gamma(0)}} \qquad |h| < n$$

Sample ACF that is an estimate of the ACF of the series can provide an initial theory that helps to find which of the well-known time series models are able to fit the data. For example, if the sample autocorrelation function of a series is close to zero for non-zero lags, the pure model (*i.i.d.* noise) may be suitable for the data, or if the sample autocorrelation function of a series has a gradual decrease (exponential or geometric), the AR model may be suitable for the data.

Definition 1.14. Partial autocorrelation function for $ARMA$ process is denoted by $\alpha(\cdot)$ and defined by $\alpha(h) = \phi_h h$ and $\alpha(0) = 1$. $\phi_h h$ is the last element of vector ϕ_h that satisfies the following equation:

$$\phi_h = \Gamma_h^{-1}\gamma(h)$$

while

$$\gamma(h) = [\gamma(h)]_{n=1}^{h} \qquad \Gamma_h = [\gamma(i-j)]_{i,j=1}^{h} \qquad \phi_h = [\phi_{h_i}]_{i=1}^{h}$$

If $\hat{\rho}(h)$ is significantly different from zero for $0 \leq h \leq q$ and is about zero for $h > q$, $MA(q)$ can be an appropriate model. For large n and lags higher than q, $\hat{\rho}(h)$ has approximately normal distribution with mean zero and variance $\frac{1}{n}$ hence if $h > q$ in a $MA(q)$ model, $\hat{\rho}(h)$ lies on interval $\frac{\pm 1.96}{\sqrt{n}}$

with probability 95%. In practice when $\hat{\alpha}$ for $0 \leq h \leq q$ is significantly different from zero and for $h > q$ is near zero, $AR(p)$ would be a good model. For large n and lags higher than p, $\hat{\alpha}(h)$ has approximately normal distribution with mean zero and variance $\frac{1}{n}$ hence if $h > p$ in a $AR(p)$ model, $\hat{\alpha}(h)$ lies on interval $\frac{\pm 1.96}{\sqrt{n}}$ with probability 95%.If ACF and $PACF$ tend to zero then they determine an $ARMA$. Stationary process X_t with finite variance can be written as $MA(\infty)$ and in these processes ACF can be used to fit an $ARMA$ model to data, thus if we work with second order moments, linear models are sufficient for analyzing the data. If the variance of X_t is finite and its mean be zero when $n \to \infty$ with probability 1:

$$\hat{\gamma}(h) \to cov(X_{t+h}, X_t)$$

$$\hat{\rho}(h) \to corr(X_{t+h}, X_t)$$

it means that the sample autocovariance function and sample autocorrelation function tend to $ACVF$ and ACF of model respectively. Therefore the main steps of fitting a model to a dataset are:

- First step: estimating the p and q using ACF and $PACF$ of sample
- Second step: estimating the parameters of model
- Third step: checking the adequacy of model using residuals

There are several methods to estimate the parameters of the model. Yule-Walker, Burg, and Maximum Likelihood are the well-known of them. We briefly explain the Yule-Walker method in the following. Let X_t be a $AR(p)$ process, we tend to estimate the coefficients vector $\phi = (\phi_1, \cdots, \phi_p)'$ and the variance of the white noise using the observation $x_1, \cdots, x_n$. Yule-Walker equations are defined as:

$$\Gamma_p \phi = \gamma_p \qquad \sigma^2 = \gamma(0) - \phi' \gamma_p$$

where Γ_p is covariance matrix $[\gamma(i-j)]_{i,j=1}^{p}$ and $\gamma_p = (\gamma(1), \cdots, \gamma(p))'$. Now if for $j = 0, 1, \cdots, p$ $\gamma(j)$ is replaced by its sample estimation, Yule-Walker equations will be:

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma_p} \qquad \hat{\sigma^2} = \gamma\hat{(0)} - \hat{\phi}' \hat{\gamma_p}$$

If $\gamma\hat{(}0) > 0$, by dividing both sides of the above equation by $\gamma\hat{(}0) > 0$ we will have:

$$\hat{\phi} = \hat{R}_p^{-1}\hat{\rho}_p \qquad \hat{\sigma}^2 = \hat{\gamma}(0)\left[1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p\right]$$

where $\hat{R}_p = [\hat{\rho}(i-j)]_{i,j=1}^p$ and $\hat{\rho_p} = (\hat{\rho}(1), \cdots \hat{\rho}(p))'$ therefore the fitted model will be:

$$X_t = \hat{\phi_1}X_{t-1} + \cdots \hat{\phi_p}X_{t-p} + Z_t$$

For large enough samples from $AR(p)$:

$$\hat{\phi} \approx N\left(\phi, n^{-1}\sigma^2\Gamma_p^{-1}\right)$$

Thus, assuming that p is identified correctly, it is possible to find confidence sets for model coefficients. If $\chi^2_{1-\alpha}(p)$ be the $(1-\alpha)$th quantile of chi-square distribution with p degrees of freedom, then

$$\left\{\phi \in \mathbb{R}^p : \left(\hat{\phi}_p - \phi\right)'\hat{\Gamma_p}\left(\hat{\phi}_p - \phi\right) \leq n^{-1}\hat{\sigma}^2\chi^2_{1-\alpha}(p)\right\}$$

contains $|phi_p$ with probability about $1-\alpha$.

1.5 Adequacy of model

In statistical analysis checking the residuals of model is a way to check the adequacy of model. The similar approach is used in time series analysis and the main steps are:

- residual plot: residuals plot versus time should not include any sign of trend or changing in variance.
- sample ACF plot of the residuals: as it was mentioned earlier, for large n sample autocorrelations of sequence of independent and identical random variable $Y_1, \cdots, Y_n$ with finite variance, approximately are distributed as $N(0, \frac{1}{n})$. Therefore, we test the matching of the observed residuals with the *i.i.d.* noise by using the sample autocorrelations of the residuals as follows. If more than five percent of the partial autocorrelation of a sample is outside the region $\frac{\pm 1.96}{\sqrt{n}}$, the assumption that the residuals are *i.i.d.* will be rejected.
- randomness of residuals: this is checked using different methods such as the test based on turning point, the difference-sign test, and the rank test.

1.5.1 Yule-Walker Estimation for $ARMA$ Process

If there exist $\gamma(h)$ values, $(h = 0, 1, \cdots, p+q+1)$, by solving a system with $p+q+1$ equations

$$\gamma(h) = \phi_1\gamma(h-1) + \cdots + \phi_p\gamma(h-p) + \sigma^2 \sum_{j=h}^{q} \theta_j\psi_{j-1} \qquad h = 0, 1, \cdots, q$$

parameters $\phi = (\phi_1, \cdots, \phi_p)$, $\theta = (\theta_1, \cdots, \theta_p)$, and σ^2 can be estimated. If X_t be a autoregressive process, then these equations appear in the simplest form. In this situation, Yule-Walker estimators $\hat{\phi} = \left(\hat{\phi}_1, \cdots, \hat{\phi}_p\right)$ are obtained by solving the following equations system:

$$\hat{\Gamma}_p\hat{\phi}_p = \hat{\gamma_p}(1)$$

where $\hat{\Gamma}_p = \left[\hat{\gamma(i-j)}\right]_{i,j=1}^{n}$ and $\hat{\gamma_p}(p) = \left(\hat{\gamma}(1), \cdots, \hat{\gamma(p)}\right)$. This is the matrix form of these equations and usually known as Yule-Walker equations.

Theorem 1.15. *Let $\phi(B) X_t = Z_t$ be a causal autoregressive time series and $4Z_t$ is a independent sequence. Then yule-Walker estimators have approximately the following distribution:*

$$\hat{\phi} \sim N_p\left(\phi, \frac{\sigma^2}{n}\Gamma_p^{-1}\right)$$

Chapter 2

Alpha-Stable Random Variables

2.1 Introduction

In this chapter, we will introduce the basics of alpha-stable distributions and explore their applications. We will also discuss the mathematical properties of alpha-stable distributions and how they can be used to model real-world data. Alpha-stable distributions are a powerful tool for modeling a variety of phenomena. Their stability property makes them useful for modeling data that is subject to noise, and their heavy tails make them able to capture the outliers that are often found in real-world datasets. As a result, alpha-stable distributions are used in a wide range of applications, including finance and physics. The theory of stable distributions was first introduced in the 1920s and 1930s by Paul Levy and Alexander Yaolich Khinchin. The works of Gandenko, Kolmograf and Fellr are among the old works that investigated the properties of these distributions, and recently Zolotarev also investigated this field. In the last decades, data with heavy tails have been collected in various fields of economics, communication, physics, etc., for which the use of non-Gaussian stable processes is suggested as possible models. Such models have more flexibility and variability than Gaussian processes. Although Gaussian processes are completely characterized by the mean function and autocorrelation function, non-Gaussian stable processes require more parameters to be identified. Gaussian distribution is always symmetrical about the mean, but Stable non-Gaussian distributions can have degree of skewness. Alpha-stable distributions include a wide class of probability distributions and are known as heavy tail distributions due to the probability mass in the tail of the density function of this group of random variables. The probability mass in the tails of the distribution depends on the number *alpha*, a value between zero and two, so that the smaller α is, the greater the probability mass in the tails. Univariate alpha-stable distributions can be defined in four different ways. The first two definitions

show the property of stability. The third definition deals with the role of stable distributions in the concept of central limit theorem and the fourth definition defines the characteristic function of a stable random variable. A univariate alpha-stable distribution is characterized by four parameters. Stability index *alpha*, scale parameter σ, skewness parameter *beta* and shift parameter μ. The proof of propositions and theorems of this chapter can be found in [2].

2.2 Alpha-Stable Distribution

Random variable X has a heavy-tailed distribution F if

$$P\left(X > x\right) = x^{-\alpha} L\left(x\right) \tag{2.1}$$

where L is a slowly varying function that is for $x > 0$

$$\lim_{t\to\infty} \frac{L\left(tx\right)}{L\left(t\right)} = 1$$

note that when $x > 0$, $\beta > \alpha \qquad E\left(X^{\beta}\right) = \infty$ and $\beta < \alpha \qquad E\left(X^{\beta}\right) < \infty$

Definition 2.1. a random variable X is said to have a stable distribution if for any positive numbers A and B, there is positive number C and real number D such that

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

where X_1 and X_2 are independent copies of X, and where $\overset{d}{=}$ denotes equality in distribution.

Note that the random variable X concentrated at one point is always stable. This degenerated case is of no special interest and, unless stated explicitly, we always assume that X is non-degenerated. A random variable X is called strictly stable if (2.2) holds with $D = 0$. A random variable X is called symmetric stable if its distribution is symmetric, that is if X and $-X$ have the same distribution. a symmetric stable random variable is obviously strictly stable.

Theorem 2.2. *For any stable random variable X, there is a number $\alpha \in (0, 2]$ such that the number C in (2.2) satisfies*

$$C^{\alpha} = A^{\alpha} + B^{\alpha}. \tag{2.3}$$

A stable random variable X with index α is called α stable.

Example 2.3. If X is a Gaussian random variable with mean μ and variance σ^2, $X \sim N\left(\mu, \sigma^2\right)$, then X is a stable random variable with $\alpha = 2$ because

$$AX_1 + BX_2 \sim N\left((A+B)\,\mu, \left(A^2+B^2\right)\sigma^2\right)$$

, and (2.2) holds with $C = \left(A^2+B^2\right)^{\frac{1}{2}}$ and $D = (A+B-C)\,\mu$.

The following definition are equivalent to definition (2.2)

Definition 2.4. 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

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} C_n X + D_n \tag{2.4}$$

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

Definition 2.5. A random variable X is said to have a stable distribution if it has a domain of attraction, i.e, if there is a sequence of *i.i.d.* random variables $Y_1, Y_2, \cdots$ and sequence of positive numbers $\{d_n\}$ and real numbers $\{a_n\}$, such that

$$\frac{Y_1 + Y_2 + \cdots Y_n}{d_n} + a_n \xrightarrow{d} X \tag{2.5}$$

The notation $\xrightarrow{d}$ denotes convergence in distribution.

Definition 2.6. 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 real μ such that its characteristic function has the following form

$$E\left(e^{itX}\right) = \begin{cases} exp\{-\sigma^\alpha |t|^\alpha \left(1 - i\beta(sign\,(t)) \tan\frac{\pi\alpha}{2}\right) + i\mu t\} & \alpha \neq 1 \\ exp\{-\sigma\,|t|\,(1 + i\beta\frac{2}{\pi}(sign\,(t)))\ln|t| + i\mu t\} & \alpha = 1 \end{cases} \tag{2.6}$$

The parameter α is the index of stability and

$$sign\,(t) = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{cases}$$

The parameters σ, β, and μ are unique.

Since (2.6) is characterized by four parameters $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\sigma \geq 0$, and $\mu \in \mathbb{R}$, stable distributions are denoted by $S_\alpha\left(\sigma, \beta, \mu\right)$. $X \sim S_\alpha\left(\sigma, \beta, \mu\right)$ is symmetric if and only if $\beta = \mu = 0$ and is written by $X \sim S\alpha S$ and it is symmetric around μ if and only if $\beta = 0$. Distribution $S_\alpha\left(\sigma, \beta, \mu\right)$ is said to be right-skewed if $\beta > 0$ and left-skewed if $\beta < 0$. The probability density function of alpha-stable random variables exist and are continuous but, with a few exceptions, they are not known in closed form.

2.3 Properties of stable random variables

A useful tool for studying α-stable distribution, is the characteristic function that can be used to derive some basic properties of stable random variables.
Let X_1 and X_2 be independent random variable with $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i)$, $i = 1, 2$. Then $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$, with

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{\frac{1}{\alpha}}, \beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \mu = \mu_1 + \mu_2$$

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ and let a be a real constant. Then $X+a \sim S_\alpha(\sigma, \beta, \mu + a)$.
Let $X \sim S_\alpha(\sigma, \beta, \mu)$ and let a be a non-zero real constant. Then

$$aX \sim S_\alpha(|a|\sigma, sign(a)\beta, a\mu) \qquad \alpha \neq 1$$

$$aX \sim S_1\left(|a|\sigma, sign(a)\beta, a\mu - \frac{2}{\pi}a\ln(|a|)\sigma\beta\right) \qquad \alpha = 1$$

Proof. Using (2.6) we have for $\alpha \neq 1$,

$$\begin{aligned}\ln\{E\exp i\theta(aX)\} &= -|\theta a|^\alpha \sigma^\alpha(1 - i\beta sign(a\theta)\tan\frac{\pi\alpha}{2}) + i\mu(\theta a)\\ &= -(\sigma|a|)^\alpha|\theta|^\alpha(1 - i\beta sign(a)sign(\theta)\tan\frac{\pi\alpha}{2}) + i(\mu a)\theta\end{aligned}$$

The proof is similar for $\alpha = 1$ □

Random variables that are totally skewed to the right can be regarded as basic building blocks because of the followings:
Let X have distribution $S_\alpha(\sigma, \beta, 0)$ with $\alpha < 2$. Then there exist two $i \cdot i \cdot d \cdot$ random variables Y_1 and Y_2 with common distribution $S_\alpha(\sigma, 1, 0)$ such that

$$X \stackrel{d}{=} (\frac{1+\beta}{2})^{\frac{1}{\alpha}}Y_1 - (\frac{1-\beta}{2})^{\frac{1}{\alpha}}Y_2 \qquad \alpha \neq 1$$

$$X \stackrel{d}{=} (\frac{1+\beta}{2})Y_1 - (\frac{1-\beta}{2})Y_2 + \sigma(\frac{1+\beta}{\pi}\ln\frac{1+\beta}{2} - \frac{1-\beta}{\pi}\ln\frac{1-\beta}{2}) \quad \alpha = 1$$

This is a direct consequence of previous properties.
Following properties concern the asymptotic behavior of the tail probabilities $p\{X > \lambda\}$ and $p\{X < -\lambda\}$ as $\lambda \to \infty$. In the Gaussian case $\alpha = 2$

$$P(X < -\lambda) = P(X > \lambda) \sim (\frac{1}{2\sqrt{\pi}\sigma\lambda})e^{\frac{-\lambda^2}{2\sigma^2}}$$

as $\lambda \to \infty$. When $\alpha < 2$, however, the tail probabilities behave like $\lambda^{-\alpha}$. (If a_λ and b_λ are real numbers, we use the notation $a_\lambda \sim b_\lambda$ to denote $\lim_{\lambda\to\infty} \frac{a_\lambda}{b_\lambda} = 1$.
Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then

$$\begin{cases} \lim\limits_{\lambda\to\infty} \lambda^\alpha P(X > \lambda) = c_\alpha \frac{1+\beta}{2}\sigma^\alpha \\ \lim\limits_{\lambda\to\infty} \lambda^\alpha P(X > -\lambda) = c_\alpha \frac{1-\beta}{2}\sigma^\alpha \end{cases}$$

where

$$c_\alpha = \left(\int_0^\infty x^{-\alpha}\sin x dx\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\frac{\pi\alpha}{2})} & \alpha \neq 1 \\ \frac{2}{\pi} & \alpha = 1 \end{cases}$$

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then

$$E|X|^p < \infty \qquad \forall \quad 0 < p < \alpha$$

$$E|X|^p = \infty \qquad \forall \quad p \geq \alpha$$

This fact that α-stable random variables with $\alpha < 2$ have an infinite second order moment means that many of the techniques valid for the Gaussian case do not apply.

2.4 series representation of stable random variables

Let $N(t)$ represent the number of costumer arrival in the time interval $[0, t]$. The process $\{N(t), t \geq 0\}$ is a Poisson process with rate λ if the interval times $\tau_{i+1} - \tau_i, i \geq 1$ are independent and exponentially distributed with mean $\frac{1}{\lambda}$, in the case, $EN(t) = \lambda t$, i.e., the rate λ equal the mean number of arrivals per unit time.

Proposition 2.7. *Let $\{\tau_i\}$ denote the arrival times of a Poisson process with rate 1 and let $\{\mathbb{R}_i\}$ be $i\cdot i\cdot d\cdot$ random variables, independent of sequence $\{\tau_i\}$. If the series*

$$\sum_{i=1}^{\infty} \tau_i^{-\frac{1}{\alpha}} R_i$$

converges almost surely (with probability 1), then it converges to a strictly α-stable random variable.

2.5 Multivariate stable distributions

A random vector is said to be Gaussian if and only if every linear combination of its elements is a Gaussian random variable. Similar to Gaussian case, if $\boldsymbol{x} = (X_1, X_2, \cdots, X_d)$ is a stable random vector then every linear combination of its elements is a stable random variable as well but, the inverse case is not true. as well as the univariate case, cumulative distribution function and density function of the stable random vectors do not have a explicit form and the characteristic functions of them are used.

2.6 Stable random vectors

Definition 2.8. A random vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_d)$ is said to be a stable random vector in $\mathbb{R}^d$ if for any positive number A, and B there is a positive number C and a vector $\boldsymbol{D} \in \mathbb{R}^d$ such that $A\boldsymbol{X}^{(1)} + B\boldsymbol{X}^{(2)} \stackrel{d}{=} C\boldsymbol{X} + \boldsymbol{D}$

where $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent copies of $\boldsymbol{X}$. The vector $\boldsymbol{X}$ is called strictly stable if (2.8) holds with $\boldsymbol{D} = 0$ for any $A > 0$, $B > 0$. the vector $\boldsymbol{X}$ is called symmetric stable if it is stable and satisfies in addition the relation

$$P(X \in A) = P(-X \in A)$$

for any Borel set A of $\mathbb{R}^d$. As in $\mathbb{R}^1$ a symmetric stable vector is strictly stable.

Theorem 2.9. *Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_d)$ be a stable random vector in $\mathbb{R}^d$*

1. *There is a constant $\alpha \in (0, 2]$ such that, in (2.8) $C = (A^\alpha + B^\alpha)^{\frac{1}{\alpha}}$. Moreover any linear combination of the components of $\boldsymbol{X}$ of the type $Y = \sum_{k=1}^{d} b_k X_k$ is an α stable.*

2. *If all linear combinations $Y = \sum_{k=1}^{d} b_k X_k$, have strictly stable distribution, then $\boldsymbol{X}$ is a strictly stable random vector in $\mathbb{R}^d$.*

3. *If all linear combinations are symmetric stable, then $\boldsymbol{X}$ is a symmetric stable random vector in $\mathbb{R}^d$.*

a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ is called α-stable if (2.8) holds with $C = (A^\alpha + B^\alpha)^{\frac{1}{\alpha}}$.

Theorem 2.10. *1. If all linear combinations* $Y = \sum_{k=1}^{d} b_k X_k$ *have strictly stable distribution, then* $\boldsymbol{X}$ *is strictly stable random vector in* $\mathbb{R}^d$.

2. *If all linear combinations are symmetric stable, then* $\boldsymbol{X}$ *is a symmetric random vector in* $\mathbb{R}^d$.

3. *If all linear combinations are stable with index of stability greater or equal to one, then* $\boldsymbol{X}$ *is a stable random vector in* $\mathbb{R}^d$.

2.7 Stable stochastic process

Random processes are a crucial part of statistics and have a wide range of applications. They are essential for understanding and modeling random phenomena, and have played a major role in developing modern probability theory. Statistical inference and hypothesis testing are closely related to random processes, and are used in a variety of fields, including industry, technology, agriculture, genetics, and other sciences.

Definition 2.11. A stochastic process $\{X_t, t \in T\}$ is stable if all its finite-dimensional distributions are stable. Moreover it is strictly stable or symmetric stable, if all its finite-dimensional distributions are respectively, strictly stable or symmetric stable. note that the finite dimensional distributions of $\{X_t, t \in T\}$ are the distribution of the vectors

$$(X(t_1), \cdots, X(t_d)) \qquad t_1, \cdots, t_d \in T, \quad d \geq 1$$

Theorem 2.12. *Let* $\{X_t, t \in T\}$ *be a stochastic process.*

1. $\{X_t, t \in T\}$ *is strictly stable if and only if all linear combinations* $b_1, b_2 \cdot \cdot \cdot, b_n \in \mathbb{R} \quad t_1, \cdots, t_d \in T \quad d \geq 1$ *are strictly stable.*

2. $\{X_t, t \in T\}$ *is symmetric stable if and only if all linear combinations, are symmetric stable.*

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

2.8 Data stability evaluation methods

For a data set, it is not possible to completely determine whether the data come from a stable distribution. The best we can do is to determine

whether a data set is consistent with the assumptions of a stable distribution. Therefore, it is necessary to provide methods to evaluate the data in order to determine their stability. In regression analysis, outliers can significantly impact the validity of test results, while deviations from normality near the center of the distribution are less crucial. While all non-Gaussian stable distributions exhibit heavy tails, not all heavy-tailed distributions are stable. Therefore, focusing solely on the tails might not provide a comprehensive assessment of the distribution's overall fit. For $0 < \alpha < 2$ and $x_0 > 0$ consider

$$g(x) = g(x\,|\alpha, x_0\,) = \begin{cases} c_1 e^{-x^2/2} & |x| \le x_0 \\ c_2\,|x|^{-(1+\alpha)} & |x| \ge x_0 \end{cases}$$

where

$$c_1 = c_1(\alpha, x_0) = [\sqrt{2\pi}(2\Phi(x_0) - 1) + (2/\alpha)x_0 \exp(-x_0^2/2)]^{-1}$$

and

$$c_2 = c_1 \exp(-x_0^2/2) x_0^{1+\alpha}$$

. A random variable X with probability density function g, for $-x_0 \le x \le x_0$ has a normal density with Pareto tail. For the normal part

$$p = P(|X| \le x_0) = c_1\sqrt{2\pi}(2\Phi(x_0) - 1)$$

and probability in tails is $(1 - p)$. For every finite value of x_0, this random variable has infinite variance and is in domain of attraction of a symmetric stable random variable with index α. Another property of non-Gaussian stable distributions is having infinite variance. According to the law of large numbers, when the variance is finite, the sample variance will approach its true value as the number of observations increases. However, for distributions with infinite variance, the sample variance will diverge, making it unsuitable for statistical analysis. So, in this case, the sample variance test can be used.This test is also not robust and can be distributions with infinite variance which are not stable, such as the Student's t distribution with two degrees of freedom. The sample variance test is as follows:

$$s_n^2 = \frac{1}{n}\sum_{t=0}^{n}(x_t - \bar{x}_n)^2$$

If the observations are from a distribution with finite variance, then as n increases, the sample variance should converge. Otherwise, it will be divergent [1]. But in general, there is no valid test to evaluate the stability of a data series.

2.9 PDF and CDF of stable random variable

In this section, the distribution functions and the density function of stable distributions for different parameters of stability α and skewness parameter β are given. Note that when $\alpha = 2$ and $\beta = 0$ we have Normal distribution. the distribution will be Cauchy distribution when $\alpha = 1$ and $\beta = 0$ and if $\alpha = \frac{1}{2}$ and $\beta = 1$ it will be Levy distribution. In these three cases the density functions have explicit form. Figure (2.9) (i) is the density functions of symmetric stable random variables with different α values and (ii) is the distribution functions of symmetric stable random variables whit different α values. Figure (2.9) (i) depicts the density functions for some asymmetric stable random variables when β varies and (ii) shows distribution functions of asymmetric stable random variables for different values of β.

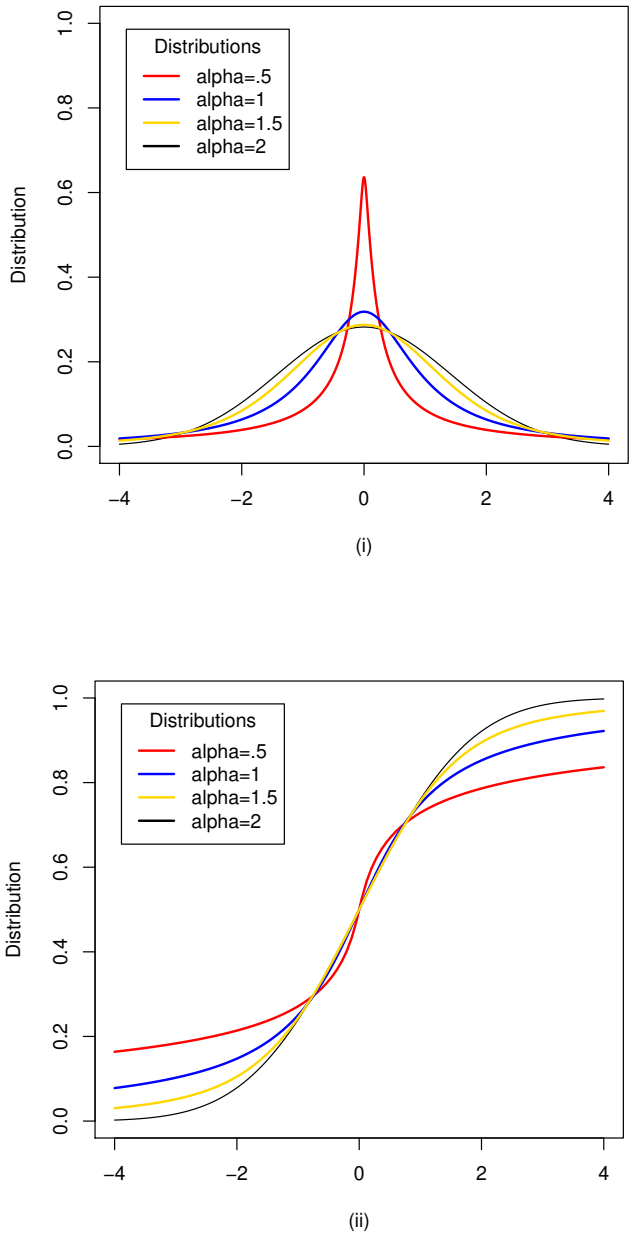

Figure 2.1: PDF and CDF for symmetric stable random variables (α varies)

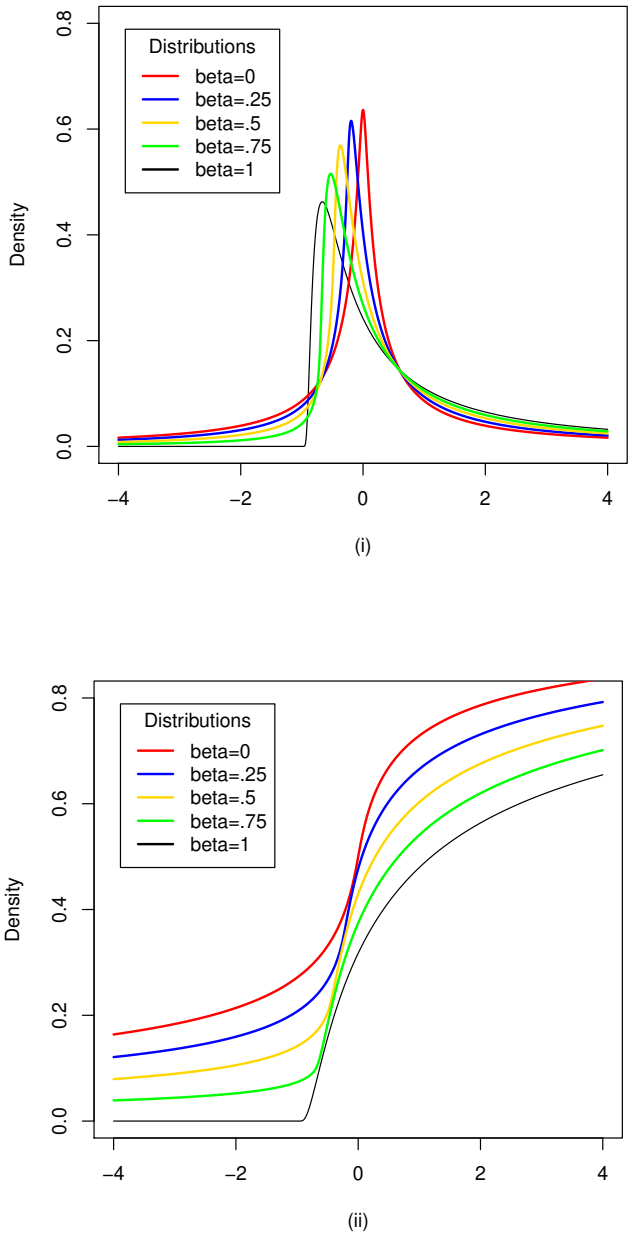

Figure 2.2: PDF and CDF for asymmetric stable random variables (β varies)

Table 2.1: Tail probability for the symmetric stable distribution

		α					
		0.6	0.9	1.2	1.5	1.8	2
x	2	0.1981578	0.1586026	0.1282269	0.1050393	0.0877029	0.0786496
	5	0.1268689	0.0755530	0.0422849	0.0206686	0.0066480	0.0203476
	10	0.0878358	0.0414320	0.0179674	0.0066393	0.0015474	0.07687299
	20	0.0598129	0.0224355	0.0077185	0.0022701	0.0004239	0.01044244

2.10 Tail probabilities

Unlike the normal distribution, whose tails decay exponentially, alpha-stable distributions exhibit varying degrees of heavy-tailedness, meaning the probability of extreme events (very large or small values) is higher compared to the normal distribution. This characteristic makes them particularly useful in modeling phenomena where extreme events hold significant importance, such as financial markets, natural disasters, and network traffic analysis. Understanding the behavior of the tails in alpha-stable distributions is crucial, as it helps quantify the likelihood of observing extreme values. The tail probability, denoted by $P(X > x)$ for the right tail and $P(X < x)$ for the left tail, represents the probability of the random variable X exceeding (or falling below) a specific value x. However, unlike the normal distribution where the tail probabilities can be expressed in terms of the standard normal cumulative distribution function, the closed-form expression for alpha-stable tail probabilities is generally unavailable. This necessitates the use of numerical methods or specialized software for calculating tail probabilities. Table 2.1 and 2.2 summarize the tail behavior for different values of the alpha parameter for symmetric and asymmetric α-stable distribution respectively. The case with $\alpha = 2$ is when the sample has come from a normal distribution.

Table 2.2: Tail probability for the asymmetric stable distribution

		α				
		0.6	0.9	1.2	1.5	1.8
x	2	0.3629263	0.2879839	0.2245698	0.1682557	0.11493
	5	0.2421764	0.1484117	0.08452	0.0413327	0.0128282
	10	0.1681291	0.0811727	0.035464	0.012955	0.002911
	20	0.113513	0.0430539	0.0147659	0.0042814	0.0007806

2.11 Infinite variance

For the non-Gaussian α-stable distribution the second order moments do not exist and when $\alpha < 1$ even mean of the distribution does not exist. In this section we try to show infinite variance using simulation. For normal distribution, as the sample size is increased and calculate the sample variance, it can be observed a fascinating phenomenon. The sample variance typically gets closer and closer to the true population variance. This convergence behavior is captured by the Law of Large Numbers, a fundamental principle in probability theory, which states that the average of a sufficiently large number of independent and identically distributed random variables tends to approach the expected value. In simpler terms, with enough data from the normal distribution, the sample variance provides a reliable estimate of the true variance. This situation is different for the non-Gaussian α-stable distribution and there is a trend towards larger and larger variances as the sample size grows. Figure 2.3 shows the typical trajectory of sample variance for standard normal distribution and a symmetric α-stable distribution (Cauchy distribution) respectively. Top plot in Figure 2.3 shows convergence of variance for standard normal distribution to the true value, while the bottom plot does not show any signs of convergence and includes jumps for an α stable distribution with $\alpha = 1$ which accumulate faster than the decay, diverging to infinity.

2.12 Maximum likelihood estimation

Due to the complex nature of α-stable distributions and the fact that the rth moment for $\alpha > r$ does not exist, estimating their parameters using traditional methods like moments is not feasible. Here, maximum likelihood estimation (MLE) emerges as a powerful alternative. MLE aims to

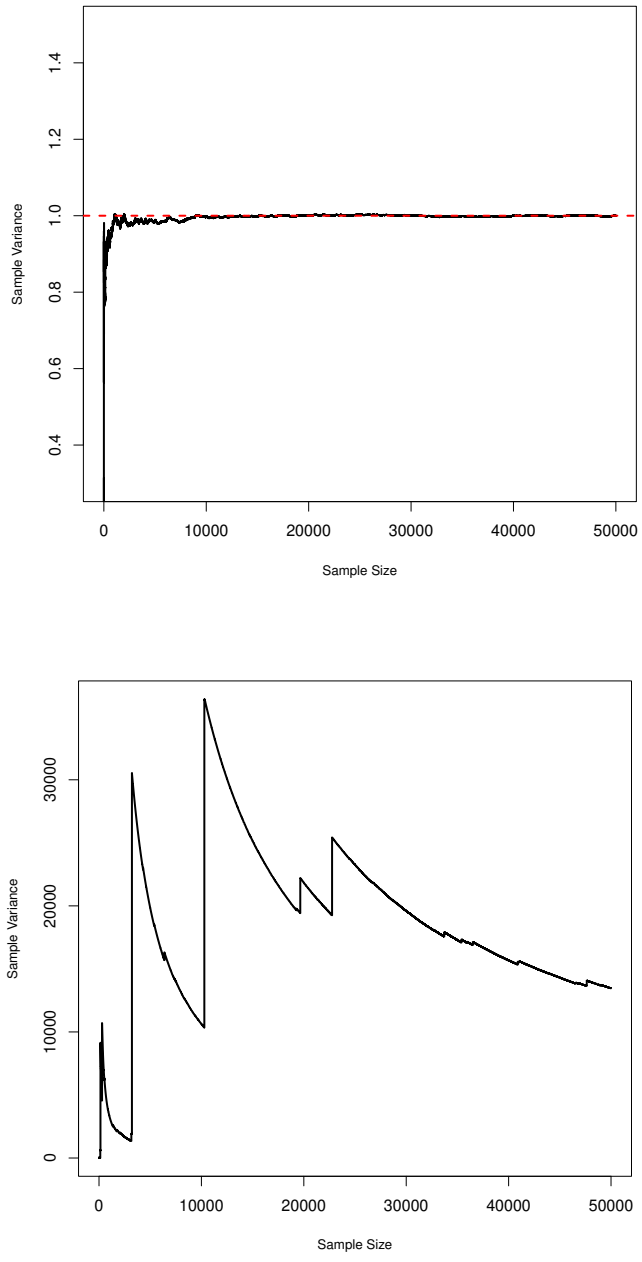

Figure 2.3: Sample variance for $N(0,1)$ (top) and $S_1(0,1,0)$ (bottom)

Table 2.3: Maximum likelihood estimation for α ($\beta = 0, 0.8$)

			α		
	0.6	0.9	1.2	1.5	1.8
			$\beta = 0$		
$\hat{\alpha}$	0.585	0.925	1.148	1.544	1.844
			$\beta = 0.8$		
$\hat{\alpha}$	0.642	0.891	1.092	1.456	1.845

find the set of parameter values that maximizes the likelihood function, which essentially represents the probability of observing the given data under those specific parameter values. On the other hand, as it was mentioned in previous sections, no general closed-form expression is known for the density function of α-stable distributions however; although, computational formulas exist that can be used to evaluate the density function. Suppose $X_1, \cdots, X_n$ is $i.i.d.$ sample from $S_\alpha(\sigma, \beta, \mu)$ and $\tilde{\theta} = (\alpha, \beta, \sigma, \mu)$and $f(x|\tilde{\theta})$ denote the parameter vector and density function respectively. The parameter space is $\Theta = (0, 2] \times [-1, 1] \times (0, \infty) \times (-\infty, \infty)$. The log-likelihood function for this sample is given by:

$$\ell(\tilde{\theta}) = \sum_{i=1}^{n} \ln f(X_i|\tilde{\theta}) \tag{2.7}$$

In order to maximize the log-likelihood function numerical methods should be implemented. Table 2.3 shows the maximum likelihood estimation of different values of α for symmetric ($\alpha = \{0.6, 0.9, 1.2, 1.5, 1.8\}, \beta = 0$) and asymmetric ($\alpha = \{0.6, 0.9, 1.2, 1.5, 1.8\}, \beta = 0.8$) simulated α-stable distribution.

Chapter 3

Diagnostic tests for non-causal time series with infinite variance

3.1 Introduction

Trying to fit a suitable model to a time series is the most important problem in assessing its performance. basis of these models is to consider random variables (white noise) that have independent or uncorrelated increments under the influence of the time parameter (discrete or continuous). In classical modeling, a distribution with finite variance is usually considered for the coefficients, and the selection of the model is done with emphasis on its autocorrelation function. Recently, statisticians have become increasingly interested in analyzing time series data that has a heavy-tailed marginal distribution. The behavior of these data for their analysis is significantly different from other time series. There are many such data, such as financial, economic, telecommunication data, etc. Autoregressive time series models with infinite variance have various practical applications. For example, Resnick [3] fitted such a model for the time intervals between computer network packet transmissions, Gallagher [4] studied sea surface temperature differences and fitted a symmetric stable autoregressive model and Ling [5] investigated the daily log-returns of the Hang Seng Index in the Hong Kong stock market. When modeling infinite variance autoregressive processes, non-Gaussian alpha-stable distributions (with an exponent parameter alpha greater than zero) are frequently employed to characterize the innovation process because they can accurately capture the sharp spikes observed in the data. Stable distributions are a powerful class of probability distributions that can model many time series that have characteristics such as heavy tails and skewness. For example time series observed in signal processing in electrical engineering, Stuck and Kleiner [6] and Sheng and Chen [7], stock portfolio selection, Rachev et al. [8], and Tokat and Schwartz [9] capital al-

location. Therefore, it seems that the assumption of stability of innovations in autoregressive models has a good theoretical and practical justification. In this thesis, portmanteau identification tests for autoregressive time series with infinite variance have been investigated. Although portmanteau tests have been extensively investigated for model identification, they have not been thoroughly investigated for time series models with infinite variance. Since the model selection methods are finally implemented through the implementation of identification tests to check the correctness of the model, the lack of appropriate diagnostic tools for time series models will be a serious concern. In the study of autoregressive processes, causality (all roots of autoregressive polynomials outside the unit circle) is usually assumed. But such an assumption is only required when the study is carried out within the framework of classical Gaussian processes in order to identify model parameters. In fact, for every Gaussian non-causal autoregressive process, there is an equivalent causal process, in the sense that both processes have the same mean and autocorrelation function [1]. Since the Gaussian distribution is uniquely characterized by the first and second order moments, both processes necessarily have the same probabilistic structure and thus will be indistinguishable. In cases where the series has non-Gaussian noise, the non-causal autoregressive process has a different probabilistic structure than its causal representation. In other words, for a non-Gaussian autoregressive process, the parameters of the model can be identified and the model can be configured without being limited to causality [10]. Andrews et al. [11] estimated the maximum likelihood of some processes. They showed that when fitting a model to Wal-Mart stock trading volume, a non-causal model can better describe the data means that the non-causal model is more consistent to the assumption of independence than the the causal representation [11]. Lanne et al. [12] studied the forecasting performance of AR models for US inflation and found that relaxing the causality constraint in AR models improved the forecasting results, especially for predicting changes in the direction of the inflation series. Andrews and Davis recently developed a model identification procedure for infinite variance AR processes. Their method utilizes Gaussian-based AIC to derive a consistent estimator for the AR order. Additionally, they adopt all-pass models to estimate the order of non-causality. The non-causal order estimation method has been used more widely to fit non-causal models with finite variance by Rosenblatt [13] and Breid et al. [14]. Compared to the efforts made in order to develop parameter estimation methods and model specification for α-stable non-Gaussian autoregressive processes, the identification of the model and checking the adequacy of the model have not been fully investigated. In this thesis, global

goodness-of-fit tests have been compared in order to identify non-causal autoregressive α-stable models, which is the result of research by [15]. They developed the methods of portmanteau tests to check the goodness of fit of non-causal α-stable autoregressive model that its parameters are estimated by maximum likelihood method presented by [11]. Due to the absence of second-order moments in infinite variance models, extracting meaningful information from sample autocorrelation of fitted model residuals for model identification proves challenging. To address this issue, Cui et al. [15] utilized trimmed residuals. Lee and Ng [16] introduced trimmed residuals to evaluate the goodness-of-fit of autoregressive models with infinite variance, and their findings motivated Cui et al. [15]. They showed that the sample autocorrelation of the trimmed residuals at a given lag fitted to a non-causal autoregressive process, has asymptotically normal distribution. Therefore, overall goodness-of-fit tests, which are used in the framework of Gaussian processes and are based on sample autocorrelations, are generalized to the case of infinite variance.

3.2 Infinite variance time series

Let $\{X_t\}$ be a $ARMA\,(p,q)$ process

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

where $\{Z_t\}$ is a sequence of random variable in domain of attraction of a stable random variable with stability index α in $(0,2]$ means that $\{Z_t\}$ has following characteristic function

$$E\left(e^{itX}\right) = \begin{cases} exp\{-\sigma^\alpha\,|t|^\alpha\,(1 - i\beta(sign\,(t))\tan\frac{\pi\alpha}{2}) + i\mu t\} & \alpha \neq 1 \\ exp\{-\sigma\,|t|\,(1 + i\beta\frac{2}{\pi}(sign\,(t)))\ln|t| + i\mu t\} & \alpha = 1 \end{cases}$$

This type of time series is called time series with infinite variance or in some cases stable time series [1]. The first step in fitting this type of series to a set of data is whether or not a heavy-tailed model fits the data, which was briefly discussed in Chapter 2.

3.2.1 Examples

Here are some examples of time series with infinite variance. Finally, since the main goal is to determine the correct model for an autoregressive time series with infinite variance, the given figures are also related to time

series simulated from the autoregressive model. The figures are related to the first-order autoregressive time series in the form of $X_t - 0.5X_{t-1} = Z_t$ and Z_t has $S_\alpha(0, 1, 0)$ distribution, where the value of α has increased from 0.6 to 2.

In figure (3.2.1) part (i) is a plot of a time series simulated from the first-order autoregressive model with a parameter of $\phi = 0.5$, and the value of the stability parameter for the sequence of its $\{Z_t\}$ is considered to be 0.6. In (ii) $\alpha = 1.2$. In figure (3.2.1) α is 1.8 and 2 for (i) and (ii) respectively. In (ii) we have a autoregressive time series with stable Gaussian noise.

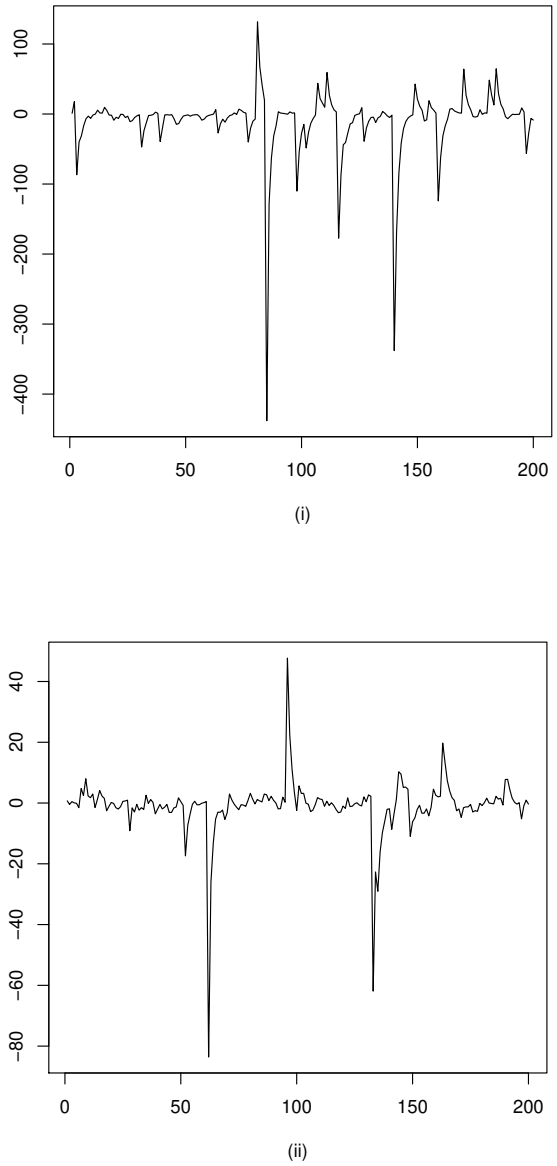

Figure 3.1: simulated $AR(1)$ model with stable noise. (i) $\alpha = 0.6$ and (ii) $\alpha = 1.2$

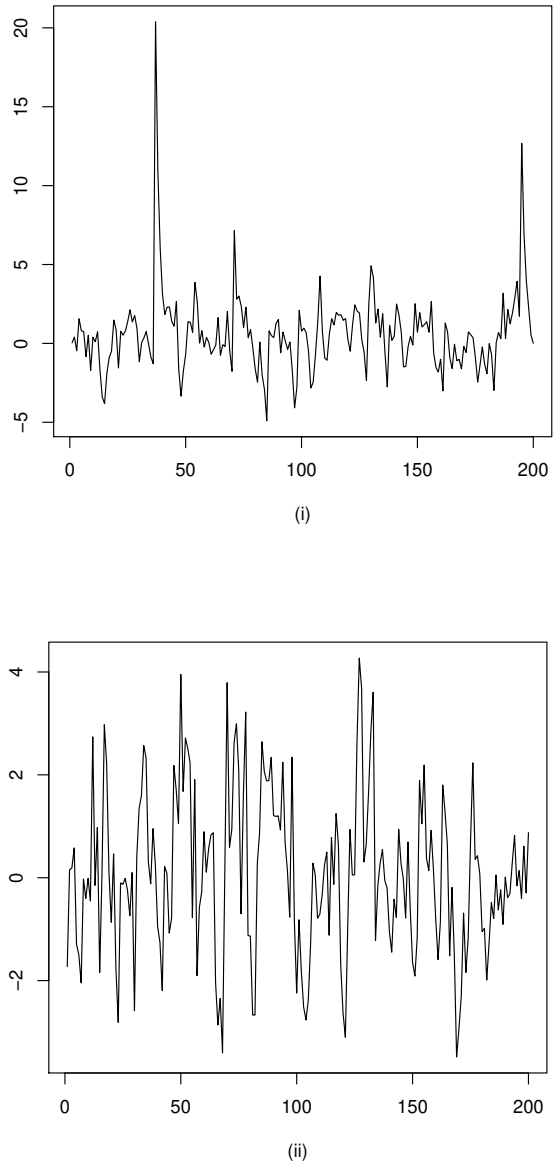

Figure 3.2: simulated $AR(1)$ model with stable noise. (i) $\alpha = 1.8$ and (ii) $\alpha = 2$

3.3 Goodness-of-fit tests

Determining the validity of the fitted model is the most important step in the analysis of a time series. In time series analysis, model selection methods usually include two steps:

1. Selecting an empirical model using model selection criteria such as AIC.

2. Checking the adequay of model

After selecting the model, the adequacy of the model should be checked. For the identification tests of ARMA models, Box and Pierce [17] proposed a test statistic based on the sum of autocorrelations of the residuals and showed that this statistic asymptotically has a chi-square distribution. Ljung and Box [18] modified the Box-Piers statistic and showed that their modified statistic is a better approximation of the chi-square distribution for a small sample. For several decades, the Living-Box test was used as a powerful test to check the adequacy of the model. Monti [19] proposed an identification test based on partial autocorrelations of the residuals. Peña and Rodríguez [20] proposed a portmanteau goodness-of-fit test based on the determinants of the autocorrelation matrix. Mahdi and Ian McLeod [21] corrected Peña-Rodrigues statistic and presented a more general statistic. For the Ljung-Box statistic, if stationary autoregressive model is specified correctly, the Ljung-Box statistic should have an almost chi-square distribution. Limit results should be used for stationary models. For example, Shin and Lee [22] showed that when the autoregressive polynomial has a unit root, the limit results are not valid and the above statistics cannot be used to check the adequacy of the model. Fisher and Gallagher [23] provided another statistic to check the adequacy of the model by including the weight for autocorrelations in Ljung-Box test statistic. McLeod and Li [24] proposed identification tests based on squared residuals, considering the usefulness of identification tests for detecting non-linear dependence. According to their analysis, the asymptotic distribution of the chi-square is independent of the autoregressive order, but it depends on the number of autocorrelations involved in the Living-Box test. In fact, the Living-Box test can determine the lack of fit considering the linear relationship, but for the non-linear relationship between the residuals, the tests based on squared residuals are a good tool to detect the lack of fit. The null hypothesis is consider to be the $ARMA$ model is identified correctly. As it was mentioned in the first chapter, for the $i.i.d.$ noise with finite variance, sample correlation $\hat{\rho}(h)$ approximately

has $N(0, \frac{1}{n})$ for large n. Hence, to find out whether the residuals can be considered as a sequence of *i.i.d.* sequence, the hypothesis test $H_0 : \rho(h) = 0$ is considered for $h = 1, \cdots, k$. To test this hypothesis, $\sqrt{n}\hat{\rho(h)}$ is used as the test statistic that has approximately normal distribution. If $\sqrt{n}\hat{\rho(h)} \geq z_{\frac{\alpha}{2}}$ or alternatively $\left|\sqrt{n}\hat{\rho(h)}\right| \geq \frac{1}{\sqrt{n}} z_{\frac{\alpha}{2}}$, then H_0 is rejected at level of α. In addition to looking at the correlation of residuals in individual lags, it is useful to have a test that considers their magnitude as a group. For example, the situation may be such that the majority Individual autocorrelations are moderate in value, some even close to their critical values, but taken together appear too large. Box and Pierce [17] suggested the statistic

$$Q_{bp} = n \sum_{k=1}^{m} \hat{\rho}_k^2 \tag{3.1}$$

to achieve this goal. They showed that if the $ARMA(p, q)$ model is correctly estimated, then for large values of n, the Q_{bp} statistic has an approximate Chi-square distribution with $m - p - q$ degrees of freedom. Fitting an incorrect model usually causes Q_{bp} to become larger. Hence, the Box-Piers statistic if the observed value If Q_{bp} is greater than a suitable critical value in the chi-square distribution with $m-p-q$ degrees of freedom, the $ARMA(p, q)$ model will be rejected. The Chi-square distribution for Q_{bp} is based on a limit theorem When $n \to \infty$, but Ljung and Box [18] showed that this limit distribution is not a good approximation for finite n. They also showed that this approximation is not satisfactory even for $n = 100$. By slightly modifying the Q_{bp} statistic, they defined a test statistic and showed that that the limiting distribution $(\chi^{2)}_{m-p-q})$ is a acceptable approximation for usual n. The modified Box-Pierce statistic or Ljung-Box statistic is

$$Q_{lb} = n(n + 2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n - k} \tag{3.2}$$

It should be noted that since for every $k \geq 0$ and $\frac{(n+2)}{(n-k)} > 1$, we have $Q_{lb} > Q$, and it explains why the Box-Pierce statistic generally ignores inadequate models. More details about the exact distributions of Q and Q_{lb} for finite samples can be found in [18]. Monti [19] suggests using the following statistics:

$$Q_{mt} = n(n + 2) \sum_{k=1}^{m} \frac{\hat{\pi}_k^2}{n - k} \tag{3.3}$$

where $\hat{\pi_k}$ is the sample partial autocorrelation at lag h and if the fitted model is a true model and residuals are a realization of white noise, then it has approximately normal distribution with mean zero and variance $\frac{(n-h)}{n(n+2)}$. If the fitted model of order p is a true model, Q_{mt} has approximately $|chi^2_{m-p}$ distribution. Using simulation, Monty showed that if the fitted $ARMA$ model has a lower moving average order than the true model, Monty's statistic detects this deficiency better compare to Ljung-Box statistic. In other words, in this case, the Monty test is more powerful than the Ljung-Box test. Other goodness-of-fit tests have been proposed to check the adequacy of the autoregressive models. Peña and Rodríguez [20] proposed a statistic based on the autocorrelation matrix, which is as follows

$$\hat{D}_m = n(1 - \left|\hat{\mathbf{R}}_m\right|^{1/m}) \tag{3.4}$$

where

$$\hat{\mathbf{R}}_m = \begin{bmatrix} 1 & \hat{\rho_1} & \cdots & \hat{\rho_m} \\ \hat{\rho_1} & 1 & \cdots & \hat{\rho_{m-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho_m} & \hat{\rho_{m-1}} & \cdots & 1 \end{bmatrix} \tag{3.5}$$

is the sample autocorrelation matrix for residuals. Peña and Rodrigues showed that if the $ARMA(p,q)$ model is correctly specified, the $\hat{D_m}$ for large values of m has an approximately gamma distribution with parameters α and β where

$$\alpha = \frac{3m[(m+1)-2p]^2}{2[2(m+1)(2m+1)-12mp]} \tag{3.6}$$

$$\beta = \frac{3m[(m+1)-2p]}{[2(m+1)(2m+1)-12mp]} \tag{3.7}$$

Mahdi-McLeod statistic, which is the modified form of $\hat{D_m}$ statistic, is as follows

$$Q_{mm} = \frac{-3n}{2m+1} \log \left|\hat{\mathbf{R}}_m\right| \tag{3.8}$$

which has a Chi-square distribution with

$$\frac{3}{2}\frac{m(m+1)}{2m+1}$$

degrees of freedom [21]. Recently, Fisher and Gallagher [23] presented the weighted statistic of Ljung and Box to check the adequacy of the pth order autoregressive model. The Fisher-Gallagher statistic is as follows

$$Q_{_{fg}} = n(n+2)\sum_{k=1}^{m}\frac{(m-k+1)\hat{\rho}_k^2}{m(n-k)} \tag{3.9}$$

which has the Gamma distribution with parameters

$$\alpha = \frac{3m(m+1)}{8m+4} \tag{3.10}$$

$$\beta = \frac{2(2m+1)}{3m} \tag{3.11}$$

It should be noted that the autocorrelation of the sample at lag 1,$\hat{\rho_1}$ is given by the weight $\frac{m}{m} = 1$. By assigning these weights, the highest and lowest emphasis is associated to the first and mth order autocorrelations, respectively. To detect nonlinear dependency between the residuals, McLeod and Li [24] developed tests based on the squared residuals. They demonstrated that the standardized autocorrelations of the squared residuals are asymptotically normally distributed. a_t be the squared sequence of $i.i.d.$ noise, the autocorrelation of a_t is defined as

$$\hat{r}_k = \frac{\sum\limits_{t-k+1}^{n}(\hat{a}_t - \hat{\sigma}^2)(\hat{a}_{t-k} - \hat{\sigma}^2)}{\sum\limits_{t=1}^{n}(\hat{a}_t - \hat{\sigma}^2)^2} \tag{3.12}$$

where

$$\hat{\sigma}^2 = \frac{\sum\limits_{t=1}^{n}\hat{a}_t}{n} \tag{3.13}$$

for positive integer m the autocorrelation of squared residuals that is defined as

$$\sqrt{n}\hat{r}(m) = \sqrt{n}(\hat{r}_1, \cdots, \hat{r}_2) \tag{3.14}$$

asymptotically distributed as $N(\mathbf{0}, \mathbf{I_m})$. In Ljung-Box, Monti, Mehdi-McLeod and Fisher-Gallagher statistics, autocorrelation of the squared residuals can be used and is denoted by Q_{slb}, Q_{smt}, Q_{smm} and g Q_{sfg} respectively.

3.4 causality and non-causality in autoregressive model

In first chapter causality for $ARMA$ model was defined. If X_t is a autoregressive time series, it is causal if there exist ψ_j with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ so that $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$.

The stationary solution of the equation $\phi(B)X_t = Z_t$ exists if the roots of the autoregressive polynomial are not on the unit circle. The meaning of the unit circle is the area defined by the complex numbers z such that $|z| = 1$ or $1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| = 1$, $z \in \mathbf{C}$ where $\mathbf{C}$ is the set of complex number (The complex number z is used here because the zeros of a polynomial of degree $p > 1$ can be real or complex).

The solution of $\phi(B)X_t = Z_t$ is causal if $1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \leq 1$ that means autoregressive polynomial has no zero inside of the unit circle. Since in this case X_t can be written as a function of past and present values of error terms $Z_t, Z_{t-1}, \cdots$, it is called causal function of Z_t. X_t is said to be non-causal if all the zeros of autoregressive polynomial are inside of the unit circle means that

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad |z| \geq 1$$

in this case X_t is a function of future values $Z_{t+1}, Z_{t+2}, \cdots$. and finally the zeros of autoregressive polynomial can be inside and outside of the unit circle hence X_t can be a function of both past and future values of error terms. Now if X_t is a non-causal autoregressive time series with identical and independent sequence Z_t with mean zero and variance σ^2, then X_t can be represented as a causal autoregressive with sequence $\tilde{Z}_t$ with mean zero and variance $\tilde{\sigma}^2$ [1]. The structure based on second order moments such as autocovariance function remain the same in both representation.Therefore, any estimation method based on second-order moment characteristics, such as Yule-Walker estimations, does not differ for causal and non-causal models. The indiscernibility of causal and non-causal models also appears in Gaussian maximum likelihood estimations [10]. Classically, Gaussian likelihood is used in time series analysis to estimate parameters for causal models [13]. Because in the Gaussian case, the probability structure of X_t is completely determined by its autocovariance function, causal and non-causal models are indistinguishable, and it is common that the parameter space is restricted only to the causal region. On the other hand, for non-Gaussian Z_t, causal and non-causal models can be distinguished using the likelihood function.

Breid et al. [14] presented a method for the maximum likelihood estimation of non-causal $AR(p)$ model parameters with independent and identical sequence Z_t with zero mean and variance σ^2 and probability density function f_σ (σ is scale parameter) and showed that the they are asymptotically normal. In this thesis, it is assumed that the model is a combination of causal and non-causal model, that is, the autoregressive polynomial has its zeros inside and outside the unit circle. Suppose the autoregressive polynomial can be written as a product of two causal and non-causal polynomials. Suppose the number of zeros of $\phi(z)$ that are outside of the unit circle is r and the number of them that are inside of the unit circle is s so that $p = r + s$ where p is the order of autoregressive and s is the order of non-causality. Therefore

$$\phi(z) = (1 - \theta_1 z - \cdots - \theta_r z^r)(1 - \theta_{r+1} z - \cdots - \theta_{r+s} z^s) \tag{3.15}$$

where $\phi(z) = (1 - \theta_1 z - \cdots - \theta_r z^r) \neq 0$ for $|z| \leq 1$ and $(1 - \theta_{r+1} - \cdots - \theta_{r+s} z^s) \neq 0$ for $|z| \geq 1$ if we define

$$(1 - \theta_1 z - \cdots - \theta_r z^r) = \phi^+(z)$$

and

$$(1 - \theta_{r+1} - \cdots - \theta_{r+s} z^s) = \phi^*(z)$$

then

$$\phi(z) = \phi^+(z)\phi^*(z)$$

where $\phi^+(z) = \prod_{i=1}^{r} (1 - m_i^{-1} z)$ and $\phi^*(z) = \prod_{i=r+1}^{p} (1 - m_i^{-1} z)$ and $m_1, \cdots, m_r, m_{r+1}, \cdots, m_p$ are the zeros of polynomial $\phi(z)$ that $|m_i| > 1$ for $i = 1, \cdots, r$ and $|m_i| < 1$ for$i = r+, \cdots, p$. For the non-causal autoregressive time series with infinite variance, Andrews et al. [11] presented the maximum likelihood estimation considering the order of non-causality order s where $s \in \{0, 1, \cdots, p\}$.

3.5 Generalization of tests for time series with infinite variance

In section (3.3) goodness-of-fit test statistics presented to check the adequacy of model. These statistics were for the case that in the autoregressive model, the White noise sequence has finite variance. In this section, these

statistics are investigated for the case where the error terms have infinite variance. As mentioned earlier, it is possible to determine whether an autoregressive model is causal or non-causal in the case that the sequence Z_t come from a stable distribution with $\alpha \in (, 2]$. Since the sequence of noise has infinite variance, the inferences based on the second moments are not valid. Lee and Ng [16] used the trimmed residuals to perform the Ljung-Box test for the autoregressive process with infinite variance. The trimming is in such a way that the estimated residuals are considered bounded. It should be noted that the trimmed or bounded random variables no longer have infinite variance, that is, the trimming methods reduce the state of infinite variance to the state of finite variance [15]. It's true that trimming methods can discard some information, but this potential drawback can be overcome by setting appropriate boundaries. Now a theorem from [15] is presented that enables us to use all of the test statistics mentioned in (3.3) for the case where the variance of the noise sequence is infinite. Suppose the $AR(p)$ model is fitted to a time series under the following condition

1. Z_t is an independent and identical sequence of random variable from a non-Gaussian stable distribution.

2. The parameters of $AR(p)$ model is estimated using maximum likelihood method.

Let $\left\{\hat{Z}_t\right\}$ be the residuals of fitted model

$$\hat{Z}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \cdots - \hat{\phi}_p Y_{t-p} \tag{3.16}$$

and $Z_{(1)} \leq \cdots \leq Z_{(n)}$ are the order statistics for the residuals of the fitted model. Let for some predetermined lower percentile λ^L and upper percentile λ^U, $\hat{M}_{nL}$ and $\hat{M}_{nU}$ be the $(n\lambda^L)$-th and the $(n\lambda^U)$-th order statistics of $\left\{\hat{Z}_t\right\}_{t=1}^{n}$, The Trimmed residuals are defined as follows:

$$\hat{\nu}_t = \hat{Z}_t \mathbf{I}_{(\hat{M}_n^L < \hat{Z}_t < \hat{M}_n^U)} \tag{3.17}$$

The aim is to test the assumption that the α-stable non-causal autoregressive model is adequately identified. For the trimmed residuals, the sample autocorrelation (ACF) at lag k, $\hat{\rho}_k$, is computed by the formula

$$\hat{\rho}_k = \frac{(\sum_{t=k+1}^{n} \hat{\nu}_t \hat{\nu}_{t-k}) - (\sum_{t=k+1}^{n} \hat{\nu}_t)(\sum_{t=k+1}^{n} \hat{\nu}_{t-k})/n - k}{(\sum_{t=1}^{n} \hat{\nu}_t^2) - (\sum_{t=1}^{n} \hat{\nu}_t)^2/n} \tag{3.18}$$

The sample partial autocorrelation ($PACF$) at lag k, $\hat{\pi}_k$, can be derived by Durbin–Levinson algorithm:

$$\hat{\pi}_k = \frac{\hat{\rho}_k - \hat{\rho}_{(k-1)}^T \hat{\mathbf{R}}_{k-2}^{-1} \hat{\rho}_{(k-1)}^*}{\mathbf{1} - \hat{\rho}_k - \hat{\rho}_{(k-1)}^T \hat{\mathbf{R}}_{k-2}^{-1} \hat{\rho}_{(k-1)}} \tag{3.19}$$

where

$$\hat{\rho}_{(k-1)} = (\hat{\rho}_1, \cdots, \hat{\rho}_{k-1})^T \tag{3.20}$$

$$\hat{\rho}_{(k-1)}^* = (\hat{\rho}_{k-1}, \cdots, \hat{\rho}_1)^T \tag{3.21}$$

and $\hat{\mathbf{R}}_{k-2} = (\hat{\rho}_{|i-j|})_{i,j=1}^{k-1}$ is the autocorrelation matrix that is a symmetric matrix.

Theorem 3.1. *If the autoregressive model is correctly identified by the MLE method, then, for any positive integer m, we have*

1) $\sqrt{n}\hat{\rho}_{(m)} \xrightarrow{D} N(0, \mathbf{I}_m)$

2) $\sqrt{n}\hat{\pi}_{(m)} \xrightarrow{D} N(0, \mathbf{I}_m)$

where where $\xrightarrow{D}$ is convergence in distribution, $\hat{\rho}_{(m)} = (\hat{\rho}_1, \cdots, \hat{\rho}_m)^T$, $\hat{\pi}_{(m)} = (\hat{\pi}_1, \cdots, \hat{\pi}_m)^T$, and $\mathbf{I}_m$ is the $m \times m$ identity matrix.

Theorem (3.1) makes it possible to use all the goodness of fit tests presented in section (3.3). The presented statistics have Chi-square distribution with m degrees of freedom. Unlike the classical case, in this case, when we use the trimmed residuals, no degree of freedom is reduced. In order to prove the first part of the theorem, first the Prerequisites of the proof are stated, then two propositions are given, the proof of which is given in [16]. The proof of the second part of the theorem can be found in [19].

We rewrite the stationary solution of the equation $\phi(B)X_t = Z_t$ which is $Y_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ as follows [1]

$$Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + \sum_{j=1}^{\infty} \bar{\psi}_j Z_{t+j} \tag{3.22}$$

now let $\varphi_t = Z_t - \hat{Z}_t$ for $t = 1, \cdots, n$, then

$$\varphi_t = \sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)Y_{t-j} = \sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)\sum_{k=0}^{\infty}\psi_k Z_{t-j-k} + \sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)\sum_{k=1}^{\infty}\bar{\psi}_k Z_{t-j+k} \tag{3.23}$$

By changing the order of summation

$$\left|\sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)\sum_{k=0}^{\infty}\psi_k Z_{t-j-k}\right| \leq \left|\sum_{j=1}^{\infty}\sum_{k=o}^{\min(j,p)}(\phi_k - \hat{\phi}_k)\psi_{j-k}Z_{t-j}\right| \leq \sum_{j=1}^{\infty}\sum_{k=0}^{\min(j,p)}\left\|\phi - \hat{\phi}\right\||\psi_{j-k}|\,|Z_{t-j}| \tag{3.24}$$

and

$$\begin{aligned}\left|\sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)\sum_{k=0}^{\infty}\bar{\psi}_k Z_{t-j+k}\right| &\leq \left|\sum_{j=0}^{\infty}\sum_{k=o}^{p}(\phi_k - \hat{\phi}_k)\bar{\psi}_{j+k}Z_{t+j}\right| \\ &+ \left|\sum_{j=1}^{p-1}\sum_{k=j+1}^{p}(\phi_k - \hat{\phi}_k)\bar{\psi}_{k-j}Z_{t-j}\right| \\ &\leq \sum_{j=0}^{\infty}\sum_{k=1}^{p}\left\|\phi - \hat{\phi}\right\||\bar{\psi}_{j+k}|\,|Z_{t+j}| \\ &+ \sum_{j=1}^{p-1}\sum_{k=j+1}^{p}\left\|\phi - \hat{\phi}\right\||\bar{\psi}_{k-j}|\,|Z_{t-j}|\end{aligned} \tag{3.25}$$

where $\left\|\phi - \hat{\phi}\right\|$ is the Euclidean distance of ϕ and $\hat{\phi}$.] *phi* is estimated by maximum likelihood method from [11], then, $n^{1/\alpha}\left(\hat{\phi} - \phi\right) \xrightarrow{D} \mathbf{S}$ that is $\hat{\phi}$ converges to a random vector in distribution. By the assumption that $0 < \alpha < 2$, it can be found a δ with $\frac{2\alpha}{\alpha+1} < \delta < \min(\alpha, 1)$ such that $n^{1/\alpha} = o(n^{-1/\delta+1/2})$. Note that for any given $\varepsilon > 0$ there always exists a $varsigma > 0$ such that $P(|\mathbf{S}| > \varsigma_1) < \varepsilon/2$ $P(|\mathbf{S}| > \varsigma_1) < \varepsilon/2$ where $|\mathbf{S}|$

denotes the maximum norm. If we define $A_n = \{\left\|\phi - \hat{\phi}\right\| < \varsigma_1 n^{1/\alpha}\}$ then there exists $N_1 > 0$ such that $P(A_n) > 1 - \varepsilon$ whenever $n > N_1$. Under the condition of A_n, we can obtain an upper bound for

$$\left|\sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)\sum_{j=0}^{\infty}\psi_k Z_{t-k}\right| \leq \varsigma_1 n^{-1/\alpha}\sum_{j=1}^{\infty}\sum_{k=1}^{\min(j,p)}|\psi_{j-k}|\,|Z_{t-j}| \tag{3.26}$$

and an upper bound for (3.25)

$$\begin{aligned}\left|\sum_{j=1}^{p}(\phi_j - \hat{\phi}_j)\sum_{k=1}^{\infty}\bar{\psi}_k Z_{t+k}\right| &\leq \varsigma_1 n^{-1/\alpha}\sum_{j=0}^{\infty}\sum_{k=1}^{p}\left|\bar{\psi}_{j+k}\right|\,|Z_{t+j}| \\ &\quad + \varsigma_1 n^{-1/\alpha}\sum_{j=1}^{p-1}\sum_{k=j+1}^{p}\left|\bar{\psi}_{k-j}\right|\,|Z_{t-j}| \end{aligned} \tag{3.27}$$

.

Let $\psi_j^* = \sum_{k=1}^{\min(j,p)}|\psi_{j-k}|$ and $\bar{\psi}_j^* = \sum_{k=1}^{p}\left|\bar{\psi}_{j+k}\right|$. Assuming A_n is true, an upper bound for $|\varphi_t|$ is given by

$$\begin{aligned}|\varphi_t| \leq \varsigma_1 n^{-1/\alpha}\sum_{j=1}^{\infty}\psi_j^*\,|Z_{t-j}| + \varsigma_1 n^{-1/\alpha}\sum_{j=0}^{\infty}\bar{\psi}_j^*\,|Z_{t+j}| \\ + \varsigma_1 n^{-1/\alpha}\sum_{j=1}^{p-1}\bar{\psi}_{1,\cdots,p}\,|Z_{t-j}| \end{aligned} \tag{3.28}$$

Proposition 3.2. *For the* (3.28)*, the following are true*

1) $E\left|\sum_{j=1}^{\infty}\psi_j^*\,|Z_{t-j}|\right|^{\delta} < \infty$

2) $E\left|\sum_{j=0}^{\infty}\bar{\psi}_j^*\,|Z_{t+j}|\right|^{\delta} < \infty$

3) $\varsigma_1 n^{-\delta/\alpha}\sum_{t=1}^{n}\left\{E\left|\sum_{j=1}^{\infty}\psi_j^*\,|Z_{t-j}|\right|^{\delta} + E\left|\sum_{j=0}^{\infty}\bar{\psi}_j^*\,|Z_{t+j}|\right|^{\delta}\right\} = o(n)$

Proof. The coefficients ψ_j and $\bar{\psi}_j$ are geometrically decaying as $j \to \infty$ that is there exist $C_1 > 0$ and $0 < D_1 < 1$ such that $|\psi_j| < C_1 D_1^{|j|}$ for all j, therefore $\sum_{j=1}^{\infty} |\psi_j|^\delta < \infty$

$\sum_{j=1}^{\infty} \left|\psi_j^*\right|^\delta < \infty$. By changing the order of summation and applying the triangle inequality, we have

$$\sum_{j=1}^{\infty} \left|\psi_j^*\right|^\delta \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\min(j,p)} |\psi_{j-k}|^\delta \leq p \sum_{j=0}^{\infty} |\psi_j|^\delta < \infty$$

and

$$\sum_{j=0}^{\infty} \left|\bar{\psi}_j^*\right|^\delta \leq \sum_{j=0}^{\infty} \sum_{k=1}^{p} \left|\bar{\psi}_{j+k}\right|^\delta \leq p \sum_{j=1}^{\infty} \left|\bar{\psi}_j\right|^\delta < \infty$$

Also by the triangle inequality and $E\left|Z_{t-j}\right|^\delta < \infty$

$$E\left|\sum_{j=1}^{\infty} \psi_j^* \left|Z_{t-j}\right|\right|^\delta \leq E\left\{\sum_{j=1}^{\infty} \left|\psi_j^*\right|^\delta \left|Z_{t-j}\right|^\delta\right\} < \infty$$

$$E\left|\sum_{j=0}^{\infty} \bar{\psi}_j^* \left|Z_{t+j}\right|\right|^\delta \leq E\left\{\sum_{j=0}^{\infty} \left|\bar{\psi}_j^*\right|^\delta \left|Z_{t+j}\right|^\delta\right\} < \infty$$

□

Given a fixed number $0 < \lambda < 1$ and $\{\beta_n\}$, a predetermined sequence of real numbers, let $\chi_t = Z_t - Z_{(\lfloor n\lambda \rfloor)} - \beta_n$ the following proposition is true for the non-causal AR sequences.

Proposition 3.3. *For every $\varsigma_2 > 0$*

$$P\{n^{-1/2} \sum_{t=1}^{n} \mathbf{1}_{(|\varphi_t| > \chi_t)} \mathbf{1}_{A_n} > \varsigma_2\} \to 0$$

Lemma 3.4. *Let q^L and q^U be the λ^L-th and λ^U-th quantiles of Z_t respectively. The mean and standard deviation of trimmed random variable $Z_t \mathbf{1}_{(q^L < Z_t < q^U)}$ are shown with μ and σ respectively and*

$$\mu = E[Z_t \mathbf{1}_{(q^L < Z_t < q^U)}]$$

$$\sigma^2 = Var[Z_t \mathbf{1}_{(q^L < Z_t < q^U)}]$$

$$\kappa^2 = \int_{q^L}^{q^U} \int_{q^L}^{t} F(s)[1 - F(t)]dsdt$$

Let $Z_t^\mu = Z_t \mathbf{1}_{(q^L < Z_t < q^U)} - \mu$, *then from [18]*

1) $n^{-1/2} \left\{ \sum_{t=k+1}^{n} Z_t^\mu Z_{t-k}^\mu \right\}_{k=1,\cdots,m} \xrightarrow{D} N(\mathbf{0}, \sigma^4 \mathbf{I}_m)$

2) $n^{-1/2} \sum_{t=1}^{n} Z_t^\mu \xrightarrow{D} N(0, \kappa^2)$

3) $n^{-1} \sum_{t=1}^{n} (Z_t^\mu)^2 \xrightarrow{P} \sigma^2$

where F *is distribution function of a heavy-tailed random variable that presented in section* (2.2)

Lemma 3.5. *Let* M_n^L *and* M_n^U *be the* $n\lambda^L$*-th and* $n\lambda^U$*-th order statistics of* $\{Z_t\}_{t=1}^n$ *respectively and* $\hat{M}_n^L$ *and* $\hat{M}_n^U$ *be the* $n\lambda^L$*-th and* $n\lambda^U$*-th order statistics of* $\left\{\hat{Z}_t\right\}_{t=1}^n$ *respectively. If We define* $\hat{Z}_t^\mu = \hat{Z}_t \mathbf{1}_{(\hat{M}_n^L < \hat{Z}_t < \hat{M}_n^U)} - \mu$, *then from [16]*

1) $n^{-1/2} \sum_{t=k+1}^{n} \left| Z_t^\mu Z_{t-k}^\mu - \hat{Z}_t^\mu \hat{Z}_{t-k}^\mu \right| \xrightarrow{P} 0 \qquad \forall k = 1, 2, \cdots, m$

2) $n^{-1/2} \sum_{t=1}^{n} \left| Z_t^\mu - \hat{Z}_t^\mu \right| \xrightarrow{P} 0$

3) $n^{-1/2} \sum_{t=1}^{n} \left| (Z_t^\mu)^2 - (\hat{Z}_t^\mu)^2 \right| \xrightarrow{P} 0$

Now the theorem (3.1) can be proved.

Proof. Using the fact that

$$\sqrt{n-k}\hat{\rho}_k = \frac{\frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \hat{Z}_t^\mu \hat{Z}_{t-k}^\mu - \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \hat{Z}_t^\mu \frac{1}{n-k} \sum_{t=k+1}^{n} \hat{Z}_{t-k}^\mu}{\frac{1}{n} \sum_{t=1}^{n} (\hat{Z}_t^\mu)^2 - \frac{1}{n^2} (\sum_{t=1}^{n} \hat{Z}_t^\mu)^2}$$

and combining the lemma (3.4) and (3.5) the poof will be completed. □

3.6 Simulation study

The purpose of this section is to verify the accuracy of the theoretical results and demonstrate the effects of the various statistics introduced in the previous sections. The primary goal is to assess the overall effectiveness of the identification tests. A classification of models, their fit, and the parameters of stable production factors are presented. Calculations related to stable distributions are thoroughly reviewed.For example, Nolan [25], [26], and Belov [27] performed complex calculations to estimate the parameters of stable α-distributions. Cui et al. Cui et al. compared Q_{lb}, Q_{mt}, Q_{mm} and, Q_{fg} statistic, to identify non-causal autoregressive time series with infinite variance, which we discuss in this section. As mentioned earlier, since the error terms have infinite variance, they must be trimmed using appropriate quantiles. The question that can mentioned is how the optimal quantiles should be chosen. Unfortunately, finding the optimal quantiles for trimming is challenging [16]. In similar analyses, Lee and Ng [16] trimmed the residuals of the autoregressive model with infinite variance for $\lambda^L = 0.01$ and confirm that approved that the results seem reasonable with this choice. Cui et al. [15] also trimmed the residuals of the fitted model in $\lambda^L = 0.01$ and $\lambda^U = 0.99$ to calculate all statistics ($0 < \lambda^L < \lambda^U < 1$). Also, each of the statistics are calculated using the squared residuals to detect the lack of fit due to the non-linear dependency. It should be noted that that the squared residuals should also be trimmed. However, unlike the ordinary residuals, only the upper quantile of the squared residuals should be trimmed. Let $\{a_t\}$ denote the squares of an $i.i.d.$ sequence of α-stable distributed random variables Z_t with $\alpha < 2$. For some predetermined upper percentile λ define M_n as the $n\lambda$-th order statistics of $\{a_t\}_{t=1}^n$ and obtain the trimmed squared residuals by

$$\xi_t = a_t \mathbf{1}_{(a_t < M_n)} \tag{3.29}$$

The sample autocorrelation function value for $\{\xi_t\}_{t=1}^n$ at lag k can be calculated using (3.18) and denoted by $\hat{\omega}_k$. Theorem (3.1) also holds in this case [15]:

$$\sqrt{n}\hat{\omega}_{(m)} \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_m) \tag{3.30}$$

where $\hat{\omega}_{(m)} = (\hat{\omega}_1, \cdots, \hat{\omega}_2)^T$ and $\mathbf{I}_{(m)}$ is the $m \times m$ identity matrix. Ljung-Box, Monty, Mehdi-McLeod and Fisher-Gallagher statistics, which are used for the square residuals to detect the nonlinear dependence of the series, have

been shown by Q_{lb}, Q_{mt}, Q_{mm}, and Q_{fg} respectively. In this section, the efficiency of the presented tests has been checked using simulation. These tests have been investigated for two different autoregressive models of first and second order non-causality with non-causality order $s = 1$. First, 500 observation from first-order non-causal autoregressive model with $\phi = 2$ and sequence of α-stable noise sequence were simulated, then model parameters including ϕ, α, β, σ, and μ were estimated by maximum likelihood method. This process has been repeated 100 times for $\alpha = \{0.9, 1.2, 15, 1.8\}$, $\beta = 0$. After estimating the parameters of the model, the sequence of $\left\{\hat{Z}_t\right\}$ has been estimated in order to reach the trimmed residuals. All the test statistics were implemented for the trimmed residuals at lags $m = \{5, 10, 15, 20\}$ and finally for every test, the number of rejection based on 100 realization has been shown in Table 3.1. Table 3.2 shows the results of similar computations for a second-order autoregressive model with non-causality order $s = 1$ and parameters $\phi_1 = -1.2$ and $\phi_2 = 1.6$. In Table 3.3, the performance of the tests with a change in the value of β has also been checked. In this case, a similar process has been carried out for simulation. Overall we see in Table 3.1 and Table 3.2 the tests all appear to have satisfactory type I error performance and many tend to be slightly conservative. We have omitted the statistics based on squared residuals and only presented the results for the first two lags and The statistics values have been calculated for two lags: $m = 10$ and $m = 20$. As the α term decreases below 1.2, the empirical size becomes liberal. Furthermore the β term appears to inflate the size when α is small. Fortunately, as noted by [28], the α values in practical applications typically exceed 1.5. Furthermore, these size studies indicate that the lag m at which the statistics are calculated has a negligible impact on their empirical size. Generally, the asymptotic distributions provide satisfactory approximations to these test statistics when $\alpha \geq 1.5$. All tests appear to perform similarly in these instances, though some may slightly conservative.

Table 3.1: number of rejections based on 100 realizations for $AR(1)$ with $s = 1$, $\phi = 2$, $n = 500$, $\beta = 0.5$, $\sigma = 1$, $\mu = 0$, α varies.

α	m	Q_{lb}	Q_{mt}	Q_{mm}	Q_{fg}	Q_{slb}	Q_{smt}	Q_{smm}	Q_{sfg}
0.9	5	5	5	5	5	4	4	4	4
	10	4	5	4	4	4	4	4	4
	15	4	5	5	4	3	6	4	3
	20	3	4	4	3	1	4	4	3
1.2	5	4	4	4	4	5	6	5	4
	10	4	5	4	3	4	5	4	4
	15	5	5	3	4	2	4	4	3
	20	4	5	3	5	2	3	3	3
1.5	5	3	3	3	3	4	5	4	3
	10	4	5	3	3	4	4	4	3
	15	4	5	3	3	2	3	3	3
	20	5	5	2	4	2	2	3	3
1.8	5	4	4	3	3	4	4	3	5
	10	5	4	4	4	4	5	4	5
	15	6	5	3	5	4	4	4	5
	20	6	4	3	5	4	3	4	5

Table 3.2: number of rejections based on 100 realizations for $AR(2)$ with $s = 1$, $\phi_1 = -1.2$, $\phi_2 = 1.6$, $n = 500$, $\beta = 0$, $\sigma = 1$, $\mu = 0$, α varies.

α	m	Q_{lb}	Q_{mt}	Q_{mm}	Q_{fg}	Q_{slb}	Q_{smt}	Q_{smm}	Q_{sfg}
0.9	5	5	5	4	5	4	4	5	4
	10	4	5	5	5	4	5	4	4
	15	4	5	5	5	2	5	4	3
	20	4	4	4	4	3	4	4	3
1.2	5	5	5	5	5	4	4	5	4
	10	4	4	4	4	5	6	4	4
	15	5	4	3	4	2	5	5	4
	20	5	3	3	4	2	4	5	3
1.5	5	4	4	4	4	5	6	5	4
	10	5	4	3	4	4	5	4	4
	15	6	3	3	5	2	4	4	3
	20	7	3	2	5	2	3	3	3
1.8	5	4	4	3	4	4	4	3	3
	10	5	5	4	5	4	4	3	3
	15	6	5	4	5	4	4	3	3
	20	7	4	3	6	4	3	2	4

Table 3.3: number of rejections based on 100 realizations for $AR(2)$ with $s = 1$, $\phi_1 = -1.2$, $\phi_2 = 1.6$, $n = 500$, $\sigma = 1$, $\mu = 0$, (α, β varies).

(α, β)	m	Q_{lb}	Q_{mt}	Q_{mm}	Q_{fg}
$\alpha = 1.5$	10	3	3	3	3
$\beta = 0$	20	5	5	4	4
$\alpha = 0.8$	10	8	8	9	9
$\beta = 0$	20	8	9	8	8
$\alpha = 1.5$	10	4	4	4	3
$\beta = 0.5$	20	4	4	3	3
$\alpha = 0.8$	10	11	11	10	11
$\beta = 0.5$	20	10	9	11	10

3.7 Conclusion

In this thesis, it was investigated how autoregressive and non-causal time series with α-stable errors can be recognized. Distinguishing between causal and non-causal models becomes virtually impossible under Gaussian noise, so it is usual to consider causality as a basic assumption. However, non-causal AR models are identifiable in the case of non-Gaussian noise. different portmanteau tests which are used for finite variance time series model, were studied for infinite variance case and the results showed that in general, the presented methods seem effective, however, when the order of non-causality is determined, the power of the tests decreases. The efficiency of tests was investigated using simulation for non-causal $AR(1)$ and $AR(2)$ with non-causality order 1 with non-Gaussian α-stable noise for different values of α and β. Almost all of the tests have good performance and tend to be conservative and in addition the value of lag m has low impact on empirical size. It is useful to check different autoregressive model (causal and non-causal) for real data with sharp spikes which can be sign of infinite variance, using the portmanteau tests.

Appendix

Appendix A

Appendix

Proof of proposition

As in Lee and Ng (2010), we can pick a constant $\varsigma_3 > 0$ such that $P\left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| > \varsigma_3\right)$ is arbitrarily small in which $s(k) = j$ if Z_j is the kth largest number among $\{Z_1, \cdots, Z_2\}$. To show the result it is sufficient to get

$$\sum_{t=1}^{n} P\left\{\left(|\varphi_t| > |\chi_t|\right) \cap A_n \cap \left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| < \varsigma_3\right)\right\} = o\left(n^{1/2}\right)$$

By Lee and Ng (2010), for any $t \in \{1, \cdots, n\}$

$$P\left\{\left(|\phi_t| > |\chi_t|\right) \cap A_n \cap \left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| < \varsigma_3\right)\right\}$$

Using the triangle inequality and (3.28)

$$\sum_{t=1}^{n} E\left\{\left|\varphi_t\right|^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{\left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| < \varsigma_3\right)} \mathbf{1}_{A_n} \left|t \neq s\left(\lfloor n\lambda \rfloor\right)\right.\right\}$$

$$\leq \sum_{t=1}^{n} E\left\{\varsigma_1^{\delta} n^{-\delta/\alpha} \left(\sum_{j=1}^{\infty} \psi_j^* \left|Z_{t-j}\right|\right)^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{\left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| < \varsigma_3\right)} \mathbf{1}_{A_n} \left|t \neq s\left(\lfloor n\lambda \rfloor\right)\right.\right\} \tag{A.1}$$

$$+ \sum_{t=1}^{n} E\left\{\varsigma_1^{\delta} n^{-\delta/\alpha} \left(\sum_{j=1}^{\infty} \bar{\psi}_j^* \left|Z_{t+j}\right|\right)^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{\left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| < \varsigma_3\right)} \left|t \neq s\left(\lfloor n\lambda \rfloor\right)\right.\right\} \tag{A.2}$$

$$+ \sum_{t=1}^{n} E\left\{\varsigma_1^{\delta} n^{-\delta/\alpha} \left(\sum_{j=1}^{p-1} \bar{\psi}_{1,\cdots,p} \left|Z_{t-j}\right|\right)^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{\left(\left|Z_{s(\lfloor n\lambda \rfloor)}\right| < \varsigma_3\right)} \left|t \neq s\left(\lfloor n\lambda \rfloor\right)\right.\right\} \tag{A.3}$$

To finish the proof, we next show (A.2) and (A.3) $o\left(n^{1/2}\right)$ Conditional on $s\left(\lfloor n\lambda \rfloor\right) = t - j$ and $s\left(\lfloor n\lambda \rfloor\right) \neq t - j$, (A.2) is bounded above by

$$\sum_{t=1}^{n} \varsigma_1^{\delta} \varsigma_3^{\delta} n^{-n/\delta} \sum_{j=1}^{\infty} \left|\psi_j^*\right|^{\delta} E\left\{ \left|\chi\right|_t^{-\delta} \left|s\left(\lfloor n\lambda \rfloor\right)\right| \neq t \right\} \tag{A.4}$$

$$+ \sum_{t=1}^{n} E\left\{ \varsigma_1^{\delta} n^{-\delta/\alpha} \sum_{j=1}^{\infty} \left|\psi_j^*\right|^{\delta} \left|Z_{t-j}\right|^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{(s(\lfloor n\lambda \rfloor) \neq t-j)} \left|s\left(\lfloor n\lambda \rfloor\right) \neq t \right\} \tag{A.5}$$

using (3.2) and $n^{-1/\alpha} = o\left(n^{-1/\delta+1/2}\right)$ to (A.5) and get

$$\sum_{t=1}^{n} \varsigma_1^{\delta} \varsigma_3^{\delta} n^{-\delta/\alpha} \sum_{j=1}^{\infty} \left|\psi_j^*\right|^{\delta} E\left\{ \left|\chi_t\right|^{-\delta} \left|s\left(\lfloor n\lambda \rfloor\right) \neq t \right\} = o\left(n^{1/2}\right)$$

for (A.5), two cases, $1 \leq j \leq t-1$ and $j \geq t$ t, are considered. When $1 \leq j \leq t-1$,

$$\sum_{t=1}^{n} E\left\{ {\varsigma_1}^{\delta} n^{-\delta/\alpha} \sum_{j=1}^{t-1} \left|\psi_j^*\right|^{\delta} \left|Z_{t-j}\right|^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{(s(\lfloor n\lambda \rfloor) \neq t-j)} \left|s(\lfloor n\lambda \rfloor) \neq t \right\}$$

$$\leq \sum_{t=1}^{n} \varsigma_1^{\delta} n^{-\delta/\alpha} \frac{n-2}{n-1} E\left\{ \sum_{j=1}^{t-1} \left|\psi_j^*\right|^{\delta} \left|Z_{t-j}\right|^{\delta} \left|\chi_t\right|^{-\delta} \left|s(\lfloor n\lambda \rfloor) \neq t, t-j \right\} = o(n^{1/2})$$

since by (5.22) of Lee and Ng (2010), $E\left\{ \sum_{j=1}^{t-1} \left|\psi_j^*\right|^{\delta} \left|Z_{t-j}\right|^{\delta} \left|\chi_t\right|^{-\delta} \left|s\left(\lfloor n\lambda \rfloor\right) \neq t, t-j \right\} = o\left(1\right)$ When $j \geq t$, Z_{t-j} is in the set $\{Z_0, Z_{-1}, \cdots\}$. Hence, Z_{t-j} j is independent of $s\left(\lfloor n\lambda \rfloor\right)$, which implies that

$$E\left\{ \left|\psi_j^*\right|^{\delta} \left|Z_{t-j}\right|^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{(s(\lfloor n\lambda \rfloor) \neq t-j)} \left|s\left(\lfloor n\lambda \rfloor\right) \neq t \right\} =$$

$$\left|\psi_j^*\right|^{\delta} E\left\{ \left|Z_{t-j}\right|^{\delta} \right\} E\left\{ \left|\chi_t\right|^{-\delta} \mathbf{1}_{(s(\lfloor n\lambda \rfloor) \neq t-j)} \left|s\left(\lfloor n\lambda \rfloor\right) \neq t \right\}$$

Now use Proposition 1 and (5.23) of Lee and Ng (2010) again

$$\sum_{t=1}^{n} E\left\{ \varsigma_1^{\delta} n^{-\delta} n^{-\delta/\alpha} \sum_{j=t}^{\infty} \left|\psi_j^*\right|^{\delta} \left|Z_{t-j}\right|^{\delta} \left|\chi_t\right|^{-\delta} \mathbf{1}_{(s(\lfloor n\lambda \rfloor)\neq t-j)} \left| s\left(\lfloor n\lambda \rfloor\right) \neq t \right.\right\}$$

$$\leq \sum_{t=1}^{n} \varsigma_1^{\delta} n_1^{\frac{-\delta}{\alpha}} \frac{n-1}{n} \sum_{j=t}^{\infty} \left|\psi_j^*\right|^{\delta} E\left|Z_{t-j}\right|^{\delta} E\left\{ \left|\chi_t\right|^{-\delta} \mathbf{1}_{(s(\lfloor n\lambda \rfloor)\neq t-j)} \mid s\left(\lfloor n\lambda \rfloor\right) \neq t \right\} = o(n^{\frac{1}{2}})$$

Bibliography

[1] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer science & business media, 1991. → pages 2, 18, 27, 28, 36, 39

[2] Gennady Samoradnitsky. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Routledge, 2017. → pages 12

[3] Sidney I Resnick. Heavy tail modeling and teletraffic data: special invited paper. *The Annals of Statistics*, 25(5):1805–1869, 1997. → pages 26

[4] Colin M Gallagher. A method for fitting stable autoregressive models using the autocovariation function. *Statistics & probability letters*, 53 (4):381–390, 2001. → pages 26

[5] Shiqing Ling. Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(3):381–393, 2005. → pages 26

[6] Bart W Stuck and Beat Kleiner. A statistical analysis of telephone noise. *Bell System Technical Journal*, 53(7):1263–1320, 1974. → pages 26

[7] Hu Sheng and YangQuan Chen. Farima with stable innovations model of great salt lake elevation time series. *Signal Processing*, 91(3):553–561, 2011. → pages 26

[8] ST Rachev, Isabella Huber, and Sergio Ortobelli. Portfolio choice with heavy tailed distributions. *Technical Report*, 2004. → pages 26

[9] Yesim Tokat and Eduardo S Schwartz. The impact of fat tailed returns on asset allocation. *Mathematical Methods of Operations Research*, 55: 165–185, 2002. → pages 26

[10] F Jay Breidt, Richard A Davis, and A Alexandre Trindade. Least absolute deviation estimation for all-pass time series models. *The Annals of Statistics*, 29(4):919–946, 2001. → pages 27, 36

[11] Beth Andrews, Matthew Calder, and Richard A Davis. Maximum likelihood estimation for α-stable autoregressive processes. 2009. → pages 27, 28, 37, 40

[12] Markku Lanne, Jani Luoto, and Pentti Saikkonen. Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28(3):623–631, 2012. → pages 27

[13] Murray Rosenblatt. *Gaussian and non-Gaussian linear time series and random fields*. Springer Science & Business Media, 2012. → pages 27, 36

[14] F Jay Breid, Richard A Davis, Keh-Shin Lh, and Murray Rosenblatt. Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36(2):175–198, 1991. → pages 27, 37

[15] Yunwei Cui, Thomas J Fisher, and Rongning Wu. Diagnostic tests for non-causal time series with infinite variance. *Journal of Statistical Planning and Inference*, 147:117–131, 2014. → pages 28, 38, 44

[16] Sangyeol Lee and Chi Tim Ng. Trimmed portmanteau test for linear processes with infinite variance. *Journal of multivariate analysis*, 101 (4):984–998, 2010. → pages 28, 38, 39, 43, 44

[17] George EP Box and David A Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970. → pages 32, 33

[18] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978. → pages 32, 33

[19] Anna Clara Monti. A proposal for a residual autocorrelation test in linear models. *Biometrika*, 81(4):776–780, 1994. → pages 32, 33, 39

[20] Daniel Peña and Julio Rodríguez. The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *Journal of Statistical Planning and Inference*, 136(8):2706–2718, 2006. → pages 32, 34

[21] Esam Mahdi and A Ian McLeod. Improved multivariate portmanteau test. *Journal of Time Series Analysis*, 33(2):211–222, 2012. → pages 32, 35

[22] Dong Wan Shin and Jong Hyup Lee. Distribution of residual autocorrelations in nonstationary autoregressive processes. *Journal of Time Series Analysis*, 17(1):105–109, 1996. → pages 32

[23] Thomas J Fisher and Colin M Gallagher. New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, 107(498):777–787, 2012. → pages 32, 35

[24] Allan I McLeod and William K Li. Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of time series analysis*, 4(4):269–273, 1983. → pages 32, 35

[25] John P Nolan. Numerical calculation of stable densities and distribution functions. *Communications in statistics. Stochastic models*, 13(4):759–774, 1997. → pages 44

[26] John P Nolan. An algorithm for evaluating stable densities in zolotarev's (m) parameterization. *Mathematical and computer modelling*, 29(10-12):229–233, 1999. → pages 44

[27] IA Belov. On the computation of the probability density function of α-stable distributions. *Math Model Anal*, 2:333–341, 2005. → pages 44

[28] Stefan Mittnik, Marc S Paolella, and Svetlozar T Rachev. Diagnosing and treating the fat tails in financial returns data. *Journal of Empirical Finance*, 7(3-4):389–416, 2000. → pages 45