Some Generalized and Heavy-tailed Time Series Models

A thesis submitted in
partial fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

by

Monika Singh Dhull

(2018MAZ0005)

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY
ROPAR

April, 2024

Monika S. Dhull: *Some Generalized and Heavy-tailed Time Series Models*
© 2024 Indian Institute of Technology Ropar
All rights reserved.

To my parents and brother, for their love, patience and consistent support.

Declaration of Originality

I hereby declare that the work which is being presented in the thesis entitled **Some Generalized and Heavy-tailed Time Series Models** has been solely authored by me. It presents the result of my own independent investigation/research conducted during the time period from January, 2019 to January, 2024 under the supervision of Dr. Arun Kumar, Associate Professor, Department of Mathematics, Indian Institute of Technology Ropar, India.

To the best of my knowledge, it is an original work, both in terms of research content and narrative, and has not been submitted or accepted elsewhere, in part or in full, for the award of any degree, diploma, fellowship, associateship, or similar title of any university or institution. Further, due credit has been attributed to the relevant state-of-the-art and collaborations with appropriate citations and acknowledgments, in line with established ethical norms and practices. I also declare that any idea/data/fact/source stated in my thesis has not been fabricated/ falsified/misrepresented. All the principles of academic honesty and integrity have been followed. I fully understand that if the thesis is found to be unoriginal, fabricated, or plagiarized, the institute reserves the right to withdraw the thesis from its archive and revoke the associated degree conferred. Additionally, the institute also reserves the right to appraise all concerned sections of society of the matter for their information and necessary action. If accepted, I hereby consent for my thesis to be available online in the institute's open access repository, inter-library loan, and the title and abstract to be made available to outside organizations.

Signature

Name: Monika Singh Dhull
Entry Number: 2018MAZ0005
Program: Ph. D.
Department of Mathematics
Indian Institute of Technology Ropar
Rupnagar, Punjab 140001, India

Date: 03/04/2024

Acknowledgement

First, I would like to express my deepest gratitude to God for His guidance, strength, and for providing me with opportunities throughout this academic journey. I am truly thankful for the blessings that have sustained me during challenging times. I take great joy to express my gratitude to all the individuals who, in one way or another, have contributed to the completion of this study.

I am indebted to my supervisor Dr. Arun Kumar, who believed in my abilities and provided me the opportunity to work under his guidance. I express my gratitude for the patience, understanding, and motivation extended to me by him throughout my Ph.D. journey. His guidance, combined with his knowledge and experience, made my Ph.D. attainable. These two lines from my guide will forever linger in my memory:

यक़ीन हो तो कोई रास्ता निकलता है
हवा की ओट भी ले कर चराग़ जलता है

"If there is belief, there is a way; the lamp burns despite the wind."

I would also like to thank our collaborators, Prof. Nikolai Leonenko and Prof. Agnieszka Wyłomańska, whose insights and contributions played a crucial role in completion of research work.

I extend my sincere gratitude to my doctoral committee members: Dr. Arvind K. Gupta, Dr. Partha S. Dutta, Dr. Arti Pandey and Dr. Satyam Agarwal for the constructive feedbacks and suggestions throughout the process. I also appreciate the help from office assistant Ms. Jaspreet Kaur, senior lab assistant Mr. Neeraj, and attendant Indrajit Ji. I would also like to thank IIT Ropar for granting the fellowship through Ministry of Education, providing travel grant for attending conferences, and providing all the necessary resources required for my research. This work was partially supported by the FIST program of the Department of Science and Technology, Government of India, reference no. SR/FST/MS-I/2018/22(C).

Completing this PhD journey would not have been possible without the support of some amazing friends. I express my appreciation to my research group, Neha Di, Niharika, Naman, Priti, Pawan, and Meenu, for their discussions on research topics and their understanding and helpful nature. I also thank Amrendra Bhaiya, Krishnendu Bhaiya, Taran Di, and Gopika Di for sharing their insights whenever needed throughout this journey. I am really grateful for the joyful moments spent with Himanshu, Ankita, Suman, Ayantika, Tanvir, Swati, Kapil, Smita, Bharti, Jasbir, Nikhil, Ravi, Nitin, and Karan. I am thankful to Niharika, Aditi, Sonam, Vikas, and Sahil for being there for me in the highs and lows of life. I have shared unforgettable moments which I will cherish in life. I am really blessed to have

Arzoo, Shagun, Bipasha, Manju, Param, Kusum and Taran Di as my friends, and they have been a source of support and encouragement throughout this journey. Special thanks to Arzoo staying by my side during difficult times and helping me with presentations, Himanshu and Sahil for always providing a listening ear. I also enjoyed the endless discussions on Mathematics with Gunjeet, which somehow kept me motivated. I would like to thank Anuradha, Neeta Chechi, Rashmi, VJ, Prerna, Sonam, Sayantan, and Yogesh for friendship that has been a source of joy and support. I acknowledge Renu Ma'am, Nayamat Ma'am, and Sannidhya Sir for their guidance. I am grateful for the long lived friendships of Ritu and Khushboo, who are now like a family. I appreciate all the good times and support over the years.

Last but not the least, I thank my family and relatives for being supportive and patient throughout this journey. My parents and brother are my pillars of strength, I could not have completed this journey without their support. Their selfless love, care and belief in my capabilities have been a constant source of motivation for me. I consider myself fortunate to have understanding parents and brother with whom I can share anything, like a friend.

Monika Singh Dhull

Certificate

This is to certify that the thesis entitled **Some Generalized and Heavy-tailed Time Series Models** submitted by **Monika Singh Dhull (2018MAZ0005)** for the award of the degree of **Doctor of Philosophy** to Indian Institute of Technology Ropar is a record of bonafide research work carried out under my guidance and supervision. To the best of my knowledge and belief, the work presented in this thesis is original and has not been submitted, either in part or full, for the award of any other degree, diploma, fellowship, associateship or similar title of any university or institution.

In my opinion, the thesis has reached the standard fulfilling the requirements of the regulations relating to the degree.

Signature

Dr. Arun Kumar
Department of Mathematics
Indian Institute of Technology Ropar
Rupnagar, Punjab 140001, India

Date: 03/04/2024

Lay Summary

This work deals with the modeling of time series data, which are defined as collection of random variables indexed by discrete time. The commonly used time series models, such as Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA) models, assume normality in the error terms. However, real-world phenomena often exhibit asymmetry, skewness, and non-Gaussian behavior, which is commonly seen in financial data, meteorological patterns, traffic flow and so on.

This thesis begins by introducing non-Gaussian innovation terms in the AR model, specifically focusing on semi-heavy-tailed and heavy-tailed distributions. We consider the AR model with normal inverse Gaussian innovation terms, exhibiting semi-heavy-tail behavior. The expectation maximization (EM) algorithm is proposed for parameter estimation, and its performance is compared with other methods through simulations and real data applications. We further study the AR model with Cauchy-distributed innovation terms, capturing heavy-tailed behavior and extreme events. The mixture representation of Cauchy distribution is used for parameter estimation using the EM algorithm. We also propose to use the empirical characteristic function method for estimation. The geometric infinitely divisible random variables are then explored, introducing AR models with geometric infinitely divisible marginals, and their distributional properties are discussed.

The focus moves to modeling the non-stationary time series data using a new model, the Humbert Fractionally Differenced Autoregressive Moving Average (HARMA) model, which is introduced with two types of Humbert polynomials. Stationarity and invertibility conditions are established, and the Whittle quasi-likelihood method is employed for parameter estimation. The effectiveness of the method is demonstrated through simulations and applied to Spain's 10-year treasury bond yield data.

Overall, this work contributes to the understanding of time series modeling by incorporating non-Gaussian innovations, heavy-tailed distributions, and addressing non-stationary data through novel models like the HARMA model. The proposed methods are evaluated through simulations and real data applications, providing insights into their performance and applicability in capturing the dynamics of various phenomena.

Abstract

The collection of random variables $\{Y_t : t \in T\}$ defined on the same sample space with time T as the index set is known as a time series. The most fundamental and easy-to-understand time series models in the literature are autoregressive (AR), moving average (MA), and the mixture of AR and MA model known as the autoregressive moving average (ARMA) model. These models are defined as the linear combination of previous terms and error terms or innovation terms, where these error terms are assumed to be normal with mean 0 and constant variance σ^2.

However, asymmetry, skewness, and non-Gaussian behavior are commonly observed in many real-life phenomena. For example, in financial data, stock prices exhibit non-Gaussian behavior with extreme values, meteorological data show asymmetry due to extreme weather events and long-term climate changes, in traffic flow, sudden congestion, accidents, or disruptions can lead to non-Gaussian behavior and asymmetry in data, and so on. To efficiently capture these events, different non-Gaussian models are considered, such as the AR model with exponential, Student's t-distribution, Laplace, Cauchy and other distributions for innovation terms.

In this thesis, we initiate our study of the AR model by considering non-Gaussian innovation terms, specifically focusing on the semi-heavy-tailed and heavy-tailed classes of distribution. First, we consider the AR(p) model with normal inverse Gaussian innovation terms, which has semi-heavy-tail behavior. We propose using the expectation-maximization (EM) algorithm for parameter estimation of the model. Further, we conduct an extensive simulation study to assess the method's performance and compare the EM method with Yule-Walker and conditional least squares methods. We also apply the proposed model to three real datasets, namely, Google equity closing price, US gasoline price and NASDAQ historical data.

In the next chapter, we consider the AR(p) model with Cauchy-distributed innovation terms. Again, this distribution is heavy-tailed with infinite mean and variance, effectively capturing extreme events. We make use of the mixture representation of the Cauchy distribution, and employ the EM algorithm for estimation. We also discuss another method based on the empirical characteristic function for parameter estimation. A simulation study is performed to compare the EM method with maximum likelihood estimation for the Cauchy distribution. Next, we delve into a class of geometric infinitely divisible random variables by examining their Laplace exponents, characterized by Bernstein functions. We introduce AR models with geometric infinitely divisible (gid) marginals, namely geometric tempered stable, geometric gamma, and geometric inverse Gaussian. We also provide some distributional properties and the limiting behavior of the probability

densities of these random variables at 0^+. Further, we present parameter estimation methods for the introduced AR(1) model, using both conditional least squares and the method of moments. The performance of estimation methods for the AR(1) model is assessed using simulated data. From empirical study on geometric tempered stable, geometric gamma, and geometric inverse Gaussian distributions, we conclude that these distributions belong to the class of semi-heavy-tailed distribution.

Until now, the focus of the work has been on one of the fundamental time series models, namely the AR model, which is applied to stationary data. The autoregressive integrated moving average (ARIMA) models accommodate non-stationary time series data by employing integer order differencing. It involves lagged innovation terms along with differencing steps. An extension of this model is referred to as the autoregressive fractionally integrated moving average (ARFIMA) model, which has a fractional differencing operator. Using similar approach, we introduce a new model by considering two different types of Humbert polynomials and call these models as type 1 and type 2 Humbert fractionally differenced autoregressive moving average (HARMA) models. We also establish the stationarity and invertibility conditions of these introduced models. The focus is particularly directed towards Pincherle ARMA, Horadam ARMA, and Horadam-Pethe ARMA processes, which are particular cases of HARMA models. The Whittle quasi-likelihood method is employed for parameter estimation of the introduced processes. This method yields consistent and normally distributed estimators, and its effectiveness is further assessed through a simulation study for the Pincherle ARMA process. Finally, the Pincherle ARMA model is applied to Spain's 10-year treasury bond yield data, demonstrating its effectiveness in capturing the dynamics of the market.

Keywords: Autoregressive models; time series models; expectation-maximization algorithm; Whittle quasi-likelihood estimation; heavy-tailed distributions; geometric infinitely divisibility; autoregressive moving average models.

List of Publications

Journal Articles

- Monika S. Dhull and Arun Kumar (2024): **Geometric infinitely divisible autoregressive models**. *Revised manuscript submitted.*

- Niharika Bhootna, Monika S. Dhull, Arun Kumar and Nikolai Leonenko (2023): **Humbert generalized fractional differenced ARMA processes**. *Communications in Nonlinear Science and Numerical Simulation*, 125, 107412.

- Monika S. Dhull, Arun Kumar, and Agnieszka Wyłomańska (2023): **The expectation-maximization algorithm for autoregressive models with normal inverse Gaussian innovations**. *Communications in Statistics - Simulation and Computation*, 1-21.

- Monika S. Dhull, and Arun Kumar (2021): **Normal inverse Gaussian autoregressive model using EM algorithm**. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, 13, 139-147.

Conference Paper

- Monika S. Dhull and Arun Kumar (2022): **Expectation-Maximization algorithm for autoregressive models with Cauchy innovations**. *Engineering Proceedings*, 18(1), 21.

Contents

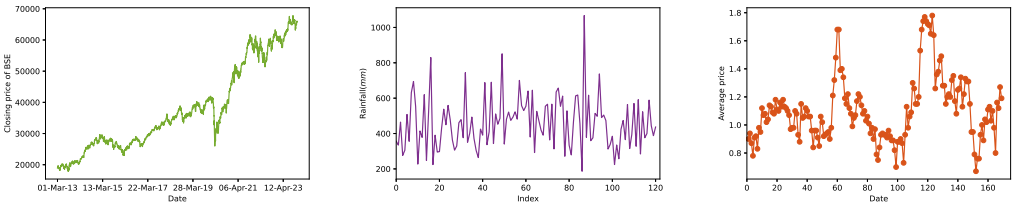
List of Figures

List of Tables

Chapter 1

Introduction

Time series data refers to a collection of data points collected sequentially at regularly spaced intervals of time [27]. Mathematically, a collection of random variables $\{Y_t : t \in T\}$ is said to be a time series, if it is defined on same sample space with index set T as discrete time. Different time series models, namely autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) models, have been extensively studied and explored in literature. These models aim to better capture the properties of data, especially the economical, meteorological, biological, and environmental data. These datasets represent time-varying random processes where extreme events can have a substantial impact on the outcome or prediction. Therefore, the study of time series models holds great importance. Data from different domains exhibit different behaviors, for example, consider the daily adjusted closing price data of Bombay Stock Exchange (BSE) for the period of March 1, 2013 to November 24, 2023 [1], the average rainfall data of the Punjab region for the years 1901 to 2016[2], and the average price of avocados [3]. Based



(a) Closing price of BSE from 2013 to 2023.

(b) Rainfall in Punjab from 1901 to 2016 (mm).

(c) Average price of avocado $(2015 - 2018)$.

Figure 1.1: Examples of time series data from different domains.

on these behaviors, we classify time series data as stationary and non-stationary. Observe that the data in Fig. 1.1 have non-stationary behavior as, BSE closing price has increasing trend, rainfall data and avocado price have spikes at random time. When we refer to stationary, we mean weak stationarity [26], where the statistical properties such as mean, variance, and autocovariance of the data are

[1] https://www.bseindia.com/Indices/IndexArchiveData.html
[2] https://www.kaggle.com/datasets/aksahaha/rainfall-india
[3] https://www.kaggle.com/datasets/shrishtitiwari04/avocado

constatnt or independent of time. For weakly stationary time series, this implies that $\mathbb{E}[Y(t)] = \mu$ is independent of t, and $\gamma_Y(t+h,t)$ is independent of t for each h [27]. The classical AR and ARMA processes are among the most foundational, easy, and interpretable models which work on stationary data. These traditional models are defined as the linear combination of lagged terms and innovation terms with the assumption that the innovation terms are serially uncorrelated with zero mean and finite variance, usually identical and independent having normal distribution. Mathematically, ARMA model is written as [26],

$$\Phi(B)Y_t = \Theta(B)\epsilon_t, \tag{1.1}$$

where ϵ_t is serially uncorrelated with mean 0 and variance σ^2, $\Phi(B)$, $\Theta(B)$ are stationary AR(p) and invertible MA(q) operators respectively, defined as,

$$\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j \text{ and } \Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j,$$

B is the lag operator, that is, $B^j(Y_t) = Y_{t-j}$. For $q = 0$, model in Eq. (1.1) represents AR model of order p and $p = 0$ corresponds to MA model of order q. The conventional time series models assume that the conditional variance of innovation terms does not depend on past information. However, many real-life phenomena exhibit time varying variance and depend on past observations. In 1982 Robert F. Engle proposed the autoregressive conditional heteroskedasticity (ARCH) model [50], which allowed the conditional variance of a time series to be time-varying and dependent on past observations. The model explicitly incorporated lagged squared observations as explanatory variables. He aimed to address the issue of volatility clustering, where instances of high volatility are followed by instances of high volatility and vice versa. The ARCH model was also applied on inflation data of the United Kingdom, demonstrating its effectiveness in capturing time-varying volatility. Later in 1986 [24], generalized ARCH model known as GARCH model was introduced. This model allow both a long memory and a more flexible lag structure by taking the lagged conditional variances. The model very well explained the uncertainty of inflation rate.

In a study undertaken by [113], it was found that among 21 real time series datasets, only 6 datasets' residuals exhibited a normal distribution. It is observed in literature that many real data has asymmetry, skewness, and non-Gaussian behavior, which is discussed further. Particularly in financial markets, the recorded time series of log returns deviates from a Gaussian distribution, exhibiting tails that are more substantial than a normal distribution but lighter than a

power law [139]. The innovation terms or errors also contribute significantly in observing the patterns in time series data. In a Gaussian time series model, the innovation terms are assumed to be Gaussian, independent and identically distributed (i.i.d.). If the distribution of innovation terms or marginal distribution of series Y_t is non-Gaussian, then it is called as non-Gaussian time series model. Non-Gaussian observations are prevalent in various domains, including the modeling of asset returns, climate data, binary outcomes, and more [28,76]. The study of non-Gaussian autoregressive models have been significantly studied for several years [14]. Lawrance [98] initially introduced AR models with exponential, gamma and mixed exponential distributions. Subsequently, Dewald and Lewis (1985) [42] studied AR(1) Laplace process. Then in 1989, Damsleth and El-Shaaravi [38] developed a time series model by incorporating Laplace noise, which worked as an alternative to normal distribution. Sim in 1990 [129], studied the AR(1) model with gamma and exponential processes. Choi and Choi [34], assumed the innovation terms to be Cauchy or a mixture of normal and Cauchy distribution for the AR(1) model with a near unit root. Later, Seethalekshmi and Jose [100,101] introduced various AR models utilizing α-Laplace and Pakes distributions. Additionally, Jose and Abraham [85] extended the count models with Mittag-Leffler waiting times. Some other well-considered non-Gaussian AR models are namely Student's t-distribution, gamma distribution, Cauchy distribution, and Laplace distribution [34,35,112,129,134,136,138]. In 2010, Trindade et al. [138] proposed ARMA and generalized autoregressive conditional heteroscedastic (GARCH) models driven by asymmetric Laplace (AL) innovations and provided the marginal distribution of proposed models. Conditional maximum likelihood based inference was also advocated and asymptotic properties were presented.

The distributions having tails heavier than a normal distribution and lighter than the power law are known as semi-heavy-tailed distributions. The normal inverse Gaussian (NIG) distribution, introduced by Barndorff-Nielsen [15], falls into this category. Note that Cauchy and Student's t-distribution with small degree of freedom ($\nu = 1$) have tails heavier than the power law and are called heavy-tailed distributions. Mathematically, the distribution of a random variable X with distribution function F is said to have a heavy (right) tail if the moment generating function of X, $M_X(t)$, is infinite for all $t > 0$ [55], that is,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\, dF(x) = \infty \text{ for all } t > 0.$$

The AR model with innovation terms from Student's t-distribution is thoroughly investigated for the multivariate case [134].

In the first segment of our work, we propose AR(p) model with innovations from

semi heavy-tailed and heavy-tailed distributions as it can model large jumps in the observed data [43,46]. Many financial data, meteorological data, biomedical data and so on exhibit extreme values, for example, see [28,106,138]. We consider an AR(p) model with NIG distribution and Cauchy distribution [44]. We apply the AR(p) NIG model on Google equity price, NASDAQ stock index and USA gasoline price data. Subsequently, we also introduce AR models using geometric infinitely divisible distributions namely, geometric inverse Gaussian, geometric gamma and geometric tempered stable [45]. We present the integral form for the probability density function (pdf) of the innovation terms in the autoregressive (AR) model, given by $Y_t = \theta Y_{t-1} + \epsilon_t$. Empirically, it is showed that these distributions are semi heavy-tailed. It is noteworthy to mention that, among the non-Gaussian AR models discussed so far, most of the research work is aligned towards the AR models with heavy-tailed distributions, which can very well handle stationary data. Recent work of Bhootna and Kumar [22] on AR(1) model with one-sided tempered stable marginals and innovations is the evidence of application of non-Gaussian and heavy-tailed times series models.

Another ubiquitous characteristic of real world time series data is non-stationarity [26]. To model the non-stationary data effectively, Box and Jenkins [26] introduced the autoregressive integrated moving average (ARIMA) model, which is formulated as follows:

$$\Phi(B)(1-B)^{\nu}Y_t = \Theta(B)\epsilon_t,\ \nu \in \mathbb{N}, \tag{1.2}$$

where ϵ_t is Gaussian white noise with variance σ^2, B is the lag operator, ν is differencing parameter, $\Phi(B)$, and $\Theta(B)$ are stationary AR(p) and invertible MA(q) operators, respectively. The ARIMA model can not capture the long range dependencies which is common in the time series data. Therefore, to overcome this issue, autoregressive fractionally integrated moving average (ARFIMA) model was introduced by Clive W.J. Granger and Robert Joyeux in 1980 [66] and later by Hosking [79]. This model is a natural extension of ARIMA framework obtained by taking a fractional differencing operator $(1-B)^{\nu},\ \nu \in \mathbb{R}$ in (1.2), that is, ν can take any real value. The ARFIMA model is stationary for $\nu \in (-0.5, 0.5)$. This model is proved to capture the memory effects in the time series data. It's statistical properties were further studied by Robinson [121], which paved the way for the estimation of long-memory parameters. The other well studied and applied time series model is the Gegenbauer autoregressive moving average (GARMA) model, which is an extension of the traditional ARMA model that use Gegenbauer type

differencing operator. GARMA model can be expressed as,

$$\Phi(B)(1-2uB+B^2)^{\nu}Y_t = \Theta(B)\epsilon_t,$$

where $\nu \in (0, 1/2), |u| < 1$, and $(1-2uB+B^2)^{-\nu}$ is generating function of Gegenbauer polynomials [67]. The second segment of our work focuses on modelling the non-stationarity and long range dependencies present in the time series data by considering some generalized differencing operators. We use the fractional differencing approach and the generating functions of Humbert polynomials to define two types of Humbert generalized fractionally differenced ARMA processes (HARMA) [21]. For $m \in \mathbb{N}$, the HARMA model of two types use the generating function of the form $(1-muB+B^m)^{-\nu}$ and $(1-2uB+B^m)^{-\nu}$, respectively. We also study the stationarity and invertibility conditions for these introduced processes. Specifically, we focus on Pincherle ARMA, Horadam ARMA and Horadam-Pethe ARMA processes and their properties.

After model selection, the subsequent step is the estimation of model parameters. For classical time series models, Yule-Walker and conditional least squares methods are widely applied. The Yule-Walker estimation method is based on the Yule-Walker equations calculated for the considered model, and utilizes the empirical autocovariance function for the analyzed data. More details of Yule-Walker method for autoregressive models can be found in [27]. The conditional least squares method estimates the model parameters for dependent observations by minimizing the sum of squares of deviations about the conditional expectation. Various estimation methods have been explored with the introduced different time series models. However, these methods can not be applied for AR models with heavy-tailed distributions, as it is based on assumption of mean zero of error terms and also need the existence of finite order moments, which is not possible for some of the heavy-tailed distributions. Choi and Choi in [34], estimated the parameters of the Cauchy AR model using maximum likelihood estimation and least square estimation methods. Christmas and Everson [35] used variational Bayes, and Nduka [112] recently used the expectation maximization (EM) algorithm to estimate the parameters of the AR model with Student's t-distribution innovations. The EM algorithm was introduced by Dempster et al. in [40] and it is considered as an alternative to numerical optimization of the likelihood function. It iterates between two steps namely, expectation step and maximization step. The EM algorithm is discussed in more details in Section 2.3. In this thesis, we consider NIG and Cauchy distributions for innovations term in AR model. Note that NIG distribution has Bessel function of third kind due to which the direct computation of derivative of log likelihood function is difficult. In second case, the Cauchy distribution does not

have finite variance, therefore we cannot apply the classical estimation methods like method of moments or the conditional least square method in case of AR model with Cauchy innovations. Keeping these behaviors in mind, we propose to use EM algorithm for parameter estimation of AR(p) NIG and compare the results with Yule-Walker and conditional least squares methods. Along with EM algorithm, we also propose to use empirical characteristic function for estimation of AR(p) Cauchy model and compare the same with maximum likelihood estimation method. However, for AR(p) models with geometric infinitely divisible distributions, we use the conditional least squares and method of moments which provide the analytical parameter estimates. Further, in second segment of our work we have fractionally differenced Humbert processes which are long memory processes, therefore we use spectral density based estimation known as Whittle quasi-likelihood estimation [125]. It was introduced by P. Whittle in 1950s [143] particularly to estimate parameters of long-memory processes. This method is based on the spectral density function, which characterizes the distribution of frequencies in a time series. It involves minimizing a quasi-likelihood function derived from the theoretical spectral density and empirical spectral density.

The subsequent chapters of the thesis are organized as follows. Chapter 2 is dedicated to introducing the terminologies, methods and results utilized throughout the thesis. Chapter 3 introduces to the work related to AR model with normal inverse Gaussian innovation terms and its applications. Chapter 4 deals with AR model with distribution of innovations as Cauchy. Chapter 5 is devoted to definitions of geometric infinitely divisible random variables and their properties. Further, we also define AR model of order 1 with geometric infinitely divisible marginals and innovations. Moreover, the Humbert polynomials based fractionally differenced ARMA models are included in Chapter 6. The last chapter concludes the work with some future ideas.

Chapter 2

Preliminaries

In this chapter, we discuss some important definitions and results applied throughout the thesis. This chapter also provides the classical time series models and prevalent estimation methods.

2.1 Definitions

Definition 2.1 (Heavy-tailed distributions [55]). The distribution of a random variable X with distribution function F is said to have a heavy (right) tail if the moment generating function of X, $M_X(t)$, is infinite for all $t > 0$, that is

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\, dF(x) = \infty, \text{ for all } t > 0.$$

Definition 2.2 (Asymptotic functions). For large x, the functions f and g are asymptotic, which is denoted by $f \sim g$, if and only if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Definition 2.3 (Characteristic function [122]). For a random variable X, the complex-valued function $\phi(s)$ on $\mathbb{R}$ is defined as characteristic function of X as,

$$\phi(s) = \mathbb{E}(\exp(i\, sX)) = \int_{-\infty}^{\infty} \exp(i\, sx) f(x)\, dx,$$

where $f(x)$ is probability density function of X.

Definition 2.4 (Bernstein function [128]). The function $g : (0, \infty) \to (0, \infty)$ is a Bernstein function if g is of class $\mathbb{C}^\infty$ and $g(s) \geq 0$, for all $s \geq 0$ and $(-1)^{n-1} g^{(n)}(s) \geq 0$, for all $n \in \mathbb{N}$ and $s > 0$.

Definition 2.5 (Infinitely divisible random variable [132]). A random variable Y is infinitely divisible if for every $n \in \mathbb{N}$ there exist independent, identically distributed (iid) random variables $Y_{1n}, \ldots, Y_{nn}$ such that, $Y \stackrel{\mathrm{d}}{=} Y_{1n} + Y_{2n} + \ldots + Y_{nn}$, where $Y \stackrel{\mathrm{d}}{=}$ represents equality in distributions.

Definition 2.6 (Laplace exponent). For a positive infinitely divisible random variable Y the Laplace transform has the following form

$$\mathbb{E}\left(e^{-sY}\right) = e^{-\psi(s)}.$$

The function $\psi(\cdot)$ is called the Laplace exponent.

Definition 2.7 (Geometric infinite divisibility [92]). The random variable Y on $(0,\infty)$ is said to have geometric infinitely divisible marginals if its Laplace transform is given by $f(s) = \frac{1}{1+g(s)}$, where $g(s)$ is a Bernstein function which is the Laplace exponent of some positive infinitely divisible random variable.

Definition 2.8 (Slowly varying function [52]). A positive (not necessarily monotone) function $L(x)$ defined on $(0,\infty)$ varies slowly at infinity if and only if $\lim_{x\to\infty} \dfrac{L(tx)}{L(x)} \to 1$ is true for all $t > 0$.

Definition 2.9 (Long memory process [48]). A stationary process $\{Y_t : t \in \mathbb{N}\}$ having finite second order moments is said to have long memory if the series $\sum_{h=0}^{n-1} |\gamma(h)| \to \infty$ as $n \to \infty$, where $\gamma(h) = \mathrm{Cov}\,(Y(t), Y(t+h))$ is the covariance function.

Definition 2.10 (Seasonal long memory [36]). The stationary time series $\{Y_t\}$ is said to have seasonal long memory if there exist $\omega_0 \in \mathbb{R}$ and $\alpha \in (0,1)$ such that the autocorrelation

$$\rho(h) \simeq h^{-\alpha}\cos(h\omega_0), \text{ as } h \to \infty,$$

and $\cos(h\omega_0) \neq 1$.

Definition 2.11 (Spectral density [48]). The spectral density of stationary time series $\{Y_t\}$ is defined as Fourier transform of autocorrelation $\rho(h)$, that is,

$$f_Y(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \rho(h)e^{-\iota\omega h},$$

where $-\pi \leq \omega \leq \pi$ are Fourier frequencies.

2.2 Classical time series models

This section deals with some well-known time series models. These models form the foundation for the models defined in the subsequent chapters.

Autoregressive process of order p (AR(p)): The time series $\{Y_t\}$ is AR process of order p, if

$$\Phi(B)Y_t = \epsilon_t, \; t = 0, \pm1, \pm2, \cdots, \tag{2.1}$$

where $\{\epsilon_t\}$ is white noise with mean 0 and variance σ^2, B is the lag operator, $\Phi(B)$ is stationary AR(p) operator defined as,

$$\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j.$$

Moving average process of order q (MA(q)): The time series $\{Y_t\}$ is MA process of order q, if

$$Y_t = \Theta(B)\epsilon_t, \ t = 0, \pm 1, \pm 2, \cdots, \tag{2.2}$$

where $\{\epsilon_t\}$ is white noise with mean 0 and variance σ^2, B is the lag operator, $\Theta(B)$ invertible MA(q) operator defined as,

$$\Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j.$$

Autoregressive moving average process of order (p, q) (ARMA(p, q)): The time series $\{Y_t\}$ is ARMA process of order (p, q), if

$$\Phi(B)Y_t = \Theta(B)\epsilon_t, \ t = 0, \pm 1, \pm 2, \cdots, \tag{2.3}$$

where $\{\epsilon_t\}$ is white noise with mean 0 and variance σ^2, B is the lag operator, $\Phi(B)$, $\Theta(B)$ are stationary AR(p) and invertible MA(q) operators respectively, defined as,

$$\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j \text{ and } \Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j, \tag{2.4}$$

where $\Phi(B)$ is called stationary operator and $\Theta(B)$ is called invertible operator if all roots of these operators lie outside the unit circle.

Autoregressive integrated moving average process of order (p, ν, q) (ARIMA(p, ν, q)): The time series $\{Y_t\}$ is ARIMA process, if

$$\Phi(B)(1 - B)^{\nu} Y_t = \Theta(B)\epsilon_t, \ t = 0, \pm 1, \pm 2, \cdots, \tag{2.5}$$

where ν is non-negative integer, $\{\epsilon_t\}$ is white noise with mean 0 and variance σ^2, B is the lag operator, $\Phi(B)$, $\Theta(B)$ are stationary AR(p) and invertible MA(q) operators respectively, defined in Eq. (2.4).

Autoregressive fractionally integrated moving average process of order (p, ν, q) (ARFIMA(p, ν, q)): The time series $\{Y_t\}$ is ARFIMA process [79], if

$$\Phi(B)(1 - B)^{\nu} Y_t = \Theta(B)\epsilon_t, \ t = 0, \pm 1, \pm 2, \cdots, \tag{2.6}$$

where ν is real number, $\{\epsilon_t\}$ is white noise with mean 0 and variance σ^2, B is the lag operator, $\Phi(B)$, $\Theta(B)$ are stationary AR(p) and invertible MA(q) operators respectively, defined in Eq. (2.4). ARFIMA is also known as long memory process as it can capture the long range dependence present in the data. The other well studied and applied time series model is the Gegenbauer autoregressive moving average (GARMA) model, which is an extension of the traditional ARMA model that uses Gegenbauer polynomials.

Gegenbauer autoregressive moving average of order (p, ν, q) (GARMA(p, ν, q)): GARMA model can be expressed as,

$$\Phi(B)(1 - 2uB + B^2)^{\nu} Y_t = \Theta(B)\epsilon_t,$$

where $\nu \in (0, 1/2)$, $|u| < 1$ and $(1-2uB+B^2)^{-\nu}$ is generating function of Gegenbauer polynomial [67]. The Gegenbauer polynomial has following form:

$$(1 - 2ux + x^2)^{-\nu} = \sum_{n=0}^{\infty} C_n^{\nu}(u)x^n,$$

where $\nu \neq 0$, $|x| < 1$ and $C_n^{\nu}(u)$ is given by:

$$C_n^{\nu}(u) = \sum_{k=0}^{n/2} (-1)^k \frac{\Gamma(n - k + \nu)(2u)^{n-2k}}{\Gamma(d)\Gamma(n + 1)\Gamma(n - 2k + 1)}.$$

The exemplary time series plot from above discussed models are shown in Fig. 2.1.

2.3 Estimation methods

In this section, we discuss some well known parameter estimation methods for time series models.

Expectation Maximization (EM) algorithm: This estimation method iterates between two steps, namely the expectation step (*E-step*) and the maximization step (*M-step*). The *E-step* computes the expectation of the complete data log-likelihood with respect to the conditional distribution of the unobserved or hidden data, given the observations and the current estimates of the parameters. Further, in the *M-step*, a new estimate for the parameters is computed which maximize the complete data log-likelihood computed in the *E-step* [108].
Consider the complete data as (X, Y) with observed data as X, hidden data as

Y and θ be the set of unknown parameters. We denote $f(X, Y|\theta)$ as the joint density function of complete data given the parameter set θ. We find the conditional expectation of log-likelihood of complete data with respect to the conditional distribution of Y given X, denoted as $Q(\theta|\theta^{(k)})$,

$$Q(\theta|\theta^{(k)}) = \mathbb{E}_{Y|X,\theta^{(k)}}[\log f(X, Y|\theta)|X, \theta^{(k)}], \tag{2.7}$$

where $\theta^{(k)}$ represents the estimates of the parameter vector at k-th iteration and $\mathbb{E}_{Y|X,\theta^{(k)}}$ is the conditional expectation with respect to Y given X. Further, in the *M-step*, we compute the parameters by maximizing the expected log-likelihood of complete data found in the *E-step* such that

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta|\theta^{(k)}).$$

Conditional least squares: The conditional least squares method estimates the model parameters for dependent observations by minimizing the sum of squares of deviations about the conditional expectation [94]. For preceding data, $\mathcal{F}_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_1)^T$ and model parameter θ, the conditional least squares function is defined as,

$$L(\theta) = \sum_{s=1}^{t} \{Y_s - \mathbb{E}[Y_s|\mathcal{F}_{s-1}]\}^2. \tag{2.8}$$

The parameter θ is estimated by minimizing the function $L(\theta)$ with respect to θ.

Yule-Walker estimation method: Consider an autoregressive (AR) model of order p, denoted as AR(p), given by the equation:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t, \tag{2.9}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive parameters, and ϵ_t is the white noise error terms. The Yule-Walker equations provide a method to estimate the autoregressive parameters by solving a system of equations based on the sample autocorrelation function (ACF). The sample ACF at lag k is denoted as $\hat{\rho}_k$. The Yule-Walker equations for an AR(p) model are given by [146]:

$$\begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix} = \begin{bmatrix} 1 & \phi_1 & \phi_2 & \ldots & \phi_{p-1} \\ \phi_1 & 1 & \phi_1 & \ldots & \phi_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{p-1} & \phi_{p-2} & \ldots & \phi_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{bmatrix}.$$

The solution to these equations provides estimates for the autoregressive parameters $\phi_1, \phi_2, \ldots, \phi_p$, and the variance of the white noise term can also be estimated.

Conditional maximum likelihood estimation: Consider an AR(p) model defined in Eq. (2.9), then the likelihood function for a time series of length n conditional on the initial values $Y_1, Y_2, \ldots, Y_p$ is given by:

$$L(\boldsymbol{\phi}, \theta | Y_1, Y_2, \ldots, Y_p) = \prod_{t=p+1}^{n} f(y_t | \mathbf{Y}_t), \tag{2.10}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)^\intercal$ is the vector of autoregressive parameters, θ is the parameter of the white noise error term, $\mathbf{Y}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^\intercal$ is the vector of lagged values, and $f(y_t|\mathbf{Y}_t)$ is the conditional distribution of time series. The conditional maximum likelihood estimates $\hat{\boldsymbol{\phi}}$ and $\hat{\theta}$ are obtained by maximizing the log-likelihood function [146],

$$\arg\max_{\boldsymbol{\phi}, \theta} \log L(\boldsymbol{\phi}, \theta | Y_1, Y_2, \ldots, Y_p).$$

The maximization can be performed using numerical optimization techniques.

Method of moments: The method of moments is a statistical technique for estimating the parameters by equating the theoretical moments (expectations) of a distribution to the sample moments computed from the data. This method provides a set of equations that can be solved to obtain estimates for the parameters [31].

Kernel density estimation method: The kernel density estimation method also known as Parzen-Rosenblatt window method, is a non-parametric approach to find the underlying probability distribution of data. It is a technique that lets one to create a smooth curve given a set of data and one of the most famous method for density estimation. For a sample $S = \{x_i\}_{i=1,2,\ldots,N}$, having distribution function $f(x)$ has the kernel density estimate $\widehat{f}(x)$ defined as [49]

$$\widehat{f}(x) = \frac{1}{N} \sum_{n=1}^{N} K_\sigma(x - x_i), \tag{2.11}$$

where K_σ is kernel function with bandwidth σ such that $K_\sigma(t) = (\frac{1}{\sigma}) K(\frac{t}{\sigma})$.

2.4 Goodness of fit tests

Following are some of the goodness of fit tests which are used in this thesis:

One-sample Kolmogorov-Smirnov (KS): The one-sample KS test is a non-parametric test based on the maximum difference between an empirical and a hypothetical cumulative distribution [107]. Consider a sample $Y = (y_1, y_2, \ldots, y_n)$ to test whether it comes from a specific distribution, say $F(y)$. The null hypothesis is that the sample comes from $F(y)$. The one-sample KS test statistic is given by:

$$D_n = \max\left(\sup_y |F_n(y) - F(y)|\right),$$

where $F_n(y)$ is the empirical distribution function of the sample, and $F(y)$ is the cumulative distribution function of the reference distribution. To perform the KS test, you can compare the calculated test statistic D_n with the critical value from the Kolmogorov-Smirnov distribution or use a significance level to determine whether to reject the null hypothesis.

Two-sample Kolmogorov-Smirnov (KS): The two-sample KS test is a non-parametric test which compares two independent samples, denoted as $X = (x_1, x_2, \ldots, x_m)$ and $Y = (y_1, y_2, \ldots, y_n)$, to determine whether they are drawn from the same distribution [107]. The test statistic for the two-sample KS test is given by:

$$D_{m,n} = \max\left(\sup_x |F_m(x) - F_n(x)|\right),$$

where $F_m(x)$ and $F_n(x)$ are the empirical distribution functions of samples X and Y, respectively. To perform the KS test, you can compare the calculated test statistic $D_{m,n}$ with the critical value from the Kolmogorov-Smirnov distribution or use a significance level to determine whether to reject the null hypothesis.

Augmented Dickey Fuller (ADF): The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a time series has a unit root, indicating non-stationarity. The null hypothesis of the test is that the time series has a unit root, while the alternative hypothesis is that the time series is stationary [47].

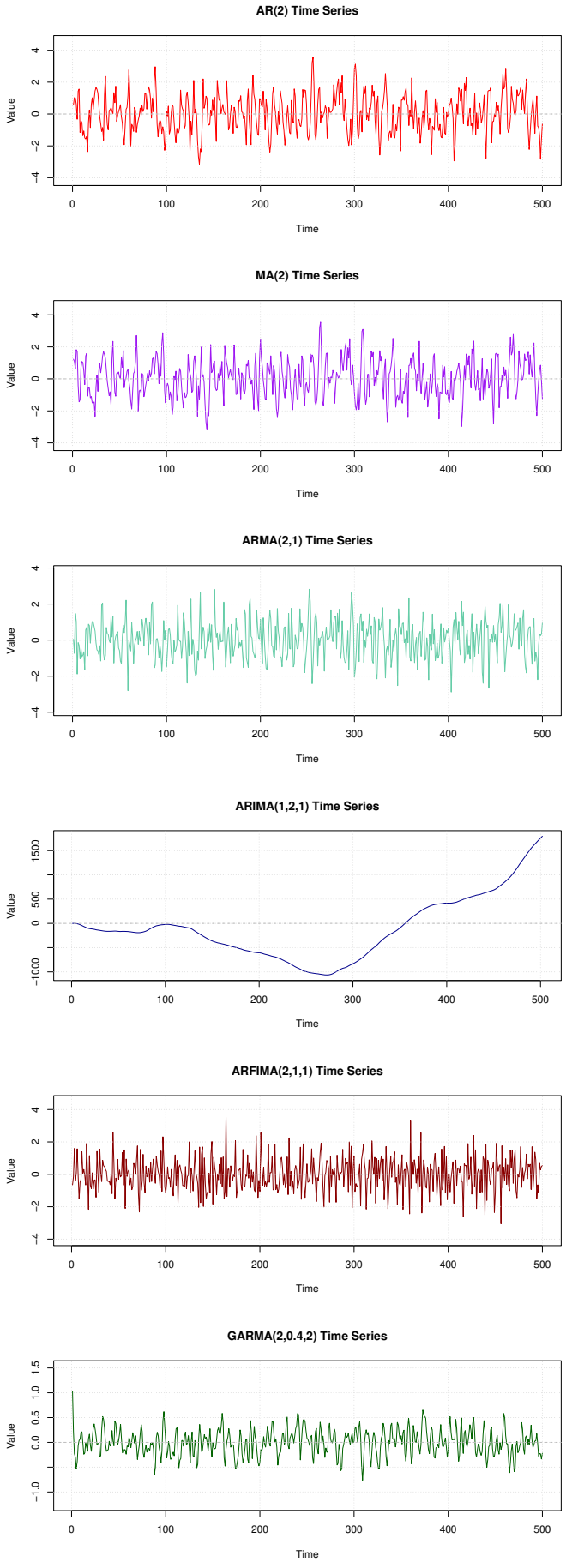


Figure 2.1: The exemplary time series of length $N = 500$ from different time series models with Gaussian innovations.

Chapter 3

Autoregressive model with normal inverse Gaussian innovations

In this chapter, we introduce an autoregressive process of order p (AR(p)) with normal inverse Gaussian innovations. Further, we apply expectation maximization (EM) algorithm to estimate the model parameters. The proposed estimation method is assessed on simulated data. We also present the real world applications of the model in different domain.

3.1 Introduction

In time series modeling, the autoregressive (AR) models are used to represent the time-varying random process in which output depends linearly on previous terms and a stochastic term, also known as error term or innovation term. In the classical approach, the marginal distribution of the innovation terms is assumed to be normal. However, the application of non-Gaussian distributions in time series allows modeling the outliers and asymmetric behavior visible in many real data. In addition, in financial markets the distribution of the observed time series (log-returns) are mostly non-Gaussian and have tails heavier than the normal distribution, however lighter than the power law. These kind of distributions are also called semi heavy-tailed, see, e.g., [37,116,118]. The AR models with non-Gaussian innovations are very well studied in the literature. Sim [129] considered AR model of order 1 with Gamma process as the innovation term. For AR models with innovations following a Student's t-distribution, see, e.g., [35,112,134,136] and references therein. Note that Student's t-distribution is used in modeling of asset returns [76].

One of the semi heavy-tailed distributions with wide range of shapes is normal inverse Gaussian (NIG), which was introduced by Barndorff-Nielsen [15]. NIG distributions were used to model the returns from the financial time-series [6,16,20,29,88,141,142]. In practical situations, we come across the data with skewness, extreme values or with missing values which can be easily assimilated by the NIG distribution. The distribution has stochastic representation, i.e., it can

be written as the normal variance-mean mixture where the mixing distribution is the inverse Gaussian distribution and a more general distribution known as generalised hyperbolic distribution is obtained by using generalised inverse Gaussian as the mixing distribution [16]. Many authors proposed various methods for analysis and estimation of NIG-distributed models, which is a testimony of their popularity, see, e.g. [32,72,89,90]. In 2017, Angelis and Viroli[7] also discussed the Markov-switching regression model with NIG innovations and illustrated through empirical studies that NIG is more flexible than Student's t-distribution and generalized error distribution. In model based clustering[114], a mixture of multivariate normal inverse Gaussian (MNIG) distributions is found to be a better alternative to a mixture of Gaussian distributions as it is able to incorporate the skewness and fatter tails of mixture components. In the recent literature we can also find various complex models with NIG distribution, such as GARCH or GARCH-based models [70,111], and exponential NIG model [4]. The mentioned models are especially useful in cryptocurrencies analysis to predict their future values. In the literature, there are also various models and stochastic processes that are based on NIG distribution. The very first example is the NIG Lévy process, see e.g, [124] which was also applied to financial data modeling [5]. In the recent literature, one can also find the interesting applications of NIG Lévy process for default of company prediction [86]. There are also numerous time series models with NIG distributed innovations and their various applications. We mention here the interesting analysis of heteroscedastic models [54,57,82,91,147] and autoregressive models [140], see also [144]. However in financial markets the observed time series (log-returns) distributions are non-Gaussian and have tails heavier than Gaussian and lighter than power law. These kind of distributions are also called semi heavy-tailed distributions (see e.g.[37,116,118]). For a literature survey on different innovation distributions or marginal distributions in the AR(1) model see [69]. The AR(1) models with non-Gaussian innovation terms are very well considered in the literature. Sim in 1990 considered AR(1) model with Gamma process as the innovation term [129]. For AR models with innovations following a Student's t-distribution see e.g. [35,112,134,136] and references therein. Note that t-distribution is used in modeling of asset returns (see [76]). Normal inverse Gaussian (NIG) distribution introduced by Barndorff-Nielsen [15] is a semi heavy tailed distribution with tails heavier than the Gaussian but lighter than the power law tails. NIG distribution is defined as the normal variance-mean mixture where the mixing distribution is the inverse Gaussian distribution. NIG distributions and processes are used to model the returns from the financial time-series see e.g.[16,88]. A more general distribution which is obtained by taking normal variance-mean mixture with mixing distribution as generalised inverse Gaussian distribution is called generalised hyperbolic distribution see [17].

It is worth mentioning that, in addition to applications in economics and finance, the NIG distributions have also found use in a wide range of other fields, including computer science [126], energy markets [19], commodity markets [152], and image analysis [150]. The multivariate NIG distributions, counterparts of univariate NIG distributions, were considered in [33,115] and were applied in various disciplines, see e.g.[1]. The current literature is also enriched by research papers with other interesting applications of NIG distribution, such that epilepsy description [75], EEG signals analysis [74] or sleep apnea detection [73]. In this chapter, we introduce an AR(p) model with NIG innovation terms. Due to heavy-tailedness of the innovation term it can model large jumps in the observed data. First, we show that AR(1) process very well model the Google stock price time-series. The non-Gaussian behaviour of the innovation terms of the Google stock price is shown using QQ plot and Kolmogorov-Smirnov test. From a market risk management perspective obtaining reasonable extreme observation levels is a crucial objective in modeling. Since, it capture the market extreme movements which a Gaussian based model doesn't capture. The parameters of the model are estimated using EM algorithm. Further, we extend the model to AR(p) model and perform the analysis for AR(2) and AR(1) models. The model applications are demonstrated on NASDAQ stock market index data and US gasoline price data.
The rest of the chapter is organised as follows. In Section 3.2, the NIG autoregressive model is defined along with important properties of the NIG distribution. Section 3.2.1 discusses the estimation procedure of the parameters of the introduced model using EM algorithm. The efficacy of the estimation procedure on simulated data and the real world financial data application is discussed in Section 3.3. Section 3.4 concludes.

3.2 NIG autoregressive model

In this section, we introduce the AR(p) model having independent identically distributed (i.i.d.) NIG residuals. However, first we remind the definition and main properties of the NIG distribution.

Definition 3.1 (NIG distribution)**.** A random variable X is said to have a NIG distribution which is denoted by X follows NIG $(\alpha, \beta, \mu, \delta)$, if its probability density function (pdf) has the following form

$$f(x;\alpha,\beta,\mu,\delta) = \frac{\alpha}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu\right)\phi(x)^{-1/2}K_1(\delta\alpha\phi(x)^{1/2})\exp(\beta x),\ x\in\mathbb{R}, \tag{3.1}$$

where $\phi(x) = 1 + [(x-\mu)/\delta]^2$, $0 \le |\beta| \le \alpha$, $\mu \in \mathbb{R}$, $\delta > 0$ and $K_\nu(x)$ denotes the

modified Bessel function of the third kind of order ν evaluated at x and is defined by

$$K_\nu(x) = \frac{1}{2}\int_0^\infty y^{\nu-1}e^{-\frac{1}{2}x(y+y^{-1})}dy.$$

The NIG distribution belongs to the four parameter family distribution. The parameter μ is location, β is an asymmetry parameter with $\beta = 0$ distribution is symmetric, $\beta > 0$ is rightly skewed, and $\beta < 0$ is left skewed, α represents the tail behavior and δ is scale parameter. We use the asymptotic properties of the modified Bessel function, $K_\nu(x) \sim \sqrt{\frac{\pi}{2}}e^{-x}x^{-1/2}$ [84] and the fact that $\phi(x) \sim (x/\delta)^2$ for large x. For definition of asymptotic functions see Def. 2.2. We get the following expression:

$$\begin{aligned}
f(x;\alpha,\beta,\mu,\delta) &\sim \frac{\alpha}{\pi}e^{(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu)}\phi(x)^{-1/2}\sqrt{\frac{\pi}{2}}e^{-\delta\alpha\sqrt{\phi(x)}}\left(\delta\alpha\sqrt{\phi(x)}\right)^{-1/2}e^{\beta x}\\
&\sim \frac{\alpha}{\pi}e^{(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu)}\sqrt{\frac{\pi}{2}}(\delta\alpha)^{-1/2}\phi(x)^{-3/4}e^{-\delta\alpha\sqrt{\phi(x)}}e^{\beta x}\\
&\sim \frac{\alpha}{\pi}e^{(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu)}\sqrt{\frac{\pi}{2}}(\delta\alpha)^{-1/2}\delta^{3/4}x^{-3/4}e^{\mu\alpha}e^{-(\alpha-\beta)x},\ \alpha>\beta,\\
&\sim \sqrt{\frac{\alpha}{2\pi}}e^{(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu)}\delta x^{-3/2}e^{-(\alpha-\beta)x},\ \alpha>\beta.
\end{aligned}$$

From the above, one can conclude that the tail probability for NIG distributed random variable X satisfies the following

$$\mathbb{P}(X > x) \sim cx^{-3/2}e^{-(\alpha-\beta)x},$$

where $c = \sqrt{\frac{\alpha}{2\pi}}\frac{\delta}{(\alpha-\beta)}e^{(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu)}$, which shows that NIG is a semi-heavy tailed distribution [37,116,118]. It is worth mentioning that the NIG distributed random variable X can be represented in the following form

$$X = \mu + \beta G + \sqrt{G}Z, \tag{3.2}$$

where Z is a standard normal random variable i.e. $Z \sim N(0,1)$ and G has an inverse Gaussian(IG) distribution with parameters γ and δ denoted by $G \sim \text{IG}(\gamma,\delta)$, having pdf of the following form

$$g(x;\gamma,\delta) = \frac{\delta}{\sqrt{2\pi}}\exp(\delta\gamma)x^{-3/2}\exp\left(-\frac{1}{2}\left(\frac{\delta^2}{x}+\gamma^2 x\right)\right),\ x>0. \tag{3.3}$$

The representation given in Eq. (3.2) is useful when we generate the NIG distributed random numbers. Also, Z and G are independent. It is also suitable to apply the EM algorithm for the maximum likelihood estimation (MLE) of the considered

model's parameters. The representation in Eq. (3.2) makes it convenient to find the main characteristics of a NIG distributed random variable X also mentioned in [89], namely, we have

$$\mathbb{E}X = \mu + \delta\frac{\beta}{\gamma} \quad \text{and} \quad \text{Var}(X) = \delta\frac{\alpha^2}{\gamma^3}.$$

Moreover, the skewness and kurtosis of X is given by

$$\text{Skewness} = \frac{3\beta}{\alpha\sqrt{\delta\gamma}} \quad \text{Kurtosis} = \frac{3(1+4\beta^2/\alpha^2)}{\delta\gamma}.$$

For $\beta = 0$, the NIG distribution is symmetric. Moreover, it is leptokurtic if $\delta\gamma < 1$ while it is platykurtic in case $\delta\gamma > 1$. Note that a leptokurtic NIG distribution is characterized by larger number of outliers than we have for normal distribution and thus, it is a common tool for financial data description.
The following properties of NIG distribution are presented in Prop. 2.3 in [95]. Let $X \sim$ NIG $(\alpha, \beta, \mu, \delta)$, then following holds.

(a) The moment generating function of X is

$$M_X(u) = e^{\mu u + \delta(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}}.$$

(b) If $X \sim$ NIG $(\alpha, \beta, \mu, \delta)$, then $X + c \sim$ NIG $(\alpha, \beta, \mu + c, \delta)$, $\quad c \in \mathbb{R}$

(c) If $X \sim$ NIG $(\alpha, \beta, \mu, \delta)$, then $cX \sim$ NIG $(\alpha/c, \beta/c, c\mu, c\delta)$, $\ c > 0$.

(d) If $X_1 \sim$ NIG $(\alpha, \beta, \mu_1, \delta_1)$ and $X_2 \sim$ NIG $(\alpha, \beta, \mu_2, \delta_2)$ are independent then the sum $X_1 + X_2 \sim$ NIG $(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$.

(e) If $X \sim$ NIG $(\alpha, \beta, \mu, \delta)$, then $\frac{X-\mu}{\delta} \sim$ NIG $(\alpha\delta, \beta\delta, 0, 1)$.

NIG autoregressive model of order p: We can define the AR(p) univariate stationary time-series $\{Y_t\}$, $t \in \mathbb{Z}$ with NIG residuals

$$\boldsymbol{Y_t} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{p}} \boldsymbol{\rho_i Y_{t-i}} + \boldsymbol{\varepsilon_t} = \boldsymbol{\rho^T Y_{t-1}} + \boldsymbol{\varepsilon_t}, \tag{3.4}$$

where $\rho = (\rho_1, \rho_2, \cdots, \rho_p)^T$ is a p-dimensional column vector

$$\mathbf{Y_{t-1}} = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})^T$$

is a vector of p lag terms and $\{\epsilon_t\}$, $t \in \mathbb{Z}$ are i.i.d. residuals distributed as NIG$(\alpha, \beta, \mu, \delta)$.

The $\{\epsilon_t\}$ represents residuals of the analyzed model. The process $\{Y_t\}$ is a stationary one if and only if the modulus of all the roots of the characteristic polynomial $(1-\rho_1 z-\rho_2 z^2-\cdots-\rho_p z^p)$ are greater than one. We assume that the error term follows a symmetric NIG distribution with mean 0 i.e. $\mu=\beta=0$. Using properties of NIG distribution, the conditional distribution of Y_t given $\rho,\alpha,\beta,\mu,\delta$ and the preceding data $\mathcal{F}_{t-1}=(Y_{t-1},Y_{t-2},\cdots,Y_1)^T$ is given by

$$p(Y_t|\rho,\alpha,\beta,\mu,\delta,\mathcal{F}_{t-1})=f(y_t;\alpha,\beta,\mu+\rho^T\mathbf{y_{t-1}},\delta),$$

where $f(\cdot)$ is the pdf given in Eq. (3.1) and $\mathbf{y_{t-1}}$ is the realization of $\mathbf{Y_{t-1}}$. We have $\mathbb{E}[Y_t]=\mathbb{E}[\varepsilon_t]=0$ and $\mathrm{Var}[Y_t]=\sigma_\varepsilon^2+\sum_{j=1}^p\rho_j\gamma_j$, where $\sigma_\varepsilon^2=\mathrm{Var}(\varepsilon_t)=\delta\alpha^2/\gamma^3$ and $\gamma_j=\mathbb{E}[Y_tY_{t-j}]=\rho_1\gamma_{j-1}+\rho_2\gamma_{j-2}+\cdots+\rho_p\gamma_{j-p},\ j\geq 1$.

3.2.1 Parameter estimation using EM algorithm

In this section, we provide a step-by-step procedure to estimate the parameters of the model proposed in Eq. (3.4). The procedure is based on EM algorithm. We provide estimates of all parameters of the introduced AR(p) with NIG distributed residuals. It is worth to mention that EM is a general iterative algorithm for model parameter estimation by maximizing the likelihood function in the presence of missing or hidden data. The EM algorithm was introduced in [40] and it is considered as an alternative to numerical optimization of the likelihood function. It is popularly used in estimating the parameters of Gaussian mixture models (GMMs), estimating hidden Markov models (HMMs) and model-based data clustering algorithms. Some extensions of EM include the expectation conditional maximization (ECM) algorithm [109] and expectation conditional maximization either (ECME) algorithm [104]. For a detailed discussion on the theory of EM algorithm and its extensions we refer the readers to [108]. The EM algorithm iterates between two steps, namely the expectation step (*E-step*) and the maximization step (*M-step*). In our case, the observed data X is assumed to be from NIG$(\alpha,\beta,\mu,\delta)$ and the unobserved data G follows IG$(\sqrt{\alpha^2-\beta^2},\delta)$. We find the conditional expectation of log-likelihood of complete data (X,G) with respect to the conditional distribution of G given X. For $\theta=(\alpha,\beta,\mu,\delta,\rho^T)$ we find

$$Q(\theta|\theta^{(k)})=\mathbb{E}_{G|X,\theta^{(k)}}[\log f(X,G|\theta)|X,\theta^{(k)}],$$

in the *E-step* where $\theta^{(k)}$ represents the estimates of the parameter vector at k-th iteration.

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta|\theta^{(k)}).$$

The algorithm is proven to be numerically stable [108]. Also, as a consequence of Jensen's inequality, log-likelihood function at the updated parameters $\theta^{(k+1)}$ will not be less than that at the current values $\theta^{(k)}$. Although there is always a concern that the algorithm might get stuck at local extrema, but it can be handled by starting from different initial values and comparing the solutions. In next proposition, we provide the estimates of the parameters of the model defined in Eq. (3.4) using EM algorithm.

Proposition 3.1. *Consider the AR(p) time-series model given in Eq. (3.4) where error terms follow NIG$(\alpha, \beta, \mu, \delta)$. The maximum likelihood estimates of the model parameters using EM algorithm are as follows*

$$\hat{\rho} = \left(\sum_{t=1}^{N} w_t \mathbf{Y_t} \mathbf{Y_{t-1}^T}\right)^{-1} \sum_{t=1}^{N} (w_t y_t - \mu w_t - \beta)\, \mathbf{Y_{t-1}},$$

$$\begin{aligned}
\hat{\mu} &= \frac{\sum_{t=1}^{N} \epsilon_t w_t - N\beta}{N\bar{w}_t}, \\
\hat{\beta} &= \frac{\sum_{t=1}^{N} (w_t \epsilon_t) - N\bar{w}_t \bar{\epsilon}_t}{N(1 - \bar{s}_t \bar{w}_t)}, \\
\hat{\delta} &= \sqrt{\frac{\bar{s}}{(\bar{s}\bar{w} - 1)}}, \quad \hat{\gamma} = \frac{\hat{\delta}}{\bar{s}}, \text{ and } \hat{\alpha} = (\hat{\gamma}^2 + \hat{\beta}^2)^{1/2},
\end{aligned} \tag{3.5}$$

where $\epsilon_t = y_t - \rho^T \mathbf{Y_{t-1}}$, $\gamma = \sqrt{\alpha^2 - \beta^2}$, $\bar{\epsilon}_t = \frac{1}{N}\sum_{t=1}^{N} \epsilon_t$, $s_t = \mathbb{E}_{G|\varepsilon,\theta^{(k)}}(g_t|\epsilon_t, \theta^{(k)})$, $\bar{s} = \frac{1}{N}\sum_{t=1}^{N} s_t$, $w_t = \mathbb{E}_{G|\varepsilon,\theta^{(k)}}(g_t^{-1}|\epsilon_t, \theta^{(k)})$ *and* $\bar{w} = \frac{1}{N}\sum_{t=1}^{N} w_t$.

Proof. For AR(p) model, let (ε_t, G_t), for $t = 1, 2, ..., N$ denote the complete data. The observed data ε_t is assumed to be from NIG$(\alpha, \beta, \mu, \delta)$ and the unobserved data G_t follows IG(γ, δ). We can write the innovation terms as follows:

$$\varepsilon_t = Y_t - \rho^{\mathbf{T}} \mathbf{Y_{t-1}}, \text{ for } t = 1, 2, ..., N.$$

Note that $\varepsilon|G = g \sim N(\mu + \beta g, g)$ and the conditional pdf is

$$f(\varepsilon = \epsilon_t|G = g_t) = \frac{1}{\sqrt{2\pi g_t}} \exp\left(-\frac{1}{2g_t}(y_t - \rho^T \mathbf{y_{t-1}} - \mu - \beta g_t)^2\right).$$

Now, we need to estimate the unknown parameters $\theta = (\alpha, \beta, \mu, \delta, \rho^T)$. We find the conditional expectation of log-likelihood of unobserved/complete data (ε, G) with respect to the conditional distribution of G given ε. Since the unobserved data is assumed to be from $\text{IG}(\gamma, \delta)$ therefore, the posterior distribution is generalised inverse Gaussian (GIG) distribution i.e.,

$$G|\varepsilon, \theta \ \sim \ \text{GIG}(-1, \delta\sqrt{\phi(\epsilon)}, \alpha).$$

The conditional first moment and inverse first moment are as follows:

$$\mathbb{E}(G|\epsilon) = \frac{\delta\phi(\epsilon)^{1/2}}{\alpha}\frac{K_0(\alpha\delta\phi(\epsilon)^{1/2})}{K_1(\alpha\delta\phi(\epsilon)^{1/2})},$$
$$\mathbb{E}(G^{-1}|\epsilon) = \frac{\alpha}{\delta\phi(\epsilon)^{1/2}}\frac{K_{-2}(\alpha\delta\phi(\epsilon)^{1/2})}{K_{-1}(\alpha\delta\phi(\epsilon)^{1/2})}.$$

These first order moments will be used in calculating the conditional expectation of the log-likelihood function. The complete data likelihood is given by

$$\begin{aligned} L(\theta) &= \prod_{t=1}^{N} f(\epsilon_t, g_t) = \prod_{t=1}^{N} f_{\varepsilon|G}(\epsilon_t|g_t) f_G(g_t) \\ &= \prod_{t=1}^{N} \frac{\delta}{2\pi g_t^2} \exp(\delta\gamma) \exp\left(-\frac{\delta^2}{2g_t} - \frac{\gamma^2 g_t}{2} - \frac{g_t^{-1}}{2}(\epsilon_t - \mu)^2 - \frac{\beta^2 g_t}{2} + \beta(\epsilon_t - \mu)\right). \end{aligned}$$

The log likelihood function will be

$$\begin{aligned} l(\theta) &= N\log(\delta) - N\log(2\pi) + N\delta\gamma - N\beta\mu - 2\sum_{t=1}^{N}\log(g_t) - \frac{\delta^2}{2}\sum_{t=1}^{N} g_t^{-1} \\ &\quad - \frac{\gamma^2}{2}\sum_{t=1}^{N} g_t - \frac{1}{2}\sum_{t=1}^{N} g_t^{-1}(\epsilon_t - \mu)^2 - \frac{\beta^2}{2}\sum_{t=1}^{N} g_t + \beta\sum_{t=1}^{N}\epsilon_t. \end{aligned}$$

Now in *E-step* of EM algorithm, we need to compute the expected value of complete data log likelihood known as $Q(\theta|\theta^k)$, which is expressed as

$$\begin{aligned} Q(\theta|\theta^{(k)}) &= \mathbb{E}_{G|\varepsilon,\theta^{(k)}}[\log f(\varepsilon, G|\theta)|\varepsilon, \theta^{(k)}] = \mathbb{E}_{G|\varepsilon,\theta^{(k)}}[L(\theta|\theta^{(k)})] \\ &= N\log\delta + N\delta\gamma - N\beta\mu - N\log(2\pi) - 2\sum_{t=1}^{N}\mathbb{E}(\log g_t|\epsilon_t, \theta^{(k)}) - \frac{\delta^2}{2}\sum_{t=1}^{N} w_t \\ &\quad - \frac{\gamma^2}{2}\sum_{t=1}^{N} s_t - \frac{\beta^2}{2}\sum_{t=1}^{N} s_t + \beta\sum_{t=1}^{N}(y_t - \rho^T\mathbf{Y_{t-1}}) - \frac{1}{2}\sum_{t=1}^{N}(y_t - \rho^T\mathbf{Y_{t-1}} - \mu)^2 w_t, \end{aligned}$$

where, $s_t = \mathbb{E}_{G|\varepsilon,\theta^{(k)}}(g_t|\epsilon_t, \theta^{(k)})$ and $w_t = \mathbb{E}_{G|\varepsilon,\theta^{(k)}}(g_t^{-1}|\epsilon_t, \theta^{(k)})$. Update the

parameters by maximizing the Q function using the following equations

$$\begin{aligned}
\frac{\partial Q}{\partial \rho} &= \sum_{t=1}^{N} w_t(y_t - \rho^T \mathbf{Y_{t-1}} - \mu)Y_{t-1}^T - \beta \sum_{t=1}^{N} Y_{t-1}^T, \\
\frac{\partial Q}{\partial \mu} &= -N\beta + \sum_{t=1}^{N} w_t(\epsilon_t - \mu), \\
\frac{\partial Q}{\partial \beta} &= -N\mu + \sum_{t=1}^{N} y_t - \beta \sum_{t=1}^{N} s_t - \sum_{t=1}^{N} \rho^T \mathbf{Y_{t-1}}, \\
\frac{\partial Q}{\partial \delta} &= N\gamma + \frac{N}{\delta} - \delta \sum_{t=1}^{N} w_t, \\
\frac{\partial Q}{\partial \gamma} &= N\delta - \gamma \sum_{t=1}^{N} s_t.
\end{aligned}$$

Solving the above equations, we obtain the following estimates of the parameters

$$\hat{\rho} = \left(\sum_{t=1}^{N} w_t Y_{t-1} Y_{t-1}^T \right)^{-1} \sum_{t=1}^{N} (w_t y_t - \mu w_t - \beta) Y_{t-1}$$

$$\begin{aligned}
\hat{\mu} &= \frac{-N\beta + \sum_{t=1}^{N} \epsilon_t w_t}{N\bar{w}_t}; \\
\hat{\beta} &= \frac{\sum_{t=1}^{N} w_t \epsilon_t - N\bar{w}_t \bar{\epsilon}_t}{N(1 - \bar{s}_t \bar{w}_t)}; \\
\hat{\delta} &= \sqrt{\frac{\bar{s}}{(\bar{s}\bar{w} - 1)}}, \quad \hat{\gamma} = \frac{\delta}{\bar{s}}, \text{ and } \hat{\alpha} = (\gamma^2 + \beta^2)^{1/2},
\end{aligned} \tag{3.6}$$

where $\bar{s} = \frac{1}{N}\sum_{t=1}^{N} s_t, \bar{w} = \frac{1}{N}\sum_{t=1}^{N} w_t$. □

The following result is a classical result also discussed in [60], so we are providing the proof for the convenience of the readers.

Proposition 3.2. *For the AR(1) time-series defined $Y_t = \rho Y_{t-1} + \epsilon_t$, we have*

$$\mathbb{E}(Y_t) = \frac{\mu\gamma + \delta\beta}{\gamma} \left(\frac{1 - \rho^{t+1}}{1 - \rho} \right); \quad \text{Var}(Y_t) = \frac{\delta\alpha^2}{\gamma^3} \left(\frac{1 - \rho^{2n+2}}{1 - \rho^2} \right).$$

Proof. For $\varepsilon_t \sim NIG(\alpha, \beta, \mu, \delta)$, we have $\mathbb{E}(\varepsilon_t) = \frac{\mu\gamma + \delta\beta}{\gamma}$ and $\text{Var}(\varepsilon_t) = \frac{\delta\alpha^2}{\gamma^3}$. For an

AR(1) time series, it follows

$$\begin{aligned}\mathbb{E}(Y_t) &= \rho\mathbb{E}(Y_{t-1}) + \mathbb{E}\epsilon_t = \mathbb{E}\epsilon_t + \rho\mathbb{E}\epsilon_{t-1} + \rho^2\mathbb{E}\epsilon_{t-2} + \cdots + \rho^t\mathbb{E}\epsilon_1 \\ &= \mathbb{E}(\epsilon_1)\left(\frac{1-\rho^{t+1}}{1-\rho}\right) = \frac{\mu\gamma + \delta\beta}{\gamma}\left(\frac{1-\rho^{t+1}}{1-\rho}\right).\end{aligned}$$

Using law of total variance to calculate $\mathrm{Var}(Y_t)$, yields to

$$\begin{aligned}\mathrm{Var}(Y_t) &= \mathbb{E}[\mathrm{Var}(Y_t|Y_{t-1} = y_{t-1})] + \mathrm{Var}[\mathbb{E}(Y_t|Y_{t-1} = y_{t-1})] \\ &= \mathbb{E}[\mathrm{Var}(\rho Y_{t-1} + \varepsilon_t|Y_{t-1} = y_{t-1})] + \mathrm{Var}[\mathbb{E}(\rho Y_{t-1} + \varepsilon_t|Y_{t-1} = y_{t-1})] \\ &= \mathbb{E}[\mathrm{Var}(\varepsilon_t|Y_{t-1} = y_{t-1})] + \mathrm{Var}[\mathbb{E}(\varepsilon_t) + \rho Y_{t-1}] \\ &= \mathrm{Var}(\varepsilon_t) + \mathrm{Var}(\rho Y_{t-1}) = \mathrm{Var}(\varepsilon_t) + \rho^2\mathrm{Var}(Y_{t-1}).\end{aligned}$$

Recursively using the above relation and since ε_t are i.i.d. we can write,

$$\begin{aligned}\mathrm{Var(Y_t)} &= \mathrm{Var}(\varepsilon_t) + (\rho^2 + \rho^4 + ... + \rho^{2t})\mathrm{Var}(\varepsilon_t) \\ &= (1 + \rho^2 + \rho^4 + ... + \rho^{2n})\mathrm{Var}(\varepsilon_t) \\ &= \frac{1-\rho^{2t-2}}{1-\rho^2}\mathrm{Var}(\varepsilon_t) = \frac{\delta\alpha^2}{\gamma^3}\left(\frac{1-\rho^{2t-2}}{1-\rho^2}\right).\end{aligned}$$

□

Similarly, we obtain the mean and variance of AR(p) series in the following way. We have $Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \cdots + \rho_p Y_{t-p} + \epsilon_t$. Assuming stationarity, it follows that

$$\mathbb{E}Y_t = \frac{\mathbb{E}(\epsilon_t)}{1 - \rho_1 - \rho_2 - \cdots - \rho_p} = \frac{\mu + \delta\beta/\gamma}{1 - \rho_1 - \rho_2 - \cdots - \rho_p}.$$

Further, for $\mu = \beta = 0$, we have $\mathbb{E}[Y_t] = \mathbb{E}[\epsilon_t] = 0$ and

$$\begin{aligned}Y_t^2 &= \rho_1 Y_{t-1}Y_t + \rho_2 Y_{t-2}Y_t + \cdots + \rho_p Y_{t-p}Y_t + \epsilon_t Y_t. \\ \mathbb{E}[Y_t^2] &= \rho_1\mathbb{E}[Y_{t-1}Y_t] + \rho_2\mathbb{E}[Y_{t-2}Y_t] + \cdots + \rho_p\mathbb{E}[Y_{t-p}Y_t] + \mathbb{E}[\epsilon_t Y_t] \\ &= \rho_1\mathbb{E}[Y_{t-1}Y_t] + \rho_2\mathbb{E}[Y_{t-2}Y_t] + \cdots + \rho_p\mathbb{E}[Y_{t-p}Y_t] + \mathbb{E}[\epsilon_t^2]. \\ \mathrm{Var[Y_t]} &= \sigma_\epsilon^2 + \sum_{j=1}^{p}\rho_j\gamma_j, \qquad (3.7)\end{aligned}$$

where $\sigma_\epsilon^2 = \mathrm{Var}[\epsilon_t]$ and $\gamma_j = \mathbb{E}[Y_t Y_{t-j}]$. Moreover,

$$\gamma_j = \rho_1\gamma_{j-1} + \rho_2\gamma_{j-2} + \cdots + \rho_p\gamma_{j-p},\ j \geq 1. \qquad (3.8)$$

For $p = 2$, using (3.7) and (3.8), it is easy to show that

$$\mathrm{Var}[\mathrm{Y_t}] = \frac{(1-\rho_2)\sigma_\epsilon^2}{1-\rho_2-\rho_1^2-\rho_2^2-\rho_1^2\rho_2+\rho_2^3},$$

where $\sigma_\epsilon^2 = \delta\alpha^2/\gamma^3$. Again for $p = 3$, using (3.7) and (3.8), it follows

$$\mathrm{Var}[\mathrm{Y_t}] = \frac{(1-\rho_2-\rho_1\rho_3-\rho_3^2)\sigma_\epsilon^2}{1-\rho_2-\rho_1\rho_3-4\rho_1\rho_2\rho_3-\rho_1^2(1+\rho_2+\rho_3^2+\rho_1\rho_3)-\rho_2^2(1+\rho_3^2-\rho_2-\rho_1\rho_2)-\rho_3^2(2-\rho_1\rho_3^2-\rho_2-\rho_3^2)}.$$

3.3 Simulation study and applications

In this section, we illustrate the performance of the proposed model and the introduced estimation technique using simulated data sets and real time series of NASDAQ stock exchange data.

3.3.1 Simulation study

We discuss the estimation procedure for AR(2) and AR(1) models. The model defined by Eq. (3.4) is simulated in two steps. In the first step, the NIG residuals are simulated using the normal variance-mean mixture form defined in Eq. (3.2). For NIG random numbers, standard normal and IG random numbers are required. The algorithm mentioned in [41] is used to generate i.i.d. IG distributed random numbers $G_i \sim IG(\mu_1, \lambda_1),\ i = 1, 2, \cdots, N$ using the following steps:

Step 1: Generate standard normal variate Z and set $Y = Z^2$.

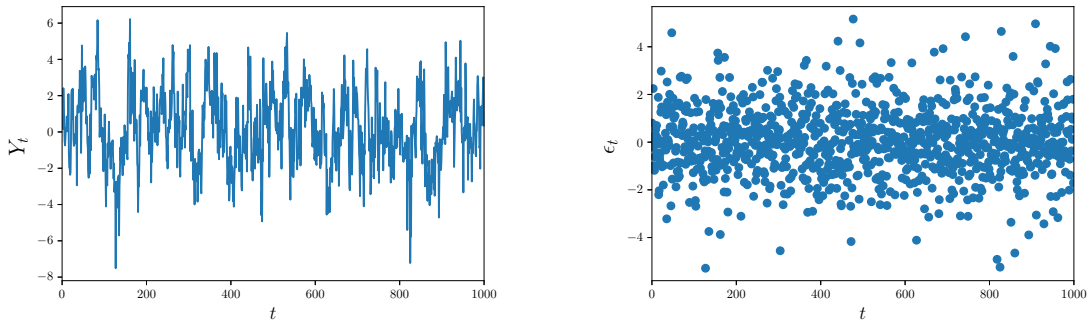
Step 2: Set $X_1 = \mu_1 + \frac{\mu_1^2 Y}{2\lambda_1} - \frac{\mu_1}{2\lambda_1}\sqrt{4\mu_1\lambda_1 Y + \mu_1^2 Y^2}$.

Step 3: Generate uniform random variate $U[0, 1]$.

Step 4: If $U \leq \frac{\mu_1}{\mu_1+X_1}$, then $G = X_1$; else $G = \frac{\mu_1^2}{X_1}$.

Note that the substitutions for parameters as $\mu_1 = \delta/\gamma$ and $\lambda_1 = \delta^2$ are required in the above algorithm because the pdf taken in [41] is different from the form given in Eq. (3.3). Again we simulate a standard normal vector of size N and use the relation defined in Eq. (3.2) with simulated IG random numbers to obtain the NIG random numbers of size N. In step 2, the simulated NIG residuals and the relation given in Eq. (3.4) are used to generate AR(p) simulated series.

Case 1: In the simulation study, first we analyze the AR(2) model with NIG residuals. In the analysis we used 1000 trajectories of length $N = 1000$ each. The used parameters of the model are: $\rho_1 = 0.5$ and $\rho_2 = 0.3$ while the residuals were



(a) The time series data plot. (b) Scatter plot of residual term.

Figure 3.1: The exemplary time series of length $N = 1000$ (left panel) and the corresponding residual term (right panel) of the AR(2) model with NIG distribution. The parameters of the model are: $\rho_1 = 0.5, \rho_2 = 0.3, \alpha = 1, \beta = 0, \mu = 0,$ and $\delta = 2$.

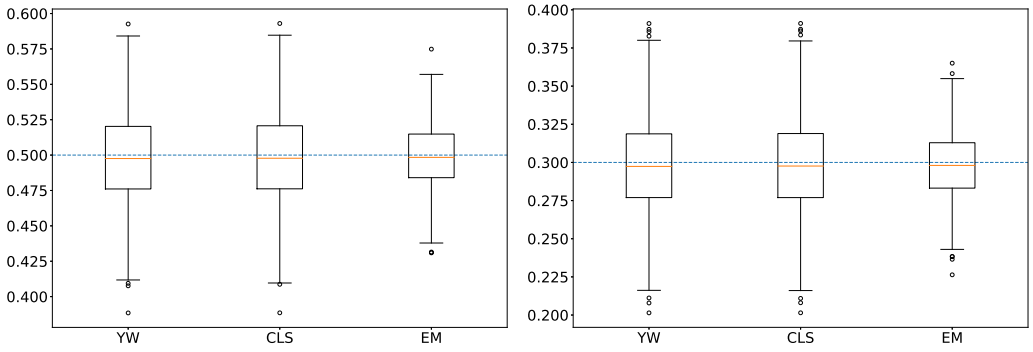
generated from NIG distribution with $\alpha = 1$, $\beta = 0$, $\mu = 0$ and $\delta = 2$. The exemplary time series data plot and scatter plot of residual terms are shown in Fig. 3.1.

Now, for each simulated trajectory, we apply the estimation algorithm presented in the previous section. For EM algorithm several stopping criteria could be used. One of the examples is the criterion based on change in the log-likelihood function which utilizes the relative change in the parameters' values. We terminate the algorithm when the following commonly used criterion for the relative change in the parameters' values is satisfied

$$\max\left\{\left|\frac{\alpha^{(k+1)}-\alpha^{(k)}}{\alpha^{(k)}}\right|, \left|\frac{\delta^{(k+1)}-\delta^{(k)}}{\delta^{(k)}}\right|, \left|\frac{\rho_1^{(k+1)}-\rho_1^{(k)}}{\rho_1^{(k)}}\right|, \left|\frac{\rho_2^{(k+1)}-\rho_2^{(k)}}{\rho_2^{(k)}}\right|\right\} < 10^{-4}. \quad (3.9)$$

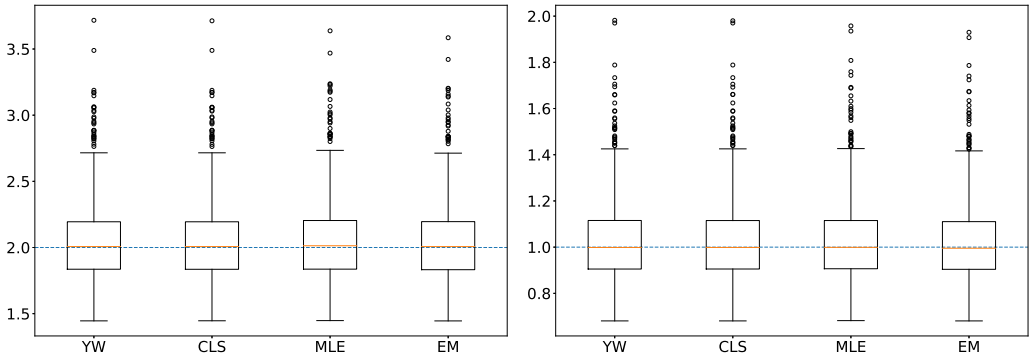
The parameters' estimates obtained from the simulated data are shown in the boxplot in Fig. 3.2. Moreover, we compared the estimation results with the classical YW algorithm, CLS method for model parameters. We remind that the YW algorithm is based on the YW equations calculated for the considered model, and utilizes the empirical autocovariance function for the analyzed data. More details of YW algorithm for autoregressive models can be found, for instance in [27] and 2.9. We remind, the CLS method estimates the model parameters for dependent observations by minimizing the sum of squares of deviations about the conditional expectation (refer 2.10). Fig. 3.2(a) and Fig. 3.2(b) represent the estimates of the model parameters ρ_1 and ρ_2 using YW, CLS and EM methods, respectively. Furthermore, using the estimated ρ_1 and ρ_2 parameters with YW and CLS methods the residuals or residual terms are obtained and then again EM algorithm is used to estimate the remaining α and δ parameters which are plotted in Fig. 3.2(c) and 3.2(d). Moreover, the estimates for α and δ using EM algorithm given in

(3.5) and the MLE by Newton-Raphson of simulated NIG residuals are also plotted in Fig. 3.2(c) and 3.2(d) for comparison. From boxplots presented in Fig. 3.2 we observe that the estimates of ρ_1 and ρ_2 parameters using the EM algorithm have less variance in comparison to the YW and CLS algorithms. Moreover, for α and δ parameters, we see that the means of the estimates for the three presented methods are close to the true values, but the range of outliers for the EM algorithm is comparatively less. Also, observe that the MLE of α and δ parameters by EM and Newton-Raphson method have the similar behavior with lesser outliers by EM as both iterative methods maximize the likelihood function to find the estimates. The convergence of EM algorithm is tested on simulated data of different sample size and results are shown in Table 3.1. The sensitivity of EM algorithm to initial guess is shown in Table 3.2 and we conclude that the proposed method is robust to initial guess.



(a) Boxplots for ρ_1 estimates.

(b) Boxplots for ρ_2 estimates.



(c) Boxplots for δ estimates.

(d) Boxplots for α estimates.

Figure 3.2: Boxplots of the estimates of the AR(2) model's parameters with theoretical values: $\rho_1 = 0.5, \rho_2 = 0.3, \delta = 2$ and $\alpha = 1$ represented with blue dotted lines. The boxplots are created using 1000 trajectories each of length 1000.

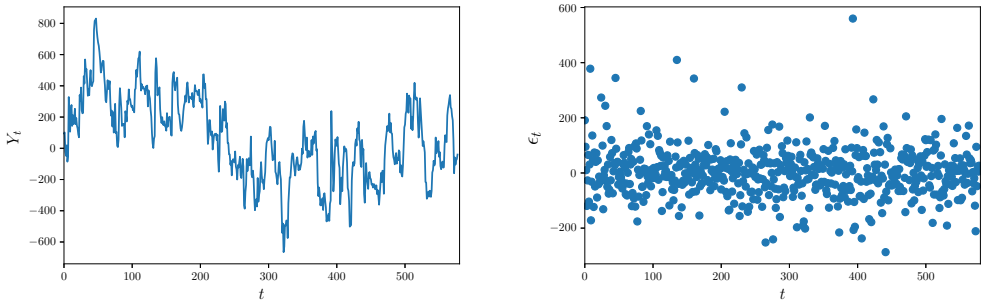
Table 3.1: Estimates of parameters with different sample size generated from AR(2) model. The initial guess used for EM algorithm is $\rho_1 = 0.1, \rho_2 = 0.1, \delta = 0.1$ and $\alpha = 0.1$. The theoretical values of the parameters are: $\rho_1 = 0.5$, $\rho_2 = 0.3$, $\delta = 2$ and $\alpha = 1$.

Parameters / Sample size	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\delta}$	$\hat{\alpha}$
100	0.399	0.328	1.961	0.952
250	0.447	0.279	2.291	1.218
500	0.499	0.272	1.934	0.952
750	0.499	0.291	1.863	0.904
1000	0.511	0.294	2.008	1.104

Table 3.2: Estimates of parameters (EP) with different initial guess (IG) using EM algorithm.

Parameters	ρ_1	ρ_2	δ	α
True values	0.5	0.3	2	1
IG 1	0.001	0.001	0.005	0.005
EP 1	0.503	0.296	2.154	1.092
IG 2	0.1	0.1	0.5	0.5
EP 2	0.503	0.296	2.154	1.092
IG 3	0.1	0.1	0.001	0.001
EP 3	0.503	0.296	2.154	1.092
IG 4	0.1	0.001	0.1	0.05
EP 4	0.503	0.296	2.154	1.092
IG 5	0.001	0.001	0.5	0.5
EP 5	0.503	0.296	2.154	1.092
IG 6	2	2	5	2
EP 6	0.503	0.296	2.172	1.102

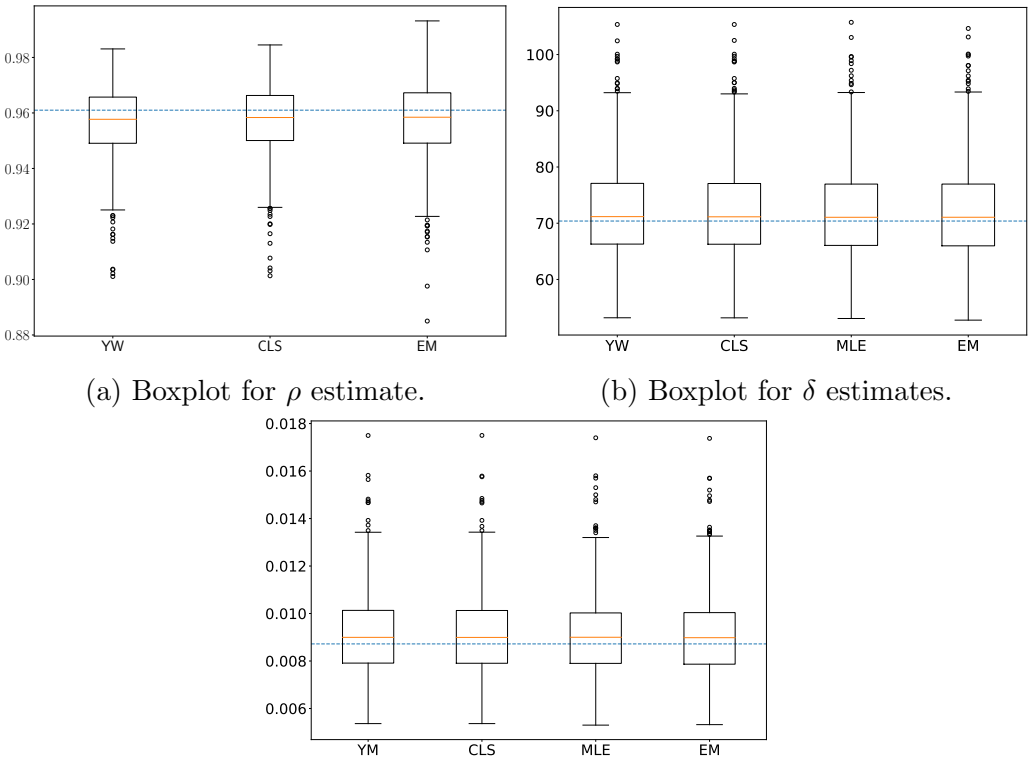
Case 2: As the second example, we analyze the AR(1) model with NIG residuals. Here we examine the trajectories of 579 data points. The same number of data points are examined in the real data analysis demonstrated in the next subsection. This exemplary model is discussed to verify the results for the segmented data 2 from the NASDAQ stock exchange data. The simulated errors follow NIG distribution with parameters $\alpha = 0.0087$, $\beta = 0$, $\mu = 0$ and $\delta = 70.3882$ while the model's parameter is $\rho = 0.9610$. In Fig. 3.3, we present the exemplary simulated trajectory and the corresponding residual terms.



(a) Time series data plot. (b) Scatter plot of residual terms.

Figure 3.3: The exemplary time series plot of the first trajectory of length $N = 579$ from AR(1) model (left panel) with the corresponding scatter plot of residual terms NIG distribution (right panel). The parameters of the model are $\rho = 0.961, \alpha = 0.0087, \beta = 0, \mu = 0$ and $\delta = 70.3882$.

Similarly as in Case 1, the introduced EM algorithm was applied to the simulated data with the stopping criteria based on the relative change in the parameter values defined in Eq. (3.9). The boxplots of the estimated parameters for 1000 trajectories each of length 579 are shown in Fig. 3.4. Similar as in the previous case, we compare the results for EM, YW and CLS algorithms.



(a) Boxplot for ρ estimate. (b) Boxplot for δ estimates.

(c) Boxplots for α estimates.

Figure 3.4: Boxplots of the estimates of the AR(1) model's parameters with theoretical values: $\rho = 0.9610, \delta = 70.3883$ and $\alpha = 0.00872$ represented with blue dotted lines. The boxplots are created using 1000 trajectories each of length 579.

From Fig. 3.4 one can observe that although the estimate of ρ has more variance compared to YW and CLS methods, but the estimates δ and α have less variance and the spread of outliers is also slightly less when compared to MLE by Newton-Raphson method. The means of the estimated parameters from 1000 trajectories of length $N = 579$ using EM algorithm are $\hat{\rho} = 0.9572$, $\hat{\delta} = 71.8647$ and $\hat{\alpha} = 0.0091$. We can conclude that the EM algorithm, also in this case, gives the better parameters' estimates for the considered model.

3.3.2 Real data applications

In this section, we assess the proposed model on three different datasets namely, Google equity price, NASDAQ index data and US gasoline price data.

Google equity price

The historical financial data of Google equity is collected from Yahoo finance [1]. The whole data set covers the period from December 31, 2014 to April 30, 2021. Initially, the data contained 1594 data points with 6 features having Google stock's *open price*, *closing price*, *highest value*, *lowest value*, *adjusted close price* and *volume* of each working day end-of-the-day values. In order to apply the proposed NIGAR(1) model, we take the univariate time series y_t to be the end-of-the-day closing prices. The Google equity closing price is demonstrated as time series data in Fig. 3.5.

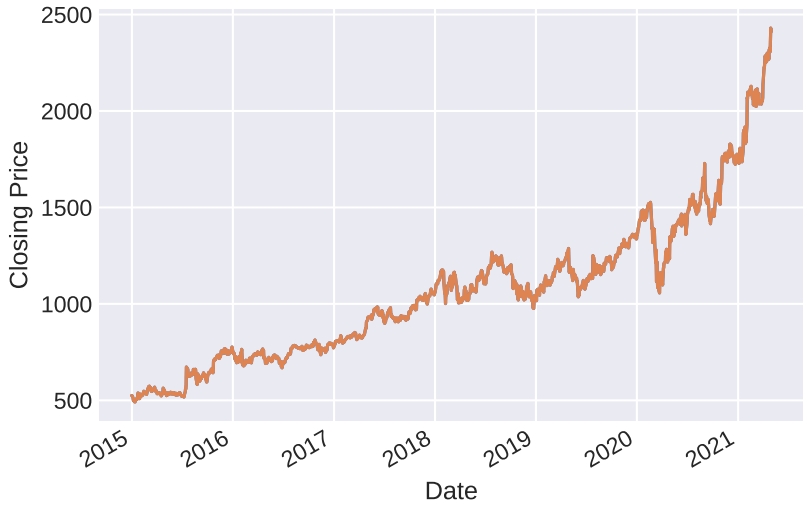
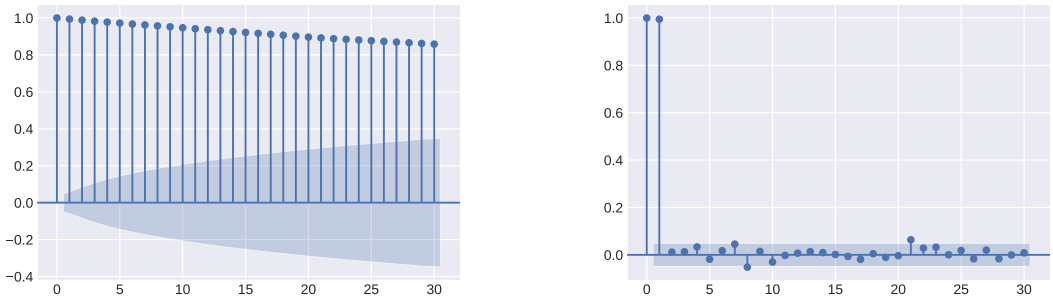


Figure 3.5: The closing price (in$) of Google equity.

We assume that the innovation terms ε_t of time series data y_t follows NIG. We use ACF and PACF plot to determine the appropriate time series model components

[1] https://finance.yahoo.com/quote/GOOGL?p=GOOGL&.tsrc=fin-srch

for closing price. Fig. 3.6 shows the ACF and PACF plot of time-series data y_t. We observe that in PACF plot there is a significant spike at lag 1, also we ignore



(a) ACF Plot. (b) PACF Plot.

Figure 3.6: The ACF and PACF plot of closing price of Google equity.

the spike at lag 0 as it represents the correlation between the term itself which will always be 0. PACF plot indicates that the closing prices follow AR model with lag 1. In ACF plot, all the spikes are significant for lags upto 30 which implies that the closing price is highly correlated. Therefore, from ACF and PACF plot the assumed NIGAR(1) model is expected to be good fit for data. First we estimate the ρ parameter using the conditional least square method as mentioned in (2.8) and obtain,

$$\hat{\rho} = \frac{\sum_{t=0}^{n}(y_t - \overline{y})(y_{t+1} - \overline{y})}{\sum_{t=0}^{n}(y_t - \overline{y})^2}.$$

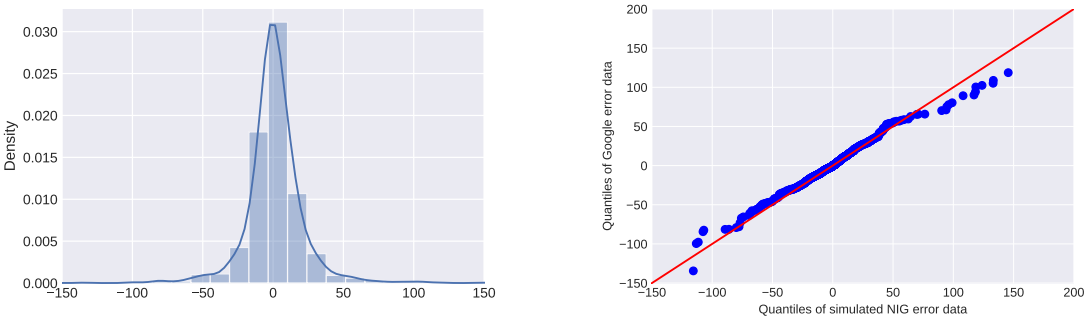
The estimated value is $\hat{\rho} = 0.9941$. Using the estimated value of ρ and relation $\varepsilon_t = y_t - \rho y_{t-1}$ we will get the innovation terms ε_t. The distribution plot of innovation terms is shown in Fig. 3.7.

The Kolmogorov-Smirnov (KS) normality test and Jarque–Bera (JB) test are performed on ε_t to test if the innovation terms are Gaussian. The p-value in both the tests was 0, which indicates that the ε_t may not be from Gaussian distribution. Therefore, we fit the proposed NIGAR(1) model on Google closing price dataset. The ML estimates of parameters using EM algorithm are $\hat{\alpha} = 0.0202, \hat{\beta} = 0.0013, \hat{\mu} = 0.226, \hat{\delta} = 9.365$, and $\hat{\gamma} = 0.0201$ with initial guesses as $\hat{\alpha}^{(0)} = 0.0141, \hat{\beta}^{(0)} = 0.01, \hat{\mu}^{(0)} = 0.01, \hat{\delta}^{(0)} = 0.01$, and $\hat{\gamma}^{(0)} = 0.01$. The relative change in the log-likelihood value with tolerance value 0.0001 is used as stopping criterion. It is worthwhile to mention that the estimated β is close to 0 and the estimated μ can be interpreted as intercept term in the AR(1) model. Based on the Google equity prices data and estimated parameters, an equivalent model to

$Y_t = \rho Y_{t-1} + \epsilon_t$ can be described as

$$Y_t = \mu + \rho Y_{t-1} + \epsilon_t,$$

where $\epsilon_t = \sqrt{G}Z$ with $G \sim \text{IG}(\gamma, \delta)$, with $\hat{\rho} = 0.9941, \hat{\mu} = 0.226, \hat{\delta} = 9.365,$ and $\hat{\gamma} = 0.0201$.



(a) Distribution of error terms.

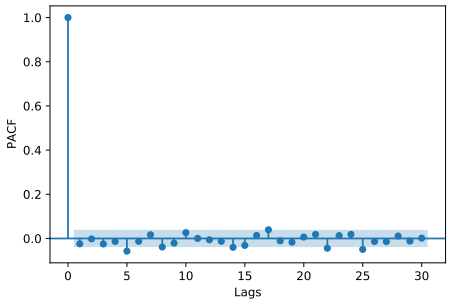
(b) QQ Plot.

Figure 3.7: Plot of error terms distribution and QQ plot between simulated and actual values.

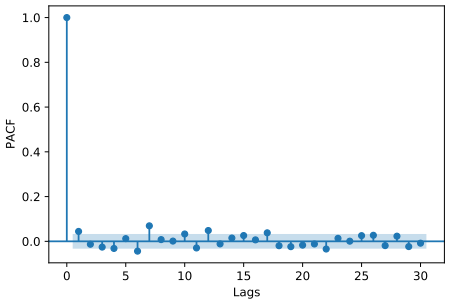
To test the reliability of the estimated results, we used the simulated data with true parameter values as $\alpha = 0.02, \beta = 0, \mu = 0.23, \delta = 9.5,$ and $\gamma = 0.02$. The QQ-plot (Fig. 5(b)) between the error terms ε of data and the NIG simulated data indicates that the innovation terms are significantly NIG with few outliers. Also, we performed the 2-sample KS test which is used to check whether the two samples are from same distribution. The 2-sample KS test on the simulated data and the innovation terms resulted in the p-value= 0.3413. Therefore, we failed to reject the null hypothesis that the samples are from same distribution and conclude that both samples follow the same distribution i.e. $\text{NIG}(\alpha = 0.02, \beta = 0, \mu = 0.23, \delta = 9.5)$.

NASDAQ index data

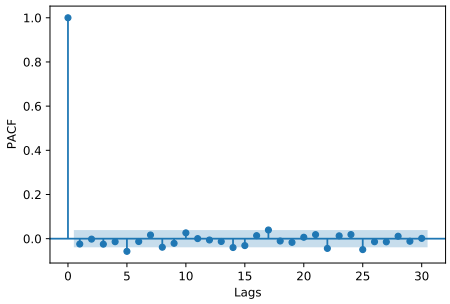
Initially we explored the four different stock market indices namely NSEI, Nikkei225, S&P/ASX200 and NASDAQ. The PACF plots of the log returns for these indices are shown in Fig. 3.8. From the charts, it is evident that there is no significant correlation across different lags of the log returns data. The empirical study on log returns of financial data also says that there may be slow decay in autocorrelation of absolute or squared log returns but there it is absent in log returns (see Section 7.1, [37]). In the following subsection, we apply the introduced model on the adjusted closed price of the NASDAQ index, which is one of these four indices.



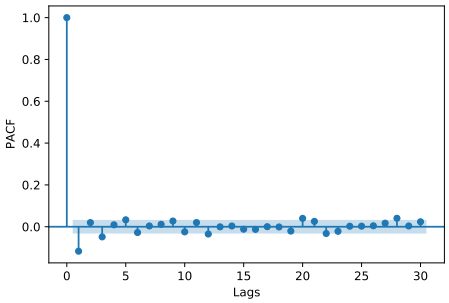
(a) PACF plot for Nasdaq data.



(b) PACF plot for NSEI data.



(c) PACF plot for Nikkei 225 data.



(d) PACF plot for Australia S&P/ASX data.

Figure 3.8: The PACF plots of the log returns of four different stock market indices.

In this part, the considered AR(p) model with NIG distribution is applied to the NASDAQ stock market index data, which is available on Yahoo finance [2]. It covers the historical prices and volume of all stocks listed on NASDAQ stock exchange from the period March 04, 2010 to March 03, 2020. The data consists of 2517 data points with features having *open price*, *closing price*, *highest value*, *lowest value*, *adjusted closing price* and *volume* of stocks for each working day end-of-the-day values. We choose the end-of-the-day adjusted closing price as a univariate time series for the analysis purpose. The residual terms of time series data is assumed to follow NIG distribution as the general one. In Fig. 3.9 we represent the adjusted closing price of NASDAQ index. Observe that the original time series data has an increasing trend. Moreover, one can easily observe that the data exhibit non-homogeneous behavior. Thus, before further analysis the analyzed time series should be segmented in order to obtain the homogeneous parts. To divide the vector of observations into homogeneous parts, we applied the segmentation algorithm presented in [58], where authors proposed to use the statistics defined as the cumulative sum of squares of the data. Finally, the segmentation algorithm is based on the specific behavior of the used statistics when the structure change point exists in the analyzed time

[2] https://finance.yahoo.com/quote/%5EIXIC/history?period1=1267660800&period2=1583193600&interval=1dfilter=history&frequency=1d&includeAdjustedClose=true

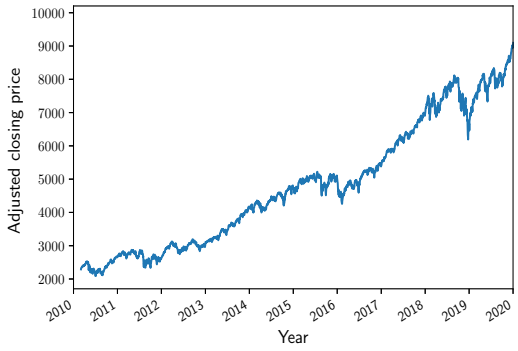
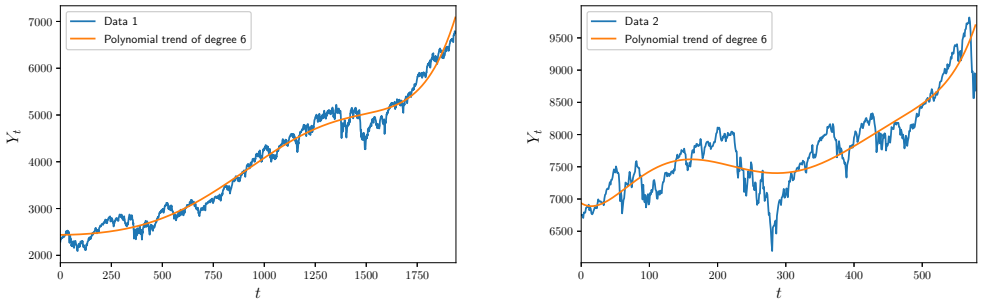


Figure 3.9: The adjusted closing price (in$) of NASDAQ index from the period March 04, 2010 to March 03, 2020 with 2517 data points.

series. More precisely, in [58] it was shown that the cumulative sum of squares is a piece-wise linear function when the variance of the data changes. Because in the considered time series we observe the non-stationary behavior resulting from the existence of the deterministic trend, thus, to find the structure break point, we applied the segmentation algorithm for their logarithmic returns. Finally, the algorithm indicates that the data needs to be divided into two segments, the first 1937 observations are considered as data 1 and rest all observations as - data 2.



(a) Data 1. (b) Data 2.

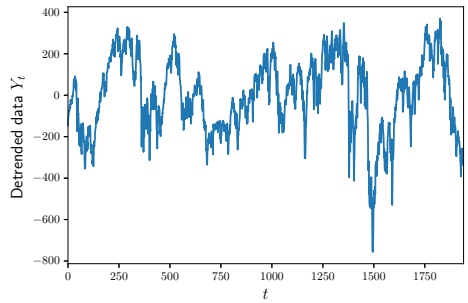
Figure 3.10: The segmented data 1 (left panel) and data 2 (right panel) together with the fitted polynomials.

The trends for both data sets were removed by using the degree 6 polynomial detrending. The trend was fitted by using the least squares method. The original data sets with the fitted polynomials are shown in Fig. 3.10. Next, for data 1 and data 2 we analyze the detrending time series and for each of them we use the partial autocorrelation function (PACF) to recognize the proper order of AR model. It is worth mentioning the PACF is a common tool to find the optimal order of the autoregressive models [27]. We select the best order that is equal to the lag corresponding to the largest PACF value (except a lag equal to zero).

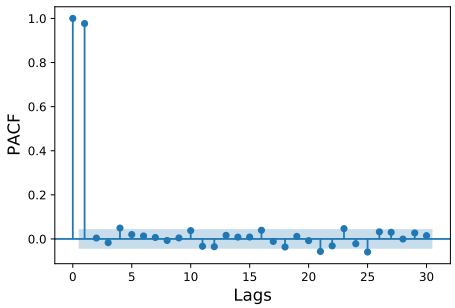
Table 3.3: Estimated parameters of data 1 and data 2 using EM algorithm.

	$\hat{\rho}$	$\hat{\delta}$	$\hat{\alpha}$
Data 1	0.9809	34.5837	0.0226
Data 2	0.9610	70.3883	0.0087

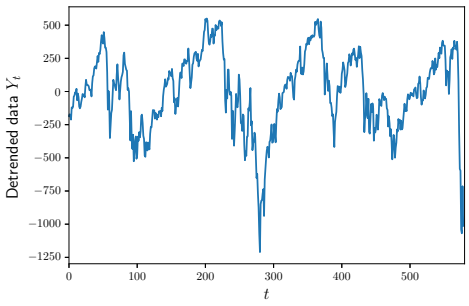
We use the PACF plots to determine the components of AR(p) model. Fig. 3.11 shows the stationary data (after removing the trend) and corresponding PACF plots indicating the optimal model - AR(1). After above-described pre-processing steps,



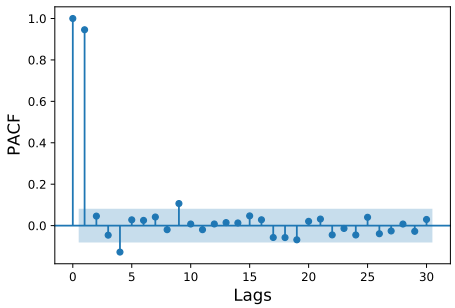
(a) Stationary data 1.



(b) PACF of data 1.



(c) Stationary data 2.



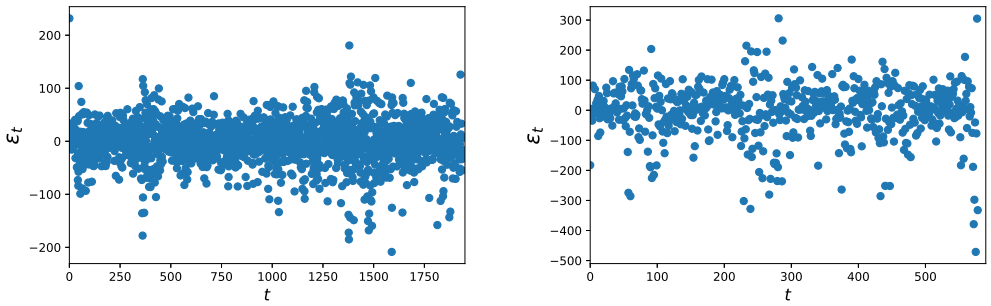
(d) PACF of data 2.

Figure 3.11: The time-series plot of stationary data 1 and data 2 (after removing the trend) - left panel and corresponding PACF plot - right panel.

the EM algorithm is used to estimate the model's parameters. The mean of NIG distribution is $\mathbb{E}X = \mu + \delta\frac{\beta}{\gamma}$. In the proposed model, we fit the data to symmetric distribution therefore, we fixed the parameters to $\mu = 0$ and $\beta = 0$. Although EM algorithm can be applied to asymmetric NIG distribution also, we tried the same on simulated data but did not include in this work as we need to compare the proposed method with Yule-Walker and conditional least squares method. The estimated values of parameters for data 1 and data 2 are summarised in Table 3.3.

As the final step, we analyze the residual terms corresponding to data 1 and data 2 to confirm they can be modeled by using NIG distribution. The scatter plots of

the residuals from the AR(1) model corresponding to data 1 and data 2 are shown in Fig. 3.12. The Kolmogorov-Smirnov (KS) test is used to check the normality of the residuals corresponding to data 1 and data 2. For details of KS tests refer Section 2.4. First, we use the one-sample KS test to reject the hypothesis of normal distribution of the residuals. The p-value of the KS test is 0 for both cases, indicating that the null hypothesis (normal distribution) is rejected for both series. Thus, we applied two-sample KS test for both residual terms with the null hypothesis of NIG distribution. The tested NIG distributions have the following parameters: $\mu = 0,\ \beta = 0,\ \delta = 34.5$ and $\alpha = 0.02$ - for residuals corresponding to data 1; and $\mu = 0,\ \beta = 0,\ \delta = 70.5$ and $\alpha = 0.008$ - for the residuals corresponding to data 2. The p-values for 2-sample KS test are 0.565 and 0.378 for data 1 and data 2, respectively, which indicates that there is no evidence to reject the null hypothesis. Therefore, we assume that both the residual series follow the same distribution, implying that data 1 follows NIG($\alpha = 0.02,\ \beta = 0,\ \mu = 0,\ \delta = 34.5$) and data 2 has NIG($\alpha = 0.008,\ \beta = 0,\ \mu = 0,\ \delta = 70.5$).

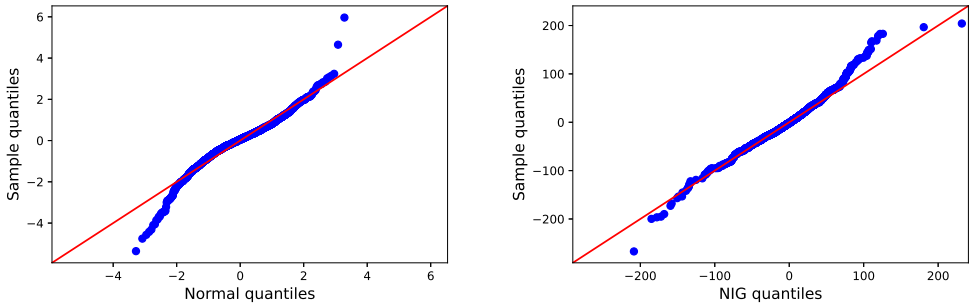


(a) Scatter plot of residuals of data 1 (b) Scatter plot of residuals of data 2.

Figure 3.12: The scatter plots of residuals of data 1 from AR(1) model with NIG($\alpha = 0.02, \beta = 0, \mu = 0, \delta = 34.5$) distribution (left panel) and data 2 with NIG($\alpha = 0.008, \beta = 0, \mu = 0, \delta = 70.5$ (right panel).

To confirm that NIG distributions (with fitted parameters) are acceptable for the residual series, in Fig. 3.13 and Fig. 3.14 we demonstrate the QQ plot for the residuals of AR(1) models for data 1 and data 2 and the simulated data from normal (left panels) and corresponding NIG distributions (right panels). Observe that the tail of both data 1 and data 2 deviates from the red line on the left panels, which indicates that the data does not follow the distribution.

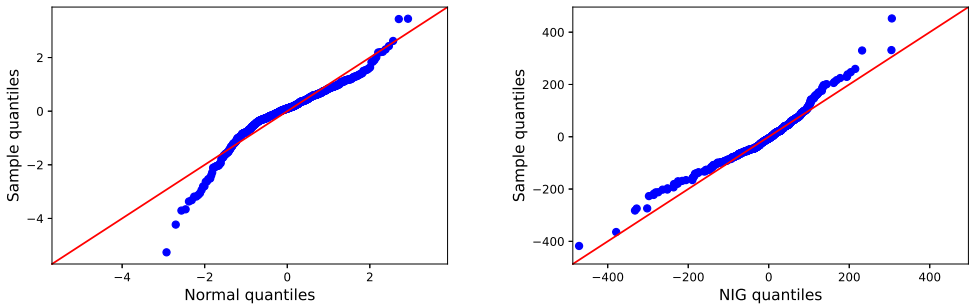
The correspondence with NIG distribution is also demonstrated in Fig. 3.15, where the kernel density estimation (KDE) plot is presented for the residual series and compared with the pdf of the normal and corresponding NIG distributions. A brief description of KDE method is given in Eq. 2.11. As one can see, the KDE



(a) QQ plot between residuals of data 1 and normal distribution.

(b) QQ plot between residuals of data 1 and NIG distribution.

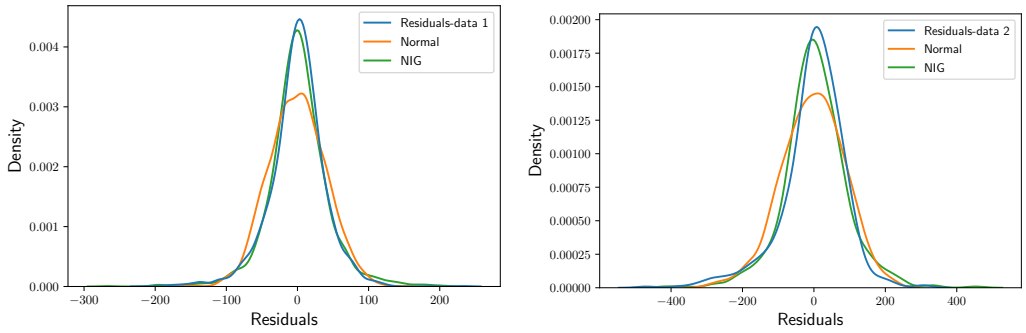
Figure 3.13: QQ plots of residual terms of data 1 compared with (a) normal distribution and (b) NIG($\alpha = 0.02, \beta = 0, \mu = 0, \delta = 34.5$) distribution.



(a) QQ plot between residuals of data 2 and normal distribution.

(b) QQ plot between residuals of data 2 and NIG distribution.

Figure 3.14: QQ plots of residual terms of data 2 compared with (a) normal distribution and (b) NIG($\alpha = 0.008, \beta = 0, \mu = 0, \delta = 70.5$) distribution.

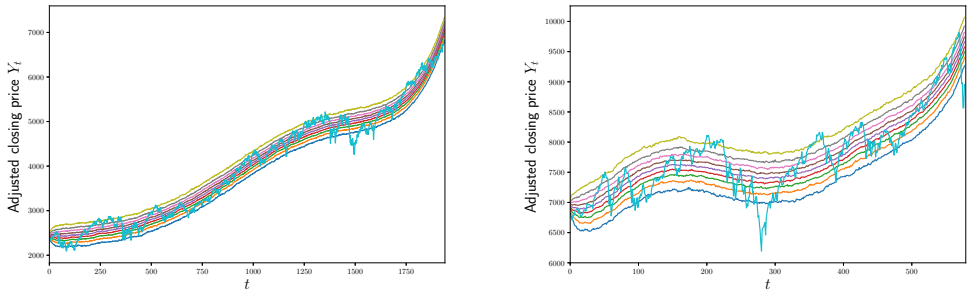


(a) KDE plot for data 1.

(b) KDE plot for data 2.

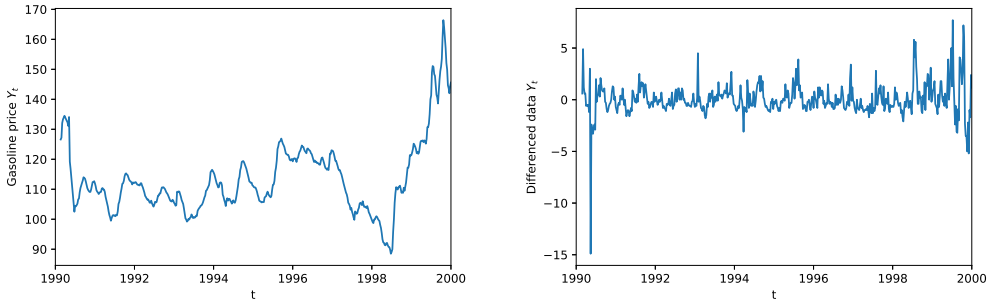
Figure 3.15: Kernel density estimation plots for comparing the residual terms of data 1 with NIG($\alpha = 0.02, \beta = 0, \mu = 0, \delta = 34.5$) distribution and normal distribution $N(\mu = -0.1170, \sigma^2 = 1513.0754)$ (left panel) and data 2 with NIG($\alpha = 0.008, \beta = 0, \mu = 0, \delta = 70.5$) and normal distribution $N(\mu = -1.6795, \sigma^2 = 89.2669)$ (right panel).

plots clearly indicate the NIG distribution is the appropriate one for the residual



(a) Data 1. (b) Data 2.

Figure 3.16: The adjusted closing price of NASDAQ index for both segments (blue lines) along with the quantile lines of $10\%, 20\%, ..., 90\%$ constructed base on the fitted AR(1) models with NIG distribution with added trends.



(a) Gasoline price data. (b) Stationary data.

Figure 3.17: The time series plot of US gasoline price data (left panel) and the corresponding stationary data after first order differencing (right panel).

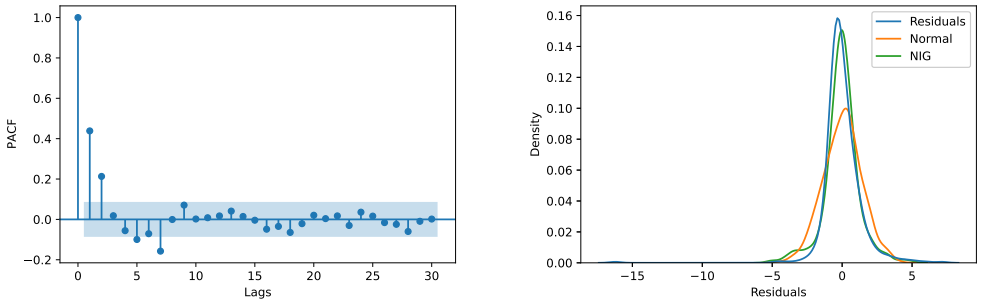
series. Finally, to confirm that the fitted models are appropriate for data 1 and data 2, we constructed the quantile plots of the data simulated from the models for which the removed polynomials were added. The quantile lines are constructed based on 1000 simulated trajectories with the same lengths as data 1 and data 2. In Fig. 3.16 we present the constructed quantile lines on the levels $10\%, 20\%, \cdots, 90\%$ and the adjusted closing price of the NASDAQ index.

The presented results for the real data indicate that the AR(1) model with residual terms corresponding to NIG distribution can be useful for the financial data with visible extreme values.

US gasoline price data

In this subsection, we fit the AR(p) model with NIG residuals on US gasoline weekly price data collected from R dataset repository [3]. The data is collected for period 1990 to 2000 with 513 data points. The Augmented Dickey Fuller (ADF) test

[3] https://vincentarelbundock.github.io/Rdatasets/datasets.html



(a) PACF of data. (b) KDE plot for data.

Figure 3.18: PACF plot of the stationary time series (left panel) and kernel density estimation plots for comparing the residual terms of data with NIG($\alpha = 0.4860, \beta = 0, \mu = 0, \delta = 0.7413$) distribution and normal distribution $N(\mu = 0.02, \sigma^2 = 1.88)$ (right panel).

[47] and the time plot of the price data in Fig. 3.17(a) suggests that the series is non-stationary. For details about ADF test Section 2.4. The stationary series is obtained by using the first order differencing as shown in Fig. 3.17(b). The PACF

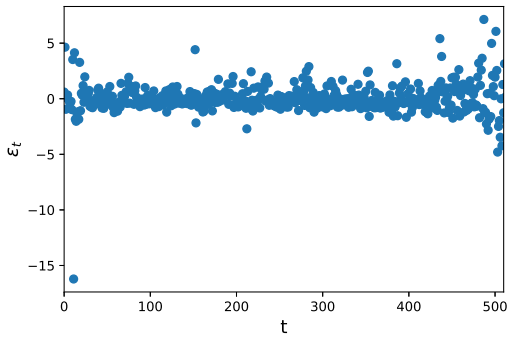


Figure 3.19: The scatter plot of NIG($\alpha = 0.4860, \beta = 0, \mu = 0, \delta = 0.7413$) distributed residuals from AR(1) model with $\rho = 0.48$ for US gasoline price data.

plot in Fig. 3.18(a) implies that the AR(1) can be a good choice. The p value of the one sample KS test for the residuals of the data is 0.00077 (less than 0.05), therefore we reject the null hypothesis of normal distribution and assume the data to be from NIG distribution. The EM algorithm is applied to estimate the parameters of AR(1) model with NIG residuals with fixed $\mu = 0$ and $\beta = 0$ parameters. The estimated parameters for the model are $\hat{\rho} = 0.48, \hat{\alpha} = 0.4860$ and $\hat{\delta} = 0.7413$. The scatter plot of the NIG($\mu = 0, \beta = 0, \alpha = 0.486, \delta = 0.7413$) residuals from AR(1) model is shown in Fig. 3.19. Again the KDE plot on residual series is used along with the NIG($\alpha = 0.4860, \mu = 0, \beta = 0, \delta = 0.7413$) distribution and normal distribution to validate the assumption of residuals. From Fig. 3.18(b) we observe that the NIG distribution is better fit than the normal distribution for the residuals of the data.

Therefore, we conclude that the AR(p) model with NIG residuals is a simple and better model for different data with extreme values.

3.4 Conclusion

The heavy-tailed and semi heavy-tailed distributions are at the heart of the financial time-series modeling. NIG is a semi-heavy tailed distribution which has tails heavier than the normal and lighter than the power law tails. The AR models with NIG distributed residuals are also discussed. We have demonstrated the main properties of the analyzed systems. The main part is devoted to the new estimation algorithm for the considered models' parameters. The technique incorporates the EM algorithm which is widely used in the time series analysis. The effectiveness of the proposed algorithm is demonstrated for the simulated data from model AR(2) and AR(1). It is shown that the new technique outperforms the classical approaches based on the YW and CLS algorithms. Also, for NIG distributed residuals the comparison of the estimation of parameters α and δ with ML estimate by Newton method and EM algorithm is also shown. Finally, we have demonstrated that an AR(1) model with NIG residuals explain well the Google equity price and NASDAQ stock market index data. The AR(1) model with NIG residuals also very well captures the behavior of data from another domain namely, US gasoline price data. The satisfactory results on two different datasets show the universality of the proposed model. We believe that the discussed model is universal and can be used to describe various real-life time series ranging from finance and economics to natural hazards, ecology, and environmental data. Also, the residuals with non-zero mean can be considered for further study.

Chapter 4

Autoregressive model with Cauchy innovations

This chapter deals with the AR(p) model with Cauchy distributed innovations. The closed form estimates of the parameters of the model with Cauchy innovations are derived using expectation-maximization (EM) algorithm. The efficacy of the estimation procedure is shown on the simulated data. Moreover, we also discuss the joint characteristic function of the AR(1) model with Cauchy innovations, which can also be used to estimate the parameters of the model using empirical characteristic function.

4.1 Introduction

Autoregressive (AR) models with stable and heavy-tailed innovations are of great interest in time series modeling. These distributions can easily assimilate the asymmetry, skewness and outliers present in time series data. The Cauchy distribution is a special case of stable distribution with undefined expected value, variance and higher order moments. For the definition of heavy-tailed distribution refer Def. 2.1, which implies that Cauchy distribution is also heavy-tailed. The Cauchy distribution and its mixture has many applications in field of economics [105], seismology [87], biology [23] and various other fields but only few study has been done with time series models with Cauchy errors. In [34], the maximum likelihood (ML) estimation of AR(1) model with Cauchy errors is studied. The standard estimation techniques for the AR(p) model with Cauchy innovations, particularly the Yule-Walker method and conditional least squares method, cannot be used due to the infinite second order moments of the Cauchy distribution. Therefore, it is worthwhile to study and assess the alternate estimation techniques for AR(p) model with Cauchy innovations. In the literature, several estimation techniques are proposed to estimate the parameters of AR models with infinite variance errors, see e.g. [83,102,133]. We propose to use the EM algorithm to estimate the distribution's and model's parameters simultaneously. It is a general iterative algorithm for model parameter estimation which iterates between two steps,

namely the expectation step (*E-step*) and the maximization step (*M-step*) [40]. It is an alternative to numerical optimization of the likelihood function which is proven to be numerically stable [108]. We also provide the formula based on characteristic function (CF) and empirical characteristic function (ECF) of Cauchy distribution for AR(p) model estimation. The idea to use ECF in time series stable ARMA model has been discussed in [148].

The remaining chapter is organized as follows: In Section 4.2, we present a brief overview of the Cauchy AR(p) model, followed by a discussion of estimation techniques namely the EM algorithm and estimation by CF and ECF. Section 4.3 checks the efficacy of the estimation procedure on simulated data. We also present the comparative study, where the proposed technique is compared with ML estimation for Cauchy innovations. Section 4.4 concludes the chapter.

4.2 Cauchy autoregressive model

We consider the AR(p) univariate stationary time-series $\{Y_t\}$, $t \in \mathbb{N}$ with Cauchy innovations defined as

$$Y_t = \sum_{i=1}^{p} \rho_i Y_{t-i} + \varepsilon_t = \rho^T \mathbf{Y_{t-1}} + \varepsilon_t, \tag{4.1}$$

where $\rho = (\rho_1, \rho_2, \cdots, \rho_p)^T$ is a p-dimensional column vector, $\mathbf{Y_{t-1}} = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})^T$ is a vector of p lag terms, $\{\epsilon_t\}$, $t \in \mathbb{N}$ are i.i.d. innovations distributed as Cauchy(α, γ). The pdf of Cauchy(α, γ) [52] is

$$f(x; \alpha, \gamma) = \frac{\gamma}{\pi}\left[\frac{1}{\gamma^2 + (x-\alpha)^2}\right], \ \gamma > 0, \ \alpha \in \mathbb{R}, \ x \in \mathbb{R}. \tag{4.2}$$

The conditional distribution of Y_t given the preceding data $\mathcal{F}_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_1)^T$ is given by [34]

$$p(Y_t|\mathcal{F}_{t-1}) = f(y_t - \rho^T \mathbf{y_{t-1}}; \alpha, \gamma) = \frac{\gamma}{\pi}\left[\frac{1}{\gamma^2 + (y_t - \rho^T \mathbf{y_{t-1}} - \alpha)^2}\right],$$

where $\mathbf{y_{t-1}}$ is the realization of $\mathbf{Y_{t-1}}$. In the next subsection, we propose the methods to estimate the model parameters ρ and innovation parameters α and γ simultaneously.

4.2.1 Parameter estimation using EM algorithm

We estimate the parameters of AR(p) model using EM algorithm which maximizes the likelihood function iteratively. Further, we discuss about the time series $\{Y_t\}$

using the characteristic function (CF) and estimation method using CF and ECF. For definition of CF see Def. 2.3. Recently, [133] the exponential-squared estimator for AR model with heavy-tailed errors is introduced and proved to be $\sqrt{n}$-consistent under some regularity conditions, similarly self-weighted least absolute deviation estimation method is also studied for the infinite variance AR model [103]. The ML estimation of AR models with Cauchy errors with intercept and with linear trend is studied and the AR coefficient is shown to be $n^{3/2}$-consistent under some conditions [34]. For AR(p) model with Cauchy innovations with n samples, the log likelihood is defined as,

$$l(\Theta) = n\log(\gamma) - n\log(\pi) - \sum_{t=1}^{n} \log(\gamma^2 + (y_t - \rho^T Y_{t-1} - \alpha)^2),$$

where $\Theta = \left(\alpha,\ \gamma,\ \rho^T\right)$.

Proposition 4.1. *Consider the AR(p) time-series model given in Eq. (4.1) where error terms follow Cauchy(α, γ). The maximum likelihood estimates of the model parameters using EM algorithm are as follows:*

$$\begin{aligned}
\hat{\rho}^T &= \left(\sum_{t=p}^{n} \frac{(y_t - \alpha)Y_{t-1}^T}{s_t}\right)\left(\sum_{t=p}^{n} \frac{\mathbf{Y_{t-1}}\mathbf{Y_{t-1}^T}}{s_t}\right)^{-1}, \\
\hat{\alpha} &= \frac{\sum_{t=p}^{n} \frac{(y_t - \rho^T Y_{t-1})}{s_t}}{\sum_{t=1}^{n} \frac{1}{s_t}}, \ \textit{and} \\
\hat{\gamma} &= \sqrt{\frac{n}{2\sum_{t=1}^{n} \frac{1}{s_t}}},
\end{aligned} \tag{4.3}$$

where $s_t = (y_t - \rho^T \mathbf{Y_{t-1}} - \alpha)^2 + \gamma^2$.

Proof. Consider the AR(p) model

$$Y_t = \rho^T \mathbf{Y_{t-1}} + \varepsilon_t,\ t = 1, 2, \cdots, n,$$

where ε_t follows Cauchy distribution Cauchy(α, γ). Let (ε_t, V_t) for $t = 1, 2, \cdots, n$ denote the complete data for innovations ε. The observed data ε_t is assumed to be from Cauchy(α, γ) and we also assume that the unobserved data V_t is from inverse gamma IG$(1/2, \gamma^2/2)$ distribution. A random variable $V \sim \text{IG}(a, b)$ if the pdf is given by

$$f_V(v; a, b) = \frac{b^a}{\Gamma(a)} \frac{e^{-b/v}}{v^{a+1}},\ a > 0,\ b > 0,\ v > 0.$$

We can rewrite ε_t as $\varepsilon_t = Y_t - \rho^{\mathbf{T}}\mathbf{Y_{t-1}}$, for $t = 1, 2, \cdots, n$. The stochastic relation $\varepsilon = \alpha + \sqrt{V}Z$ with $Z \sim N(0,1)$ i.e. standard normal and $V \sim \text{IG}(1/2, \gamma^2/2)$ is used to generate Cauchy(α, γ) distribution. First we show that this relation holds. The conditional distribution $\epsilon|V$ has following form:

$$f(\varepsilon = \epsilon_t | V = v_t) = \frac{1}{\sqrt{2\pi v_t}} \exp\left(-\frac{1}{2v_t}(\epsilon_t - \alpha)^2\right).$$

The density function of $f_\epsilon(e) = \int_0^\infty f_{\epsilon|V}(e|v) f(v)\, dv$, where $f_V(v) = \frac{\gamma}{\sqrt{(2\pi)}v^{3/2}} \exp(-\gamma^2/2v)$.

$$\begin{aligned} f_\epsilon(e) &= \int_0^\infty f_{\epsilon|V}(e|v) f(v)\, dv \\ &= \int_0^\infty \frac{\exp\Big(-\dfrac{(\epsilon-\alpha)^2}{2v}\Big)}{\sqrt{(2\pi v)}} \frac{\gamma}{\sqrt{(2\pi)}} \frac{\exp(-\gamma^2/2v)}{v^{3/2}}\, dv \\ &= \int_0^\infty \frac{\gamma v^{-2} \exp(-\frac{(e-\alpha)^2-\gamma^2}{2v})}{2\pi}\, dv \end{aligned}$$

We substitute $t = \frac{-1}{2v}$ in the above integral and solve to obtain the following form:

$$\begin{aligned} f_\epsilon(e) &= \int_{-\infty}^0 \frac{\gamma}{2\pi} \exp\Big(\frac{t}{2}((e-\alpha)^2 + \gamma^2)\Big)\, dt \\ &= \frac{\gamma}{\pi}\Big[\frac{1}{(e-\alpha)^2 + \gamma^2}\Big]. \end{aligned}$$

Therefore, we say that the stochastic relation $\varepsilon = \alpha + \sqrt{V}Z$ with $Z \sim N(0,1)$ and $V \sim \text{IG}(1/2, \gamma^2/2)$ holds. Now, we need to estimate the unknown parameters $\Theta = (\alpha, \gamma, \rho^T)$. To apply the EM algorithm for estimation we first find the conditional expectation of log-likelihood of complete data (ε, V) with respect to the conditional distribution of V given ε. Since the unobserved data is assumed to be from IG$(1/2, \gamma^2/2)$ therefore, the posterior distribution is again an inverse gamma i.e.,

$$V|\varepsilon = e, \Theta \sim \text{ IG}\left(1, \frac{(e-\alpha)^2 + \gamma^2}{2}\right).$$

The following conditional inverse first moment and $\mathbb{E}(\log(V)|\varepsilon = e)$ will be used in calculating the conditional expectation of the log-likelihood function:

$$\begin{aligned} \mathbb{E}(V^{-1}|\varepsilon = e) &= \frac{2}{(e-\alpha)^2 + \gamma^2}, \\ \mathbb{E}(\log(V)|\varepsilon = e) &= \log((e-\alpha)^2 + \gamma^2) - \log 2 + 0.5776. \end{aligned}$$

The complete data likelihood is given by

$$\begin{aligned} L(\Theta) &= \prod_{t=1}^{n} f(\epsilon_t, v_t) = \prod_{t=1}^{n} f_{\varepsilon|V}(\epsilon_t|v_t) f_V(v_t) \\ &= \prod_{t=1}^{n} \frac{\gamma}{2\pi v_t^2} \exp\left(-\frac{(\epsilon_t - \alpha)^2 + \gamma^2}{2v_t}\right). \end{aligned}$$

The log likelihood function will be

$$l(\Theta) = n\log(\gamma) - n\log(2\pi) - 2\sum_{t=1}^{n}\log(v_t) - \sum_{t=1}^{n}\frac{(\epsilon_t - \alpha)^2 + \gamma^2}{2v_t}.$$

We use the relation $\epsilon_t = Y_t - \rho^T Y_{t-1}$ in further calculations. In the first step (*E-step*) of EM algorithm, we need to compute the expected value of complete data log likelihood known as $Q(\Theta|\Theta^{(k)})$, which is expressed as,

$$\begin{aligned} Q(\Theta|\Theta^{(k)}) &= \mathbb{E}_{V|\varepsilon,\Theta^{(k)}}[\log f(\varepsilon, V|\Theta)|\epsilon_t, \Theta^{(k)}] = \mathbb{E}_{V|\varepsilon,\Theta^{(k)}}[l(\Theta|\Theta^{(k)})] \\ &= n\log\gamma - n\log 2\pi - \sum_{t=p}^{n}\mathbb{E}(\log v_t|\epsilon_t, \Theta^{(k)}) \\ &\quad - \frac{\gamma^2}{2}\sum_{t=p}^{n}\mathbb{E}(v_t^{-1}|\epsilon_t, \Theta^{(k)}) - \frac{1}{2}\sum_{t=p}^{n}(\epsilon_t - \alpha)^2\mathbb{E}(v_t^{-1}|\epsilon_t, \Theta^{(k)}) \\ &= n\log\gamma - n\log 2\pi - n\log 2 + 0.5776n - \sum_{t=p}^{n}\log((\epsilon_t - \alpha^{(k)})^2 + {\gamma^{(k)}}^2) \\ &\quad - \frac{\gamma^2}{2}\sum_{t=p}^{n}\frac{1}{(\epsilon_t - \alpha^{(k)})^2 + {\gamma^{(k)}}^2} - \sum_{t=p}^{n}\frac{(\epsilon_t - \alpha)^2}{(\epsilon_t - \alpha^{(k)})^2 + {\gamma^{(k)}}^2} \\ &= n\log\gamma - n\log 2\pi - n\log 2 + 0.5776n - \sum_{t=p}^{n}\log(s_t^{(k)}) \\ &\quad - \frac{\gamma^2}{2}\sum_{t=p}^{n}\frac{1}{s_t^{(k)}} - \sum_{t=p}^{n}\frac{(\epsilon_t - \alpha)^2}{s_t^{(k)}}. \end{aligned}$$

where, $s_t = (y_t - \rho^T \mathbf{Y_{t-1}} - \alpha)^2 + \gamma^2$. In the next *M-step*, we estimate the parameters α, γ and ρ^T by maximizing the Q function using the equations below:

$$\begin{aligned} \frac{\partial Q}{\partial \rho} &= 4\sum_{t=p}^{n}\frac{(y_t - \rho^T Y_{t-1} - \alpha Y_{t-1}^T)}{s_t^{(k)}}, \\ \frac{\partial Q}{\partial \alpha} &= \sum_{t=p}^{n}\frac{(y_t - \rho^T Y_{t-1} - \alpha)}{s_t^{(k)}}, \\ \frac{\partial Q}{\partial \gamma} &= \frac{n}{\gamma} - 2\gamma\sum_{t=p}^{n}\frac{1}{s_t^{(k)}}. \end{aligned}$$

Solving the above equations, we get the following closed form estimates of the parameters:

$$\hat{\rho}^T = \left(\sum_{t=p}^{n} \frac{(y_t - \alpha)Y_{t-1}^T}{s_t}\right)\left(\sum_{t=p}^{n} \frac{\mathbf{Y_{t-1}Y_{t-1}^T}}{s_t}\right)^{-1},$$

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{t=p}^{n} \frac{(y_t - \rho^T Y_{t-1})}{s_t}}{\sum_{t=1}^{n} \frac{1}{s_t}}, \text{ and} \\ \hat{\gamma} &= \sqrt{\frac{n}{2\sum_{t=1}^{n} \frac{1}{s_t}}}, \end{aligned} \tag{4.4}$$

where $s_t = (y_t - \rho^T \mathbf{Y_{t-1}} - \alpha)^2 + \gamma^2$. □

4.2.2 Characteristic function for estimation

So far, we have considered the conditional distribution of Y_t given the preceding data $\mathcal{F}_{t-1}$. Now, we include the dependency of time series $\{Y_t\}$ by defining the variable $d_j = (y_j, \cdots, y_{j+p})$ for $j = 1, \cdots, n-p$. In each variable $\{d_j\}$ there are p terms same as adjacent variable. The distribution of $\{d_j\}$ will be multivariate Cauchy with dimension $r = p + 1$ [53]. The CF of each d_j is $c(\Theta, s) = \mathbb{E}(\exp(i\, s^T d_j))$ and the ECF is $c_n(s) = \frac{1}{n}\sum_{j=1}^{n-p} \exp(i\, s^T d_j)$, where $s = (s_1, \cdots, s_{p+1})^T$.

To estimate the parameters using CF and ECF we make sure that the joint CF of AR(p) model has closed form. In the next result the closed form expression for the joint CF of the AR(1) model with Cauchy innovations is given.

Proposition 4.2. *The joint CF of stationary AR(1) model with Cauchy innovations is*

$$c(s_1, s_2; \Theta) = \exp\left[i\alpha(s_1 + s_2)\left(\frac{1}{1-\rho}\right)\right] \times \exp\left[-\gamma\left(|s_2| + |s_1 + \rho s_2|\left(\frac{1}{1-|\rho|}\right)\right)\right].$$

Proof. For stationary AR(1) model $y_t = \rho y_{t-1} + \varepsilon_t$, we can rewrite it as,

$$y_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3\varepsilon_{t-3} + \cdots.$$

Note that $\{\varepsilon_t\}$ are i.i.d from Cauchy(α, γ) distribution and CF of Cauchy(α, γ) is $\mathbb{E}(\exp(i\ s\varepsilon_t)) = \exp(i\ \alpha s - \gamma|s|)$ [52]. Then the joint CF of (y_t, y_{t-1}) is calculated as

follows:

$$
\begin{aligned}
c(s_1, s_2; \Theta) &= \mathbb{E}[\exp(i\ s_1 y_{t-1} + i\ s_2 y_t)] \\
&= \mathbb{E}[\exp(i\ s_1(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \cdots)) + i\ s_2(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots)] \\
&= \mathbb{E}[\exp(i\ s_2\varepsilon_t + i(s_1 + s_2\rho)\varepsilon_{t-1} + i\rho(s_1 + \rho s_2)\varepsilon_{t-2} + \cdots)] \\
&= \exp(i\alpha s_2 - \gamma|s_2|) \times \exp(i\alpha(s_1 + \rho s_2) - \gamma|s_1 + \rho s_2|) \\
&\times \exp(i\alpha(\rho s_1 + \rho^2 s_2) - \gamma|\rho s_1 + \rho^2 s_2|) \times \cdots . \\
&= \exp(i\alpha(s_1 + s_2)(1 + \rho + \rho^2 + \rho^3 + \cdots)) \\
&\times \exp(-\gamma(|s_2| + |s_1 + \rho s_2| + |\rho||s_1 + \rho s_2| + |\rho^2||s_1 + \rho s_2| + \cdots)) \\
&= \exp\left(i\alpha(s_1 + s_2)\left(\frac{1}{1-\rho}\right)\right) \times \exp(-\gamma(|s_2| + |s_1 + \rho s_2|(1 + |\rho| + |\rho^2| + \cdots))) \\
&= \exp\left[i\alpha(s_1 + s_2)\left(\frac{1}{1-\rho}\right)\right] \times \exp\left[-\gamma\left(|s_2| + |s_1 + \rho s_2|\left(\frac{1}{1-|\rho|}\right)\right)\right].
\end{aligned}
$$

□

The joint CF for higher dimension can be obtained in similar manner. The model parameters can be estimated by solving the following integral with CF and ECF as defined in [148]:

$$\int \cdots \int w_\Theta(s)(c_n(s) - c(s;\Theta))ds = 0. \tag{4.5}$$

where optimal weight function

$$w^*_\Theta(s) = \frac{1}{2\pi}\int \cdots \int \exp(-i\, s^T d_j)\frac{\partial \log f(y_{j+p}|y_j, \cdots, y_{j+p-1})}{\partial\Theta} dy_j \ldots dy_{j+p}. \tag{4.6}$$

Remark 4.1. For stationary AR(l) process $\{Y_t\}$ with $p = l$ the ECF estimator defined by Eq. (4.5) with optimal weight function defined in Eq. (4.6) is a conditional ML (CML) estimator and hence asymptotically efficient. The conditional log pdf for Cauchy distribution is:

$$\log f(y_{j+p}|y_j, \cdots, y_{j+p-1}) = \log\gamma - \log\pi - \log\left(\gamma^2 + (y_t - \rho^T Y_{t-1} - \alpha^2)^2\right).$$

The proof is similar to the proof of Proposition 2.1 in [148].

4.3 Simulation study

In this section, we assess the proposed model and the introduced estimation technique using simulated data set. We discuss the estimation procedure for AR(2) model with Cauchy innovations. The AR(2) model defined by Eq. (4.1) is simulated

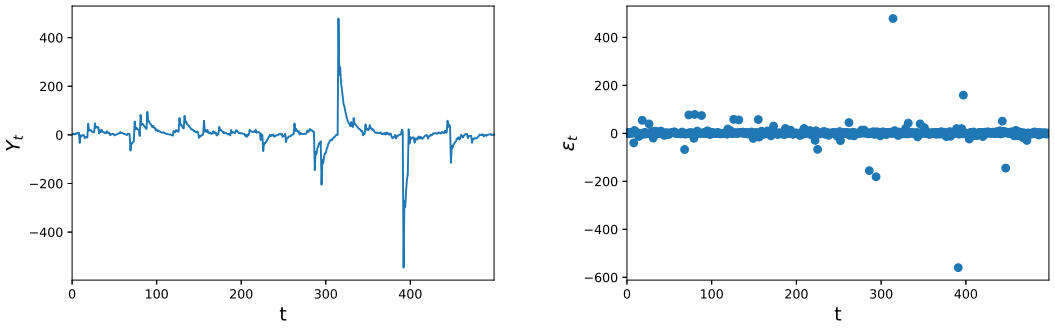
with ρ_1 and ρ_2 as model parameters. We generate 1000 trajectories each of size $N = 500$ of Cauchy innovations using the normal variance-mean mixture form $\varepsilon = \alpha + \sqrt{V}Z$ with $Z \sim N(0,1)$ i.e. standard normal and $V \sim \text{IG}(1/2, \gamma^2/2)$. Use the following simulation steps to generate the Cauchy innovations:

Step 1: Generate standard normal variate Z;

Step 2: Generate inverse gamma random variate $\text{IG}(1/2, \gamma^2/2)$ with $\gamma = 2$;

Step 3: Using the relation $\varepsilon = \alpha + \sqrt{V}Z$, we simulate the Cauchy innovations with $\alpha = 1$;

Step 4: The time series data y_t is generated with model parameters $\rho_1 = 0.5$ and $\rho_2 = 0.3$.



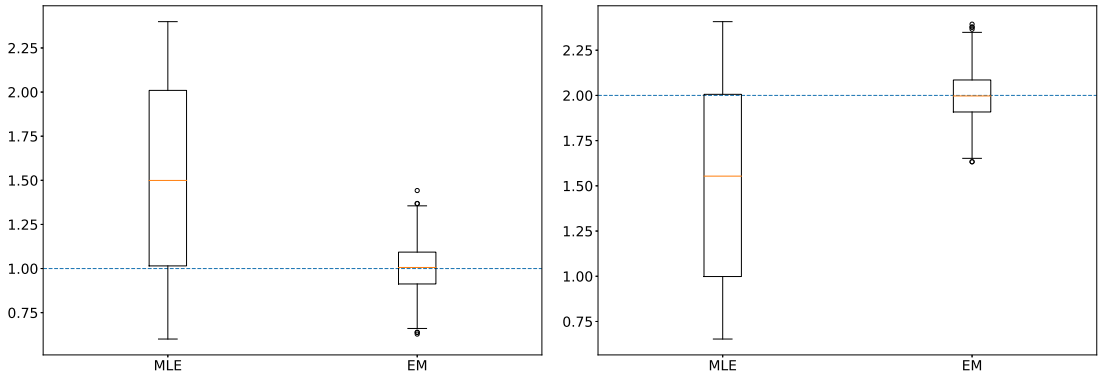
(a) The time series data plot. (b) Scatter plot of innovation terms.

Figure 4.1: The exemplary time series of length $N = 500$ (left panel) and the corresponding innovation terms (right panel) of the AR(2) model with Cauchy innovations. The chosen parameters of the model are: $\rho_1 = 0.5$, $\rho_2 = 0.3$, $\alpha = 1$, and $\gamma = 2$.

The exemplary time series data plot and scatter plot of innovation terms are shown in Fig. 4.1. We apply the discussed EM algorithm to estimate the model parameters and distribution parameters. The relative change in the parameters is considered to terminate the algorithm. Following is the stopping criteria which is commonly used in literature:

$$\max\left\{\left|\frac{\alpha^{(k+1)}-\alpha^{(k)}}{\alpha^{(k)}}\right|, \left|\frac{\gamma^{(k+1)}-\gamma^{(k)}}{\gamma^{(k)}}\right|, \left|\frac{\rho_1^{(k+1)}-\rho_1^{(k)}}{\rho_1^{(k)}}\right|, \left|\frac{\rho_2^{(k+1)}-\rho_2^{(k)}}{\rho_2^{(k)}}\right|\right\} < 10^{-4}. \quad (4.7)$$

We compare the estimation results of Cauchy(α, γ) with EM algorithm and maximum likelihood (ML) estimation. The ML estimates are computed using the inbuilt function "mlcauchy" in R which uses the exponential transform of the location parameter and performs non-linear minimization by Newton-type



(a) Boxplot for α estimate.

(b) Boxplot for γ estimate.

Figure 4.2: Boxplots of the estimates of the AR(2) model's parameters with theoretical values: $\alpha = 1$ and $\gamma = 2$ represented with blue dotted lines. The boxplots are created using 1000 trajectories each of length 500.

algorithm. The comparison of estimates of Cauchy(α, γ) are shown in boxplot in Fig. 4.2. From the boxplots, we find that the EM algorithm converges near to the true value of the Cauchy(α, γ) as compared to the ML estimation. There is possibility of getting better result from ML method if different algorithm or inbuilt function for optimization are used for estimation.

4.4 Conclusion

We derive the closed form of estimates of AR model with Cauchy innovations using EM algorithm. The performance of the proposed algorithm is compared with the ML method using simulated data. The ML estimation is found using inbuilt function in R. Another benefit of using EM algorithm is that it calculates the model as well as the innovation parameters simultaneously. It is evident from the boxplot that EM algorithm outperforms the ML method. Further, we discuss another approach based on CF to estimate the AR model parameters with stable distribution. In the future, we study and compare the proposed algorithm and ECF based estimation method with the existing techniques in [83,102,133] for AR model with infinite variance. Further, the real life phenomena can be studied using the proposed model and methods.

Chapter 5

Geometric infinitely divisible autoregressive models

This chapter focuses on defining AR(1) processes using the Laplace transform. First, we introduce the geometric infinitely divisible (gid) distributions and study their properties. Specifically, we consider geometric tempered stable, geometric gamma, and geometric inverse Gaussian distributions. For AR(1) model defined by Eq. (5.2) with assumption of gid marginals, the integral form of the density function of innovation terms is obtained. Further, we also provide the estimation of AR(1) model defined in Prop. 5.9 using conditional least squares (CLS) and method of moments (MOM) and provide the simulation study for the same.

5.1 Introduction

Classical autoregressive (AR) processes are the most popular time series models because these models are easy to understand, interpret, and have applications in different fields. In the AR process, the present value is a linear combination of the past values plus some innovation term. The distribution of innovation terms plays an important role in capturing extreme events. The classical AR model assumes that the innovation term has Gaussian distribution, which further leads to the marginals becoming Gaussian. However, many real life time series data exhibit heavy-tailed or semi-heavy-tailed behaviours. Various AR models have been introduced with non-Gaussian innovations and marginals. The data with non-negative observations, binary outcomes, series of counts, and proportions are some examples of non-Gaussian real life time series (see e.g. [68]). The first order AR models with different marginal distributions have been extensively studied in literature. In 1980, Gaver and Lewis [61] considered an AR(1) model with marginals having gamma distribution. The authors discussed that the innovation terms have the same distribution as a non-negative random variable, which is exponential for positive values and has a point mass at 0. Later, Dewald and Lewis [42] studied a second order autoregressive model in which Laplace distributed marginals are assumed. An inverse Gaussian autoregressive model in which the marginals are

inverse Gaussian distributed was also studied [2]. The authors show that after fixing the parameter of the inverse Gaussian to 0 the innovation term is also inverse Gaussian distributed. Recently, [22] studied the AR(1) model with one-sided tempered stable marginals and innovations, and demonstrated that the model fits very well real-world data. The first order AR processes were also studied by Lekshmi and Jose [100,101] with geometric α-Laplace and geometric Pakes generalized Linnik marginals. For a literature survey on unified view of AR(1) models, see [69].

In this study, we use the concept of geometric infinitely divisible (gid) distributions to define a new class of generalized AR(1) processes. We first define a class of geometric infinitely divisible random variables with a Laplace transform of the form $\frac{1}{1+g(s)}$, where $g(s)$ is a Bernstein function which is also known as Laplace exponent of general subordinators. A large class of distributions, for example, the geometric α-stable (or Mittag Leffler), geometric tempered stable, geometric gamma, geometric inverse Gaussian, and one-sided geometric Laplace distributions, come under this umbrella. This class of generalised AR(1) models have ability to handle intricate and versatile time series patterns. These models can accommodate more complex underlying dynamics of time series data. Earlier work related to AR(1) models primarily focused on specific distributions of innovation terms or assumed particular forms for the marginals see [100,101]. We discuss the AR(1) model within a general class of geometric infinitely divisible distributions. We mainly focus on geometric tempered stable, geometric inverse Gaussian and geometric gamma distributions and corresponding AR models. These distributions have different properties which make them interesting. We empirically show that these distributions fall under the category of semi heavy-tailed or heavy-tailed distributions. We plot the $\log(\mathbb{P}(X > x))$, where $\mathbb{P}(X > x)$ is survival function. The survival function is the probability that an event has not occurred by a threshold x which is essentially a tail probability [93]. To study the tail behavior of these three distributions, we compare them with stable, standard normal and standard exponential distributions. All the three distributions lies between standar normal and stable distribution, which imply that they belong to the class of semi heavy-tailed distributions. This property make them suitable for modeling data with extreme events and outliers, particularly relevant in finance and risk management. The corresponding time series models derived from this distribution can effectively capture temporal dependencies. The study of the general AR(1) model based on geometric infinitely divisible distributions provides a unified platform to address the behaviors of wide range of data.

To the best of our knowledge these models and corresponding AR processes have not been explored in the literature. First, the properties of gid marginals are studied in Section 5.2. Section 5.3 describes the AR(1) model with gid marginals. We also obtain the distribution of innovation terms, which are from the class of gid

distributions. We then evaluate the pdf of innovation terms in integral form for the geometric tempered stable, geometric gamma, and geometric inverse Gaussian cases. Further, we generalize the results obtained in this section to the kth order autoregressive processes. In Section 5.4, the conditional least square (CLS) and method of moments (MOM) are used to estimate the parameters of the AR(1) model defined in Prop. 5.9 and provide the simulation study for two different datasets. Section 5.5 concludes the chapter.

5.2 Geometric infinitely divisible distributions

In this section, we discuss gid distributions. Infinitely divisible random variables (or distributions) with positive supports play a crucial role in the study of subordinators, see [11] and Def. 2.5. The subordinators are real-valued non-decreasing stochastic processes having independent and stationary increments. The properties of subordinators are well studied in the literature (e.g. see, [11,96,97]). Bernstein functions play a crucial role in the theory of non-decreasing Lévy processes or subordinators. The Bernstein functions are defined in Def. 2.4. For a subordinator $S(t)$, it follows that $\mathbb{E}[e^{-sS(t)}] = e^{-t\phi(s)}$, where $\phi(s)$ is also called the Laplace exponent. Next we define the gid' marginals.
Consider a positive random variable Y with Laplace transform $f(s)$ that is $\mathbb{E}[e^{-sY}] = f(s)$. Let X be a positive infinitely divisible random variable with Laplace transform $\mathbb{E}[e^{-sX}] = e^{-g(s)}$, where $g(s)$ is a Bernstein function ([128]). The Laplace transform of X can be written as,

$$\phi_X(s) = e^{-g(s)} = e^{1-g(s)-1} = e^{1-\frac{1}{(1+g(s))-1}} = e^{1-\frac{1}{f(s)}},\ s > 0. \tag{5.1}$$

It is known from [92] that a distribution with Laplace transform $f(s)$ is geometrically infinitely divisible if and only if the distribution with Laplace transform $e^{1-\frac{1}{f(s)}}$ is infinitely divisible. One can define a class of geometrically infinitely divisible random variables by choosing $g(s)$ as the Laplace exponent of general subordinators and we call these random variables as the gid based random variables. We introduce a class of random variables Y on $(0,\infty)$ with Laplace transform of the form $f(s) = \frac{1}{1+g(s)}$, where $g(s)$ is the Laplace exponent of positive infinitely divisible random variables and study the relevant properties for time series autoregressive models. Additionally, we extend the concept of geometric α-Laplace ([100]), geometric Pakes generalized Linnik marginals ([101]) based autoregressive models to general class of autoregressive models with gid marginals.

Definition 5.1. The random variable Y on $(0,\infty)$ is said to have gid marginals if its Laplace transform is given by $f(s) = \frac{1}{1+g(s)}$, where $g(s)$ is a Bernstein function

and the Laplace exponent of a positive infinitely divisible random variable.

In particular, we consider different functions for $g(s)$ and define the Laplace transform of corresponding gid marginals:

(a) Geometric tempered stable: The Laplace transform of tempered stable random variable is $e^{-g(s)}$ with $g(s) = (s+\lambda)^\beta - \lambda^\beta$, then $f(s) = \frac{1}{1+(s+\lambda)^\beta - \lambda^\beta}$, $\lambda > 0$ and $\beta \in (0,1)$.

(b) Geometric gamma: The Laplace transform of gamma random variable is $e^{-g(s)}$ with $g(s) = \alpha \log(1+\frac{s}{\beta})$, then $f(s) = \frac{1}{1+\alpha \log\left(\frac{s+\beta}{\beta}\right)}$, $\alpha > 0$ and $\beta > 0$.

(c) Geometric inverse Gaussian: The Laplace transform of inverse Gaussian random variable is $e^{-g(s)}$ with $g(s) = \delta\gamma\left\{\sqrt{1+\frac{2s}{\gamma^2}} - 1\right\}$, then $f(s) = \frac{1}{1+\delta\gamma\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$, $\delta > 0$ and $\gamma > 0$.

Proposition 5.1. *Consider the identically and independently distributed random variables $X_1, X_2, \ldots,$ with gid marginals defined in Def. 5.1. Also consider $N(\theta)$ be a geometric random variable with mean $\frac{1}{\theta}$ and $\mathbb{P}[N(\theta) = r] = \theta(1-\theta)^{r-1}$, $r = 1, 2, \cdots$, $0 < \theta < 1$. Then $Y = \sum_{i=1}^{N(\theta)} X_i$ also have gid marginals with $f(s) = \frac{1}{1+\frac{g(s)}{\theta}}$.*

Proof. Let the Laplace transform of each X_i for $i = 1, 2, \ldots$ is denoted by $\phi_X(s)$. Then, the Laplace transform of random variable Y can be written as,

$$\phi_Y(s) = \sum_{r=1}^{\infty} [\phi_X(s)]^r \theta(1-\theta)^{r-1} = \sum_{r=1}^{\infty} \left[\frac{1}{1+g(s)}\right]^r \theta(1-\theta)^{r-1} = \frac{\frac{\theta}{1+g(s)}}{1-\left(\frac{1-\theta}{1+g(s)}\right)}$$
$$= \frac{1}{1+\frac{g(s)}{\theta}}.$$

Hence, Y is also a gid random variable. □

Next, we study the limiting behavior of the densities of gid random variables near 0^+ using Laplace transform. As the pdf of gid random variables lack closed forms and are expressed through integral representations, we employ the Tauberian theorem defined in [52] (pp. 446, Theorem 4). The Tauberian theorem is stated as follows:

Theorem 5.2. *Let $0 < \rho < \infty$. If U has ultimately monotone derivative u then as $\lambda \to 0$ and $x \to \infty$, respectively,*

$$\omega(\lambda) \sim \frac{1}{\lambda^\rho} L(1/\lambda) \text{ iff } u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x).$$

In the following theorem, we apply the Tauberian theorem to the geometric tempered stable and geometric inverse Gaussian distributions.

Theorem 5.3. *Let $W(s)$ be the Laplace transform and $f(x)$ be the probability density function for a gid random variable, then we obtain the asymptotic behavior of $f(x)$ at 0^+.*

(a) For geometric tempered stable, $W(s) = \frac{1}{1+\theta((s+\lambda)^\beta - \lambda^\beta)}$, then as $x \to 0^+$ density $f(x) \sim \frac{x^{\beta-1}}{\theta\Gamma(\beta)}$, where $\lambda > 0$ and $0 < \beta < 1$.

(b) For geometric inverse Gaussian, $W(s) = \frac{1}{1+\theta\delta\{\sqrt{\gamma^2+2s}-\gamma\}}$, then as $x \to 0^+$ density $f(x) \sim \frac{x^{-1/2}}{\theta\delta\sqrt{2\pi}}$, where $\delta > 0$, and $\gamma > 0$.

Proof. (a) The Laplace transform of geometric tempered stable is $W(s) = \frac{1}{1+\theta((s+\lambda)^\beta - \lambda^\beta)}$, where $\lambda > 0$, $0 < \beta < 1$. For large s, $W(s) \sim \frac{1}{\theta s^\beta}$. We apply Tauberian theorem with exponent β and the slowly varying function $L(1/s) = \frac{1}{\theta}$. Thus, the density $f(x) \sim \frac{x^{\beta-1}}{\theta\Gamma(\beta)}$, as $x \to 0^+$.

(b) The Laplace transform of geometric inverse Gaussian is $W(s) = \frac{1}{1+\theta\delta\{\sqrt{\gamma^2+2s}-\gamma\}}$, where $\delta > 0, \gamma > 0$. Again for large s, $W(s) \sim \frac{1}{\theta\delta\sqrt{2s}}$. The slowly varying function $L(1/s) = \frac{1}{\theta\delta\sqrt{2}}$ and exponent $1/2$. By Tauberian theorem, $f(x) \sim \frac{x^{-1/2}}{\theta\delta\sqrt{2\pi}}$, as $x \to 0^+$.

□

The density plots of the gid distributions, including geometric tempered stable and geometric inverse Gaussian, align with the results obtained in the above theorem. We apply the Laplace transform method, as described in [120], to generate datasets of length 1000 using the R programming language. The density plots for gid with various parameter values are illustrated in Fig. 5.1.

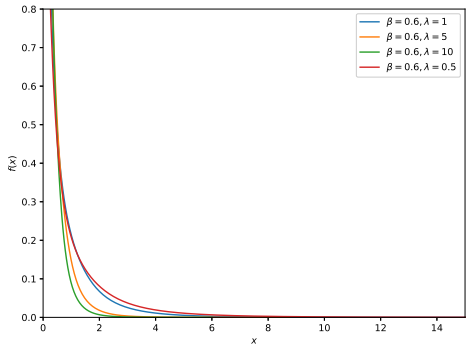
5.2.1 Mixture of gid random variables

In this subsection, we introduce a new random variable M which is a mixture of two independent gid random variables denoted as Y_1 and Y_2. The definition of mixture random variable is given in [52] (pp. 53) as follows:

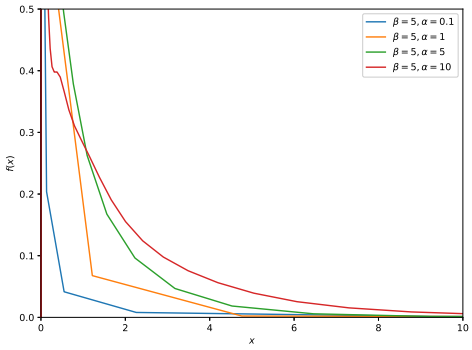
Definition 5.2. The random variable X is said to be mixture random variable if the density of X has the following form: $w(x) = \sum_k f(x, \theta_k)p_k$, where $p_k \geq 0, \sum_k p_k = 1$, and $f(x, \theta_k)$ are component densities.

We define $g(s) = cg_1(s) + (1-c)g_2(s)$, where $0 < c < 1$, g_1 and g_2 are Laplace exponent functions for Y_1 and Y_2 respectively. We compute $\phi_M(s)$ the Laplace for each case as follows:

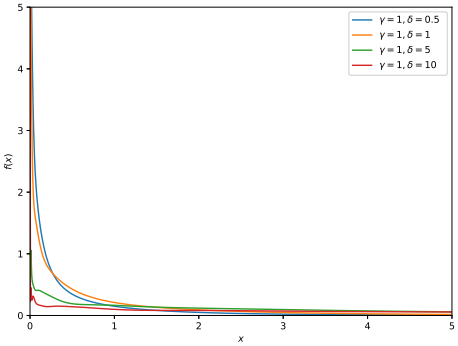
$$e^{-g(s)} = e^{-(cg_1(s)+(1-c)g_2(s))} = e^{1-1-(cg_1(s)+(1-c)g_2(s))} = e^{1-\frac{1}{(1+cg_1(s)+(1-c)g_2(s))^{-1}}}.$$



(a) Geometric tempered stable.



(b) Geometric gamma.



(c) Geometric inverse Gaussian.

Figure 5.1: The density plot of (a) geometric tempered stable with parameters $\beta = 0.6$ and different $\lambda = 1, 5, 10, 0.5$ (b) geometric gamma with parameters $\beta = 5$ and different $\alpha = 0.1, 1, 5, 10$ (c) geometric inverse Gaussian with parameters $\gamma = 1$ and different $\delta = 0.5, 1, 5, 10$.

The Laplace exponent of gid mixture M is,

$$\phi_M(s) = \frac{1}{1 + cg_1(s) + (1-c)g_2(s)} = \frac{1}{1 + g_2(s) + c(g_1(s) - g_2(s))},$$

for $0 < c < 1$. We substitute the values of functions $g_1(s)$ and $g_2(s)$ to obtain the Laplace transform of the following:

(a) Geometric mixture tempered stable: for $0 < \beta_1, \beta_2 < 1$, and $\lambda_1, \lambda_2 > 0$,

$$\phi_M(s) = \frac{1}{1 + (s+\lambda_2)^{\beta_2} - \lambda_2^{\beta_2} + c((s+\lambda_1)^{\beta_1} - \lambda_1^{\beta_1} - (s+\lambda_2)^{\beta_2} + \lambda_2^{\beta_2})}.$$

(b) Geometric mixture gamma: for $\beta_1, \beta_2 > 0$ and $\alpha_1, \alpha_2 > 0$,

$$\phi_M(s) = \frac{1}{1 + \alpha_2 \log(\frac{s+\beta_2}{\beta_2}) + c\{\alpha_1 \log(\frac{s+\beta_1}{\beta_1}) - \alpha_2 \log(\frac{s+\beta_2}{\beta_2})\}}.$$

(c) Geometric mixture inverse Gaussian: for $\delta_1, \delta_2 > 0$ and $\gamma_1, \gamma_2 > 0$,

$$\phi_M(s) = \frac{1}{1 + \delta_2\left(\sqrt{2s+\gamma_2^2} - \gamma_2\right) + c\left(\delta_1\left(\sqrt{2s+\gamma_1^2} - \gamma_1\right) - \delta_2\left(\sqrt{2s+\gamma_2^2} - \gamma_2\right)\right)}.$$

Remark 5.1. One can study the asymptotic behavior of gid mixture as done in Theorem 5.3. Also, further these gid mixtures can be used to define autoregressive models as discussed in following section.

5.3 Autoregressive models

In this section, we develop two type of processes namely the switching AR(1) process and AR(1) process with $\{Y_n\}$ as gid based marginals. Consider the following switching AR(1) process:

$$Y_n = \begin{cases} \epsilon_n, & \text{with probability } \theta, \\ Y_{n-1} + \epsilon_n, & \text{with probability } 1-\theta, \end{cases} \tag{5.2}$$

where $0 < \theta < 1$.

Theorem 5.4. *Consider a stationary switching AR(1) process $\{Y_n\}$ as defined in Eq. (5.2). If $\{Y_n\}$ has gid marginals, then $\{Y_n\}$ and $\{\epsilon_n\}$ have similar distribution.*

Proof. Assume that the marginal distribution of $\{Y_n\}$ is gid. Then the Laplace transform of $\{Y_n\}$ is $\phi_{Y_n}(s) = \frac{1}{1+g(s)}$ and innovation terms $\{\epsilon_n\}$ is denoted by $\phi_{\epsilon_n}(s)$. We can write Eq. (5.2) in terms of Laplace transform as,

$$\phi_{Y_n}(s) = \theta\phi_{\epsilon_n}(s) + (1-\theta)\phi_{Y_{n-1}}(s)\phi_{\epsilon_n}(s).$$

Using the stationarity condition and the Laplace transform $\phi_Y(s) = \frac{1}{1+g(s)}$, we rewrite above equation as,

$$\phi_Y(s) = \theta\phi_\epsilon(s) + (1-\theta)\phi_Y(s)\phi_\epsilon(s) = \phi_\epsilon(s)\{\theta + (1-\theta)\phi_Y(s)\} \tag{5.3}$$

$$\phi_\epsilon(s) = \frac{\phi_Y(s)}{\theta + (1-\theta)\phi_Y(s)} = \frac{1}{1+\theta g(s)}. \tag{5.4}$$

Hence, we conclude that innovation terms $\{\epsilon_n\}$ are gid with Laplace transform as $\frac{1}{1+\theta g(s)}$.

□

We find the joint Laplace transform of Y_n and Y_{n-1} and check for time reversibility of the process in the following result. The definition of time reversible time series is

defined in [99]. A time series, modelled by random variables $X_t, t = 0, \pm1, \pm2, \ldots$, is said to be reversible when, for all $r = 1, 2, \cdots$ and $t = 0, \pm1, \pm2, \ldots$, the joint distribution of $X_t, X_{t+1}, \cdots, X_{t+r}$ is equal to the joint distribution of $X_{t+r}, X_{t+r-1}, \cdots, X_t$. Using the joint Laplace transform to time series data, we can analyze the frequency-domain characteristics and dependencies between different variables.

Proposition 5.5. *Consider the stationary switching AR(1) process as defined in Eq. (5.2). Y_n can be written as $Y_n = I_n Y_{n-1} + \epsilon_n$, where $P[I_n = 0] = \theta = 1 - P[I_n = 1], 0 < \theta < 1$. Then the AR(1) process is not time reversible.*

Proof. The Laplace transform of joint variables (Y_n, Y_{n-1}) is calculated as follows:

$$\begin{aligned}\phi_{Y_{n-1},Y_n}(s_1, s_2) &= \mathbb{E}[\exp(-s_1 Y_{n-1} - s_2 Y_n)] = \mathbb{E}[\exp(-s_1 Y_{n-1} - s_2[I_n Y_{n-1} + \epsilon_n]] \\ &= \mathbb{E}[\exp(-(s_1 + s_2 I_n) Y_{n-1} - s_2 \epsilon_n)] = \mathbb{E}[\exp(-(s_1 + s_2 I_n) Y_{n-1})] \phi_{\epsilon_n}(s_2) \\ &= \frac{1}{1 + \theta g(s_2)} \Big[\frac{\theta}{1 + g(s_1)} + \frac{1 - \theta}{1 + g(s_1 + s_2)} \Big].\end{aligned}$$

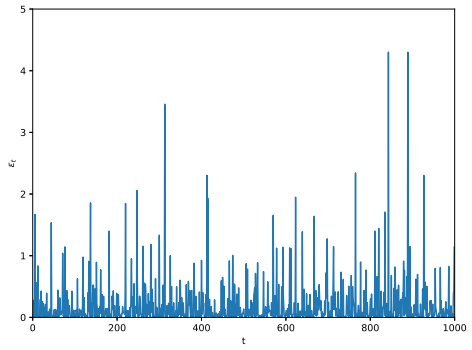
Since the joint Laplace transform is not a symmetric function, the process is not time reversible. □

To simulate the switching AR(1) model, first generate the i.i.d. innovation terms $\{\epsilon_t\}$ using the Laplace transform as discussed in [120]. Then generate $\{Y_t\}$ from AR(1) model defined in Eq. (5.2) with $\theta = 0.3$. The plot of innovation terms $\{\epsilon_t\}$ for geometric tempered stable marginals, geometric gamma marginals and geometric inverse Gaussian are shown in Fig. 5.2. The time series $\{Y_t\}$ from switching AR(1) model with $\theta = 0.3$ for all the three cases are shown in Fig. 5.3. In the next results, we discuss the form of the pdf of innovation terms using Laplace transform and complex inversion formula. The Cauchy inversion formula is a tool for computing the Laplace inverse of $F(s)$. For a continuous function f possessing a Laplace transform $F(s)$, the Laplace inverse $f(t)$ is given by

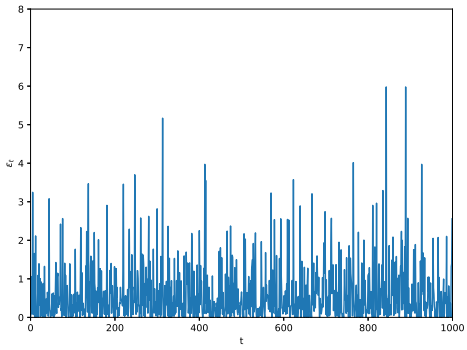
$$f(t) = \lim_{y \to \infty} \frac{1}{2\pi\iota} \int_{x-\iota y}^{x+\iota y} e^{st} F(s)\, ds,$$

where $s = x + \iota y$. For theory of complex inversion formula, one can refer Chap. 4 of [127].

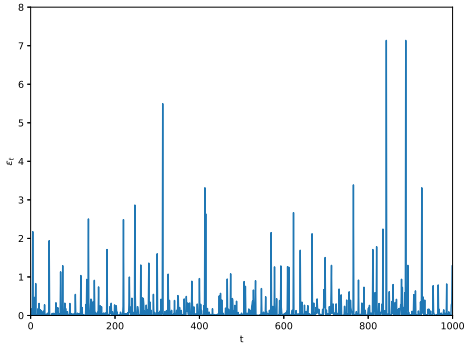
Theorem 5.6. *Consider a stationary switching AR(1) process $\{Y_n\}$ as defined in Eq. (5.2). If $\{Y_n\}$ is marginally distributed as geometric tempered stable with Laplace transform $\phi_Y(s) = \frac{1}{1+(s+\lambda)^\beta - \lambda^\beta}$, for $\beta = \frac{1}{m}, m = 2, 3, \cdots$, then the innovation terms $\{\epsilon_n\}$ also follow geometric tempered stable distribution with Laplace transform*



(a) Geometric tempered stable.



(b) Geometric gamma.



(c) Geometric inverse Gaussian.

Figure 5.2: The plot of innovation terms $\{\epsilon_t\}$ from AR(1) model with $\theta = 0.3$ for (a) geometric tempered stable with parameters $\beta = 0.6$ and $\lambda = 1$ (b) geometric gamma with parameters $\beta = 5$ and $\alpha = 10$ (c) geometric inverse Gaussian with parameters $\gamma = 1$ and $\delta = 0.5$.

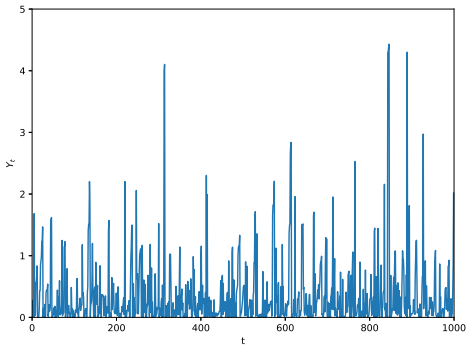
$\phi_\epsilon(s) = \frac{1}{1+\theta((s+\lambda)^\beta - \lambda^\beta)}$. *Moreover, the pdf of innovation terms has the following integral form:*

$$f_\epsilon(x) = \frac{e^{(s_0-\lambda)x}}{\theta\beta s_0^{\beta-1}} + \frac{1}{\pi}\int_0^\infty \frac{\theta e^{-x(y+\lambda)} y^\beta \sin(\pi\beta)}{1 + 2\theta(y^\beta\cos(\pi\beta) - \lambda^\beta) + \theta^2(y^{2\beta} - 2\lambda^\beta y^\beta \cos(\pi\beta) + \lambda^{2\beta})}\, dy. \tag{5.5}$$

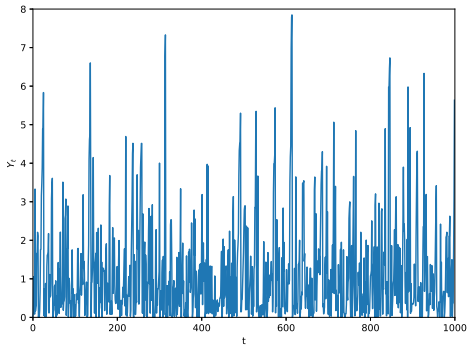
where $s_0 = (\lambda^\beta - 1/\theta)^{\beta-1}, \beta \in (0,1), \beta = \frac{1}{m}, m = 2, 3, \cdots$ *and* $\lambda > 0$.

Proof. We substitute $g(s) = \frac{1}{1+(s+\lambda)^\beta - \lambda^\beta}$ in Eq. (5.4) to obtain the Laplace transform of innovation terms $\phi_\epsilon(s) = \frac{1}{1+\theta((s+\lambda)^\beta - \lambda^\beta)}$, $\beta = \frac{1}{m}$, $m = 2, 3, 4, \ldots$.
Consider the function $G(s) = \frac{1}{1+\theta(s^\beta - \lambda^\beta)}$. For the function $G(s)$, with $\beta = \frac{1}{m}$, $m = 2, 3, \ldots$, it follows that $s_0 = (\lambda^\beta - 1/\theta)^{1/\beta}$ is a simple pole and $s_1 = (0, 0)$ is the branch point. We use complex inversion formula to compute the Laplace inverse of $G(s)$ which in turn gives the pdf $f_\epsilon(x)$ of innovation terms

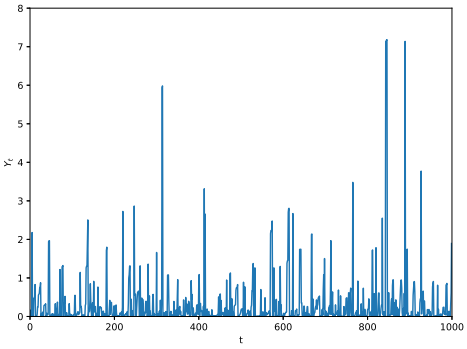
$$f_\epsilon(x) = \frac{1}{2\pi\iota}\int_{x_0-i\infty}^{x_0+i\infty} e^{sx}\phi_\epsilon(s)\, ds,$$



(a) Geometric tempered stable.



(b) Geometric gamma.



(c) Geometric inverse Gaussian.

Figure 5.3: The plot of time series $\{Y_t\}$ of length 1000 from AR(1) model with $\theta = 0.3$ for (a) geometric tempered stable with parameters $\beta = 0.6$ and $\lambda = 1$ (b) geometric gamma with parameters $\beta = 5$ and $\alpha = 10$ (c) geometric inverse Gaussian with parameters $\gamma = 1$ and $\delta = 0.5$.

where $x_0 > a$ is chosen such that the integrand is analytic for $Re(s) > a$. Consider the contour $ABCDEF$ in Fig. 5.4 with branch point $s_1 = (0, 0)$, circular arcs AB and EF of radius R, arc CD of radius r, line segments BC and DE parallel to $x-$axis and AF is line segment from $x_0 - iy$ to $x_0 + iy$. For closed curve C and poles s_i inside C, Cauchy residue theorem states that,

$$\frac{1}{2\pi\iota}\oint_C e^{sx}G(s)\,ds = \sum_{i=1}^{n} \text{Res}(e^{sx}G(s), s_i).$$

In limiting case, for contour $ABCDEF$, the integral on circular arcs AB and EF tend to 0 as $R \to \infty$. The integral over CD also tends to 0 as $r \to 0$. We need to compute the following integral:

$$\frac{1}{2\pi\iota}\int_{x_0-\iota\infty}^{x_0+\iota\infty} e^{sx}G(s)\,ds = \text{Res}(e^{sx}G(s), s_0) - \frac{1}{2\pi\iota}\int_{BC} e^{sx}G(s)\,ds - \frac{1}{2\pi\iota}\int_{DE} e^{sx}G(s)\,ds. \tag{5.6}$$

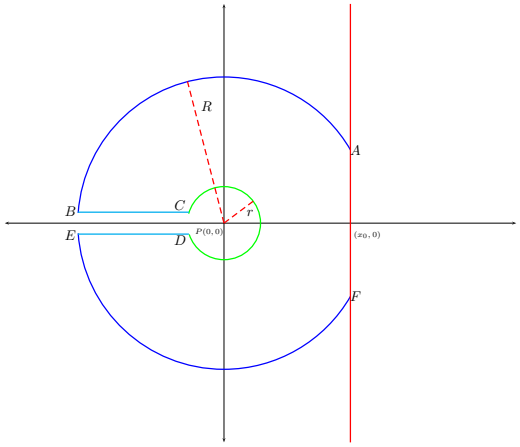


Figure 5.4: Contour plot with branch point at $P = (0, 0)$ in anti-clockwise direction

Now, along BC, let $s = ye^{\iota\pi}$, then $ds = -dy$ and,

$$\frac{1}{2\pi\iota}\int_{-R}^{-r}\frac{e^{sx}}{1+\theta(s^\beta-\lambda^\beta)}\,ds = \frac{1}{2\pi\iota}\int_{r}^{R}\frac{e^{-xy}}{1+\theta(y^\beta e^{\iota\pi\beta}-\lambda^\beta)}\,dy = \frac{1}{2\pi\iota}\int_{0}^{\infty}\frac{e^{-xy}}{1+\theta(y^\beta e^{\iota\pi\beta}-\lambda^\beta)}\,dy. \tag{5.7}$$

Along DE, let $s = ye^{-\iota\pi}$, then $ds = -dy$ and,

$$-\frac{1}{2\pi\iota}\int_{-R}^{-r}\frac{e^{sx}}{1+\theta(s^\beta-\lambda^\beta)}\,ds = -\frac{1}{2\pi\iota}\int_{r}^{R}\frac{e^{-xy}}{1+\theta(y^\beta e^{-\iota\pi\beta}-\lambda^\beta)}\,dy \tag{5.8}$$

Now substitute Eq. (5.7) and (5.8) in Eq. (5.6).

$$\begin{aligned}\frac{1}{2\pi\iota}\int_{x_0-\iota\infty}^{x_0+\iota\infty}e^{sx}G(s)\,ds &= \mathrm{Res}(e^{sx}G(s), s_0) - \frac{1}{2\pi\iota}\Big[\int_0^\infty\frac{e^{-xy}}{1+\theta(y^\beta e^{\iota\pi\beta}-\lambda^\beta)}\,dy\\ &\quad -\int_0^\infty\frac{e^{-xy}}{1+\theta(y^\beta e^{-\iota\pi\beta}-\lambda^\beta)}\,dy\Big]\\ &= \mathrm{Res}(e^{sx}G(s), s_0) - \frac{1}{2\pi\iota}\int_0^\infty\frac{e^{-xy}(\theta y^\beta(e^{-\iota\pi\beta}-e^{\iota\pi\beta}))}{(1+\theta(y^\beta e^{\iota\pi\beta}-\lambda^\beta))(1+\theta(y^\beta e^{-\iota\pi\beta}-\lambda^\beta))}\,dy\\ &= \mathrm{Res}(e^{sx}G(s), s_0) + \frac{1}{\pi}\int_0^\infty\frac{e^{-xy}\theta y^\beta\sin(\pi\beta)}{1+2\theta(y^\beta\cos\pi\beta-\lambda^\beta)+\theta^2(y^{2\beta}-2\lambda^\beta y^\beta\cos\pi\beta+\lambda^{2\beta})}\,dy.\end{aligned}$$

Now we find the $\mathrm{Res}(e^{sx}G(s))$ for poles at $s_0 = (\lambda^\beta - 1/\theta)^m$, $\beta = 1/m$, $m = 2, 3, 4, \cdots$. Also, note that all the poles $s_0 < 0$ are outside the analytic region, therefore we will calculate the residue at all those s_0 where $(\lambda^\beta - 1/\theta)^m > 0$ i.e. $\lambda > \frac{1}{\theta^m}$. We evaluate the residue as,

$$\lim_{s\to s_0}\frac{(s-s_0)e^{sx}}{1+\theta(s^\beta-\lambda^\beta)} = \frac{e^{s_0x}}{\theta\beta(s_0)^{\beta-1}},$$

where $s_0 = (\lambda^\beta - 1/\theta)^m$. The inverse Laplace of $G(s)$ is

$$\mathcal{L}^{-1}\{G(s)\} = \frac{e^{s_0x}}{\theta\beta(s_0)^{\beta-1}} + \frac{1}{\pi}\int_0^\infty\frac{e^{-xy}\theta y^\beta\sin(\pi\beta)}{1+2\theta(y^\beta\cos\pi\beta-\lambda^\beta)+\theta^2(y^{2\beta}-2\lambda^\beta y^\beta\cos\pi\beta+\lambda^{2\beta})}\,dy.$$

We obtain the pdf by using the first translational property of inverse Laplace.

$$f_\epsilon(x) = \mathcal{L}^{-1}\{G(s+\lambda)\} = e^{-\lambda x}\mathcal{L}^{-1}G(s)$$
$$= e^{-\lambda x}\Big[\frac{e^{s_0 x}}{\theta\beta(s_0)^{\beta-1}} + \frac{1}{\pi}\int_0^\infty \frac{e^{-xy}\theta y^\beta \sin(\pi\beta)}{1+2\theta(y^\beta\cos\pi\beta-\lambda^\beta)+\theta^2(y^{2\beta}-2\lambda^\beta y^\beta\cos\pi\beta+\lambda^{2\beta})}\,dy\Big].$$

Hence, we obtain the form of pdf as mentioned in Eq. (5.5). □

Remark 5.2. For $\lambda = 0, \phi_Y(s) = \frac{1}{1+s^\beta}$ and $\phi_\epsilon(s) = \frac{1}{1+\theta s^\beta}$, which is the Laplace transform of one side geometric stable distribution also known as Mittag-Leffler distribution. For $\theta = 1$, poles will be $s_0 = (-1)^m$, $m = 2, 3, 4\cdots$. The residue for the pole corresponding to $s_0 = 1$ is e^x. Then, the pdf becomes,

$$f_\epsilon(x) = e^x + \frac{1}{\pi}\int_0^\infty \frac{e^{-xy}y^\beta\sin(\pi\beta)}{1+2y^\beta\cos(\pi\beta)+y^{2\beta}}\,dy, \text{ where } s_0 = 1. \tag{5.9}$$

Remark 5.3. In Eq. (5.9), we take $\beta = 1/2$ or equivalently $m = 2$ and obtain the density using the results from [3](pp. 303-304),

$$f_\epsilon(x) = e^x + \frac{1}{\pi}\int_0^\infty \frac{e^{-xy}\sqrt{y}}{1+y}\,dy = e^x + \frac{1}{\sqrt{\pi x}} - e^x + e^x Erf(\sqrt{x}) = \frac{1}{\sqrt{\pi x}} + e^x Erf(\sqrt{x}).$$

As $x \to 0$, $Erf(\sqrt{x}) \to 0$ and $f_\epsilon(x) \sim \frac{1}{\sqrt{\pi x}}$, which match with the asymptotic behaviour discussed in Theorem 5.3.

Proposition 5.7. *For a stationary AR(1) process with $\{Y_n\}$ defined as in Eq. (5.2) with marginals distributed as geometric gamma with Laplace transform $\phi_Y(s) = \frac{1}{1+\alpha\log(\frac{s+\beta}{\beta})}$, then the innovation terms $\{\epsilon_n\}$ also follow geometric gamma with Laplace transform $\phi_\epsilon(s) = \frac{1}{1+\alpha\theta\log(\frac{s+\beta}{\beta})}$. Moreover, the pdf of innovation terms has the following integral form:*

$$f_\epsilon(x) = \frac{e^{s_0 x}(s_0+\beta)^{\alpha\theta}}{\beta^{\alpha\theta}} - e^{-\beta x}\int_0^\infty \frac{\alpha\theta e^{-xy}}{1+2\alpha\theta\log(y/\beta)+\pi^2\alpha^2\theta^2+\alpha^2\theta^2(\log(y/\beta))^2}\,dy,$$

where $s_0 = \beta(e^{-1/\alpha\theta}-1)$, $\alpha > 0$ and $\beta > 0$.

Proof. Again we substitute $g(s) = \frac{1}{1+\alpha\log(\frac{s+\beta}{\beta})}$ in Eq. (5.4) and get the Laplace transform of innovation terms $\phi_\epsilon(s) = \frac{1}{1+\alpha\theta\log(\frac{s+\beta}{\beta})}$. Again we use the complex inversion formula to obtain the pdf of innovation terms. We consider the function $G(s) = \frac{1}{1+\theta\alpha\log(\frac{s+\beta}{\beta})}$ and the pole for $G(s)$ at $s_0 = \beta(e^{-1/\alpha\theta}-1)$ and branch point at $s = -\beta$. The residue corresponding to s_0 is $\frac{e^{s_0 x}(s_0+\beta)^{\alpha\theta}}{\beta^{\theta\alpha}}$. Using a similar contour as given in Fig. 5.4 and the same steps to obtain the pdf as

$$f_\epsilon(x) = \frac{e^{s_0 x}(s_0+\beta)^{\alpha\theta}}{\beta^{\alpha\theta}} - e^{-\beta x}\int_0^\infty \frac{\alpha\theta e^{-xy}}{1+2\alpha\theta\log(y/\beta)+\pi^2\alpha^2\theta^2+\alpha^2\theta^2(\log(y/\beta))^2}\,dy.$$

□

Proposition 5.8. *For a stationary AR(1) process with* $\{Y_n\}$ *defined as in Eq.* (5.2) *with marginals distributed as geometric inverse Gaussian with Laplace transform* $\phi_Y(s) = \frac{1}{1+\gamma\delta\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$, *then the innovation terms* $\{\epsilon_n\}$ *also follows geometric inverse Gaussian with Laplace transform* $\phi_\epsilon(s) = \frac{1}{1+\theta\gamma\delta\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$. *Moreover, the pdf of innovation terms has the following integral form:*

$$f_\epsilon(x) = \frac{e^{s_0 x}\sqrt{2s_0+\gamma^2}}{\theta\delta} + \frac{\gamma^2 e^{-x}}{2\pi}\int_0^\infty \frac{\theta\delta\sqrt{y}e^{-\frac{2xy}{\gamma^2}}}{1-2\theta\delta\gamma+\theta^2\delta^2(y+\gamma^2-2\gamma\sqrt{y}\cos\theta)}\,dy,$$

where $s_0 = \frac{\gamma^2}{2}(1-\frac{1}{\theta\gamma\delta})^2 - \frac{\gamma^2}{2}$, $\delta > 0$ *and* $\gamma > 0$.

Proof. The Laplace transform of innovation terms $\phi_\epsilon(s) = \frac{1}{1+\theta\gamma\delta\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$ is straight forward from Eq. (5.4). Now we use the complex inversion formula to obtain the pdf of innovation terms as done in previous theorem. We consider the function $G(s) = \frac{1}{1+\theta\delta\gamma\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$ and the pole for $G(s)$ is $s_0 = \frac{\gamma^2}{2}\left\{1-\frac{1}{\theta\delta\gamma}\right\} - \frac{\gamma^2}{2}$. The residue corresponding to s_0 is $\frac{e^{s_0 x}\sqrt{2s_0+\gamma^2}}{\theta\delta}$. We use the same steps to obtain the pdf as

$$f_\epsilon(x) = \frac{e^{s_0 x}\sqrt{2s_0+\gamma^2}}{\theta\delta} + \frac{\gamma^2 e^{-x}}{2\pi}\int_0^\infty \frac{\theta\delta\sqrt{y}e^{-\frac{2xy}{\gamma^2}}}{1-2\theta\delta\gamma+\theta^2\delta^2(y+\gamma^2-2\gamma\sqrt{y}\cos\theta)}\,dy.$$

□

We know that for a random process, the moments (if exist) uniquely describe the properties of the random variable X. In the following result, we provide the kth order moments using the Laplace transform. The kth order moment of random variable X is evaluated as: $\mathbb{E}(X^k) = (-1)^k\phi^{(k)}(s)$ for $s = 0$ and $k \in \mathbb{N}$. We obtain the first and second order moments for these cases as follows:

(a) Geometric tempered stable innovations: $\phi_\epsilon(s) = \dfrac{1}{1+\theta\{(s+\lambda)^\beta - \lambda^\beta\}}$. Then,
$\mathbb{E}(\epsilon) = \theta\beta\lambda^{\beta-1}, \quad \mathbb{E}(\epsilon^2) = \theta\beta(\beta-1)\lambda^{\beta-2} - 2\theta^2\beta^2(\lambda)^{2\beta-2}, \lambda > 0.$

(b) Geometric gamma innovations: $\phi_\epsilon(s) = \dfrac{1}{1+\alpha\theta\log\left(\dfrac{s+\beta}{\beta}\right)}$. Then,
$\mathbb{E}(\epsilon) = \dfrac{\theta\alpha}{\beta}, \quad \mathbb{E}(\epsilon^2) = \dfrac{\theta\alpha+2\theta^2\alpha^2}{\beta^2}.$

(c) Geometric inverse Gaussian innovations: $\phi_\epsilon(s) = \dfrac{1}{1+\theta\gamma\delta\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$.

Then,
$\mathbb{E}(\epsilon) = \dfrac{\theta\delta}{\gamma}, \quad \mathbb{E}(\epsilon^2) = \dfrac{2\theta^2\delta^2}{\gamma^2} + \dfrac{\theta\delta}{\gamma^3}.$

5.3.1 Generalisation to kth order autoregressive processes

We define the generalized form of the process developed in previous section by Eq. (5.2) as follows:

$$Y_n = \begin{cases} \epsilon_n, & \text{with probability } \theta, \\ Y_{n-1} + \epsilon_n, & \text{with probability } 1-\theta_1, \\ \vdots & \\ Y_{n-k} + \epsilon_n, & \text{with probability } 1-\theta_k, \end{cases}, \tag{5.10}$$

where $\sum_{i=1}^{k}\theta_i = 1-\theta,\ 0<\theta<1,\ i=1,2,\ldots,k$. Also, $\{\epsilon_n\}$ are independent of $\{Y_{n-1}, Y_{n-2}, \cdots,\}$. We write the Laplace transform $\phi_{Y_n}(s)$ for the model defined in Eq. (5.10),

$$\begin{aligned}\phi_{Y_n}(s) &= \theta\phi_{\epsilon_n}(s) + \theta_1\phi_{Y_{n-1}}(s)\phi_{\epsilon_n}(s) + \cdots + \theta_k\phi_{Y_{n-k}}(s)\phi_{\epsilon_n}(s) \\ &= \phi_{\epsilon_n}(s)\{\theta + \theta_1\phi_{Y_{n-1}(s)} + \cdots + \theta_k\phi_{Y_{n-k}(s)}\}.\end{aligned}$$

Since series is stationary, we get,

$$\phi_Y(s) = \phi_\epsilon(s)\theta + \sum_{i=1}^{k}\theta_i\phi_Y(s) = \phi_\epsilon(s)\{\theta+(1-\theta)\phi_Y(s)\} \implies \phi_\epsilon(s) = \frac{\phi_Y(s)}{\theta+(1-\theta)\phi_Y(s)}.$$

Hence, we obtain the similar form for innovation terms $\{\epsilon_n\}$. Next, we define the stationary AR(1) process with $|\theta|<1$ and find the Laplace transform of the innovation terms.

Proposition 5.9. *Consider the AR(1) process $Y_n = \theta Y_{n-1} + \epsilon_n$, $|\theta|<1$ is strictly stationary with Laplace transform of marginals as $\phi_Y(s) = \frac{1}{1+g(s)}$ then the Laplace transform of innovation terms $\{\epsilon_n\}$ is $\phi_\epsilon(s) = \frac{1+g(\theta s)}{1+g(s)}$.*

Proof. We have stationary AR(1) process with $|\theta|<1$, then Laplace is defined as

$$\phi_{Y_n}(s) = \phi_{Y_{n-1}}(\theta s)\phi_\epsilon(s);\ \phi_\epsilon(s) = \frac{\phi_Y(s)}{\phi_Y(\theta s)} = \frac{1+g(\theta s)}{1+g(s)}.$$

□

The Laplace transform of innovation terms $\{\epsilon_n\}$ for three cases are as follows:

(a) Geometric tempered stable: for $g(s) = (s+\lambda)^{\beta} - \lambda^{\beta}$, then $\phi_{\epsilon}(s) = \frac{1+(\theta s+\lambda)^{\beta}-\lambda^{\beta}}{1+(s+\lambda)^{\beta}-\lambda^{\beta}}$, $\lambda > 0, \beta \in (0,1)$.

(b) Geometric gamma: for $g(s) = \alpha \log(1+\frac{s}{\beta})$, then $\phi_{\epsilon}(s) = \frac{1+\alpha \log\left(\frac{\theta s+\beta}{\beta}\right)}{1+\alpha \log\left(\frac{s+\beta}{\beta}\right)}$, $\alpha > 0$ and $\beta > 0$.

(c) Geometric inverse Gaussian: for $g(s) = \delta\gamma\left\{\sqrt{1+\frac{2s}{\gamma^2}} - 1\right\}$, then $\phi_{\epsilon}(s) = \frac{1+\delta\gamma\left\{\sqrt{1+\frac{2\theta s}{\gamma^2}}-1\right\}}{1+\delta\gamma\left\{\sqrt{1+\frac{2s}{\gamma^2}}-1\right\}}$, $\delta, \gamma > 0$.

5.4 Parameter estimation and simulation study

In this section, we estimate the parameters of the model defined in Prop. 5.9 using conditional least square (CLS) and method of moments (MOM). We first apply the CLS method to estimate the parameter θ and then use the MOM for the parameters of marginals in next subsection.

5.4.1 Estimation by conditional least squares and method of moments

The function for conditional least square is given by,

$$L(\theta, \lambda, \beta) = \sum_{t=1}^{n} (Y_t - \mathbb{E}(Y_t|Y_{t-1}))^2, \text{ where } \mathbb{E}[Y_t|Y_{t-1}] = \theta Y_{t-1} + \mathbb{E}(\epsilon_t).$$

(a) Geometric tempered stable: For the model defined in Prop. 5.9, first we assume that innovations are from distribution with Laplace transform $\phi_{\epsilon}(s) = \frac{1+(\theta s+\lambda)^{\beta}-\lambda^{\beta}}{1+(s+\lambda)^{\beta}-\lambda^{\beta}}$, $\lambda > 0$ and $\beta \in (0,1)$. The first order and second order theoretical moments of innovation terms $\{\epsilon_t\}$ are approximated by the empirical moments, which are given by

$$\hat{m}_1 = \sum_{t=1}^{n} \frac{\epsilon_t}{n} = (1-\theta)\beta\lambda^{\beta-1}; \ \hat{m}_2 = \sum_{t=1}^{n} \frac{\epsilon_t^2}{n} = \beta(\beta-1)(\theta^2-1)\lambda^{\beta-2} - 2\beta^2\lambda^{2\beta-2}(\theta-1). \tag{5.11}$$

We substitute $\hat{m}_1$ from Eq. (5.11) to the function $L(\theta, \lambda, \beta)$ and obtain,

$$L = \sum_{t=1}^{n} (Y_t - \theta Y_{t-1} - (1-\theta)\beta\lambda^{\beta-1})^2.$$

Take derivative with respect to unknown parameters θ, λ and β and obtain the following relation,

$$\frac{\partial L}{\partial \lambda} = \sum_{t=1}^{n} 2(Y_t - \theta Y_{t-1} - (1-\theta)\beta\lambda^{\beta-1})((1-\theta)\beta\lambda^{\beta-1}), \tag{5.12}$$

$$\frac{\partial L}{\partial \beta} = \sum_{t=1}^{n} 2(Y_t - \theta Y_{t-1} - (1-\theta)\beta\lambda^{\beta-1})((1-\theta)\lambda^{\beta-1} + (1-\theta)\beta\lambda^{\beta-1}\log(\beta-1)), \tag{5.13}$$

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^{n} 2(Y_t - \theta Y_{t-1} - (1-\theta)\beta\lambda^{\beta-1})(\beta\lambda^{\beta-1} - Y_{t-1}). \tag{5.14}$$

Solving the above equations, we get the estimate, $\hat{\theta} = \frac{\sum_{t=1}^{n} Y_t Y_{t-1} - n\bar{Y}_{t-1}^2}{\sum_{t=1}^{n}(Y_{t-1} - \bar{Y}_{t-1})^2}$, where $\bar{Y}_{t-1} = \frac{\sum_{t=1}^{n} Y_{t-1}}{n}$. Now we use first and second order moments to estimate the remaining parameters λ and β.

$$\hat{m}_1 = (1-\hat{\theta})\beta\lambda^{\beta-1};\ \hat{m}_2 = \beta(\beta-1)(\hat{\theta}^2-1)\lambda^{\beta-2} - 2\beta^2\lambda^{2\beta-2}(\hat{\theta}-1).$$

After rearranging the terms we get the following non-linear relation between λ and β,

$$\beta = 1 - \lambda\frac{\hat{m}_2(1-\hat{\theta}) - 2\hat{m}_1^2}{\hat{m}_1(1-\hat{\theta}^2)}.$$

To solve it further we use *fsolve()* function available in python scipy package. Also, note that the estimate for θ using CLS will be same for all the cases.

(b) Geometric gamma: For this case the Laplace transform of innovation terms are $\phi_\epsilon(s) = \frac{1+(\log(\theta s+\beta/\beta))^\alpha}{1+(\log(s+\beta/\beta))^\alpha}$, $\alpha > 0$ and $\beta > 0$. The first and second order moments of innovation terms will be,

$$\hat{m}_1 = (1-\hat{\theta})\frac{\alpha}{\beta};\ \hat{m}_2 = \frac{\alpha(1-\hat{\theta}^2)}{\beta^2} + \frac{2\alpha^2(1-\hat{\theta})}{\beta^2}.$$

Using these moments we get the the estimates as $\hat{\beta} = \frac{\hat{m}_1(1-\hat{\theta}^2)}{\hat{m}_2(1-\hat{\theta}-2\hat{m}_1^2)}$ and $\hat{\alpha} = \frac{\hat{m}_1\hat{\beta}}{1-\hat{\theta}}$.

(c) Geometric inverse Gaussian: The Laplace transform of innovation terms are $\phi_\epsilon(s) = \frac{1+\delta(\sqrt{2\theta s+\gamma^2}-\gamma)}{1+\delta(\sqrt{2s+\gamma^2}-\gamma)}$, $\gamma > 0$ and $\delta > 0$. The first and second order

moments of innovation terms will be,

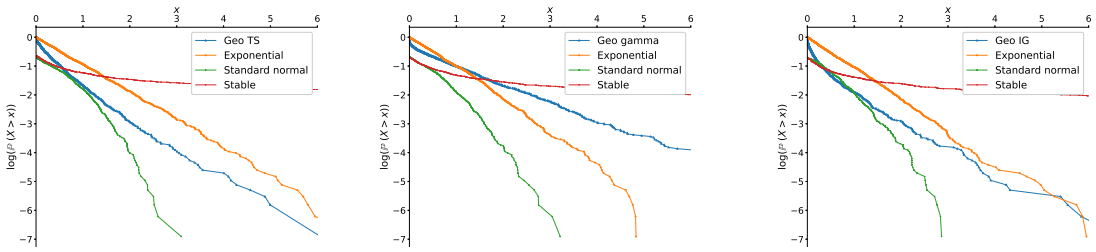
$$\hat{m}_1 = (1-\hat{\theta})\frac{\delta}{\gamma};\ \hat{m}_2 = 2(1-\hat{\theta})\frac{\delta^2}{\gamma^2} + (1-\hat{\theta}^2)\frac{\delta}{\gamma^3}.$$

Using these moments we get the the estimates as $\hat{\gamma} = \sqrt{\dfrac{\hat{m}_1(1-\hat{\theta}^2)}{\hat{m}_2(1-\hat{\theta}) - 2\hat{m}_1^2}}$ and $\hat{\delta} = \dfrac{\hat{m}_1\hat{\gamma}}{1-\hat{\theta}}$.

In the next subsection we observe the tail behavior of geometric tempered stable, geometric gamma and geometric inverse Gaussian empirically. We also present the simulation study for AR(1) model with these distributions.

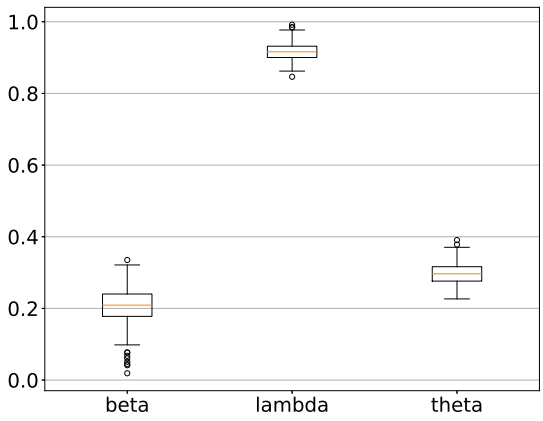
5.4.2 Simulation study

We first generate the random variables from stable, standard normal and standard exponential distribution each of size 1000. We generate 1000 random variables from all the proposed three distributions. We compare the log of tail probability of these distribution with the stable, standard normal and standard exponential distribution. Observe from Fig. 5.5 that geometric tempered stable and geometric inverse Gaussian and geometric gamma distributions lie between standard normal and stable distribution. This behavior depicts that all the three distributions are from semi heavy-tailed distributions. Also, observe that geometric gamma distribution is heavy-tailed as it lies above standard normal and standard exponential distribution. Therefore, we can empirically conclude that geometric gamma, geometric tempered stable and geometric inverse Gaussian belong to the class of semi heavy-tailed distribution.

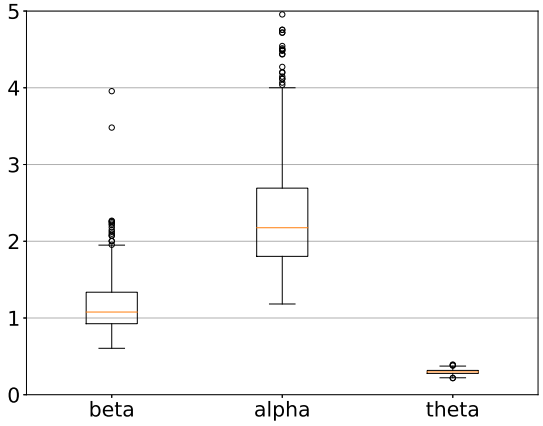


(a) Geo tempered stable. (b) Geo gamma. (c) Geo inverse Gaussian.

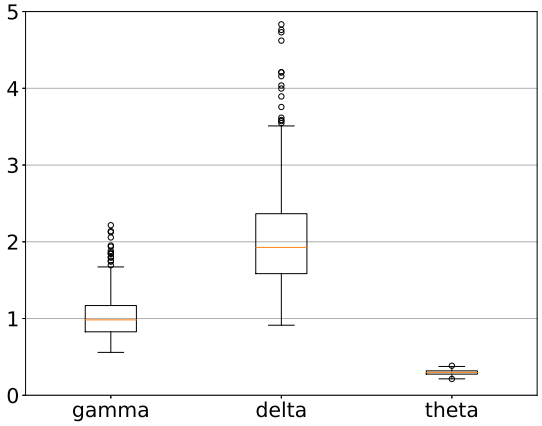
Figure 5.5: The plot of $\log(\mathbb{P}(X > x))$, where X is random variables of length 1000 (a) geometric tempered stable with parameters $\beta = 0.6$ and $\lambda = 1$ (b) geometric gamma with parameters $\beta = 1$ and $\alpha = 1$ (c) geometric inverse Gaussian with parameters $\gamma = 1$ and $\delta = 0.5$.



(a) Geometric tempered stable.



(b) Geometric gamma.



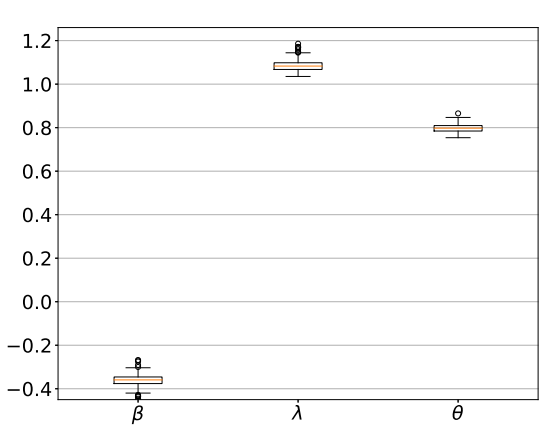
(c) Geometric inverse Gaussian.

Figure 5.6: The boxplots for model parameter estimates from AR(1) model with true value of $\theta = 0.3$ for (a) geometric tempered stable with true parameter values of $\beta = 0.6$ and $\lambda = 1$ (b) geometric gamma with true parameter values of $\beta = 1$ and $\alpha = 2$ (c) geometric inverse Gaussian with parameters $\gamma = 1$ and $\delta = 2$

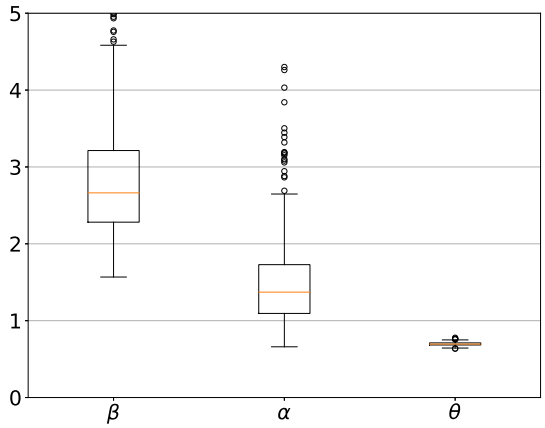
Table 5.1: Case 1: Estimation of parameters using CLS and MOM for AR(1) model with three distributions.

Distributions		parameter 1	parameter 2	parameter 3
Geometric tempered stable	True values	$\beta = 0.6$	$\lambda = 1$	$\theta = 0.3$
	Est. values	$\hat{\beta} = 0.206$	$\hat{\lambda} = 0.917$	$\hat{\theta} = 0.297$
Geometric gamma	True values	$\beta = 1$	$\alpha = 2$	$\theta = 0.3$
	Est. values	$\hat{\beta} = 1.189$	$\hat{\alpha} = 2.422$	$\hat{\theta} = 0.298$
Geometric inverse Gaussian	True values	$\gamma = 1$	$\delta = 2$	$\theta = 0.3$
	Est. values	$\hat{\gamma} = 1.028$	$\hat{\delta} = 2.054$	$\hat{\theta} = 0.296$

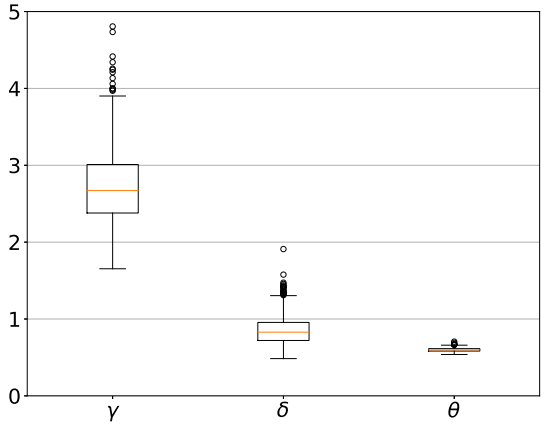
We use simulation to further assess the performance of the estimation method.



(a) Geometric tempered stable.



(b) Geometric gamma.



(c) Geometric inverse Gaussian.

Figure 5.7: Case 2: The boxplots for model parameter estimates from AR(1) model (a) geometric tempered stable with true parameter values of $\theta = 0.8$, $\beta = 0.3$ and $\lambda = 3$ (b) geometric gamma with true parameter values of $\theta = 0.7$, $\beta = 2.5$ and $\alpha = 1.3$ (c) geometric inverse Gaussian with parameters $\theta = 0.6$, $\gamma = 1.6$ and $\delta = 0.5$.

We simulate the 500 trajectories each of length 1000 for geometric tempered stable, geometric gamma and geometric inverse Gaussian. We simulate two sets of data for assessment. We use the method of Laplace transform to simulate the innovation terms $\{\epsilon_n\}$ as described in [120] and then generate the time series $\{Y_n\}$ from AR(1) model. The true values for the parameters of both the cases are tabulated in Table 5.1 and 5.2. For case 1, the Table 5.2 and boxplots in Fig. 5.6 we observe that the CLS method gives good estimate for parameter θ, whereas the estimation of other parameters by MOM has variance. The estimated value of β from geometric tempered stable is not good. The relation of β and λ is non-linear therefore we solved it by using *fsolve* function defined in python's scipy library. The method to solve non-linear relationship amongst parameters are based on numerical methods. For case 2, the simulated data do not have randomness which results in poor estimation of parameters. The results using boxplots are shown in Fig. 5.7 and numerical

Table 5.2: Case 2: Estimation of parameters using CLS and MOM for AR(1) model with three distributions.

Distributions		parameter 1	parameter 2	parameter 3
Geometric tempered stable	True values	$\beta = 0.3$	$\lambda = 3$	$\theta = 0.8$
	Est. values	$\hat{\beta} = -0.362$	$\hat{\lambda} = 1.085$	$\hat{\theta} = 0.797$
Geometric gamma	True values	$\beta = 2.5$	$\alpha = 1.3$	$\theta = 0.7$
	Est. values	$\hat{\beta} = 2.870$	$\hat{\alpha} = 1.531$	$\hat{\theta} = 0.697$
Geometric inverse Gaussian	True values	$\gamma = 1.6$	$\delta = 0.5$	$\theta = 0.6$
	Est. values	$\hat{\gamma} = 2.734$	$\hat{\delta} = 0.859$	$\hat{\theta} = 0.599$

values are also given in Table 5.2. We observe that for geometric inverse Gaussian distribution, the estimates are not satisfactory.

The reason for these outcomes may depend on the simulated data because the simulation technique is based on the inversion method using a modified Newton-Raphson method. The values of the distribution and density functions are obtained by numerical transform inversion. Therefore, there is non randomness in the simulated data which results in poor estimates for some of the parameters.

From the simulation study we conclude that we can rely on the estimation techniques for modeling AR(1) model with these distributions. Further, we also need another simulation method for these distributions for future use.

5.5 Conclusion

We use the Bernstein function $g(s)$ which is the Laplace exponent of a positive infinitely divisible random variable to define gid random variables with Laplace transform of the form $\frac{1}{1+g(s)}$. We also find the Laplace transform of mixtures of some particular gid random variables which is a new class of marginals to study. A new autoregressive process of order 1 with gid distribution is considered. We deduce that if marginals of AR(1) defined in Eq. (5.2) are gid then the innovation terms are also gid. We find the integral form of the pdf of innovation terms using the Laplace transform and complex inversion method for three cases namely, geometric tempered stable, geometric gamma, and geometric inverse Gaussian subordinators. Further, we have also calculated the first and second order moments for these three gid random variables which play an important role in studying the characteristics of pdf. Next, we generalized the AR process to k^{th} order and also proposed the AR(1)

model defined in Prop. 5.9 with marginals having Laplace transform of the form $\frac{1}{1+g(s)}$. At last, we have estimated the parameters of the model defined in Prop. 5.9 using CLS and MOM and the simulation study for two cases implies that the estimates are satisfactory.

Chapter 6

Humbert generalized fractional differenced ARMA processes

In this chapter, we use the generating functions of Humbert polynomials to define two categories of Humbert generalized fractional differenced ARMA processes. We establish conditions for stationarity and invertibility for the introduced models. Additionally, we explore the singularities in the spectral densities of these models, with a specific focus on Pincherle ARMA, Horadam ARMA, and Horadam-Pethe ARMA processes. It is shown that the Pincherle ARMA process exhibits long memory properties when $u = 0$. Furthermore, we use the Whittle quasi-likelihood technique to estimate the parameters of the introduced processes, yielding results on the consistency and normality of the parameter estimators. To validate the efficacy of our estimation technique, particularly for the Pincherle ARMA process, we conduct a comprehensive simulation study. Moreover, we apply the Pincherle ARMA model to real data namely, Spain's 10-year treasury bond yield data, to show its utility in capturing market dynamics.

6.1 Introduction

The study of fractionally differenced time series by Granger and Joyeux (1980) [66] and Hosking in 1981 [79] provided an impetuous to a new research direction in time series modelling. The fractionally differenced time series called the autoregressive fractionally integrated moving average (ARFIMA) model generalizes the autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) models defined respectively by Yule (1926) [149], Slutsky (1937) [130] and Wold (1938) [145]. Also, the ARFIMA model is an extension of the autoregressive integrated moving average (ARIMA) process defined by Box and Jenkins (1976) [25] to model non-stationary time series by assuming the order of differencing $\nu \in \mathbb{R}$. The fractionally differenced time series is useful to model the data exhibiting long-range dependence (LRD). The data exhibiting LRD behaviours or long memory have a high correlation after a significant lag. For large sample inference for long-memory processes see Giraitis et al. [63]. Anh et al. [9] proposed some continuous

time stochastic processes with seasonal long-range dependence and these kinds of long memory processes have spectral pole at non-zero frequencies. In subsequent years, Andel (1986); Gray, Zhang and Woodward (1989, 1994) introduced the concept of Gegenbauer ARMA (GARMA) process. GARMA process also possesses seasonal long-range dependence [39]. The study on the usefulness of the Gegenbauer stochastic process is done by Dissanayake et al. [48]. The limit theorems for stationary Gaussian processes and their non-linear transformations with covariance function

$$\rho(h) \simeq \sum_{k=1}^{r} A_k \cos(h\omega_k) h^{-\alpha_k}, \sum_{k=1}^{r} A_k = 1,$$

where $A_k \geq 0, \alpha_k > 0, \omega_k \in [0, \pi), k = 1, \cdots, r$ have been considered in [81]. For seasonal long memory process X_t, the autocorrelation function for lag h denoted by $\rho(h)$ behaves asymptotically as $\rho(h) \simeq \cos(h\omega_0)h^{-\alpha}$ as $h \to \infty$ for some positive $\alpha \in (0, 1)$ and $\omega_0 \in (0, \pi)$ (see [36]). In literature, many tempered distributions and processes are studied using the exponential tempering in the original distribution or process see e.g. and references therein [12,65,97,123,125,137,151]. The fractionally integrated process with seasonal components are studied and maximum likelihood estimation is done by Reisen et al. [119]. The parametric spectral density with power-law behaviour about a fractional pole at the unknown frequency ω is analysed and Gaussian estimates and limiting distributional behavior of estimate is studied by Giraitis and Hidalgo [62]. The autoregressive tempered fractionally integrated moving average (ARTFIMA) process is obtained by using exponential tempering in the original ARFIMA process [125]. The ARTFIMA process is semi LRD and has a summable autocovariance function. In ARIMA process the fractional differencing operator $(1 - B)^{\nu}$, $|\nu| < 1$ is considered instead of $(1 - B)$, where B is the shift operator. In defining ARTFIMA model the tempered fractional differencing operator $(1-e^{-\lambda}B)^{\nu}$ is used where $\lambda > 0$ is the tempering parameter. The Gegenbauer process uses $(1-2uB+B^2)^{\nu}$, $|u| \leq 1$, $|\nu| < \frac{1}{2}$ as a difference operator, which can be written in terms of Gegenbauer polynomials.

In this chapter, we study Humbert polynomials based time series models. The Gegenbauer and Pincherle polynomials are the particular cases of Humbert polynomials. The Gegenbauer polynomials based time series model, namely GARMA process, is already studied and has been applied in several real world applications emanating from different areas. These processes possess seasonal long memory which helps to capture autocorrelation present in the data, leading to improved forecasting accuracy. We introduce and study two types of Humbert autoregressive fractionally integrated moving average (HARMA) models which are

defined by considering Humbert polynomials. We also obtain the spectral density, stationarity and invertibility conditions of the process. In particular, Pincherle ARMA, Horadam ARMA and Horadam-Pethe ARMA processes are studied. These new class of time series models generalizes the existing models like ARMA, ARIMA, ARFIMA, ARTFIMA and GARMA in several directions. The possible areas of applications of proposed model includes sales forecasting in e-commerce industries as it can capture seasonality, trends, and other patterns in historical sales data [135]. Also, the long memory property can capture autocorrelation patterns observed in financial returns, volatility and other indicators. These models can be applied to analyze environmental monitoring data, such as water quality parameters, air pollution levels, and ecosystem dynamics [30]. Further, we also provide the Whittle quasi-likelihood estimation for HARMA processes and applied on simulated data. Also, we applied the Pincherle ARMA model to Spain's 10-year treasury bond yield data.

The rest of the chapter is organized as follows. In Section 6.2, we introduce the Type 1 HARMA (p, ν, u, q) process, where p and q are autoregressive and moving average lags respectively and ν is a differencing parameter. This section includes the study of the stationarity property and spectral density of the introduced processes. Section 6.2 also includes the study of a particular case of type 1 HARMA(p, ν, u, q) process by taking $m = 3$, which is Pincherle ARMA (p, ν, u, q) process. Moreover, the spectral density of the Pincherle ARMA (p, ν, u, q) process is obtained and it is shown that for $u = 0$ the model exhibits seasonal long memory property. Section 6.3 deals with the type 2 HARMA process (p, ν, u, q). In this section, the particular cases namely, the Horadam ARMA process and the Horadam-Pethe ARMA process are discussed. In Section 6.4, we provide the Whittle quasi-likelihood method to estimate the parameters of type 1 and type 2 HARMA processes and it is shown that the estimators are consistent. The simulation study of Pincherle ARMA and its applications are discussed in Section 6.5. Overall, our study contributes to the advancement of time series analysis by introducing novel Humbert generalized fractional differenced ARMA models, investigating their properties, providing parameter estimation techniques, and showcasing their efficacy through simulations and real data applications.

6.2 Type 1 HARMA(p, ν, u, q) process

In this section, we introduce a new time series model namely type 1 HARMA(p, ν, u, q) process with the help of Humbert polynomials which we call hereafter type 1 Humbert polynomials. For Humbert polynomials and related

properties see e.g. [64,80,110]. A detailed discussion on special functions including Humbert polynomials is given in [64,131].

Definition 6.1 (Type 1 Humbert polynomials)**.** The Humbert polynomials of type 1 $\{\Pi^{\nu}_{n,m}\}_{n=0}^{\infty}$ are defined in terms of generating function as

$$(1 - mut + t^m)^{-\nu} = \sum_{n=0}^{\infty} \Pi^{\nu}_{n,m}(u)t^n, m \in \mathbb{N}, |t| < 1, |u| \leq 1 \text{ and } |\nu| < \frac{1}{2}. \tag{6.1}$$

For the table of special cases of Eq. (6.1), including Gegenbauer, Legendre, Tchebysheff, Pincherle, Kinney polynomials, see Gould (1965) [64]. In above definition, polynomial $\Pi^{\nu}_{n,m}(u)$ is explicitly can be written as follows [80]:

$$\Pi^{\nu}_{n,m}(u) = \sum_{k=0}^{\lfloor \frac{n}{m} \rfloor} \frac{(-mu)^{n-mk}}{\Gamma((1-\nu-n)+(m-1)k)(n-mk)!k!}, \text{ where } \left\lfloor \frac{n}{m} \right\rfloor \text{ is floor function.}$$

The hypergeometric representation of $\Pi^{\nu}_{n,m}(u)$ is given as follows:

$$\Pi^{\nu}_{n,m}(u) = \frac{(\nu)_n (mu)^n}{n!} {}_mF_{m-1}\left[\frac{-n}{m}, \frac{-n+1}{m}, \cdots, \frac{-n-1+m}{n}; \right.$$
$$\left. \frac{-\nu-n+1}{m-1}, \frac{-\nu-n+2}{m-2}, \cdots, \frac{-\nu-n+m-1}{m-1}; \frac{1}{(m-1)^{m-1}u^m}\right].$$

For more properties and results on hypergeometric functions see Srivastava and Manocha (1984) [131]. The type 1 Humbert polynomial satisfies the following recurrence relation

$$(n+1)\Pi^{\nu}_{n+1,m}(u) - mu(n+\nu)\Pi^{\nu}_{n,m}(u) - (n+m\nu-m+1)\Pi^{\nu}_{n-m+1,m}(u) = 0.$$

For $m = 2$ the Humbert polynomials reduces to Gegenbauer polynomials generally denoted as $\{C^{\nu}_n(u)\}_{n=0}^{\infty}$ and for $m = 3$ the polynomials reduce to Pincherle polynomials $\{P^{\nu}_n(u)\}_{n=0}^{\infty}$, see Pincherle (1891) [117]. The generating function of Pincherle polynomials have the following form

$$(1 - 3ut + t^3)^{-\nu} = \sum_{n=0}^{\infty} P^{\nu}_n(u)t^n,$$

where $P^{\nu}_n(u)$ has the following representation in terms of hypergeometric function [117]

$$P_n^{\nu}(u) = \frac{(\nu)_n (3x)^n}{n!} {}_3F_2\left[\frac{-n}{3}, \frac{-n+1}{3}, \frac{-n+2}{3}; \frac{-n-\nu+1}{2}, \frac{-n-\nu+2}{2}; \frac{-1}{4x^3}\right],$$

where ${}_3F_2(a_1, a_2, a_3; b_1, b_2; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k}{(b_1)_k (b_2)_k} \frac{x^k}{k!}$ and $(a_1)_k = \frac{\Gamma(a_1+k)}{\Gamma(a_1)}$ see e.g. [3].

Definition 6.2 (Type 1 HARMA process)**.** The type 1 HARMA(p, ν, u, q) process X_t is defined by

$$\Phi(B)(1 - muB + B^m)^{\nu} X_t = \Theta(B)\epsilon_t, \tag{6.2}$$

where ϵ_t is Gaussian white noise with variance σ^2, B is the lag operator, $0 \leq u < 2/m$, and $\Phi(B)$, $\Theta(B)$ are stationary AR and invertible MA operators respectively, defined as,

$$\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j, \Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j, \text{ and } B^j(X_t) = X_{t-j}.$$

In next result, the stationarity and invertibility conditions of the type 1 HARMA process are given. Also, the Abel's test which will be used in next theorem is stated below as proposition.

Proposition 6.1 (Abel's tests [18])**.** *If the series* $\sum_{n=0}^{\infty} a_n$ *is convergent and* $\{b_n\}$ *is monotone and bounded sequence then series* $\sum_{n=0}^{\infty} a_n b_n$ *is also convergent.*

Theorem 6.2. *Let* $\{X_t\}$ *be the type 1* HARMA(p, ν, u, q) *process defined in Eq. (6.2) and all roots of* $\Phi(B) = 0$ *and* $\Theta(B) = 0$ *lie outside the unit circle then the* HARMA(p, ν, u, q) *process is stationary and invertible for* $|\nu| < 1/2$ *and* $0 \leq u \leq 2/m$.

Proof. Using Eq. (6.2), one can write

$$X_t = \frac{\Theta(B)}{\Phi(B)}(1 - muB + B^m)^{-\nu}\epsilon_t, \text{ where } \frac{\Theta(B)}{\Phi(B)} = \sum_{j=0}^{\infty} \psi_j B^j.$$

Further,

$$(1 - muB + B^m)^{-\nu} = \sum_{n=0}^{\infty} \frac{(\nu)_n}{n!} (muB - B^m)^n.$$

Then Eq. (6.2) can be written as,

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty}\sum_{n=0}^{\infty}\psi_j\frac{(\nu)_n}{n!}(muB - B^m)^n\epsilon_{t-j-n} \\ &= \sum_{j=0}^{\infty}\sum_{n=0}^{\infty}\psi_j\frac{(\nu)_n}{n!}\sum_{r=0}^{n}(-1)^r\binom{n}{r}(mu)^{n-r}B^{mr}\epsilon_{t-n} \\ &= \sum_{j=0}^{\infty}\sum_{n=0}^{\infty}\psi_j\frac{(\nu)_n}{n!}(mu-1)^n\epsilon_{t-n-mj}. \end{aligned}$$

The variance of the process X_t is given by

$$\mathrm{Var}(X_t) = \sigma^2\sum_{j=0}^{\infty}\psi_j^2\sum_{n=0}^{\infty}\left(\frac{(\nu)_n}{n!}\right)^2(mu-1)^{2n} = \sigma^2\sum_{j=0}^{\infty}\psi_j^2\sum_{n=0}^{\infty}\left(\frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}\right)^2(mu-1)^{2n}.$$

Let $a_n = (mu-1)^{2n}$ and $\{b_n\} = \left(\frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}\right)^2$, then using Abel's test $\sum_{n=0}^{\infty}a_n$ converges for $0 < u < \frac{2}{m}$ and using Stirling's approximation, for large n, $b_n \simeq \frac{n^{2\nu-2}}{(\Gamma(\nu))^2}$, which implies that the sequence is bounded for $\nu < \frac{1}{2}$. We can write $b_n = \binom{\nu+n-1}{n}$ and it is known that $\binom{n}{x}$ is decreasing for $x \geq \lfloor\frac{n}{2}\rfloor$ this implies that $\{b_n\}$ is decreasing for $\nu \leq 1$. This indicates that the sequence is bounded and monotone for $\nu < 1/2$. Also, $\sum_{j=0}^{\infty}\psi_j^2$ is convergent, hence the $\mathrm{Var}(X_t)$ converges for the defined range. Similarly to prove the invertibility condition we define the process Eq. (6.2) as,

$$\epsilon_t = \pi(B)X_t,$$

where $\pi(B) = \frac{\Phi(B)}{\Theta(B)}(1 - muB + B^m)^\nu$ and again using the same argument discussed above the $\pi(z)$ will converge for $-\frac{1}{2} < \nu < 1$ and $0 < u < \frac{2}{m}$. For $u = 0$ and $u = \frac{2}{m}$ the variance can be defined as follows:

$$\begin{aligned} \mathrm{Var}(X_t) &= \sigma^2\sum_{n=0}^{\infty}\left(\frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}\right)^2 \\ &= \sigma^2\sum_{n=0}^{N}\left(\frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}\right)^2 + \sum_{n=N+1}^{\infty}\left(\frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}\right)^2. \end{aligned}$$

In the above equation, the first summation is finite and the terms inside the second summation behave like $\frac{n^{2\nu-2}}{\Gamma(\nu)^2}$ for large n and it is bounded for $\nu < \frac{1}{2}$. The $\mathrm{Var}(X_t)$ converges for $|\nu| < 1/2$ and $0 \leq u \leq \frac{2}{m}$. Hence, the HARMA process is stationary and invertible for $|\nu| < 1/2$ and $0 \leq u \leq \frac{2}{m}$. □

In the next result, we derive the spectral density of the type 1 HARMA(p, ν, u, q) process. The spectral density of process is defined in Def. 2.11.

Theorem 6.3. *For a type 1* HARMA(p, ν, u, q) *process defined in Eq. (6.2), under the assumptions of Theorem 6.2, the spectral density takes the following form*

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + m^2u^2 - 2mu(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu},$$

where $z = e^{-\iota\omega}$, $\omega \in (-\pi, \pi)$.

Proof. Rewrite Eq. (6.2) as follows:

$$X_t = \Psi(B)\epsilon_t,$$

where $\Psi(B) = \frac{\Theta(B)}{\Phi(B)}(1 - muB + B^m)^{-\nu}$. Then using the definition of spectral density of linear process, we have

$$f_x(\omega) = |\Psi(z)|^2 f_\epsilon(\omega), \tag{6.3}$$

where $z = e^{-\iota\omega}$ and $f_\epsilon(\omega)$ is spectral density of the innovation term. The spectral density of the innovation process ϵ_t is $\sigma^2/2\pi$. Then Eq. (6.3) becomes,

$$\begin{aligned} f_x(\omega) = \frac{\sigma^2}{2\pi}|\Psi(z)|^2 &= \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}\left|1 - mue^{-\iota\omega} + e^{-m\iota\omega}\right|^{-2\nu} \\ &= \frac{\sigma^2|\Theta(e^{-\iota\omega})|^2\left|1 - mue^{-\iota\omega} + e^{-m\iota\omega}\right|^{-2\nu}}{2\pi|\Phi(e^{-\iota\omega})|^2}. \end{aligned}$$

Here, $|1 - mue^{-\iota\omega} + e^{-m\iota\omega}|^{-2\nu} = (2 + m^2u^2 - 2mu(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu}$ and the spectral density takes the following form

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + m^2u^2 - 2mu(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu}. \tag{6.4}$$

□

Definition 6.3 (Singular point [59])**.** The point $\omega = \omega_0$ is said to be singular point of function f if at $\omega = \omega_0$, f fails to be analytic, that is $f(\omega_0) = \infty$.

Next, the definition of seasonal or cyclic long-memory is given in 2.9, which is characterized by having a spectral pole at a frequency $\kappa \in \mathbb{R}$ different from 0, see, e.g., [9,39].

Theorem 6.4. *Let* $\{X_t\}$ *be the stationary type 1* HARMA(p, ν, u, q) *process and all the assumptions of Theorem 6.2 hold then the spectral density of* HARMA(p, ν, u, q) $\{X_t\}$ *has singular spectrum*

(a) at $u = 0$ *and* $\omega = \frac{4n\pi \pm \pi}{m}$ *for* $-\frac{m\pm1}{4} < n < \frac{m\mp1}{4}$;

(b) at $u = \frac{2}{m}(-1)^n \cos(\frac{4n\pi}{m-2})$ *and* $\omega = \pm\frac{2n\pi}{m-2}$ *for* $m \neq 2$ *and* $-\frac{(m-2)}{4} < n < \frac{(m-2)}{4}$;

(c) at $\omega = \cos^{-1}(u)$ *for* $m = 2$.

Proof. From Eq. (6.4), the spectral density of the process $\{X_t\}$ is

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2+m^2u^2+2\cos(m\omega)-2mu(\cos(\omega)+\cos((m-1)\omega)))^{-\nu}, \text{ where } z = e^{-\iota\omega}.$$

We consider the denominator and find the zeros as follows,

$$\begin{aligned}
&2 + m^2u^2 + 2\cos(m\omega) - 2mu(\cos(\omega) + \cos((m-1)\omega)) \\
&= 2 + 2\cos(m\omega) + [mu - \{\cos(\omega) + \cos((m-1)\omega)\}]^2 - [\cos(\omega) + \cos((m-1)\omega)]^2 \\
&= 4\cos^2\left[\frac{m\omega}{2}\right] + [mu - \{\cos(\omega) + \cos((m-1)\omega)\}]^2 - 4\cos^2\left[\frac{m\omega}{2}\right]\cos^2\left[\frac{(m-2)\omega}{2}\right] \\
&= 4\cos^2\left[\frac{m\omega}{2}\right]\sin^2\left[\frac{(m-2)\omega}{2}\right] + \left[mu - 2\cos\left(\frac{m\omega}{2}\right)\cos\left(\frac{(2-m)\omega}{2}\right)\right]^2.
\end{aligned}$$

We have the following two cases.

(a) The first term, $4\cos^2\left[\frac{m\omega}{2}\right]\sin^2\left[\frac{(m-2)\omega}{2}\right] \geq 0$ for all m and $-\pi < \omega < \pi$.

$$\begin{aligned}
4\cos^2\left[\frac{m\omega}{2}\right]\sin^2\left[\frac{(m-2)\omega}{2}\right] &= 0 \\
\text{if } \cos^2\left(\frac{m\omega}{2}\right) &= 0 \text{ or } \sin^2\left(\frac{(m-2)\omega}{2}\right) = 0 \text{ or both} \\
\Rightarrow \cos\left(\frac{m\omega}{2}\right) &= 0 \text{ for } \omega_1 = (4n \pm 1)\frac{\pi}{m}, \text{ for all } m \in \mathbb{N} \text{ and} \\
n &= 0, \pm1, \pm2, \cdots.
\end{aligned}$$

We find the condition of singularity by solving the second term, $\left[mu - 2\cos\left(\frac{m\omega}{2}\right)\cos\left(\frac{(2-m)\omega}{2}\right)\right]^2$ at ω_1, which yields $u = 0$.
Also, the singular point $\omega_1 \in (-\pi, \pi)$ for $-\frac{m\pm1}{4} < n < \frac{m\mp1}{4}$.
Therefore, the type 1 HARMA(p, ν, u, q) process $\{X_t\}$ will have singular points for $u = 0$ and $\omega_1 = (4n \pm 1)\frac{\pi}{m}$ for all m and $-\frac{m\pm1}{4} < n < \frac{m\mp1}{4}$.
This proves the part (a).

(b) Again the term

$$4\cos^2\left(\frac{m\omega}{2}\right)\sin^2\left(\frac{(m-2)\omega}{2}\right) = 0$$

when

$$\sin^2\left(\frac{(m-2)\omega}{2}\right) = 0$$

$$\sin\left(\frac{(m-2)\omega}{2}\right) = 0 \text{ for } \omega_2 = \frac{\pm 2n\pi}{m-2} \text{ for all } m \in \mathbb{N}-\{2\}, \text{ and } n = 0, \pm 1, \pm 2, \cdots .$$

At ω_2, the second term will become zero if and only if,

$$\begin{aligned}\left[mu - \left(2\cos\left(\frac{m\omega_2}{2}\right)\cos\left(\frac{(2-m)\omega_2}{2}\right)\right)\right]^2 &= 0\\ \Rightarrow mu - 2(-1)^n \cos\left(\pm\frac{4n\pi}{m-2}\right) &= 0\\ \Rightarrow u = \frac{2(-1)^n}{m}\cos\left(\frac{4n\pi}{m-2}\right).&\end{aligned}$$

This proves the part (b).

(c) In Eq. (6.4) let $U(\omega) = (2 + m^2u^2 - 2mu(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))$. For different values of $m = 1, 2, 3, 4$ and $0 \leq u < 2/m$ in Fig. 6.1, we observe that the function $U(\omega)$ does not attain 0 for $\omega \in (-\pi, \pi)$. This signifies that the spectral density defined by Eq. (6.4) has no singularity for $m = 1, 3, 4$ and $0 \leq u < 2/m$. For $m = 2$, the spectral density is unbounded since $U(\omega)$ takes value 0 at $\omega = \cos^{-1}(u)$. Therefore, we conclude that for $m = 2$ the singularities are at $\omega = \cos^{-1}(u)$.

□

In Fig. 6.1, observe the behaviour of function $U(\omega)$ for $\omega \in (-\pi, \pi)$ and for different values of m and u. For $m = 2$, the function U touches the x-axis for all values of u. Further, for $m = 1$ it touches the x-axis only for $u = 0$. For other cases, see Theorem 6.4.

Definition 6.4 (Slowly varying function [67])**.** A function $b(\omega)$ is said to be slowly varying at ω_0 if for $\delta > 0$, $(\omega-\omega_0)^{\delta}b(\omega)$ is increasing and $(\omega-\omega_0)^{-\delta}b(\omega)$ is decreasing in some right-hand neighborhood of ω_0. Also, $(\omega-\omega_0)^{\delta}b(\omega)$ is decreasing and $(\omega-\omega_0)^{-\delta}b(\omega)$ is increasing in some left-hand neighbourhood of ω_0.

We need the following lemma which is given in [67] to prove our next result.

Lemma 6.5 (Gray et al.[67])**.** *Let $R(\tau) = \int_0^{\pi} P(\omega)\cos(\tau\omega)d\omega$ where τ is an integer and $P(\omega)$ is spectral density. Suppose $P(\omega)$ can be expressed as*

$$P(\omega) = b(\omega)|\omega - \omega_0|^{-\beta} \tag{6.5}$$

with $0 < \beta < \frac{1}{2}$ and $\omega_0 \in (0, \pi)$. Further, suppose that $b(w)$ is non-negative and of bounded variation in $(0, \omega_0 - \epsilon) \cup (\omega_0 + \epsilon, \pi)$ for $\epsilon > 0$. Also suppose that $b(\omega)$ is slowly varying at ω_0, then as $\tau \to \infty$

$$R(\tau) \simeq \tau^{2\beta-1}cos(\tau\omega_0).$$

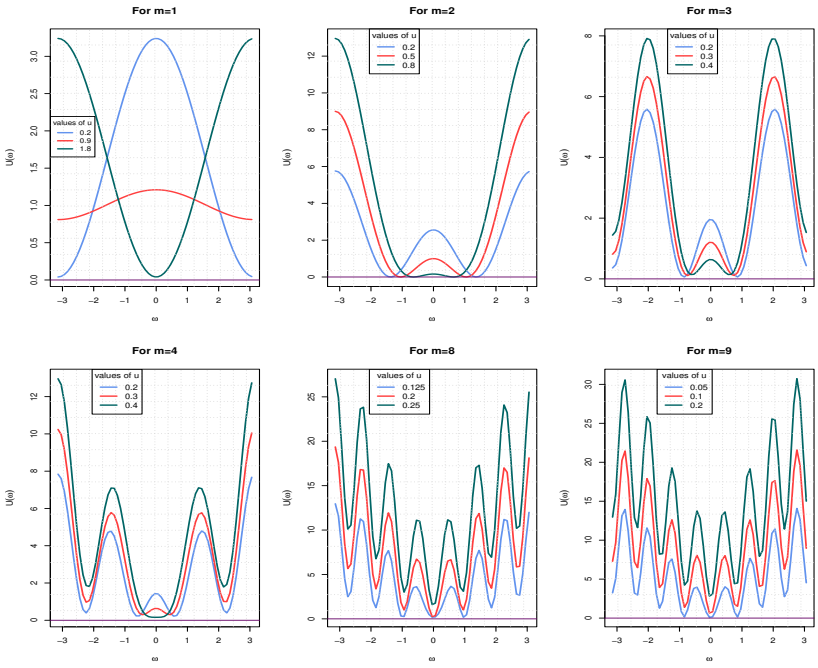


Figure 6.1: Plot of the function $U(\omega)$ for different values of $m \in \{1, 2, 3, 4, 8, 9\}$ and $0 \leq u \leq 2/m$.

Theorem 6.6. *The stationary type* 1 *HARMA*$(p, \nu, 0, q)$ *process has seasonal long memory for* $0 < \nu < 1/2$.

Proof. The spectral density of type 1 HARMA$(0, \nu, u, 0)$ process is given by

$$f_x(\omega) = \frac{\sigma^2}{2\pi}(2 + m^2u^2 - 2mu(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu}.$$

For $u = 0$ the spectral density has the form

$$f_x(\omega) = \frac{\sigma^2}{2\pi}(2 + 2\cos(m\omega)). \tag{6.6}$$

Also, the spectral density is unbounded at $\omega_0 = (4n \pm 1)\frac{\pi}{m}, -\frac{m\pm1}{4} < n < \frac{m\mp1}{4}$. which implies that the covariance is not absolutely summable for $u = 0$ at frequency ω_0. To prove the process is seasonal long memory we use lemma 6.5 defined by Gray et al. [67]. Now Eq. (6.6) can be rewritten as

$$f_x(\omega) = \frac{\sigma^2(2 + 2\cos(m\omega))^{-\nu}|\omega - \omega_0|^{-2\nu}}{2\pi|\omega - \omega_0|^{-2\nu}}.$$

Comparing the above equation with Eq. (6.5)

$$b(\omega) = \frac{\sigma^2(2\cos(m\omega) + 2)^{-\nu}}{2\pi|\omega - \omega_0|^{-2\nu}}.$$

Now to show $b(\omega)$ is slowly varying at ω_0, consider the case $\omega > \omega_0$ and for $\delta > 0$ define,

$$l(\omega) = b(\omega)(\omega - \omega_0)^{\delta} = \frac{\sigma^2}{2\pi}(2 + 2\cos(m\omega))^{-\nu}(\omega - \omega_0)^{\delta+2\nu}$$

and

$$l'(\omega) = \frac{\sigma^2}{2\pi}(\omega - \omega_0)^{\delta+2\nu-1}(2 + 2\cos(m\omega))^{-\nu-1}((\delta + 2\nu)(2 + 2\cos(m\omega) + 2\nu m \sin(m\omega)(\omega - \omega_0)).$$

For $\omega > \omega_0$ the terms $(\omega - \omega_0)^{\delta+2\nu-1}$, $(\delta + 2\nu)$ and $(2 + 2\cos(mw))$ are positive. It can be easily shown that

$$\lim_{\omega \to \omega_0} (2\nu m \sin(m\omega)(\omega - \omega_0) + (\delta + 2\nu)(2 + 2\cos(m\omega))) > 0.$$

Thus in some right hand neighbourhood of ω_0, i.e. for $\omega \to \omega_0^+$, $l'(\omega) > 0$ and $(\omega - \omega_0)^{\delta} b(\omega)$ is increasing and similarly $(\omega - \omega_0)^{-\delta} b(\omega)$ is decreasing when $\omega \to \omega_0^+$. Similarly, it can be easily shown that for $\omega < \omega_0$, $(\omega - \omega_0)^{\delta} b(\omega)$ is decreasing and $(\omega - \omega_0)^{-\delta} b(\omega)$ is increasing in some left hand neighbourhood of ω_0. Thus the function is slowly varying at ω_0.
Also, it can be easily verified that the function $b(w)$ has bounded derivative in $(0, \omega_0 - \epsilon) \cup (\omega_0 + \epsilon, \pi)$, hence it is of bounded variation in $(0, \omega_0 - \epsilon) \cup (\omega_0 + \epsilon, \pi)$. Using the above two results and the lemma 6.5 the autocorrelation function $R(h)$ of the type 1 Humbert ARMA process takes the following asymptotic form

$$R(h) \simeq h^{2\nu-1} \cos(h\omega_0), \text{ as } h \to \infty. \tag{6.7}$$

The result in Eq. (6.7) implies that the process is seasonal long memory for $0 < \nu < 1/2$.

□

6.2.1 Pincherle ARMA (p, ν, u, q) process

This section deals with the special case of the type 1 HARMA process for $m = 3$. The Pincherle polynomials are polynomials introduced by Pincherle (1891) [117]. The Pincherle polynomials were generalized to Humbert polynomials by Humbert (1920) [80].

Definition 6.5 (Pincherle polynomials)**.** The Pincherle polynomials $P_n^{\nu}(u)$ are defined as the coefficient of t in the expansion of $(1 - 3ut + t^n)^{-\nu}$. The Pincherle

polynomials are defined by taking $m = 3$ in type 1 Humbert polynomials that is $P_n^\nu(u) = \Pi_{n,3}^\nu(u)$. Also, the generating function relation for Pincherle polynomials is given by

$$(1 - 3ut + t^n)^{-\nu} = \sum_{n=0}^{\infty} P_n(u)t^n, \ |t| < 1, |\nu| < 1/2, \ |u| \leq 1.$$

The polynomials satisfy the following difference equation [13]

$$(n+1)P_{n+1}^\nu(u) - 3u(n+\nu)P_n^\nu(u)u + (n+3\nu-2)P_{n-2}^\nu(u) = 0.$$

The coefficient of Pincherle polynomials can be written as $P_0^\nu(u) = 1$, $P_1^\nu(u) = 3\nu, P_2^\nu(u) = 9\nu(\nu+1)u^2/2$ and the nth coefficient takes the form [13]

$$\frac{\Gamma(n+\nu)\Gamma(1/3)\Gamma(2/3)}{\Gamma(\nu)\Gamma((n+1)/3)\Gamma((n+2)/3)\Gamma((n+3)/3)}.$$

Definition 6.6 (Pincherle ARMA process)**.** The Pincherle ARMA (p, ν, u, q) process is defined by taking $m = 3$ in type 1 HARMA process defined in Eq. (6.2) and the process has the form defined below

$$\Phi(B)(1 - 3uB + B^3)^\nu X_t = \Theta(B)\epsilon_t, \tag{6.8}$$

where ϵ_t is Gaussian white noise with variance σ^2, $0 \leq u < 2/3$ and B, $\Phi(B)$ and $\Theta(B)$ are lag, stationary AR and invertible MA operators, respectively defined in Def. 6.2.

Theorem 6.7. *Let* $\{X_t\}$ *be the Pincherle* ARMA(p, ν, u, q) *process defined by Eq.* (6.8) *and all roots of* $\Phi(B) = 0$ *and* $\Theta(B) = 0$ *lie outside the unit circle then the Pincherle* ARMA(p, ν, u, q) *process is stationary and invertible for* $|\nu| < 1/2$ *and* $0 \leq u \leq 2/3$.

Proof. The proof can be easily done by taking $m = 3$ in the proof of Theorem 6.2. □

Theorem 6.8. *The stationary Pincherle HARMA*$(p, \nu, 0, q)$ *process has seasonal long memory for* $0 < \nu < 1/2$ *at* $\omega_0 = \pi/3$.

Proof. According to Theorem 6.4 the spectral density of the Pincherle ARMA process has singularity at $u = 0$ for $\omega_0 = \pi/3$. Also, similar to the proof of Theorem 6.6 the autocovariance function of Pincherle ARMA process $\gamma(h)$ has the asymptotic form $R(h) \simeq h^{2\nu-1}\cos(h\omega_0)$. This proves that the process has a seasonal long memory for $0 < \nu < 1/2$ at $\omega_0 = \pi/3$. □

Theorem 6.9. *For a Pincherle ARMA(p, ν, u, q) process defined by Eq. (6.8), the spectral density takes the following form*

$$f_x(\omega) = \frac{\sigma^2}{2\pi} \frac{|\Theta(z)|^2}{|\Phi(z)|^2} (8\cos^3(\omega) - 12u\cos^2(\omega) - C\cos(\omega) + D)^{-\nu},$$

where $z = e^{-\iota\omega}$, $C = 6 + 6u$, *and* $D = 2 + 6u + 9u^2$.

Proof. Taking $m = 3$ in Eq. (6.4) gives us the desired spectral density.

□

Theorem 6.10. *The autocovariance function for the Pincherle ARMA process takes the following form*

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \psi_j \psi_{j+h} P_n^{\nu}(u) P_{n+h}^{\nu}(u).$$

Proof. For lag h the autocovariance of the process $\{X_t\}$ and $\{X_{t+h}\}$ using the Eq. (6.8) is given by

$$\mathrm{Cov}(X_t X_{t+h}) = \mathrm{E}[X_t X_{t+h}],$$

where X_t can be written as

$$X_t = \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \psi_j P_n^{\nu}(u) \epsilon_{t-j-n}$$

and

$$\mathrm{E}[X_t X_{t+h}] = \sigma^2 \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \psi_j \psi_{j+h} P_n^{\nu}(u) P_{n+h}^{\nu}(u).$$

□

6.3 Type 2 HARMA(p, ν, u, q) process

Milovanovic and Dordevic in 1987 [110] considered the following generalization of Gegenbaur polynomials, which we call type 2 Humbert polynomials and are used to define the type 2 HARMA process.

Definition 6.7 (Type 2 Humbert polynomials)**.** The type 2 Humbert polynomials are defined by considering the polynomials $Q_{n,m}^{\nu}(u)$ defined by the following

generating function

$$(1-2ut+t^m)^{-\nu}=\sum_{n=0}^{\infty}Q_{n,m}^{\nu}(u)t^n,\ |t|<1,\ |\nu|<1/2,\ |u|\leq 1. \tag{6.1}$$

Here $Q_{n,m}^{\nu}(u)=\Pi_{n,m}^{\nu}(\frac{2u}{m})$ (see *Eq.*(6.1)).

The explicit form of the polynomials $Q_{n,m}^{\nu}(u)$ is defined by

$$Q_{n,m}^{\nu}(u)=\sum_{k=0}^{[\frac{n}{m}]}(-1)^k\frac{(\nu)_{(n-(m-1)k)}}{k!(n-mk)!}(2u)^{n-mk},$$

where $\nu_0=1$ and $(\nu)_n=\nu(\nu+1)\cdots(\nu+n-1)$.

Definition 6.8 (Type 2 HARMA process)**.** The type 2 HARMA process is defined by using the above-defined generation function as follows

$$\Phi(B)(1-2uB+B^m)^{\nu}X_t=\Theta(B)\epsilon_t, \tag{6.2}$$

where ϵ_t is Gaussian white noise with variance σ^2, $0\leq u<1$, and B, $\Phi(B)$, $\Theta(B)$ are lag, stationary AR and invertible MA operators, respectively defined in Def. 6.2.

For $m=2$ the above polynomials in Eq. (6.1) is Gegenbauer polynomials and $Q_{n,2}^{\nu}(u)=C_n^{\nu}(u)$. Also, for $m=3$ the polynomials in Eq. (6.1) are known as Horadam-Pethe polynomials and for $m=1$ they are known as Horadam polynomials, see Gould (1965) [64], Horadam (1985) [77] and Horadam and Pethe (1981) [78].

Theorem 6.11. *Let* $\{X_t\}$ *be the type 2* HARMA(p,ν,u,q) *process and all roots of* $\Phi(B)=0$ *and* $\Theta(B)=0$ *lies outside the unit circle then the* HARMA(p,ν,u,q) *process is stationary and invertible for* $|\nu|<1/2$ *and* $0\leq u\leq 1$.

Proof. The process is stationary and invertible for $|\nu|<1/2$ and $0\leq u\leq 1$ can be easily proved using the proof for the stationarity of type 1 HARMA process defined in Theorem 6.2. □

Theorem 6.12. *For a type 2 Humbert ARMA*(p,ν,u,q) *process defined in Eq.* (6.2)*, under the assumptions of Theorem 6.11 the spectral density takes the following form*

$$f_x(\omega)=\frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2+4u^2-4u(\cos(\omega)+\cos((1-m)\omega))+2\cos(m\omega))^{-\nu}, \tag{6.3}$$

where $z=e^{-\iota\omega}$.

Proof. Rewrite Eq. (6.2) as follows

$$X_t = \Psi(B)\epsilon_t,$$

where $\Psi(B) = \frac{\Theta(B)}{\Phi(B)}\Delta^\nu$ and $\Delta^\nu = (1 - 2uz + z^m)^{-\nu}$. The spectral density of the innovation process ϵ_t is given by $\sigma^2/2\pi$, which implies

$$f_x(\omega) = \frac{\sigma^2}{2\pi}|\Psi(z)|^2 = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}\left|1 - 2uz + z^m\right|^{-2\nu},$$

where $z = e^{-\iota\omega}$. Furthermore,

$$\left|1 - 2ue^{-\iota\omega} + e^{-m\iota\omega}\right|^{-2\nu} = (2 + 4u^2 - 4u(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu},$$

and the spectral density takes the following form

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu}.$$

□

Theorem 6.13. *Under the assumption of Theorem 6.11 let* $\{X_t\}$ *be the type 2* HARMA(p, ν, u, q) *process then the spectral density of* HARMA(p, ν, u, q) *process has singularities*

(a) at $u = 0$ *and* $\omega = \frac{4n\pi \pm \pi}{m}$ *for* $-\frac{m\pm1}{4} < n < \frac{m\mp1}{4}$.

(b) at $u = (-1)^n \cos(\frac{4n\pi}{m-2})$ *and* $\omega = \pm\frac{2n\pi}{m-2}$ *for* $m \neq 2$ *and* $-\frac{(m-2)}{4} < n < \frac{(m-2)}{4}$.

Proof. The spectral density of the type 2 HARMA process is

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega))^{-\nu}.$$

Similar to the proof in Theorem 6.4, we find the zeros by writing the denominator as follows

$$\begin{aligned} &2 + 4u^2 - 4u(\cos(\omega) + \cos((1-m)\omega)) + 2\cos(m\omega) = \\ &4\cos^2\left[\frac{m\omega}{2}\right]\sin^2\left[\frac{(m-2)\omega}{2}\right] + \left[2u - 2\cos\left(\frac{m\omega}{2}\right)\cos\left(\frac{(2-m)\omega}{2}\right)\right]^2 \end{aligned} \tag{6.4}$$

The proof of part *(a)* is the same as the part *(a)* of Theorem 6.4. To prove the part *(b)* the term $4\cos^2\left[\frac{m\omega}{2}\right]\sin^2\left[\frac{(m-2)\omega}{2}\right] = 0$ at $\omega_0 = \frac{\pm 2n\pi}{m-2}$. For this ω_0, the

second term of Eq. (6.4) is zero for $u = (-1)^n \cos(\frac{4n\pi}{m-2})$ for all $m \in \mathbb{N}-\{2\}$ and $-\frac{(m-2)}{4} < n < \frac{(m-2)}{4}$.

□

The particular cases of the type 2 Horadam ARMA process is discussed as follows:

6.3.1 Horadam ARMA(p, ν, u, q) process

Definition 6.9 (Horadam polynomials)**.** In Eq. (6.1) by taking $m = 1$ the reduced polynomials are known as Horadam polynomials. The Horadam polynomials are defined as the coefficient of t in the expansion of $(1 - 2ut + t)$ and the generating function relation is given as follows:

$$(1 - 2ut + t)^{-\nu} = \sum_{n=0}^{\infty} Q_{n,1}^{\nu}(u)t^n, \ |t| < 1, \ |\nu| < 1/2, \ |u| \leq 1.$$

Definition 6.10 (The Horadam ARMA process)**.** The time series process defined using the generating function of Horadam polynomials are defined by the Horadam ARMA process, which is a special case of type2 HARMA process for m=1 and the process takes the following form:

$$\Phi(B)(1 - 2uB + B)^{\nu} X_t = \Theta(B)\epsilon_t, \tag{6.5}$$

where ϵ_t is Gaussian white noise with variance σ^2, $0 \leq u \leq 1$, and B, $\Phi(B)$, $\Theta(B)$ are lag, stationary AR and invertible MA operators, respectively defined in Def. 6.2.

Theorem 6.14. *For a Horadam ARMA(p, ν, u, q) process defined in Eq. (6.5), the spectral density takes the following form*

$$f_x(\omega) = \frac{\sigma^2}{2\pi} \frac{|\Theta(z)|^2}{|\Phi(z)|^2} (2 + 4u^2 - 4u - 4u\cos(\omega) + 2\cos(\omega))^{-\nu}, \ z = e^{-\iota\omega}. \tag{6.6}$$

Proof. This can be easily proved by taking $m = 1$ in the spectral density of type 2 HARMA process defined in Eq. (6.3). □

6.3.2 Horadam-Pethe ARMA(p, ν, u, q) process

Taking $m = 3$ in Eq. (6.1) the reduced form of the polynomials is known as Horadam-Pethe polynomials and the corresponding time series defined using the generating function of Horadam-Pethe polynomials is known as Horadam-Pethe ARMA process defined as follows

$$\Phi(B)(1 - 2uB + B^3)^{\nu} X_t = \Theta(B)\epsilon_t, \tag{6.7}$$

where $(1-2uB+B^3)^{-\nu}=\sum_{n=0}^{\infty}Q_{n,3}^{\nu}(u)t^n$.

Theorem 6.15. *Under the assumptions of Theorem 6.11 for a Horadam-Pethe ARMA(p,ν,u,q) process defined in Eq. (6.7), the spectral density takes the following form*

$$f_x(\omega)=\frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2+4u^2-4u(\cos(\omega)+\cos(2\omega))+2\cos(3\omega))^{-\nu},$$

where $z=e^{-\iota\omega}$.

Remark 6.1. Taking $m=2$ the polynomials in Eq. (6.1) reduced to Gegenbauer polynomials and $Q_{n,2}^{\nu}(u) = C_n^{\nu}(u)$. Moreover, the corresponding time series using the generating function of Gegenbauer polynomials namely the Gegenbauer Autoregressive Moving Average (GARMA) process is studied by Gray and Zhand in 1989 (see [67]).

Remark 6.2. The stationarity and invertibility condition for Horadam ARMA and Horadam-Pethe ARMA process is the same as the type 2 HARMA process, which is the process is stationary and invertible if all roots of $\Phi(B)=0$ and $\Theta(B)=0$ lies outside the unit circle and $|\nu|<1/2$ and $0\leq u<1$.

The time-series plots for simulated Pincherle, Horadam, Horadam-Pethe and Gegenbauer ARMA processes are given in Fig. 6.2. We simulated time-series of size 1000 from each process. All these series have in theory infinite differencing terms. We consider only finite terms by truncating the binomial expansions of the different shift operators. For Pincherle ARMA process, the relation defined in Eq. (6.8) is used, that is

$$X_t=\frac{\Theta(B)}{\Phi(B)}(1-3uB+B^3)^{-\nu}\epsilon_t. \tag{6.8}$$

The series $Z_t=(1-3uB+B^3)^{-\nu}\epsilon_t$ is generated using the simulated innovation series $\epsilon_t\sim\mathcal{N}(0,\sigma^2)$. Further, we approximate Z_t by considering first 4 terms in the binomial expansion of $(1-3uB+B^3)^{-\nu}$, which is

$$\begin{aligned}Z_t=(1-3uB+B^3)^{-\nu}\epsilon_t&=\sum_{n=0}^{\infty}\sum_{j=0}^{n}(-1)^j\frac{(\nu)_n}{n!}\binom{n}{j}(3u)^{n-j}B^{2j+n}\epsilon_t\\&\approx\sum_{n=0}^{4}\sum_{j=0}^{n}(-1)^j\frac{(\nu)_n}{n!}\binom{n}{j}(3u)^{n-j}\epsilon_{t-n-2j}.\end{aligned}$$

Now by generating the series Z_t the Eq. (6.8) takes the following form

$$X_t=\frac{\Theta(B)}{\Phi(B)}Z_t,$$

which is nothing but the ARMA process which is simulated using the "nloptr" library in R by passing the Z_t as innovation series. Using the same approach, we simulate the Horadam, Gegenbauer and Horadam Pethe ARMA processes by taking the binomial expansion of $(1 - 2uB + B^m)^{-\nu}$, for $m = 1, 2$ and 3, respectively.

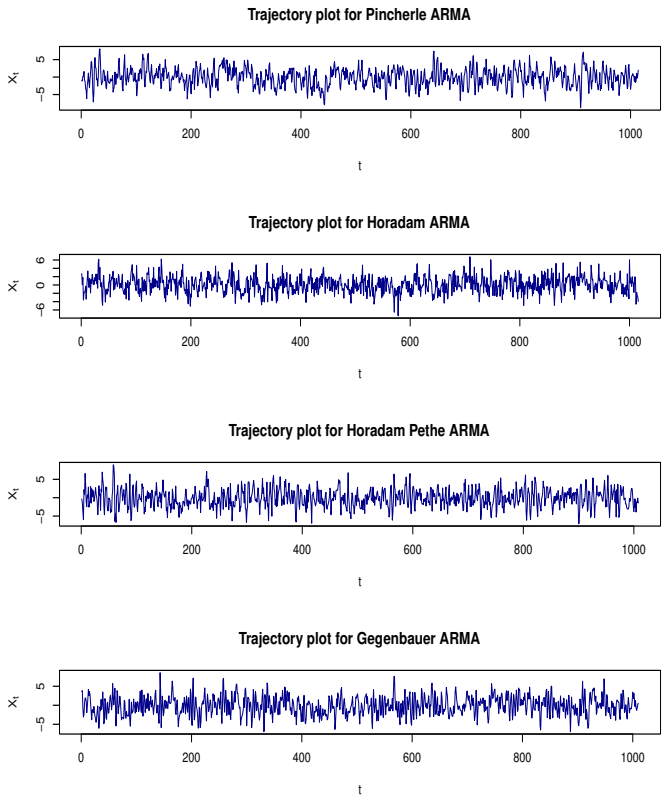


Figure 6.2: Trajectory plots for Pincherle, Horadam, Horadam-Pethe and Gegenbauer ARMA processes for $p = 1, q = 0, \nu = 0.3$ and $u = 0.1$.

6.4 Parameter estimation

In this section, we introduce the Whittle quasi-likelihood estimation method for the type 1 and type 2 Humbert ARMA Processes. The Whittle quasi-likelihood technique leverages the empirical spectral density and theoretical spectral density to estimate the model parameters of the time series X_t, where $t \in \{0, 1, \ldots, n\}$ and n denotes the sample size. The estimation process involves minimizing the likelihood function. Consider the set of harmonic frequencies $\omega_j, j = 0, 1, \ldots, n/2$. These frequencies are selected to define the empirical spectral density, which plays a crucial role in the Whittle quasi-likelihood estimation. The empirical spectral density provides a representation of the distribution of frequencies in the time series data. The estimation process starts by calculating the empirical spectral density. This involves computing the periodogram, which is a commonly used estimator of

the spectral density expressed as follows

$$I_x(\omega_j) = \frac{1}{2\pi}\{R(0) + \sum_{s=1}^{n-1} R(s)\cos(s\omega_j)\},\ \omega_j = \frac{2\pi j}{n},\ j = 0, 1, \ldots, n/2, \tag{6.9}$$

where $R(s) = \frac{1}{n}\sum_{i=1}^{n-s}(X_i - \bar{X})(X_{i+s} - \bar{X})$, $s = 0, 1, \ldots, (n-1)$, is the sample autocovariance function with sample mean $\bar{X}$. The Whittle quasi-likelihood estimation method aims to find the model parameters that minimize the discrepancy between the empirical and theoretical spectral densities that is $I_x(\omega_j)$ and $f_x(\omega_j)$ respectively [143]. This is achieved by optimizing the likelihood function

$$W_n(\theta) = \sum_{j=1}^{n}\left(\frac{I_x(\omega_j)}{f_x(\omega_j)} + \log(f_x(\omega_j))\right),$$

where θ represents unknown parameters $\theta = (\nu, u)$, which is a row vector for both type 1 and type 2 HARMA processes. We estimate the parameters by minimizing the likelihood function $W_n(\theta)$ with respect to unknown parameter θ.

Pincherle ARMA Process: Let us assume $S = \{\nu, u : |\nu| < 1/2, 0 \leq u \leq 2/3\}$ and $S_0 \subset S$ is a compact set. From Theorem 6.9, the spectral density of Pincherle ARMA process is

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(8\cos^3(\omega) - 12u\cos^2(\omega) - C\cos(\omega) + D)^{-\nu},$$

where $z = e^{-\iota\omega}$, $C = 6 + 6u$, and $D = 2 + 6u + 9u^2$ and empirical spectral density can be calculated using Eq. (6.9). The estimate of θ is given by

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\ W_n(\theta),\ \theta \in S_0.$$

Horadam and Horadam-Pethe ARMA Process: To estimate the parameters of Horadam and Horadam-Pethe ARMA processes we assume $S' = \{\nu,\ u : |\nu| < 1/2,\ 0 \leq u \leq 1\}$ and $S_0' \subset S'$ is a compact set. From Theorem 6.12, the spectral density for the Horadam process takes the following form:

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u - 4u\cos(\omega) + 2\cos(\omega))^{-\nu},\ z = e^{-\iota\omega}.$$

From Theorem 6.15 the spectral density for the Horadam-Pethe ARMA process has the following form:

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u(\cos(\omega) + \cos(2\omega)) + 2\cos(3\omega))^{-\nu}.$$

The estimate of θ is obtained by minimizing the likelihood

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\, W_n(\theta),\ \theta \in S_0'.$$

To prove the consistency and asymptotic normality of the Whittle quasi-likelihood estimators for the defined processes, we use the following results discussed by Hannan [71]:

Conditions A:

1. The time series $\{X_t\}$ has the following moving average representation:

$$X_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k},, \text{ where } \sum_{k=0}^{\infty} a_k^2 < \infty \text{ and } a_0 = 1. \tag{6.10}$$

2. The spectral density $(f_x(\omega))$ of the process $\{X_t\}$ can be written as:

$$f_x(\omega) = \frac{\sigma^2}{2\pi} K(\omega)$$

   and $\frac{1}{K(\omega)+a}$ is a continuous function, for $\omega \in (-\pi, \pi)$, $\forall\ a > 0$.

3. Assuming the parameter space (Ω_0) is compact. Then the parameter vector $\theta \in \Omega_0$ defines the spectral density uniquely.

Conditions B:

1. $K(\omega) > 0$, for all $\omega \in (-\pi, \pi)$ and $\theta \in \Omega_0$.

2. $K(\omega)$ twice differentiable of all the parameters in parameter vector θ.

3. The condition in (6.10) holds.

To prove the consistency of the Whittle quasi-likelihood using *Conditions A*, Fox and Taqqu gave a theorem (see Theorem 8.2.1 [56]), which is as follows:

Theorem 6.16. *Suppose an observable moving-average process* $\{X_t, t \in \mathbb{Z}\}$, *of* (6.10) *is ergodic and has the spectral density* $f_x(w) = \frac{\sigma^2}{2\pi} K(\omega)$, *and suppose the functions* $K(\omega)$, $\theta \in \Omega_0$ *satisfy Conditions A. Further, if additionally,* $1/f_x(\omega)$ *is continuous on* $(-\pi, \pi)$, *then,*

$$\hat{\theta}_n \xrightarrow{a.s.} \theta, \quad \text{as } n \to \infty.$$

The asymptotic normality of the Whittle quasi-likelihood can be proved using Theorem 2 defined by Hannan [71], which is as follows:

Theorem 6.17. *Under Conditions B, $\sqrt{n}(\hat{\theta}-\theta)$ converges in distribution to Gaussian random vectors with zero mean and covariance matrix W^{-1}, where*

$$W=\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial \log K(\omega)}{\partial\theta}\right\}\left\{\frac{\partial \log K(\omega)}{\partial\theta}\right\}' d\omega.$$

We use the above theorems to prove the properties of the Whittle likelihood estimators.

Theorem 6.18. *Assume the conditions of Theorem 6.7 holds, then the Whittle quasi-likelihood estimators for the Pincherle ARMA process are consistent. That is, $\lim_{n\to\infty}\hat{\theta}_n=\theta \quad a.s.$*

Proof. To prove the consistency of the Whittle quasi-likelihood method we use the result defined by Hannan (see theorem 1 in [71]). Assume that the parameters lie in the compact space S_0. We can rewrite the spectral density of the Pincherle ARMA process defined in Theorem 6.9 as, $f_x(\omega)=\frac{\sigma^2}{2\pi}K(\omega)$ and $K(\omega)$ is given as follows:

$$K(\omega)=\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(8\cos^3(\omega)-12u\cos^2(\omega)-C\cos(\omega)+D)^{-\nu},$$

where $z=e^{-\iota\omega}$, $C=6+6u$, and $D=2+6u+9u^2$. First, we prove that the time series defined in Eq. (6.8) can be represented as $X_t=\sum_{k=0}^{\infty}a_k\epsilon_{n-k}$, where $\sum_{k=1}^{\infty}a_k^2<\infty$ and $a_0=1$. The Eq. (6.8) can be stated as

$$\begin{aligned}X_t&=\left(\sum_{j=0}^{\infty}\psi_jB^j\right)\left(\sum_{n=0}^{\infty}\sum_{r=0}^{\infty}(-1)^r\frac{\Gamma(\nu+n)}{\Gamma(n+1)\Gamma n}\binom{n}{r}(3v)^{n-r}B^{n+2r}\right)\epsilon_t\\&=\left(\sum_{j=0}^{\infty}\psi_jB^j\right)\left(\sum_{i=0}^{\infty}\rho_iB^i\right)\epsilon_t,\end{aligned}$$

where

$$\rho_i=\frac{\Gamma(\nu+i)(3u)^i}{\Gamma(\nu)\Gamma(i+1)}-\frac{\Gamma(\nu+i-2)(3u)^{i-3}}{\Gamma(\nu)\Gamma(2)\Gamma(i-2))}. \tag{6.11}$$

In the following way, X_t can be reformulated as

$$X_t=\sum_{k=0}^{\infty}a_kB^k\epsilon_t=\sum_{k=0}^{\infty}a_k\epsilon_{t-k},$$

where $a_k=\sum_{s=0}^{k}\psi_{k-s}\rho_s$ and $a_0=1$. To prove $\sum_{k=0}^{\infty}a_k^2<\infty$ we can show $\sum_{k=0}^{\infty}|a_k|<\infty$. The operators $\Theta(B)$ and $\Phi(B)$ can be characterized as

stationary autoregressive and invertible moving average operators, respectively. Their corresponding polynomial representations $\frac{\Theta(z)}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$, where the series $\sum_{j=0}^{\infty} |\psi_j| < \infty$ for $|z| \leq 1+\epsilon$. Consequently, we can deduce that the absolute values of the coefficients ψ_j decrease polynomially with increasing j, bounded by the inequality $|\psi_j| < C(1+\epsilon)^{-j}$, where C represents a constant. Also, in Eq. (6.11) for large i using Stirling's approximation ρ_i can be approximated as follows

$$\rho_i \sim i^{\nu-1}(3u)^i - (i-2)^{\nu-1}(3u)^{i-3},$$

which clearly indicates that the $\sum_{i=0}^{\infty} |\rho_i| < \infty$ for $|\nu| < 1/2$ and $0 \leq u \leq 2/3$. Further,

$$\sum_{k=0}^{\infty} |a_k| \leq \sum_{k=0}^{\infty} \sum_{s=0}^{k} |\psi_{k-s}||\rho_s| = \sum_{s=0}^{\infty} \sum_{k=s}^{\infty} |\psi_{k-s}||\rho_s| = \sum_{s=0}^{\infty} \sum_{r=0}^{\infty} |\psi_r||\rho_s|. \tag{6.12}$$

Since, we have proved that $\sum_{i=0}^{\infty} |\rho_i|$ is finite which implies $\sum_{k=0}^{\infty} |a_k| < \infty$.

Next we show $\frac{1}{a+K(\omega)}$ is continuous for $\omega \in (-\pi, \pi)$ and for all $a > 0$. Observe that

$$\frac{1}{a+K(\omega)} = \frac{\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(8\cos^3(\omega) - 12u\cos^2(\omega) - C\cos(\omega) + D)^{\nu}}{a\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(8\cos^3(\omega) - 12u\cos^2(\omega) - C\cos(\omega) + D)^{\nu} + 1}.$$

It is easy to see that $(8\cos^3(\omega) - 12u\cos^2(\omega) - C\cos(\omega) + D)^{\nu}$ is continuous for $0 < \nu < 1/2$. Hence, $\frac{1}{a+K(\omega)}$ is continuous $\omega \in (-\pi, \pi)$, $|\nu| < 1/2$ and $0 \leq u \leq 2/3$ for all $a > 0$. Also, for $\theta \in S_0$ and $\omega \in (-\pi, \pi)$, the spectral density $f_x(\omega)$ is uniquely defined. Therefore, we conclude the Whittle quasi-likelihood estimators for the Pincherle ARMA process are consistent. □

Theorem 6.19. *Assuming the conditions of Theorem 6.11 holds, then the Whittle quasi-likelihood estimators for the Horadam ARMA process are consistent. That is,* $\lim_{n\to\infty} \hat{\theta}_n = \theta \quad a.s.$

Proof. Assume that the parameter vector lies in the compact space S_0'. Similar to the previous theorem, taking $\frac{\Theta(B)}{\Phi(B)} = \sum_{j=0}^{\infty}(\psi_j B^j)$ and $(1 - 2uB + B) = \sum_{k=0}^{\infty}\sum_{r=0}^{k}(-1)^r \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k+1)}(2u)^{k-r}B^k$ in Eq. (6.7), the process X_t can be expressed as follows:

$$X_t = \frac{\Theta(B)}{\Phi(B)}(1 - 2uB + B)\epsilon_t = \sum_{j=0}^{\infty}(\psi_j B^j)\sum_{n=0}^{\infty}\sum_{r=0}^{n}(-1)^r \frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}(2u)^{n-r}B^n\epsilon_t$$
$$= \sum_{j=0}^{\infty}(\psi_j B^j)\sum_{n=0}^{\infty}(\beta_n B^n),$$

where $\beta_n = \sum_{r=0}^{n}(-1)^r \frac{\Gamma(\nu+n)}{\Gamma(\nu)\Gamma(n+1)}\binom{n}{r}(2u)^{n-r}$. Further, the X_t can have following representation

$$X_t = \sum_{j=0}^{\infty} a_k \epsilon_{t-k},$$

where $a_k = \sum_{s=0}^{k} \psi_{k-s}\beta_s$. Similar to the previous theorem using Stirling's approximation, it can be easily proved that $\sum_{k=0}^{\infty}|a_k| < \infty$ and $a_0 = 1$ for $|u| \leq 1$ and $|\nu| < 1/2$. Also, the spectral density for the Horadam ARMA process using Eq. (6.6) can be written as

$$f_x(\omega) = \frac{\sigma^2}{2\pi}K(\omega),$$

where $K(\omega) = \frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u - 4u\cos(\omega) + 2\cos(\omega))^{-\nu}$, $z = e^{-\iota\omega}$. To prove that $\frac{1}{K(\omega)+a}$ is continuous for $a > 0$ we have

$$\frac{1}{K(\omega)+a} = \frac{\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u - 4u\cos(\omega) + 2\cos(\omega))^{\nu}}{a\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u - 4u\cos(\omega) + 2\cos(\omega))^{\nu} + 1},$$

here $(2+4u^2-4u-4u\cos(\omega)+2\cos(\omega))^{\nu}$ is continuous for $0 < \nu < 1/2$ and $|u| < 1$ implying $\frac{1}{K(\omega)+a}$ is continuous for $|\nu| < 1/2$ and $|u| \leq 1$. Moreover, the parameter space $\theta \in S_0'$ defines the spectral density uniquely. These conditions satisfy the results given by Hannan [71] and hence prove the consistency. □

Theorem 6.20. *Assuming the conditions of Theorem 6.11 holds, then the Whittle quasi-likelihood estimators for the Horadam-Pethe ARMA process are consistent. That is,* $\lim_{n\to\infty}\hat{\theta}_n = \theta \quad a.s.$

Proof. Assume that the parameter vector lies in the compact space S_0'. The Horadam-Pethe ARMA process X_t has the following moving average representation

$$\begin{aligned} X_t &= \left(\sum_{j=0}^{\infty}\psi_j B^j\right)\left(\sum_{n=0}^{\infty}\sum_{r=0}^{\infty}(-1)^r\frac{\Gamma(\nu+n)}{\Gamma(n+1)\Gamma n}\binom{n}{r}(2u)^{n-r}B^{n+2r}\right)\epsilon_t \\ &= \left(\sum_{j=0}^{\infty}\psi_j B^j\right)\left(\sum_{i=0}^{\infty}\zeta_i B^i\right)\epsilon_t, \end{aligned}$$

where

$$\zeta_i = \frac{\Gamma(\nu+i)}{\Gamma(\nu)\Gamma(i+1)}(2u)^i - \frac{\Gamma(\nu+i-2)(2u)^{i-3}}{\Gamma(\nu)\Gamma(2)\Gamma(i-2))}. \tag{6.13}$$

We can rewrite X_t as follows

$$X_t = \sum_{k=0}^{\infty} a_k B^k \epsilon_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k},$$

where $a_k = \sum_{s=0}^{k} \psi_{k-s}\zeta_s$. The proof for $\sum_{k=0}^{\infty} |a_k| < \infty$ is similar to Theorem 6.18. Also, it can be proved that $\frac{1}{K(\omega)+a}$ is continuous for $a > 0$, here it can be expressed as

$$\frac{1}{K(\omega)+a} = \frac{\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2+4u^2-4u(\cos(\omega)+\cos(2\omega))+2\cos(3\omega))^{\nu}}{a\frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2+4u^2-4u(\cos(\omega)+\cos(2\omega))+2\cos(3\omega))^{\nu}+1}.$$

Again using the same argument as Theorem 6.18 it can be proved that $\frac{1}{K(\omega)+a}$ is continuous for $a > 0$ and the parameter vector θ defines the spectral density uniquely. These all conditions satisfies the results defined by Hannan (1973) [71] hence prove the consistency of the Whittle quasi-likelihood estimators. □

Remark 6.3. In order to establish the normality of the estimators, we exclude the scenario where the spectral density is unbounded that is at u = 0 and u = 1. For a detailed examination of the unbounded spectral density case at these points, refer to the paper by Fox and Taqqu [56].

In the next results, $\frac{\partial \log K(\omega)}{\partial \theta}$ is 2×1 column vector which represents the derivative of $\log K(\omega)$ with respect to both the parameters ν and u and $\left\{\frac{\partial \log K(\omega)}{\partial \theta}\right\}\left\{\frac{\partial \log K(\omega)}{\partial \theta}\right\}'$ will be a 2×2 matrix, where $\left\{\frac{\partial \log K(\omega)}{\partial \theta}\right\}'$ represents the transpose of a 2×1 column vector.

Theorem 6.21. *Let the Whittle quasi-likelihood estimate for the Pincherle ARMA process be defined as follows*

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\, W_n(\theta),\ \theta \in \Omega_0,$$

where $\Omega_0 \subset \Omega = \{\nu, u : |\nu| < 1/2, 0 < u < 2/3\}$ *is a compact set then for the Whittle quasi-likelihood estimators for Pincherle ARMA process the* $n^{1/2}(\hat{\theta}_n - \theta) \sim \mathcal{N}(0, W^{-1})$, *where* W *represents the variance-covariance matrix having the following form*

$$W = \frac{1}{4\pi}\int_{-\pi}^{\pi} \left\{\frac{\partial \log K(\omega)}{\partial \theta}\right\}\left\{\frac{\partial \log K(\omega)}{\partial \theta}\right\}' d\omega.$$

Proof. Using the results defined by Hannan (1973) (see theorem 2 in [71]) we need to verify the following conditions to check the asymptotic normality of the parameters.

(a) $K(\omega) > 0$ for all $\omega \in (-\pi, \pi)$ and $\theta \in \Omega_0$.

(b) $K(\omega)$ twice differentiable of parameters ν and u.

(c) The time series defined in Eq. (6.8) can be written as $X_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k}$, $\sum_{k=0}^{\infty} a_k < \infty$ and $a_0 = 1$.

The condition (a) can be proved by rewriting $K(\omega)$ as follows

$$\begin{aligned}
K(\omega) &= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 9u^2 + 2\cos(3\omega) - 6u(\cos(\omega) + \cos(2\omega))^{-\nu} \\
&= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}((2 + 2\cos(3\omega) + [3u - \{\cos(\omega) + \cos(2\omega)\}]^2) - [\cos(\omega) + \cos(2\omega)]^2)^{-\nu} \\
&= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}\left(4\cos^2\left[\frac{3\omega}{2}\right] + [3u - \{\cos(\omega) + \cos(2\omega)\}]^2 - 4\cos^2\left[\frac{3\omega}{2}\right]\cos^2\left[\frac{\omega}{2}\right]\right)^{-\nu} \\
&= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}\left(4\cos^2\left[\frac{3\omega}{2}\right]\sin^2\left[\frac{\omega}{2}\right] + \left[3u - 2\cos\left(\frac{3\omega}{2}\right)\cos\left(\frac{\omega}{2}\right)\right]^2\right)^{-\nu}.
\end{aligned}$$

This indicates that $K(\omega) > 0$ for $0 < u < 2/3$ and $|\nu| < 1/2$. The condition (b) can be easily verified as the function does not have any singularity for $0 < u < 2/3$ hence continuous and differentiable. Moreover, the condition (c) is proven in the Theorem 6.18. Thus the Whittle quasi-likelihood estimates for $|\nu| < 1/2$ and $0 < u < 2/3$ are asymptotically normal. □

Theorem 6.22. *Let the Whittle quasi-likelihood estimate for the Horadam ARMA process be defined as follows*

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\, W_n(\theta), \theta \in \Omega_0',$$

where $\Omega_0' \subset \Omega' = \{\nu, u : |\nu| < 1/2, 0 < u < 1\}$ *is a compact set then for the Whittle quasi-likelihood estimators for Horadam ARMA process the* $n^{1/2}(\hat{\theta}_n - \theta) \sim \mathcal{N}(0, W^{-1})$, *where* W *represents the variance-covariance matrix having the following form*

$$W = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial \log K(\omega)}{\partial\theta}\right\}\left\{\frac{\partial \log K(\omega)}{\partial\theta}\right\}' d\omega.$$

Proof. This can be proved again using the conditions defined in Theorem 6.21. The $K(\omega)$ for the Horadam ARMA process can be written as follows

$$\begin{aligned}
K(\omega) &= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u - 4u\cos(\omega) + 2\cos(\omega))^{-\nu} \\
&= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}\left(4\cos^2\left[\frac{\omega}{2}\right]\sin^2\left[\frac{\omega}{2}\right] + \left[2u - 2\cos^2\left(\frac{\omega}{2}\right)\right]^2\right)^{-\nu},
\end{aligned}$$

From this expression, it is evident that $K(\omega)$ is strictly greater than zero and that $K(\omega)$ is twice differentiable for all values of $\theta \in \Omega_0'$. Furthermore, the moving average representation for the Horadam process is provided in Theorem 6.19. By establishing this moving average representation, we have successfully proven the desired result. □

Theorem 6.23. *Let the Whittle quasi-likelihood estimate for the Horadam-Pethe ARMA process be defined as follows*

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\, W_n(\theta), \theta \in \Omega_0',$$

where $\Omega_0' \subset \Omega' = \{\nu, u : |\nu| < 1/2, 0 < u < 1\}$ *is a compact set then for the Whittle quasi-likelihood estimators for Horadam-Pethe ARMA process the* $n^{1/2}(\hat{\theta}_n - \theta) \sim \mathcal{N}(0, W^{-1})$*, where* W *represents the variance-covariance matrix having the following form*

$$W = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial \log K(\omega)}{\partial\theta}\right\}\left\{\frac{\partial \log K(\omega)}{\partial\theta}\right\}' d\omega.$$

Proof. The $K(\omega)$ for the Horadam-Pethe ARMA process can be written as follows

$$\begin{aligned} K(\omega) &= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}(2 + 4u^2 - 4u(\cos(\omega) + \cos(2\omega)) + 2\cos(3\omega))^{-\nu} \\ &= \frac{|\Theta(z)|^2}{|\Phi(z)|^2}\left(4\cos^2\left[\frac{3\omega}{2}\right]\sin^2\left[\frac{(\omega}{2}\right] + \left[2u - 2\cos\left(\frac{3\omega}{2}\right)\cos\left(\frac{\omega}{2}\right)\right]^2\right)^{-\nu}. \end{aligned}$$

This expression indicates that $K(\omega)$ is greater than zero and possesses two continuous derivatives for all values of $\theta \in \Omega_0'$. Furthermore, Theorem 6.20 provides the moving average representation for the Horadam-Pethe ARMA process. Thus, by establishing the aforementioned moving average representation and considering the expression for $K(\omega)$, the desired result has been successfully demonstrated. □

6.5 Simulation study for Pincherle ARMA process and its application

In order to evaluate the efficacy of the parameter estimation techniques introduced, we employ simulated data. The simulation study serves as a valuable tool in the evaluation of parameter estimation techniques. It enables us to empirically examine the accuracy and reliability of the estimation method by comparing the estimated parameters to the actual parameters obtained from synthetic data. The use of

simulated data provides several advantages for performance assessment. Firstly, it allows us to create controlled experiments where the true parameters are known, facilitating a direct comparison. Secondly, simulations provide the flexibility to generate data with specific properties or characteristics, allowing us to investigate the behavior of estimation techniques under different scenarios. Finally, by repeating the simulation process multiple times, we can obtain statistical measures of performance, such as average estimation error, providing a more comprehensive evaluation. By conducting a simulation study, we can gather empirical evidence that sheds light on the effectiveness of these statistical techniques. Through the utilization of appropriate simulation methods, we generate a synthetic time series based on an initial set of parameters. Subsequently, we apply the defined parameter estimation techniques to the simulated series, aiming to estimate the underlying parameters.

By comparing the estimated parameters with the actual parameters used in the simulation, we can assess the performance of the applied techniques. This comparison serves as a fundamental metric to evaluate the accuracy and reliability of the estimation methods. If the estimated parameters closely align with the actual parameters, it indicates that the techniques effectively capture the underlying characteristics of the data. On the other hand, significant discrepancies between the estimated and actual parameters may indicate limitations or potential areas for improvement in the estimation techniques.

The data from the Pincherle ARMA process is simulated by first simulating the i.i.d. innovations $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. The simulation and estimation study is done using *R*. Now we use the relation defined in Eq. (6.8) as,

$$X_t = \frac{\Theta(B)}{\Phi(B)}(1 - 3uB + B^3)^{-\nu}\epsilon_t. \tag{6.14}$$

The series $Z_t = (1 - 3uB + B^3)^{-\nu}\epsilon_t$ is generated using the simulated innovation series ϵ_t in Eq. (6.14). The generation of the series is done by taking the binomial expansion of $(1 - 3uB + B^3)^{-\nu}$ up to 4 terms, which is given as follows

$$\begin{aligned} Z_t = (1 - 3uB + B^3)^{-\nu}\epsilon_t &= \sum_{n=0}^{\infty}\sum_{j=0}^{n}(-1)^j\frac{(\nu)_n}{n!}\binom{n}{j}(3u)^{n-j}B^{2j+n}\epsilon_t \\ &= \sum_{n=0}^{\infty}\sum_{j=0}^{n}(-1)^j\frac{(\nu)_n}{n!}\binom{n}{j}(3u)^{n-j}\epsilon_{t-n-2j}. \end{aligned}$$

Now by generating the series Z_t the Eq. (6.14) takes the following form

$$X_t = \frac{\Theta(B)}{\Phi(B)}Z_t,$$

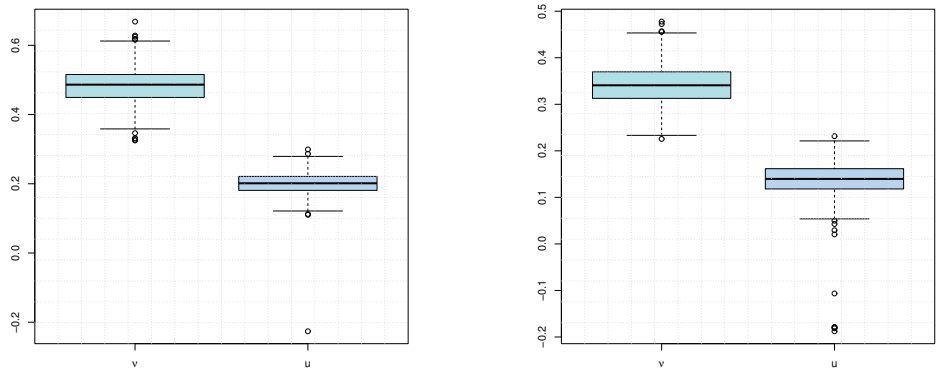
which is nothing but the ARMA process which is simulated using the inbuilt R library by passing the Z_t as innovation series. Further to check the effectiveness of the model the parameter estimation is done on the simulated series. By assuming two different combinations of the initial set of parameters that is $u = 0.2, \nu = 0.4$ and $u = 0.1, \nu = 0.3$ the series Z_t is generated and the two series Pincherle ARMA$(1, 0.2, 0.15, 0)$ and Pincherle ARMA$(1, 0.3, 0.1, 0)$ is generated using the above-defined procedure. The results of parameter estimation using the Whittle quasi-likelihood approach are summarized in the following Table 6.1.

	Actual	**Estimated**
Case 1	$\nu = 0.2$, $u = 0.15$	$\hat{\nu} = 0.21$, $\hat{u} = 0.13$
Case 2	$\nu = 0.3$, $u = 0.1$	$\hat{\nu} = 0.32$, $\hat{u} = 0.07$

Table 6.1: Actual and estimated parameter values for two different choices of parameters estimated by the Pincherle ARMA process using the Whittle quasi-likelihood approach.

From the above Table 6.1 it is clearly shown that the estimate of ν and u from the Whittle quasi-likelihood approach is good. In order to assess the effectiveness of the Whittle quasi-likelihood technique based on empirical spectral density, we construct box plots for different parameters. To create these box plots, we perform a simulation of 1000 series, assuming fixed values for the parameters $\nu = 0.4$ and $u = 0.1$. Each simulated series consists of 1000 observations. Using the Whittle quasi-likelihood estimation method, we estimate the parameters ν and u from each simulated series. By repeating this process for all 1000 simulated series, we obtain a distribution of estimated parameters for each parameter. we construct box plots. Each box plot represents the variability and central tendency of the estimated parameters across the 1000 simulations. The box plot displays the median value, the interquartile range (IQR), and any potential outliers for each parameter. The box plots provide a comprehensive visual representation of the estimated parameters' variability and the overall performance of the Whittle quasi-likelihood technique. The following Fig. 6.3 displays the box plots for the estimated parameters obtained from each simulation, allowing for a visual assessment of their distribution and variability.

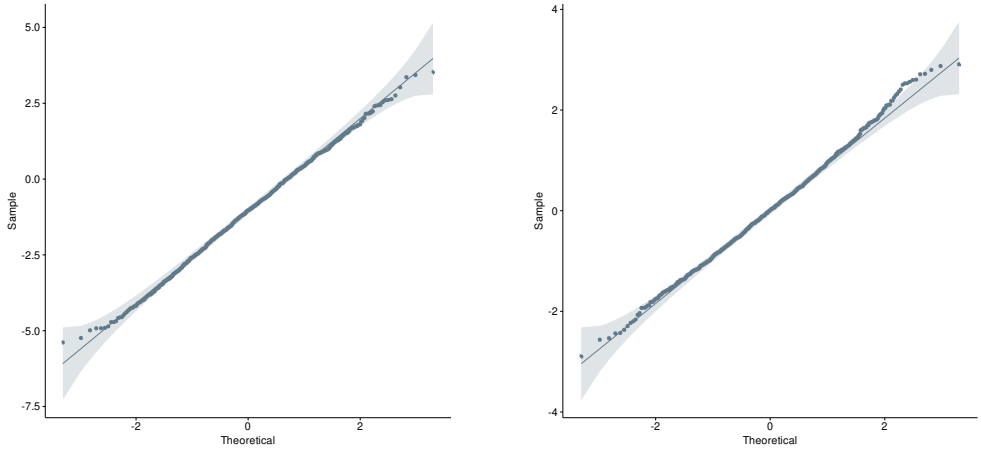
To assess the asymptotic normality of the estimates, we performed a comprehensive simulation analysis, which consists of 1000 datasets. Each dataset consisted of a sequence of length 1000, with the parameter values set at $\nu = 0.45$ and $\nu = 0.2$. The estimated parameters were denoted as $\hat{\nu}$ and $\hat{u}$. In order to visualize the results, we constructed QQ plots for the standardized differences, namely $\sqrt{n}(\hat{\nu} - \nu)$ and $\sqrt{n}(\hat{u} - u)$. These plots provide a graphical representation of the comparison between the observed quantiles and the theoretical quantiles of the standard normal



(a) (b)

Figure 6.3: Box plot of parameters using 1000 samples for $\nu = 0.45$ and $u = 0.2$ (left) and for $\nu = 0.35$ and $u = 0.1$ (right) based on Whittle quasi-likelihood approach

distribution. The QQ plots for $\sqrt{n}(\hat{d} - d)$ and $\sqrt{n}(\hat{u} - u)$ are depicted in Fig. 6.4.



(a) (b)

Figure 6.4: QQ plots using 1000 samples for $\sqrt{n}(\hat{\nu} - \nu)$ (left) and $\sqrt{n}(\hat{u} - u)$ (right)

Furthermore, we demonstrate the normality of the estimated parameters by conducting the Shapiro-Wilk normality test. The resulting p-values for both parameters, ν and u are found to be greater than 0.05. This indicates that the variables $\sqrt{n}(\hat{d} - d)$ and $\sqrt{n}(\hat{u} - u)$ follow a normal distribution.

Real Data Application: We conduct an analysis using the Pincherle ARMA model on the daily percentage yield data of Spain's 10-year treasury bond. The data covers the period from October 7th, 2011 to June 7th, 2018. The Pincherle ARMA model is compared with other existing models, namely autoregressive integrated moving average (ARIMA), autoregressive fractionally integrated moving average (ARFIMA), autoregressive tempered fractionally integrated moving average(ARTFIMA), and Gegenbauer autoregressive moving average(GARMA).

The daily percentage yield is a commonly used metric in financial markets to measure the return on investment for fixed-income securities, such as government bonds. It represents the change in the bond's yield, expressed as a percentage, from one day to the next. This yield data is of particular interest to investors, traders, and policymakers as it provides insights into the performance and market dynamics of long-term government debt. By analyzing this dataset, we can gain valuable insights into the behavior and patterns of Spain's 10-year treasury bond yield over the given time frame. The objective of applying various models, including the Pincherle ARMA model, is to accurately capture and forecast the future movements and trends in the bond yield, thereby assisting in decision-making processes related to investment strategies, risk management, and financial planning. The trajectory plot for the introduced dataset is given in Fig. 6.5

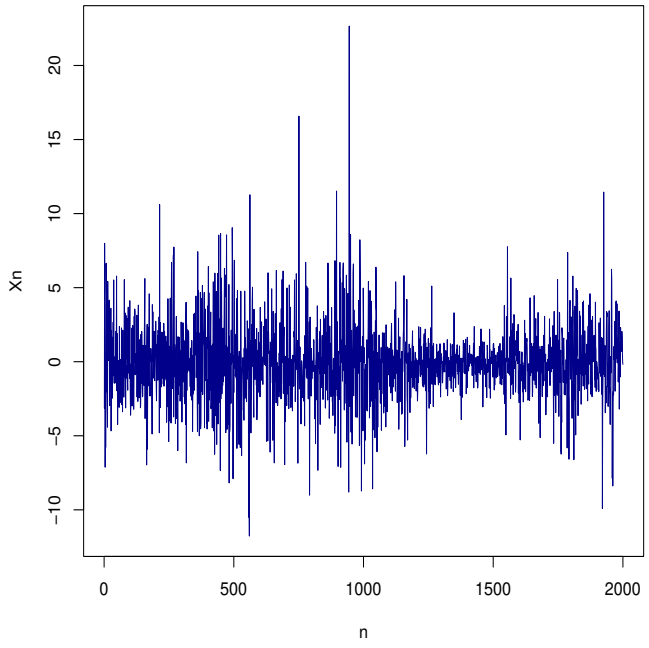


Figure 6.5: Trajectory plot for Spain's 10-year treasury yield series dataset.

To evaluate the accuracy of these models, we utilize two common metrics: the root mean square error (RMSE) and the mean absolute error (MAE). These metrics provide measures of the deviation between the predicted values and the actual values of the bond yield data.

Table 6.2 presents the results of the accuracy assessment for each model. From the table, we observe that the RMSE of the Pincherle ARMA process is lower than that of the other models. This indicates that the Pincherle ARMA model performs better in terms of capturing the overall variability in the bond yield data. However,

it is worth noting that the MAE for the Pincherle ARMA process is slightly higher compared to the ARFIMA and ARTFIMA models. The MAE represents the average magnitude of the errors made by the models, regardless of their direction. In this case, the slightly higher MAE suggests that the Pincherle ARMA model may have a slightly larger average error in its predictions when compared to the ARFIMA and ARTFIMA models.

Based on these findings, we can conclude that the Pincherle ARMA process demonstrates good accuracy in predicting Spain's 10-year treasury bond yield when compared to the other models evaluated in this study. This suggests that the Pincherle ARMA model can be a valuable tool in other fields where accurate forecasting is required. It may be particularly useful in financial applications where predicting bond yields is crucial for investment decisions and risk management.

Models	**RMSE**	**MAE**
ARIMA	2.588	1.972
ARFIMA	2.476	1.831
ARTFIMA	2.463	1.801
GARMA	2.584	1.962
Pincherle ARMA	2.406	1.840

Table 6.2: The goodness-of-fit measures of different models using RMSE and MAE metrics.

6.6 Conclusion

We study the general Humbert polynomials based autoregressive moving average called here HARMA (p, ν, u, q) time series models. Initially, type 1 HARMA (p, ν, u, q) process defined in Eq. (6.2) and it's stationarity and invertibility conditions are derived. We also compute the spectral density of the above process. For $m = 3$ in Eq. (6.8), we focus on a particular case Pincherle ARMA (p, ν, u, q) process, by obtaining the spectral density and also prove that for $u = 0$ and $0 < \nu < 1/2$, the process also exhibits seasonal long memory property. In the subsequent section, we study similar properties of particular cases of type 2 HARMA (p, ν, u, q) process defined in Eq. (6.2) for $m = 1$ and $m = 3$ named as Horadam ARMA process and Horadam-Pethe ARMA process respectively. We also provide the Whittle quasi-likelihood estimation method to estimate the parameters of the HARMA process. We also provide the results for the consistency and normality of the estimators. The simulation study on 1000 series each of length 1000 is performed and real data of Spain's 10-year treasury bond daily percentage yield is used to show the application of the Pincherle ARMA model.

Further, we believe that the proposed time series models will be helpful in the modeling of real-world data from other fields. Also, the estimation techniques for example minimum contrast estimation [8,10] will be applied for the discussed models. This technique estimates the parameters by minimizing the spectral density and empirical spectral density of the process. Maximum likelihood estimation is the particular case of minimum contrast estimation. Apart from this, Pincherle, Horadam and Horadam-Pethe random fields will be interest of study on the line of Gegenbauer random fields [51].

Chapter 7

Conclusion and Future Work

This thesis majorly revolves around the autoregressive (AR) models with non-Gaussian innovation terms. Amongst the non-Gaussian distributions, we focus on semi-heavy-tailed and heavy-tailed distributions for innovation terms. This class of distributions form the foundational basis of modeling financial time series. Normal inverse Gaussian is a semi-heavy tailed distribution which has tails heavier than the normal and lighter than the power law tails. First, the AR models with normal inverse Gaussian distributed residuals are studied. We have illustrated the key properties of the examined models. The main part is devoted to the estimation algorithm for the considered models' parameters. The technique incorporates the EM algorithm which is widely used in the time series analysis. The effectiveness of the proposed algorithm is showcased through simulations using data generated from AR(2) and AR(1) models. Our results suggest that EM estimation approach outperforms traditional methods based on the YW and CLS algorithms. Additionally, we compare the estimation of parameters α and δ for NIG-distributed residuals using the ML estimate by the Newton method and the EM algorithm. Finally, we demonstrate that an AR(1) model with NIG residuals explain well the Google equity price and NASDAQ stock market index data. The AR(1) model also effectively captures the behavior of another data namely, US gasoline price data. The satisfactory results on three distinct datasets show the versatility of the proposed model. We believe that the discussed model can be used to describe various real-life time series ranging from finance and economics to natural hazards, ecology, and environmental data. Furthermore, residuals with non-zero mean can be considered for future exploration.

Further, we consider the heavy-tailed distribution for innovation terms of AR(p) model, that is Cauchy distribution. We obtain the closed-form estimates for the model using the EM algorithm. We evaluate the performance of EM algorithm by comparing it to the maximum likelihood method, implemented through an R inbuilt function, on simulated data. An added advantage of the EM algorithm is its simultaneous computation of both model and innovation parameters. The boxplot analysis clearly demonstrates the superior performance of the EM algorithm over the ML method. The ML estimates also depend on the underlying numerical

optimization method used. Furthermore, we explore an alternative approach based on the empirical characteristic function to estimate the parameters of the AR model with Cauchy distribution.

Moreover, we extend the work by considering the other class of distributions known as geometric infinitely divisible. We use the Bernstein function $g(s)$, representing the Laplace exponent of a positive infinitely divisible random variable, to define geometric infinitely divisible random variables. The Laplace transform of these variables takes the form $\frac{1}{1+g(s)}$. Additionally, we determine the Laplace transform of mixtures involving specific geometric infinitely divisible random variables, establishing a new class of marginals for exploration. The autoregressive process of order 1 with geometric infinitely divisible distribution is introduced. We derive that if the marginals of the AR(1), as defined in Eq. (5.2), are geometric infinitely divisible, then the innovation terms also exhibit geometric infinitely divisible characteristics. The integral form of the probability density function for innovation terms is obtained using the Laplace transform and the complex inversion method. This analysis is conducted for three cases, namely, geometric tempered stable, geometric gamma, and geometric inverse Gaussian. Furthermore, we calculate the first and second order moments for these three geometric infinitely divisible random variables, crucial for studying characteristics of distribution. Next, we extend the AR process to the k^{th} order and propose the AR(1) model defined in Prop. 5.9, with marginals having a Laplace transform of the form $\frac{1}{1+g(s)}$. Finally, we estimate the parameters of the model defined in Prop. 5.9 using conditional least squares and method of moments. The simulation study for two cases indicates that the estimates are satisfactory. Empirically it is observed that these three distributions also have semi-heavy-tailed behavior.

We further extend the work available in literature on fractionally integrated autoregressive moving average processes. We utilize the generating functions of Humbert polynomials and define two types of Humbert generalized fractional differenced ARMA processes (HARMA), with a specific emphasis on Pincherle ARMA, Horadam ARMA, and Horadam-Pethe ARMA processes. Then, we establish stationarity and invertibility conditions for these processes. Also, the long memory property of the Pincherle ARMA process is achieved when the fractionally differencing parameter $u = 0$. Parameter estimation is carried out using the Whittle quasi-likelihood method, yielding consistent and normally distributed estimators. The method is validated through a simulation study focusing on the Pincherle ARMA process. Further, the Pincherle ARMA model is applied to Spain's 10-year treasury bond yield data which effectively capture the market dynamics.

In conclusion, we believe that the proposed models will significantly contribute to the field of statistical modeling with ability to better capture and incorporate the dynamics of real-life data.

In the future, we plan to delve into a comprehensive study and comparison of EM algorithm and the empirical characteristic function based estimation method with existing techniques presented in [83,102,133] for AR models with infinite variance. Additionally, the applicability of the AR model to real-life phenomena will be explored in detail. Moreover, we can explore tempered variations of Humbert, Pincherle, Horadam, and Horadam-Pethe ARMA processes, akin to the approach taken by Sabzikar et al. [125]. The presence of singularities in the spectral density of HARMA models adds complexity to the implementation of estimation techniques. It is crucial to acknowledge that the HARMA processes in this study assume constant volatility. Consequently, these models may not be well-suited for capturing heteroscedastic data that demonstrates persistence in the conditional variance of the innovation term. In the future, it would be interesting to extend the concept of the Humbert processes to incorporate volatility modeling, specifically by developing a Humbert-GARCH process. Further, the heavy-tailed time series models can be extended to regime switching modeling where the innovation terms have different distribution parameters across different regimes. Also, the application of EM algorithm can be explored to estimate the parameters of distributions which lack analytic form of probability density function.

Appendix A

Python and R Codes

A.1 EM algorithm for AR(2) model with NIG innovations

Following is the Python code to implement EM algorithm on the simulated data from AR(2) model with NIG innovations with fixed $\mu = 0$, $\beta = 0$ parameters. The file NIGError_data.csv is the error data simulated from NIG distribution and X_AR2_Errordata.csv is the simulated time series data. The estimated parameter values are saved in Para_est_MBfix.csv file.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.special as sp

#################### 1. Initializations of parameters. ###############

E = pd.read_csv('NIGError_data.csv', header=None )
Y = pd.read_csv('X_AR2_Errordata.csv')
E.drop([0], axis=0, inplace=True)
E = E.reset_index(drop=True)

est_values = []
for k in range(1000):
    y = E.iloc[:,k]
    y = y.to_numpy()
    y.reshape((1000,1))
    n = len(y)-1
    y = y.ravel()
    N = len(y)
    Y1 = Y.iloc[:,k]

#### 2. y array has Error data which follows NIG distribution.####
```

```
    rho = [0.001, 0.001]
    p = len(rho)
    mu = 0.0        # actual mu = 0 fix
    beta = 0.0      # actuual beta = 0 fix
    delta = 0.005   # actual delta = 2
    gamma = 0.005   # gamma = 1
    alpha = np.sqrt(gamma**2+beta**2)
    def phi_update(mu, delta, y):
        phi = []
        for i in range(N):
            phi.append(1 + ((y[i]-mu)/delta)**2)
        phi = np.array(phi)
        return phi

############## 3.Function to define Y_(t-1) p-vector. ##########

    Y1 = Y1.to_numpy()
    Y1.resize((N, ))
    def y_vector(Y, p):

        B=[]
        for i in range(len(Y1)-p+1):
            l=[]
            for j in range(i, i+p):
                l.append(Y1[j])
            B.append(l)
        return B

    B1 = y_vector(Y1, p)
    B1 = np.array(B1)
    B = B1[:, [1,0]]

    phi = phi_update(mu, delta, y)

################## 4. Functions to implement E-step ##########

    def exp_G(delta, mu, beta, alpha, gamma, y):
        phi = []
```

```
    s = []
    for i in range(N):
        phi.append(1 + ((y[i]-mu)/delta)**2)
        s.append(delta*(np.sqrt(phi[i]))*sp.k0(delta*alpha*\
        np.sqrt(phi[i]))/(alpha*sp.k1(delta*alpha*np.sqrt(phi[i]))))
    phi = np.array(phi)
    s = np.array(s)
    s_b = np.average(s)
    return s, s_b

def inv_exp_G(delta, mu, beta, alpha, gamma, y):
    phi = []
    w = []
    for i in range(N):
        phi.append(1 + ((y[i]-mu)/delta)**2)
        w.append((alpha*(sp.k0(delta*alpha*np.sqrt(phi[i])) + \
            (2/(delta*alpha*np.sqrt(phi[i])))*\
            sp.k1(delta*alpha*np.sqrt(phi[i]))))/(delta\
            *(np.sqrt(phi[i]))* sp.k1(delta*alpha*np.sqrt(phi[i]))))
    w = np.array(w)
    w_b = np.average(w)
    return w, w_b

################ 5. Function to update parameters ###############

def terms(delta, mu, beta, alpha, gamma, rho, y):
    s, s_b = exp_G(delta, mu, beta, alpha, gamma, y)
    w, w_b = inv_exp_G(delta, mu, beta, alpha, gamma, y)
    d_new = np.sqrt(s_b/(s_b*w_b-1))
    g_new = d_new/s_b
    a_new = np.sqrt(g_new**2)
    C = []
    sum1 = 0
    sum2 = 0
    for i in range(N-p):
        C = B[i].reshape(p,1)
        term1 = w[i+p]*np.dot(C, C.T)
        term2 = (w[i+p]*Y1[i+p]-mu*w[i+p]-beta)*C
        sum1 = sum1 + term1
```

```
            sum2 = sum2 + term2
            inv_sum = np.linalg.inv(sum1)

        rho_est = (inv_sum)@sum2
        return d_new, g_new, a_new, rho_est

    d_new, g_new, a_new, rho_est = terms( delta, mu, beta, alpha,\
    gamma, rho, y)

    while (max(abs(delta-d_new)/abs(delta), abs(gamma-g_new)/\
    abs(gamma),abs(rho[0]-rho_est[0])/abs(rho[0]), \
    abs(rho[1]-rho_est[1])/abs(rho[1]))) > 0.0001:

        mu = 0
        beta = 0
        delta = d_new
        gamma = g_new
        alpha = a_new
        rho = rho_est
        d_new, g_new, a_new, rho_est = terms(d_new, mu, beta, a_new,\
        g_new, rho, y)
    l = [d_new, g_new, rho_est[0], rho_est[1]]
    est_values.append(l)

print("Output:\ndelta", d_new, "\ngamma", g_new, "\nalpha", a_new,\
"\nrho", rho_est)

########################### 6. Estimated data to csv #############

df = pd.DataFrame(est_values)
df.to_csv('Para_est_MBfix.csv', index=False, header=False)
```

A.2 EM algorithm for AR(2) model with Cauchy innovations

Following is the Python code to implement EM algorithm on the simulated data from AR(2) model with Cauchy innovations and $\rho_1 = 0.5, \rho_2 = 0.3$. File Cauchy_Errors_EE1.csv has the error data simulated from Cauchy with parameters

$\alpha = 1, \gamma = 2.$

```
############## Code to implement EM algorithm.###########

import numpy as np
import pandas as pd
import scipy.special as sp
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.ar_model import ar_select_order
import matplotlib.pyplot as plt
import datetime as dt

############ 1. Initializations of parameters. ###############

start_time = dt.datetime.now()
Y = pd.read_csv('yy_seriesAR2_try1.csv')
y1 = Y.iloc[:,0]
N = Y.shape[0]
M = len(Y.columns)
est_values = []

############# 2. Loop for each trajectory ##################

for k in range(M):
    Y1 = Y.iloc[:,k]
    rho = [0.1, 0.1]
    p = len(rho)
    rho = np.array(rho)
    rho = rho.reshape((1,2))
    gamma = 0.02
    alpha = 0.01

####### 3. Function to define Y_(t-1) p-vector.##################

    Y1 = Y1.to_numpy()
    Y1.resize((N, ))
    def y_vector(Y, p):
```

```
        B=[]
        for i in range(len(Y1)-p+1):
            l=[]
            for j in range(i, i+p):
                l.append(Y1[j])
            B.append(l)
        return B

    B1 = y_vector(Y1, p)
    B1 = np.array(B1)
    B = B1[:, [1,0]]

    def AR_residuals(Y1, rho, B):
        s_vec = []
        for i in range(N-p):
            s = Y1[i+p] - np.sum((rho*B[i,:]))
            s_vec.append(s)
        return s_vec

    def inv_exp_G(alpha, gamma, Y1, rho, B):
        y = AR_residuals(Y1, rho, B)
        w = []
        for i in range(N-p):
            s = (y[i] - alpha)**2 + gamma**2
            w.append(2/s)
        w = np.array(w)
        return w

############ 4. Function to get estimates ##################

    def terms(alpha, gamma, rho, Y1):
        y = AR_residuals(Y1, rho, B)
        w = inv_exp_G(alpha, gamma, Y1, rho, B)
        a_new = (np.sum(y*w))/(np.sum(w))
        g_new = np.sqrt((N-p)/np.sum(w))
        C = []
        sum1 = 0
        sum2 = 0
        for i in range(N-p):
```

```
            C = B[i].reshape(p,1)
            term1 = w[i]*np.dot(C, C.T)
            term2 = (w[i]*Y1[i+p]-alpha*w[i])*(C.T)
            sum1 = sum1 + term1
            sum2 = sum2 + term2
            inv_sum = np.linalg.pinv(sum1)
            #print(inv_sum)

        rho_est = sum2@(inv_sum)
        return a_new, g_new, rho_est

    a_new, g_new, rho_est = terms(alpha, gamma, rho, Y1)

############ 5. Using the relative change in the parameters as
##stopping criterion. ############

    while (max(abs(alpha-a_new)/abs(alpha),abs(gamma-g_new)/abs(gamma),
    abs(rho[0,0]-rho_est[0,0])/abs(rho[0,0]),
    abs(rho[0,1]-rho_est[0,1])/abs(rho[0,1]))) > 0.0001:

        gamma = g_new
        alpha = a_new
        rho = rho_est
        s_vec=[]
        a_new, g_new, rho_est = terms(a_new, g_new, rho, Y1)
    l = [a_new, g_new, rho_est[0,0], rho_est[0,1]]
    est_values.append(l)
print("Output:\nalpha", a_new, "\ngamma", g_new, "\nrho", rho_est)
end_time = dt.datetime.now() - start_time
print("Execution time:\t", end_time)

############### 6. Estimated data to csv file ########################

df = pd.DataFrame(est_values)
df.to_csv('EM_Para_est_Cauchyerror1.1.csv', index=False, header=False)
```

A.3 R Code to generate geometric infinitely divisible random variables

Following is the R code to simulate geometric tempered stable random variables using Laplace transform.

```
# ================================================
# To simulate geometric tempered stable random variables with
#parameters \beta=0.6 and \lambda=1.
# ================================================

#---------------------
# Define the transform
#---------------------
lt.gammapdf <- function(s, alpha, beta) {
  1/(1+((s+alpha)^(beta)-alpha^beta))
}

#---------------------
# Set parameter values
#---------------------
beta <- 0.6
alphavals = c(1)
#-------------------
# Get the generators
#-------------------
source("/home/monika/Documents/WORK/Work_Bernstein/rcode/rlaptrans.r")
#---------------------------------
# Initialise random number generator
#---------------------------------
set.seed(682)

#-------------------------------------
# Generate 500 trajectories of geometric tempered stable
#distribution each of size 1000.
#-------------------------------------
par(mfrow=c(2,2))
df_traj = list()
```

```
for (i in 1:500){
  for (alpha in alphavals) {
    x.rlap <- rlaptrans(1000, lt.gammapdf, alpha, beta)
    df_traj[[i]] = x.rlap
    w.rlap <- density(log(x.rlap)) # kernel density estimation
    u.rlap <- exp(w.rlap$x)
    v.rlap <- w.rlap$y / u.rlap
    df_u = as.data.frame(u.rlap)
    df_u$vrlap = c(v.rlap)

  }
}

df_x = data.frame(df_traj)
names(df_x) = NULL
write.csv(df_x,file = "/home/monika/Documents/WORK/Work_Bernstein/
rcode/AR1_Y_traj_ts.csv")

# ================================================
# Generate error terms of AR(1) model with geometric tempered stable
#marginal with parameters rho =0.5, beta = 1, and lambda = 2.
# ================================================

#---------------------
# Define the transform
#---------------------
lt.gammapdf <- function(s, alpha, beta) {
  (1+(0.5*s+alpha)^beta-alpha^beta)/(1+(s+alpha)^beta-alpha^beta)
}

#---------------------
# Set parameter values
#---------------------
beta <-1         #beta >0
alphavals = c(2)    #lambda > 0
#-------------------
# Get the generators
#-------------------
source("/home/monika/Documents/WORK/Work_Bernstein/rcode/rlaptrans.r")
```

```
#----------------------------------
# Initialise random number generator
#----------------------------------
#set.seed(602)

par(mfrow=c(2,2))
df_traj = list()
for (i in 1:500){
  for (alpha in alphavals) {
    x.rlap <- rlaptrans(1000, lt.gammapdf, alpha, beta)
    df_traj[[i]] = x.rlap
    w.rlap <- density(log(x.rlap)) # kernel density estimation
    u.rlap <- exp(w.rlap$x)
    v.rlap <- w.rlap$y / u.rlap
    df_u = as.data.frame(u.rlap)
    df_u$vrlap = c(v.rlap)

  }
}

df_x = data.frame(df_traj)
names(df_x) = NULL
write.csv(df_x,file = "/home/monika/Documents/WORK/Work_Bernstein/
rcode/AR1_traj_df_x_ts1.2.csv")
```

A.3.1 Python code to estimate parameters of AR(1) model with geometric tempered stable marginals.

```
"""
Parameter estimation for rho = 0.8, beta = 0.3, and lambda = 3 from
geometric tempered stable errors 1000x500.
"""

from scipy.optimize import fsolve
import math
import numpy as np
from scipy.optimize import minimize
import pandas as pd
from scipy.optimize import Bounds
```

```
E = pd.read_csv('AR1_traj_df_x_ts1.3.csv', header = None)
E[E<0.000000001] = 0
E = E.iloc[:,1:501]
E_arr1 = np.array(E)  #error series with beta=0.3 and lambda=3.
shape = E_arr1.shape
n = shape[0]
rho_est_list = []
beta_est_list = []
lamda_est_list = []
df_est = pd.DataFrame()

def generate_yseries(E,n, rho = 0.8):
    Y =[]
    Y.append(E[0])
    for i in range(1,n):
        Y.append(rho*Y[i-1] + E[i])
    return Y

def rho_estimation_CLS(Y):
    Y = np.array(Y)
    Y_t = Y[1:1000]
    Y_t_1 = Y[0:999]
    Yt_bar = np.mean(Y_t)
    Yt1_bar = np.mean(Y_t_1)
    Yt_sum = np.sum(Y_t)
    Yt1_sum = np.sum(Y_t_1)
    rho_est = (np.sum(Y_t*Y_t_1)-n*(Yt1_bar)**2)/(np.sum((Y_t_1)**2)-\
    n*(Yt1_bar)**2)
    return rho_est


def obj(p,*args): ###in terms of beta.
    (b) = p   #theta=0.3, beta=0.6, lambda=1,
    m_1 = np.mean(E_arr)
    m_2 = np.mean(E_arr**2)
    A = (2*(m_1)**2-m_2*(1-rho_est))/(m_1*(1-rho_est**2))
    term = (m_1/1-rho_est) - b*(((b-1)*(1-rho_est**2)*m_1)/(2*m_1**2-
    (1-rho_est)*m_2))**(b-1)
```

```
    return term

def lambda_para_estimate(E_arr, r, rho_est):
    m_1 = np.mean(E_arr)
    m_2 = np.mean(E_arr**2)
    A = (2*(m_1)**2-m_2*(1-rho_est))/(m_1*(1-rho_est**2))
    return 1+r*A

for i in range(shape[1]):
    E_arr = E_arr1[:,i]
    Y = generate_yseries(E_arr, n, rho = 0.8)
    rho_est = rho_estimation_CLS(Y)
    rho_est_list.append(rho_est)
    initial_point = np.random.random_sample(size = 500)
    beta = fsolve(obj, 0.7, args=(E_arr, rho_est))
    lamda = lambda_para_estimate(E_arr, beta, rho_est)
    beta_est_list.append(beta[0])
    lamda_est_list.append(lamda[0])

df_est['beta'] = beta_est_list
df_est['lambda'] = lamda_est_list
df_est['rho'] = rho_est_list
df_est.to_csv('Para_est_ts_500.3.csv')
```

1

[1] https://drive.google.com/drive/folders/1kbE4ghyfsr5I8UdQGuQCm9AddgF3S1ry?usp=drive_link

References

[1] Aas, K., Haff, I. H., and Dimakos, X. K. Risk estimation using the multivariate normal inverse Gaussian distribution. *J. Risk 8*, 2 (2006), 39–60.

[2] Abraham, B., and Balakrishna, N. Inverse Gaussian autoregressive models. *J. Time Ser. Anal. 20-6* (1999), 605–618.

[3] Abramowitz, M., and Stegun, I. A. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Conference on Mathematical Tables, Cambridge (UK), 1954.

[4] Aguilar, J. P. Explicit option valuation in the exponential NIG model. *Quantitative Finance 21*, 8 (2021), 1281–1299.

[5] Albrecher, H., and Predota, M. On Asian option pricing for NIG Lévy processes. *J. Comput. Appl. Math. 172*, 1 (2004), 153–168.

[6] Alfonso, L., Mansilla, R., and Terrero-Escalante, C. A. On the scaling of the distribution of daily price fluctuations in the Mexican financial market index. *Physica A 391*, 10 (2012), 2990–2996.

[7] Angelis, L. D., and Viroli, C. A markov-switching regression model with non-Gaussian innovations: estimation and testing. *Stud. Nonlinear Dyn. E. 21* (2017).

[8] Anh, V., Leonenko, N. N., and Sakhno, L. M. On a class of minimum contrast estimators for fractional stochastic processes and fields. *J. Statist. Plann. Inference. 123*, 1 (2004), 161–185.

[9] Anh, V. V., Knopova, V. P., and Leonenko, N. N. Continuous-time stochastic processes with cyclical long-range dependence. *Aust. N. Z. J. Stat. 46*, 2 (2004), 275–296.

[10] Anh, V. V., Leonenko, N. N., and Sakhno, L. M. Minimum contrast estimation of random processes based on information of second and third orders. *J. Statist. Plann. Inference. 137*, 4 (2007), 1302–1331.

[11] Applebaum, D. *Lévy Processes and Stochastic Calculus. 2nd ed.* Cambridge University Press, Cambridge, U.K., 2009.

[12] Baeumer, B., and Meerschaert, M. Tempered stable Lévy motion and transient super-diffusion. *J. Comput. Appl. Math 233* (2010), 2438–2448.

[13] Baker, B. B. On the relation between pincherle's polynomials and the hypergeometric function. *Proc. Edinburgh Math. Soc. 39*, 4 (1920), 58–62.

[14] Balakrishna, N., et al. *Non-Gaussian autoregressive-type time Series*. Springer, 2021.

[15] Barndorff-Nielsen, O. E. Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London Ser. A. 353* (1977), 401–409.

[16] Barndorff-Nielsen, O. E. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat. 24* (1977), 1–13.

[17] Barndorff-Nielsen, O. E., Mikosch, T., and Resnick, S. I. Lévy processes: Theory and applications. *Proc. Roy. Soc. London Ser. A.* (2013).

[18] Bartle, R. G., and Donald, R. S. *Introduction to real analysis*. New York: John Wiley, 2000.

[19] Benth, F. E., and Benth, J. S. The normal inverse Gaussian distribution and spot price modelling in energy markets. *Int. J. Theor. Appl. Finance 7*, 2 (2004), 177–192.

[20] Benth, F. E., Groth, M., and Kettler, P. C. A quasi-Monte Carlo algorithm for the normal inverse Gaussian distribution and valuation of financial derivatives. *Int. J. Theor. Appl. Finance 9*, 6 (2006), 843–867.

[21] Bhootna, N., Dhull, M. S., Kumar, A., and Leonenko, N. N. Humbert generalized fractional differenced ARMA processes. *Communications in Nonlinear Science and Numerical Simulation 125* (2023).

[22] Bhootna, N., and Kumar, A. Tempered stable autoregressive models. *Commun. Stat. Theory Methods* (2022), 1–21.

[23] Bjerkedal, T. Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *Am J Hyg. 72* (1960), 130–148.

[24] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom. 31*, 3 (1986), 307–327.

[25] Box, G. E. P., and Jenkins, G. M. *Time series analysis : forecasting and control*. Holden-Day, San Francisco, 1976.

[26] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. *Time Series Analysis Forecasting and Control*. Wiley, 2008.

[27] Brockwell, P. J., and Davis, R. A. *Introduction to Time Series and Forecasting*. Springer, 2016.

[28] Broszkiewicz-Suwaj, E., and Wyłomańska, A. Application of non-Gaussian multidimensional autoregressive model for climate data prediction. *Int. J. Adv. Eng. Sci. Appl. Math. 13* (2021), 236–247.

[29] Burnecki, K., Gajda, J., and Sikora, G. Stability and lack of memory of the returns of the Hang Seng index. *Physica A 390*, 18–19 (2011), 3136–3146.

[30] Can, M., Gucuyener, E., and Bostanci, A. Time series analysis of atmospheric pollutants using ARMA models. *Fresenius Environ. Bull. 28(8)*, 3 (2019), 6107–6115.

[31] Casella, G., and Berger, R. L. *Statistical Inference*. Duxbury Press, 2001.

[32] Chang, Y.-P., Hung, M.-C., Liu, H., and Jan, J.-F. Testing symmetry of a NIG distribution. *Commun. Stat. - Simul. Comput. 34*, 4 (2005), 851–862.

[33] Chang, Y.-P., Hung, M.-C., Wang, S.-F., and Yu, C.-T. An EM algorithm for multivariate NIG distribution and its application to value-at-risk. *Int. J. of Inf. 21*, 3 (2010), 265–283.

[34] Choi, J., and Choi, I. Maximum likelihood estimation of autoregressive models with a near unit root and Cauchy errors. *Ann. Inst. Stat. Math. 71* (2019), 1121–1142.

[35] Christmas, J., and Everson, R. Robust autoregression: Student-t innovations using variational Bayes. *IEEE Trans. Signal Process. 59* (2011), 48–57.

[36] Chung, C.-F. A generalized fractionally integrated autoregressive moving-average process. *J. Time Series Anal. 17*, 2 (1996), 111–140.

[37] Cont, R., and Tankov, P. *Financial Modeling With Jump Processes*. Chapman & Hall, CRC Press, London, UK, 2004.

[38] Damsleth, E., and El- Shaarawi, A. H. ARMA models with double-exponentially distributed noise. *Journal of the Royal Statistical Society: Series B (Methodological) 51* (1989), 61–69.

[39] del Barrio Castro, T., and Rachinger, H. Aggregation of seasonal long-memory processes. *Econometrics and Statistics 17* (2021), 95–106.

[40] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. Ser. B. 39* (1977), 1–38.

[41] Devroye, L. *Non-Uniform Random Variate Generation, 1ed.* Springer, New York, 1986.

[42] Dewald, L. S., and Lewis, P. A. W. A new Laplace second-order autoregressive time series model. *IEEE Trans. Inf. Theory 31-5* (1985), 645–651.

[43] Dhull, M. S., and Kumar, A. Normal inverse Gaussian autoregressive model using EM algorithm. *Int. J. Adv. Eng. Sci. Appl. Math. 13* (2021), 139–147.

[44] Dhull, M. S., and Kumar, A. Expectation-Maximization algorithm for autoregressive models with Cauchy innovations. *Engineering Proceedings 18* (2022).

[45] Dhull, M. S., and Kumar, A. Geometric infinitely divisible autoregressive models.

[46] Dhull, M. S., Kumar, A., and Wyłomańska, A. The expectation-maximization algorithm for autoregressive models with normal inverse Gaussian innovations. *Commun. Stat. Simul. Comput.* (2023), 1–21.

[47] Dickey, D. A., and Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc. 74*, 366a (1979), 427–431.

[48] Dissanayake, G., Peiris, M. S., and Proietti, T. Fractionally differenced Gegenbauer processes with long memory: A review. *Statist. Sci. 33* (2018), 413–426.

[49] Elgammal, A., Duraiswami, R., Harwood, D., and Davis, L. Background and foreground modeling using non-parametric kernel density estimation for visual surveillance. *Proceedings of the IEEE 90*, 7 (2002), 1151–1163.

[50] Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 4 (1982), 987–1007.

[51] Espejo, R. M., Leonenko, N. N., and Ruiz-Medina, M. D. Gegenbauer random fields. *Random Operators and Stochastic Equations 22*, 1 (2014), 1–16.

[52] Feller, W. *An Introduction to Probability Theory and Its Applications, 2nd Ed.*, vol. 2. Wiley, 1991.

[53] Ferguson, T. S. A Representation of the Symmetric Bivariate Cauchy Distribution. *The Annals of Mathematical Statistics 33*, 4 (1962), 1256 – 1266.

[54] Forsberg, L., and Bollerslev, T. Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): the GARCH-NIG model. *J. Appl. Econom. 17*, 5 (2002), 535–548.

[55] Foss, S., Korshunov, D., and Zachary. *An introduction to heavy-tailed and subexponential distributions*, vol. 6. New York: Springer, 2011.

[56] Fox, R., and Taqqu, M. S. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Stat. 14*, 2 (1986), 517–532.

[57] Gajda, J., Bartnicki, G., and Burnecki, K. Modeling of water usage by means of ARFIMA–GARCH processes. *Physica A 512* (2018), 644–657.

[58] Gajda, J., Sikora, G., and Wyłomańska, A. Regime variance testing-A quantile approach. *Acta Phys. Pol. B 44*, 5 (2013), 1015–1035.

[59] Gamlin, T. W. *Complex Analysis*. New York: Springer, 2001.

[60] Gary K. Grunwald, Rob J. Hyndman, L. T. . R. L. T. Theory & methods: Non-Gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics, 42* (2000), 479–495.

[61] Gaver, D., and Lewis, P. First-order autoregressive gamma sequences and point processes. *Adv. Appl. Probab. 12-3* (1980), 727–745.

[62] Giraitis, L., Hidalgo, J., and Robinson, P. M. Gaussian estimation of parametric spectral density with unknown pole. *Ann. Statist 29*, 4 (2001), 987–1023.

[63] Giraitis, L., Koul, H. L., and Surgailis, D. *Large sample inference for long memory processes*. World Scientific, 2012.

[64] Gould, H. W. Inverse series relations and other expansions involving Humbert polynomials. *Duke Math. J. 32*, 4 (1965), 697–711.

[65] Grabchak, M. *Tempered stable distributions: stochastic models for multiscale processes*. Springer, New York, 2016.

[66] Granger, C. W. J., and Joyeux, R. An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal. 1* (1980), 15–29.

[67] Gray, H. L., Zhang, N., and Woodward, W. A. On generalized fractional processes. *J. Time Ser. Anal. 10* (1989), 233–257.

[68] Grunwald, G., Adrian, E. R., and Guttorp, P. Time series of continuous proportions. *J. R. Stat. Soc., Ser. B, Methodol. 55*, 1 (1993), 103–116.

[69] Grunwald, G., Hyndman, R., and Tedesco, L. A unified view of linear AR(1) models. *D.P. Monash University, Clayton* (1996).

[70] Guo, Z.-Y. Risk management of Bitcoin futures with GARCH models. *Finance Research Letters 45* (2022).

[71] Hannan, E. J. The asymptotic theory of linear time-series models. *J. Appl. Probab. 10*, 1 (1973), 130–145.

[72] Hanssen, A., and Oigard, T. A. The normal inverse Gaussian distribution: A versatile model for heavy-tailed stochastic processes. In *Proc.-ICASSP IEEE Int. Conf. Acoust. Speech Signal Process.* (2001), vol. 6, IEEE, pp. 3985–3988.

[73] Hassan, A. R. Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting. *Biomed. Signal Process. Control. 29* (2016), 22–30.

[74] Hassan, A. R., and Bhuiyan, M. I. H. An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting. *Neurocomputing 219* (2017), 76–87.

[75] Hassan, A. R., Subasi, A., and Zhang, Y. Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise. *Knowl. Based Syst. 191* (2020).

[76] Heyde, C. C., and Leonenko, N. N. Student processes. *Adv. Appl. Probab. 37* (2005), 342–365.

[77] Horadam, A. F. Gegenbauer polynomials revisited. *Fibonacci Quart 23*, 4 (1985), 294–299.

[78] Horadam, A. F., and Pethe, S. Polynomials associated with Gegenbauer polynomials. *Fibonacci Quart 19*, 5 (1981), 393–398.

[79] Hosking, J. R. Fractional differencing. *Biometrika 68* (1981), 165–176.

[80] Humbert, P. Some extensions of Pincherle's polynomials. *Proc. Edinburgh Math. Soc. 39*, 4 (1920), 21–24.

[81] Ivanov, A. V., Leonenko, N. N., Ruiz-Medina, M. D., and Savich, I. N. Limit theorems for weighted nonlinear transformations of Gaussian stationary processes with singular spectra. *Ann. Probab. 41*, 2 (2013), 1088–1114.

[82] Jensen, M. B., and Lunde, A. The NIG-S&ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econom. J. 4*, 2 (2001), 319–342.

[83] Jiang, Y. An exponential-squared estimator in the autoregressive model with heavy-tailed errors. *Stat. Interface 9* (2016), 233–238.

[84] Jorgensen, B. Statistical properties of the generalized inverse Gaussian distribution. Lecture Notes in Statistics. *Springer-Verlag, New York 9* (1982).

[85] Jose, K. K., and Abraham, B. A count model based on Mittag-Leffler interarrival times. *Statistica 71*, 4 (2011), 501–514.

[86] Jovan, M., and Ahčan, A. Default prediction with the Merton-type structural model based on the NIG Lévy process. *J. Comput. Appl. Math. 311* (2017), 414–422.

[87] Kagan, Y. Y. Correlations of earthquake focal mechanism. *Geophys. J. Int. 110* (1992), 305–320.

[88] Kalemanova, A., Schmid, B., and Werner, R. The normal inverse Gaussian distribution for synthetic CDO pricing. *J. Deriv. 14*, 3 (2007), 80–94.

[89] Karlis, D. An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Stat. Probab. Lett. 57* (2002), 43–52.

[90] Karlis, D., and Lillestöl, J. Bayesian estimation of NIG models via Markov chain Monte Carlo methods. *Appl. Stoch. Models Bus. Ind. 20*, 4 (2004), 323–338.

[91] Kiliç, R. Conditional volatility and distribution of exchange rates: GARCH and FIGARCH models with NIG distribution. *Stud. Nonlinear Dyn. Econom. 11*, 3 (2007).

[92] Klebanov, L. B., Maniya, G. M., and Melamed, I. A. A problem of Zolotarev and analogs of infinitely divisible and stable distributions in a scheme for summing a random number of random variables. *Theory Probab. its Appl. 29*, 4 (1985), 791–794.

[93] Kleinbaum, D. G., and Klein, M. *Survival Analysis: Self-Learning Text.* Springer, New York, NY, 2012.

[94] Klimko, L. A., and Nelson, P. I. On conditional least squares estimation for stochastic processes. *Ann. Stat. 6*, 3 (1978), 629–642.

[95] Kucharska, M., and Pielaszkiewicz, J. M. Nig distribution in modelling stock returns with assumption about stochastic volatility: Estimation of parameters and application to var and etl, 2009.

[96] Kumar, A., and Nane, E. On the infinite divisibility of distributions of some inverse subordinators. *Mod. Stoch.: Theory Appl. 5* (2018), 509–519.

[97] Kumar, A., and Vellaisamy, P. Inverse tempered stable subordinators. *Stat. Probab. Lett. 103* (2015), 134–141.

[98] Lawrance, A. J. Some autoregressive models for point processes. *In Colloquia Mathematica Societatis Janos Bolyai 24* (1978), 257–275.

[99] Lawrance, A. J. Directionality and reversibility in time series. *International Statistical Review / Revue Internationale de Statistique 59*, 1 (1991), 67–79.

[100] Lekshmi, V. S., and Jose, K. K. An autoregressive process with geometric $\alpha-$ Laplace marginals. *Statistical Papers 45* (2004), 337–350.

[101] Lekshmi, V. S., and Jose, K. K. Autoregressive processes with Pakes and geometric Pakes generalized Linnik marginals. *Stat. Probab. Lett. 76* (2006), 318–326.

[102] Li, J., Liang, W., He, S., and Wu, X. Empirical likelihood for the smoothed LAD estimator in infinite variance autoregressive models. *Statist. Probab. Lett. 80* (2010), 1420–1430.

[103] Ling, S. Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. R. Stat. Soc. Ser. B 67* (2005), 381–393.

[104] Liu, C. H., and Rubin, D. B. ML estimation of t-distribution using EM and its extensions, ECM and ECME. *Statist. Sinica. 5* (1995), 19–39.

[105] Liu, T., Zhang, P., Dai, W.-S., and Xie, M. An intermediate distribution between Gaussian and Cauchy distributions. *Phys. A: Stat. Mech. 391* (2012), 5411–5421.

[106] Maleki, M., Mahmoudi, M. R., Wraith, D., and Pho, K.-H. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease 37* (2020), 101742.

[107] Massey(Jr.), F. J. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc. 46* (1951), 68–78.

[108] McLachlan, G. J., and Krishnan, T. *The EM Algorithm and Extensions, 2ed.* John Wiley & Sons, 2007.

[109] Meng, X. L., and Rubin, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 80* (1993), 267–278.

[110] Milovanovic, G. V., and Djordjevic, G. B. On some properties of Humbert's polynomials. *Fibonacci Quart. 25* (1987), 356–360.

[111] Mostafa, F., Saha, P., Islam, M. R., and Nguyen, N. GJR-GARCH volatility modeling under NIG and ANN for predicting top cryptocurrencies. *J. Risk Financial Manag. 14*, 9 (2021).

[112] Nduka, U. C. EM-based algorithms for autoregressive models with t-distributed innovations. *Commun. Statist. Simul. Comput. 47* (2018), 206–228.

[113] Nelson(Jr.), H. L., and Granger, C. Experience with using the Box-Cox transformation when forecasting economic time series. *J. Econom. 10*, 1 (1979), 57–69.

[114] OHagan, A., Murphy, T. B., and et. al., E. C. G. Clustering with the multivariate normal inverse Gaussian distribution. *Comput. Statist. Data Anal. 93* (2016), 18–30.

[115] Oigard, T. A., Hanssen, A., Hansen, R. E., and Godtliebsen, F. EM-estimation and modeling of heavy-tailed processes with the multivariate normal inverse Gaussian distribution. *Signal Process. 85* (2005), 1655–1673.

[116] Omeya, E., Gulcka, S. V., and Vesilo, R. Semi-heavy tails. *Lith. Math. J. 58* (2018), 480–499.

[117] Pincherle, S. Una nuova estensione delle funzioni sferiche. *Memorie della Accademia Reale di Bologna (in Italian)* (1891), 337–369.

[118] Rachev, S. T. *Handbook of Heavy Tailed Distributions in Finance: Hand-books in Finance*. Amsterdam, The Netherlands: Elsevier, 2003.

[119] Reisen, V., Rodrigues, A., and Palma, W. Estimation of seasonal fractionally integrated processes. *Comput. Statist. Data Anal. 50*, 2 (2006), 568–582.

[120] Ridout, M. S. Generating random numbers from a distribution specified by its Laplace transform. *Statistics and Computing 19*, 4 (2009), 439–450.

[121] Robinson, P. M. Semiparametric analysis of long memory time series. *Ann. Stat. 22* (1994), 515–539.

[122] Rohatgi, V. K., and Saleh, A. M. E. *An introduction to probability and statistics*. John Wiley & Sons, 2015.

[123] Rosiński, J. Tempering stable processes. *Stochastic Process. Appl. 117* (2007), 677–707.

[124] Rydberg, T. H. The normal inverse Gaussian Lévy process: simulation and approximation. *Commun. Stat. Stoch. Model. 13*, 4 (1997), 887–910.

[125] Sabzikar, F., McLeod, A. I., and Meerschaert, M. M. Parameter estimation for ARTFIMA time series. *J. Stat. Plan. Inference 200* (2019), 129–145.

[126] Sadreazami, H., Ahmad, M. O., and Swamy, M. N. S. Multiplicative watermark decoder in contourlet domain using the normal inverse Gaussian distribution. *IEEE Trans. Multimed. 18*, 2 (2016), 196–207.

[127] Schiff, J. *The Laplace Transform: Theory and Applications*. Springer, New York, 1999.

[128] Schilling, R. L., Song, R., and Vondracek, Z., Eds. *Bernstein Functions: Theory and Applications*. De Gruyter, Berlin, Boston, 2012.

[129] Sim, C. H. First-order autoregressive models for gamma and exponential processes. *J. Appl. Prob. 27* (1990), 325–332.

[130] Slutsky, E. The summation of random causes as the source of cyclic processes. *Econometrica 5* (1937), 105–146.

[131] Srivastava, H., and Manocha, H. *Treatise on generating functions*. New York: John Wiley & Sons., 1984.

[132] Steutel, F. W., Kent, J. T., Bondesson, L., and Barndorff-Nielsen, O. Infinite divisibility in theory and practice [with discussion and reply]. *Scandinavian Journal of Statistics 6*, 2 (1979), 57–64.

[133] Tang, L., Zhou, Z., and Wu, C. Efficient estimation and variable selection for infinite variance autoregressive models. *J. Appl. Math. Comput. 40* (2012), 399–413.

[134] Tarami, B., and Pourahmadi, M. Multi-variate t autoregressions: Innovations, prediction variances and exact likelihood equations. *J. Time Ser. Anal. 24* (2003), 739–754.

[135] Thawonmas, R., and Damrongthammasakul, S. Sales forecasting in fashion retail using ARIMA models. *Int. J. Adv. Comput. Sci. Appl. 10*, 1 (2019), 77–81.

[136] Tiku, M. L., Wong, W. K., Vaughan, D. C., and Bian, G. Time series models in non-normal situations: Symmetric innovations. *J. Time Ser. Anal. 21* (2000), 571–596.

[137] Torricelli, L., L., C., and A. Tempered positive Linnik processes and their representations. *Electron. J. Stat. 16*, 2 (2022), 6313–6347.

[138] Trindade, A. A., Zhu, Y., and Andrews, B. Time series models with asymmetric Laplace innovations. *J. Stat. Comput. Simul. 80* (2010), 1317–1333.

[139] Tsay, R. S. *Analysis of Financial Time Series*, vol. 2nd ed. 2005.

[140] Weiss, J., Bernardara, P., Andreewsky, M., and Benoit, M. Seasonal autoregressive modeling of a skew storm surge series. *Ocean Model(Oxf) 47* (2012), 41–54.

[141] Weron, R. *Computationally intensive value at risk calculations*. Springer, Berlin, 2004.

[142] Weron, R., Weron, K., and Weron, A. A conditionally exponential decay approach to scaling in finance. *Physica A 264*, 3-4 (1999), 551–561.

[143] Whittle, P. *Hypothesis testing in time series analysis*. Almqvist and Wiksells boktr, Uppsala, 1951.

[144] Wilhelmsson, A. Value at risk with time varying variance, skewness and kurtosis—the NIG-ACD model. *Econom. J. 12*, 1 (2009), 82–104.

[145] Wold, H. *A Study in the Analysis of Stationary Time Series*. Almgrist & Wiksell, Stockholm, 1938.

[146] Woodward, W., Gray, H., and Elliott, A. *Applied Time Series Analysis*. CRC Press, 2011.

[147] Wyłomańska, A. How to identify the proper model. *Acta Phys. Pol. B 43*, 5 (2012), 1241–1253.

[148] Yu, J., and Knight, J. L. Empirical characteristic function in time series estimation. *Econom. Theory 18* (2002), 691–721.

[149] Yule, G. Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. *R. Stat. Soc 89* (1926), 1–64.

[150] Zhang, X., and Jing, X. Image denoising in contourlet domain based on a normal inverse Gaussian prior. *Digit. Signal Process. 20*, 5 (2010), 1439–1446.

[151] Zheng, M., and Karniadakis, G. Numerical methods for SPDEs with tempered stable processes. *SIAM J. Sci. Comput. 37*, 3 (2015).

[152] Zou, Q. Z., Dou, W. Y., and Li, J. L. Study on distribution of oil price return based on NIG distribution. In *2010 International Conference on Management and Service Science* (2010), pp. 1–5.