# UNIVERSITY OF NAIROBI

# Self-Exciting Threshold Autoregressive Modelling of the NSE 20 Share Index Using the Bayesian Approach

By

**Jacinta Wayua Muindi**

**I56/40894/2021**

A Thesis Submitted to the Department of Mathematics for Examination for Partial Fulfilment of the Requirements for Reward of a Degree of Master of Science in Mathematical Statistics.

Department of Mathematics

University of Nairobi

November 2023

# Self-Exciting Threshold Autoregressive Modelling of the NSE 20 Share Index Using the Bayesian Approach

**Research Report in Mathematics, Number XX, 2023**

Jacinta Wayua Muindi

Department of Mathematics
Faculty of Science and Technology
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the Department of Mathematics in partial fulfilment for a degree in Master of Science in Mathematical Statistics

Submitted to: The Department of Mathematics, University of Nairobi, Kenya

# Abstract

This study is on Self-Exciting Threshold Autoregressive (SETAR) modeling of the NSE 20 Share Index using the Bayesian approach. The objectives of the study are to analyze the properties of the NSE 20 Share Index data, to determine the estimates of SETAR model parameters using the Bayesian approach, to forecast the NSE 20 Share Index for the next 12 months using the fitted model, and to compare the forecasting performance of the Bayesian SETAR with the frequentist SETAR and ARIMA model. A Bayesian SETAR model is developed to model the NSE 20 Share Index. MCMC techniques, that is, Gibbs sampling and the Metropolis-Hastings Algorithm, are used to estimate the model parameters. SETAR (2;4,4) model is fitted and used to forecast the NSE 20 Share Index for the next 12 months. The model's forecasting precision is compared to that of a SETAR model that employs the Frequentist approach and to ARIMA. The findings revealed a downward trajectory in the NSE 20 Share Index until April 2024, followed by a gradual upward trend. Additionally, in comparison, the Bayesian SETAR model outperformed its frequentist counterpart in forecasting accuracy. While the ARIMA model performed better compared to the Bayesian SETAR for a shorter forecasting horizon, Bayesian SETAR performed better for a longer forecasting horizon.

**Master Thesis in Mathematics at the University of Nairobi, Kenya.**
**ISSN 2410-1397: Research Report in Mathematics**

**DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya**

# Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

05/12/2023

Signature Date

JACINTA WAYUA MUINDI
Reg No. I56/40894/2021

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Dr. George Muhua
Department of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: muhuageorge@gmail.com

05/12/2023

Signature Date

Dr. Ronald Wanyonyi
Department of Mathematics
Egerton University,
Box 536, 20115 Egerton, Kenya.
E-mail: ronaldwaliaula@gmail.com

08/12/2023

Signature Date

# Dedication

To my incredible family— my dad, Maurice Muindi, and my siblings Richard, Mbithe, Mumo and Mbatha—your unwavering support and enduring love have been my driving force. To my late mum, your enduring spirit remains my inspiration. This research is dedicated to the memory of your remarkable love and belief in me.

# Acknowledgments

First, I thank God, whose guidance, grace, and strength have been the cornerstone of my journey.

I am immensely thankful to Dr. Ronald Wanyonyi and Dr. George Muhua for their unwavering support, invaluable guidance, and continuous encouragement. Their expertise and insights have significantly shaped the direction and depth of this research.

A special acknowledgment is extended to the administrative support from the Department of Mathematics, especially to the Chairman, Prof. Stephen Luketero, for his constant support and willingness to assist, which have been instrumental in overcoming challenges.

To my family, whose boundless love, understanding, and unwavering support have been my source of strength—thank you for standing by me through the challenges and triumphs of this journey.

I am also grateful to a remarkable group of friends from the Department of Mathematics, with whom I have shared this Master's journey. Your companionship, shared experiences, and collaborative efforts have made this journey meaningful and enjoyable.

This research would not have been possible without these individuals' collective efforts and support; for that, I am genuinely grateful.

Jacinta Wayua Muindi

Nairobi, 2023.

# Contents

# Acronyms

| | |
|---|---|
| SETAR | Self-Exciting Threshold Autoregressive |
| NSE | Nairobi Securities Exchange |
| NSE20 | NSE 20 Share Index |
| pdf | Probability Density Function |
| AR | Autoregressive |
| MA | Moving Average |
| ARIMA | Autoregressive Integrated Moving Average |

# 1 Introduction

## 1.1 Background Information

The analysis and interpretation of time series data hold paramount significance across different fields, including economics, finance, and engineering, among other fields. This kind of data, characterized by sequential observations over time, sometimes exhibits complex patterns and trends that some commonly used models cannot capture. One key challenge of modeling this data is capturing the volatility and nonlinear dynamics, which can significantly impact the future behavior of the series. Commonly used approaches such as linear autoregressive (AR) and simple moving average (MA) models often fail to capture the intricate nature of volatility, characterized by time-varying patterns, regime shifts, and nonlinear relationships.This limitation calls for the development of more sophisticated and flexible models that can effectively capture the complexity of time series data.

Over the past years, nonlinear time series models have been introduced, with many researchers aiming at capturing the dynamics and complexities of time series data, especially economic and financial data. One class of nonlinear models introduced to capture the nonlinear dynamics in time series data and as an evolution from the linear models is the Threshold Autoregressive (TAR) models. The TAR models were first introduced by Tong (1978). This class of models incorporates regime shifts in time series data, dividing the data into distinct regimes. Each regime has a specific set of parameters that capture the underlying dynamics. A threshold variable governs the regime transitions, allowing the model to capture changes in market conditions. Self-Exciting Threshold Autoregressive (SETAR) is an extension of Tong's TAR, which was proposed to incorporate the concept of self-exciting behavior (Tong and Lim, 1980). These models consider the past behavior of a series in determining the regime shifts, allowing for the number of regimes to be specified using the data instead of being pre-specified. This feature makes SETAR models a more promising choice in modeling the nonlinear dynamics in time series data like financial data, where the thresholds and number of regimes may change over time.

When using SETAR models, there are typically many parameters to be estimated due to the different regimes, and the estimation process can be challenging, especially when the researcher is dealing with high-dimensional or noisy financial data. This requires sophisticated yet flexible and simple approaches, such as the Bayesian approach, to be incorporated into the Self-Exciting Threshold Autoregressive modeling. In this study, the strengths of SETAR models in capturing nonlinear dynamics are combined with the Bayesian approach, which provides a flexible and robust way of parameter estimation that also incorporates uncertainty.

## 1.2 Problem Statement

Accurate modeling and forecasting of time series is critical in different areas. For example, in the financial markets, it directly impacts risk management strategies, investment decisions, and portfolio performance. While linear models like AR, MA, and ARIMA have been used before in cases where volatility is evident, they struggle to capture the threshold behavior and nonlinearity, which often characterize financial and economic data. Studies that have been conducted to explore, model, and forecast Nairobi Securities Exchange (NSE) data patterns mostly use ARIMA models despite the unique characteristics of the NSE data, including nonlinearity and high volatility. This calls for models that can capture these complex patterns. This study aims to bridge this gap by developing a SETAR model suitable for non-linear data, essential for enhancing forecasting accuracy.

Additionally, while SETAR models have been used in different studies, parameter estimation is usually through the frequentist approach. The frequentist approach in model parameter estimation provides point estimates, which may not fully capture the uncertainty inherent in the parameters. Further, the approach operates within the confines of observed data, lacking a mechanism to incorporate prior knowledge. This is a limitation, particularly in scenarios with limited data, where reliance solely on observed information may lead to less reliable estimates. Therefore, the adoption of the Bayesian approach in this study will provide a substantial improvement. By embracing Bayesian methods, this study capitalizes on seamlessly integrating prior information into parameter estimation, thereby overcoming the constraints posed by limited data. This integration also allows for a more holistic and informed estimation process, offering probability distributions for parameters comprehensively representing uncertainty. The full potential of the Bayesian approach is yet to be exploited in the context of SETAR modeling.

## 1.3 Study Objectives

### 1.3.1 General Objective

To model the NSE 20 Share Index using Self-Exciting Threshold Autoregressive (SETAR) model and the Bayesian approach of parameter estimation.

### 1.3.2 Specific Objectives

i. To analyze the dynamics of NSE 20 Share Index data through correlation analyses as well as linearity and stationarity tests.

ii. To determine the estimates of SETAR model parameters using the Bayesian framework.

iii. To forecast the NSE 20 Share Index for the next 12 months using the model.

iv. To compare the forecasting performance of the Bayesian SETAR with the performance of the common SETAR and ARIMA models.

## 1.4 Research Questions

The study will investigate the following questions:

i. What are the key characteristics of the NSE 20 data? For example, does the data portray linearity or stationarity, and are there significant correlations among lagged values?

ii. How can Bayesian methods be applied for parameter estimation in the SETAR model, and what are the parameter estimates in the SETAR model?

iii. What are the projected trends and expected fluctuations in the NSE 20 Share Index over the next 12 months, as determined by the forecasting capabilities of the employed model?

iv. How well does the Bayesian SETAR model perform in forecasting the NSE 20 Share Index compared to the commonly used ARIMA models and SETAR models fitted with the frequentist approach?

## 1.5 Justification of the Study

The application of a SETAR model in the context of the NSE 20 Share Index could offer a robust framework for modeling the volatility and nonlinearity usually inherent in financial time series data. By combining the flexibility of Bayesian methods with the adaptability of SETAR models, this study aims to provide more robust, contextually enriched parameter estimates, enhancing the accuracy and depth of understanding within the model's dynamics. This, in turn, has the potential to lead to a more robust and accurate model that will help with decision-making and risk management in the financial markets.

The study will lead to an enhanced comprehension of the market and, hence, better predictability. The regimes and thresholds that characterize the behavior of the NSE 20 Share Index will be estimated, giving insights into the market conditions and price movements. Understanding these regimes will lead to better market predictability, providing information crucial to financial institutions, policymakers, and investors in making informed decisions. By improving forecasting accuracy and risk management in financial markets, the research can indirectly contribute to stability and economic growth since better financial decision-making and risk management directly impact investments, businesses, and overall economic prosperity.

This study will contribute to academic knowledge by extending the existing knowledge on time series and econometrics analysis. It will enrich the literature on Bayesian inference and non-linear modeling approaches, providing valuable insights for academics and researchers interested in exploring these areas. The application of Bayesian methods in the SETAR model contributes to advancing Bayesian statistics in finance. Also, the study's findings can stimulate further research in Bayesian analysis and non-linear modeling in finance.

# 2 Literature Review

## 2.1 Introduction

Research about nonlinear time series models has been conducted where they have been compared to other models, and they have been applied in many areas as well. Below are some of the studies that have been conducted on Threshold Autoregressive models and the Bayesian approach, and some of the comparative studies that have been conducted.

## 2.2 Comparative Studies of SETAR with Other Models

Gibson and Nur (2011) conducted a comparative study to investigate the efficiency of threshold autoregressive (TAR) models in modeling in the domain of finance. Using weekly Nikkei 225 Index data from January 2000 up to September 2010, the authors fitted three models: SETAR model, STAR model, and AR model. The Threshold Autoregressive Models provided improved fit and forecasting performance compared to the linear model. The conclusion drawn was that threshold models, SETAR in particular, were an improvement on linear models in regards to capturing volatility and reflecting the overall process.

Aydin and Güneri (2015) carried out a comparative study to investigate the prediction performance of ARM, SETAR, AR, and hybrid models AR&SETAR, AR&AAR, SETAR&AR, AAR&SETAR, AAR&AR, and SETAR&AAR. Two datasets were used for the study, that is Turkey's monthly export volume index numbers for January 1997 to December 2014 and the monthly domestic producer price index for January 2006 to December 2014 period. The two datasets were divided into training sets, which were used to train the models, and the forecast sets, which were used to evaluate the models' performance. RMSE, MSE, MAPE, and MAE were calculated as performance indicators. The findings of the study were that the AAR-SETAR hybrid model performed the best, supporting the idea that hybrid models perform well in forecasting problems in time series.

Boero and Lampis (2017) conducted a study to investigate the accuracy of SETAR models in forecasting compared to seasonal ARMA and linear autoregressive models. The authors used the monthly unadjusted Industrial Production Index (IPI) data from January 1975 to December 2011 for four EU countries: the UK, Spain, Italy, and France. Data subsets from January 1975 to December 2005 were used to estimate the models, while the rest were used as forecasting samples. For every 12 months, the models were fully re-specified, and they were also re-estimated for every new monthly observation added to the sample. 1, 3, 6, and 12-step-ahead forecasts were calculated. While the results did not show that the SETAR model had a homogeneous forecast superiority across all horizons, sample periods, and countries, the point forecast findings suggested that SETAR had better forecast performance, especially in the one-step-ahead forecasts, which was linked to the re-specification. The results also showed that the performance tended to deteriorate for longer horizons for both linear and non-linear models.

Firat (2017) investigated SETAR performance in currency modeling. The author used Euro-US Dollar parity, Euro-Turkish Lira parity, and US Dollar-Turkish parity data for the study. The EUR/USD parity data ran from December 1999 to February 2015. The EUR/TRY parity data and USD/TRY parity data run from May 2010 to February 2015. This data was used to fit SETAR, linear, NNETTs, AAR, and LSTAR, and a comparison was made amongst these models using Akaike Information Criteria (AIC) values. The SETAR model showed superior performance compared to the other models in regard to the relevant parities.

Oyewale et al. (2017) carried out a study to assess the forecasting accuracy of two time series models - the linear Seasonal ARIMA model and the non-linear SETAR model. Nigerian inflation rate data from 1993 to 2013 was used for the study. The authors then compared the in-sample and out-of-sample forecasting performance of the SARIMA and SETAR models using various error measures, including MAPE, MSE, MAE, and Theil's U inequality coefficient. The findings showed that the SETAR model made better forecasts than the linear SARIMA model, both for in-sample and out-of-sample predictions.

## 2.3 The Bayesian Approach of Parameter Estimation

Pan et al. (2017) proposed a Bayesian approach to analyzing possible threshold values in a TAR model with several possible thresholds. Instead of assuming a fixed number of regimes, the authors introduced a Bayesian stochastic search selection method to pick out the threshold values in the model. The primary concept behind the approach involved introducing a series of random variables that took the value 1 at positions that were associated with threshold values and value 0 elsewhere. Within the Bayesian framework, the authors estimated the threshold-dependent parameters that were unknown by utilizing their posterior distributions through maximum a posteriori (MAP) estimation. To make an estimate of the threshold-dependent variables and estimate of the remaining model parameters, the authors employed a hybrid MCMC method, combining Metropolis–Hastings (M–H) algorithm and Gibbs sampler. The authors then applied the methodology to analyze sunspot number data for the period 1700 to 1979. The authors concluded that the approach was effective and feasible in practice.

Agiwal and Kumar (2020) proposed a Bayesian approach to analyzing TAR models with multiple regimes and several structural breaks. The authors obtained the full conditional posterior distributions for all the model parameters, assuming suitable prior information. The threshold and breakpoint variables did not have standard form distributions, so they used the Gibbs sampler with the Metropolis-Hastings algorithm to compute the posterior distributions. The authors carried out a simulation study to demonstrate the performance of the Bayesian estimators under different loss functions. The results showed that the estimates were close to the actual values, and most parameters had minimum posterior standard deviation under the absolute loss function. They also applied the suggested methodology to an annual real tree ring data from China for the years 1079 to 2009 to determine the breakpoint and estimate the model parameters. The results showed that the Bayesian approach can appropriately determine the breakpoints and estimate the associated parameters.

Ojo (2021) conducted a study to analyze prior sensitivity to see how the posterior estimates were sensitive to changes in prior assumptions using the Bayesian threshold autoregressive model. The inflation rate data for Nigeria from 1960 to 2019 was used for the study. The author also tried different values of the delay parameter $d$. The results were that as the prior $c \to \infty$, infinite posterior estimates were obtained for the model parameters. However, assigning a small value to c gave an estimation that was more accurate. The delay parameter $d$ assigned the highest probability to $d = 1$ when the prior $c$ was set at 0.05 or 1. Conversely, when $c$ was set to 100 (noninformative), $d$ assigned the highest probability to $d = 4$. The conclusion drawn was that posterior estimates were sensitive to changes in the prior.

## 2.4 Recent Applications of SETAR Models

Tobechukwu et al. (2022) conducted a study to model daily Nigerian COVID-19-confirmed cases using the SETAR model. The data used for the study was Covid-19 daily confirmed cases from January 2020 to September 2022. The authors identified SETAR (2, 4, 1) model as the best-fitted model for the data based on metrics like the MSE and Akaike Information Criteria (AIC). The specified SETAR model was used to get one-month period predictions of daily confirmed COVID-19 cases in Nigeria. The results showed that the cumulative number of daily confirmed COVID-19 cases was expected to rise from around 281,526 cases in 2022 to around 312,776 cases in 2023 based on the forecasts from the SETAR model.

## 2.5 Gaps Identified

Despite the promising results in the above studies in regard to the application of SETAR models and the Bayesian framework, the application of the Bayesian approach to parameter estimation in SETAR models is yet to be fully explored. In most studies, the parameter estimation is done using the frequentist approach, which is limited because it does not provide a mechanism to incorporate prior knowledge that the researcher may have before collecting or seeing the data. Also, the estimates in the frequentist approach are fixed estimates, not accommodating for uncertainty. These limitations can be addressed using the Bayesian approach, which is promising according to the studies above. Additionally, the SETAR model is yet to be investigated in modelling the NSE 20 Share index. To close these gaps, Bayesian SETAR will be used to model and forecast the NSE 20 Share index.

# 3 Research Methodology

## 3.1 Introduction

This study focuses on TAR models, particularly SETAR models, which are non-linear models first proposed by Tong (1978). This class of models incorporates regime shifts in time series data, dividing the data into distinct regimes. The methodology of this study centers on the application of Bayesian Inference to SETAR models for the purpose of modeling the NSE 20 Share Index. This chapter will focus on the methodology of the study.

## 3.2 Threshold Autoregressive (TAR) Models

Within the framework of threshold time series models, it is assumed that the process has different regimes, determined by a threshold. The basic idea is that the process will behave differently when a variable's values go beyond a certain threshold, meaning that different models apply when the values are below and above the threshold.

Suppose a series $\{y_t\}$ is observed at discrete time points t. TAR(p), a TAR model of order p and two regimes can be written as

$$y_t = \begin{cases} \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + a_t^{(1)}, & \text{if } z_t \leq r \\ \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + a_t^{(2)}, & \text{if } z_t > r \end{cases} \tag{1}$$

where $z_t$ is the threshold variable, and $a_t^{(j)}$ are independent Gaussian white noise processes with mean zero and variance $\sigma_j^2$, with j = 1,2. $\phi_i$ and $\theta_j$ are real-valued parameters and $r$ is the threshold.

The two-regime TAR model above can be expanded to incorporate numerous regimes and can be formulated as:

$$y_t = \phi^j + \sum_{i=1}^{p} \theta_i^j y_{t-i} + a_t^{(j)} \tag{2}$$

where $j \in \{1,...,k\}$ is an indicator of regime switching. Each regime has it's set of coefficients $\{\theta_i^j\}'s$ and different regimes can also have different orders p.

### 3.2.1 SETAR Models

SETAR model is a type of TAR models. Consider the following TAR model with two regimes:

$$y_t = \begin{cases} \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + a_t^{(1)}, & \text{if } z_t \leq r \\ \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + a_t^{(2)}, & \text{if } z_t > r \end{cases} \tag{3}$$

where r is the threshold parameter.
If the $z_t$ value is replaced with previous values of the time series $y_t$ as below, the TAR model is then referred to as SETAR model.

$$y_t = \begin{cases} \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + a_t^{(1)}, & \text{if } y_{t-d} \leq r \\ \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + a_t^{(2)}, & \text{if } y_{t-d} > r \end{cases} \tag{4}$$

The switch of regimes is therefore influenced by the past values of $y_t$, that is, $y_{t-d}$, where $d$ represents the delay parameter.

The SETAR model can be generalised to include multiple regimes.
Let $n > 1$ be a positive finite integer and $\{r_j | j = 0,...,n\}$ a real number sequence with $r_0 = -\infty < r_1 < r_2 < ... < r_{n-1} < r_n = \infty$ where $-\infty$ and $\infty$ are regarded as real numbers. Then a time series $y_t$ is a $m$-regime SETAR(p) model if it satisfies

$$y_t = \begin{cases} \theta_{0,1} + \sum_{i=1}^{p} \theta_{i,1} y_{t-i} + a_t^{(1)}, & \text{if } y_{t-d} \leq r_1 \\ \theta_{0,2} + \sum_{i=1}^{p} \theta_{i,2} y_{t-i} + a_t^{(2)}, & \text{if } r_1 < y_{t-d} \leq r_2 \\ \dots & \\ \theta_{0,m} + \sum_{i=1}^{p} \theta_{i,m} y_{t-i} + a_t^{(m)}, & \text{if } r_{m-1} < y_{t-d} \end{cases} \tag{5}$$

where $y_{t-d}$ is the threshold variable, $d > 0$ is the delay parameter, $a_t^{(j)}$ are independent Gaussian white noise processes with mean zero and variance $\sigma_j^2$, $\sigma_i$'s are positive real numbers, and $\theta_{i,j}$ are real parameters.
In a compact form, equation 5 can be written as

$$y_t = \theta_{0,j} + \theta_{1,j} y_{t-1} + ... + \theta_{p,j} y_{t-p} + a_t^{(j)} \quad if \quad r_{j-1} < y_{t-d} < r_j \tag{6}$$

with $j = 1,...,n$. Each regime is represented by a AR(p) model, governed by a different set of coefficients $\theta^{(j)}$.

Despite the simplicity that SETAR models provide in handling non-linear time series processes, one key challenge in using these models is the existence of many free parameters that need to be chosen and approximated in building the models. These parameters include the the number of regimes/ thresholds, threshold values, orders of the AR models, and the model coefficients. The Bayesian approach provides a promising way of estimating the model parameters. It allows the incorporation of prior information that the researcher may have about variables before seeing the data, allowing sequential learning. Additionally, the approach is a natural way of making predictions since we take account of all parameters and the model uncertainty.

## 3.3 Bayesian Estimation of the Model Parameters

In the Bayesian statistical perspective, parameters are viewed as random variables to account for uncertainty in their values. Bayesian analysis involves specifying a likelihood, which is the conditional density of data given the parameters, together with a prior distribution for the parameters derived from past knowledge or beliefs. The joint density of data and parameters is obtained by multiplying the prior and likelihood. To get the marginal density of the data, the parameters have to be integrated out. The posterior density, which represents the parameters given the data, is then derived by dividing the joint density by the marginal density. This posterior contains all the information about parameter values and serves as the foundation for Bayesian inference. Different point estimators like median, mean, or mode are derived.

Suppose $m$ observations $y_{1:m}$ are collected of a time series $y_t$. Suppose that each data point, $y_t$, is associated with a probability distribution which can be expressed as a function of a parameter or many parameters $\phi$ so that the relationship between $y_t$ and $\phi$ is described by a pdf $p(y_t|\phi)$. When $p(y_t|\phi)$ is considered as a function of $\phi$ instead of $y_t$, it is referred to as the *likelihood function*. By applying Bayes' theorem, the posterior pdf $p(\phi|y_t)$ can be derived, that is, the posterior pdf of $\phi$ given $y_t$, by multiplying the likelihood with the prior density, $p(\phi)$. That is,

$$p(\phi|y_t) = \frac{p(\phi)p(y_t|\phi)}{p(y_t)} \tag{7}$$

where $p(y_t) = \int p(\phi)p(y_t|\phi)d\phi$

$p(y_t)$ defines what is known as predictive density function. The prior distribution provides a means to include researcher's initial beliefs or assumptions regarding $\phi$, and Bayes' theorem enables the revision and updating of these assumptions once the data is observed.

Bayes' theorem can also be applied sequentially as follows:

- Before gathering any data, the initial beliefs on $\phi$ can be expressed in the probabilistic form $p(\phi)$.
- After collecting the first observation $y_1$ at time $t = 1$, suppose that $p(\phi|y_1)$ is obtained through Bayes' theorem.
- When $y_2$ is observed, $p(\phi|y_{1:2})$ can be obtained as $p(\phi|y_{1:2}) \propto p(\phi)p(y_{1:2}|\phi)$ using Bayes' theorem.
- If $y_1$ and $y_2$ are conditionally independent with respect to $\phi$, $p(\phi|y_{1:2})$ can be expressed as $p(\phi|y_{1:2}) \propto p(\phi|y_1)p(\phi|y_2)$, essentially turning the posterior of $\phi$ given $y_1$ into a prior distribution before observing $y_2$.
- Likewise, in a sequential manner, $p(\phi|y_{1:m})$ can be obtained if all the values are independent. However, the values are not typically independent in time series analysis. For instance, it is commonly assumed is that the value at time $t$ relies solely on $\phi$ and the value observed at $t-1$. In this scenario, we have

$$p(\phi|y_{1:m}) \propto p(\phi)p(y_1|\phi)\prod_{t=2}^{m} p(y_t|y_{t-1},\phi) \tag{8}$$

For example, consider an AR(1) process. Let the model parameters be $\phi = (\theta, v)'$. The conditional likelihood for every $t > 1$ is $p(y_t|y_{t-1},\phi) = N(y_t|\theta y_{t-1}, v)$. It can also be proved that $y_1 \sim N(0, v/1-\theta^2)$ for a stationary process. The likelihood is hence given by

$$p(y_{1:m|\phi}) = \frac{(1-\theta^2)^{\frac{1}{2}}}{(2\pi v)^{\frac{m}{2}}} exp\left\{-\frac{Q^*(\theta)}{2v}\right\} \tag{9}$$

where

$$Q^*(\theta) = y_1^2(1-\theta^2) + \sum_{t=2}^{m}(y_t - \theta y_{t-1})^2 \tag{10}$$

Using Bayes' rule, the posterior density is

$$p(\theta|y_{1:m}) \propto p(\theta)\frac{(1-\theta^2)^{\frac{1}{2}}}{(2\pi v)^{\frac{m}{2}}} exp - \frac{Q^*(\theta)}{2v} \tag{11}$$

### 3.3.1 Obtaining Priors

For this study, conjugate priors will be used. Just as Chen and Lee (1995), the natural conjugate priors are selected as follows; $\theta_1$ and $\theta_2$ are taken as independent and normally distributed as $N(\theta_{0i}, M_i^{-1})$, while $\sigma_1^2$ and $\sigma_2^2$ are taken as independent with inverse gamma distribution *inverse gamma* $(\nu_i/2, \nu_i\lambda_i/2)$. The hyper-parameters d and r are assumed to be known. Next, following Geweke and Terui (1993) approach, *r* is presumed to have a uniform distribution on $(\alpha, \beta)$ while *d* is presumed to take a discrete uniform distribution on $1, 2, ..., D$.

### 3.3.2 Obtaining Posterior Distributions

The primary interest of the analysis is to obtain the marginal posterior distributions of the parameters $\theta_i$'s, $\sigma^2$s, *r*, and *d*. Determining the posterior distribution is frequently a challenging task due to the need for complex numerical integration in high-dimensional spaces. As a result, modern techniques for computing posteriors have emerged, such as the Gibbs Sampler and the Metropolis-Hastings algorithm. These are the techniques that will be used in this study to find the conditional posterior distributions of the unknown parameters.

#### Gibbs Sampler

The Gibbs sampler, which is a Markov Chain Monte Carlo (MCMC) technique, is used to estimate target posterior distributions from conditional distributions. MCMC provides a way to draw samples from the posterior distribution, allowing the approximation of the posterior and making inferences about the parameters and other quantities of interest. The fundamental concept of MCMC involves creating a Markov chain with a stationary distribution, and this distribution is the targeted posterior distribution. Through simulating a Markov chain, convergence is achieved towards the desired posterior distribution over time. Gibbs Sampling is specifically designed to sample from multivariate distributions by sequentially updating each variable while conditioning on the current values of the other variables.

Consider a scenario where the random variable $\phi$ can be broken down into components $\phi = (\phi_1, ..., \phi_r)$ and the conditional densities
$\phi_j|\phi_1, ..., \phi_{j-1}, \phi_{j+1}, ..., \phi_r \sim f_j(\phi_j|\phi_1, ..., \phi_{j-1}, \phi_{j+1}, ..., \phi_r)$ can be simulated for $j = 1, ..., r$.

Then, to sample from the joint density of $(\phi_1, ..., \phi_r)$ using Gibbs sampler, the following algorithm is followed:

1. Given the sample $(\phi_1^{(m)}, ..., \phi_r^{(m)})$, generate
2. $\phi_1^{(m+1)} \sim f_1(\phi_1|\phi_2^{(m)}, \phi_3^{(m)}, ..., \phi_r^{(m)})$,
3. $\phi_2^{(m+1)} \sim f_2(\phi_2|\phi_1^{(m)}, \phi_3^{(m)}, ..., \phi_r^{(m)})$,
   $\vdots$

r. $\phi_r^{(m+1)} \sim f_r(\phi_r|\phi_1^{(m)}, \phi_2^{(m)}, ..., \phi_{r-1}^{(m)})$.

Gibbs Sampling is specifically designed to sample from multivariate distributions by iteratively updating each variable conditioned on the current values of the other variables.

Chen and Lee (1995) derived the conditional posterior distributions of the parameters $\theta_i$'s, $\sigma^2$s, $r$, and $d$ of the SETAR (2;$p_1$,$p_2$) model as follows;
Considering a TAR (2;$p_1$,$p_2$)

$$y_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} y_{t-i} + a_t^{(1)}, & \text{if } y_{t-d} \leq r \\ \phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} y_{t-i} + a_t^{(2)}, & \text{if } y_{t-d} > r \end{cases} \tag{12}$$

Consider $p$ as the maximum among $p_1$ and $p_2$. Assuming that the initial $p$ cases are predetermined, denoted as $(y_1, ..., y_p)$, and letting $\pi_i$ represent the index of the $i^{th}$ smallest observation in $(y_{p+1-d}, ..., y_{n-d})$, the likelihood function, conditioned on the first $p$ values, can be expressed as follows:

$$L(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, r, d|Y) \propto \sigma_1^{-s} \sigma_2^{-(n-p-s)} exp\Bigg\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{s} \left( y_{\pi_i+d} - \phi_0^{(1)} - \sum_{k=1}^{p_1} \phi_k^{(1)} y_{\pi_i+d-k} \right)^2 - \frac{1}{2\sigma_2^2} \sum_{i=s+1}^{n-p} \left( y_{\pi_i+d} - \phi_0^{(2)} - \sum_{k=1}^{p_2} \phi_k^{(2)} y_{\pi_i+d-k} \right)^2 \Bigg\} \tag{13}$$

where $Y = (y_{\pi_1+d}, y_{\pi_2+d}, ..., y_{\pi_{n-p}+d})'$
$\theta^{(1)} = (\theta_0^{(1)}, \theta_1^{(1)}, ..., \theta_{p_1}^{(1)})'$
$\theta^{(2)} = (\theta_0^{(2)}, \theta_1^{(2)}, ..., \theta_{p_2}^{(2)})'$

and $s$ satisfies $y_{\pi_s} < r \leq y_{\pi_{s+1}}$

For the model, the parameters to be estimated are $\theta^{(1)}, \theta^{(2)}, \sigma_1^2, \sigma_2^2, d$ and $r$.

Let $Y_1^* = (y_{\pi_1+d}, y_{\pi_2+d}, ..., y_{\pi_s+d})'$ and $Y_2^* = (y_{\pi_{s+1}+d}, ..., y_{\pi_{n-p}+d})'$ be observations that are generated by regimes I and II, respectively.
Additionally, let
$x_{1,t} = (1, y_{\pi_i+d-1}, ..., y_{\pi_i+d-p_1})'$
$x_{2,t} = (1, y_{\pi_i+d-1}, ..., y_{\pi_i+d-p_2})'$
$X_1^* = (x_{1,\pi_1+d,}, x_{1,\pi_2+d,}, ..., x_{1,\pi_s+d,})'$
$X_2^* = (x_{2,\pi_{s+1}+d,}, x_{2,\pi_{s+2}+d,}, ..., x_{2,\pi_{n-p}+d,})'$
These represents an arranged auto-regression where the first $s$ observations of Y are in the first regime while the other n-p-s cases are in the second regime.

The conditional posterior distributions for the unknown parameters conditioned on all the other parameters were derived as follows;

- The conditional posterior distribution of $\theta_i$ is

$$p(\theta_i|Y, \sigma_1^2, \sigma_2^2, r, d) \sim N(\theta_i^*, V_i^{*-1}) \tag{14}$$

  where for $i \neq j$, the conditional probability of $\theta_i$ is independent of $\theta_j$.
  $\theta_i^* = (\frac{X_i^{*'} X_i^*}{\sigma_i^2} + V_i)^{-1} (\frac{X_i^{*'} X_i^*}{\sigma_i^2} \hat{\theta}_i + V\theta_{0i})$ where $\hat{\theta}_i = (X_i^{*'} X_i^*)^{-1} X_i^{*'} Y_i^*$.

  $V_i^* = \frac{X_i^{*'} X_i^*}{\sigma_i^2} + V_i$

- The conditional posterior distribution of $\sigma_i^2$ is

$$p(\sigma_i^2|Y, \theta_1, \theta_2, r, d) \sim inverse\, gamma(\frac{v_i + n_i}{2}, \frac{v_i \lambda_i + n_i s_i^2}{2}) \tag{15}$$

  where for $i \neq j$, the conditional probability of $\sigma_i^2$ is independent of $\sigma_j^2$.

  $\frac{v_i \lambda_i + n_i s_i^2}{\sigma_i^2} \sim \chi^2_{v_i + n_i}$, where $i$ = 1,2

  $s_i^2 = n_i^{-1} (Y_i^* - \hat{Y}_i)' (Y_i^* - \hat{Y}_i)$, where $\hat{Y}_i = X_i^{*'} \theta_i$

  $n_1 = \sum_{t=1}^{n-p} I_{\{y_{\pi_i} \leq r\}}, \quad n_2 = \sum_{t=1}^{n-p} I_{\{y_{\pi_i} > r\}}$

- For $d$, the conditional posterior probability function is a multinomial distribution whose probability is given by

$$p(d|Y,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,r) = \frac{L(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,r,d|Y)}{\sum_{d=1}^{D} L(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,r,d|Y)} \tag{16}$$

  with d = 1,2,...,D and

  $L(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,r,d|Y) = \frac{exp\{-\sum_{i=1}^{2}(1/2\sigma_i^2)(Y_i^*-X_i^{*'}\theta_i)'(Y_i^*-X_i^{*'}\theta_i)\}}{\sigma_1^{n_1}\sigma_2^{n_2}}$

- For $r$, the conditional posterior probability function is

$$p(r|Y,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,d) \propto \frac{exp\{-\sum_{i=1}^{2}(1/2\sigma_i^2)(Y_i^*-X_i^{*'}\theta_i)'(Y_i^*-X_i^{*'}\theta_i)\}}{\sigma_1^{n_1}\sigma_2^{n_2}} \tag{17}$$

  with $n_1$ and $n_2$ being functions of $r$.

The conditional densities of the unknown parameters have been identified above except for the conditional density of $r$. Therefore, these conditional distributions will be used alongside the Gibbs sampling technique described above to obtain the marginal posterior distributions for the unknown parameters. Regarding $r$, the Metropolis algorithm will be used.

### The Metropolis-Hastings Algorithm

Let the conditional density in equation 17 be denoted by *f(k)*, and as before, the assumption that $k$'s prior distribution is uniform over $(\alpha,\beta)$ is made. Hence, a transition kernel $h(k,k^*)$ with $k^* = \log(k-\alpha)/(\beta-k)$ can be utilized to map $(\alpha,\beta)$ into $(-\infty,\infty)$. The Metropolis algorithm will then work in the following manner;

- Begin with an initial value, $k^{(0)}$ drawn from the prior $U(\alpha,\beta)$ and set the indicator $j$ to 0.
- Utilizing the transition kernel $h(k^{(j)},k^*)$, generate a new point $k^*$.
- With a probability of $p = \min 1, f(k^*)/f(k^{(j)})$, update $k^{(j)}$ to $k^{(j+1)} = k^*$ and remain at $k^{(j)}$ with a probability of $(1-p)$.
- By increasing the indicator, repeat steps 2 and 3 until a stationary distribution is attained.

## 3.4 Test for Linearity

SETAR models were introduced to handle non-linear time series data, and hence, before proceeding with modeling, it is essential to check for non-linearity to ensure the model selected is appropriate. The non-linearity tests that will be used are Tsay's F test and BDS test.

### 3.4.1 Tsay's F Test

Tsay's F test will be utilised to examine the existence of threshold-type non-linearity. In testing for non-linearity, the null hypothesis of linearity is tested against the alternative hypothesis that there exists a threshold model.
Considering a TAR(p) process with two regimes, the test is,

$$H_0: \quad y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \tag{18}$$

$$H_1: \quad y_t = \begin{cases} \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t, & \text{if } y_{t-d} \leq r \\ \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + \varepsilon_t, & \text{if } y_{t-d} > r \end{cases} \tag{19}$$

### 3.4.2 BDS Test for Nonlinearity

The BDS test, introduced in 1987 by Brock, Dechert and Scheinkman as part of chaos theory, stands out as a widely used assessment for detecting nonlinearity. Originally conceived as a nonparametric test to examine independence and identical distribution (iid), it has proven to be effective in identifying both linear and nonlinear patterns. The BDS test investigates the spatial dependence of a time series. The series is represented in a multi-dimensional space (m-space), and the dependence of x is investigated by counting close points, which are data points that are within a distance of 'eps' from each other.

$H_0$: The data is independently and identically distributed (iid).

$H_1$: The residuals exhibit an underlying structure, potentially of a non-linear nature.

## 3.5 Test for Stationarity

It is also important to test for stationarity. Stationarity tests assess whether the time series is stationary, i.e, whether it exhibits stable statistical properties over time, including the mean, variance, and autocorrelation. Many time series models, Autoregressive models included, assume stationarity as it simplifies the analysis and allows for more reliable forecasts. The stationarity tests that will be employed are the ADF Test and the Zivot and Andrews (1992) test.

### 3.5.1 ADF Test

ADF is a commonly utilised statistical test that tests for the existence of unit roots in the data, which are indicators of non-stationarity. The ADF test's null hypothesis suggests that the time series possesses unit roots, rendering it non-stationary, while the alternative hypothesis suggests the absence of unit roots, indicating stationarity.

$H_0 : \rho = 1$ (has unit roots)
$H_1 : \rho < 1$ (no unit roots)
If the p-value obtained from the test is below a selected significance level, the null hypothesis is rejected, indicating stationarity.

### 3.5.2 Zivot-Andrews Unit Root Test

This is a unit root test that was introduced to incorporate an endogenous structural break. The null hypothesis is that the time series represents an integrated process devoid of structural changes. Conversely, the alternative hypothesis proposes that the process is trend-stationary, characterized by a singular break occurring at an unspecified moment in time.

$H_0 : \theta_i = 1$ series has a unit root with drift
$H_1 : \theta_i < 1$ series is stationary with break(s)

# 4 Data Analysis and Results

## 4.1 Introduction

In this chapter, the NSE20 time series data will be explored and the Bayesian SETAR model will be developed. The data will be explored for attributes including stationarity, linearity, and correlation, which will, in turn, lay the groundwork for fitting of a SETAR model. The model, once established, will serve as a predictive tool, offering insights into future trends. Furthermore, the performance of the SETAR model will be evaluated to assess the predictive power of the employed methodologies.

## 4.2 Data

The data that is utilized in this study was obtained from the Nairobi Securities Exchange (NSE), the leading securities exchange in Kenya, serving as a marketplace for buying and selling different financial instruments, which include stocks, exchange-traded funds (ETFs), bonds, and other securities. The exact data is the NSE 20 Share Index historical data from December 1997 to August 2023. The NSE 20 Share Index is basically a price-weighted Index, which is calculated as the mean of the top 20 best-performing companies. The basis of the selection of the companies is their weighted market performance during the review period. The criteria for selection of the constituent companies are that: The company should have at least 20 percent of its shares quoted on the NSE, the primary listing of the company's shares must be on the NSE, the company should meet the Kes.20 million minimum market capitalization, the company must have been quoted continuously for at least one year, and it should be a 'blue chip' company, portraying a superior record of profits and dividends.

The NSE20 Index generally tracks the performance of the 20 largest and best-performing companies listed on the NSE in Kenya. It is one of the main indices used to gauge the overall health and performance of the Kenyan stock market. It is often seen as a barometer for the broader Kenyan economy and investor sentiment, and analyzing and forecasting this index can yield valuable insights for investors and shareholders.

Figure 1 shows the NSE 20 Share Index data from December 1997 to August 2023. The time series is asymmetrical, and this insinuates a lack of stable statistical properties in the up and down phases, which is a challenge when doing estimation with both linear and non-linear models.

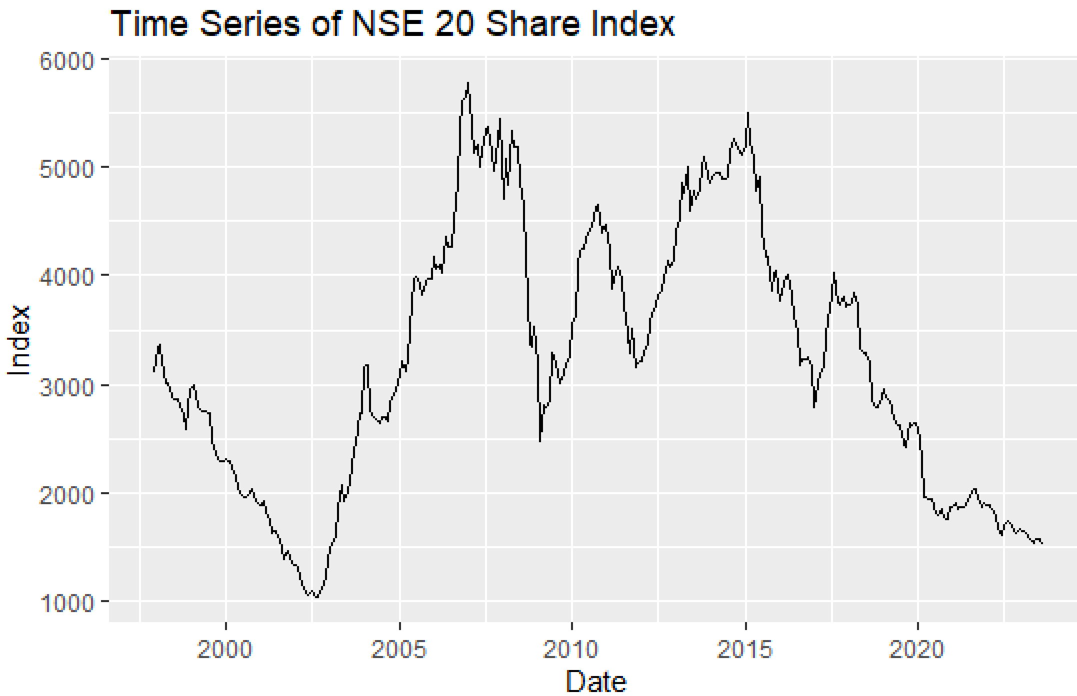


**Figure 1. Time Series of NSE 20 share index monthly: Dec 1997 - Aug 2023**

## 4.3 Non-Linearity Test

### 4.3.1 Tsay's F-Test

Tsay's threshold non-linearity test is conducted using different working orders ( 1 to 10). The following table shows the results of the tests showing the order of the model used for testing, the F test statistic, and the p-value obtained.

**Table 1. Test for Threshold Non-Linearity using Tsay's F-Test**

| Working Orders | test statistic | p-value |
|---|---|---|
| 6 | 2.249 | 0.001699 |
| 7 | 2.408 | 0.0001683 |
| 8 | 2.221 | 0.0001926 |
| 9 | 2.131 | 0.0001425 |
| 10 | 2.205 | 2.511e-05 |

The results support threshold non-linearity when orders 6,7,8, and 10 are used. Therefore, at 5% level of significance, the null hypothesis that suggests linearity is rejected.

### 4.3.2 BDS Test

To confirm the above results, the BDS test is run. The results obtained from running the BDS test are as shown below.

Table 2. **Test for Non-Linearity using BDS Test**

| P-Values | | | | |
|---|---|---|---|---|
| Embedding Dimension | | $\varepsilon$*(standard deviation)* | | |
| | **0.5** | **1.0** | **1.5** | **2.0** |
| **2** | 2.2 e-16 | 2.2 e-16 | 2.2 e-16 | 2.2 e-16 |
| **3** | 2.2 e-16 | 2.2 e-16 | 2.2 e-16 | 2.2 e-16 |

For the output above, the test employs default values of $\varepsilon = (0.5, 1.0, 1.5, 2.0)$, which are then transformed into the original data's units. From the results, the null hypothesis, stating that the data is i.i.d is rejected for all $m$ and $\varepsilon$ combinations at the standard significance levels. Given the lack of apparent linear patterns in the data, the outcomes of the BDS test indicate nonlinear structures within the data.

## 4.4 Test for Stationarity

### 4.4.1 ADF Test

The test yields a p-value of 0.73, leading to the retention of the null hypothesis as the obtained p-value exceeds the 0.05 significance level. Consequently, it can be inferred that the time series is characterized by non-stationary behavior.

### 4.4.2 Zivot-Andrews Unit Root Test

To confirm these results, the Zivot and Andrews Test is run.

Table 3. **Test for Stationarity using Zivot-Andrews Unit Root Test**

| Test Statistic | Critical Values | | |
|---|---|---|---|
| | **0.01** | **0.05** | **0.1** |
| -3.0983 | -5.34 | -4.8 | -4.58 |

Based on these findings, the t-statistic of the Z-A test exceeds the critical values, leading to the retention of the null hypothesis, signifying the presence of a unit root in the time series. This implies that the time series is non-stationary.

The time series data is non-stationary and hence it will be transformed to stationarity by logarithmic transformation of the series and then differencing. Obtaining stationarity in non-stationary time series is crucial for ensuring that the underlying assumptions of time series models and various statistical tests are met, leading to more accurate and reliable analyses, predictions, and interpretations. Logarithmic transformation is used to stabilize the variance of a time series.

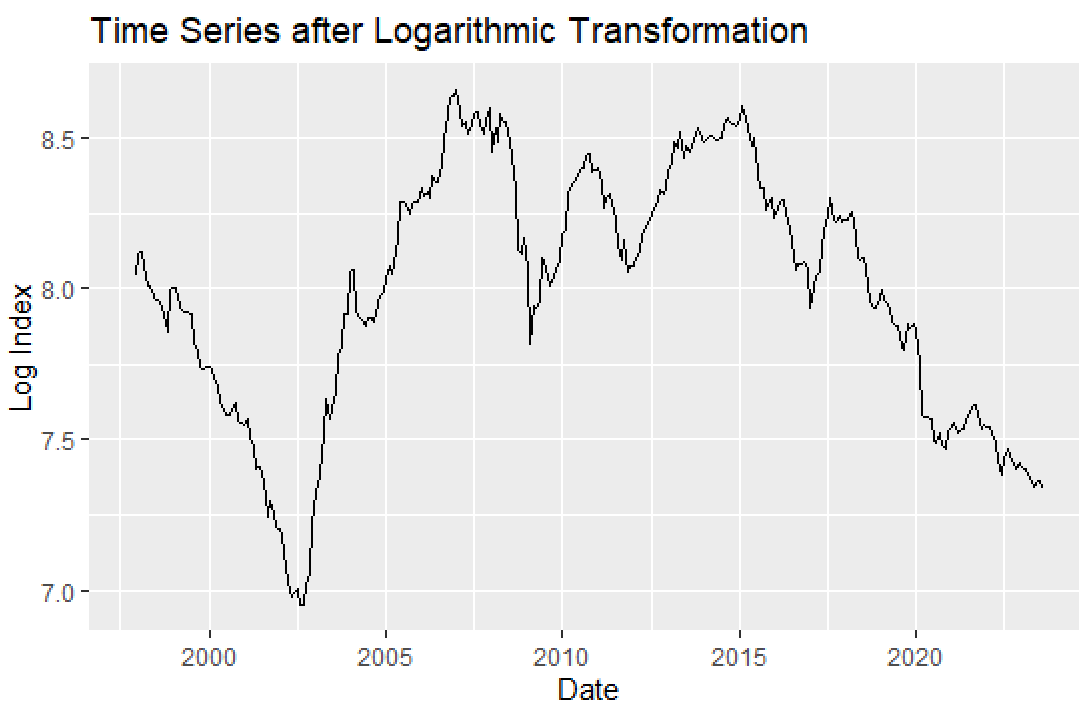


Figure 2. **Time Series after Logarithmic Transformation**

Differencing involves subtracting consecutive observations from each other. Figure 3 shows the transformed time series.

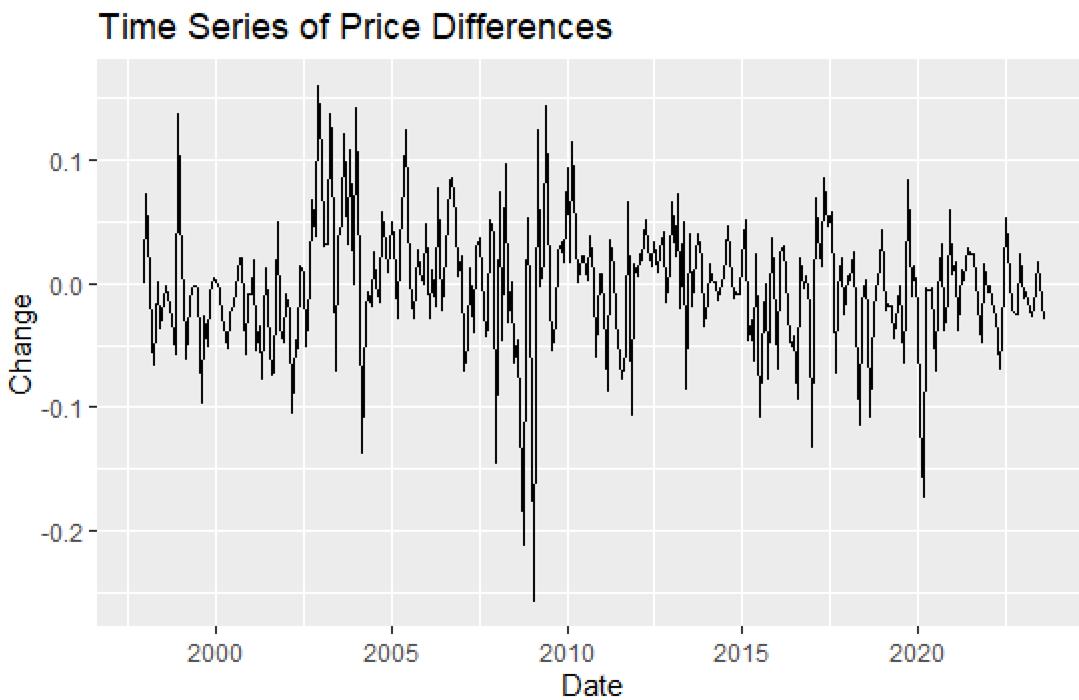


Figure 3. **Time Series of Change in NSE 20 Share Index**

Next, stationarity of the series is tested again to ensure that stationarity has been attained. The results of ADF test show that the series is now stationary, with a p-value of smaller than 0.01. Testing for stationarity using the Zivot and Andrews Test, these following results are obtained.

Table 4. **Test for Stationarity using Zivot-Andrews Unit Root Test**

| **Test Statistic** | | **Critical Values** | |
|---|---|---|---|
| | **0.01** | **0.05** | **0.1** |
| -11.7046 | -5.34 | -4.8 | -4.58 |

From the table above, t-statistic of the test falls below the critical values, and therefore the null hypothesis supporting the existence of a unit root is rejected.

## 4.5 Identification of Regimes

Looking at figure 3, positive values indicate growth in the NSE20 Index, which can be interpreted as good health or performance of the Kenyan stock market, while negative values and values close to zero indicate poor performance or stagnation. Seasons of good performance are characterized by sharp rises, while those of poor performance are indicated by sharp drops and stagnation around the zero line.

With these observations of growth and drops, a two-regime model will be suitable for modeling these growths and declines. This translates to just one $r$ value, which can informally be envisioned as a horizontal line separating the growth from the drop seasons. Therefore, our SETAR model will have two regimes and will take the form $SETAR(2; p_1, p_2)$. Further, using an alternative method of non-parametric approach of local polynomial fitting where we plot $x_t$ against its lags, visual inspection of the lagged plot below show regression curve that do not appear fairly straight. The selection of two regimes is also supported by the lagged plot where the fitted line seems to change around $x_{t-1} = 0$.

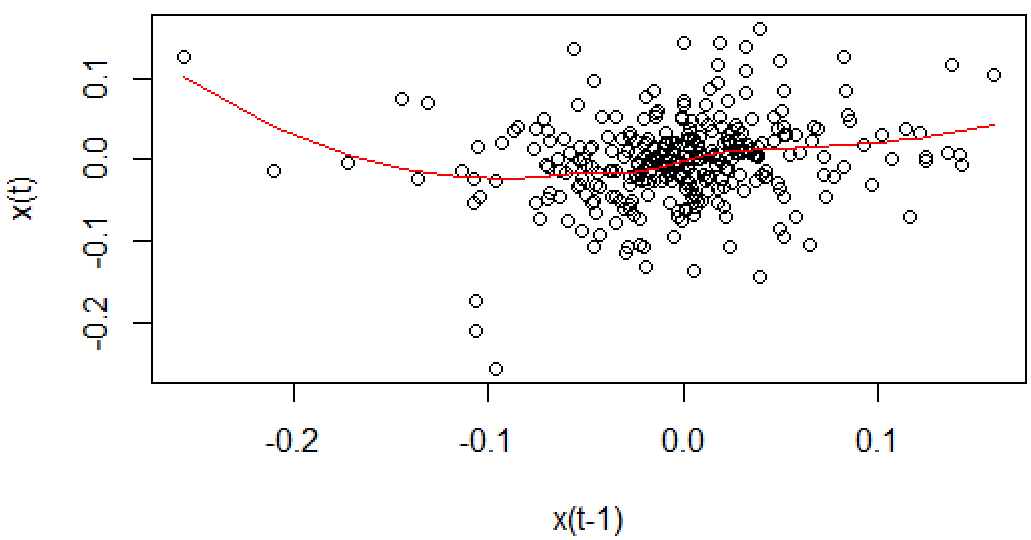


**Figure 4. Scatter Plots of $x_t$ against its Lags**

## 4.6 Selection of the Orders for the Regimes and the Delay Parameter

To develop the SETAR model, we need to find the orders of the two models, for the two identified regimes. That is, $p_1$ and $p_2$. We also need to find the lag parameter *d*. We use ACF and PACF to get the lag order. From the ACF and PACF plots figure 5 and figure 6 below, an order of 4 will be used for both models.

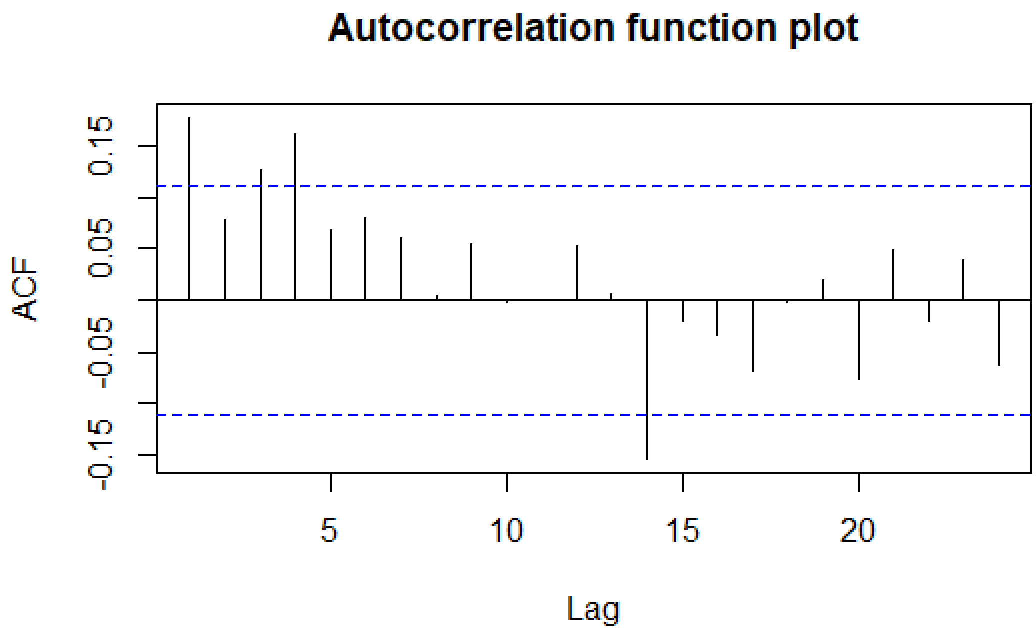


**Figure 5.** **ACF Plot**

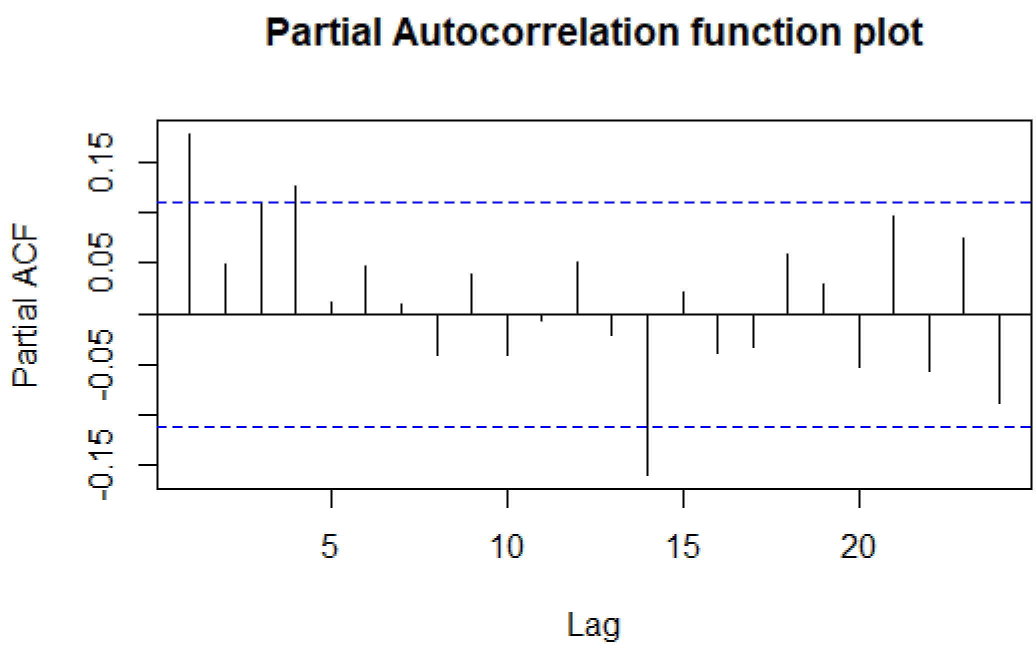


**Figure 6.** **PACF Plot**

To get the delay parameter, Tsay's threshold non-linearity test together with the $p$ value obtained is first tried, and later the Bayesian approach will be utilised. The delay parameter $d$ is assumed to satisfy $1 \leq d \leq p$. Next, with $p = 4$ and $d$ taking values from the set $\{1,2,3,4\}$, Tsay's threshold nonlinearity test for SETAR models is applied. The test statistics and the corresponding $p$ values are as follows;

Table 5. **Test for Threshold Non-Linearity using (p,d) Pairs**

| (p,d) | F-Stat | p-value |
|---|---|---|
| (4,1) | 2.796836 | 0.01764542 |
| (4,2) | 2.74101 | 0.01965312 |
| (4,3) | 0.9194889 | 0.4688014 |
| (4,4) | 0.4189912 | 0.8353437 |

Tsay suggests a way of selecting $d$ for a particular AR order $p$ such that,

$$d =_{i \in D} F(p,i)$$

where D is a set of all $d$ values being considered, and F(p,I) is the F-statistic obtained for the auxiliary regression with delay parameter equal to $I$ and AR order $p$. Following this rule and considering output obtained above for threshold nonlinearity test, $d$ is set to be $d = 1$. From this analysis, SETAR(2;4,4) with d=1 is obtained.

## 4.7 Bayesian Parameter Estimation

Next, the unknown parameters $\theta^{(1)}, \theta^{(2)}, \sigma_1^2, \sigma_2^2, d$ and $r$ are estimated using the Bayesian approach. The Gibbs sampling technique is used to find marginal posterior distributions of $\theta^{(1)}, \theta^{(2)}, \sigma_1^2, \sigma_2^2, d$ by working with conditional distributions obtained in Chapter 3. To find the marginal posterior distribution of $r$, the Metropolis algorithm is used.

The Gibbs sampler was executed for 1,000 iterations, and the initial 500 iterations were disregarded as the burn-in sample. $N = 1000$ is the total MCMC sample. The estimates of the parameters alongside their standard errors are as shown in the table below;

Table 6. **Parameter Estimation for the SETAR Model**

| Parameter | Mean | Median | Standard Deviation |
|---|---|---|---|
| $\theta_0^{(1)}$ | -0.0341 | -0.0338 | 0.0108 |
| $\theta_1^{(1)}$ | -0.2903 | -0.2889 | 0.1542 |
| $\theta_2^{(1)}$ | 0.0550 | 0.0529 | 0.1410 |
| $\theta_3^{(1)}$ | -0.0955 | -0.0960 | 0.1397 |
| $\theta_4^{(1)}$ | 0.3342 | 0.3368 | 0.1248 |
| $\theta_0^{(2)}$ | 0.0016 | 0.0016 | 0.0044 |
| $\theta_1^{(2)}$ | 0.1581 | 0.1577 | 0.1000 |
| $\theta_2^{(2)}$ | 0.0420 | 0.0434 | 0.0570 |
| $\theta_3^{(2)}$ | 0.1270 | 0.1216 | 0.0646 |
| $\theta_4^{(2)}$ | 0.0567 | 0.0572 | 0.0638 |
| $\sigma_1^2$ | 0.0034 | 0.0034 | 0.0005 |
| $\sigma_2^2$ | 0.0023 | 0.0023 | 0.0002 |
| r | -0.0216 | -0.0211 | 0.0037 |
| d | 1 | 1 | 0.0000 |

Using the above results, the following model is obtained;

$$y_t = \begin{cases} -0.0341 - 0.2903y_{t-1} + 0.0550y_{t-2} - 0.0955y_{t-3} + 0.3342y_{t-4} + a_t^{(1)}, & \text{if } y_{t-1} \leq -0.0216 \\ 0.0016 + 0.1581y_{t-1} + 0.0420y_{t-2} + 0.1270y_{t-3} + 0.0567y_{t-4} + a_t^{(2)}, & \text{if } y_{t-1} > -0.0216 \end{cases} \tag{20}$$

where $\hat{\sigma_1^2} = 0.0034$ and $\hat{\sigma_2^2} = 0.0023$. The estimated value of $d$ is 1, similar to the value previously estimated.

## 4.8 Forecasting

After estimating the SETAR model to capture the intricate patterns and nonlinear relationships within the NSE 20 Share Index, the focus of this study naturally extends to generating forecasts. Forecasting future values of the time series is paramount for informed decision-making in financial markets. In the forthcoming sections, we will employ the Bayesian SETAR (2;4,4) model to predict future values of the NSE20. To gauge the effectiveness of the Bayesian SETAR model, its forecasting performance is compared with that of a SETAR model estimated through a frequentist approach and an ARIMA model. This comparative analysis will shed light on the advantages and potential gains in accuracy and reliability offered by Bayesian SETAR modeling in the context of stock market forecasting.

### 4.8.1 Forecasting with the Bayesian SETAR (2;4,4) Model

For the forecasting process, the one-step-ahead recursive method was employed. This method involves a step-by-step prediction process, where the first forecast is initiated by estimating the next value using our Bayesian SETAR (2;4,4) model. Subsequently, this predicted value is integrated into the existing data, and the model is re-estimated. This cycle continues, step by step, to predict future values.

First is to forecast the NSE 20 Share Index values for the next 12 months starting from September 2023 to August 2024. The modeling data spanned from January 1998 to August 2023, and the forecasts were generated for the subsequent period from September 2023 onward. Table 7 and figure 7 and below show the predictions for the next 12 months using the Bayesian SETAR model.

Table 7. **Predicted Values of the NSE 20 Share Index for September 2023 to August 2024**

| | Predictions | | |
|---|---|---|---|
| Month/Year | Mean | Lower | Upper |
| Sep 2023 | 1479.60 | 1442.50 | 1509.65 |
| Oct 2023 | 1434.30 | 1363.80 | 1499.86 |
| Nov 2023 | 1384.13 | 1286.95 | 1477.09 |
| Dec 2023 | 1334.12 | 1214.67 | 1455.83 |
| Jan 2024 | 1310.46 | 1154.62 | 1450.02 |
| Feb 2024 | 1312.03 | 1138.23 | 1469.58 |
| Mar 2024 | 1294.95 | 1092.29 | 1464.44 |
| Apr 2024 | 1278.36 | 1048.83 | 1459.33 |
| May 2024 | 1280.79 | 1034.87 | 1482.12 |
| Jun 2024 | 1282.58 | 1021.09 | 1503.62 |
| Jul 2024 | 1285.41 | 1008.21 | 1528.02 |
| Aug 2024 | 1288.24 | 994.29 | 1551.43 |

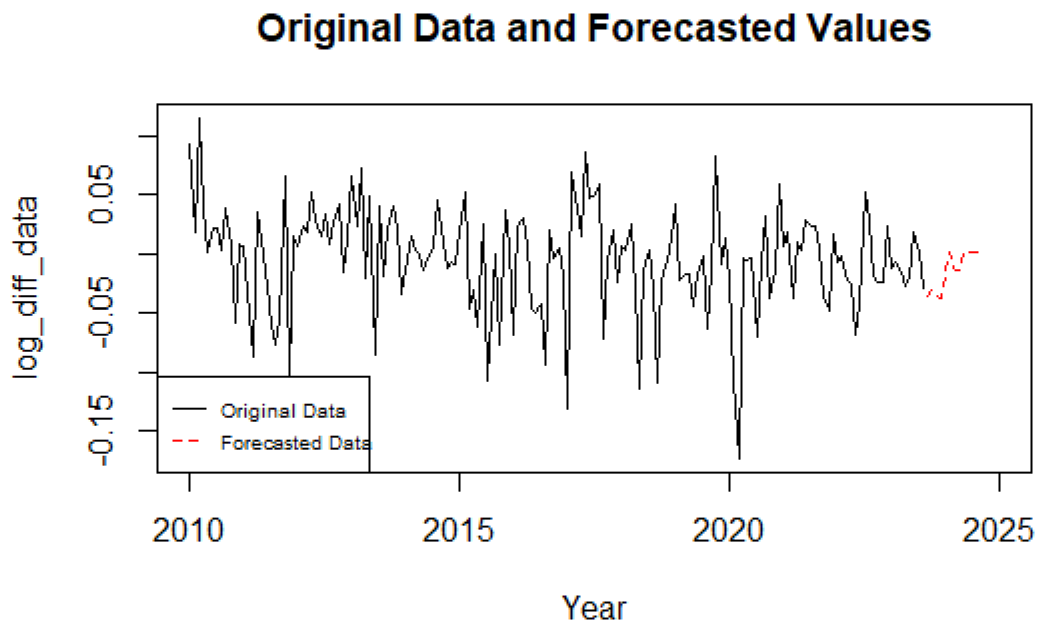


Figure 7. **Predicted Future Changes of the NSE20 Index are Indicated by the Red line**

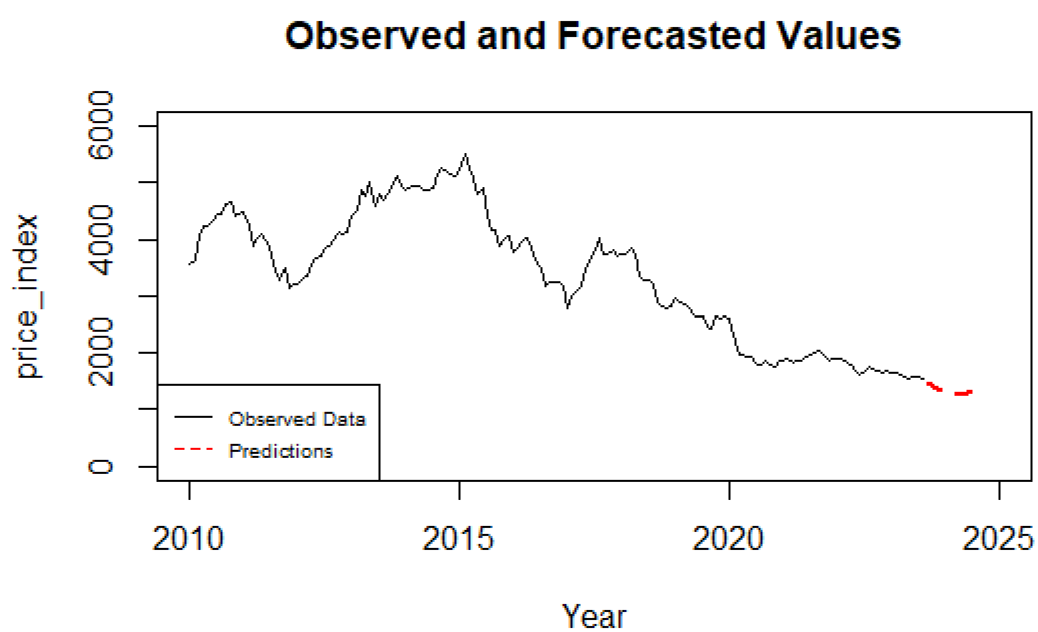


Figure 8. **Predictions of the NSE20 Index for the Next 12 Months are Indicated by the Red line**

### 4.8.2 Evaluation of the Model's Forecasting Ability

In evaluating the forecasting precision of the model, the Bayesian SETAR model, developed with data spanning from January 1998 to December 2022, was compared with a SETAR model estimated using the frequentist approach and an ARIMA model. The primary interest was forecasting the NSE 20 Share Index changes for January 2023 to August 2023. To assess the accuracy of these forecasts, Root Mean Square Errors (RMSE) were calculated for the three models. The RMSE provides a quantitative measure of the forecast error, with smaller values indicating better predictive accuracy. The results are as shown in the following table 8.

**Table 8. RMSE for the Bayesian SETAR, Normal SETAR, and ARIMA Models**

| | **Bayesian SETAR** | **Frequentist SETAR** | **ARIMA** |
|---|---|---|---|
| **RMSE** | 0.0213 | 0.0239 | 0.0176 |

The results indicate that the Bayesian SETAR model outperformed its frequentist counterpart in forecasting the NSE20 changes. The RMSE for the Bayesian SETAR was 0.0213, while that of the other SETAR model was 0.0239.

However, compared with ARIMA forecasts, the ARIMA model performed better. It is important, however, to note that this comparison was for forecasts for a short period of time (only eight months). Next, a comparison is made for a longer forecasting horizon where structural breaks are expected. The models were fitted for data spanning from January 1998 to Aug 2019, and forecasts were made for September 2019 and August 2023 (48 months). During this forecast period, there were periods of abrupt and high changes, and the goal was to investigate whether the tested models captured this behavior.

The performance results are as shown in the following table 9 and the forecasts are as shown in figure 9.

Table 9. **RMSE for the Bayesian SETAR, Normal SETAR, and ARIMA Models**

| | Bayesian SETAR | Frequentist SETAR | ARIMA |
|---|---|---|---|
| RMSE | 0.0388 | 0.0515 | 0.0415 |

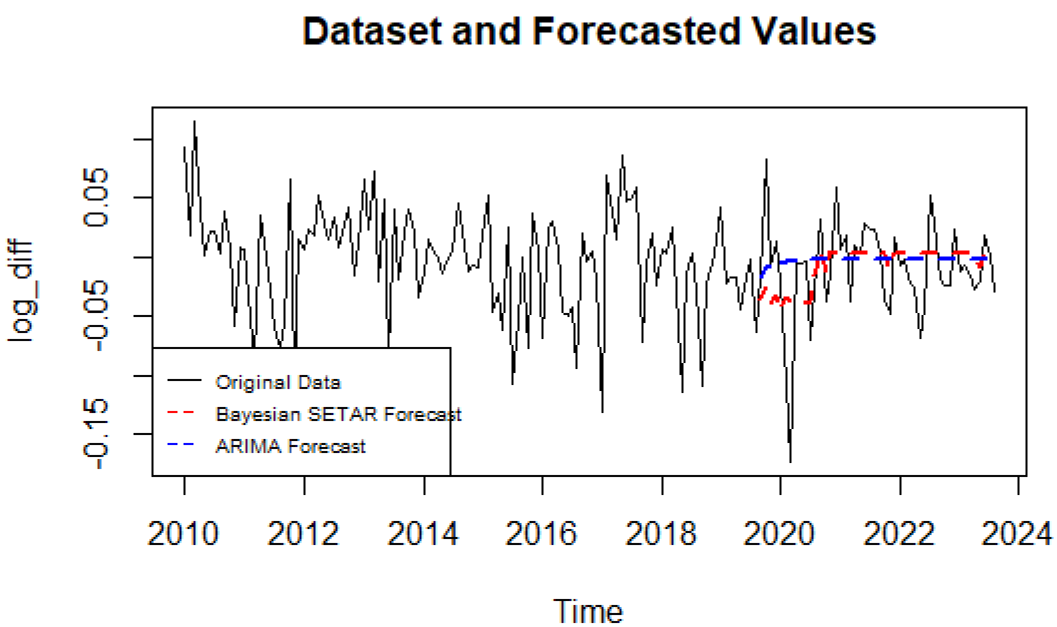


Figure 9. **Original Data and Forecasted Values using Bayesian SETAR and using ARIMA**

Here, the Bayesian SETAR had a better predictive power. The RMSE for Bayesian SETAR model was 0.0388, while that of ARIMA was 0.0415.

# 5 Summary, Conclusions, and Recommendations for Further Research

## 5.1 Summary

The main goal of this research was to explore the Bayesian framework for SETAR models with application to Nairobi Securities Exchange (NSE) 20 Share Index data. To this end, the four specific goals were to analyze the properties of the NSE20 Index data, to determine the estimates of SETAR model parameters using the Bayesian approach, to forecast the NSE 20 Share Index for the next 12 months using the fitted model, and to compare the performance of the Bayesian SETAR with the frequentist SETAR and ARIMA model in forecasting.

SETAR modeling was used to model the changes in NSE 20 Share index. The study began with stationarity and non-linearity tests, which confirmed the presence of non-linearity in the NSE20 data. This justified the use of a SETAR model to grasp the behavior of the data. The data was not stationary and to ensure the data had stable statistical properties for convenience in analysis, logarithmic transformation and differencing was conducted. The Bayesian approach of parameter estimation was used to find the estimates of the model parameters.

Through an analysis of the data, SETAR(2;4,4) was the most adequate for the data. Using the Bayesian approach, the model parameters were estimated, including autoregressive coefficients $\theta^{(1)}$ and $\theta^{(2)}$, regime switching parameter ($r$), error variances $\sigma_1^2$ and $\sigma_2^2$, and the delay parameter ($d$). The estimated Bayesian SETAR model was then used to generate forecasts for the next 12 months of the NSE20 Index. Considering the forecasts obtained, the NSE 20 Share Index is expected to continue in a downward trend in the coming months till April 2024 where the NSE 20 Share Index is expected to be around 1278.36 , after which they index is expected to start going up slowly.

In terms of forecasting performance, the Bayesian SETAR model outperformed the SETAR model that utilized the Frequentist approach. This highlights the advantages of incorporating Bayesian techniques for modelling and forecasting non-linear financial time series. Compared to the ARIMA model, Bayesian SETAR performed better for longer forecasting horizons and during periods characterized by high fluctuations. It showed that

Bayesian SETAR performed better where there were high changes or fluctuations, compared to ARIMA. These findings underscore the Bayesian SETAR model's resilience and adaptability, particularly in scenarios characterized by heightened volatility.

## 5.2 Conclusions

In conclusion, this study has successfully applied the Bayesian framework to SETAR model, specifically in the context of modeling the Nairobi Securities Exchange (NSE) 20 Share Index data. The study achieved its primary objectives, including the analysis of key data properties, estimation of SETAR model parameters through Bayesian techniques, and the generation of forecasts for the NSE20 Index over the next 12 months. The findings highlight the applicability of SETAR modeling in capturing the non-linear dynamics inherent in financial time series data, particularly observed in the NSE20 Index. The forecasted trends indicate a projected downward trajectory in the NSE 20 Share Index until April 2024, followed by a gradual upward trend. The implications of these findings extend to different stakeholders in the financial market, including investors, investment firms and policymakers.

The application of the Bayesian approach for parameter estimation proved instrumental in refining the SETAR model, emphasizing its adaptability in handling non-linearity. In comparison with the frequentist SETAR and ARIMA models, the Bayesian SETAR model exhibited superior performance, particularly in scenarios marked by high volatility and fluctuations, outperforming its counterparts. This underscores the model's resilience and adaptability, positioning it as a robust tool for forecasting in financial environments characterized by dynamic and unpredictable changes.

In general, the Bayesian SETAR model offers a valuable tool for more accurate and adaptive forecasting in the face of market uncertainties. Its demonstrated superiority in handling non-linearities and volatile conditions suggests practical applications in decision-making processes, risk management, and strategic planning within the financial sector. As the Kenyan financial market continues to navigate complex and evolving conditions, the Bayesian SETAR model emerges as a valuable asset for stakeholders seeking reliable insights into market behavior and trends, and hence this model is recommended.

## 5.3 Recommendations for Further Research

The first recommendation is on model comparison. Future research can extend this study's comparison to include other non-linear time series models. Further comparison can be done to compare the accuracy of the Bayesian SETAR model with other nonlinear models such as Markov-switching models, STAR (Smooth Transition Autoregressive

Models) and nonparametric models, such as kernel regression or neural network-based models. This broader model comparison can provide a more comprehensive understanding of which modeling approach is most suitable for NSE 20 Share Index data or similar financial data.

Natural conjugate priors were chosen for this study. There are various methods of obtaining the priors, including subjective priors, objective priors, empirical priors, and conjugate priors, and in this case conjugate priors were used because of mathematical convenience. Future studies can also investigate how variations in prior distributions for the model parameters affect the results.

In this study, two regimes were selected for the study. However, future studies can extend the study to incorporate more than two regimes. Expanding the scope to include multi-regimes may lead to better capturing of the data's underlying patterns and nonlinear relationships that may not be fully revealed in a two-regime framework.

# References

1. Agiwal V. and Kumar, J. (2020), Bayesian Estimation for Threshold Autoregressive Model with Multiple Structural Breaks, *Metron*, 78(3), 361-382.

2. Aydin D. and Güneri Ö. İ. (2015), Time Series Prediction Using Hybridization of AR, SETAR and ARM Models, *International Journal of Applied*, 5(6), 87-96.

3. Boero G. and Lampis F. (2017), The Forecasting Performance of SETAR Models: An Empirical Application, *Bulletin of Economic Research*, 69(3), 216-228.

4. Carroll R. J., Ruppert D. and Stefanski L. A. (1995), Measurement Error in Nonlinear Models (Vol. 105), CRC press.

5. Firat E. H. (2017), SETAR (Self-Exciting Threshold Autoregressive) Non-Linear Currency Modelling in EUR/USD, EUR/TRY and USD/TRY Parities, *Mathematics and Statistics*, 5(1), 33-55.

6. Gibson D. and Nur D. (2011), Threshold Autoregressive Models in Finance: A Comparative Approach. *In Proceedings of the Fourth Annual ASEARC Conference*, 17, 18-23.

7. Ojo O. O. (2021), Bayesian Modelling of Inflation in Nigeria with Threshold Autoregressive Model, *Rattanakosin Journal of Science and Technology*, 3(1), 10-18.

8. Oyewale A. M., Adelekan O. G. and Innocient O. O. (2017), Forecast Comparison of Seasonal Autoregressive Integrated Moving Average (SARIMA) and Self Exciting Threshold Autoregressive (SETAR) Models, *American Journal of Theoretical and Applied Statistics*, 6(6), 278-283.

9. Palma W. (2016), Time Series Analysis, John Wiley & Sons.

10. Pan J., Xia Q. and Liu J. (2017), Bayesian Analysis of Multiple Thresholds Autoregressive Model, *Computational Statistics*, 32, 219-237.

11. Prado R. and West M. (2010), Time Series: Modeling, Computation, and Inference, Chapman and Hall/CRC.

12. Tobechukwu N. M., Emmanuel B. O. and Ibienebaka B. T. (2022), Self-Exciting Threshold Autoregressive Modelling of COVID-19 Confirmed Daily Cases in Nigeria, *International Journal of Data Science and Analysis*, 8(6), 182-186.

13. Tong H. and Lim K. S. (1980), Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(3), 245-268.

14. Tsay R. S. and Chen R. (2019), Nonlinear Time Series Analysis (Vol. 891). John Wiley & Sons.