Universität Paderborn
Faculty of Business Administration and Economics
Department of Economics

A.184.4095 Masterarbeit

Modeling and forecasting trend- or scale-stationary time series under short or long memory applied to economic, financial and environmental data.

by
Mohammad Niaj Uddin Ahmed
Matriculation Number: 6900003
niaj@mail.uni-paderborn.de

submitted to
Prof. Dr. Yuanhua Feng

10.01.2025

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig, ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken, Tabellen, Skizzen, Zeichnungen, bildliche Darstellungen usw. sind ausnahmslos als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Paderborn, 10.01.2025

MD. Niaj uddin

Mohammad Niaj Uddin Ahmed

Abstract

This thesis shows in details the modeling and forecasting of trend- or scale-stationary time series with short or long memory characteristics, and the methods are applied to economic, financial, and environmental data. The thesis shows models, such as ARMA and ARIMA, for short-memory processes, and also shows semiparametric approaches like SEMIFAR for long-memory processes. The fractional differencing parameter plays a crucial role in capturing long-range dependence in this thesis. Real-world data are included economic indicators, financial ETF volume data, and environmental datasets like natural gas consumption and vehicle miles driven. The thesis has shown the trends LM and SM trends.

Keywords: Short Memory, Long Memory, tsmoothlm, ARMA, SEMIFAR, FARIMA, Semiparametric

Contents

List of Figures

List of Tables

1 Introduction

In this thesis I have concluded every aspects of the topic. The analysis of time series data is significant in various fields, including economics, finance, and environmental studies. Time series analysis focuses on understanding and forecasting patterns, which are crucial for decision-making processes in these domains. For instance, economists do their research on time series models to predict inflation trends, and it can be seen in real life when a govt try to mitigate impacts on economics. Analyzing unemployment rates, and forecasting GDP growth, all fall into in same area. In finance, investors can use such models for portfolio optimization, risk management, anlysing UHF data. Similarly, environmental scientists also use models for research and apply these models to monitor climate changes, forecast energy consumption, and design sustainable resource management systems.

In the field of time series, one fundamental distinction we can observe in the nature of memory exhibited by the data. As my thesis topic has explicitly mentioned to cover the either Short or Long memory processes where Short-memory processes characterized by rapidly decaying correlations, meaning we have data with limited influence of past events on future values. On the other hand, long-memory processes do the opposite meaning it is defined by slow-decaying correlations, indicate persistent effects that extend over time. Understanding these distinctions is very important for correct modeling and forecasting.

In This thesis I have explored both parametric and semiparametric approaches to time series modeling. Parametric models, such as Autoregressive Moving Average (ARMA) and its possible extensions, are well-suited for short-memory processes. These models rely on mathematical structures to capture relationships within the data, which offers us a robust framework for time sereis analysis.

For long-memory processes, fractional models like Fractionally Integrated Autoregressive Moving Average (FARIMA) provide us a nuanced approach, which is incorporating fractional differencing to account for persistent correlations meaning stroing previous memory. Additionally, I have explored semiparametric models, such as the Semiparametric Fractional Autoregressive (SEMIFAR) model, which combines the strengths of parametric methods with the flexibility of nonparametric techniques, which allows the modeling of complex patterns and trends.

The methodologies explored in this thesis have wide-ranging applications. In economics, time series models are essential for predicting macroeconomic indicators and informing policy decisions, such as macroeconmics metrics Consumer Price Index, Producer Price Index, and Employent rate in an economy. Financial applications include the modeling of Financia index, interest rates, and asset volatilities, all of which exhibit varying memory characteristics. I have picked up three finacial index in this thesis, Such as ETF Volume for S&P500, Nikkei225, and NASDAQ. For Environmental datasets, I have used as natural gas consumption, Vehicle miles. and waste management where all these three activites have higher GHG emmision to the environment. This thesis seeks to address systematically evaluating existing methodologies to specific datasets.

My thesis is structures starts with Introduction; Where I have provided an overview of the research problem, objectives, and relevance, and setting up the stage for following chapters. Parametric Time Series and Short Memory: Explores the theoretical foundations and applications of short-memory models, with a focus on ARMA and related techniques. Semiparametric Time Series and Long Memory: Examines models designed for long-memory processes, including SEMIFAR and FARIMA, and their extensions. Implementation of Models on Datasets; Where I have demonstrated the application of these models to real-world economic, financial, and environmental datasets. For these processes I have used LM trend extractor function such `tsmoothlm` from esemifar package.

I have bridged theoretical insights with practical applications, and my this thesis contributes to the field of time series analysis, offering tools and methodologies that enhance our ability to model and forecast complex datasets.

2 Short-Memory Models and Parametric Approaches

2.1 Time Series under Short Memory

In this chapter we will discuss about the possible models that are related to short memory in details. Starting from some of the most notable models in short memory time series area. Given that our topic focuses on modeling and forecasting trend or scale-stationary time series under short and long memory time series memory in economic, financial, and environmental data, many of the suitable models in short memory will be discussed further, nevertheless the applicability and most relevant models only depend on the research topics and the very specific characteristics of the data. We have divided this chapter into parts such as, related models under short memory, its components etc. Further we will also discuss in further chapter for potential models which can also be used under long memory process.

In time series context the topic of short memory has not been explored extensively while we can find a vast amount of researches when it comes to time under long memory. From general knowledge we can define that models under short memory are those models where we can observe the impact of a shock or change diminish rather quickly. As the lag increases the correlation between variables decreases. Some of the most common models like ARMA (Autoregressive Moving Average) and ARMAX (Autoregressive Moving Average with Exogenous variables) are used as short memory as these models are often viewed as sufficient for research and practical purposes. ARMA model family often serve the purposes of short memory time series analysis.

2.2 Short-Memory Processes according to Local Polynomial Fitting

The paper from Beran and Feng (2002a) explores various aspects of memory processes in time series analysis which are including short-memory processes. In this section, we will provide a in depth breakdown of concepts related to short memory and depth discussion of parametric time series, suitable for inclusion in this thesis.

Characteristics of Short Memory

Short-memory processes, also known as in time series arena as weakly dependent processes, refer to time series where the autocorrelations decay exponentially fast. This means that observations separated by a large time gap have minimum or no correlation. Mathematically, for a time series X_t, short-memory implies:

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty,$$

where $\gamma(k)$ is the autocovariance function of X_t.

Alternatively, a process can be called short-memory when its autocorrelation function $\rho(k)$ satisfies:

$$\lim_{k\to\infty} |\rho(k)| = 0 \quad \text{and} \quad \sum_{k=-\infty}^{\infty} |\rho(k)| < \infty.$$

This condition ensures that the dependence between distant observations diminishes sufficiently fast.

Contrast with Long Memory

Short-memory processes are contrasted with long-memory processes, which have slowly decaying autocorrelations. In long-memory processes, the sum of autocovariances which does not converge, but in short-memory processes, the sum is finite. Specifically, in long-memory processes, the autocorrelation function $\rho(k)$ decays hyperbolically, often following a power-law of the form:

$$\rho(k) \sim L(k)k^{-(2d-1)}, \quad \text{as } k \to \infty,$$

where $0 < d < 0.5$ and $L(k)$ is a slowly varying function.

This distinction is critical when applying statistical models, as models that assume short memory may not be appropriate for long-memory data. Wrongly selecting the memory contents of a time series can lead to wrong model and misleading inference.

Modeling Short Memory

Short-memory processes can often be perfectly modeled while using Autoregressive Moving Average (ARMA) models. These models catch the varieties of time series with faster diminishing correlations, as the ARMA structure allows for a limited number of lags to make impacts on the current observation. An ARMA(p, q) model is defined as:

$$\Phi(L)X_t = \Theta(L)\varepsilon_t,$$

where:

- $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is the autoregressive (AR) polynomial,
- $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ is the moving average (MA) polynomial,
- L is the lag operator ($LX_t = X_{t-1}$),
- ε_t this is white noise error term with zero mean and constant variance.

Because here MA or the AR involve finite orders p and q, the influence of past observations diminishes beyond these lags.

Estimation Methods for Short-Memory Processes

The paper also discusses the local polynomial fitting method for estimating regression functions in the presence of short-memory errors. This method is adapted for short-memory data to improve the accuracy of estimation by minimizing the bias which could occur due to avoiding memory characteristics.

In the context of nonparametric regression, consider the model:

$$Y_t = m(X_t) + \varepsilon_t,$$

where $m(\cdot)$ is not known regression function, X_t is the predictor variable, and ε_t are short-memory errors. Local polynomial fitting involves estimating $m(\cdot)$ by lowering a weighted sum of squared residuals in a neighborhood of a target point.

When The presence of short-memory errors affects the variance of the estimator but it does not affects the bias. When properly accounting for the dependence structure in

ε_t it is crucial for accurate estimation of the variance and also for constructing valid confidence intervals.

Statistical Techniques in Short Memory

For time series with short memory, the asymptotic properties of estimators, such as variance and bias, behave somewhat differently than for long-memory processes. The convergence rates for estimators are typically much faster due to the rapid decay of dependencies.

Specifically, under certain regularity conditions, estimators of parameters in short-memory processes we can observe achievable the standard $\sqrt{n}$ convergence rate, where n is the sample size. The central limit theorem applies, which is allowing for normal approximation of the sampling distribution of estimators.

Implementation

In practical applications, short-memory models are widely used in fields like economics and finance, where time series are often modeled assuming short-range dependence. Examples where we can include modeling stock returns, interest rates, and economic indicators.

The results from the paper Beran and Feng (2002a) shows the importance of choosing the correct model to account for short or long memory, surely depending on the observed data characteristics. Using a short-memory model when the data exhibit long-memory properties can lead to underestimated standard errors and overconfident inferences, hence it is important to keep an eye on estimation of d.

Reasoning

By integrating these points into my thesis, a comprehensive view of short-memory processes as explored in the paper is provided, particularly in contrast what we have seen with long-memory models and their role in local polynomial fitting. It is important for understanding the differences between short and long memory is crucial for accurate modeling and inference in time series analysis.

Also in the book, *Statistics for Long memory process* has extensively described the short memory process in the context of time series. The book is published by Beran (2017), where the author has explained short memory dependence as short range memory or weak dependence process where these act as a type of stochastic process. The faster decay implies that the correlation can be summed up. Mathematically, a short memory can be observed if auto correlations $\rho(k)$ decays to zero at a rate at which it will be fast enough for summing up the absolute values of auto correlations where overall lags are finite, so the equation stands at,

$$\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty.$$

- $\rho(k)$ is the autocorrelation function where k is the time series lags.
- and also we can state that k represents the lag, which shows us the difference in time when two observations are being compared.

When comparing to long memory processes, we can observe the differences and contrasts where correlations will decay very slowly. The impact of past values on future values becomes unrecognizable in a short memory process, this happens while after a short period. This features make them simple to handle statistically. The books from Beran (2017) and Brockwell and Davis (1991) have showed the short memory process in more details in the context of time series, which is demonstrated below,

Mathematical Characterization

Auto correlation Function (ACF)

In stationary time series $\{X_t\}$, the autocorrelation function $\rho(k)$ measures the correlation nuances between observations X_t and X_{t-k} which is separated separated by lag k. We can process has short memory if:

$$\lim_{k\to\infty} |\rho(k)| \leq C\phi^k,$$

where $C > 0$ as constant and $0 \leq \phi < 1$. This nuance demonstrate that auto correlations will be decaying exponentially faster whereas k will be increasing.

Summability of Auto covariance Function

The auto covariance function is $\gamma(k)$ which can be defined as summable :

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty.$$

This condition dictates that the cumulative impacts of auto covariance lags is finite.

Spectral Density Function

The spectral density $f(\lambda)$ in the case of short memory process, it is bounded by the frequencies λ in the interval $[-\pi, \pi]$, hence the equation goes as,

$$0 < \sup_{\lambda} f(\lambda) < \infty.$$

This implies there are no infinite traces in the frequency domain.

Properties of Short Memory Processes

- **Rapid Decay of Autocorrelations**: In the short memory time series, the influence of past values decreases quickly, which makes distant observations practically free from current data points in other words independent.
- **Finite Variance of Partial Sums**: In the time series short memory context The variance of the sum $S_n = \sum_{t=1}^{n} X_t$ increases linearly along with n, i.e., $\mathrm{Var}(S_n) = O(n)$.
- **Central Limit Theorem Applies**: Because of the finite variance of partial total summation, there will be holding of standard statistical inference techniques and asymptotic normality.
- **Transitory Shocks**: Any shocks or innovations in the data are to be the system that have effects that dissipate quickly over the time span.

Examples of Short Memory Processes

Autoregressive Moving Average (ARMA) Models: ARMA processes, which will be discussed in detail below, under stationarity conditions, exhibit short memory.

AR(1) Process

$$X_t = \phi X_{t-1} + \epsilon_t$$

where $|\phi| < 1$ and ϵ_t is white noise. The autocorrelation function is $\rho(k) = \phi^k$, showing exponential decay. **Moving Average (MA) Processes** which have autocorrelation beyond lag q are zero which demonstrates short memory model.

Contrasting with Long Memory

In contrast, time series under long memory (or long-range dependencies) processes have autocorrelations which decay over the time with much more slow pace, often known as a power-law decay. The equation as follows,

$$\rho(k) \sim k^{-d} \text{ as } k \to \infty, \quad 0 < d < 0.5.$$

For long memory processes:

- The autocorrelations can not be an absolutely summable process.
- The spectral density function most often tends to be an infinity as frequency reaches near zero.
- Any shocks and changes have persistent effects over the time span.

Implications in Time Series Modeling

- **Model Selection**: Short memory processes most often by standard ARMA models as these models are seen as sufficient models but Long memory processes may require ARFIMA (Autoregressive Fractionally Integrated Moving Average) models.

- **Forecasting**: In short memory processes, recent observations carry more weight or more impact full data points in forecasting the future values.
- **Statistical Inference**: Standard statistical methods are often seen as valid for short memory model creation because of the applicability in term of central limit theorem.

As we dive deep into the time series under short and long memory, which will be discussed further with more relevant models in details and context.

2.3 Models under Short Memory in details

2.3.1 AR, MA, and Combined ARMA models

Autoregressive Models are suitable for short memory processes as the autocorrelation function exponentially decays because AR model uses a time series with its own past values.

The generic form of the AR model p is:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

where:

- X_t is the time that is represented in time series t,
- $\phi_1, \phi_2, \ldots, \phi_p$ are represented as the autoregressive coefficients,
- ϵ_t is the error term for the time series t.

similar to AR, MA is also appropriate for short memory. MA models can be expressed as past white noise errors. The most common uses for both AR and MA is for financial time series. The generic formula structure is give below as q,

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

where:

- X_t is the time that is represented in time series t,

- $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$ this is represented as the white noise error terms at times $t, t-1, \ldots, t-q$,
- $\theta_1, \theta_2, \ldots, \theta_q$ coefficients for the moving average.

With the combination of both AR and MA models we can have the ARMA model. This model is frequently used in economic time series which shows short-term momentum along with mean reverts. Below we can find the basic form of the ARMA model with p and q,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

where:

- X_t is the time that is represented in time series t.
- $\phi_1, \phi_2, \ldots, \phi_p$ the relationship between the current value of the series and its past p values.
- ϵ_t for time t,this is the white noise error, at assumed mean of zero and constant variance.
- $\theta_1, \theta_2, \ldots, \theta_q$ coefficients for the the moving average (MA), and the past q error terms.
- p indicating number of past values of the series in the model.
- q indicating number of past error terms are included in the model.

The ARMA model combines of both the AR and MA components, AR part deals with the short-term dependencies and MA part deals with the effects of past shocks.

2.3.2 ARIMA

The paper from Benvenuto et al. (2020) used ARIMA model on the covid-2019 data set to show it application where the author has explained the ARIMA model that adds autoregressive(AR) model, moving average model(MA), and seasonal autoregressive integrated moving average (SARIMA). Further on the top ARIMA model, the authors have also used the augmented dickey-fuller(ADF) to employ unit root test to understand the stationary situation within the data. To make the data steady, preferred techniques like log transformation and differences were also used. Besides seasonal and non seasonal differences helped the model to stabilize the periodicity and the trend. The reason of this explanation is to show the flexibility ARIMA model has. The authors Ariyo et al. (2014) have explained the ARIMA model and it's elements,

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \tag{2.1}$$

where,

- Y_t is value for the actual entities and whereas ϵ_t is the random error at t.
- ϕ_i and θ_j can be noted as coefficients.
- p and q are integers which very often referred as autoregressive (AR) and moving average (MA), respectively.

In addition of building an ARIMA model consists of several factors such as *identification for the model*, *checking the diagnostic*, and also *estimation of the parameter*.

Determining ARIMA Model

Further Ariyo et al. (2014) have shown that determining ARIMA depends on some of the following criteria, these criteria were used in the stock market data which are general used most often for building an ARIMA model.

- Relatively small of Bayesian Information Criterion was used because it is relevant for making an ARIMA model.
- Standard error of regression is relevant to the model building.
- Adjusted R-square is also significant element.
- Finally, Q statistics is analysed in the study because of its importance. Besides autocorrelation functions and (ACF) and partial autocorrelation function is also used.

ARIMA and Neural Network

Author Pai and Lin (2005) stated that, although traditionally the autoregressive integrated moving average model is used and modified widely in the academic fields and it has been the mostly used model, but it has a lacking which is that ARIMA can not properly apprehend the nonlinear patterns in the data. to tackle that Support Vector Machines (SVM) is used for capturing much more accurate nonlinear patterns in the data because of it is based on neutral network capabilities. This method is highly sought after because of its performance on stock price data. We will explain SVM further in later sections.

2.4 Fundamentals of Parametric Time Series Models

Definition of Parametric

Parametric models play a important role in time series analysis, this offers us in the field a structured approach to modeling and forecasting temporal data which means it lose dependencies slowly. According to Chen et al. (1997) These models assume a specific functional form which is needed for behind the scene data-generating process, characterized by a finite set of parameters. The parametric model mainly focuses on short memory time series. With leveraging the parametric structure, many researchers have caught the essential dynamics of time series data, for examples autocorrelation structures, seasonality, and volatility clustering.

This part of my thesis will discuss briefly about parametric time series models, which focusing on their theoretical foundations, estimation methods, and practical applications. An emphasis is placed on ensuring consistency in notation and variables throughout the discussion. The key models such as ARMA, ARIMA, and GARCH are explored in depth, with detailed explanations of their properties and estimation techniques. Since my thesis topic is explicitly about parametric and semiparametric method then I will not discuss anything about GARCH model in this thesis. I have added relevant literature sources to provide a complete context and to guide further reading.

Parametric time series models are defined by a specific functional form that is related past observations and random shocks to the future values. The primary goal here is to model the autocorrelation structure of the series using a finite number of parameters.

Estimation Methods for Parametric Models

While Estimating the parameters of parametric time series models, it is important for us to accurate modeling and forecasting. There are common estimation techniques in this topic where we include include the Method of Moments, Least Squares Estimation, and Maximum Likelihood Estimation.

Maximum Likelihood Estimation (MLE)

MLE is widely used because of its desirable statistical properties, for example it has consistency and asymptotic efficiency. The likelihood function is usually constructed where it is based on the assumed distribution of the error terms.

MLE for ARMA Models

For ARMA models, the likelihood function is typically evaluated by using the conditional likelihood approach or we can also evaluate it with the exact likelihood when the sample size is small.

Least Squares Estimation

For AR models, Ordinary Least Squares (OLS) can be applied directly because the model is linear in parameters. However, for MA and ARMA models, we must have nonlinear optimization techniques.

Method of Moments

The Method of Moments is involved when equating the theoretical moments of the model to the sample moments and when we are solving for the parameters.

2.5 Properties of Parametric Models

Stationarity and Invertibility

For valid modeling, We have ARMA processes that must satisfy stationarity and invertibility conditions. The AR part needs to have the roots of the characteristic equation to lie outside the unit circle, while the MA part needs to have roots inside the unit circle.

Autocorrelation and Partial Autocorrelation Functions

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are necessary tools for identifying the order of ARMA models.

- **AR(p) Model**: PACF cuts off after lag p, ACF tails off.
- **MA(q) Model**: ACF cuts off after lag q, PACF tails off.
- **ARMA(p, q) Model**: Both ACF and PACF tail off.

Diagnostic Checking is when after fitting a parametric model, diagnostic itself checks which are necessary to validate the model. **Residual Analysis** is about analyzing the residuals for randomness and the constant variance makes sure that the model adequately captures the data structure.

Ljung-Box Test test assesses the null hypothesis that is the residuals and they are independently distributed.

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k},$$

where $\hat{\rho}_k$ is the autocorrelation of residuals at lag k.

2.6 Applications of Parametric Models

Economic Time Series; Parametric models are extensively used in modeling economic indicators for example GDP growth or PPI, CPI in my case or inflation rates, and un/employment levels. **Financial Time Series**; In finance, ARIMA models are widely used to model asset returns and volatility, aiding in risk management and option pricing. **Engineering and Environmental Sciences**; Time series models are applied in signal processing, climatology, and hydrology for forecasting and also analyzing temporal patterns.

3 Long-Memory and Semiparametric Time Series Models

As it was cited in Beran (2017) that a process will express **long memory** or also known as long-range dependence if its correlations decay at a very slow rate and they are not summable. This interpretation can be described by the equation:

$$\rho(k) \sim c_p \, |k|^{-\alpha}. \tag{3.1}$$

where:

- $p(k)$ this represents the autocorrelation function at which the lag is k,
- c_p, a finite positive constant,
- $\alpha \in (0, 1)$, a parameter which determines the rate at which the memory decays the correlations.

In long memory processes, the dependence between events or variables that are far apart and which diminishes at a very slow rate, which can defined as the notation $k \to \infty$. This definition itself contrasts the short memory processes, where the correlations decay much faster, often simultaneously and exponentially.

3.1 Long-Range Dependence in Time Series

Long-range dependence in short LRD, also known as long memory or long-term persistence, is a time series related phenomenon observed in data where correlations between variables decay at a slow pace than it is anticipated in short-memory time series. We can redefined also as, events or values which remain far apart in time frame as they are significantly correlated.

The concept of Long range dependence originated from studies that were done to understand the field of hydrology, particularly from the study of the author Hurst (1951), who has who has studied the Nile River's water levels, where he observed persistent correlations between the observations over the time scales which are much longer than he expected under traditional or classical short-memory models. This revolutionary discovery later led to the development of new statistical software tools and thoroughly studied models to capture long memory behavior.

Long Range dependence has since been identified in various fields:

- **Economics and Finance:** In this field we can define the aspects of asset returns, interest rates, and volatility which most often exhibit long memory, and it is affecting modeling and forecasting.
- **Telecommunications:** This field also exhibits long range dependence behaviour as internet traffic and network packet arrivals show long range dependence, this behaviour is impacting network design and congestion control.
- **Climate Science:** Studies on climate change or weather has also used a lot of long range dependence methods or models to understands the climate's most unpredictable behaviour. Temperature and precipitation records data show long-term persistence thus it is influencing climate modeling.
- **Chemistry and Biology:** In chemistry ot biology fields we can observe processes like DNA sequences and physiological signals which can exhibit long-range correlations between observations.

Defining Long Memory Processes

Long memory can be defined in multiple ways but it is depending on the properties of the process and also depending the context of the analysis. The definitions generally diverted from the following categories:

Second-Order Definitions Based on a Stationary Process

Long memory is typically defined using second-order properties for stationary processes, for example the autocorrelation function and the spectral function of the density. This topic is studied by authors Granger and Joyeux (1980) and Hosking (1981) where they clarified the followings,

Autocorrelation Function Approach

A stationary time series can be said to have (X_t) exhibited long memory if its autocorrelation function $\rho(k)$ decayed hyperbolically rather than exponentially because of the lag k exhibits increase, hence the equation stands as,

$$\rho(k) \sim L(k)k^{-(2d-1)}, \quad \text{as } k \to \infty,$$

where:

- $0 < d < 0.5$ is the parameter for the long memory process.
- $L(k)$ is a slowly varying function at infinity level (i.e., for the constant at any point is $c > 0$, $\lim_{k\to\infty} \frac{L(ck)}{L(k)} = 1$).

Implications:

- The sum of autocorrelations diverges:

$$\sum_{k=1}^{\infty} \rho(k) = \infty.$$

- The process shows a significant memory for the past values over long time-periods.

Spectral Density Function Approach Alternatively we can define, in the frequency domain, long memory depends on the of the density of spectral $f(\lambda)$ changes near the zero frequency:

$$f(\lambda) \sim G|\lambda|^{-2d}, \quad \text{as } \lambda \to 0,$$

where:

- $G > 0$, a constant.
- The spectral density has a pole at frequency which is zero, that initiates to infinite power at frequencies of low level.

Second-Order Definitions Based on Non-Stationary Process

Authors Gray et al. (1989) and Robinson (1995) have explained in their papers where the showed thorough explanation of Gaussian semiparametric estimation for the long range dependencies for the process of non-stationary and also the authors have showed fractional autoregressive moving-average (FARMA) model. Long memory is often defined using the concept of fractional integration.

Fractional Integration A time series (X_t) has the behaviour of integrated of order d $(X_t \sim I(d))$ if the applying of the fractional differencing operator as $(1-L)^d$ yields a process for stationary then,

$$(1-L)^d X_t = \varepsilon_t,$$

where:

- L, a lag operator $(LX_t = X_{t-1})$.
- ε_t, a stationary and short-memory process.
- d known as real number which allows for fractional differencing.

Properties:

- When $0.5 < d < 1$, X_t is known as non-stationary but it is still observed as mean-reverting.
- The autocovariance function $\gamma(k)$ behaves as:

  $$\gamma(k) \sim \frac{\sigma^2}{\Gamma(1-2d)} k^{2d-1}, \quad \text{as } k \to \infty,$$

  where $\Gamma(\cdot)$ is the gamma function.

Implications for Non-Stationary Processes

- Non-stationary processes with long memory have the infinite variance properties which is significant.

- Standard differencing ($d = 1$) may cause over difference the series, that might facilitate fractional differencing.

Continuous-Time Process Definition

Long memory is the context of continuous-time processes which can also be explained with fractional Brownian motion.

Fractional Brownian Motion (fBm) Fractional Brownian motion, which is published by the author Mandelbrot and Van Ness (1968), where $B_H(t)$ is a continuous-time Gaussian process known and tagged by the Hurst parameter $H \in (0, 1)$:

- $B_H(0) = 0$.
- Mean zero: $\mathbb{E}[B_H(t)] = 0$.
- Covariance function:

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t - s|^{2H}\right).$$

Properties:

- When $H = 0.5$, $B_H(t)$ that reduces to standard Brownian motion.
- For $H > 0.5$, increments have positive correlation that indicates long memory.
- For $H < 0.5$, increments have negative correlation.

Spectral Density of fBm The spectral density of fBm shows behaviour such as:

$$f(\lambda) \propto |\lambda|^{1-2H}, \quad \text{as } \lambda \to 0.$$

- A zero frequency pole when $H > 0.5$, it is indicating a long memory porcess.

Alternate Definitions of Long Memory

Other definitions focus on different statistical properties of the time series process in the context of long memory. Author Lo (1991) has shown a test for long run memory where a robust short range dependencies has also been exhibited.

Variance of Partial Sums For a process in (X_t) with long memory, the variance of the partial sum will be $S_n = \sum_{t=1}^{n} X_t$ grows faster than linear process:

$$\mathrm{Var}(S_n) \sim n^{2d+1}, \quad \text{as } n \to \infty.$$

- For $0 < d < 0.5$, the variance increases at a rate higher than n, indicating persistent behavior.

Rescaled Range (R/S) Analysis The rescaled range topic is introduced by Hurst (1951) where R/S analysis measures the range for the cumulative deviations from the mean which is rescaled by the standard deviation. Hence, the formula stands at,

$$\frac{R(n)}{S(n)} \propto n^{H}, \quad \text{as } n \to \infty,$$

where:

- $R(n)$, defined as range of the first n number of observations.
- $S(n)$ which is the standard deviation of the first n number of observations.
- H is known as the Hurst exponent with range $(0 < H < 1)$.

Interpretation of Hurst Exponent:

- $H = 0.5$: has no long-range dependencies or we can state as random walk).
- $H > 0.5$: has positive long-range dependencies also can be defined as the persistent behavior.
- $H < 0.5$: has negative long-range dependencies or behavior of the anti-persistent.

Finally when we seek for a fitted long memory concept then we can state that long memory is a fundamental concept in time series analysis which is reflecting persistent correlations over long time variations and periods. The various definitions provided which are essential for:

- **Model Selection:** while choosing appropriate models like Autoregressive Fractionally Integrated Moving Average (ARFIMA) for data exhibiting long memory.
- **Forecasting:** In the case of forecasting, we have to Understand long-term dependencies which improves forecasting accuracy.
- **Statistical Inference:** while correctly estimating the time series parameters and testing hypotheses we need proper statistical interference.

When we intend to comprehensively defining long memory through different lenses, situations of stationarity, or fractional integration, or even continuous-time processes, and alternative statistical properties we gain a deeper knowledge and experiences of the behavior of complex time series meta.

3.2 Semiparametric Modeling with SEMIFAR Models

In time series analysis, capturing both long-range dependence and nonstationarity depicts some significant challenges to face. Traditional parametric models may not be successful to adequately represent complex data structures, particularly when we have underlying processes which exhibit trends or long memory. To address these problems, **Semiparametric Fractional Autoregressive (SEMIFAR)** models have been developed as a flexible framework that mashes up both parametric and nonparametric approaches together.

The SEMIFAR model, discussed by Beran and Feng (2002b), which extends to the conventional fractional autoregressive integrated moving average (FARIMA) models by incorporating a nonparametric component with the model deterministic trends or with the smooth functions. This section of the chapter provides an in-deep understanding of exploration of SEMIFAR models, discussing their formulation, properties, estimation methods, and applications.

3.3 Background on Semiparametric Models

Semiparametric models connects the gap between fully parametric and nonparametric models by adding both parametric and nonparametric elements into one process. This hybrid approach lets introduce a greater flexibility in modeling complex data structures without creating rigid functional forms.

Advantages of Semiparametric Models

- **Flexibility**: They can catch nonlinear trends and patterns that parametric models might miss.
- **Efficiency**: With this modeling we only require only components that are flexibility nonparametrically, they maintain the efficiency of parametric models for other components.
- **Robustness**: They are less sensitive to model wrong specification, leading to more reliable inference.

Challenges in Time Series Analysis

Time series data often exhibit:

- **Long-Range Dependence (LRD)**: Autocorrelations decay slowly, indicating persistent effects over time.
- **Nonstationarity**: Statistical properties such as mean and variance can change over time.
- **Trends and Seasonality**: Deterministic patterns that can confound analysis if they are not properly modeled.

Semiparametric models, like the SEMIFAR model, are well-suited to address these mentioned challenges by allowing for a more flexible representation of trends and of long memory.

3.4 Definition of the SEMIFAR Model

The SEMIFAR model combines a nonparametric function to model deterministic trends with a fractional autoregressive component to have long-range dependence caught.

*Model Specification

Let $\{Y_t\}_{t=1}^n$ be a time series. The SEMIFAR model is defined as:

$$Y_t = m\left(\frac{t}{n}\right) + u_t,$$

where:

- $m(\cdot)$, this component is an unknown smooth deterministic function which representing trends or seasonality.
- u_t, this is a stationary stochastic process exhibiting long-range dependence, which we have as modeled as a fractionally integrated process.

The stochastic component u_t follows a fractional autoregressive integrated moving average (FARIMA) model:

$$\Phi(L)\left(1 - L\right)^d u_t \;=\; \Theta(L)\,\varepsilon_t.$$

where:

- L, known here as the backshift (lag) operator ($LY_t = Y_{t-1}$).
- d, which is the fractional differencing parameter ($0 < d < 0.5$) capturing long memory.
- $\Phi(L)$, and $\Theta(L)$ are used for polynomials which representing the autoregressive (AR) and moving average (MA) components, respectively.
- ε_t is white noise with zero mean and constant variance.

Components of the Model

- **Nonparametric Trend Function ($m(\cdot)$):** Where it captures deterministic trends without specifying a parametric form. We can observe here estimated used techniques like local polynomial regression.
- **Fractional Integration ($(1 - L)^d$):** Accounts for long-range dependence in the stochastic component.
- **ARMA Structure ($\Phi(L), \Theta(L)$):** Already discussed models which show short-term dependencies and shocks in the data.

3.5 Properties of the SEMIFAR Model

Long-Range Dependence

- The fractional differencing parameter d determines the degree of long memory.
- When $0 < d < 0.5$, the process u_t is stationary with autocorrelations decaying hyperbolically.

Nonparametric Trend

- The function $m(\cdot)$ allows us for flexible modeling of trends, accommodating the nonlinearity and smooth changes over time.
- By scaling time as t/n, the function adapts to the sample size, ensuring consistency in estimation.

Stationarity and Invertibility

- The stochastic component u_t is stationary if $d < 0.5$ and the roots of $\Phi(L)$ lie outside the unit circle.
- Invertibility requires the roots of $\Theta(L)$ to lie outside the unit circle.

3.6 Estimation Methods

When it comes to estimating the SEMIFAR model we can observe involvement where jointly estimating the nonparametric trend $m(\cdot)$ and the parameters of the stochastic component u_t.

Estimation of the Nonparametric Trend $m(\cdot)$

Local Polynomial Regression:

- It fits a polynomial of degree p locally around each point.
- This uses a kernel function $K(\cdot)$ to weight observations based on their distance from the target point.

Estimator:

$$\hat{m}(x) = \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{Y},$$

where:

- $\mathbf{X}$, is the design matrix for the polynomial terms.
- $\mathbf{Y}$, is the vector of observations.
- $\mathbf{W}$, is a diagonal weight matrix with elements as it can be written as $K\left(\frac{t/n - x}{h}\right)$.
- h, is the bandwidth controlling the smoothness of the estimate.
- $\mathbf{e}_1$, this component selects the intercept term as the estimate at point x.

Bandwidth Selection:

- Crucial for balancing bias and variance.
- Methods can be used which include cross-validation and plug-in selectors.

Estimation of the Stochastic Component u_t

After detrending the data:

$$\hat{u}_t = Y_t - \hat{m}\left(\frac{t}{n}\right),$$

the residuals $\hat{u}_t$ are used to estimate the parameters of the FARIMA model.

Estimating d:

- **Semiparametric Methods**: This utilizes the spectral density properties of long-memory processes.
  - **Log-Periodogram Regression**:

    $$\log I(\lambda_j) = \alpha - 2d \log \lambda_j + \epsilon_j,$$

    where:

    * $I(\lambda_j)$ is the periodogram at frequency λ_j.
    * ϵ_j is an error term.
  - **Whittle Estimator**: These components tend to minimizes an approximation of the likelihood function in the domain of the frequency.

Estimating ARMA Parameters:

- Use methods like maximum likelihood estimation (MLE) or conditional least squares on the residuals $\hat{u}_t$.

Iterative Estimation Procedure

Due to the interdependence between $m(\cdot)$ and u_t, an iterative procedure is often employed:

1. **Initial Estimation**:
    - Estimate $m(\cdot)$ using a preliminary bandwidth.
    - Compute residuals $\hat{u}_t$.
2. **Estimate d and ARMA Parameters**:
    - Estimate d and other parameters using $\hat{u}_t$.
3. **Update Bandwidth**:
    - Adjust the bandwidth h based on the estimated long-memory parameter d.
4. **Re-estimate $m(\cdot)$ and u_t**:
    - Repeat steps 1–3 until convergence.

3.7 Asymptotic Properties

Consistency and Asymptotic Normality

- Under regularity conditions, the estimator $\hat{m}(x)$ is consistent.
- The bias and variance of $\hat{m}(x)$ changes as per as the bandwidth h and the long-memory parameter d.

Mean Squared Error (MSE)

Bias:

$$\text{Bias}[\hat{m}(x)] = \frac{h^{p+1}}{(p+1)!} m^{(p+1)}(x)\mu,$$

where μ is a constant depending on the kernel function.

Variance:

$$\text{Var}[\hat{m}(x)] \approx \frac{\sigma^2}{nh^{2d} f_X(x)},$$

where:

- σ^2 is the variance of ϵ_t.
- $f_X(x)$ is the density of t/n at point x.

The presence of long memory ($d > 0$) tends to increase the variance, indicating the need for larger sample sizes or bandwidth adjustments.

3.8 Applications

Modeling Economic and Financial Time Series

- **Economic Indicators**: Involves in capturing trends and cyclical behavior in GDP, inflation rates, etc.
- **Financial Markets**: Involves in modeling asset prices and returns that exhibit long memory and nonstationarity.

Environmental Data Analysis

- **Climate Studies**: Involves in analyzing temperature records with long-term trends and dependencies.
- **Hydrology**: Involves in modeling river flows and precipitation patterns over time.

Signal Processing

- **Telecommunications**: Handling data with long-range dependence and time-varying patterns.

3.9 Advantages and Limitations

Advantages

- **Flexibility**: Simultaneously models deterministic trends and stochastic dependencies.
- **Improved Forecasting**: Better captures underlying patterns, leading to more accurate predictions.
- **Theoretical Foundations**: Well-established asymptotic properties under regularity conditions.

Limitations

- **Computational Complexity**: Iterative estimation can be computationally intensive.
- **Bandwidth Selection**: Choosing the optimal bandwidth is challenging, especially in the presence of long memory.
- **Interdependence of Components**: Estimation of $m(\cdot)$ and u_t are interrelated, complicating inference.

3.10 Extensions and Related Models

Seasonal SEMIFAR Models

- Incorporate seasonal effects by extending the nonparametric component or adding seasonal ARMA terms.

Multivariate SEMIFAR Models

- Extend the framework to handle multiple time series simultaneously, meaning multiple variables, capturing cross-dependencies.

Nonstationary Fractional Models

- Allow the fractional differencing parameter d to vary over time or depend on covariates.

The SEMIFAR model provides a robust and flexible framework for modeling time series data with complex structures, including trends and long-range dependence. The model integrates nonparametric regression techniques with fractional time series models, it addresses the limitations and lackings of traditional models in capturing nonstationarity and persistent dependencies.

Understanding and properly implementing the SEMIFAR model can lead to significant improvements in analyzing and forecasting time series data across various fields and various types of datasets, from economics to environmental sciences.

3.11 FARIMA Model

3.11.1 FARIMA: Historical Context

The **Fractional Autoregressive Integrated Moving Average (FARIMA)** model, also known as **ARFIMA**, is an extension of the traditional or classical ARIMA model that forces to take fractional values from the differencing parameter. This innovations or errors exist to enable the model to capture long memory or long-range dependence in time series datasets, which traditionally captured by ARIMA models that cannot address adequately.

The origin of the FARIMA model are properly depicted in the papers published by the prominent writers from the pioneering works of Granger and Joyeux (1980) and Hosking (1981), the authors have independently acknowledged the concept of fractional differencing in time series analysis spectrum. They observed that the data driven phenomenon in many real-world time series which exhibit persistent autocorrelations which decay at a hyperbolic rate hence this is indicative of long memory, rather than the exponential decay phenomenon that we observe in ARIMA models.

Limitations of Traditional Time Series Models

The conventional time series models such as ARMA and ARIMA have limitations when dealing with long memory time series processes. So the limitations are as follow,

1. **Inadequate Modeling of Long Memory:**
   - ARIMA models tend to decay the autocorrelations exponentially, as this is not suitable for the crucial long memory data because the autocorrelations decay at a hyperbolic rate.
2. **Over-Differencing:**
   - while working with the given process, applying integer differencing to a series that needs fractional differencing as we can observe that it can lead to over-differencing, as a result the loss of important information can be noticed on the underlying process.
3. **Inaccurate Forecasting:**
   - In time series models which do not account for long memory tend to provide weak forecasts, as they fail to read through the persistence behaviour in the data.
4. **Misleading Inference:**
   - In thime series process statistical tests and confidence intervals which are based on improper models which may be invalid, in doing so the model proceses leads to ambiguous results.

Key Features of FARIMA

The FARIMA model is used from the ARIMA model which shows a relation with ARIMA model as the differencing parameter d tend to be an significant fractional value as it is noted as FARIMA(p, d, q), where:

- p: is the order of the autoregressive (AR) function.
- d: Known as the fractional differencing parameter ($d \in \mathbb{R}$).
- q: is the order of the moving average (MA) function.

Derivation of the FARIMA Process

The FARIMA model can be defined by the equation:

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t,$$

where:

- $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is the AR polynomial.
- $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ is the MA polynomial.
- L is the lag operator object ($LX_t = X_{t-1}$).
- ε_t which is the white noise error term with zero mean and variance σ^2.

The fractional differencing operator $(1-L)^d$ is known using the binomial expansion:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k,$$

where $\Gamma(\cdot)$ which is the gamma function, and the generalized for the binomial coefficient $\binom{d}{k}$ is:

$$\binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

Properties:

- **Reduction to ARMA Models:** When $d = 0$, the model steps back to an ARMA(p, q) model.
- **Reduction to ARIMA Models:** When d is an integer, the model steps into an ARIMA model.
- **Introduction of Long Memory:** The fractional differencing operator introduces long memory as it enters the process when $0 < d < 0.5$.

This whole interpretation has been taken from Hosking (1981).

Spectral Density, Autocovariance, and Autocorrelation of the Process

Spectral Density Function According to Beran (2017), the spectral density $f(\lambda)$ of a FARIMA which we will know as the process near zero frequency and that behaves as:

$$f(\lambda) \propto |\lambda|^{-2d}, \quad \text{as } \lambda \to 0.$$

- For $0 < d < 0.5$, the spectral density turns into an unbounded at zero frequency that starts to introduce long memory.

Autocovariance Function The autocovariance function $\gamma(k)$ as we know that decays hyperbolically:

$$\gamma(k) \sim \frac{\Gamma(1-2d)\sigma^2}{\Gamma(d)\Gamma(1-d)} k^{2d-1}, \quad \text{as } k \to \infty.$$

Autocorrelation Function The autocorrelation function $\rho(k)$ which also decays at a hyperbolic rate:

$$\rho(k) \sim k^{2d-1}, \quad \text{as } k \to \infty.$$

- The sum of autocorrelations are not convergence rather it diverges when $0 < d < 0.5$, consistent with long memory behavior.

3.11.2 Fractional Brownian Motion: Relevance and Insights

Fractional Brownian Motion (fBm) From Mandelbrot and Van Ness (1968) and Samorodnitsky et al. (1996), it is known for its component of continuous-time Gaussian process $B_H(t)$ characterized by the Hurst parameter $H \in (0, 1)$:

- **Initial Condition:** $B_H(0) = 0$.
- **Mean:** $\mathbb{E}[B_H(t)] = 0$.
- **Covariance Function:**

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

Relation to FARIMA Models:

- FARIMA processes can be defined as a discrete-time analogs of fBm when we have $d = H - \frac{1}{2}$.
- Both fBm and FARIMA models exhibit time series that contains long memory properties when $H > 0.5$.

Properties:

- **Standard Brownian Motion:** For $H = 0.5$, fBm step backs into the standard Brownian motion.
- **Long-Range Dependence:** Incremental persistence of fBm are rather stationary and have long-range dependence when we have value of $H \neq 0.5$.

3.11.3 FARIMA Extensions and Generalizations

Several extensions which we collected from few studies such as Porter-Hudak (1990) and Lobato (1999) from what the FARIMA model addresses various complexities in time series data, such as;

1. **Seasonal FARIMA Models (SARFIMA):**
    - These models incorporate seasonal patterns by adding extension to the FARIMA model with seasonal differencing components and seasonal ARMA components.
2. **Non-linear FARIMA Models:**
    - As this addresses a non-linearities by including the non-linear functions.
3. **Multivariate FARIMA Models:**
    - With extending to multivariate settings to model multiple time series models with the long memory and potentials to the co-movements.
4. **FARIMA-GARCH Models:**
    - A combination process of FARIMA with Generalized Autoregressive Conditional Heteroskedasticity as stands for (GARCH) models work hand in hand to capture both of the long memory in the mean and also the volatility clustering.

5. **Fractionally Integrated Processes with Stable Innovations:**
   - We can say that it allows for heavy-tailed error distributions to the model data which has significantly noticeable large outliers.

3.11.4 Estimating the Fractional Differencing Parameter d

As the authors from Beran (1995) and Geweke and Porter-Hudak (1983) demonstrate the parameter d where it is a very significant and when we have to adjust the FARIMA model. The parameter involves in fractional differencing. There are several methods exist to find out the d estimation;

1. **Parametric Estimation:**
   - **Maximum Likelihood Estimation (MLE):**
     - This estimates almost all model jointly which is also including d, we have that by maximizing or topping the likelihood function.
   - **Exact and Whittle Likelihood Methods:**
     - The Whittle estimator nears the likelihood in the domain of the frequency as a result it activates the reduction function for computational complexity.
2. **Semiparametric Estimation:**
   - **Geweke and Porter-Hudak (GPH) Estimator:**
     - According to author that the log-periodogram regression is:

       $$\log I(\lambda_j) = \beta_0 \; - \; 2d \, \log(\lambda_j) \; + \; \epsilon_j,$$

       where $I(\lambda_j)$ is the periodogram at frequency λ_j.
   - **Robinson's Estimator:**
     - This estimator utilizes the local Whittle estimation for d which in return does not include the short-term components.
3. **Wavelet-Based Estimation:**

- This step explores the time-scale decomposition of wavelets to estimate d.

4. **Rescaled Range (R/S) Analysis:**
   - Rescaled range involves in estimating the Hurst exponent H, from which d can be derived as $d = H - \frac{1}{2}$.

Difficulties:

- Since there are long memory components within the datasets which complicate further the estimation because of the infinite variance and slow decay of autocorrelations.
- Also the difficulties lie on numerical optimization which may be challenging as there are complex likelihood surface.

3.11.5 Maximum Likelihood Estimation Techniques

Authors Rohatgi and Saleh (2015) have offered an intensive study on probability and statistics which are filled with profound deep knowledge and the study has also dedicated chapters on MLE where Exact Maximum-Likelihood Estimation (MLE) demonstrates efficient and consistent parameters which are impactful for FARIMA models. However, from general perspectives we know that the computational complexity tend to increase when there are increase in the sample sizes because of the need to invert large covariance matrices.

Cholesky Decomposition Method

The Cholesky Decomposition is used widely to factorize the covariance matrix $\mathbf{\Sigma}$ of the observations numbers:

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\top,$$

where $\mathbf{L}$ is amatrix that is lower triangular in shape. This decomposition which is less completed, the computation of the determinant and inverse of $\mathbf{\Sigma}$, which are needed in the likelihood function, the formula goes as,

$$\mathcal{L}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}),$$

where:

- θ which represents the parameters to be estimated.
- $\mathbf{X}$, this is the vector of observations.
- $\boldsymbol{\mu}$, this is the mean vector.

Advantages and cons:

- It can be effective in reduction of computational burden by leveraging matrix factorization.
- It ensures numerical stability during the computation of the likelihood.

Levinson-Durbin Algorithm

The Levinson-Durbin algorithm is known for its effectiveness in solving the Yule-Walker equations when we need to structure autoregressive models. For FARIMA models:

- The algorithm is suited for working with the infinite AR representation of the FARIMA process.
- It recursively calculates the parameters and since there will be prediction error variances.

Algorithm processes:

1. **Initialization:** Set $\phi_0 = 1$ and prediction error variance σ_0^2.
2. **Recursion:** For $k = 1$ to n, calculation reflecting coefficients κ_k.
3. **Update:** Update of AR coefficients ϕ_k and the prediction error variances σ_k^2.

Pros:

- Computational efficiency: $O(n^2)$ this includes operations instead of $O(n^3)$ for matrix inversion.
- It reduces the computational load for large datasets which will make the process much faster.

Exact State Space Approach

From the work of Sowell (1992), we can feature that the state-space representation of the FARIMA model will show us the necessity of the Kalman filter for likelihood evaluation.

State-Space Representation:

- **State Equation:**

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\varepsilon_t,$$

- **Observation Equation:**

$$X_t = \mathbf{C}\mathbf{x}_t + \varepsilon_t,$$

  where $\mathbf{x}_t$ is the state vector.

Kalman Filter Steps:

1. **Prediction:**

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1},$$

2. **Update:**

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(X_t - \mathbf{C}\hat{\mathbf{x}}_{t|t-1}),$$

   where $\mathbf{K}_t$ is the Kalman gain.

Pros:

- This is efficient in working with missing data and irregular time series.
- It also exhibit efficiency for large datasets because of its recursive computation behaviour.

Asymptotic Results for the Exact MLE

When it comes to understanding the asymptotic properties of the MLE it is crucial to deal with statistical inference. This part is dicussed in detailes in the paper from Fox and Taqqu (1986)

Consistency and Asymptotic Normality:

- Under regularity conditions, the MLE $\hat{\theta}$ is shows a consistent and asymptotically normal component:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{I}^{-1}(\theta_0)),$$

  where θ_0 used as the true parameter value, and $\mathbf{I}(\theta_0)$, which is the Fisher information matrix.

Cons:

- Long memory introduces complexities in proving the asymptotic normality when it shows dependence structure.
- The standard regularity conditions may not hold as the model pregresses.

3.11.6 Using Autoregressive Approximations for FARIMA

Autoregressive (AR) approximations simplify the estimation of FARIMA models by approximating them with high-order AR models.

Haslett-Raftery Method

As name of the section suggest, hence this part is populated with help from the study of Haslett and Raftery (1989) where the publisher suggested that this method proposes a approximate of the FARIMA process with a high-order AR model. Which also assembles the fact that the infinite AR representation of FARIMA can be deflected.

Steps:

1. **AR Approximation:**

- This method fits an AR(m) model, where m is large enough to hold or capture the long memory traces.

2. **Parameter Estimation:**
   - The Yule-Walker equations are used or Burg's method is used to estimate AR coefficients.

3. **Estimate d:**
   - At this point it relates the estimated AR coefficients to the fractional differencing parameter d.

Pros:

- It makes the process computationally much more simpler in compared to exact MLE.
- It is well equipped in handling large datasets effectively.

Perspective of Beran

The author Beran (1995) demonstrated that in practical applications of Box-Jenkins autoregressive integrated moving average (ARIMA) models in his studies, where he showed a great deal of the number of times that the observed time series which in return must be differenced to achieve approximate stationarity and is usually determined by careful steps, but they largely remain informal, and analysis of the differenced series also remain in this way. Author suggested in his study which is later known as a semiparametric estimation method using autoregressive approximations.

Methodology:

1. **Estimate d:**
   - The study shows the usage of a semiparametric estimator like the log-periodogram regression or also known as Whittle estimator.

2. **AR Approximation:**
   - it fits an AR(p) model to the fractionally differenced series $(1-L)^d X_t$.

3. **Parameter Estimation:**

- This calculate the AR coefficients using standard methods.

Pros:

- The steps separates the estimation of d from the short-memory parameters.
- It can also help lessening the bias in the estimation of d because it shows short-memory components.

—

Summary of the FARIMA model

The FARIMA model provides a flexible and a smooth framework for modeling time series in which long memory characteristics are preserved in the datasets. As we deep dive, understanding its origin, properties, and estimation methods is crucial for working with the model and also very important to effectively capture the dynamics of such processes. Exact maximum-likelihood estimation also offers us a very precise parameter estimates process which may be intensive resource hunger computationally. Methods that are based on autoregressive approximations do provide practical alternatives for large datasets to work with smoothly.

3.12 Advanced Nonparametric Regression Techniques and the SEMIFAR Model

3.12.1 Fundamentals of Nonparametric Regression

Nonparametric regression is a flexible approach when we need to create a structure which shows the relationships between variables and without assuming a specific functional form for the regression function. We can say that it is not similar as the parametric models, in parametric model we can specify a method is applied for a fixed structure, and on the other hand the nonparametric methods let the data to determine the shape of the regression curve, making them suitable for capturing complex patterns.

In the context of time series analysis, the concept of nonparametric regression is used to estimate trends and smooth out noise in the data. However, when there is presence of dependent errors, particularly those exhibiting long-range dependence, pop up challenges for estimation and inference.

Kernel Estimators for Short-Range Dependent Errors

When the error terms in a regression model are short-range dependent (i.e., their autocorrelations decay exponentially at a faster rate), kernel estimators can be effectively used for nonparametric regression. Hence, the basic regression model is:

$$Y_t = m(X_t) + \varepsilon_t,$$

where:

- Y_t is the response variable,
- X_t is the predictor variable,
- $m(\cdot)$ is the unknown regression function,
- ε_t are the error terms with short-range dependence.

The Nadaraya-Watson Kernel Estimator

The Nadaraya-Watson kernel estimator $g(x)$ for a regression function $m(x)$ is introduced as follows:

The Nadaraya-Watson kernel regression estimator from the paper Bierens (1988) is defined by:

$$g(x) = \frac{\sum_{j=1}^{n} Y_j \, K\left(\frac{x - X_j}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)}$$

where:

- Y_j known as the dependent variable values,
- X_j represents the regressor of variable values,
- $K(\cdot)$ is the kernel function, this is generally chosen as a uni-modal function centered at zero for example a Gaussian function,

- h_n know as the bandwidth or window width which gives us the control of the smoothness of the estimator.

The discussed estimator provides a weighted average of the observed Y_j values, which along with provides weights depending on the proximity of the components X_j to x, as a result it is smoothing the function $m(x)$ without assuming a specific parametric form.

For short-range dependent errors, standard asymptotic results hold with slight adjustments to account for the dependence in variance calculations. The estimator as we know that remains consistent, and in the case of independent convergence rate shows the similar convergence rate.

Kernel Estimators for Long-Range Dependent Errors

As we have known that the Long-range dependence (LRD) in error terms implies that autocorrelations decay slowly, very often following a power-law decay. This distinguished characteristic significantly affects the components of kernel estimators. In the presence of LRD errors, the variance of the estimator rises at rate on the other hand the convergence rate slows down.

The error process ε_t can be modeled as a fractional Gaussian noise model with a previously discussed Hurst parameter H $(0.5 < H < 1)$. The autocovariance function behaves as the model below:

$$\gamma(k) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t+k}) \sim k^{2H-2}, \quad \text{as } k \to \infty.$$

Under LRD processes, the kernel estimator's variance is inflated due to the strong dependence among errors. As a result, the estimator may not acquire the same convergence rate as we have in the short-range dependent case, rather we can observe special techniques are required to adjust for this effect.

Model Selection and Local Polynomial Fitting

When it comes to selecting model we first have to define the Local polynomial, which is a fitting model that extends kernel methods by fitting a polynomial function within a local neighborhood around each and every point. This approach involves in reducing boundary bias and provides a well performing theoretical properties.

The local polynomial estimator of degree p is obtained by minimizing, as it is shown below by the model :

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{t=1}^{n} \left[Y_t - \beta_0 - \beta_1(X_t - x) - \cdots - \beta_p(X_t - x)^p\right]^2 K\left(\frac{X_t - x}{h}\right).$$

The estimated regression function at point x is $\hat{m}(x) = \hat{\beta}_0$.

Model Selection Considerations:

- **Bandwidth Selection (h)**: The bandwidth controls and influences the trade-off between bias and variance. A smaller h holds more local objectives but in return it increases variance. On the other hand a larger h smooths the function but may make bias appear.
- **Polynomial Degree (p)**: These are higher-degree polynomials as they can reduce bias but also the can make them increase variance and overfitting.

Cross-validation, plug-in methods, or information criteria are most common and are used widely used when it comes to selecting optimal bandwidth and polynomial degree.

Automatic Kernel Method

Automatic kernel methods has a goal to select the optimal bandwidth and kernel function without manual influence and manual tweaks and changes. These methods typically involve:

- **Data-Driven Bandwidth Selection**: There are techniques such as like cross-validation or plug-in methods which estimate the bandwidth that help minimizing the estimation error.
- **Adaptive Procedures**: Adaptive proceudure is a process that adjust bandwidth locally based on the data density or based on the data variability.

In the context of dependent data, especially with LRD errors, this mentioned automatic methods mainly need to check for the dependence structure to ignore and remove underestimating variance and also to avoid selecting sub-optimal bandwidths.

3.12.2 Local Polynomial Regression Introduction to SEMIFAR

Definition and Estimators of the SEMIFAR Model

The **SEMIFAR (Semiparametric Fractional Autoregressive)** is a model that combines nonparametric regression with fractional autoregressive models to estimate and capture the both trends and long-range dependence in time series data. This topic is deeply structured and studied by authors Beran & Feng in their prestigious papers Beran et al. (1999) & Beran and Feng (2002b). According to them, Time series tends to exhibit local and global trends in several areas. These trends are explained in statistical perspective with the help of polynomial regression, smooth bounded which are trends that are estimated nonparametrically, as these difference-stationary processes shows integrated ARIMA processes hence from this base we can write the model as:

$$Y_t = m(X_t) + u_t,$$

where:

- $m(\cdot)$, this known to be an unknown smooth function which depicts the trend,
- u_t, this is a stationary error term in the case of long-range dependence.

The error term u_t follows a fractional ARIMA process:

$$(1 - L)^d u_t = \varepsilon_t,$$

with:

- L known as the lag operator,
- d which is the fractional differencing at parameter $(0 < d < 0.5)$,
- ε_t known as white noise.

Estimation Steps:

1. **Estimate** $m(\cdot)$: this component is used as a local polynomial regression which can obtain $\hat{m}(x)$.
2. **Calculate Residuals**: $\hat{u}_t = Y_t - \hat{m}(X_t)$.

3. **Estimate** d: This is the application methods like the Whittle estimator to $\hat{u}_t$.

Asymptotic Results for Bias and Variance for the SEMIFAR Model

When there are existence of long-range dependence affects the asymptotic properties of the nonparametric estimator.

Bias:

The bias of the local polynomial estimator stays as similar in the case independent:

$$\mathrm{Bias}[\hat{m}(x)] = \frac{h^{p+1}}{(p+1)!} m^{(p+1)}(x)\mu,$$

where:

- h is the bandwidth,
- p is the degree of the polynomial,
- μ, known as a constant depending on the kernel function.

Variance:

Under LRD case, the variance tend to increases:

$$\mathrm{Var}[\hat{m}(x)] \approx \frac{\sigma^2}{nh^{2d} f_X(x)},$$

where:

- σ^2 is the variance of ε_t,
- $f_X(x)$ is the density of X_t at x,
- n is the sample size.

The factor h^{-2d} this component indicates that a variance is inflated becasue of the long-range dependence.

Asymptotic Mean Integrated Squared Error (MISE) for the SEMIFAR Model

The Mean Integrated Squared Error (MISE) is:

$$\text{MISE} = \int \left[\text{Bias}[\hat{m}(x)]^2 + \text{Var}[\hat{m}(x)]\right] dx.$$

When we conduct balancing the bias and variance terms and that leads to the optimal bandwidth:

$$h_{\text{opt}} \propto n^{-1/(2p+2d+1)}.$$

This is the formula which shows that the rate at which h reduces depends in case of both the polynomial degree p and the memory parameter d.

IPI-Bandwidth Selection Algorithm for the SEMIFAR Model

The **Iterative Plug-In (IPI) bandwidth selection algorithm** is designed to select the optimal bandwidth in the presence of LRD, process also studied thoruoughly by both authors, Beran et al. (1999) & Beran and Feng (2002b).

Algorithm Steps:

1. **Initial Estimation**:
   - First we choose an initial bandwidth h_0 and estimate $m(x)$ in order to obtain residuals $\hat{u}_t$.
2. **Estimate Memory Parameter** d:
   - we use $\hat{u}_t$ to estimate d via semiparametric methods.
3. **Update Bandwidth**:
   - We have to compute an updated bandwidth h_1 by using:

$$h_1 = \left(\frac{R(K)\sigma^2}{\mu_2^2 [m^{(p+1)}(x)]^2 n} \right)^{1/(2p+2d+1)},$$

     where:
     - $R(K)$ and μ_2 both are known as kernel-dependent constants,
     - σ^2, this is the estimated variance.

4. **Iteration**:
    - We have to repeat steps 1-3 using h_1 until we observe convergence.

Advantages:

- The main pros is it adjusts for the impact of LRD on variance.
- It provides a data-driven approach to bandwidth selection.

3.12.3 Estimation Methods for SEMIFAR

When it we have to work the estimation of the SEMIFAR model it involves combining the nonparametric regression techniques with the methods for estimating long-memory parameters.

Estimation Processes:

1. **Nonparametric Estimation of $m(x)$**:
    - First we must apply local polynomial regression with the bandwidth selected via the IPI algorithm.
2. **Residual Calculation**:
    - Then comes the process computing residuals $\hat{u}_t = Y_t - \hat{m}(X_t)$.
3. **Estimation of d**:
    - We have to use semiparametric estimators like the Whittle estimator or log-periodogram regression on $\hat{u}_t$ to estimate d.
4. **Model Checking**:
    - Important steps is to analyze residuals for independence and stationarity.
    - Then Performing diagnostic tests to validate the model.
5. **Iterative Refinement**:
    - If necessary, we have to refine the estimates of $m(x)$ and d iteratively accroding to goal of desired resutls.

Challenges:

- **Interdependence of $m(x)$ and d**: Estimating the trend and memory parameter simultaneously can be complex.
- **Computational Complexity**: The iterative procedures may be computationally intensive for large datasets.
- **Choice of Estimators**: Selecting appropriate methods for estimating d is crucial for accuracy.

3.12.4 Integrating Core Semiparametric Concepts

This section discussed the nonparametric regression methods, while focusing on kernel estimators and local polynomial fitting when we have short-range and long-range dependent errors present in the dataset .The SEMIFAR model was introduced as a semiparametric approach that as we can observe, effectively combines trend estimation along with the modeling long-range dependence by using fractional integration.

Key points include:

- **Impact of Dependence Structure**: As Long-range dependence significantly affects, the variance of estimators and which needed to be checked for in both case of estimation and bandwidth selection.
- **Bandwidth Selection**: The IPI-bandwidth selection algorithm gies us a systematic method for in the help of choosing the optimal bandwidth when LRD exists.
- **Estimation Challenges**: Simultaneously estimating the regression function and memory parameter needs a thorough and careful consideration when we have to ensure accurate modeling bases.

We must understand these concepts as it is essential for analyzing time series data that depicts complex trends and dependence structures within the datasets.

3.13 ACD Models: EFARIMA and ESEMIFAR

3.13.1 Extensions of the ACD Framework

Description of the ARCH Model

The **Autoregressive Conditional Heteroskedasticity (ARCH)** model was introduced by Engle (1982) to model time-varying volatility in financial time series. Previous we have discussed the ARCH and GARCH model in the context of short memory time series with the extensions of ARMA. Now the ARCH model which we will discuss the author Engle (1982) that captures the clustering of volatility, where we will observe large changes in asset returns and that tend to be followed by large changes, and on the other hand small changes tend to be followed by small changes.

The basic ARCH(q) model is defined as:

$$
\begin{aligned}
Y_t &= \mu + \epsilon_t, \\
\epsilon_t &= \sigma_t Z_t, \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2,
\end{aligned}
$$

where:

- Y_t is the asset return at the time t,
- μ is the mean return,
- ϵ_t is the residual term,
- σ_t^2 is the conditional variance,
- Z_t is known for sequence of independent and identically distributed (i.i.d.) random variables with zero mean and unit variance,
- $\alpha_0 > 0$ and $\alpha_i \geq 0$ are considered as the parameters which to be estimated.

Key Features:

- The model captures time-varying volatility when the modeling the conditional variance is involved as a function of past squared errors.
- This model assumes that volatility clusters over a time parameter duration.

Limitations:

- The cons can be the requirements as non-negativity constraints on parameters to ensure positive variance.
- This might require a large number of lags q to perfectly depict the volatility dynamics.

Description of the GARCH Model

The **Generalized Autoregressive Conditional Heteroskedasticity (GARCH)** model which ws initially discussed in the paper of Bollerslev (1986) as this extends the ARCH model by including the must observed significant lagged conditional variances in the model specification process. The GARCH(p, q) model is showed below as:

$$
\begin{aligned}
Y_t &= \mu + \epsilon_t, \\
\epsilon_t &= \sigma_t Z_t, \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,
\end{aligned}
$$

where:

- $\beta_j \geq 0$ are additional parameters whcih will let us observe the persistence of volatility.

Key Features:

- As we have more parsimonious than ARCH models when we try to model volatility clustering.
- The model incorporates both short-term (ARCH terms) and long-term (GARCH terms) in the process of capturing persistence in volatility.

Limitations:

- The cons are being, non-negativity constraints are needed to ensure positive variance.
- This might not be capturing leverage impacts or asymmetries in volatility.

Description of the ACD Model

The **Autoregressive Conditional Duration (ACD)** model, inaugurated by Engle and Russell (1998), this typically extends the ARCH/GARCH framework to a model where time durations between events are observed, such as the trades are observed in financial markets. The duration's X_t are known as positive random variables, and the ACD model particularly explained the conditional expectation of X_t given past information.

Hence the basic ACD model is defined as:

$$X_t = \psi_t \epsilon_t,$$
$$\psi_t = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{j=1}^{p} \beta_j \psi_{t-j},$$

where:

- X_t is the duration between events at time t,
- ψ_t is the conditional expected duration,
- ϵ_t is an i.i.d. error term along with unit mean,
- $\omega > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$ are known as parameters.

Key Features:

- These models the clustering of events over time, in the process of doing so we can capture the phenomenon where events occur in bursts.
- These are useful for high-frequency financial data analysis or UHF Data.

Limitations:

- The model needs positivity constraints on parameters.
- The model might not perfectly capture long-memory properties observed in the duration data.

A Few Extensions to the ACD Model

There are several extensions to the basic ACD model that have been proposed to address its limitations and amis to capture more complex dynamics attributes:

1. **Log-ACD Model:**

    - It applies a logarithmic transformation to the time duration to make sure positivity without applying parameter constraints.

    - This is defined as:

    $$\ln X_t = \ln \psi_t + \epsilon_t,$$

    where ϵ_t is known as a zero-mean error term. The Log-ACD Model is depicted here from the paper Bauwens and Giot (2000).

2. **Fractionally Integrated ACD (FIACD) Model:** The FICAD is introduced in the paper Jasiak (1999) where,

    - It incorporates long-memory effects by letting appear the differencing parameter which becomes a fractional.

    - The FICAD extended model captures the persistence in durations over long time horizons.

3. **Exponential ACD (EACD) Model:** In paper from Fernandes and Grammig (2006) discuss the EACD model as a foundational and fundamental model for structuring the durations between events. They acknowledge that the EACD model is useful, but while it has some mention able limitations in capturing certain empirical features in the case of duration data, for example of over dispersion and long memory effects. To make sure these limitations are appeared, they propose a new family of ACD models that generalize the EACD

    - This model uses an exponential specification to the model to precess the conditional duration.

    - Defined as:

$$\psi_t = \exp\left(\omega + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{j=1}^{p} \beta_j \psi_{t-j}\right).$$

where:

- ψ_t, which is the conditional expected duration at the parameter time t,
- ω, known as a constant term,
- α_i are parameters for the exogenous variables X_{t-i},
- β_j are parameters for the expected durations og lagged positions ψ_{t-j},
- X_{t-i} are exogenous variables,
- p and q are the orders of the autoregressive terms.

This term makes sure that $\psi_t > 0$ for all t, which is a must have condition when modeling durations.

These three extensions address that

- **Long Memory**: while incorporating fractional integration.
- **Positivity Constraints**: Through a logarithmic process or exponential transformations.
- **Flexibility**: By modeling nonlinear effects and accommodating the key feature which is heavy-tailed distributions.

3.13.2 EFARIMA in the Multiplicative Error Framework

Defining the EFARIMA Based on Multiplicative Error Model (MEM)

The **Fractionally Integrated Log-Autoregressive Conditional Duration (FI-Log-ACD)** model, also known to us as the **EFARIMA** model (Extended Fractionally Integrated Autoregressive Moving Average), which have combines features of long memory and the logarithmic transformation to the model's positive-valued time series and finally it does exhibiting long-range dependence. The papers by Beran et al. (2015) & Feng and Zhou (2015) addresse the key components and challenges of modeling duration series that shows the both long-range dependence and trends. Duration series shows us the time intervals in between consecutive events, for example financial transactions or meteorological occurrences can be predicted by the model adequately. Accurately modeling these durations is very importand and should be conducted with cautions for understanding the underlying processes and for making the proper and reliable forecasts.

Multiplicative Error Model (MEM):

- The MEM framework structures the model with non-negative time series data by specifying the observed data or series as the product of a conditional scale factor and an innovation term.
- For durations X_t:

$$X_t = \psi_t \epsilon_t,$$

  where:

  - ψ_t, known as the conditional expectation of X_t,
  - ϵ_t, known as a positive i.i.d. error term with unit mean.

Defining the EFARIMA Model:
The model have Parameters which are estimated by using methods adapted for fractionally integrated processes, such as Maximum Likelihood Estimation (MLE) or Whittle estimation also.

- By taking logarithms, the model becomes additive as follows:

$$\ln X_t = \ln \psi_t + \ln \epsilon_t.$$

- The conditional expectation $\ln \psi_t$ is modeled using the traditional model FARIMA specification to capture long memory:

$$\Phi(L)\left(1-L\right)^d\left(\ln \psi_t - \mu\right) \; = \; \Theta(L)\,\eta_t.$$

  where:

  - $\Phi(L)$ and $\Theta(L)$ are known to the model as polynomials in the lag operator L,

- d can be defined as the fractional differencing parameter ($0 < d < 0.5$),
- μ which is the mean of $\ln \psi_t$,
- η_t as a white noise error term.

Key Features:

- **Long Memory**: The fractional differencing operator $(1 - L)^d$ which is very effective to capture the long-range dependence in durations of the datasets.
- **Log Transformation**: The model makes ensuring steps as the positivity of the durations without imposing parameter constraints.
- **Flexibility**: The model combines ARMA dynamics with fractional integration.

Properties of the FI-Log-ACD or EFARIMA Model

Long-Range Dependence:

- As we can observe in the formula that the EFARIMA model captures long memory through the fractional differencing parameter d.
- The autocorrelation function (ACF) of $\ln X_t$ which depicts a rate of decaying at the rate on hyperbolically, hence;

$$\rho(k) \sim k^{2d-1}, \quad \text{as } k \to \infty.$$

Stationarity and Invertibility:

- The model is stationary when we have the values within $d < 0.5$ and the roots of $\Phi(L)$ lie outside the unit circle.
- In the model, invertibility requires the roots of $\Theta(L)$ which also lie outside the unit circle.

Marginal Distribution:

- The marginal distribution of X_t can illustrate heavy tails because of the multiplicative structure and long memory.

Estimation:

- The model's parameters can be estimated by using maximum likelihood or semi-parametric methods.
- The estimation of d is very significant because it adequately captures the long-range dependence.

Advantages:

- **Captures Persistence**: Capturing persistence is one of the pros because it properly models the persistent autocorrelations that are observed in duration data.
- **Positivity of Durations**: Log transformation secures that the modeled durations stay as positive.

Applications:

- When it comes to applying the model we can state that it is suitable for modeling financial durations, such as time between trades or price changes.
- The model can be applied to other positive-valued time series exhibiting long memory.

3.13.3 Semiparametric Fractional Log-ACD Analysis

Further the paper from Feng and Zhou (2015) has included some insights in addition to the information above. They contributed to the literature by introducing a semiparametric fractionally which is an integrated Log-ACD model that is very crucial for adding both short-term dynamics and long-range dependence in financial data which have impactful durations. Their empirical results show that the proposed model provides superior forecasting performance compared to traditional models. This work underscores the importance of incorporating long-memory features and adopting flexible modeling approaches in financial econometrics. Few steps to be noticed below as far as model goes;

1. Log Transformation

As we will observe in data that the durations δ_t are transformed by using the natural logarithm:

$$y_t = \log(\delta_t).$$

This transformation secures that the durations stay as positive and the variance stays steady.

2. Fractional Integration

The transformed durations y_t are modeled under using a fractionally integrated process:

$$(1 - L)^d (y_t - \mu) = \varepsilon_t.$$

where:

- L is the lag operator.
- d, the fractional differencing parameter ($0 < d < 0.5$ for stationarity).
- μ, defined as the mean of y_t.
- ε_t, known as a stationary error term.

3. Semiparametric Estimation

The authors have depicted that the error term ε_t is modeled semiparametrically, in the process it avoids strict distributional assumptions. This approach gives us great deal of flexibility in capturing the underlying data structure.

3.13.4 Examination and Interpretation of the ESEMIFAR Model

The **Semi-Fractionally Integrated Log-Autoregressive Conditional Duration (Semi-FI-Log-ACD)** model, also referred in many papers as the **ESEMIFAR** model (Extended Semiparametric Fractionally Integrated Autoregressive), as it extends the EFARIMA model when it includes a nonparametric component to model trends or nonlinear effects in the data. The paper from Letmathe et al. (2023) inaugurates an extended version of the Exponential Semiparametric Fractional Autoregressive (ESEMIFAR) model. This extended model is designed to effectively be used in positive-valued time series data that illustrate both short-range and long-range dependence, which are noticed very often in the fields like financial econometrics and environmental studies.

Model Specification:

- The structure shows that durations X_t are modeled as:

$$\ln X_t = m\left(\frac{t}{n}\right) + u_t,$$

where:

- $m(\cdot)$ is an unknown smooth function effectively capturing deterministic trends or also known as seasonal effects,
- u_t, which is a stationary process modeled by the traditional FARIMA model.

- Here the stationary object u_t is specified as:

$$\Phi(L)(1-L)^d u_t = \Theta(L)\eta_t,$$

with η_t known as white noise.

Key Features:

- **Semiparametric Approach**: The model combines nonparametric regression (for $m(\cdot)$) along with we can observe the parametric modeling of u_t.
- **Captures Nonlinear Trends**: The function $m(\cdot)$ allows for flexible modeling to catch the trends, cycles, or any seasonal patterns that exist in the datasets.
- **Long Memory**: The fractional differencing parameter d checks for long-range dependence in u_t.

Estimation Procedure:

1. **Nonparametric Estimation of $m(\cdot)$**:
   - We can depict that it takes into account the local polynomial regression or kernel smoothing techniques.
   - The bandwidth selection is necessary for balancing bias and variance.
2. **Estimation of d and Other Parameters**:
   - While detrending the data, we can estimate d by using semiparametric methods for example the Geweke-Porter-Hudak (GPH) estimator or the Whittle estimator.

  - We can estimate ARMA parameters $\Phi(L)$ and $\Theta(L)$ with the help of using standard techniques.

3. **Iterative Procedure**:
  - The estimation of $m(\cdot)$ and d may be iterated to improve accuracy as the prediction processes progress.

Advantages:

- **Flexibility**: This we can assume as one of the pros because the model captures complex data structures combining trends and long memory.
- **Improved Forecasting**: The methods accounts for both nonstationarity and long-range dependence, as the step progresses it enhances predictive performance in better accuracy.

Applications:

- This model is useful in financial econometrics for modeling ultra high-frequency data with trends and robust volatility.
- The model is applicable to any positive-valued time series with similar features.

Overall, the EFARIMA and ESEMIFAR models provide us with the powerful tools for modeling positive-valued time series data, which contain both long-range dependence and nonstationarity, and finally extending traditional ACD models, these tools offer us greater flexibility and improved modeling capabilities for much more complex and more compact financial data.

4 Implementation of the Models to the Datasets

4.1 Implementation on Economic data

In this section, three key monthly economic indicators from the United States are analysed:

1. **Producer Price Index (PPI)**, from January 1974 to December 2019.
2. **Consumer Price Index (CPI)**, from January 1947 to December 2019.
3. **Employment Level**, from January 1948 to December 2019.

These three monthly frequency data have been collected from FRED websites and they have initially been seasonally adjusted. All these three series are transformed using the natural logarithm to stabilize variance and facilitate interpretation in terms of growth rates. I applied a two-step modeling strategy:

- First, I have estimated a smooth time-varying trend via local linear regression (the `tsmoothlm` function from the `esemifar` R package).
- Second, I fit a FARIMA-type (Fractionally Integrated ARMA) model to the *detrended* residuals, thereby I have captured any long-memory dependence.

4.1.1 The relations among Producer Price, Consumer Price, and Employment

[1], [2], [3] The reason I have chosen these three data is because, from an economic standpoint, these three data are closely interlinked:

- **Producer Price (PPI)**: It reflects the average price changes that domestic producers receive for their goods. While rising PPI often precedes rising CPI because higher production costs can translate into higher retail prices.

[1] Employment level data; https://fred.stlouisfed.org/series/CE16OV
[2] Producer Price Index data; https://fred.stlouisfed.org/series/WPSFD4131
[3] Consumer Price Index data; https://fred.stlouisfed.org/series/CPIAUCSL

- **Consumer Price (CPI)**: It measures changes in the prices paid by urban consumers for a market basket of consumer goods and services. When I have sustained rises in CPI indicate inflation.
- **Employment**: This metric captures the total number of employed persons. As the economy grows, employment typically increases; that growth can place upward pressure on both producer and consumer prices when demand outstrips supply, leading to inflationary trends.

Time Series Plots

Log-Transformed Series

Figure 4.1 shows the three monthly economic indicators (in logs). It is noted here the strong upward trends in each series, though they begin at different dates.

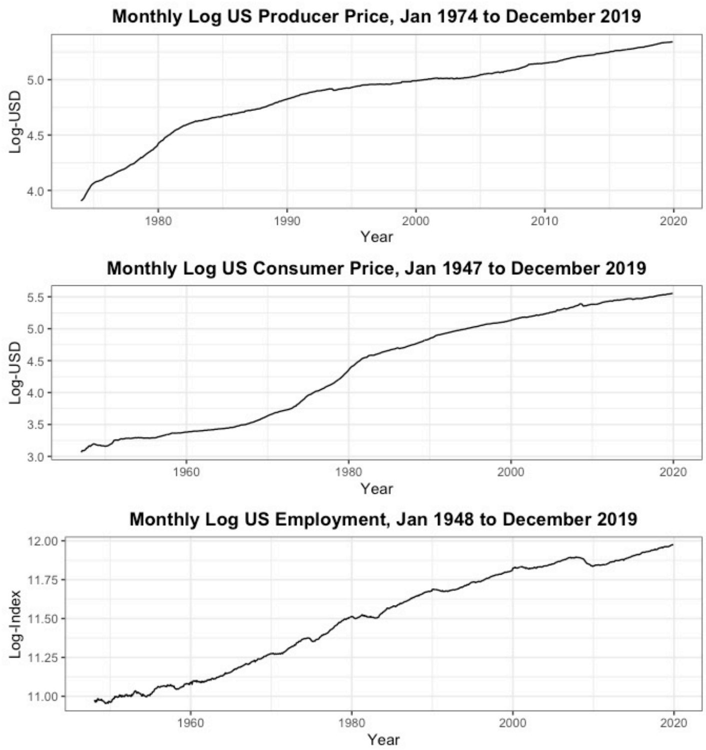

Figure 4.1: Monthly Log US Producer Price, Log US Consumer Price, and Log US Employment.

Trend Estimates

Next, I have extracted the trends using `tsmoothlm`. The resulting trends (in red) and observations (in grey) are displayed in Figure 4.2.

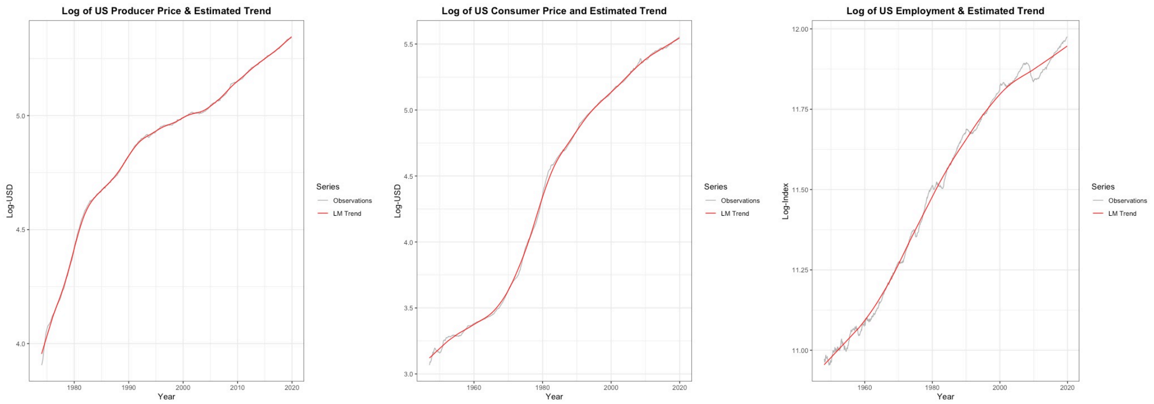

Figure 4.2: Side-by-side comparison of the log-transformed series and the estimated trends for (a) Producer Price, (b) Consumer Price, and (c) Employment.

FARIMA Model Summary

Table 4.1: FARIMA Model Summary for Log-Transformed US Economic Data (Trend-Adjusted).

Series	Time Span	FARIMA Model	$\hat{d}$	AR / MA Coeffs (Std. Error)
Producer Price (PPI)	1974–2019	FARIMA$(2, 0.243, 2)$	0.243	AR Coeffs: $\phi_1 = 1.788\,(\pm 0.0022)$ $\phi_2 = -0.815\,(\pm 0.0019)$ MA Coeffs: $\theta_1 = 1.149\,(\pm 0.0019)$ $\theta_2 = -0.289\,(\pm 0.0019)$
Consumer Price (CPI)	1947–2019	FARIMA$(3, 0.376, 0)$	0.376	AR Coeffs: $\phi_1 = 1.001\,(\pm 0.0026)$ $\phi_2 = -0.147\,(\pm 0.0026)$ $\phi_3 = 0.081\,(\pm 0.0025)$ MA Coeffs: None
Employment	1948–2019	FARIMA$(3, 0.282, 1)$	0.282	AR Coeffs: $\phi_1 = -0.197\,(\pm 0.0037)$ $\phi_2 = 0.754\,(\pm 0.0037)$ $\phi_3 = 0.327\,(\pm 0.0041)$ MA Coeff: $\theta_1 = -0.782\,(\pm 0.0035)$

The smooth trend is subtracted, as I fit FARIMA models to the residuals. Table 4.1 summarizes the parameter estimates.

Residual Diagnostics and Plots

I examine the FARIMA residuals to check for any remaining trends or autocorrelation. Figure 4.3 displays the residuals for Producer Price, Consumer Price, and Employment. Each series appears roughly stationary with no obvious trend, indicating that my combination of local linear smoothing plus FARIMA captures both the long-term trend and the dependence structure.

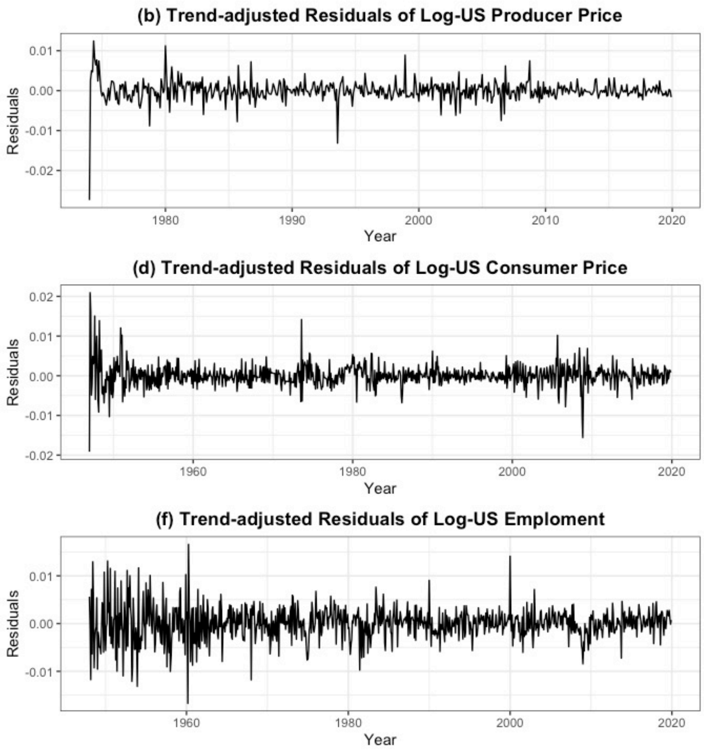

Figure 4.3: Trend-Adjusted Residuals of Log-US Producer Price, Consumer Price, & Employment.

Forecasting and Economic Insights

After obtaining the detrended FARIMA fits, I generated 20-step-ahead forecasts. Figures 4.4 shows point forecasts and 95%/99% prediction intervals for each series on the log scale.

Observations:

- **Producer Price:** Forecasts suggest continued gradual growth in PPI, indicating modest upward pressure on costs faced by producers.

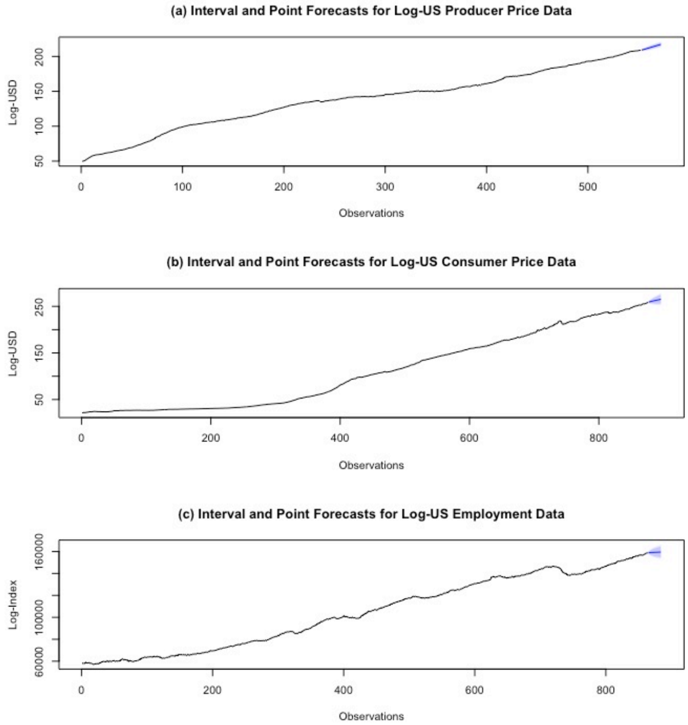

Figure 4.4: Interval and Point Forecasts for all log series

- **Consumer Price:** CPI forecasts predict ongoing inflation. The higher $\hat{d}$ implies that shocks to consumer prices can have somewhat more persistent effects on future values.

- **Employment:** Here we can see Employment also shows a positive trend, with short-term cycles captured by AR and MA terms. As labor market conditions tighten, higher employment can reinforce inflationary pressures in both PPI and CPI.

Overall, there are relations among these three data as the interplay among Producer Price, Consumer Price, and Employment reflects classic macroeconomic linkages: as growing employment fuels demand, raising both wholesale and retail prices; supply-side shocks feed into longer-run inflationary dynamics. We can say that the fractional differencing parameter $\hat{d}$ underscores the long-memory, which is already present, in the residual processes, meaning shocks do not fully dissipate over short time horizons.

4.2 Implementation on Financial data

The core steps I used here are:

1. **Data Preparation**: Filteing non-positive volumes and log-transform the series.
2. **Trend Extraction**: Using `tsmoothlm` with an optimal bandwidth (`InfR="Opt"`) to extract a smooth trend.
3. **FARIMA Modeling**: Fitting a fractional ARIMA model on the residuals, identify the differencing parameter d, and check for any AR/MA terms.
4. **Forecasting**: Combining linear trend extrapolation with FARIMA residual forecasts to generate point forecasts and 80%/95% intervals.
5. **Visualization**: Finally Displaying a six-panel figure of original vs. log scale, full vs. selected range, and stationary residual views.

4.2.1 FARIMA Modeling of NASDAQ ETFs Volume Data

This section we can see the *trend extraction plus FARIMA* approach for a monthly dataset of NASDAQ ETFs trading volumes, approximately covering January 2000 to December 2019.

[4]

Data Preparation and Log Transformation

Reading and Filtering

I loaded the Excel file, and select only positive volumes, All of the three financial data are collected by R scripts. I defined a pseudo-time axis (2000–2019) via

Trend Extraction with `tsmoothlm`

- $\hat{b}_{\text{opt}} = 0.1318$: optimal bandwidth,
- $\hat{y}_e(t)$: smooth log-volume trend,

[4] Financial Indices data have been extraced with R scripts, which are provided with the submission.

- $\varepsilon_t = \log(\text{Volume}) - \hat{y}_e(t)$: residuals.

FARIMA Modeling of Residuals

Parameter Estimates

From `results$FARMA.BIC`, I observe $\hat{d} \approx 0.4449$, with no AR or MA terms:

$$\text{FARIMA}(0,\ 0.4449,\ 0),$$

and a tiny standard error for d ($\approx 9.67 \times 10^{-6}$). I then confirm these estimates with `fracdiff` on the residuals.

Summary of Key Outputs

Table 4.2 highlights the main code outputs:

Table 4.2: Key Code Outputs for NASDAQ ETFs Data

Item	**Value / Estimate**	**Comment**
Optimal Bandwidth (b_{opt})	0.1317796	For `tsmoothlm`
Differencing Param (d)	0.4449251	FARIMA(0,d,0) selected by BIC
St. Error of d	9.67×10^{-6}	Very tight CI $\big(0.4449062,\ 0.4449441\big)$
AR order ($p.BIC$)	0	No AR terms
MA order ($q.BIC$)	0	No MA terms
σ_ε	0.2903131	Std. dev. of residuals
Log-likelihood	-917.1	AIC = 1838.182

Thus, here the residual series requires only fractional integration, with $\hat{d} \approx 0.445$. No autoregressive or moving-average components are selected.

Forecasting

I forecast $K = 50$ steps by:

1. Extrapolating the final slope of $\hat{y}_e(t)$ (log-trend) linearly,
2. Adding the FARIMA(0,d,0) forecast of the residuals ε_t,
3. Exponentiating to revert to original (volume) scale,

4. Computing 80%/95% intervals via `forecast(fracdiff,...)`.

Visualization and Key Observations

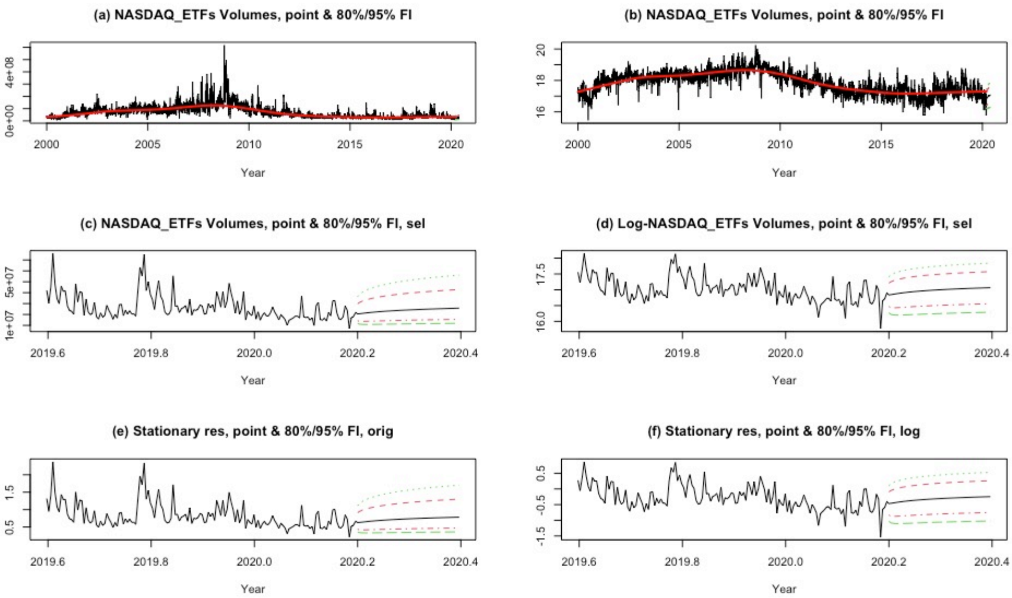

Figure 4.5: Six-panel figure displaying: (a) Original volumes + 80%/95% forecast intervals, (b) Same on log scale, (c) Zoomed subset (original scale), (d) Zoomed subset (log scale), (e) Stationary residuals in original scale, (f) Stationary residuals in log scale. The red line here denotes the estimated trend; green/red dashed lines show the 80%/95% intervals.

Key Observations from Figure 4.5:

- *(a), (b)*: The LM trend (red) tracks the overall log-volume shape, revealing a moderate decline after the mid-2000s peak, stabilizing afterward.
- *(c), (d)*: Over the recent subset, forecast intervals widen gradually, reflecting moderate uncertainty in the long-memory residual process.
- *(e), (f)*: Residuals appear mean-stationary with no obvious AR/MA dependence, consistent with FARIMA(0,d,0).
- The tight CI for d suggests that fractional differencing is highly significant, implying mild but persistent memory in volume fluctuations.

In summary, the `tsmoothlm` approach leaves a residual series best characterized by FARIMA(0,d,0), with $d \approx 0.445$. Forecast intervals expand modestly over 50 steps, indicating moderate long memory but minimal short-run autocorrelation. Future extensions could compare alternative models (ARIMA, random walk) or incorporate external market covariates to refine the forecasts.

4.2.1 FARIMA Modeling of Nikkei225 ETFs Volume Data

This Section applies our *trend extraction plus FARIMA* framework to a monthly dataset of Nikkei225 ETFs trading volumes from January 2000 to December 2019. Any zero or negative volumes were excluded prior to analysis.

Data Preparation and Log Transformation

Reading and Filtering

I load the data file and kept only positive volumes. A pseudo-time axis (2000–2019) is defined as:

$$\text{Year}_t = 2000 + \frac{t}{n} \times 20, \quad t = 0, \ldots, n-1.$$

Trend Extraction with `tsmoothlm`

Using:

I obtained:

- $\hat{b}_{\text{opt}} \approx 0.1697$ as the optimal bandwidth,
- $\hat{y}_e(t)$ as the smooth log-volume trend,
- $\varepsilon_t = \log(\text{Volume}) - \hat{y}_e(t)$ as the residuals.

FARIMA Modeling of Residuals

Parameter Estimates

From the BIC-selected model, I have:

$$d \approx 0.3414, \quad \text{FARIMA}(0, d, 0),$$

with a small standard error $\approx 2.07 \times 10^{-5}$. Table 4.3 summarizes key results from the code output. Fitting ε_t via `fracdiff` confirms no AR/MA terms are selected; only fractional differencing is needed.

Table 4.3: Key FARIMA Model Outputs for Nikkei225 ETFs Residuals.

Item	**Value / Estimate**	**Comment**
Optimal Bandwidth (b_{opt})	0.1696865	from `tsmoothlm`
Differencing Param (d)	0.3413805	FARIMA(0,d,0) chosen
St. Error of d	2.07×10^{-5}	95% CI: (0.34134, 0.34142)
AR order ($p.BIC$)	0	No AR terms
MA order ($q.BIC$)	0	No MA terms
σ_ε	0.5042035	Standard deviation of residuals
Log-likelihood	≈ -1967	AIC $\approx$ 3938.42

Forecasting and Combining with Trend

To forecast $K = 50$ steps ahead, I:

1. Linearly extrapolate the final slope of the log-trend $\hat{y}_e(t)$.
2. Forecast the residual series via FARIMA(0,d,0).
3. Add the extrapolated trend and residual forecasts in the log scale.
4. Exponentiate to return to original (volume) scale.
5. Obtained 80% and 95% intervals using `forecast(...)`.

Visualization and Observations

Figure 4.6 presents the resulting plots. **Key observations**:

- *(a), (b)*: The log-trend (red) captures the general downward drift in volumes since the early 2000s.
- *(c), (d)*: Zooming in on a recent subrange, forecast intervals expand gradually, reflecting the fractional differencing influence.
- *(e), (f)*: Residuals appear stationary with no visible AR/MA structure, validating the FARIMA(0,d,0) fit.
- Perfect Differencing parameter $d \approx 0.34$ indicates long-memory behavior, but not enough to require AR or MA terms.

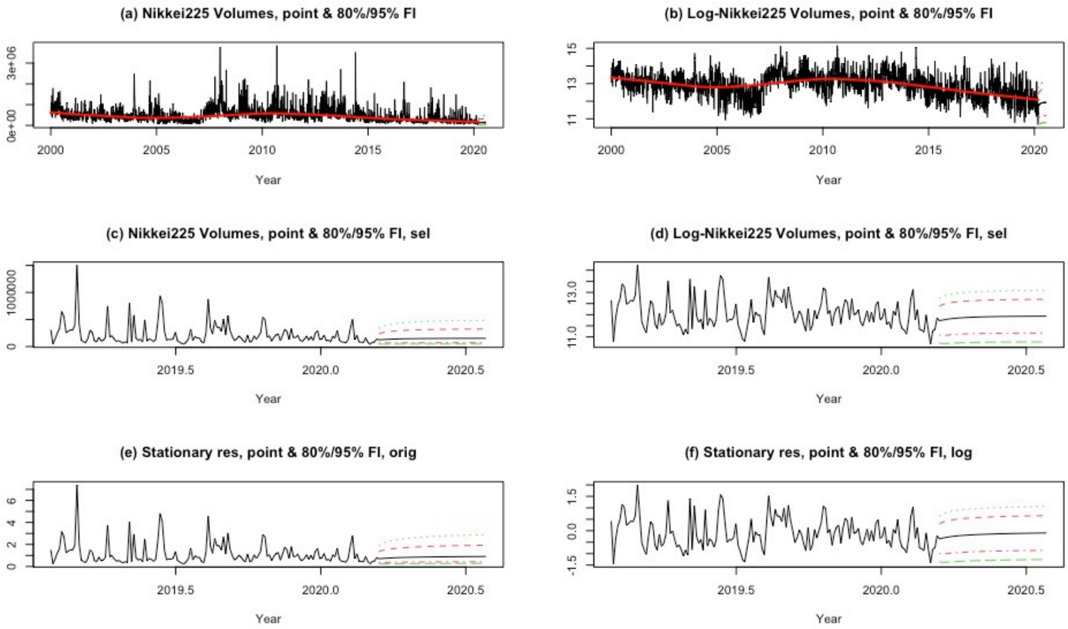

Figure 4.6: Six-panel figure for Nikkei225 Volumes: (a) original scale & 80%/95% forecast intervals, (b) log scale & trend (red), (c) selected subrange (original scale), (d) selected subrange (log scale), (e) stationary residuals (original scale), (f) stationary residuals (log scale). Dashed green/red lines indicate 80%/95% intervals.

In summary, the `tsmoothlm` approach for log-volumes leaves a stationary residual series best described by FARIMA(0,d,0) with $d \approx 0.34$. Forecast intervals widen marginally over the 50-step horizon, suggesting moderate but not strong memory.

4.2.1 FARIMA Modeling of S&P500 ETFs Volume Data

In this section, I applied simlilar to previously, *trend extraction plus FARIMA* methodology to **S&P500 ETFs** trading volumes. The dataset covers roughly January 2000 to December 2019, with non-positive volumes removed.

Data Preparation and Log Transformation

Reading and Filtering

Similar to prvious data handling, herr I also loaded data and kept only positive values. I defined a pseudo-time axis spanning 20 years (2000–2019):

FARIMA Modeling of Residuals

Parameter Estimates

From `results$FARMA.BIC`, I have $d \approx 0.4461$ with no AR/MA terms, i.e. FARIMA$(0, d, 0)$. The standard error is about 9.83×10^{-6}, yielding a narrow confidence interval $\big(0.44607,\ 0.44610\big)$. Fitting the residuals ε_t via `fracdiff` confirms these parameter values.

Model Summary

Table 4.4 lists key findings from the R output:

Table 4.4: Key FARIMA Outputs for S&P500 ETFs Data

Item	**Value / Estimate**	**Comment**
Optimal Bandwidth (b_{opt})	0.1635284	from `tsmoothlm`
d (FARIMA param)	0.4460846	FARIMA(0, d, 0)
Standard Error of d	9.83×10^{-6}	95% CI: $(0.44607,\ 0.44610)$
AR order ($p.BIC$)	0	no AR terms
MA order ($q.BIC$)	0	no MA terms
σ_ε	0.2911448	std. dev. of residuals
Log-likelihood	-931.7	AIC = 1867.355

Forecasting

To predict $K = 50$ steps:

1. Extrapolate the *final slope* of $\hat{y}_e(t)$ (log-trend).
2. I forecasted FARIMA(0,d,0) residuals over K steps.
3. I added them in log scale, exponentiate to original units.
4. It generates 80% and 95% prediction intervals using `forecast(...)`.

Visualization and Observations

Figure 4.7 illustrates the fitted trend and forecasts:

- **Panels (a), (b)**: The log-trend (red) aligns with the main volume changes, peaking around 2010–2011 and gradually declining afterward.

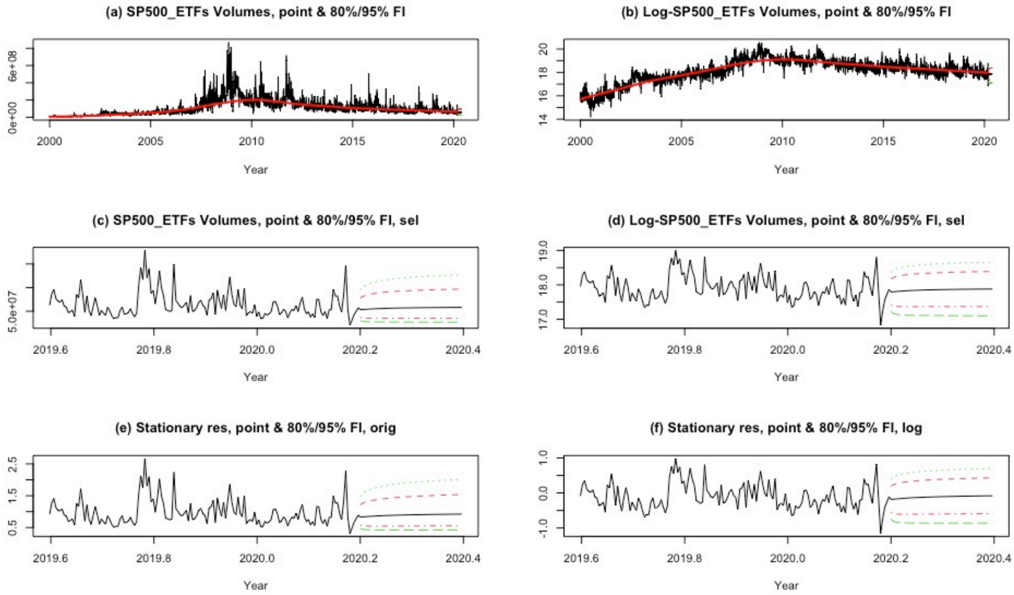

Figure 4.7: Six-panel figure: (a) Original volumes + 80%/95% forecast intervals, (b) Log volumes + smooth trend (red), (c) Subrange (original scale), (d) Subrange (log scale), (e) Residuals in original scale, (f) Residuals in log scale. Dashed green/red lines show 80%/95% intervals.

- **Panels (c), (d)**: In the selected recent subrange, forecast intervals expand gently, consistent with the moderate fractional differencing parameter.
- **Panels (e), (f)**: Residuals appear stationary with minimal AR/MA structure, supporting FARIMA(0,d,0).
- The differencing parameter $d \approx 0.446$ suggests persistent but not overly strong long memory.

For S&P500 ETFs volume data, `tsmoothlm` produces a detrended residual series well described by FARIMA(0,d,0), with $d \approx 0.446$. The forecast intervals widen slowly over 50 steps, reflecting stable volumes and moderate fractional integration.

4.2.2 Comparison of ETF Volume Results: S&P500, Nikkei225, and NASDAQ

In this section, I summarized and compared the outcomes of our *trend extraction plus FARIMA* analyses for three major ETF volume datasets: **S&P500**, **Nikkei225**, and **NASDAQ**. Each dataset is presented in a six-panel figure labeled (a)–(f), illustrating:

- *(a)* Original-scale volumes with point and 80%/95% forecast intervals.

- *(b)* Log-scale volumes plus the smooth trend (red).
- *(c)* Recent subrange (original scale) forecast.
- *(d)* Recent subrange (log scale) forecast.
- *(e)* Stationary residual series (original scale).
- *(f)* Stationary residual series (log scale).

S&P500 ETFs (Figure 4.7)

- **Trend Extraction:** Panels (a) and (b) show a pronounced rise in log-volumes peaking near 2010–2011, followed by a gradual decline. The red `tsmoothlm` trend captures the main structural changes in volumes.
- **Recent Subrange Forecast:** Panels (c) and (d) zoom in on the most recent period. The green/red dashed lines (80%/95% forecast intervals) expand slightly but they do not show abrupt changes.
- **Residual Stationarity:** Panels (e) and (f) confirm that subtracting the log-trend leaves a near-stationary process (FARIMA$(0, d, 0)$ with $d \approx 0.446$). Shocks persist moderately, but no AR/MA terms are needed.

Nikkei225 ETFs (Figure 4.6)

- **Trend Extraction:** Panels (a) and (b) reveal a largely downward or subdued trend in log-volumes over the 2000–2019 span. And the red curve (LM approach) highlights a gentle decline.
- **Short-Horizon Forecasts:** Panels (c) and (d) focus on the tail end of the series. The intervals widen but remain moderate, reflecting a fractional differencing parameter $d \approx 0.34$.
- **Stationary Residuals:** Panels (e) and (f) show the detrended volumes, now largely stationary. We see no strong AR/MA patterns are evident, consistent with FARIMA$(0, d, 0)$.

NASDAQ ETFs (Figure 4.5)

- **Trend Extraction:** Here panels (a) and (b) indicate volumes peaked in the mid-to-late 2000s, then trended downward. The local linear (LM) trend in red offers a smooth representation of this evolution.
- **Forecast Intervals (Recent Data):** The panels (c) and (d) again show modestly expanding forecast bands for the short-term horizon, with no steep changes predicted.
- **Residuals and Long Memory:** Panels (e) and (f) confirm the stationarity of the detrended data. The best-fitting model is FARIMA$(0, d, 0)$ with $d \approx 0.44$, meaning moderate long memory without AR/MA components.

Overall Observations and Takeaways

- **Trends:** Each of the three ETF volumes (S&P500, Nikkei225, NASDAQ) exhibits a significant rise at some point in the 2000–2010 range, followed by a decline or stabilization.
- **FARIMA$(0, d, 0)$:** In all of these three cases, the optimal model suggests only *fractional differencing* with no AR/MA terms. This points to moderate long memory in the residual processes.
- **Forecast Intervals:** For all three datasets, we see the 80%/95% bands widen *gradually*, indicating that while volumes are somewhat persistent, the short-horizon uncertainty remains contained once the main trend is removed.

4.3 Implementation on Environmental data

For three Environmental data we can assume,

- X_t denote the raw monthly data points,
- $X_t^{(\mathrm{SA})}$ the seasonally adjusted (SA) series obtained via STL decomposition,
- $\hat{y}_e^{(\mathrm{LM})}(t)$ or $\hat{y}_e^{(\mathrm{SM})}(t)$ the LM or SM trend estimates,
- $\varepsilon_t^{(\mathrm{LM})}$ and $\varepsilon_t^{(\mathrm{SM})}$ the residuals after subtracting each respective trend from $X_t^{(\mathrm{SA})}$.

4.3.1 Implementing Models to Natural Gas Consumption impact data

Reading and Filtering the Data

In this section, I have applied a similar modeling strategy as in the economic data, but now to an *environmental* dataset. Specifically, I analyze monthly natural gas consumption from January 2000 to December 2019. The data was not sesasonally adjusted. I loaded the monthly natural gas consumption data from an Excel file and restrict attention to the period January 2000 through December 2019. I then converted the observations to a monthly time series object with frequency 12.

5

STL Decomposition and Seasonally Adjusted Series

I employ STL (Seasonal and Trend decomposition using Loess) to isolate and remove the seasonal component. Figure 4.8 shows the seasonally adjusted data over the specified time range.

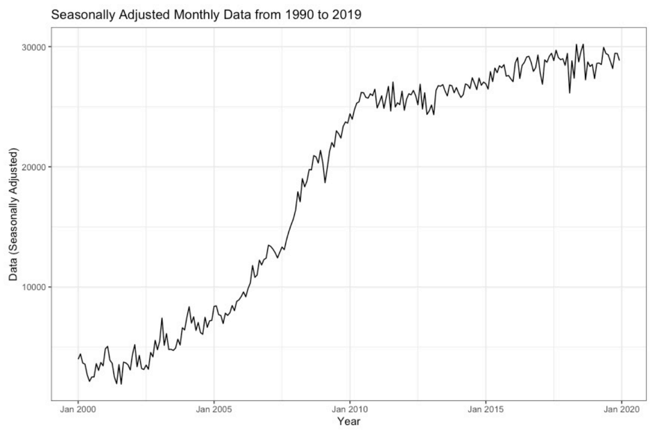

Figure 4.8: Seasonally Adjusted Monthly Natural Gas Consumption, 2000–2019.

Trend Extraction: LM vs. SM

I next used two smoothing approaches: LM Smoother, via `tsmoothlm`, which simultaneously estimates a fractional ARMA model on the residuals, and SM Smoother via `tsmooth`. Figure 4.9 plots the seasonally adjusted data in black and overlays both the LM (red) and SM (blue) trend estimates.

[5] Natural Gas Consumption data; https://fred.stlouisfed.org/series/NATURALGASD11.

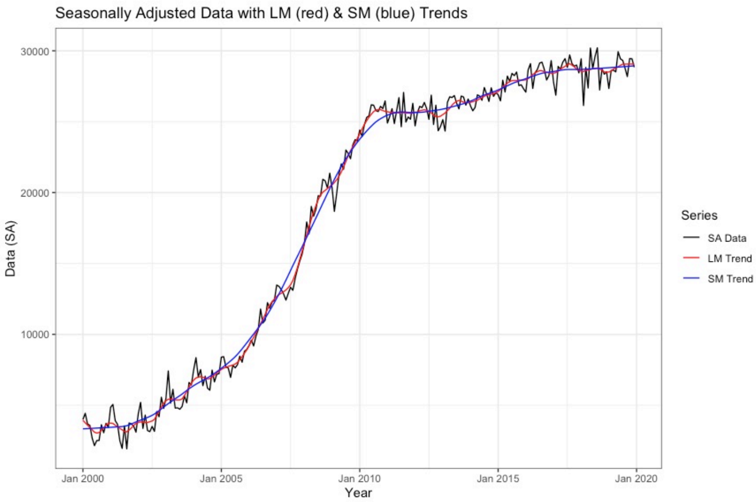

Figure 4.9: Seasonally Adjusted Data with LM (red) and SM (blue) Trend Estimates.

Residual Diagnostics

Subtracting each estimated trend from the seasonally adjusted series yields residuals:

$$\varepsilon_t^{(\mathrm{LM})} \;=\; X_t^{(\mathrm{SA})} - \hat{y}_e^{(\mathrm{LM})}(t), \quad \varepsilon_t^{(\mathrm{SM})} \;=\; X_t^{(\mathrm{SA})} - \hat{y}_e^{(\mathrm{SM})}(t).$$

Figure 4.10 compares the residual series for LM (top panel) and SM (bottom panel).

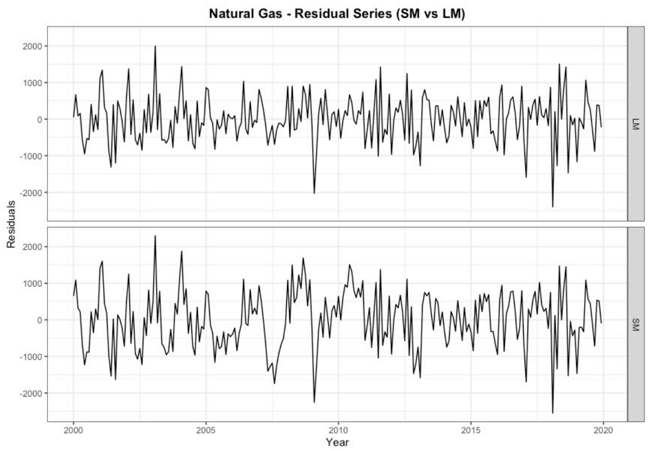

Figure 4.10: Natural Gas Residual Series: LM vs. SM Approaches.

FARIMA Modeling and Forecasting (LM Trend)

Using the LM approach, `tsmoothlm` internally fits a FARIMA model to the residuals. The output suggests:

```
$d = 0.03475764
$ar = (1.2912890, -0.3342144, -0.2351624)
$ma = (1.6287819, -0.6677783)
```

indicating a small fractional differencing parameter ($\hat{d} \approx 0.035$).

FARIMA Residuals

The residuals after fitting the FARIMA model on $\varepsilon_t^{(\mathrm{LM})}$ appear relatively stationary without strong autocorrelation. we can also inspect the ACF/PACF or periodogram to confirm no remaining seasonality.

Forecasting with FARIMA + LM Trend

I generated a 10-step-ahead forecast by

1. Forecasting the FARIMA residuals forward,
2. Extrapolating the LM trend as a linear continuation,
3. Summing the two forecasts to get the final predictions of $\hat{X}_t^{(\mathrm{SA})}$.

Figure 4.11 shows the seasonally adjusted data (in black) plus the forecast region (blue ribbon for 95% prediction intervals).

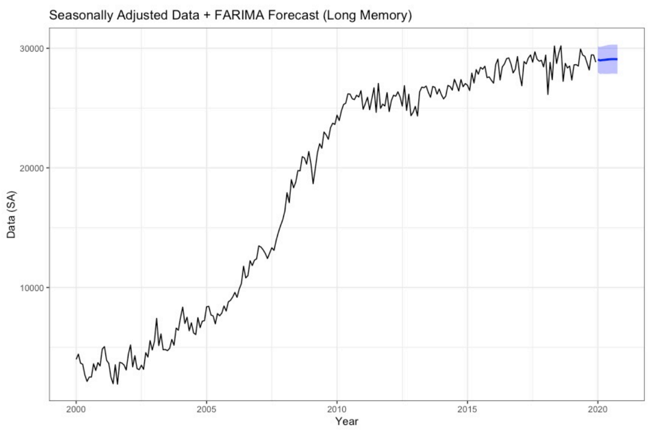

Figure 4.11: Seasonally Adjusted Data + FARIMA Forecast (LM Long Memory).

Observations:

- The fractional differencing parameter $\hat{d} \approx 0.035$ is relatively small, suggesting only mild long-memory effects in the residual series after detrending.

- The linear extrapolation of the LM trend shows us a gentle upward slope, consistent with historical growth rates for natural gas consumption.
- Prediction intervals widen moderately for the 10-month horizon, reflecting uncertainty in both the residual process and the trend continuation.

By applying STL-based seasonal adjustment, then extracting a LM or SM trend, and finally fitting a FARIMA model to the residuals, I capture both the *seasonal*, *trend*, and *long-memory* aspects of the data.

1. STL effectively removes the strong monthly/annual seasonality.
2. The LM or smoothed (SM) trend captures structural growth patterns.
3. FARIMA accommodates potential long-memory behavior in the detrended, deseasonalized residuals.

The resulting forecasts can help us assess near-term trends in natural gas consumption, with intervals that reflect both short-term fluctuations and long-run dependencies. From the environment perspective, future research could explore alternative robust trend estimators, incorporate external covariates (e.g., weather, economic indicators), or this allows for structural breaks in the trend if policy or technological shifts alter consumption patterns.

4.3.2 Implementation of the Models to Vehicle Miles Driven

Data Preparation

In this section, I applied the same modeling framework to a monthly time series of vehicle miles driven in the United States. The dataset spans from January 2000 to December 2019. Let I begin by reading the data from an Excel file and selecting observations from January 2000 to December 2019. We then convert these into a `ts` object with monthly frequency.

[6]

[6] Vehicle Miles driven; https://fred.stlouisfed.org/series/TRFVOLUSM227SFWA.

STL Decomposition

Using `stl(..., s.window = "periodic")`, we remove the seasonal component. Figure 4.12 shows the resulting seasonally adjusted vehicle miles.

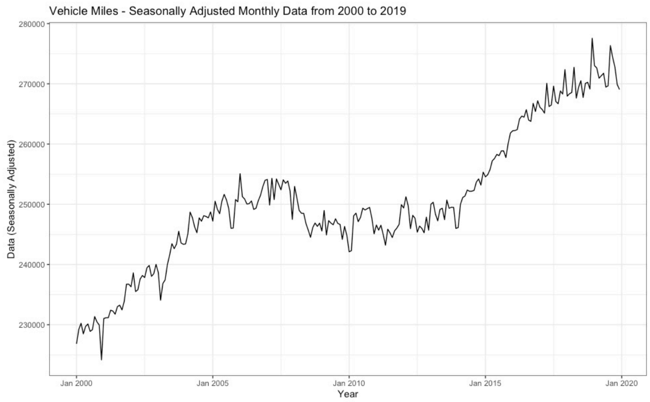

Figure 4.12: Vehicle Miles – Seasonally Adjusted Monthly Data from 2000 to 2019.

Trend Extraction: LM vs. SM

Next, we estimate two smooth trends on the seasonally adjusted series $X_t^{(\mathrm{SA})}$:

1. **LM (Local Memory) Trend** via `tsmoothlm`,
2. **SM (Simple/Nonparametric) Trend** via `tsmooth`.

Figure 4.13 displays the seasonally adjusted data (black) with the LM trend (red) and SM trend (blue).

Residual Diagnostics

Here I Subtracted each trend from $X_t^{(\mathrm{SA})}$ yields two sets of residuals:

$$\varepsilon_t^{(\mathrm{LM})} = X_t^{(\mathrm{SA})} - \hat{y}_e^{(\mathrm{LM})}(t), \quad \varepsilon_t^{(\mathrm{SM})} = X_t^{(\mathrm{SA})} - \hat{y}_e^{(\mathrm{SM})}(t).$$

Figure 4.15 compares the LM and SM residuals. Both appear roughly mean-stationary with no major seasonal pattern remaining.

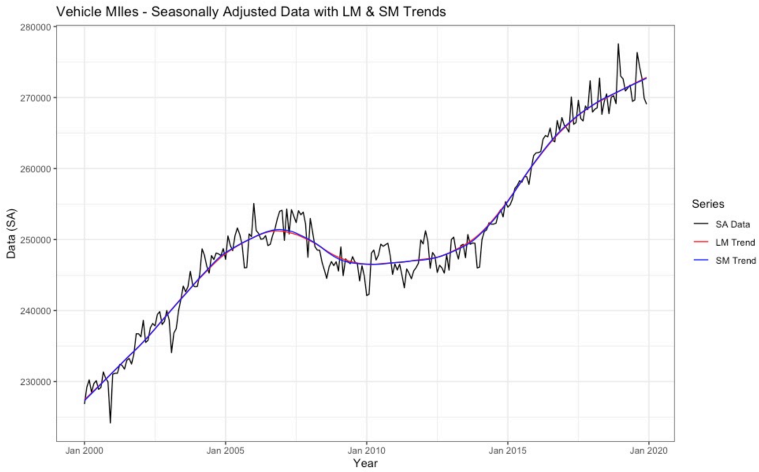

Figure 4.13: Vehicle Miles – Seasonally Adjusted Data with LM (red) & SM (blue) Trends.

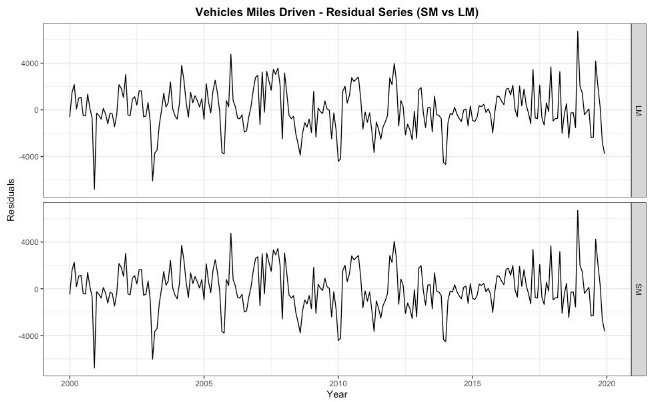

Figure 4.14: Vehicles Miles Driven – Residual Series (SM vs. LM).

FARIMA Modeling (LM Trend)

Parameter Estimates

`tsmoothlm` automatically attempts to fit a fractional ARMA model to $\varepsilon_t^{(\mathrm{LM})}$. From the console output:

```
$d = 0.3818971
$ar = (-0.1239251, -0.6833763, 0.2739873)
$ma = (-0.2181509, -0.9191390, 0.4473195)
```

Hence we have a FARIMA(3, 0.382, 3)-type specification with a moderately large fractional differencing parameter

FARIMA Residuals

After fitting the FARIMA model on $\varepsilon_t^{(\mathrm{LM})}$, we obtain FARIMA residuals that appear largely stationary and free of strong autocorrelations. One may verify this via the ACF/PACF or a periodogram. Below FARIMA residual is shown,

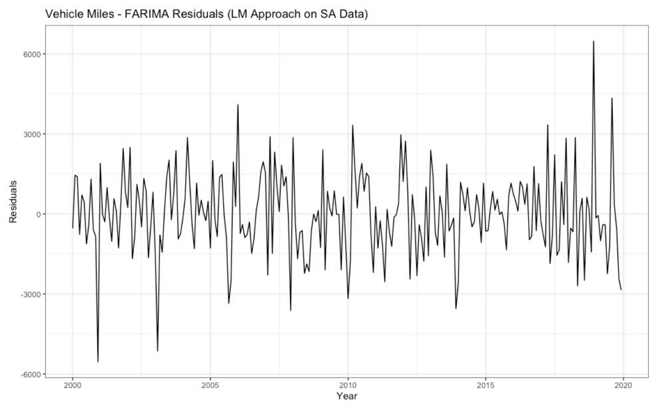

Figure 4.15: Vehicles Miles Driven – FARIMA Residual Series

Forecasting with FARIMA + LM Trend

Forecast Procedure

I forecasted 10 months ahead as follows:

1. Forecast the FARIMA residuals $\varepsilon_t^{(\mathrm{LM})}$,
2. LM trend $\hat{y}_e^{(\mathrm{LM})}(t)$ using its last slope,
3. Sum the trend forecast and the FARIMA residual forecast to obtain $\hat{X}_t^{(\mathrm{SA})}$.

Forecast Results

Figure 4.16 shows the final forecast (blue line) with 95% prediction intervals (blue shaded region) overlaid on the seasonally adjusted data.

Observations:

- The fractional differencing parameter $\hat{d} \approx 0.38$ which indicates the residuals within the long-memory behavior.
- The upward slope of the LM trend from 2015 to 2019 continues into the forecast horizon, which is suggesting ongoing growth in vehicle miles driven.

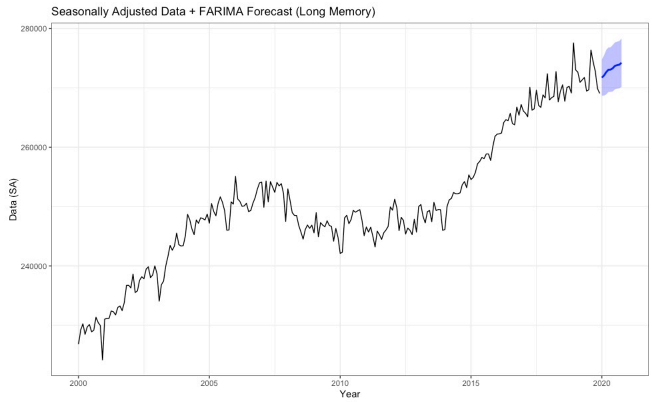

Figure 4.16: Seasonally Adjusted Data + FARIMA Forecast (Long Memory).

- Prediction intervals widen gradually, reflecting both the short-term fluctuations (captured by the AR/MA components) and the uncertainty around $\hat{d}$.

While Using STL to remove seasonal effects, smoothing (LM or SM) for trend extraction, and a FARIMA model for residual dynamics provides a flexible approach for vehicle miles driven data. The notable fractional differencing parameter shows that shocks or deviations in driving behavior may persist longer than in a purely short-memory framework.

Overall, these results align well with the general economic trend of increasing driving activity and underscore the utility of combining seasonal adjustment, smoothing, and FARIMA to capture both long-run and short-run dynamics.

4.3.3 Implementation of the Models to Waste Management

In this section, I applied the same methodology (trend extraction plus FARIMA modeling) to a monthly waste management dataset, measured in USD per ton from January 1990 to December 2019. The steps include:

1. Reading and filtering the data,
2. Estimating smooth trends via `tsmoothlm` (LM) and `tsmooth` (SM),
3. Inspecting the resulting residuals,
4. Fitting a FARIMA model for potential long-memory behavior,
5. Generating point and interval forecasts.

Data Preparation

Reading and Filtering

I loaded the dataset from an Excel file and restrict our attention to observations from January 1990 through December 2019. After converting the `observation_date` column to a `Date` object, I created a monthly time series (`ts` with frequency = 12).

7

Figure 4.17 shows the raw monthly time series.

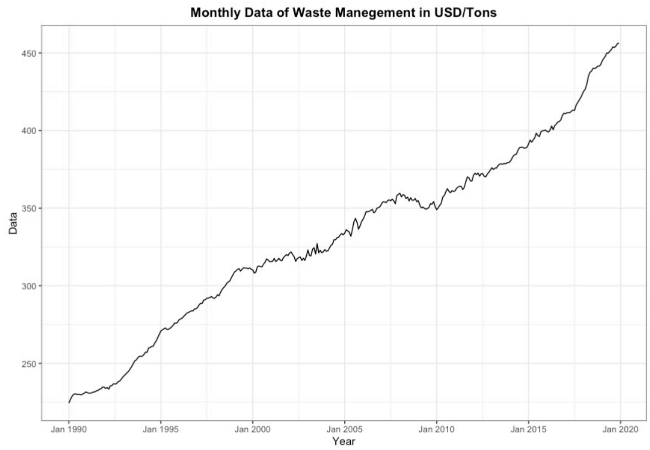

Figure 4.17: Monthly Data of Waste Management in USD/Tons, 1990–2019.

Trend Extraction with LM and SM

Fitting `tsmoothlm` and `tsmooth`

I consider two smoothing approaches for the raw data X_t:

- **LM Approach**: The `tsmoothlm()` function estimates both a trend and a fractional ARMA model for the detrended residuals.
- **SM Approach**: The `tsmooth()` function uses local polynomial smoothing without fitting a FARIMA model to the residuals.

Figure 4.18 plots the original data (gray) overlaid with the SM trend (blue) and the LM trend (red).

[7] Waste Management data; https://fred.stlouisfed.org/series/CES6056200001.

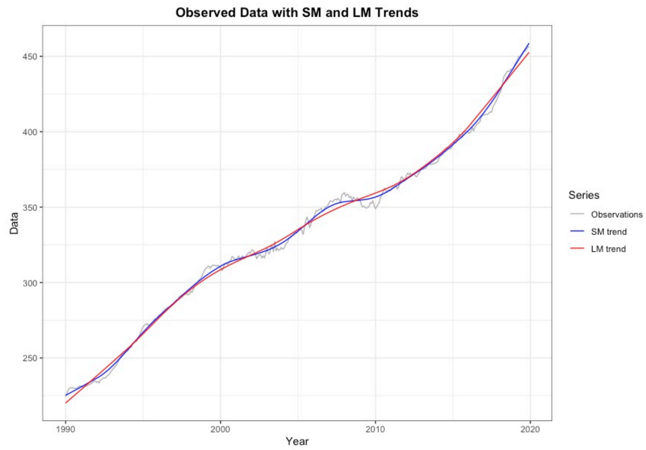

Figure 4.18: Observed Data with SM and LM Trends.

Residual Inspection

Subtracting each estimated trend from the raw series yields two sets of residuals:

$$\varepsilon_t^{(\mathrm{SM})} = X_t - \hat{y}_e^{(\mathrm{SM})}(t), \quad \varepsilon_t^{(\mathrm{LM})} = X_t - \hat{y}_e^{(\mathrm{LM})}(t).$$

Figure 4.19 shows the residuals from SM (bottom panel) and LM (top panel). Both appear mean-stationary with no strong seasonal component.

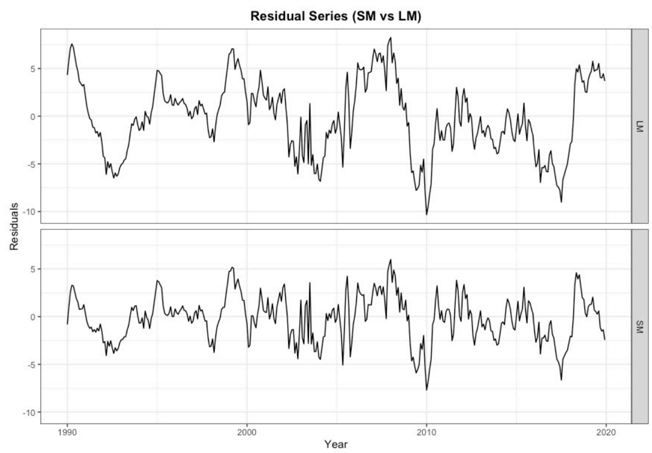

Figure 4.19: Residual Series (SM vs. LM).

FARIMA Modeling on LM Residuals

Fractional Integration Parameters

From `tsmoothlm`, I obtain a FARIMA model for the LM residuals. The output indicates:

```
d = 0.4623
ar = (1.1673, -1.1065, 0.7776)
ma = (0.7933, -1.2240, 0.6440)
```

Hence, I have something akin to FARIMA(3, 0.4623, 3).

Residual Diagnostics

I can inspect the FARIMA residuals for any leftover patterns. Figure 4.20 shows the final residual series after fitting the FARIMA model to the LM-based detrended data. Generally, here we can see, no large-scale structure remains, which suggesting that the model is capturing both the trend and the dependence in the data.

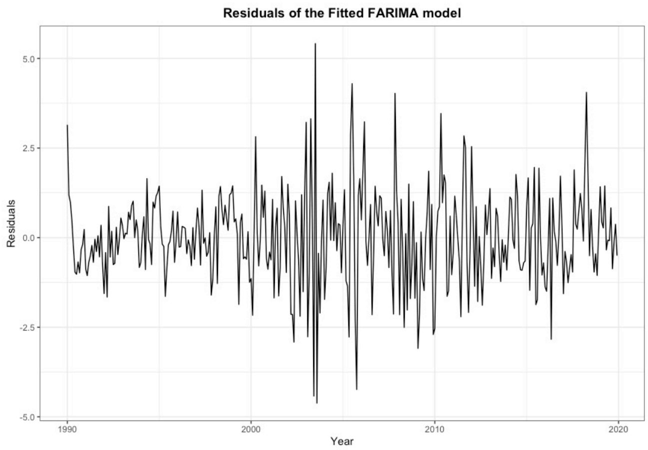

Figure 4.20: Residuals of the Fitted FARIMA Model.

Forecasting

To forecast h steps ahead, I Extrapolate the LM trend linearly based on its last slope, The predict the FARIMA residuals forward are shown, and the sum these two forecasts for the final predictions. I, then plotted point forecasts with a 95% interval.

Forecast Results

Figure 4.21 depicts the observed data (black), the forecasted path (blue line), and the associated 95% interval (blue ribbon).

Key Points:

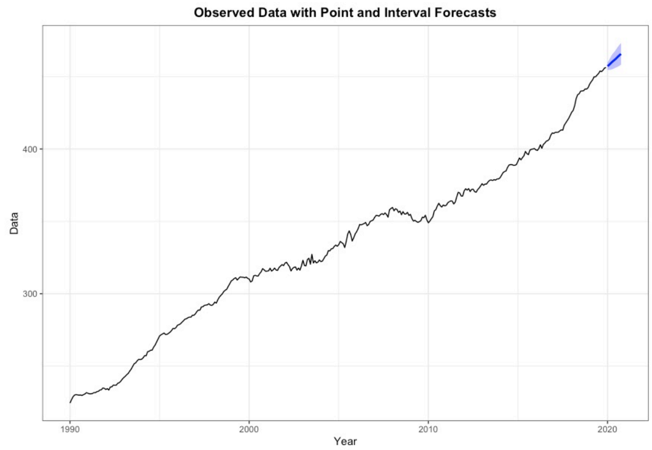

Figure 4.21: Observed Data with Point and Interval Forecasts.

- The estimated differencing parameter $d \approx 0.4623$ is above 0.5's threshold for stationary short-memory, suggesting the possibility of persistent long-memory-like behavior.
- The linear extrapolation of the LM trend captures the ongoing growth pattern observed in the data.
- Here widening intervals reflect uncertainty in both the fractionally integrated residuals and we see here the assumed linear extension of the trend.

In summary, I estimated a smooth trend (`tsmoothlm` or `tsmooth`) to capture the structural increase in waste management costs/values over time. The LM approach shows a FARIMA$(3, d, 3)$ with $d \approx 0.46$, indicative of moderate fractional dependence, and the Forecasts suggest continued upward movement, consistent with historical trends, while I accounted for the long memory probability within residuals.

These three data are interrelated through their environmental implications: for instances, higher natural gas usage and increased vehicle travel can contribute to pollution as we can see in the data they are given by units which we can further explore by converting those units into GHG emissions hence we will be able able observe impacts on environment, similarly while rising waste management costs may reflect growing disposal challenges or stricter regulations. In future work, one could compare SM vs. LM forecasting accuracy out-of-sample or explore exogenous factors (e.g., weather, economic shocks) to better explain shifts in these environment-related series.

5 Conclusion

This thesis has explored the modeling and forecasting of trend- or scale-stationary time series under short and long memory, with applications in economic, financial, and environmental contexts. I have systematically analyzed parametric and semiparametric models, in this work I provided insights into the complexities of time series data, for significantly I have focused on memory contents and their implications for in the time series arena.

The key contributions from my theis are included such as in-depth evaluation of short-memory processes, such as using ARMA and ARIMA, which demonstrated their effectiveness in capturing rapidly diminishing correlations. For long-memory processes, the thesis showed semiparametric approaches, including the SEMIFAR model, and I have emphasized their flexibility in accommodating persistent correlations and trends. The analysis has also included the FARIMA models.

The practical applications of these models were showed through their implementation on economic, financial, and environmental datasets. In Economic datasets I have highlighted relationships among key indicators like producer prices, consumer prices, and employment, while in financial models I have explored ETF volume data across major financial indices. Finally Environmental models were shown with long-memory methods in understanding natural gas consumption, vehicle miles driven, and waste management trends. Despite the robust methodologies that were discussed, There were still major challenges remain in the computational intensity of semiparametric models and of course there were the difficulty of selecting optimal parameters for example bandwidth and differencing factors.

In conclusion, My thesis contributes to the field of time series analysis by offering a coherent framework for modeling and forecasting under memory conditions. The findings have significance as the implications for both theoretical advancements and practical applications exists in this thesis.

Bibliography

Ariyo, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*, pages 106–112. IEEE.

Bauwens, L. and Giot, P. (2000). The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks. *Annales d'Économie et de Statistique*, (60):117–149.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., and Ciccozzi, M. (2020). Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, 29:105340.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(4):659–672.

Beran, J. (2017). *Statistics for long-memory processes.* Routledge.

Beran, J. and Feng, Y. (2002a). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics*, 54:291–311.

Beran, J. and Feng, Y. (2002b). Semifar models—a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, 40(2):393–419.

Beran, J., Feng, Y., and Ghosh, S. (2015). Modelling long-range dependence and trends in duration series: an approach based on efarima and esemifar models. *Statistical Papers*, 56:431–451.

Beran, J., Feng, Y., and Ocker, D. (1999). Semifar models. Technical report, Technical Report.

Bierens, H. J. (1988). The nadaraya-watson kernel regression function estimator.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. Springer science & business media.

Chen, G., Abraham, B., and Bennett, G. W. (1997). Parametric and non-parametric modelling of time series—an empirical study. *Environmetrics: The official journal of the International Environmetrics Society*, 8(1):63–74.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingrom inflation. *Econometrica*, 50:391–407.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162.

Feng, Y. and Zhou, C. (2015). Forecasting financial market activity using a semiparametric fractionally integrated log-acd. *International Journal of Forecasting*, 31(2):349–363.

Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, 130(1):1–23.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14(2):517–532.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4):221–238.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29.

Gray, H. L., Zhang, N.-F., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of time series analysis*, 10(3):233–257.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 38(1):1–21.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American society of civil engineers*, 116(1):770–799.

Jasiak, J. (1999). Persistence in intertrade durations. *Available at SSRN 162008*.

Letmathe, S., Beran, J., and Feng, Y. (2023). An extended exponential semifar model with application in r. *Communications in Statistics-Theory and Methods*, pages 1–13.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica: Journal of the Econometric Society*, pages 1279–1313.

Lobato, I. N. (1999). A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, 90(1):129–153.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.

Pai, P.-F. and Lin, C.-S. (2005). A hybrid arima and support vector machines model in stock price forecasting. *Omega*, 33(6):497–505.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association*, 85(410):338–344.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, pages 1630–1661.

Rohatgi, V. K. and Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.

Samorodnitsky, G., Taqqu, M. S., and Linde, R. (1996). Stable non-gaussian random processes: stochastic models with infinite variance. *Bulletin of the London Mathematical Society*, 28(134):554–555.

Sowell, F. (1992). Modeling long-run behavior with the fractional arima model. *Journal of monetary economics*, 29(2):277–302.