



Anton de Kom Universiteit van Suriname

Bibliotheek

Universiteitscomplex, Leysweg 86, Paramaribo, Suriname, Postbus 9212
Telefoon (597)464547, Fax (597)434211, E-mail: adekbib@uvs.edu

APPROVAL

NAAM: ..*Abdoelgefoer Traisha*.....

verleent (wel) / niet aan de AdeKUS kosteloos de niet-exclusieve toestemming om haar / zijn Drs. (B.Sc.) / M.Sc.
afstudeerscriptie online beschikbaar te stellen aan gebruikers binnen en buiten de AdeKUS.

Plaats en datum, ..*Paramaribo, 21-7-2025*.....

Handtekening ..*J. Abdoelgefoer*.....



Analysis of Four Types of Time Series Models

Anton de Kom Universiteit van Suriname

Faculteit der Wis- en Natuurkundige Wetenschappen

Study: Mathematics

Author:

Abdoelgafoer Iraisha

Student number: 201700582

Supervisor:

prof. dr. S. Venetiaan

Date:

July 31, 2025

Preface

When I began my bachelorproject, I honestly had no idea what topic I should best base my analysis on. With the guidance of professor dr. S. Venetiaan, I gradually became more familiar with time series analysis. As I delved deeper into the subject, I encountered many statistical concepts that were initially quite challenging, especially on my own. However, thanks to the continuous guidance and references provided by Miss Venetiaan, the subject became increasingly interesting as well. For that and for her patience with me, I would sincerely like to thank Miss Venetiaan. Furthermore, I could not have completed this thesis without the support of the faculty. I would like to extend my heartfelt gratitude to the entire team of the faculty of Wis- en Natuurkundige Wetenschappen, and in particular to the lecturers K. Hagens, D. Getrouw, R. Mahadewsing, S. Hiwat, C. Nijman and the student counsellor P. van Dijk. Last but certainly not least the encouragement and motivation of my fellow students, friends and family have meant a great deal to me throughout this process. Above all, I am deeply grateful to the Almighty for granting all that was needed and more to complete this thesis.

I hereby declare that for the research and writing of this thesis, I have committed no plagiarism, have not violated the copyrights of another, and have cited all sources used.

I. Abdoelgafoer

Abstract

This thesis analyzes four types of time series models, namely:

- Autoregressive processes
- Moving Average processes
- Autoregressive Moving Average processes
- Autoregressive Integrated Moving Average processes

Additionally, key statistical concepts such as white noise, stationarity, autocorrelation functions, and model properties like causality and invertibility are explored to understand the theoretical underpinnings of each model.

The research question is: *Which of the four studied time series models provides the best fit for the dataset of monthly USD–SRD exchange rates from 2004 to 2017?*

The analysis was conducted using RStudio, where stationarity was assessed and models were compared based on diagnostic plots and the Akaike Information Criterion (AIC). Out-of-sample forecasting was performed to evaluate predictive accuracy using actual exchange rate data from 2018 to 2022.

The results show that the ARIMA(1,2,1) model provided the best fit for the in-sample data. However, the model underperformed in out-of-sample forecasting due to its inability to capture structural breaks and external economic shocks. Therefore, the ARIMA model is suitable for historical modeling but limited in forecasting accuracy for future periods without incorporating additional factors.

Contents

Preface	i
Abstract	ii
1 Introduction	1
2 Fundamental Concepts	4
3 Four Univariate Time Series Processes	21
3.1 Autoregressive Processes	21
3.2 Moving Average Processes	31
3.3 Autoregressive Moving Average Processes	37
3.4 Autoregressive Integrated Moving Average Processes	42
4 Methodology	44
5 Results and Discussion	48
6 Conclusions and Recommendations	60
6.1 Conclusions	60
6.2 Recommendations	61
Appendices: R-scripts	62
References	66

Chapter 1

Introduction

Right from the start of this program, my interest lied in applied mathematics. Subjects such as Introduction to Statistics, Introduction to Operations Research and Introduction to Modeling, depicted the direct applications of mathematics in the “real” world. In other words, they reflected the practices of mathematics in various fields of work. The analytical and solution-oriented mapping of real-world factors into a mathematical model, and vice versa, leading to reality-applicable predictions and recommendations, is what encouraged me to base my graduate research on one of the subfields of applied mathematics. However, finding a graduation subject on my own was not easy. With guidance from prof. dr. S. Venetiaan, I discovered lecture notes on time series analysis, a subfield of applied mathematics that spans multiple disciplines, making it an ideal area for my thesis. Thus, my focus has been on **Time Series Analysis**.

A **time series** is a sequence of observations (data) recorded throughout a certain time unit. In the analysis of time series, time is a significant variable of the data. And the data points (observations) are collected at regular time intervals, rather than intermittently or randomly.

Some examples of time series are yearly birth rates, monthly unemployment rates, weekly stock market quotations or even simple observations like, the daily number of hours a student studies for, etc.

So time series analysis, in other words, is the analysis of data collected at consistent intervals over a set period of time.

This produces meaningful insights and predictions of future values of the sequence i.e. data forecasting. It showcases how the observed data changes over time and that there are dependencies between the data points.

This, predicting future events based on historical data, can be and in fact is being applied to fields such as :

- **Economics and Finance.** For example, banks or other finance institutions can predict interest rates, currency exchange rates or stock prices based on the past rates and therefore adjust their investments to ensure the highest of returns.
- **Retail.** For example, traders can predict their sales or the demand of their goods and services to adjust their purchasing strategies.
- **Agriculture.** For example, forecasting of the weather, which, amongst other, can influence when to plant or harvest.
- **Meteorology.** For example, forecasting of climate change for the coming years.
- **Politics.** For example, predicting the probability of a certain party to be elected based on previous election results.
- **Medicine.** For example, blood pressure measurements traced over time (for evaluating drugs).
- **Education.** For example, schools can analyze yearly student achievements to identify at-risk students or track progress over time.

Just as there are many applications of time series analysis, there are many types of time series and therefore various methods for modeling and analyzing those different time series.

One of the main ways in which one distinguishes time series models is based on the number of observed variables in the data. The two main types of models are:

- **Univariate models:** these include various models, which are used for analyzing sequences that consist of observations of one single factor recorded over time. So, here we only have one time-dependent variable. And the prediction is based on its own previous values and/ or (past and current) error terms. This class of models functions on the assumption that the time series data is stationary and it does not allow for trends or seasonalities to occur.
- **Multivariate models:** these models, on the other hand, are used for time series, where more than one (i.e. multiple) time-dependent variables are involved. So, in this case, two or more variables are measured over time to forecast their outcomes and besides including the dependency they have on their own past values, these models also consider the dependencies the variables have on each other (i.e. interdependencies among the variables). (*Time Series Analysis: Definition, Types, Techniques, and When It's Used, n.d.*)

The focus of this research is limited to univariate time series models. Through a comprehensive literature review, I have studied four prominent univariate time series processes: Autoregressive processes, Moving Average processes, Autoregressive Moving Average processes and Autoregressive Integrated Moving Average processes.

Based on these models, my research seeks to answer the following question:

Which of the four studied time series models provides the best fit for the monthly USD–SRD exchange rate data from 2004 to 2017?

The following chapters will first provide definitions and explanations of key statistical concepts, followed by a detailed discussion of the four univariate time series processes. The final chapter will determine which model best fits the given dataset, with the analysis conducted using RStudio.

Chapter 2

Fundamental Concepts

For us to understand time series models, it is imperative to consider each observation of such a series as a realized value of certain random variables. The random variables, i.e. all that what was observed throughout time, can be denoted by X_t and each observation, i.e. their realized values at given times, by x_t . Let us now note that according to the following definition, a time series is therefore nothing other than the realization of a stochastic process (Brockwell & Davis, 2009, p. 8).

Definition 2.1 (Stochastic Process). A stochastic process is a collection of random variables $\{X_t\}$ indexed by time (Brockwell & Davis, 2009, p. 8). $\{X_1, X_2, \dots\}$ a stochastic process and $\{x_1, x_2, \dots\}$ a single realization

Again, a time series is thus a realization of the family of random variables $\{X_t\}$, which is nothing other than a stochastic process and a time series model specifies the joint distribution of the sequence of random variables (Wang, 2015, p. 4). But we will not have to deal with these multivariate distributions, as implied earlier. We will only view time series $\{X_t\}$ consisting of observations of one random variable X_t , i.e. one factor and its values over time ($t= 0,1,2,\dots$).

According to Wang (2015, p. 4), a typical description of such a time series model can be given as:

Definition 2.2 (Time Series Model).

$$X_t = m_t + s_t + Y_t$$

,

where:

$$\begin{aligned} m_t &= \text{trend component} \\ s_t &= \text{seasonal component} \\ Y_t &= \text{Zero-mean error} \end{aligned}$$

It can also be rewritten as:

$$X_t = \text{Signal} + \text{Noise} ,$$

where:

Signal = the predictable part for which a model can be formulated.

Noise = White Noise = residuals part (unpredictable)

Definition 2.3 (White Noise). White Noise is a time series that satisfies the following criteria (Brooks, 2008, p. 209):

1. The variables are independent and identically distributed (IID)
2. The mean is equal to zero
3. The variance is constant over time
4. The correlation between lags is equal to zero (i.e., the autocorrelations between values at different time lags are zero), implying that there are no predictable patterns or dependencies in the residuals.

It is ideal for the residuals of a time series model to be white noise, because this indicates that the model has captured all the predictable (i.e. signal) information, leaving no further structure or patterns to explain. If the residuals show non-zero autocorrelation, changing mean or variance, this suggests the model could be improved. White noise series and other series, which also have a zero-mean structure, are therefore an important building block for modeling and forecasting of time series (Brownlee, 2018).

The following models represent examples of zero-mean time series components, i.e., different ways in which the noise term of a time series model can be modeled. In these models, it is assumed that the mean is zero, the variance is constant, and there are no predictable trends or seasonality. To further illustrate these zero-mean time series components, we include corresponding plots that visually confirm these underlying assumptions of the models. When the error term of a time series model exhibits these characteristics, it serves as a fundamental component in constructing effective forecasting models.

1. **IID Noise:** It is the simplest time series model with no trend or seasonal components and the observations (x_t 's) are independent and identically distributed (IID) with zero mean and constant variance. Its notation is: $X_t \sim IID(0, \sigma^2)$.

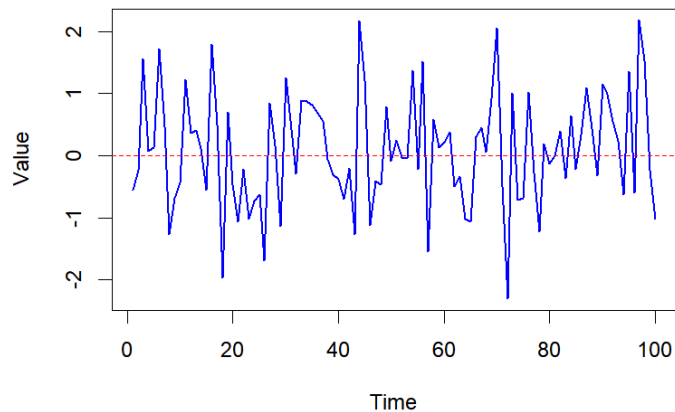


Figure 2.1: IID Noise Process $X_t \sim N(0,1)$

2. **Binary process (discrete):** $P(X_t = 1) = 0.5, P(X_t = -1) = 0.5$

In this discrete model, each observation x_t of the random variable X_t arbitrarily takes on only two values (1 and -1) with equal probability ($p = 0.5$), resulting in a series with a constant mean ($E(X_t) = 0$) and showing no autocorrelations (if the observations are independent).

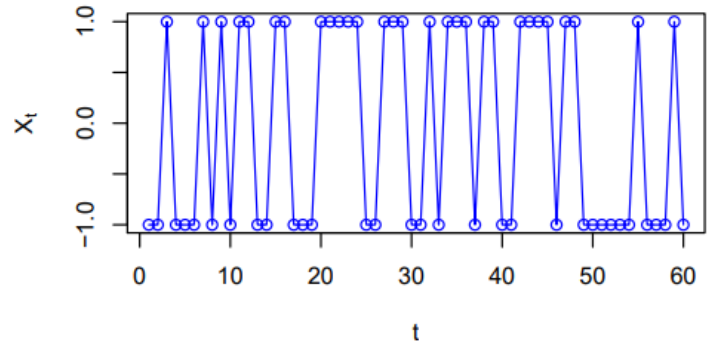


Figure 2.2: Binary Process. Retrieved from Wang (2015, p. 5).

3. **Gaussian noise (continuous):** $X_t \sim N(0, \sigma^2)$ iid.

In this continuous model, X_t is distributed according to a normal distribution with zero mean and constant variance. It is most commonly used as a baseline model for the residuals of time series models. Gaussian noise is a specific type of white noise that follows a normal distribution, without systematic patterns or correlations between observations.

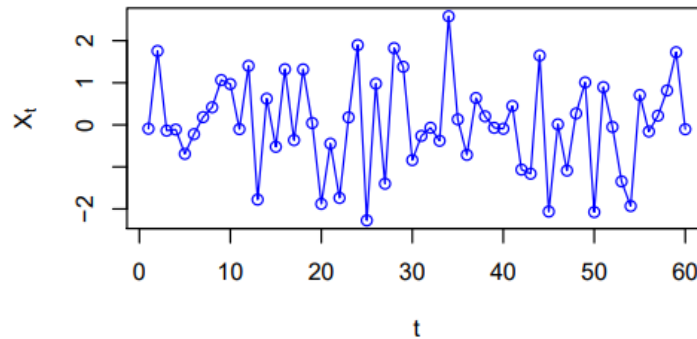


Figure 2.3: Gaussian Noise. Retrieved from Wang (2015, p. 5).

4. **Random Walk:** $S_t = X_1 + \dots + X_t$; **where $\{X_t\}$ is IID noise**

A random walk model has a recursive structure and assumes that the current value S_t is the cumulative sum of the previous values.

In this case, it's the sum of noise terms, which are IID, meaning that X_t follows a zero-mean structure. Hence, this random walk model is an example of a zero-mean model.

Further we note that:

- The noise term can be recovered by differencing, as this reveals that the series is driven by IID noise, emphasizing that the noise term X_t has a zero-mean structure:

$$\nabla S_t = S_t - S_{t-1} = X_t$$

This connection to differencing is important as it shows how random walks, despite appearing to trend, can be modeled as zero-mean processes when appropriately differenced.

- $\{X_t\}$ is IID $\Rightarrow E(X_t) = 0$, therefore:

$$E(S_t) = E\left(\sum_t X_t\right) = \sum_t E(X_t) = \sum_t 0 = 0$$

A random walk model depends on the type of process from which the values of X_t are drawn. For instance, a binary random walk, where each step is a binary outcome (0 or 1), will differ significantly from a Gaussian random walk, where each step follows a normal distribution. (This implies that while random walk processes can be zero-mean, the type of underlying process will influence the behavior and fluctuations over time.)

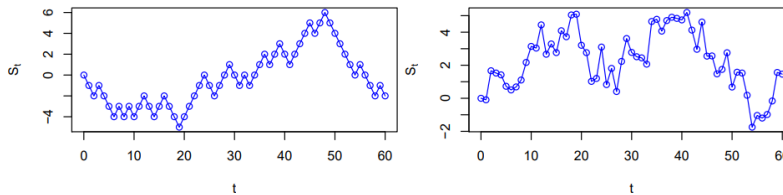


Figure 2.4: Random Walk of a binary process (left) and a Gaussian noise (right). Retrieved from Wang (2015, p. 5).

5. **White noise:** $X_t \sim WN(0, \sigma^2)$, when $\{X_t\}$ is uncorrelated and has constant variance. As defined earlier, white noise is a type of zero-mean process because each observation X_t has an expected value of zero, constant variance, and no autocorrelation. This randomness and lack of structure are characteristic of white noise and it often serves as the baseline model for the residuals of more complex time series models.

It is important to note that while all IID noise processes are white noise, the reverse is not always true (Wang, 2015, p. 6). White noise only requires uncorrelated observations, whereas IID noise is a stronger condition that also implies statistical independence between observations.

Both white noise and IID noise properties depend on the underlying distribution. For example, normally distributed white noise shows random fluctuations around the mean, while uniformly distributed white noise produces more evenly spread values:

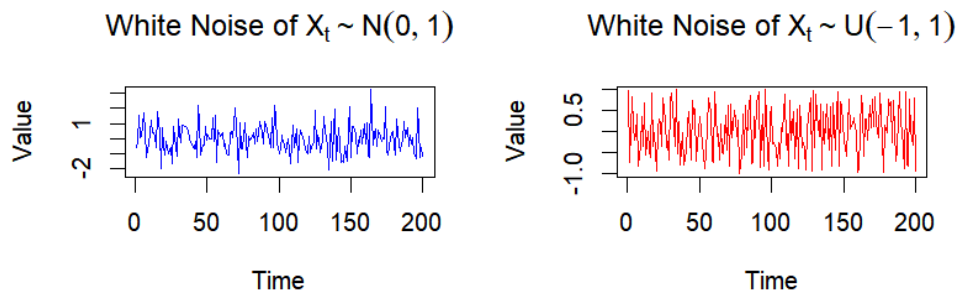


Figure 2.5

In a time series model, important information from the joint distributions can be summarized using means, variances, and covariances. These properties describe the dependencies between observed values over time.

Consider a time series model with the sequence $\{X_t: t = 0, 1, 2, \dots\}$. Then, the following properties hold (Wang, 2015, pp. 11-12):

- Mean of $\{X_t\}$:

$$E(X_t) = \mu_X(t)$$
- Autocovariance function (ACVF) of $\{X_t\}$:

$$\gamma_X(s, t) = \text{Cov}(X_s, X_t) = E[\{X_s - \mu_X(s)\}\{X_t - \mu_X(t)\}]$$
- Autocorrelation function (ACF) of $\{X_t\}$:

$$\rho_X(s, t) = \text{Corr}(X_s, X_t) = \frac{\gamma_X(s, t)}{\sqrt{\gamma_X(s, s) \cdot \gamma_X(t, t)}}$$

Definition 2.4 (Autocovariance Function). The Autocovariance Function (ACVF) measures the linear dependence (covariance) between different points in a time series. It quantifies how much past values influence future values and is a key tool for identifying underlying structures such as trends and seasonality.

For a time series X_t , the autocovariance function is defined as:

$$\gamma_X(s, t) = \text{Cov}(X_s, X_t) = E[\{X_s - \mu_X(s)\}\{X_t - \mu_X(t)\}] \quad (2.1)$$

where $\mu_X(t) = E(X_t)$ is the mean of the series at time t .

Interpreting the ACVF:

- High positive autocovariance: A large positive value of $\gamma_X(s, t)$ suggests that large values of X_t are likely to be followed by large values of X_{t+k} , and small values by small values.
- High negative autocovariance: A large negative value of $\gamma_X(s, t)$ indicates that if X_t is high, X_{t+k} is likely to be low, and vice versa.
- Near-zero autocovariance: This implies little to no linear dependence between observations at different time points meaning past values do not provide much information about future values.

Autocovariance is an essential concept in time series analysis, as it helps identify patterns and dependencies that can inform the choice of time series models.

Definition 2.5 (Autocorrelation Function). The Autocorrelation Function (ACF) standardizes autocovariance to provide a measure of correlation that is independent of the scale of the data. It quantifies the strength of the linear relationship between a time series and its lagged values.

The ACF is given by:

$$\rho_X(s, t) = \text{Corr}(X_s, X_t) = \frac{\gamma_X(s, t)}{\sqrt{\gamma_X(s, s) \cdot \gamma_X(t, t)}} \quad (2.2)$$

which normalizes the autocovariance values to the range $[-1, 1]$.

Interpreting the ACF:

- Strong positive autocorrelation ($\rho_X \approx 1$): Indicates that high values tend to be followed by high values, and low values by low values.
- Strong negative autocorrelation ($\rho_X \approx -1$): Suggests that high values tend to be followed by low values, and vice versa.
- Autocorrelation near zero: Indicates little to no linear dependence at the given lag.

Using R software, we can generate ACF plots (correlograms), which not only display the autocorrelation coefficients with confidence bands but also help identify trends, seasonality, and the appropriate type of time series model to use.

For example, when a dataset exhibits a trend, the autocorrelations for small lags tend to be large and positive because consecutive observations are likely to have similar values, reflecting the linear trend in the data. Therefore, the ACF of a time series with a trend typically shows large positive values that gradually decrease as the lags increase, capturing the persistence of the trend over time.

When the data exhibit seasonality, the autocorrelations will be larger at seasonal lags (multiples of the seasonal frequency), indicating repeating patterns or cycles in the data.

Lastly, when the data exhibit both a trend and seasonality, the ACF will show a combination of these effects, with large positive autocorrelations that decrease gradually, along with significant peaks at the seasonal lags.

While the autocorrelation function (ACF) provides insights into how observations in a time series are linearly related at different lags, it does not distinguish between direct and indirect dependencies. To better understand the direct relationship between an observation and its lagged values, we introduce the *Partial Autocorrelation Function (PACF)*.

Definition 2.6 (Partial Autocorrelation Function). The partial autocorrelation function (PACF) measures the direct relationship between a time series observation and its lagged values, removing the influence of intermediate lags. Unlike the autocorrelation function (ACF), which accounts for both direct and indirect correlations between observations, the PACF isolates only the direct linear dependence at each lag (Brooks, 2008, pp. 222-223).

Mathematically, the PACF at lag k is the correlation between X_t and X_{t-k} after removing the effects of the intermediate lags $1, 2, \dots, k - 1$.

For example: pacf for lag 3 (τ_{33}) would measure correlation between x_t and x_{t-3} after controlling effects of x_{t-1} and x_{t-2} .

Using PACF in Practice:

Similar to the ACF, the PACF can be visualized in a partial correlogram, where significant lags help identify the direct dependencies in the time series. In R, the PACF can be computed and plotted using the function `pacf()`, aiding in model selection and time series analysis.

Definition 2.7 (Stationarity). Stationarity is a property of certain time series processes, where the statistical properties of the process remain constant over time. This property simplifies analysis, forecasting, trend estimation, and causal inference. More specifically, for a process to be stationary, its statistical characteristics—such as the mean, variance, and autocovariance—do not change over time. While the values of the series may vary, the way the values change stays consistent, which makes the series predictable. When a time series has a *unit root*, its values do not tend to revert to a mean (i.e. there is no constant mean). These type of series often resemble a random walk, where shocks or random events have a persistent impact and the mean and variance do not remain constant over time, leading to unpredictable fluctuations and trends.

Stationarity can be classified into two types: **strict stationarity** and **weak stationarity**. A process is strictly stationary if its entire distribution remains unchanged over time. This means that the probability that a value falls within a particular interval is the same at any point in time. However, for the purposes of this study, we will focus on **weak stationarity**, which is more commonly used in practice.

There are three conditions to be met for a process to be called (**weakly**) **stationary** (Brooks, 2008, p. 208):

- $E(X_t) = \mu \forall t$ (constant mean)
- $\text{Var}(X_t) = E(X_t - \mu)(X_t - \mu) = \sigma^2 < \infty$ (constant variance)
- $\text{Cov}(X_{t_1}, X_{t_2}) = E(X_{t_1} - \mu)(X_{t_2} - \mu) = \gamma(t_2 - t_1) \forall t_1, t_2$ (a constant autocovariance structure)

The third condition, a constant autocovariance structure, means that the relationship between X_t and its past values depends only on the time difference $t_2 - t_1$, not on the specific time points. Therefore, for a stationary series, the covariance between X_t and X_{t-1} is the same as between X_{t-2} and X_{t-1} , and so on. Autocovariances are often difficult to interpret directly, as they depend on the units of measurement of X_t . For this reason, it is more convenient to use autocorrelations, which are normalized autocovariances. (Autocorrelations are always between ± 1 , providing a standardized measure of the strength and direction of the relationship between a time series and its lagged values, as mentioned earlier.)

Non-stationarity

Now that we know about the importance of stationarity for time series data, it is also necessary to learn about recognizing non-stationary data. This is because, while stationarity is a crucial assumption for many time series models, real-world data often exhibit non-stationary behavior due to trends, seasonality, or changing variance.

The most convenient method for identifying it would thus be plotting the data. (Wang, 2015, p. 20)

Some examples of non-stationary time series are (Wang, 2015, pp. 20-21):

1.

$$X_t = 1 + 0.5t + Y_t \quad (2.3)$$

where $\{Y_t\} \sim N(0, 1)$ iid.

Here $\mu_X(t)$ would be equal to the trend component, which is $1 + 0.5t$.

Since this depends on t , we conclude that the mean is non-constant.

This is also noticeable from the plot.

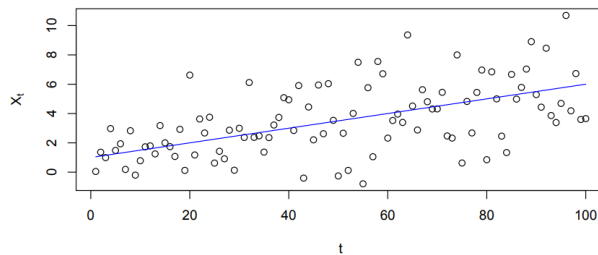


Figure 2.6: Example of non-constant trend. Retrieved from Wang (2015, p. 20).

2. Time series with seasonal trend: if $X_t = s_t + Y_t$ with seasonal trend s_t and zero-mean error Y_t , then $\mu_X(t) = s_t$ is not a constant.

For example:

$$X_t = 1 + 0.5t + 2 \cos(\pi t/5) + 3 \sin(\pi t/3) + W_t \quad (2.4)$$

where $\{W_t\} \sim N(0, 1)$ iid.

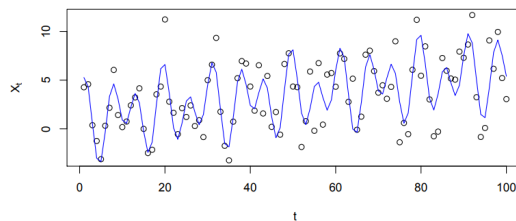


Figure 2.7: Example of seasonal trend. Retrieved from Wang (2015, p. 21).

3. Time series with non-constant variance. For example, random walk:

$$S_t = \sum_{j=1}^t X_j \quad (2.5)$$

where X_j is iid $N(0, 1)$.

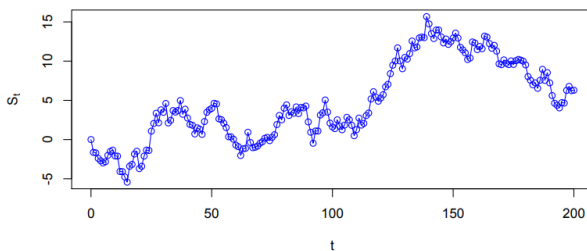


Figure 2.8: Example of non-constant variance. Retrieved from Wang (2015, p. 21).

The shape of the autocorrelation function (ACF) is also a way of identifying stationarity, as insinuated earlier. For a stationarity series, the ACF decreases quickly. However, for a non-stationarity series (such as a random walk), the ACF declines slowly or does not decay at all.

But in practice the true ACF remains unknown, so we make use of a sample ACF. For that we need estimators of the mean (μ_x) and the autocovariance function (ACVF).

So, for observations x_1, \dots, x_n of a time series X_t , where $t = 1, \dots, n$:

- the sample mean is: $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$
- the sample ACVF is: $\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_t - \bar{X})$
This way (i.e. by normalizing it by n instead of $n-h$ and by subtracting full sample mean), the sample covariance matrix is nonnegative definite.
- the sample ACF is: $\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$, for $-n < h < n$

All this is necessary to be able to follow the procedure of time series modeling, which can typically be described in the following three steps (Wang, 2015, p. 25):

1. Plotting the time series to look for trends, seasonality, step changes or outliers. And examining summary statistics and visualizations such as the autocorrelation function (ACF) and partial autocorrelation function (PACF) to assess dependencies.
2. Transforming the data so that residuals become stationary.
This can be done by:
 - estimating and subtracting the trend component and/or seasonal component
 - differencing: applying the difference operator to eliminate trends.
 - nonlinear transformations (\log , $\sqrt{\cdot}$): applying logarithms or power transformations to stabilize variance

Stationarity can be checked using statistical tests such as the Augmented Dickey-Fuller (ADF) test.

3. Fitting the model to residuals: In this step, an appropriate time series model (such as AR, MA, ARMA or ARIMA) is fitted to the residuals. The goal is to model any remaining dependence structure and ensure that the residuals resemble white noise, meaning they contain no further predictable patterns.

Model parameters are estimated using methods like Maximum Likelihood Estimation (MLE) or Ordinary Least Squares (OLS). A well-fitted model should yield residuals that resemble white noise, meaning they are independently and identically distributed (iid) with a mean of zero and constant variance.

Model selection criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) can be used to compare different models.

Finally yet importantly, follows the introduction of the difference operator and backshift operator, also known as lag operator (Wang, 2015, p. 25).

- lag-1 difference operator: $\nabla X_t = \frac{X_t - X_{t-1}}{t - (t-1)} = X_t - X_{t-1} = (1 - B)X_t$;
 where B is the backshift operator, $BX_t = X_{t-1}$
 (So what the backshift/ lag operator here does is, it shifts the series back by one time unit and X_{t-1} is called the first lag.)
- lag-s difference operator: $\nabla_s X_t = X_t - X_{t-s} = (1 - B^s)X_t$;
 where B^s is the backshift operator applied s times, $B^s X_t = B(B^{s-1} X_t)$

The backshift operator is useful in time series modeling because it simplifies notation and makes it easier to express differencing and lag structures. For example, the first-order difference operator $\nabla X_t = X_t - X_{t-1}$ can be written compactly as $(1 - B)X_t$. This is especially helpful in ARMA and ARIMA models, where differencing is often required to achieve stationarity.

Linear Processes:

Linear processes form the mathematical foundation of many standard time series models (such as AR, MA, and ARMA), because many of these models can be expressed as linear combinations of past disturbances. Understanding the properties of linear processes, especially mean square convergence and stability conditions, is crucial for analyzing stationarity, invertibility, and causality in ARMA-type models. These properties are used in later sections when defining and estimating AR, MA, and ARMA processes.

Definition 2.8 (Linear Process). A time series $\{X_t\}$ is called a **linear process** if it can be expressed as:

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j W_{t-j}, \quad (2.6)$$

where:

- $\{W_t\}$ is a white noise process with mean zero and variance σ^2 , i.e., $W_t \sim WN(0, \sigma^2)$,
- $\{\psi_j\}$ are the weight coefficients, which determine how past noise terms influence X_t ,
- $\sum_{j=0}^{\infty} |\psi_j| < \infty$ converges and therefore ensures stationarity and that X_t is well-defined.

(Wang, 2015, p. 26)

In other words, X_t is composed of a constant μ and a weighted sum of past white noise terms. The sum of absolute values of the coefficients must be finite to ensure that past shocks do not have an infinitely accumulating effect on the present value of X_t (i.e. the effect of past disturbances on X_t should diminish over time). In practical terms, this ensures that the process is **stable** and that forecasting methods relying on linear processes remain valid.

Linear processes are crucial in time series analysis for the following reasons:

- As mentioned before, many standard models (AR, MA, ARMA) are special cases of linear processes.
- The stationarity of a time series can often be analyzed through the properties of its corresponding linear process.
- Concepts such as **causality** and **invertibility** are tied to the stability of linear processes, ensuring well-defined and unique representations of ARMA models.
- Linear processes allow for the application of fundamental statistical results, such as convergence properties, which are crucial for estimating and predicting time series behavior.

Convergence in Mean Square (Wang, 2015, p. 26)

A fundamental property of linear processes is their convergence behavior, particularly **convergence in mean square**. A sequence of random variables X_n is said to **converge in mean square** to a random variable X (denoted as $X_n \xrightarrow{L^2} X$) if:

$$E(X_n - X)^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (2.7)$$

This ensures that the expected squared difference between X_n and X becomes arbitrarily small for large n , meaning that the sequence stabilizes in a probabilistic sense. For linear processes, convergence in mean square is particularly useful for proving the existence of well-behaved solutions to models and ensuring the stability of estimators.

Cauchy Criterion and Riesz-Fisher Theorem (Wang, 2015, p. 27)

The concept of mean square convergence is often established using the **Cauchy criterion**, which states that X_n converges in mean square if and only if:

$$\lim_{m, n \rightarrow \infty} E(X_m - X_n)^2 = 0. \quad (2.8)$$

This criterion is essential because it allows us to verify convergence without needing explicit knowledge of the limiting variable X .

A key result related to mean square convergence is the **Riesz-Fisher Theorem**, which guarantees that a Cauchy sequence in mean square converges to a well-defined limit in L_2 space. This result is fundamental in time series analysis because it ensures that processes modeled as infinite linear combinations of past noise terms remain well-defined.

Chapter 3

Four Univariate Time Series Processes

3.1 Autoregressive Processes

The name gives it away already, these are processes where the prediction of future values is based on its own past values. So the lagged value(s), i.e. value(s) which the observed variable took during the previous period(s), is (/are) taken into consideration. Here the current value, x_t , depends solely on past value(s) and an error term, as shown in the model below:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + W_t \quad (3.1)$$

This is the standard representation of an AR model of the order p , i.e AR(p) model; wherein p is the number of lags. (Wang, 2015, p. 31)

We can use the sigma notation and backshift (/lag) operator to rewrite the above equation more compactly:

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + W_t \quad (3.2)$$

With use of the backshift operator, we can understand that the above is same as the following:

$$\phi(L)X_t = W_t; \tag{3.3}$$

as:

$$\phi(L) = 1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p \tag{3.4}$$

Let us study two concrete examples of AR-models.

Example 1: $x_t = x_{t-1} + w_t$ (Brooks, 2008, p. 217)

We first notice the order of the model. This is an AR(1) example. Secondly, we check for stationarity. This is done by transforming the equation with use of the backshift operator. What we then get, is the characteristic equation i.e. a polynomial in terms of z).

Before we dive further into our example, let us pause for a while to look into what stationarity means for AR-models.

The condition for testing stationarity of a general AR(p) model is that all roots of the characteristic equation should lie outside the unit circle, meaning their absolute values must be greater than 1.

The general form of the characteristic equation is:

$$\phi(z) = 1 - \phi_1z - \phi_2z^2 - \dots - \phi_pz^p \tag{3.5}$$

So, the stationarity condition for AR(p) processes is: $|z| > 1$

This requirement ensures that past shocks in the process decay over time rather than persist indefinitely.

Now, back to a step by step analysis of our example:

1. $x_t - x_{t-1} = w_t$
2. backshift operator $\Rightarrow (1 - B)x_t = w_t$
3. Let $1 - B = \phi(B)$, then characteristic equation $\Rightarrow \phi(z) = 1 - z$
(we plug in z for the backshift operator.)
4. we solve $\phi(z) = 0$ to find the root(s) $\Rightarrow 1 - z = 0 \Rightarrow z = 1$
This doesn't lie outside the unit circle, but on the unit circle.
Conclusion: The above process is non-stationary.
(This one is known as a Random Walk process.)

Example 2: Check $X_t = 3x_{t-1} - 2.75x_{t-2} + .75x_{t-3} + w_t$ for stationarity.
(Brooks, 2008, p. 217)

$$\Rightarrow x_t = 3Bx_t - 2.75B^2x_t + 0.75B^3x_t + w_t$$

$$\Rightarrow x_t(1 - 3B - 2.75B^2 + 0.75B^3) = w_t$$

$$\Rightarrow -0.75z^3 + 2.75z^2 - 3z + 1 = 0 \text{ (Finding roots of characteristic equation)}$$

\Rightarrow By testing rational candidates $(\pm 1, \pm \frac{1}{2}, \pm \frac{1}{3}, \dots)$, we find that $z = 1$ is a root.

\Rightarrow Dividing $-0.75z^3 + 2.75z^2 - 3z + 1$ by $(1 - z)$ using polynomial division yields:

$$(1 - z)(-0.75z^2 + 2z - 1) = 0$$

\Rightarrow Solving the quadratic equation $-0.75z^2 + 2z - 1 = 0$ using the quadratic formula, yields:

$$z = \frac{-2 \pm \sqrt{4-3}}{-1.5} = \frac{-2 \pm 1}{-1.5} \text{ which results in the roots: } z_1 = \frac{2}{3}, \quad z_2 = 2$$

\Rightarrow The three roots we got are: $z = 1 \vee z = \frac{2}{3} \vee z = 2$

This process is also not stationary, because not all the roots lie outside the unit circle (only $z = 2$ does).

The characteristic equation of an AR(p) process plays a crucial role in determining stationarity. If any root lies inside or on the unit circle ($|z| \leq 1$), then the process is non-stationary, as shocks do not dissipate over time. These examples illustrate how verifying the roots of the characteristic equation provides a concrete method for assessing stationarity.

Example 3 (Brooks, 2008, p. 218):

AR(1) model: $y_t = \mu + \phi_1 y_{t-1} + u_t$; where:

- μ is the drift or constant term, which determines whether the process has a mean different from zero. In some formulations, this term is omitted (e.g., if the process fluctuates around a zero mean).
- ϕ_1 is the autoregressive parameter, which indicates how strongly y_t depends on its previous value y_{t-1} . The magnitude of ϕ_1 affects the stationarity and stability of the process. Stationarity requires that $|\phi_1| < 1$.
- u_t is the error term or white noise term, representing a random disturbance at time t . It is often modeled as an independent and identically distributed (iid) noise term with mean zero and variance σ^2 , i.e. $u_t \sim WN(0, \sigma^2)$.

Unlike the previous two examples, where we first checked for stationarity, in this example, we will derive the statistical properties of the process by computing its mean, variance, and autocorrelation.

$$\Rightarrow E(y_t) = \mu + \phi_1 E(y_{t-1})$$

$$\begin{aligned} \text{We know: } y_{t-1} &= \mu + \phi_1 y_{t-2} + u_{t-1} \\ \Rightarrow E(y_t) &= \mu + \phi_1 E(\mu + \phi_1 y_{t-2} + u_{t-1}) \\ &= \mu + \phi_1 (\mu + \phi_1 E(y_{t-2})) \\ &= \mu(1 + \phi_1) + \phi_1^2 E(y_{t-2}) \\ &= \mu + \phi_1 \mu + \phi_1^2 E(y_{t-2}) \end{aligned}$$

Again, we know: $y_{t-2} = \mu + \phi_1 y_{t-3} + u_{t-2}$
 $\Rightarrow E(y_t) = \mu + \phi_1 \mu + \phi_1^2 (\mu + \phi_1 E(y_{t-3}))$
 $= \mu + \phi_1 \mu + \phi_1^2 \mu + \phi_1^3 E(y_{t-3})$

Making n-substitutions (y_{t-n}), we get:
 $\Rightarrow E(y_t) = \mu + \phi_1 \mu + \dots + \phi_1^{n-1} \mu + \phi_1^n E(y_{t-n})$
 $= \mu(1 + \phi_1 + \phi_1^2 + \dots + \phi_1^{n-1}) + \phi_1^n E(y_{t-n})$

For stationarity holds: $|\phi_1| < 1$

$\Rightarrow \phi_1^n \rightarrow 0 \Rightarrow$ as $n \rightarrow \infty$, then $\lim_{n \rightarrow \infty} \phi_1^n E(y_{t-n}) = 0$

\Rightarrow So: $E(y_t) = \mu(1 + \phi_1 + \phi_1^2 + \dots)$

We know that finite sum of an infinite number of geometrically declining terms in a series is given by "first term in series divided by one minus the common difference".

Common difference = the quantity that each term in the series is multiplied by to arrive at the next term.

$\Rightarrow E(y_t) = \frac{\mu}{1-\phi_1}$

By above example is demonstrated that the expected or mean value of an AR(1) process is equal to intercept parameter (μ) divided by one minus the AR-coefficient. (This result shows that the mean of an AR(1) process depends on both the intercept parameter μ and the AR-coefficient ϕ_1 .)

Notice that when $\mu = 0$, the mean of the process is also zero, i.e. $E(y_t) = 0$ and the AR(1) process can then be referred to as a zero-mean time series.

Let us now look at the calculations for the variance of y_t , while μ is set to zero for simplicity:

$\Rightarrow y_t = \phi_1 y_{t-1} + u_t$

using the lag-operator we can rewrite the above as:

$\Rightarrow y_t(1 - \phi_1 L) = u_t$

Next, we will apply Wold's Decomposition theorem. This theory implies that any stationary series can be decomposed into the sum of two unrelated processes, namely: a purely deterministic part (, which is perfectly predictable from past values) and a purely stochastic part (, which is an infinite moving average (MA) representation of a white noise process). In other words, any stationary AR(p) process with no constant and no other terms can be expressed as an infinite order MA model (MA(∞)). This allows us to derive variance formulas and understand the structure of the process in terms of past shocks.

We know, using lag polynomial (/backshift-operator) on an AR(p) model with $\mu=0$, gives:

$$\phi(L)y_t = u_t \quad (3.6)$$

where: $\phi(L) = 1 - \phi_1L - \dots - \phi_pL^p$

\Rightarrow Wold's Decomposition theorem:

$$y_t = \psi(L)u_t \Rightarrow \psi(L) = \frac{1}{\phi(L)} = (1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p)^{-1} \quad (3.7)$$

By the use of the properties of an infinite geometric series,:

$$\frac{1}{1-r} = \sum_{j=0}^{\infty} r^j = 1 + r + r^2 + \dots \quad (3.8)$$

we can rewrite (3.7) as:

$$\psi(L) = 1 + \phi_1L + \phi_2L^2 + \dots + \phi_pL^p \Rightarrow y_t = (1 + \phi_1L + \dots + \phi_pL^p)u_t \quad (3.9)$$

which, as we can see, expresses our process as an infinite sum of white-noise terms (i.e. Wold's Decomposition theorem with the geometric series expansion as a useful tool).

Applying this to the AR(1) process in our example, we obtain the following MA(∞) representation:

$$\begin{aligned} \Rightarrow y_t &= (1 - \phi_1L)^{-1}u_t \\ \Rightarrow y_t &= (1 + \phi_1L + \phi_1^2L^2 + \dots)u_t \\ \Rightarrow y_t &= u_t + \phi_1u_{t-1} + \phi_1^2u_{t-2} + \phi_1^3u_{t-3} + \dots \end{aligned}$$

By definition of the variance, we know: $Var(y_t) = E[y_t - E(y_t)]^2 \wedge E(y_t) = 0$, as μ is set to zero.

So, for $Var(y_t) = \gamma_0$ holds:

$$\begin{aligned} \Rightarrow \gamma_0 &= E(y_t^2) = E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \phi_1^3 u_{t-3} + \dots)^2] \\ \Rightarrow &= E[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \phi_1^6 u_{t-3}^2 + \dots + \text{cross-products}] \end{aligned}$$

Note that "cross-products" is the catchall expression for all the terms in u , which have different time subscripts, such as: $u_t u_{t-1}$; $u_{t-1} u_{t-2}$, etc., i.e. the auto-covariances of u_t

We know: $E(\text{cross-products}) = 0$, as $Cov(u_t, u_{t-s}) = 0 \forall s \neq 0$

The properties of the distribution of our error term here (i.e. u_t , which follows a white noise process distribution ($WN(0, \sigma^2)$)), ensure that:

- $E(u_t) = 0$, which holds for every lag value of u_t
- $E(u_t^2) = \sigma^2$, meaning that the variance of u_t remains constant over time.

Since white noise is identically and independently distributed, this also implies that the expectation of any lagged squared value is the same, i.e:

$$E(u_{t-s}^2) = \sigma^2 \forall s \geq 0$$

Carrying on, we now have:

$$\Rightarrow Var(y_t) = \gamma_0 = \sigma^2 + \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + \dots = \sigma^2(1 + \phi_1^2 + \phi_1^4 + \dots)$$

Provided that $|\phi_1| < 1$, i.e., y_t is stationary, the sum will converge, allowing us to apply the geometric series formula. Thus, the infinite sum can be rewritten as:

$$\sum_{j=0}^{\infty} \phi_1^{2j} = \frac{1}{1 - \phi_1^2} \Rightarrow Var(y_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2} \quad (3.10)$$

Now, to compute the autocorrelation function (ACF), the autocovariances must first be calculated, as the ACF values are determined by dividing these covariances by the variance.

autocovariance at lag 1 $\Rightarrow \gamma_1 = \text{Cov}(y_t, y_{t-1})$:

$$\Rightarrow \gamma_1 = E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))]$$

We know: $\mu = 0 \Rightarrow E(y_t) = E(y_{t-1}) = 0$

$$\Rightarrow \gamma_1 = E[(y_t)(y_{t-1})]$$

$$\Rightarrow = E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \dots)(u_{t-1} + \phi_1 u_{t-2} + \phi_1^2 u_{t-3} + \dots)]$$

$$\Rightarrow = E[\phi_1 u_{t-1}^2 + \phi_1^3 u_{t-2}^2 + \dots + \text{cross-products}]$$

$$\Rightarrow = \phi_1 \sigma^2 (1 + \phi_1^2 + \phi_1^4 + \dots) = \frac{\phi_1 \sigma^2}{1 - \phi_1^2}$$

Following the same steps, we also obtain:

$$\Rightarrow \gamma_2 = \text{cov}(y_t, y_{t-2}) = E[(y_t)(y_{t-2})] = \frac{\phi_1^2 \sigma^2}{1 - \phi_1^2}$$

$$\Rightarrow \gamma_3 = \frac{\phi_1^3 \sigma^2}{1 - \phi_1^2}$$

$$\Rightarrow \gamma_s = \frac{\phi_1^s \sigma^2}{1 - \phi_1^2}$$

With the autocovariances computed, we can now determine the autocorrelations by normalizing them with the variance.

$$\Rightarrow \tau_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\Rightarrow \tau_1 = \frac{\gamma_1}{\gamma_0} = \left(\frac{\phi_1 \sigma^2}{1 - \phi_1^2}\right) \left(\frac{1 - \phi_1^2}{\sigma^2}\right) = \phi_1$$

$$\Rightarrow \tau_2 = \phi_1^2$$

$$\Rightarrow \tau_3 = \phi_1^3$$

$$\Rightarrow \tau_s = \phi_1^s \text{ (= autocorrelation at lag s)}$$

ACF plot:

A typical ACF plot of a stationary AR(p) process displays a gradual, exponential decay and may oscillate, depending on the parameters. (This distinguishes it from non-stationary processes, where the ACF remains high across many lags without a clear tendency to decay to zero.) This (exponentially decaying) pattern arises from the autoregressive nature of AR models, where past values influence both the present and future values, meaning there is correlation between lags. This correlation is most pronounced at the beginning of the ACF plot, and as we move to later lags, the effect diminishes, reflecting the weakening influence of past values over time. (Haque, 2015)

PACF plot:

Similarly, in a stationary AR(p) process, the PACF plot typically shows a sharp cutoff after lag p. This means that only the first p lags have significant partial autocorrelations, while the values at higher lags drop to (or fluctuate around) zero. This pattern occurs because an AR(p) process directly depends on its p most recent past values, but not on earlier lags once the effects of the first p lags have been accounted for. In other words, any correlation at higher lags is explained indirectly through the dependence on the first p lags, and therefore, those higher-lag correlations are weak or nonexistent in the PACF. The PACF plot is valuable for identifying the order p of the AR model, as it provides a clear indication of how many past values (lags) should be included in the model. (Haque, 2015)

Causality (Wang, 2015, p. 31):

A stochastic process $\{X_t\}$ is a *causal* autoregressive process if it can be expressed as a function of past white noise terms $\{W_t\}$.

Mathematically, this means there is a $\psi(B)$ such that:

$$X_t = \psi(B)W_t \tag{3.11}$$

where $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$; i.e. a convergent power series in the backshift operator B.

This definition ensures that the present value of the process depends only on past and current innovations and not on future values.

Causality is important because it guarantees that the process can be generated from past shocks in a stable manner, making it useful for forecasting and interpretation. In practical terms, a causal AR process does not rely on future values of the error term, which aligns with the natural assumption that past information influences the present rather than the other way around.

For an AR(1) process of the form

$$X_t = \phi X_{t-1} + W_t \tag{3.12}$$

causality holds if and only if $|\phi| < 1$. This condition ensures that the infinite MA representation of the process

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j} \tag{3.13}$$

converges, meaning that the influence of past shocks diminishes over time. If $|\phi| > 1$, the process is non-causal because the past values would explode rather than decay, making the process unstable. Importantly, this specific criterion applies only to AR(1) models. For higher-order AR(p) processes, causality conditions are determined by the roots of the characteristic equation lying outside the unit circle in the complex plane.

3.2 Moving Average Processes

These processes have the current x_t term be depending on the current and past error terms. It's the simplest class of time series models, as it's nothing other than a linear combination of white noise processes (i.e. disturbance terms).

An MA(q) model can be represented by the following equation:

$$X_t = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q} \quad (3.14)$$

where q denotes the number of lagged error terms. (Brooks, 2008, p. 211)

Moving Average (MA) processes are useful in modeling time series where past shocks have a transient effect on the present values. They are particularly effective in capturing short-term dependencies and are often used to model time series data with a fixed memory of past disturbances.

To gain a deeper understanding of MA processes, we now consider a specific example: an MA(2) process. By deriving its mean, variance, and autocorrelation function (ACF), we illustrate its fundamental properties.

Example MA(2) process (Brooks, 2008, p. 212):

$$\begin{aligned} y_t &= w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} ; \\ w_t &= \text{zero-mean white noise} \\ &\text{with variance equal to } \sigma^2 \\ &\text{i.e.: } w_t \sim \text{WN}(0, \sigma^2) \end{aligned}$$

$$\begin{aligned} E(y_t) &= ? \Rightarrow \text{if } E(w_t) = 0 \Rightarrow E(w_{t-i}) = 0 \forall i \\ E(y_t) &= E(w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}) = E(w_t) + \theta_1 E(w_{t-1}) + \theta_2 E(w_{t-2}) = 0 \end{aligned}$$

$$\begin{aligned} \text{Var}(y_t) &= E\{[y_t - E(y_t)][y_t - E(y_t)]\} = E\{[y_t][y_t]\} \\ &= E\{[w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}][w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}]\} \\ &= E\{w_t^2 + \theta_1^2 w_{t-1}^2 + \theta_2^2 w_{t-2}^2 + \text{cross-products}\} \end{aligned}$$

(As explained earlier, the expectation of the cross-products can be set to zero.)

$$\begin{aligned}
\text{Var}(y_t) &= E(w_t^2) + \theta_1^2 E(w_{t-1}^2) + \theta_2^2 E(w_{t-2}^2) \\
&= \sigma^2 + \theta_1^2 \sigma^2 + \theta_2^2 \sigma^2 \\
&= \sigma^2(1 + \theta_1^2 + \theta_2^2) = \gamma_0 \quad (= \text{Autocovariance at lag zero})
\end{aligned}$$

ACF:

$$\tau_1 = \frac{\gamma_1}{\gamma_0} \text{ and } \tau_2 = \frac{\gamma_2}{\gamma_0} ;$$

$\gamma_1 =$ autocovariance at lag 1

$$\begin{aligned}
&= E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})] \\
&= E[y_t][y_{t-1}] \\
&= E[w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}][w_{t-1} + \theta_1 w_{t-2} + \theta_2 w_{t-3}] \\
&= E[\theta_1 w_{t-1}^2 + \theta_1 \theta_2 w_{t-2}^2 + \text{cross-products}] \\
&= \theta_1 E(w_{t-1}^2) + \theta_1 \theta_2 E(w_{t-2}^2)
\end{aligned}$$

$$\Rightarrow \gamma_1 = \theta_1 \sigma^2 (1 + \theta_2)$$

$\gamma_2 =$ autocovariance at lag 2

$$\begin{aligned}
&= E[y_t - E(y_t)][y_{t-2} - E(y_{t-2})] \\
&= E[y_t][y_{t-2}] \\
&= E[w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}][w_{t-2} + \theta_1 w_{t-3} + \theta_2 w_{t-4}] \\
&= E[\theta_2 w_{t-2}^2 + \text{cross-products}]
\end{aligned}$$

$$\Rightarrow \gamma_2 = \theta_2 \sigma^2$$

Note that all autocovariances γ_s for the MA(2) process will be zero for any lag length (s) greater than 2.

Carrying on with the ACF's, we have now got:

$$\Rightarrow \tau_1 = \frac{\theta_1 \sigma^2 (1 + \theta_2)}{\sigma^2 (1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} \text{ and } \tau_2 = \frac{\sigma^2 \theta_2}{\sigma^2 (1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2} ;$$

$$\tau_s = \frac{\gamma_s}{\gamma_0} = 0 \quad \forall s > 2$$

ACF plot:

A key characteristic of a MA(q) process is that its ACF cuts off after lag q . This means that the autocorrelations are nonzero for the first q lags but drop to approximately zero for all higher lags. This behavior arises because an MA(q) process is defined as a finite sum of white noise terms, meaning that past values beyond q lags have no direct influence on the present value. (If the ACF does not cut off and instead decays gradually, the series may not be an MA process but rather an AR or ARMA process.)

Mathematically, for a MA(1) process given by 3.17, where W_t is white noise, the autocorrelation function is given by:

$$\tau(1) = \frac{\theta}{1 + \theta^2} ; \quad \tau(k) = 0 \quad \forall k > 1 \tag{3.15}$$

This demonstrates the cutoff property, where only the first lag exhibits significant autocorrelation. For a general MA(q) process, the ACF remains significant up to lag q and then drops to zero. This makes the ACF particularly useful in identifying the order of an MA process.

PACF plot:

Unlike the ACF, the PACF of a MA(q) process exhibits a gradual, tapering decay rather than a sharp cutoff. This is because the PACF measures the direct dependence between observations while removing the influence of intermediate lags. Since an MA process does not include past values of X_t , its partial autocorrelations extend indefinitely, gradually decreasing in magnitude.

For an MA(1) process, the PACF does not cut off at lag 1 but instead decays gradually, showing a damped sinusoidal or exponential pattern depending on the parameter values.

Invertibility (Wang, 2015, p. 32):

While a MA process expresses the observed time series as a function of current and past error terms, *invertibility* ensures that the process can be rewritten as a convergent infinite AR representation in terms of past observed values. (Similarly, causality for AR processes ensures they can be expressed in terms of past white noise terms.)

A stochastic process $\{X_t\}$ is an *invertible* moving average process if it can be expressed in terms of past values of the observed series $\{X_t\}$. Mathematically, this means there is a $\pi(B)$ such that:

$$W_t = \pi(B)X_t \tag{3.16}$$

where $\pi(B) = \pi_0 + \pi_1 B + \pi_2 B^2 + \dots$; i.e. a convergent power series in the backshift operator B.

This condition is essentially the inverse of the way an MA process is originally defined. This transformation is particularly useful in estimation and forecasting, as it ensures a unique and stable representation of the time series. Invertibility is crucial for the identification and estimation of MA models. In practice, a non-invertible MA model is problematic because different sets of parameters could describe the same data, leading to non-uniqueness in model selection. The invertible representation is preferred because it ensures that past values of $\{X_t\}$ provide a meaningful and finite representation of the white noise disturbances.

MA processes are stationary by definition, as they are expressed as a finite sum of white noise terms. This means that we do not need to check for stationarity. Instead, we examine invertibility, which determines whether the process can be rewritten as a convergent infinite AR representation.

Consider a simple MA(1) process given by:

$$X_t = W_t + \theta W_{t-1} \tag{3.17}$$

where $W_t \sim WN(0, \sigma^2)$
(Wang, 2015, p. 32)

To determine invertibility, we check whether we can express W_t in terms of past values of X_t :

$$W_t = X_t - \theta W_{t-1} \tag{3.18}$$

(Note: we have not yet reached an expression solely in terms of X_t 's, so we will apply repeated substitutions for W_{t-1} .)

Since we know that the MA(1) equation holds for every time point t , we can apply the same equation to W_{t-1} , as well. Meaning:

$$\begin{aligned} \text{Since: } X_{t-1} &= W_{t-1} + \theta W_{t-2} &\Rightarrow & W_{t-1} = X_{t-1} - \theta W_{t-2} \\ \text{Since: } X_{t-2} &= W_{t-2} + \theta W_{t-3} &\Rightarrow & W_{t-2} = X_{t-2} - \theta W_{t-3} \\ & & & \vdots \end{aligned}$$

By recursively solving and substituting for 3.18, we obtain:

$$W_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \dots \tag{3.19}$$

This series representation converges if and only if $|\theta| < 1$. This constraint ensures that the coefficients of the infinite past values decay exponentially, preventing the series from diverging.

- if $|\theta| < 1$, the MA(1) process is invertible, meaning we can represent the process in terms of past values of X_t .
- if $|\theta| > 1$, the process is not invertible, because the infinite expansion does not converge.
- if $|\theta| = \pm 1$, the process is also not invertible, as it results in an unstable and non-decaying representation.

(Wang, 2015, pp. 32-33)

This was one way of checking invertibility, by explicitly rewriting the MA process as an infinite AR representation. Another approach is, to examine the characteristic equation, as follows (Wang, 2015, p. 33):

The invertibility condition for a general MA(q) model is that all roots of the characteristic equation should lie outside the unit circle, meaning the modulus of each root must be greater than 1. The general form of the characteristic equation for an MA(q) model is:

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q \quad (3.20)$$

The invertibility condition requires that the roots satisfy $|z| > 1$

Now, let's analyze our specific MA(1) example step by step:

1. Given process:

$$X_t = W_t + \theta W_{t-1} \quad (3.21)$$

2. Rewrite using backshift operator:

$$X_t = W_t(1 + \theta B) \quad (3.22)$$

3. Define the characteristic equation:

$$\theta(z) = 1 + \theta z \quad (3.23)$$

4. Solve $\theta(z) = 0$ to find the root(s):

$$1 + \theta z = 0 \Rightarrow z = -\frac{1}{\theta} \quad (3.24)$$

5. The invertibility condition requires $|z| > 1$, meaning:

$$\left| -\frac{1}{\theta} \right| > 1 \Rightarrow |\theta| < 1 \quad (3.25)$$

Conclusion: For an MA(1) process to be invertible, the parameter θ must satisfy $|\theta| < 1$. Otherwise, the model cannot be expressed as a stable AR(∞) process.

For higher-order MA(q) processes, the general principle remains the same: an MA(q) process is invertible if the roots of its characteristic equation lie outside the unit circle in the complex plane.

3.3 Autoregressive Moving Average Processes

An ARMA(p,q) process is based on both previous values and past errors (, where the errors are a White Noise series). Again, p tells us the order of the AR part of the process and q tells us the order of the MA part of the process. These processes thus have the characteristics of both AR-processes and MA-processes combined (Brooks, 2008, p. 224).

The general notation of a stationary ARMA(p,q) process X_t is given as:

$$X_t - \phi_1 X_{t-1} \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \quad (3.26)$$

wherein W_t has a White Noise distribution with zero mean and constant variance, i.e.:

$$\{W_t\} \sim WN(0, \sigma^2)$$

(Wang, 2015, p. 48)

[Note that the left hand side of the general notation denotes the AR(p) part and the right hand side denotes the MA(q) part.]

This can be rewritten as:

$$\phi(B)X_t = \theta(B)W_t \quad (3.27)$$

where : $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$

In other words for B, the backshift operator, holds:

$$B^j X_t = X_{t-j}; j = 0, 1, \cdots, p$$

$$B^j W_t = W_{t-j}; j = 0, 1, \cdots, q$$

(Which was shown in the previous chapters.)

For an ARMA(p,q) process, stationarity, causality, and invertibility are determined by the roots of the characteristic polynomials (Wang,2015, pp. 50-51):

- For **Stationarity** the roots of $\phi(z)$ must lie outside the unit circle.
- **Causality:** $\{X_t\}$ is a causal function of $\{W_t\}$ if there exists a sequence of constants $\{\psi_j\}$ such that $\{X_t\}$ can be expressed as a linear sum of past white noise terms:

$$X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j} \quad ; \quad \sum_{j=0}^{\infty} |\psi_j| < \infty. \quad (3.28)$$

This holds if and only if $\phi(z) \neq 0$ for all $|z| > 1$.

Here too, the roots of $\phi(z)$ must lie outside the unit circle for an ARMA(p,q) process to be causal.

- **Invertibility:** $\{X_t\}$ is an invertible function of $\{W_t\}$ if there exists a sequence of constants $\{\pi_j\}$ such that:

$$W_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad ; \quad \sum_{j=0}^{\infty} |\pi_j| < \infty. \quad (3.29)$$

This holds if and only if $\theta(z) \neq 0$ for all $|z| > 1$.

For an ARMA(p,q) process to be invertible, all roots of $\theta(z)$ must lie outside the unit circle.

Here, neither causality nor invertibility is an intrinsic property of the given time series $\{X_t\}$ alone, but rather a characteristic of the relationship between $\{X_t\}$ and $\{W_t\}$. Both properties are essential for ensuring a well-defined and meaningful representation of the ARMA process.

Let us study how this works for two concrete examples.

Example 1 (Brooks, 2008, p. 227):
ARMA(1,0) process (= AR(1) process):

$$X_t = 0.9X_{t-1} + W_t \quad (3.30)$$

Rewriting with backshift operator and finding characteristic equation $\phi(z)$:

$$(1 - 0.9B)X_t = W_t \quad \Rightarrow \quad \phi(z) = 1 - 0.9z \quad (3.31)$$

Solving $\phi(z) = 0$:

$$1 - 0.9z = 0 \quad \Rightarrow \quad z = \frac{1}{0.9} \quad (3.32)$$

Since $z > 1$, the process is causal and stationary. The MA component has $\theta(z) = 1$, meaning no roots, so the process is also invertible.

Example 2:
ARMA(2,1) process:

$$X_t - 0.75X_{t-1} + 0.5625X_{t-2} = W_t + 1.25W_{t-1} \quad (3.33)$$

Rewriting with backshift operator:

$$(1 - 0.75B + 0.5625B^2)X_t = (1 + 1.25B)W_t \quad (3.34)$$

Characteristic equations are:

$$\phi(z) = 1 - 0.75z + 0.5625z^2 \quad \wedge \quad \theta(z) = 1 + 1.25z \quad (3.35)$$

To check for stationarity, causality and invertibility, we solve the characteristic equations for $\phi(z)$ and $\theta(z)$:

$$\Rightarrow \text{roots of } \phi(z) \text{ are: } z_{1,2} = \frac{2}{3} \pm \frac{2\sqrt{3}i}{3} \quad \Rightarrow |z| = \frac{4}{3}$$

As $|z| > 1$, we may conclude 3.33 is stationary and causal.

$$\Rightarrow \text{root of } \theta(z) \text{ is: } z = -\frac{4}{5}$$

The process is not invertible, as $|z| < 1$

Thus the ARMA(2,1) process given in 3.33 is stationary and causal but not invertible.

ACF and PACF plots in ARMA Models

The autocovariance function (ACVF) and its visualization through the autocorrelation function (ACF) and partial autocorrelation function (PACF) are essential for understanding ARMA models.

- ACF of an ARMA model shows an initial decay that depends on the AR and MA components.
- PACF of an ARMA model displays cutoffs after a certain lag, which helps determining the AR order (i.e.: p).

One of the most widely used methodologies for estimating and applying ARMA models is the **Box-Jenkins approach**. Although ARMA models had already been utilized before, Box and Jenkins (1976) were the first to develop a systematic and practical method for their identification and estimation. Their approach consists of the following three steps (Brooks, 2008, pp. 230-233):

1. Identification: Determine the model order by analyzing time series plots, ACF, and PACF plots.
2. Estimation: Estimate parameters using methods such as least squares or maximum likelihood.

3. Diagnostic Checking: Evaluate model adequacy through residual diagnostics, overfitting tests, and statistical tests such as the Ljung-Box test.

Traditionally, the identification stage of ARMA modeling relied on graphical inspection of ACF and PACF plots. However, in practice, real-world data often does not display the clear-cut patterns seen in theoretical examples, making these plots difficult to interpret. To address this issue and reduce subjectivity, **information criteria** are commonly used instead.

Information criteria balance two factors: the residual sum of squares (RSS) and a penalty for adding extra parameters, which reduces degrees of freedom. The goal is to select the model order that minimizes the information criterion, ensuring that any added parameter improves model fit sufficiently to justify the complexity.

The most widely used information criteria are given below with their algebraic expressions.

- Schwarz's Bayesian Information Criterion (SBIC) = $\ln(\hat{\sigma}^2) + \frac{k}{T} \ln(T)$
- Hannan-Quinn Criterion (HQIC) = $\ln(\hat{\sigma}^2) + \frac{2k}{T} \ln(\ln(T))$
- Akaike's Information Criterion (AIC) = $\ln(\hat{\sigma}^2) + \frac{2k}{T}$

wherein:

- $\hat{\sigma}^2$ = residual variance $\equiv \frac{RSS}{T}$
- $k = p + q + 1$ = total number of parameters estimated
- T = sample size (number of observations)

(The criteria are presented in descending order of penalty stringency, from the strictest to the least strict.)

To determine which criterion is preferable when they suggest different model orders, it is important to note that the SBIC is strongly consistent but inefficient, whereas the AIC is not consistent but generally more efficient. The SBIC asymptotically selects the correct model order, while the AIC tends to select a model that is too large on average, even with an infinite amount of data.

3.4 Autoregressive Integrated Moving Average Processes

Autoregressive Integrated Moving Average (ARIMA) processes extend ARMA models by incorporating differencing to handle non-stationarity in time series data. An $ARIMA(p, d, q)$ process consists of an autoregressive (AR) component, a moving average (MA) component, and an integration (d) component that accounts for the number of times differencing is applied to achieve stationarity (Wang, 2015, p. 105).

An $ARIMA(p, d, q)$ process is defined as:

$$\phi(B)(1 - B)^d X_t = \theta(B)W_t, \quad (3.36)$$

where:

- B is the backshift operator,
- d represents the number of differencing operations required to make the process stationary,
- $\phi(B)$ defines the $AR(p)$ component,
- $\theta(B)$ defines the $MA(q)$ component,
- $W_t \sim WN(0, \sigma^2)$ is a white noise process.

The differencing operation transforms a non-stationary series $\{X_t\}$ into a stationary series:

$$Y_t = (1 - B)^d X_t, \quad (3.37)$$

which is then modeled as an $ARMA(p, q)$ process.

Stationarity and Differencing

A key assumption for ARIMA modeling is that differencing is applied only as needed to achieve stationarity. The process is considered:

- **Non-stationary** if differencing is required ($d > 0$),
- **Stationary** if no differencing is needed ($d = 0$), reducing to an ARMA(p, q) model.

The value of d is chosen based on the presence of trends in the data, often determined through visual inspection of time series plots and statistical tests such as the Augmented Dickey-Fuller (ADF) test.

ACF and PACF in ARIMA Models

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are useful for identifying the orders p , d , and q in an ARIMA model:

- The ACF of an ARIMA model typically shows slow decay due to non-stationarity before differencing.
- The PACF helps determine the AR order p by identifying significant lags.
- After differencing, the transformed series should exhibit stationary behavior, making the ACF and PACF more interpretable for ARMA modeling.

Chapter 4

Methodology

The methodology of this research combines literature review, practical application of time series analysis, and model fitting using RStudio. The first step in this process involved a comprehensive review of the time series literature, which included Brooks' *Econometrics for Finance*, Brockwell and Davis' *Time Series: Theory and Methods*, and lecture notes by Dewei Wang from the University of South Carolina. These resources provided the theoretical foundation needed for understanding univariate time series processes.

Following the review of the literature, I completed several sub-assignments suggested by my supervisor, Prof. Dr. S. Venetiaan. These tasks included reviewing a specific AR and MA process, computing their statistical properties and constructing ACF and PACF plots in RStudio.

After all this literature research, the need of the hour then was to find a quantitative dataset to apply the studied theory to. In other words it is sufficient, for this paper, to make use of a secondary (/existing) dataset and determine the best fitted type of time series model for it with RStudio. For this, my supervisor suggested links to various statistics and I arbitrarily chose to work with exchange rates series (the monthly USD purchase rates from 2004-2017, to be more specific), which were accessible via the site of the Centrale Bank van Suriname (CBvS).

The steps taken in the analysis are outlined below::

1. Importing the dataset
2. Checking that the data is considered as a time series. The data is most likely to be classified as a dataframe, so it would be necessary to first convert it to the class: "time series".
3. Installing the Time Series and Forecast packages from the library. The `ts()` function in R is used to convert a data frame (or vector) into a time series object. Here we need to specify the start time and frequency (e.g., yearly, quarterly, or monthly). This transformation allows R to treat the data as a sequence of observations over time. The forecast package (and others) provides tools for time series analysis. It includes functions like `auto.arima()` and `forecast()` that automate and simplify complex modeling.
4. Analyzing the statistical attributes by plotting the data, ACF and PACF functions and drawing conclusions about the data's stationarity. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are generated using `acf()` and `pacf()` functions. ACF shows the correlation of the series with its own lagged values, while PACF isolates the direct correlation at each lag. These plots help determine whether the series is stationary or if transformations (like differencing) are needed.
5. Performing the augmented dickey-fuller (adf) test for stationarity. The `adf.test()` function (from the `tseries` package) tests for stationarity. If the p-value is below a certain threshold (e.g., 0.05), the null hypothesis of a unit root (non-stationarity) is rejected, indicating the series is stationary. In case of non-stationary data, the data can be converted to stationary. The `auto.arima()` function from the forecast package automatically identifies the best ARIMA model by testing different combinations of parameters and selecting the one that minimizes an information criterion. After fitting a model, the `forecast()` function is used to generate future values based on the fitted model. This function provides point forecasts and confidence intervals, which are visualized using `plot()`.

The reliability of this research is ensured by using officially published exchange rate data from the Centrale Bank van Suriname (CBvS). Since CBvS is the primary financial institution responsible for reporting exchange rates, the dataset can be considered accurate and free from measurement errors. Additionally, the time series analysis techniques applied, such as stationarity tests (ADF test), autocorrelation analysis (ACF, PACF), and model selection (AR, MA, ARMA, ARIMA), are well-established statistical methods used in econometrics and financial research.

The validity of the study is ensured through the systematic application of time series modeling steps, aligning with the theoretical framework established in the literature. By following a structured approach that includes data visualization, stationarity testing, and model evaluation, the research ensures that conclusions drawn from the analysis are meaningful.

Furthermore, the selection of the best-fitting model is based on objective criteria such as information criteria (e.g., AIC, BIC). (The `auto.arima()` function in R selects the best-fitting ARIMA model based on the Akaike Information Criterion (AIC), ensuring a balance between model complexity and goodness-of-fit.)

Additionally, out-of-sample forecasting is used as a key validation method. The dataset consists of monthly USD exchange rates from 2004 to 2022, but only data from 2004 to 2017 is used for model estimation. (Data from this period is called training data.) The fitted model is then employed to forecast monthly exchange rates for the next five years. The validity of the model is assessed by comparing these forecasts with the actual observed exchange rates from 2018 to 2022. (Data from this period is called test data or out-of-sample data.)

To quantitatively evaluate the forecast accuracy, several performance metrics (accuracy measures) are computed in RStudio using the test data. These include:

- Mean Absolute Error (MAE): the average of the absolute differences between forecasted and actual values.
- Root Mean Squared Error (RMSE): the square root of the average of squared errors, which gives more weight to larger errors.
- Mean Absolute Percentage Error (MAPE): the average of absolute percentage differences, useful for understanding relative error.
- Mean Percentage Error (MPE): shows the average bias of forecasts, indicating whether predictions tend to over- or underestimate.
- Mean Absolute Scaled Error (MASE): a scale-independent measure that allows comparison across series or methods.
- ACF1: the autocorrelation of residuals at lag 1, which helps assess whether residuals are independent.
- Theil's U-statistic: a relative accuracy measure that compares forecast performance against a naïve benchmark.

Chapter 5

Results and Discussion

We will systematically analyze the results by examining each plot and its implications for the time series model selection.

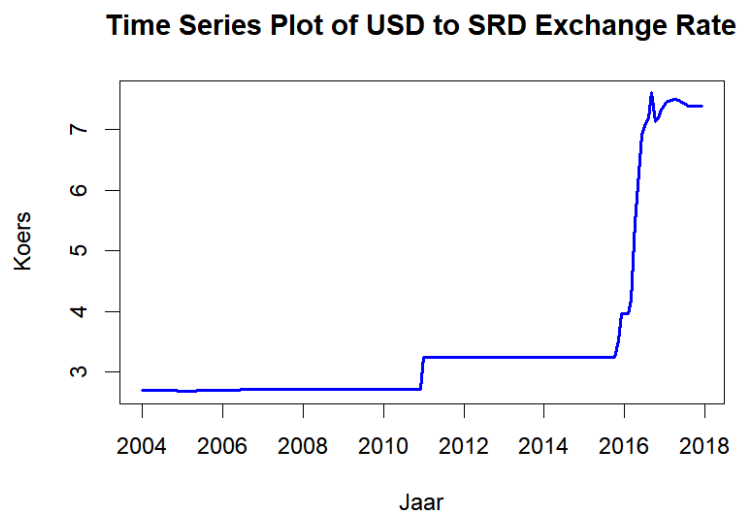


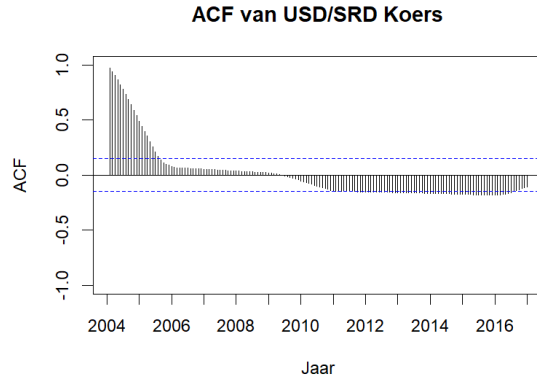
Figure 5.1: Time series plot of the exchange rate data

The time series plot of the exchange rate data reveals several important characteristics that influence the choice of an appropriate time series model.

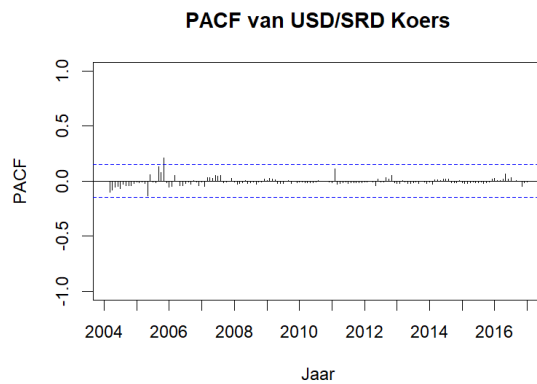
Firstly, the data exhibits clear structural breaks, particularly around 2011 and 2015-2016, where the exchange rate experiences sudden and significant increases. These abrupt changes suggest that the series is non-stationary, meaning its statistical properties, such as mean and variance, are not constant over time. Additionally, the series contains sharp jumps instead of a smooth trend, indicating the presence of external shocks or policy interventions that have caused abrupt fluctuations in the exchange rate. Such behavior is inconsistent with a purely autoregressive or moving average process, which typically assumes gradual changes over time. Furthermore, after the sharp rise in 2016, the exchange rate appears to stabilize at a higher level, with relatively smaller fluctuations. This suggests a potential shift in the underlying economic conditions that led to a new equilibrium level.

Given these characteristics, a standard ARMA model is unlikely to be appropriate without first transforming the data to achieve stationarity. This can be done by applying differencing techniques to remove trends and stabilize variance, ensuring that the model assumptions are met.

Having analyzed the time series plot, we now turn to the ACF and PACF plots to further investigate the data's dependence structure and identify potential model specifications.



(a) acf plot



(b) pacf plot

Figure 5.2: ACF and PACF plots of the exchange rate data

The ACF plot (Figure 5.2a) displays the correlation between the time series and its own lagged values. Each spike in the ACF plot represents the autocorrelation between the exchange rate at a given time and its past values at different lags (i.e. at previous points in time). Since the data are monthly, a lag of 1 corresponds to the correlation with the value one month earlier, lag 2 to two months earlier, and so on. In this case, the ACF shows high positive autocorrelations for the first several lags, which is indicative of a strong persistence in the series. This pattern is consistent with trending behavior and non-stationarity. Unlike seasonal time series, no distinct peaks are observed at lag 12 or its multiples (e.g., lag 12, 24, 36,...), which would

represent correlations with the same month in previous years. This suggests that the exchange rate data does not follow a yearly seasonal pattern. The ACF shows large spikes that persist across many lags and decay very slowly over time. While the decay is gradual, it does not follow a typical exponential pattern, which is often expected in stationary autoregressive processes. The fact that many of these autocorrelations lie outside the blue dashed confidence bands (which represent the approximate 95% confidence interval, typically $\pm \frac{1.96}{\sqrt{n}}$) suggests that the correlations at those lags are statistically significant, i.e. they are unlikely to have occurred by chance.

This persistence of significant autocorrelations over a long range of lags is a strong indicator of non-stationarity in the time series. In particular, the pattern is consistent with the presence of a unit root, meaning the process likely has a stochastic trend and requires differencing to become stationary. This behavior is typical of integrated processes such as ARIMA models, where the series needs to be differenced one or more times before a suitable model can be identified.

Additionally, the high autocorrelation across all lags makes it unlikely that a pure moving average (MA) model would be sufficient to model the data adequately. MA models typically exhibit a sharp cut-off in autocorrelation after a certain lag, which is not observed here. Instead, the data likely requires an autoregressive (AR) component or a combination of both AR and MA terms after differencing, making an ARIMA model a more appropriate choice.

The PACF plot (Figure 5.2b) illustrates the partial autocorrelations of the exchange rate series at different lags. Each spike represents the direct correlation between the exchange rate at a given time and its value at a specific lag, while controlling for the influence of the intervening lags. For example, the spike at lag 3 reflects the correlation between the current value and the value three months earlier, excluding the effects of months 1 and 2.

In this plot, we observe that most spikes fall within the 95% confidence bands, indicating that they are not statistically significant. There are no large spikes followed by a sharp cutoff, which would typically suggest an autoregressive (AR) structure of a particular order. Instead, the absence of dominant partial correlations and the generally low magnitude of the coefficients reinforce the idea that the series is non-stationary.

Combined with the ACF plot, this suggests that the original series contains a

stochastic trend and likely requires differencing to achieve stationarity. The lack of a clear cutoff in the PACF further implies that a pure AR model may not be suitable, and that a differenced model such as ARIMA is more appropriate for capturing the dynamics of the series.

Although the plots indicate that our data is non-stationary, we perform a stationarity test, namely the Augmented Dickey-Fuller (ADF) test.

The null hypothesis (H_0) of this test states that the data contains a unit root, implying non-stationarity, while the alternative hypothesis (H_1) suggests that the data is stationary.

```
data: ts_data_numeric
Dickey-Fuller = -1.7984, Lag order = 5, p-value = 0.6604
alternative hypothesis: stationary

> #augmented dickey fuller test:
> adf.test(ts_data_numeric)

Augmented Dickey-Fuller Test

data: ts_data_numeric
Dickey-Fuller = -1.7984, Lag order = 5, p-value = 0.6604
alternative hypothesis: stationary
```

Figure 5.3: ADF Test

The test yielded a p-value of 0.6604, which is greater than the standard significance level of 0.05.

Since the p-value exceeds 0.05, we fail to reject the null hypothesis, confirming that the exchange rate time series is non-stationary (Monigatti, 2023).

As stationarity is a key assumption for ARIMA modeling, transformations are required before fitting a suitable model.

To identify the best-fitting ARIMA model, the `auto.arima()` function was employed. This function evaluates a wide range of candidate ARIMA models by minimizing the Akaike Information Criterion (AIC), which balances model fit and complexity.

```

> rates_USD = auto.arima(ts_data, ic = "aic", trace = TRUE)

Fitting models using approximations to speed things up...

ARIMA(2,2,2)(1,0,1)[12]          : Inf
ARIMA(0,2,0)                    : -128.8969
ARIMA(1,2,0)(1,0,0)[12]        : -144.7512
ARIMA(0,2,1)(0,0,1)[12]        : -172.5668
ARIMA(0,2,1)                    : -174.5371
ARIMA(0,2,1)(1,0,0)[12]        : -160.1374
ARIMA(0,2,1)(1,0,1)[12]        : -158.1374
ARIMA(1,2,1)                    : -179.169
ARIMA(1,2,1)(1,0,0)[12]        : -164.6668
ARIMA(1,2,1)(0,0,1)[12]        : -177.1704
ARIMA(1,2,1)(1,0,1)[12]        : -162.6668
ARIMA(1,2,0)                    : -159.0531
ARIMA(2,2,1)                    : Inf
ARIMA(1,2,2)                    : Inf
ARIMA(0,2,2)                    : -174.7176
ARIMA(2,2,0)                    : -160.2448
ARIMA(2,2,2)                    : Inf

Now re-fitting the best model(s) without approximations...

ARIMA(1,2,1)                    : -185.2801

Best model: ARIMA(1,2,1)

```

Figure 5.4: Best Fitted Model

The differencing parameter (d) is automatically determined, ensuring that the transformed series meets the stationarity requirement. The model with the lowest AIC value was ARIMA(1,2,1), indicating that one autoregressive term, two levels of differencing, and one moving average term best capture the dynamics of the series. The re-evaluation of the best model without approximation confirmed this selection, resulting in an AIC of -185.2801. This low AIC value suggests a good trade-off between simplicity and explanatory power, and supports the use of ARIMA(1,2,1) for forecasting the series.

```

> summary(rates_USD)
Series: ts_data
ARIMA(1,2,1)

Coefficients:
      ar1      ma1
    0.3996 -0.9711
s.e.  0.0783  0.0279

sigma^2 = 0.01849:  log likelihood = 95.64
AIC=-185.28  AICc=-185.13  BIC=-175.94

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.008273302 0.1343559 0.03580651 0.1966385 0.7156306 0.09733083
              ACF1
Training set -0.08026811

```

Figure 5.5: Model summary

Once the model was selected, the corresponding coefficients were estimated using maximum likelihood estimation. For the fitted ARIMA(1,2,1) model, the estimated parameters are:

- AR(1) = 0.3996 (SE = 0.0783)
- MA(1) = -0.9711 (SE = 0.0279)
- $\sigma^2 = 0.01849$

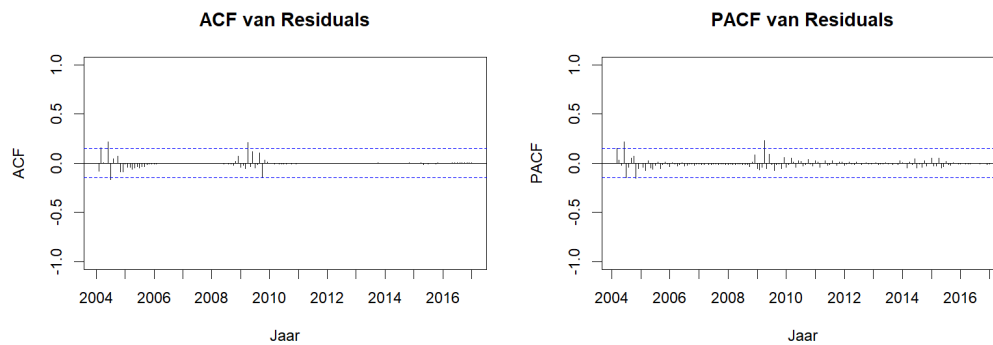
These coefficients were then used to generate the out-of-sample forecasts for the period 2018–2022. The accuracy and behavior of the forecast depend on these estimated values, making them a crucial part of the model fitting process.

After fitting the ARIMA(1,2,1) model and differencing the data to achieve stationarity, it is important to evaluate whether the model has adequately captured the structure in the time series. This is done by examining the ACF and PACF plots of the model residuals.

Residuals are the differences between the observed values and the values predicted by the model. If the residuals resemble white noise — meaning they are uncorrelated, have constant variance, and a zero mean — this suggests that the model has successfully extracted all systematic patterns from the data.

Specifically, the ACF and PACF plots of the residuals help detect any remaining autocorrelation. If significant spikes remain outside the confidence

bands, it indicates that some temporal dependence is still present, and the model may require further refinement. However, if no significant autocorrelations are observed, it implies that the residuals are essentially random, and the model is well-specified.



(a) acf plot

(b) pacf plot

Figure 5.6: ACF and PACF plots for the transformed series

The ACF plot (Figure 5.6a) exhibits that most autocorrelations fall within the 95% confidence bounds, with only minor deviations at a few early lags. Similarly, the PACF plot (Figure 5.6b) shows no significant autocorrelations beyond the first few lags, indicating that the selected ARIMA model adequately captures the dependence structure and temporal dynamics in the data.

The next step involves utilizing the fitted model for forecasting future exchange rate values and evaluating its predictive performance against out-of-sample data.

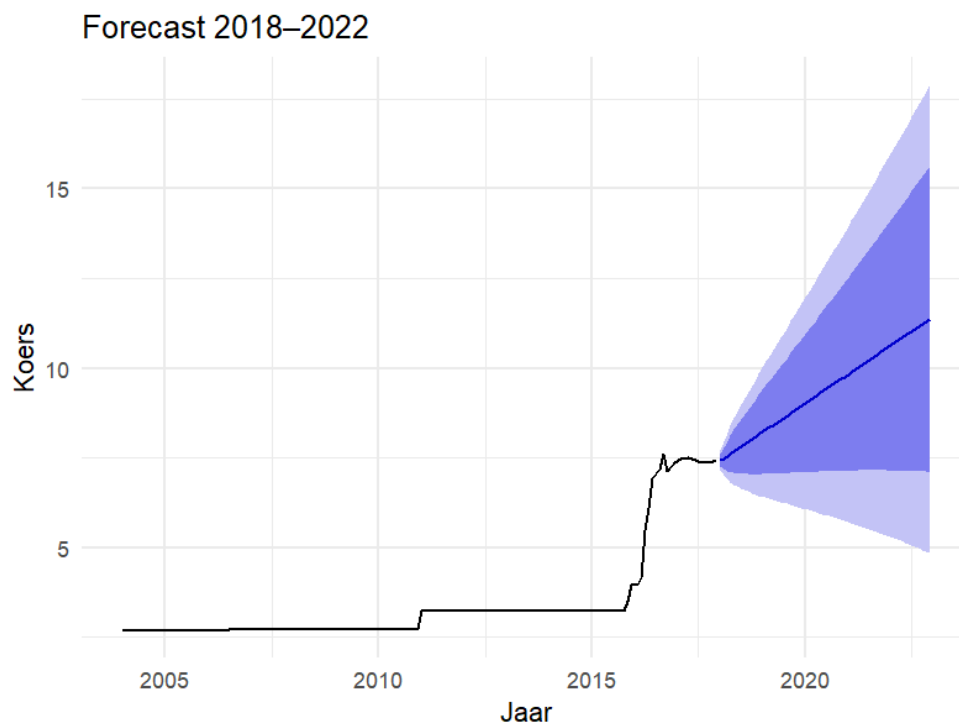


Figure 5.7: Forecast

Figure 5.7 presents the forecast of the time series for the period 2018–2022, including 80% and 95% confidence intervals. The forecast (blue line) continues the historical trend observed in the data, while the shaded areas represent the uncertainty associated with the predictions. The widening intervals reflect increasing uncertainty as the forecast horizon extends. This visualization demonstrates the model’s anticipated range of future values under the assumption that historical patterns persist.

```

> accuracy(forecast_result, ts_test)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.008273302 0.1343559 0.03580651 0.1966385 0.7156306 0.09733083
Test set     4.073262416 7.4788913 5.19764795 13.8418245 29.0295845 14.12847326
      ACF1 Theil's U
Training set -0.08026811 NA
Test set     0.93562051 4.416756

```

Figure 5.8: Accuracy Measures

Figure 5.8 shows the accuracy measures after comparing the forecasts produced by the selected ARIMA(1,2,1) model for the period 2018–2022 with the actual out-of-sample values.

The training set metrics (based on data from 2004–2017) indicate very low error across all measures, with a MAE of 0.036 and a MAPE of only 0.72%. This suggests a good model fit on the in-sample data.

In contrast, the test set metrics (2018–2022) reveal a substantial increase in forecast error. The RMSE jumps to 7.48 and the MAPE to 29.03%, indicating that the model significantly underperforms when predicting future data. The large MPE (13.84%) and high MASE (14.13) further confirm this.

The ACF1 value of 0.94 in the test set suggests that residuals from the forecasts are strongly autocorrelated, which violates the assumption of white noise residuals. (Although the residuals on the training data resemble white noise, as shown by the ACF and PACF plots, the high ACF1 value in the forecast errors indicates that the model fails to generalize well to unseen data, suggesting poor out-of-sample performance.) The Theil’s U statistic of 4.42 also indicates poor predictive accuracy, as values greater than one typically mean that a naive forecast would outperform the fitted model.

These findings suggest that although the ARIMA(1,2,1) model fits the historical (training) data well, it fails to capture the dynamics of the exchange rate during the out-of-sample (test) period. This discrepancy may stem from structural breaks or external shocks (e.g., inflation, policy changes, or economic instability) that the model does not account for.

Forecast vs Werkelijke Waarden (2018–2022)

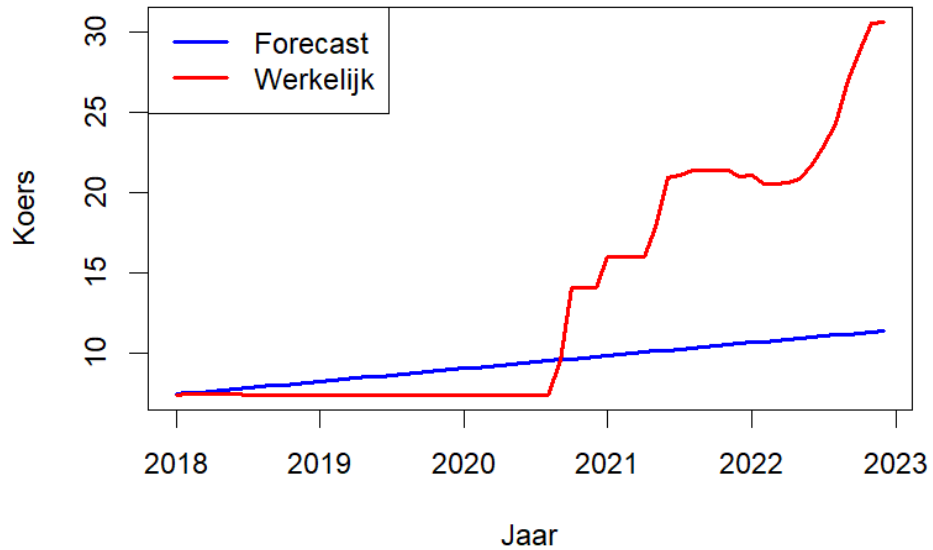


Figure 5.9: accuracy

Figure 5.9 compares the forecasted values (blue) with the actual observed values (red) over the period 2018–2022. While the model captures the general upward trend, it significantly underestimates the step increases observed from 2020 onwards. This again demonstrates that the model was unable to anticipate abrupt changes or structural shifts, highlighting its limitations in adapting to unexpected dynamics or exogenous shocks during the forecast period.

Interpretation of Forecast Performance and Economic Context

The results indicate that while the ARIMA(1,2,1) model provides a reasonable in-sample fit for the USD to SRD exchange rate data from 2004 to 2017, it fails to produce reliable forecasts for the out-of-sample period (2018–2022). This forecasting inaccuracy can be attributed to the model's inability to account for external shocks and structural changes that occurred in the Surinamese economy during and after 2015.

From 2015 onwards, Suriname experienced several significant macroeconomic disruptions, including sharp increases in inflation, heavy government borrowing, currency devaluations, and changing monetary policies. These developments likely introduced structural breaks in the exchange rate dynamics, changes that a purely univariate model such as ARIMA is not equipped to handle, as it relies solely on past values of the variable itself.

These insights highlight the limitations of relying solely on univariate models for forecasting in volatile economic environments, and point to the need for more flexible or multivariate approaches, as explored in the recommendations section.

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

To conclude, a time series, formally defined as a stochastic process, represents a sequence of random variables indexed by time, and is particularly useful for identifying temporal patterns and making future predictions. Time series analysis thus holds significant value in various real-world domains, including finance, economics, and policy planning. In this thesis we investigated the research question:

Which of the four studied univariate time series models; AR, MA, ARMA, and ARIMA, best fits the given dataset? The data consist of monthly USD to SRD buying exchange rates from January 2004 to December 2017, obtained from the Central Bank of Suriname.

Throughout the study, we examined several key statistical concepts such as white noise, zero-mean processes, stationarity, autocovariance and autocorrelation functions (ACVF and ACF), the partial autocorrelation function (PACF), the backshift operator, as well as the notions of causality and invertibility. The best fitting model, identified through the `auto.arima()` function in RStudio, was the ARIMA(1,2,1) model. This model performed well in capturing the dynamics of the in-sample data (2004–2017), as confirmed by residual diagnostics that indicated approximate white noise be-

havior. However, when evaluated through out-of-sample forecasting for the period 2018–2022, the model’s performance declined significantly. The predicted values deviated considerably from the actual observed exchange rates, with large forecast errors and high accuracy metrics (RMSE and MAPE). These results suggest that while the ARIMA(1,2,1) model effectively fits the historical data, it lacks robustness in forecasting future values, possibly due to structural changes, external economic shocks, or non-linear effects not captured by the model.

6.2 Recommendations

In light of these findings, the following recommendations can be proposed for future research:

- **Incorporate exogenous variables:** Future models should include relevant external factors (e.g., inflation rate, foreign debt levels, commodity prices, or interest rates) to better capture the economic environment. This could be achieved using ARIMAX or regression with ARIMA errors.
- **Consider multivariate models:** Employing models such as the Vector Autoregressive (VAR) model would allow for the simultaneous modeling of multiple interrelated economic variables, thus capturing dynamic interdependencies more effectively.
- **Explore non-linear and data-driven methods:** Since financial and economic data often exhibit non-linear patterns and sudden regime shifts, the use of machine learning techniques (e.g., neural networks, random forests, or support vector machines) may enhance predictive performance, particularly in the presence of structural instability.
- **Test for structural breaks:** Future studies may also benefit from explicitly identifying and modeling structural breaks or using regime-switching models (such as Markov switching models) that can accommodate changes in underlying dynamics over time.

Appendices: R-scripts

```
1 #Benodigde pakketten installeren en laden
2 install.packages("readxl", dependencies = TRUE)
3 library(readxl)
4 install.packages("forecast")
5 install.packages("tseries")
6 library(forecast)
7 library(tseries)
8 install.packages("ggplot2")
9 library(ggplot2)
10 getwd()
11 setwd("/Users/Gebruiker/Downloads/Afstudeerprocess")
12 # Pad naar het bestand
13 file_path <- "thesis data.xlsx"
14 library(readxl)
15 # Data inlezen, alleen Sheet1
16 df <- read_excel(file_path, sheet = "Sheet1")
17
18 # Bekijk de eerste rijen om te controleren of alles goed is
   ingelezen
19 head(df)
20
21 # Controleer de kolomnamen en pas ze aan indien nodig
22 colnames(df) <- c("Datum", "Koers")
23
24 # Zet de "Datum" kolom om naar een datumformaat
25 df$Datum <- as.Date(paste0(df$Datum, "-01"), format="%Y-%m
   -%d") # Maandelijks data
26
27 # Koers kolom van character omzetten naar numeric
28 df$Koers <- as.numeric(df$Koers)
```

```

29
30 # Zet de dataset om naar een tijdreeksobject (ts)
31 start_year <- as.numeric(format(min(df$Datum), "%Y")) #
   Startjaar bepalen
32 start_month <- as.numeric(format(min(df$Datum), "%m")) #
   Startmaand bepalen
33 ts_data <- ts(df$Koers, start = c(start_year, start_month),
   frequency = 12)
34 #ts_data <- ts(df$Koers, start = c(2004, 1), frequency =
   12)
35
36 # Controleer de tijdreeks
37 print(ts_data)
38 class(ts_data)
39 # Plot de tijdreeks om een eerste indruk te krijgen
40 plot(ts_data, main = "Time Series Plot of USD to SRD
   Exchange Rate",
41       xlab = "Jaar", ylab = "Koers", col = "blue", lwd = 2)
42
43 acf_result <- acf(ts_data, lag.max = 156, plot = FALSE)
44
45 # Plot zelf met correcte x-as labels (maanden = lags)
46 plot(acf_result$lag[-1] * 12, acf_result$acf[-1], type = "h
   ",
47       main = "ACF van USD/SRD Koers", xlab = "Jaar", ylab =
   "ACF",
48       xaxt = "n", ylim = c(-1, 1))
49 abline(h = 0)
50 abline(h = c(-1.96/sqrt(length(ts_data)), 1.96/sqrt(length(
   ts_data))), col = "blue", lty = 2)
51
52 # X-as: elk jaar als label (12 maanden = 1 jaar)
53 axis(1, at = seq(0, 156, by = 12), labels = 2004:2017)
54
55 pacf_result <- pacf(ts_data, lag.max = 156, plot = FALSE)
56
57 plot(pacf_result$lag[-1] * 12, pacf_result$acf[-1], type =
   "h", main = "PACF van USD/SRD Koers", xlab = "Jaar",
   ylab = "PACF", xaxt = "n", ylim = c(-1, 1))
58 abline(h = 0)

```

```

59 abline(h = c(-1.96/sqrt(length(ts_data)), 1.96/sqrt(length(
    ts_data))), col = "blue", lty = 2)
60
61 # X-as: elk jaar als label (12 maanden = 1 jaar)
62 axis(1, at = seq(0, 156, by = 12), labels = 2004:2017)
63
64
65 #Comment on Stationarity and use ARIMA for converting non-
    stationary data to stationary data
66 library(tseries)
67 #augmented dickey fuller test:
68 adf.test(ts_data)
69 #H_0 = Data is non-stationary
70 #H_1 = Data is stationary
71 #p-value = 0.6604 > 0.05, so non-stationary data
72 library(forecast)
73 rates_USD = auto.arima(ts_data, ic = "aic", trace = TRUE)
74 summary(rates_USD)
75 acf_newresult <- acf(rates_USD$residuals, lag.max = 156,
    plot = FALSE)
76
77 plot(acf_newresult$lag[-1] * 12, acf_newresult$acf[-1],
    type = "h",
78     main = "ACF van Residuals", xlab = "Jaar", ylab = "ACF
    ",
79     xaxt = "n", ylim = c(-1, 1))
80 abline(h = 0)
81 abline(h = c(-1.96/sqrt(length(ts_data)), 1.96/sqrt(length(
    ts_data))), col = "blue", lty = 2)
82
83 # X-as: elk jaar als label (12 maanden = 1 jaar)
84 axis(1, at = seq(0, 156, by = 12), labels = 2004:2017)
85
86 pacf_newresult <- pacf(rates_USD$residuals, lag.max = 156,
    plot = FALSE)
87
88 plot(pacf_newresult$lag[-1] * 12, pacf_newresult$acf[-1],
    type = "h", main = "PACF van Residuals", xlab = "Jaar",
    ylab = "PACF", xaxt = "n", ylim = c(-1, 1))
89 abline(h = 0)

```

```

90 abline(h = c(-1.96/sqrt(length(ts_data)), 1.96/sqrt(length(
    ts_data))), col = "blue", lty = 2)
91
92 # X-as: elk jaar als label (12 maanden = 1 jaar)
93 axis(1, at = seq(0, 156, by = 12), labels = 2004:2017)
94
95 forecast_result <- forecast(rates_USD, h = 60)
96 library(ggplot2)
97 # Plot van forecast
98 #plot(forecast_result, main = "Forecast -20182022", xlab =
    "Jaar", ylab = "Koers", col = "blue", lwd = 2)
99 forecast::autoplot(forecast_result) + ggtitle("Forecast -
    20182022") + xlab("Jaar") + ylab("Koers") + theme_
    minimal()
100
101 test_file <- "out_of_sample_data.xlsx"
102 test_data <- read_excel(test_file, sheet = "Sheet1")
103
104 colnames(test_data) <- c("Datum", "Koers")
105
106 # Zet kolom om naar numeriek
107 test_data$Koers <- as.numeric(test_data$Koers)
108 test_data$Datum <- as.Date(paste0(test_data$Datum, "-01"),
    format="%Y-%m-%d")
109
110 # Maak tijdreeks van testdata
111 ts_test <- ts(test_data$Koers, start = c(2018, 1),
    frequency = 12)
112
113 # Bereken foutmaten van de forecast
114 accuracy(forecast_result, ts_test)
115
116 # Combineer beide in één plot
117 plot(forecast_result$mean, type = "l", col = "blue", lwd =
    2, ylim = range(c(ts_test, forecast_result$mean)), ylab
    = "Koers", xlab = "Jaar", main = "Forecast vs Werkelijke
    Waarden -(20182022)")
118 lines(ts_test, col = "red", lwd = 2)
119 legend("topleft", legend = c("Forecast", "Werkelijk"), col
    = c("blue", "red"), lty = 1, lwd = 2)

```

References

- Brockwell, P. J., & Davis, R. A. (2009). *Time series: Theory and methods* (2nd ed.). Springer.
- Brooks, C. (2008). *Introductory econometrics for finance* (2nd ed.). Cambridge University Press.
- Brownlee, J. (2020, August, 14). *White noise time series with Python*. Machine Learning Mastery. Retrieved from <https://machinelearningmastery.com/white-noise-time-series-python/>
- Chyun, C. (2019, March 19). *Time series and moving average with R*. Medium. Retrieved from <https://medium.com/@chyun55555/time-series-and-moving-average-with-r-f318af5b62a4>
- Coghlan, A. (n.d.). *A little book of R for time series*. Retrieved from <https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html>
- Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis: With applications in R* (2nd ed.). Springer.
- Wang, D. (2015). *Time series analysis: Lecture notes*. Department of Statistics, University of South Carolina. <https://people.stat.sc.edu/wang528/Stat%20720/STAT720%20Notes.pdf>
- Hyndman, R. J., & Athanasopoulos, G. (n.d.). *Forecasting: Principles and practice*. OTexts. Retrieved from <https://otexts.com/fpp3/accuracy.html>
- Hyndman, R. J., & Athanasopoulos, G. (n.d.). *Seasonal ARIMA models*. OTexts. Retrieved from <https://otexts.com/fpp2/seasonal-arma.html>

- Tableau. (n.d.). *Time Series Analysis: Definition, Types, Techniques, and When It's Used* Retrieved from <https://www.tableau.com/learn/articles/time-series-analysis>
- Monigatti, L. (2023, April 11). *Stationarity in time series: A comprehensive guide*. Towards Data Science. Retrieved from <https://towardsdatascience.com/stationarity-in-time-series-a-comprehensive-guide-8beabe20d68>
- Pandian S., (n.d.). *A comprehensive guide to time series analysis*. Retrieved from <https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/>
- Radečić D., (2023, November 22). *Time Series Analysis in R: How to Read and Understand Time Series Data*. Retrieved from <https://www.appsiilon.com/post/r-time-series-analysis>
- Haque, M. M. (2015). *Behaviour of ACF and PACF for stationary process [Figure]*. In *Time Series Analysis: Forecasting and Control*. ResearchGate. Retrieved from https://www.researchgate.net/figure/Behaviour-of-ACF-and-PACF-for-stationary-process_tbl3_280832347
- Centrale Bank van Suriname. (n.d.). *Gemiddelde maandkoersen*. Retrieved from <https://www.cbvs.sr/statistieken/financiele-markten-statistieken/gemiddelde-maandkoersen>