

# Time Series Analysis and Forecasting

Matthew Juan z5259434

# Contents

<b>1</b>	<b>Introduction to Time Series'</b>	<b>4</b>
<b>2</b>	<b>Properties of Time Series Data</b>	<b>5</b>
2.1	Trend . . . . .	5
2.2	Seasonality . . . . .	5
2.3	Residual . . . . .	5
2.4	Additive vs Multiplicative . . . . .	5
2.5	Stationarity . . . . .	6
2.5.1	How to test for stationarity . . . . .	6
2.5.2	Methods of making a time series stationary . . . . .	6
<b>3</b>	<b>Time Series Forecasting Models</b>	<b>11</b>
3.1	ARIMA . . . . .	11
3.1.1	Autoregressive Models . . . . .	11
3.1.2	Moving Average Models . . . . .	11
3.1.3	Putting It All Together . . . . .	12
3.1.4	Seasonal Variation . . . . .	12
3.2	Exponential Smoothing . . . . .	12
3.2.1	Formulas . . . . .	12
3.2.2	Seasonal Variation . . . . .	12
<b>4</b>	<b>Implementation</b>	<b>12</b>
4.1	ARIMA . . . . .	12
4.2	Exponential Smoothing . . . . .	12
<b>5</b>	<b>Results</b>	<b>12</b>
<b>6</b>	<b>Conclusion</b>	<b>12</b>



## 1 Introduction to Time Series'

Time series are found all over the place and are extremely useful to analyse and more importantly, forecast. There are many situations where you want to predict past your last data point. Use cases include:

- Traders predicting the stock market so they know when to buy or sell
- Data centre operators identifying when a critical system may go down or is acting irregularly
- Retail store managers knowing how to maximise what stock they order to maximise profit and minimise items that won't sell

Unfortunately, just like humans, it is impossible to predict the future. However, by analysing the time series, it is possible to formulate statistical properties that can be used to give a rough estimate of where the data is heading.

## 2 Properties of Time Series Data

A time series can fundamentally be thought of as a random variable consisting of a signal and noise. This signal and noise are formed by 3 components which are core to understanding time series', trend, seasonality, and residual. Understanding their role and what they represent as well as some other key definitions are needed before diving in to analysing time series data.

### 2.1 Trend

The first component is trend. Trend can be viewed as how the data is changing overall. Is the data increasing or decreasing.

### 2.2 Seasonality

Seasonality refers to a regular occurring pattern that emerges within a given period. E.g peaks or valleys in the data during certain months or hours in the day.

### 2.3 Residual

Residual are the parts left over after fitting a model to the data. In many time series models, it is defined as the difference between the actual observation and the predicted value.

### 2.4 Additive vs Multiplicative

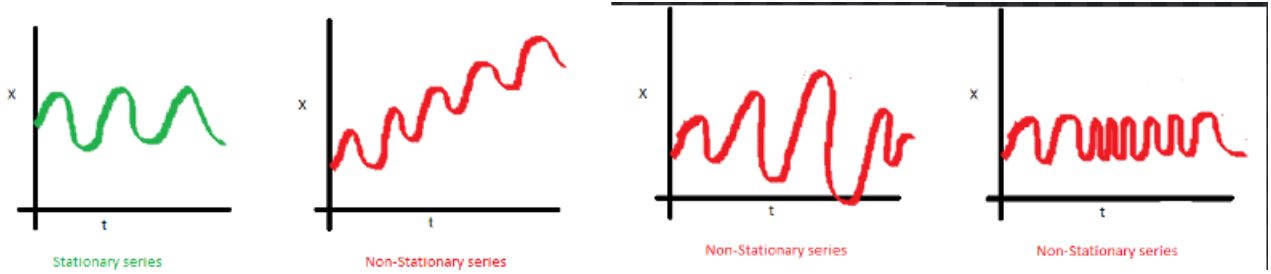
The trend and seasonality in a time series can be classified as being additive or multiplicative. In additive time series', the data exhibits a general upwards or downwards trend but does so at a constant rate. Peaks and valleys in your data are roughly the same size. In contrast, multiplicative time series data exhibits peaks or valleys that are amplified within the seasonal activity. The data becomes exaggerated and the difference between peaks at the start are very different than at the tail.

## 2.5 Stationarity

The concept of stationarity is crucial for time series analysis and For time series analysis to be most effective, the data needs to form a stationary process. A random variable that forms a time series is classified stationary if it's probability distribution is constant throughout the full range of times. Formally, if  $T$  is the set of all times in your time series of length  $n$ , a random process  $X$  is stationary if the joint probability distribution at times  $t_1, \dots, t_n \in T$  and  $t_1 + \Delta, \dots, t_n + \Delta \in T$ ,  $\Delta \in \mathbb{R}$  are equal. That is

$$F_X(x_{t_1}, \dots, x_{t_n}) = F_X(x_{t_1+\Delta}, \dots, x_{t_n+\Delta}) \quad \forall x_{t_1}, \dots, x_{t_n} \in T$$

Intuitively, this means a stationary time series has constant propertues such as mean, variance, covariance, etc. Below shows the differences between stationary and non-stationary processes.



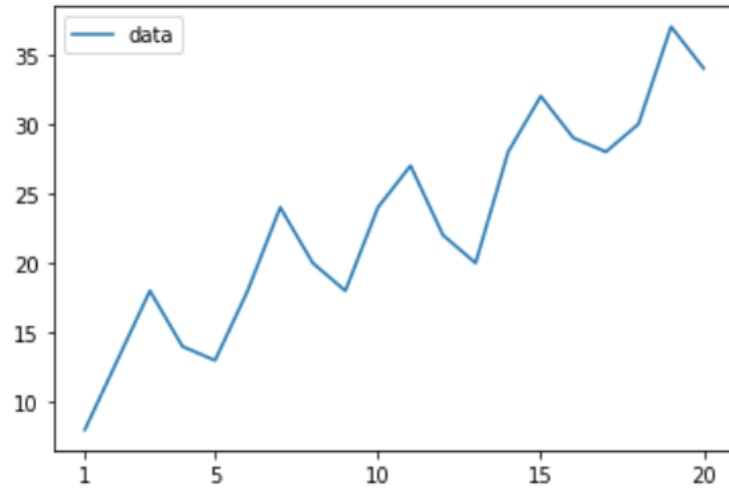
### 2.5.1 How to test for stationarity

To determine whether a time series is stationary or not, the Augmented Dickey Fuller(ADF) test can be performed. ADF performs a hypothesis test and upon observing p-value  $p < \alpha$  where  $\alpha$  is commonly 0.05, the hypothesis is rejected and the data is considered to be stationary.

### 2.5.2 Methods of making a time series stationary

If you have a non-stationary time series, there are a few ways to make it stationary. Let's use the below data as an example.

data  
8  
13  
18  
14  
13  
18  
24  
20  
18  
24  
27  
22  
20  
28  
32  
29  
28  
30  
37  
34

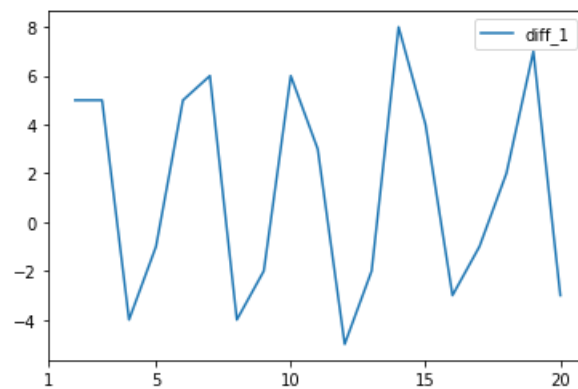


The most common is differencing. Differencing involves creating a new dataset where the values at time  $t$  is equal to the difference between the original value at time  $t$  and original value at  $t - 1$ . So if  $y'$  is the new dataset and  $y$  is the original data, then

$$y'_t = y_t - y_{t-1}.$$

This process can be repeated multiple times to obtain a potentially more stationary series known as higher order differencing. As you can see, the data looks much more stationary than before.

data	diff_1
8	NaN
13	5.0
18	5.0
14	-4.0
13	-1.0
18	5.0
24	6.0
20	-4.0
18	-2.0
24	6.0
27	3.0
22	-5.0
20	-2.0
28	8.0
32	4.0
29	-3.0
28	-1.0
30	2.0
37	7.0
34	-3.0



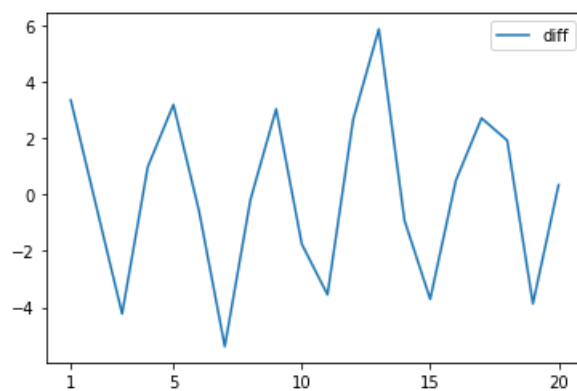


Another common technique is to detrend by model fitting. You simply fit a regression model, such as linear or exponential, and create a new dataset where each value at time  $t$  is equal to the difference between the predicted value and the original value.

$$y'_t = pred_t - y_t.$$

Again, the data is much more stationary than the original.

data	predicted	diff
8	11.357143	3.357143
13	12.566917	-0.433083
18	13.776692	-4.223308
14	14.986466	0.986466
13	16.196241	3.196241
18	17.406015	-0.593985
24	18.615789	-5.384211
20	19.825564	-0.174436
18	21.035338	3.035338
24	22.245113	-1.754887
27	23.454887	-3.545113
22	24.664662	2.664662
20	25.874436	5.874436
28	27.084211	-0.915789
32	28.293985	-3.706015
29	29.503759	0.503759
28	30.713534	2.713534
30	31.923308	1.923308
37	33.133083	-3.866917
34	34.342857	0.342857



Other techniques include seasonal differencing, where you perform differencing but with value at time  $t$  and  $t - \rho$  where  $\rho$  is the period of your data, or performing non-linear transformation your dataset such as logging. These can be combined with previous techniques to increase stationarity.

## 3 Time Series Forecasting Models

There are many types of forecasting models but this paper will look arguably the most popular two. ARIMA (Auto-Regressive Integrated Moving Average) and Exponential Smoothing models.

### 3.1 ARIMA

ARIMA is an extension of the ARMA or Box-Jenkin model popularised by George E. P. Box and Gwilym Jenkins in 1970. The main difference is that ARIMA simply allows for an extra parameter  $d$  which defines how much differencing should be done before fitting the model, whereas ARMA assumes the data is stationary already. ARIMA is actually a combination of two models, an autoregressive model (AR) and a moving averages model (MA). We will dive into what makes up both models and how ARIMA is able to effectively combine the two to create an even stronger model. As previously mentioned, a random variable in the form of a time series can be viewed as being a combination of signal and noise. The goal of ARIMA is to filter the noise and extrapolate the signal.

#### 3.1.1 Autoregressive Models

The term *autoregressive* defines that the model uses a linear combination of previous values of a variable and learned predictors. This differs from a typical regression model which predicts based on a linear combination of the input features and learned predictors. Autoregressive models has one parameter  $p$  which defines the order of the model. This parameter defines how many previous values influence the one being predicted. An  $AR(1)$  model is simply defines a model in which the predicted value is equal to some linear combination of the value at the previous time step

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t,$$

where  $\varepsilon_t$  is some white noise. An  $AR(2)$  model uses the previous two timesteps as so on. Thus, an  $AR(p)$  model can be generalised to

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

We can see that we are using previous known values as the features to predict the future.

#### 3.1.2 Moving Average Models

In a sort of similar manner, Moving Average (MA) models use the errors of past forecasts to predict. An MA model instead has parameter  $q$  so an  $MA(q)$  model

can be written similarly as

$$y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}.$$

### **3.1.3 Putting It All Together**

### **3.1.4 Seasonal Variation**

## **3.2 Exponential Smoothing**

### **3.2.1 Formulas**

### **3.2.2 Seasonal Variation**

## **4 Implementation**

### **4.1 ARIMA**

### **4.2 Exponential Smoothing**

## **5 Results**

## **6 Conclusion**

## **7 Resources**

[https://www.probabilitycourse.com/chapter10/10\\_1\\_4\\_stationary\\_processes.php](https://www.probabilitycourse.com/chapter10/10_1_4_stationary_processes.php)  
<https://www.statology.org/detrend-data/>