

Time Series Analysis and Forecasting

Matthew Juan z5259434

Contents

1	Introduction to Time Series'	4
2	Properties of Time Series Data	5
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
3	Time Series Forecasting Models	9
3.1	ARIMA	9
3.1.1	Autoregressive Models	9
3.1.2	Moving Average Models	10
3.1.3	Putting It All Together	10
3.1.4	Seasonal Variation	11
3.2	Exponential Smoothing	11
3.2.1	Formulas	11
3.2.2	Seasonal Variation	11
4	Implementation	11
4.1	ARIMA	11
4.2	Exponential Smoothing	11
5	Results	11
6	Conclusion	11

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, also commonly referred to as errors, 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

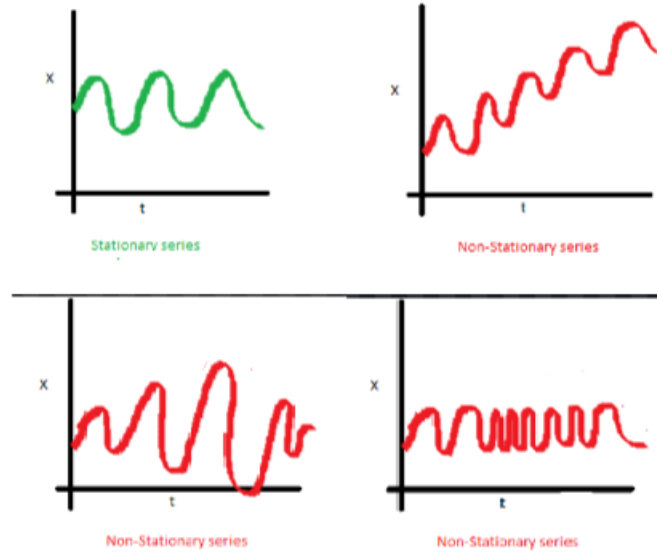
The trend and seasonality in a time series can be classified as being additive or multiplicative. In additive time series', the data exhibits a general upwards or downwards trend but does so at a constant rate. Peaks and valleys in your data are roughly the same size. In contrast, multiplicative time 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

Before going into analysis, the time series needs to be stationary. A random variable that forms a time series is classified stationary if its 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 properties such as mean, variance, covariance, etc. Below shows the differences between stationary and non-stationary processes.



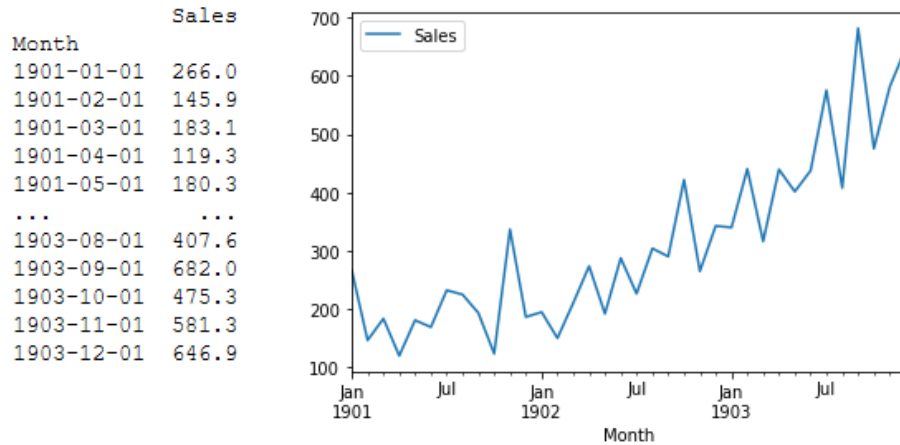
2.5.1 How to test for stationarity

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 α 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 Shampoo Sales dataset which describes the monthly number

of sales of shampoo over a 3 year period. The data here has a distinct multiplicative increasing trend and is clearly non stationary. Using the ADF test, the p-value comes out to $1 > 0.05$ so there is weak evidence against the null hypothesis so the data is non-stationary.



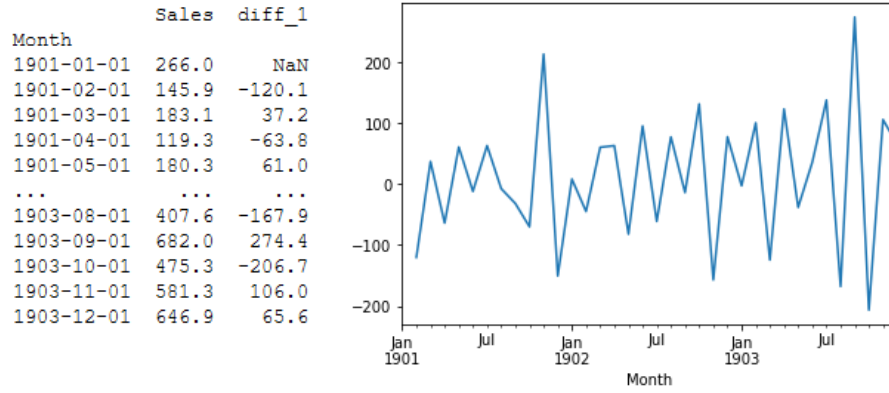
The most common is differencing. Differencing involves creating a new dataset where the values at time t is equal to the difference between the original value at time t and original value at $t - 1$. So if y is the new dataset and Y is the original data, then

$$y_t = Y_t - Y_{t-1}.$$

Sometimes the differenced data will not make the data stationary enough and it may be necessary to difference the data again known as 2nd order differencing. This process is applicable to higher orders Note, the 2nd order difference is not $y_t = Y_t - Y_{t-2}$ but it is the first difference of the first difference so it comes out to

$$y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}).$$

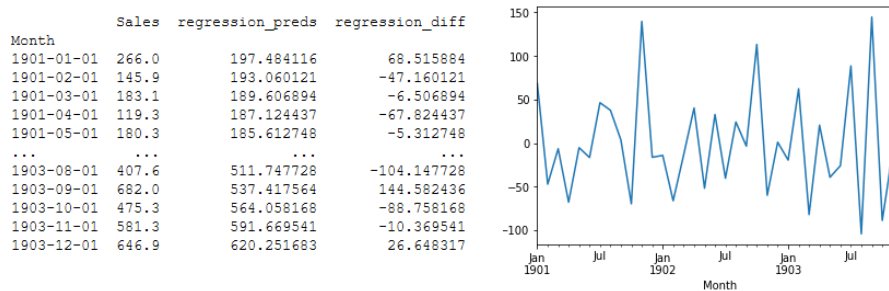
The p-value from the ADF test is about $1.8 \times 10^{-10} < 0.05$ so the data is now stationary.



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 original value and the predicted value. When fitted with a 2nd order polynomial regression model, the resulting differences produced a p-value of about $3.4 \times 10^{-10} < 0.05$ so again, the data is now stationary.

$$y_t = Y_t - \text{pred}_t.$$

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



There are various other techniques that can be performed and combined with one another to make your data more stationary. If your data has a seasonal pattern to it, seasonal differencing can be performed where you take the difference between t and $t - \rho$ where ρ is the period of your data. Non-linear transformations your dataset such as logging can also prove effective at making data stationary.

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

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

where μ is the average period to period change and ε_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

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

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

Let's see how this works for the first differenced data we saw before. For my implementation, to estimate the parameters I used a linear least squares solver. Firstly, the lags are calculated and then we estimate parameters using these lags and our original time series values. We simply do a dot product between the coefficients and add it to our value for μ to retrieve our predictions.

This is the resulting predictions of a few $AR(p)$ models.

There isn't too much change apart from the beginning of the data series where no lags could be properly calculated so they were filled with 0.

The already known lags are used to forecast one time step ahead. This forecast is then assumed to be the observed value and further values can be forecasted in a similar manner.

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

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

If your data is properly stationary, the error terms produced, ε_t , should form white noise and form a normal distribution with 0 mean and σ^2 variance.

Parameters were again estimated using a linear least squares solver and this is the results of a few $MA(q)$ models on the same data. Observe that the residuals produced do generally form a normal distribution with 0 mean.

3.1.3 Putting It All Together

When we combine differencing, an AR, and an MA model, we obtain an ARIMA model. It's general equation is as follows

$$y'_t = \mu + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon'_{t-1} + \dots + \theta_q \varepsilon'_{t-q} + \varepsilon_t$$

where y'_t is the differenced data.

An ARIMA model has 3 parameters and can fully be expressed as $ARIMA(p, d, q)$.

- p is order of autoregressive component
- d is the order of differencing to be performed on the data
- q is the order of moving averages component

This generalised model is great since we don't have to difference the data ourselves and allows for the combination of $AR(p)$ and $MA(q)$ models. The implementation of this is quite simple as you just have to add the results of the AR and MA components 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.statology.org/detrend-data/bibliography\[1\]](https://www.statology.org/detrend-data/bibliography[1])

References

- [1] ABC News. “Fact check: Do more people die in Australia than Sweden due to poorly heated homes?” In: *JAMA Dermatol* (2018). DOI: <https://www.abc.net.au/news/2017-08-09/fact-check-australia-sweden-winter-fatalities-heating/8780588>. (accessed: 29.09.2021).