Time Series Analysis

An Introduction and Overview

Selin Karabulut[1] Shuying Yu[2] Sophia Arabadjis[3]

[1]Political Science
UCSB

[2]Psychological and Brain Sciences
UCSB

[3]Geography
UCSB

Time Series Consulting, April 2021

Table of Contents

Table of Contents

Time Series Data

A time series is a set of data measured or observed at different points in time [4] [5]. In this way, data have an implicit *order* or flow, and questions about time series data often involve questions about change over time. Examples of time series data come from both physical sciences and social sciences:

- Quarterly earnings of a company over time
- Yearly global average temperature
- Monthly birth counts over a decade
- Heart rate, oxygen and CO_2 levels during a VO_2 test
- Particulate matter concentration and weekly cardiovascular deaths

Key Concepts of Time Series

Time series can be expressed as univariate or multivariate models. We will introduce examples in the exploratory data analysis (EDA) section.

Univariate time series are those that have only one variable varying over time.

Multivariate time series are those in which multiple variables are varying simultaneously over time.

Types of Time Series

In a **multiplicative** time series, the components multiply together to make the time series. If you have an increasing trend, the amplitude of seasonal activity increases. Everything becomes more exaggerated. This is common when you're looking at web traffic.

In an **additive** time series, the components add together to make the time series. If you have an increasing trend, you still see roughly the same size peaks and troughs throughout the time series. This is often seen in indexed time series where the absolute value is growing but changes stay relative. [6]

Plots of Time Series

Examples of time series plots

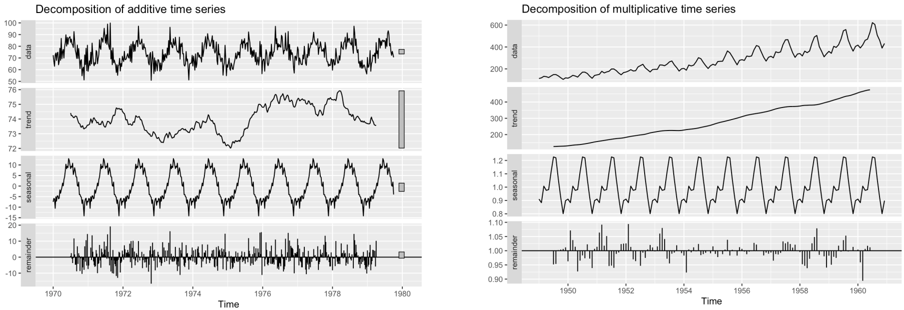

Figure: An additive (left) and multiplicative (right) decomposition of time series data.

Concepts in Time Series

There are three components to a time series [6]:

- Trend: how things are overall changing (linear)
- Seasonality: how things change within a given period
- Error: also known as residuals or irregular activity; not explained by the trend or seasonality values

How these three components interact (and particularly their correlation) determines the difference between a multiplicative and an additive time series [3]:

$$\text{Multiplicative}: Y_t = trend_t * season_t * \epsilon_t \tag{1}$$

$$\text{Additive}: Y_t = trend_t + season_t + \epsilon_t \tag{2}$$

Strict Stationarity: A time series is said to be strictly stationary if the following condition is true [3] [2]:

The joint probability density associated with n random variable's $X_{t1}, ..., X_{tn}$ for any set of times $t_1, t_2, ...t_n$ is the same as that associated with the n random variable's $X_{t_1+k}, ..., X_{t_n+k}$. This means it is devoid of

trend or seasonal patterns, which makes it looks like a random white noise irrespective of the observed time interval.
In reality, this rarely happens.

Weak Stationarity

A time series is said to be weakly stationary if the following conditions are true:

- The mean value of time-series is constant over time, which implies, the trend component is nullified i.e. $E(X_t) = \mu \ \forall \ t \in T$
- The covariances exist i.e. $E|X_t|^2 < \infty$
- The covariance of X_t with any X_s is equivalent to the covariance of X_{t+r} with X_{s+r}, in other words the covariance should be constant across shifting.
  $Cov(X_t, X_s) = Cov(X_{t+r}, X_{s+r}) = Cov_X(t-s) \ \forall \ t, s, r \in T$
- For Guassian Time Series, weak stionarity = strict stationarity because Gaussian distribution is completely determined by its mean and covariance.

Non-stationarity

Any periodicity means the time series is non-stationary. Any increasing variance indicates the time series is non-stationary.

However, is possible to observe stationary effects in a weakly stationary model, which can be done by removing the trend through differencing or detrending the time series.

Detrending occurs when the trend is subtracted from the time series.

Differencing occurs when an observation at time t is subtracted from the observation at time $t-1$. Differencing is used when the trend is fixed and when the goal is to coerce the data to stationarity.

Practicality of Non-Stationarity

"[I]t is the local behavior of a process, and not the global behavior of a process, that is of concern" [4]

While there are methods to remove the trend in a time series data, it may not be practical as studying trends or patterns over time is the goal of time series analysis. Often, the piece of interest is the trend.

However, we use still stationary tools as a building block to help us understand non stationary patterns.

Analysis Approaches for Time Series

Frequency Domain One approach to studying time series is to focus on periodic variation or systematic sinusoidal variations from physical, biological or environmental phenomena. (In the social sciences, we would consider monthly unemployment.) Approaches in this vein use spectral analysis in which frequency is expressed as the number of cycles per unit of time [5]. (We won't talk about this today.)

Time Domain Another (more common) approach to studying time series focuses on "modeling some future value of time series as a parametric function of the current and past values" (p.2 [4]). This is the approach we will focus on today.

It is interesting to note that these approaches have converged in recent years; as Shumway and Stoffer write, "In our view, no schism divides time domain and frequency domain methodology," but suggest that the time domain performs better over shorter series [4]

Table of Contents

Univariate: Case Data

We will use the air passenger (The AirPassenger dataset in R) as described in Shumway and Stoffer (2006) [4].

Visualizing the Data: Periodicity and Trend

The AirPassenger dataset in R provides monthly totals of a US airline passengers, from 1949 to 1960. This dataset is already of a time series class therefore no further class or date manipulation is required.

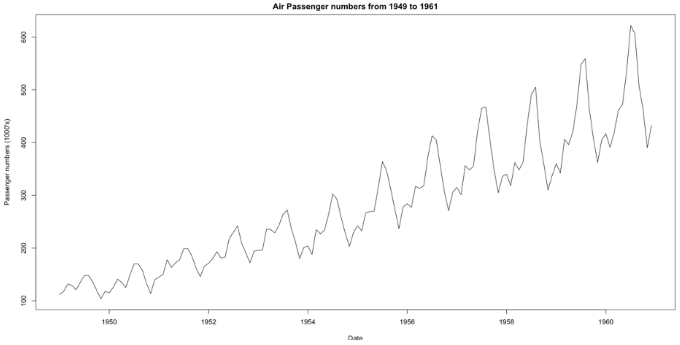

Figure: Plot the raw data using the base plot function

Visualizing the Data: Assessing Seasonality

Next, we might look at seasonal variation in the Air Passenger's data. We can group by month, and plot the variation.

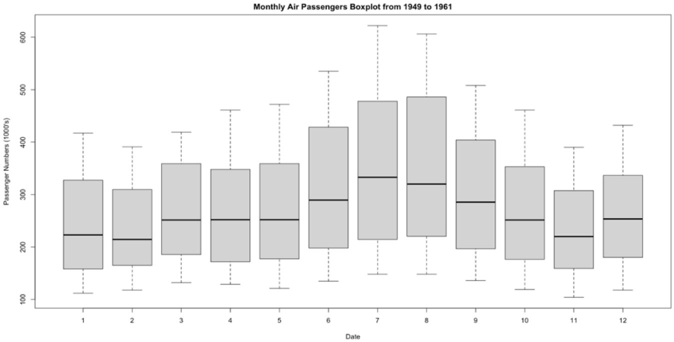

Figure: The boxplot function to see any seasonal effects

Visualizing the Data: Decomposition

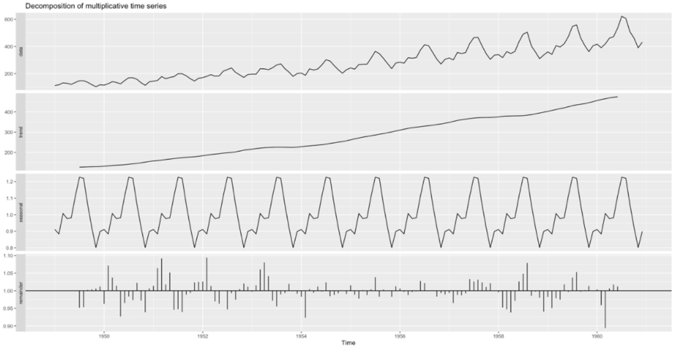

Figure: Time series decomposition

Summary Statistic: ADF Test for Stationarity

2 ways to test the stationarity of the series

- Augmented Dickey-Fuller Test

```
```{r}
adf.test(AP, k=12)
```
```

```
	Augmented Dickey-Fuller Test

data:  AP
Dickey-Fuller = -1.5094, Lag order = 12, p-value = 0.7807
alternative hypothesis: stationary
```

Figure: the Augmented Dickey-Fuller Test using the adf.test function from the tseries R package

Autocorrelation Plot for Stationarity

2 ways to test the stationarity of the series

- Autocorrelation

```{r}
autoplot(acf(AP,plot=FALSE))+ labs(title="Correlogram of Air Passengers from 1949 to 1961")
```

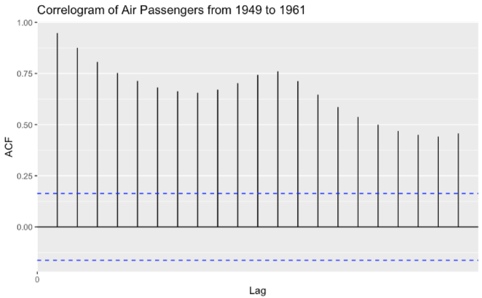

Figure: autocorrelation function (acf) in from the base stats R package

Multivariate Case: Data

Transition to Markdown!

Table of Contents

Strengths

"Time series are analysed to understand the past and to predict the future, enabling managers or policy makers to make properly informed decisions. A time series analysis quantifies the main features in data and the random variation. These reasons, combined with improved computing power, have made time series methods widely applicable in government, industry, and commerce" [1].

Strengths

- Forecasting data
  A major benefit of time series analysis is that it can be the basis to forecast data. This is because time series analysis — by its very nature — uncovers patterns in data, which can then be used to predict future data points.
  For example, autocorrelation patterns and seasonality measures can be used to predict when a certain data point can be expected. Further, stationarity measures can be used to estimate what the value of that data point will be.
- Discovering the mechanism by which the data are generated

Limitations

- Forecasting data
  The central logical problem in forecasting is that the "cases" (that is, the time periods) which you use to make predictions never form a random sample from the same population as the time periods about which you make the predictions
- Analyzing the Impact of Single Events
  When you try to assess the impact of a single event, the major problem is that there are always many events occurring at any one time.
- Causality
  it is rarely reasonable to assume that the time sequence of the causal patterns matches the time periods in the study.

References

[1] Paul S.P. Cowpertwait and Andrew V. Metcalfe. *Time Series Analysis and Its Applications: With R Examples*. Springer New York, New York, NY, 2009.

[2] Wendy Meiring. *Time Series: Data Collected Over Time: Concepts Based on PSTAT 174/274*, 2020.

[3] Selva Prabhakaran. Time series analysis, 2017.

[4] Robert Shumway and David Stoffer. *Time Series Analysis and Its Applications: With R Examples*, volume 03 of *Springer Texts in Statistics*. Springer New York, New York, NY, 2006.

[5] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Springer International Publishing, Cham, 2017.

[6] Steph. Is my time series additive or multiplicative? | R-bloggers, 2017.