4/18/2023

Decomposition of Time Series

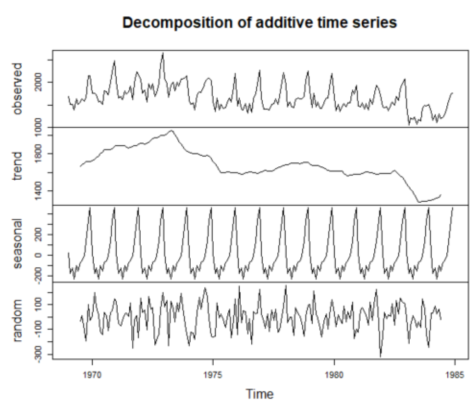
Seasonality in time series refers to patterns that repeat at fixed intervals of time, typically within a year or a week. For example, sales of winter clothes tend to increase in the winter months and decrease in the summer months, while sales of summer clothes tend to increase in the summer months and decrease in the winter months.

Seasonality can be identified by analyzing the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the time series data. The ACF and PACF can help identify the frequency and strength of the seasonal pattern. Seasonality can also be visualized by plotting the time series data over time, highlighting the patterns that repeat at fixed intervals.

To model seasonality in time series data, seasonal components can be added to the model. One common approach is to use seasonal autoregressive integrated moving average (SARIMA) models, which are a variation of the ARIMA model that include seasonal components. Another approach is to use seasonal decomposition of time series (STL) methods, which decompose the time series into trend, seasonal, and residual components.

Removing seasonality from time series data can be useful for forecasting purposes, as it can help eliminate the effects of the seasonal pattern and make it easier to identify other underlying patterns in the data. This can be done by detrending the data or by differencing the data by the length of the seasonal period. However, in some cases, seasonality may be an important factor in the time series data and should not be removed.

Decomposing a time series involves separating it into different components, typically a trend component, a seasonal component, and a residual component. This can be useful for understanding the underlying patterns in the time series data and for making forecasts.



There are several methods for decomposing a time series, including:

Seasonal Decomposition of Time Series (STL): This method decomposes a time series into its trend, seasonal, and residual components using a moving window approach. The method is useful for identifying both additive and multiplicative seasonality.

Classical Decomposition: This method decomposes a time series into its trend, seasonal, and residual components using a simple average approach. The method assumes that the seasonal component is fixed and known in advance.

Moving Average (MA) Decomposition: This method decomposes a time series into its trend and residual components using a moving average approach. The method assumes that the trend component is a moving average process.

Exponential Smoothing Decomposition: This method decomposes a time series into its trend, seasonal, and residual components using an exponential smoothing approach. The method is useful for identifying both additive and multiplicative seasonality.

After decomposing a time series, the individual components can be analyzed separately to identify patterns and relationships. The trend component can be used to identify long-term changes in the time series, while the seasonal component can be used to identify periodic patterns. The residual component represents the unexplained variation in the time series.

Once the time series has been decomposed, the individual components can be modeled separately using appropriate methods, such as regression models or time series models, to make forecasts.

Regression can be used to find a trend in time series data. The basic idea is to fit a regression line to the time series data, where the independent variable is time and the dependent variable is the time series variable of interest. The slope of the regression line represents the trend (in the linear model case), and the intercept represents the level of the time series.

There are different types of regression models that can be used to model trends in time series data, depending on the specific characteristics of the data. One common approach is to use simple linear regression, where a straight line is fitted to the data. However, if the data shows nonlinear trends, polynomial regression or other nonlinear regression models may be more appropriate.

It is important to note that regression models are not always the best choice for modeling trends in time series data, as they assume that the residuals (the differences between the observed values and the predicted values) are independent and identically distributed (IID), which may not be true for time series data. Therefore, more sophisticated time series models such as ARIMA, exponential smoothing, or state space models may be better suited for modeling trends in time series data.

Seasonality in time series data can be modeled using different methods, depending on the specific characteristics of the data. Here are some common approaches:

Seasonal decomposition: This method involves decomposing the time series into its seasonal, trend, and residual components. Once the seasonal component is isolated, it can be modeled separately using appropriate methods, such as a seasonal ARIMA model or a seasonal regression model.

Seasonal ARIMA model: This method involves using an ARIMA model with a seasonal component to model the time series. A seasonal ARIMA model includes additional terms to account for seasonal patterns, such as seasonal differences, seasonal autoregressive terms, and seasonal moving average terms.

Exponential smoothing: This method involves using an exponential smoothing model with a seasonal component to model the time series. Exponential smoothing models use a weighted average of past observations to make forecasts and can be adapted to include seasonal patterns.

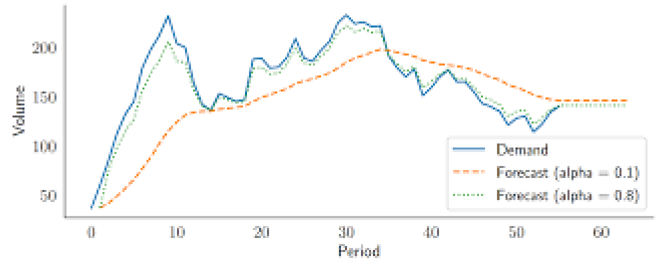


Figure 3.2: Simple smoothing

Seasonal regression model: This method involves using a regression model to model the time series, where the independent variables include time-related variables (e.g., month of the year, day of the week, etc.) to account for seasonal patterns.

Fourier analysis: This method involves using a Fourier transform to identify and model the seasonal component of the time series. The Fourier transform converts the time series into a series of sine and cosine waves, which can be used to estimate the seasonal component.

It is important to note that modeling seasonality in time series data can be challenging, especially when the seasonal patterns are complex or changing over time. In these cases, more sophisticated models, such as state space models or neural network models, may be required.

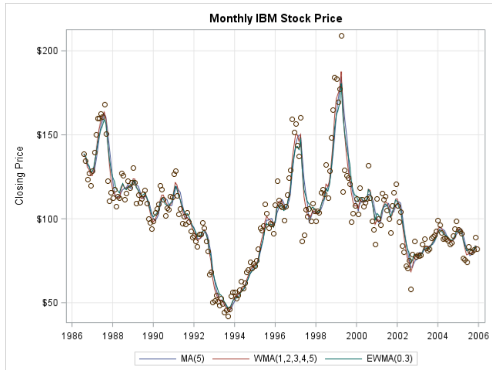
We can transform variables in a time series model to improve its fit or to satisfy certain assumptions of the modeling approach. Here are some examples of variable transformations that can be applied to time series data:

Log transformation: This transformation involves taking the natural logarithm of the time series variable. Log transformations can be useful when the time series exhibits exponential growth or decay, as they can make the data more linear and easier to model.

Box-Cox transformation: This transformation involves raising the time series variable to a power that depends on a parameter lambda. The Box-Cox transformation can be used to transform the data to have a more normal distribution, which can be helpful when using certain modeling approaches, such as linear regression.

Seasonal differencing: This transformation involves taking the difference between the time series variable and the value of the variable from the same season in the previous year. Seasonal differencing can be useful when the time series exhibits a seasonal pattern that is not well modeled using other methods.

Moving averages: This transformation involves taking the average of a rolling window of the time series variable. Moving averages can be useful for smoothing out short-term fluctuations in the data, which can make it easier to identify longer-term trends and patterns.



It is important to note that variable transformations should be chosen carefully, as they can change the interpretation of the data and the modeling approach. Moreover, the transformation may introduce new complexities that require additional modeling steps. It is recommended to explore different transformation options and to compare the results of the different models before selecting a final model.

Choosing a time series model can be challenging, as it depends on a variety of factors related to the data, the modeling objectives, and the available resources. Here are some important factors to consider when choosing a time series model:

Stationarity: The stationarity of the time series is an important factor to consider, as many time series models assume stationarity. If the time series is non-stationary, then it may need to be transformed or differenced to achieve stationarity, or a non-stationary model may be required.

Seasonality: The presence of seasonality in the time series is an important factor to consider, as it can affect the choice of model and the estimation approach. Some models, such as seasonal ARIMA and seasonal regression, are specifically designed to capture seasonality, while other models may require additional transformations or differencing to account for seasonal patterns.

Trend: The presence of a trend in the time series is another important factor to consider, as it can affect the choice of model and the estimation approach. Some models, such as exponential smoothing and regression, are specifically designed to capture trend, while other models may require additional transformations or differencing to account for trends.

Data frequency: The frequency of the time series data (e.g., hourly, daily, monthly, quarterly, yearly, etc.) is an important factor to consider, as it can affect the choice of model and the estimation approach. Some models, such as ARIMA, are designed for low-frequency data, while other models, such as exponential smoothing, are designed for high-frequency data.

Modeling objectives: The modeling objectives, such as forecasting, anomaly detection, or causal analysis, are important factors to consider, as they can affect the choice of model and the evaluation criteria. Some models, such as regression and causal models, are designed for causal analysis, while other models, such as exponential smoothing and ARIMA, are designed for forecasting.

Available resources: The available resources, such as computational power and data availability, are important factors to consider, as they can affect the choice of model and the estimation approach.

Some models, such as state space models and neural network models, require significant computational resources, while other models, such as regression and exponential smoothing, can be implemented with less computational power.

Overall, the choice of a time series model depends on the specific characteristics of the data and the modeling objectives. It is recommended to explore multiple modeling approaches and to evaluate the performance of the models using appropriate evaluation metrics before selecting a final model.

Resources:

1. https://towardsdatascience.com/seasonality-of-time-series-5b45b4809acd
2. https://machinelearningmastery.com/time-series-seasonality-with-python/
3. https://otexts.com/fpp2/tspatterns.html
4. https://otexts.com/fpp2/regression.html
5. https://www.sas.upenn.edu/~fdiebold/Teaching104/Ch14_slides.pdf
6. https://bookdown.org/mpfoley1973/time-series/time-series-regression.html
7. https://machinelearningmastery.com/machine-learning-data-transforms-for-time-series-forecasting/
8. https://web.stat.tamu.edu/~jnewton/stat626/topics/topics/topic5.pdf