Berkeley Engineering | BerkeleyHaas

PROFESSIONAL CERTIFICATE IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Office Hour #10 with Matilde D'Amelio

May 20, 2022 at 9 pm UTC

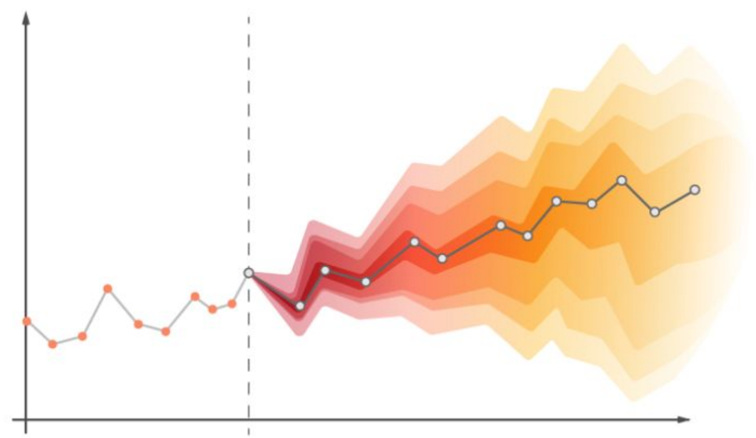
Capstone

Berkeley Engineering | BerkeleyHaas

Time Series & Forecasting

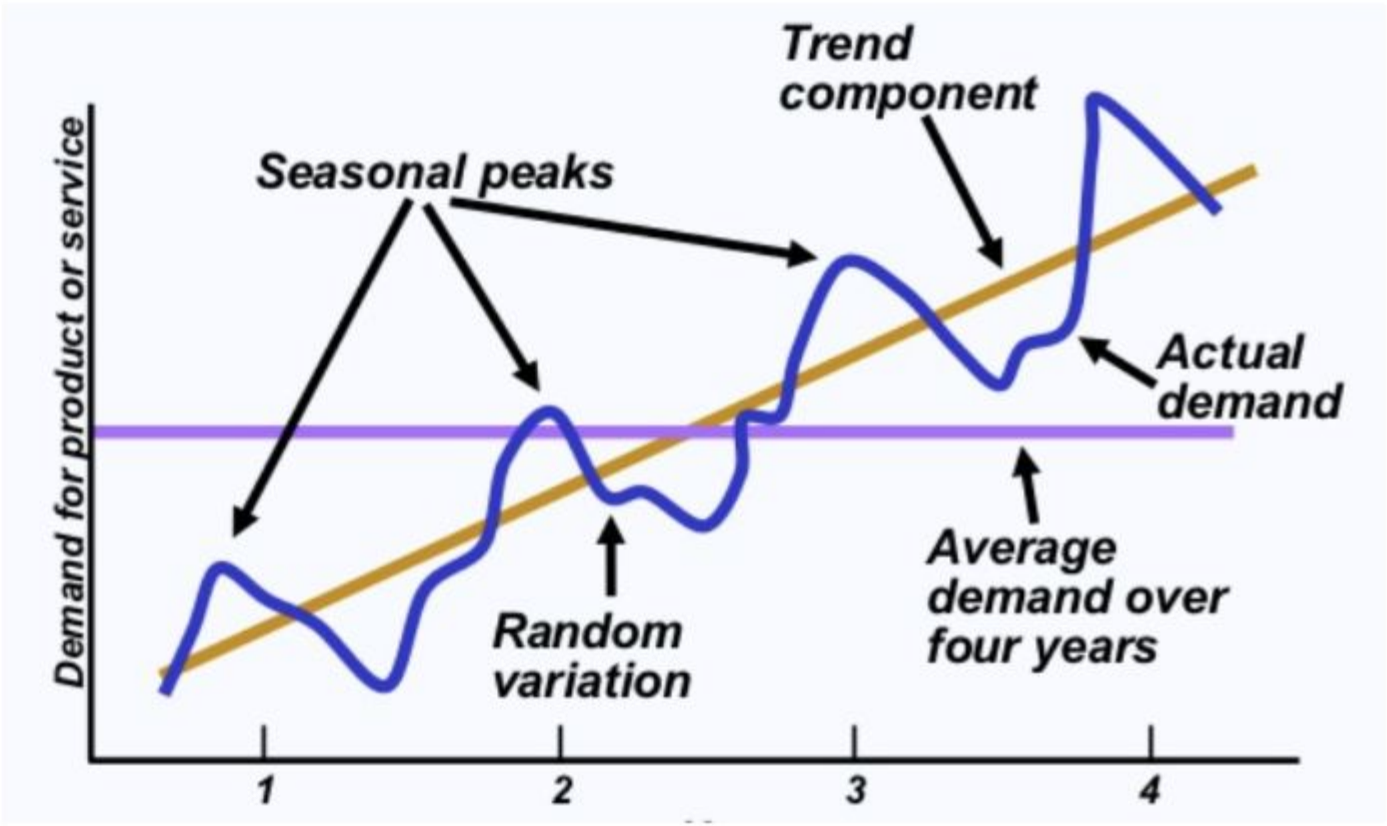
Time series forecasting occurs when you make **scientific predictions based on historical time stamped data**. It involves building models through historical analysis and using them to make observations and drive future strategic decision-making.

Time series analysis involves developing models to gain an understanding of the data to understand the underlying causes. Analysis can provide the **"why"** behind the outcomes you are seeing. Forecasting then takes the next step of what to do with that knowledge and the predictable extrapolations of what might happen in the future.



Data teams should use time series forecasting when they **understand the business question and have the appropriate data and forecasting capabilities** to answer that question. Good forecasting works with clean, time stamped data and can identify the genuine trends and patterns in historical data. Analysts can tell the difference between random fluctuations or outliers, and can separate genuine insights from seasonal variations.

Time Series & Forecasting



Tips for Good Forecasting

1. **Granularity**: More aggregated your forecast is more accurate you will be. Its simply because aggregated data has less variance and thus less noise. For example: If we want to predict airline passengers for next month, then forecasting the total number of travellers next month will be more accurate than forecasting travellers on a specific route. Again this is all derived by business requirement.
2. **Frequency**: How frequent you want to update your forecasts to keep them relevant. As time passes, we add more information (maybe new information) which needs to be incorporated to keep predictions relevant. Let's say we want to forecast the number of TV views and frequency of updating forecast is 3 months. Due to COVID-19, people are locked in their houses for 2–3 months and has increased TV views significantly during this time. We might miss this opportunity because the frequency of forecast update is more than event duration.
3. **Horizon**: Forecasts in earlier time frame are more accurate than far future. Say we are forecasting for the next 6 months sales, it will be more accurate in the first few months as compared to later months in future

Autocorrelation

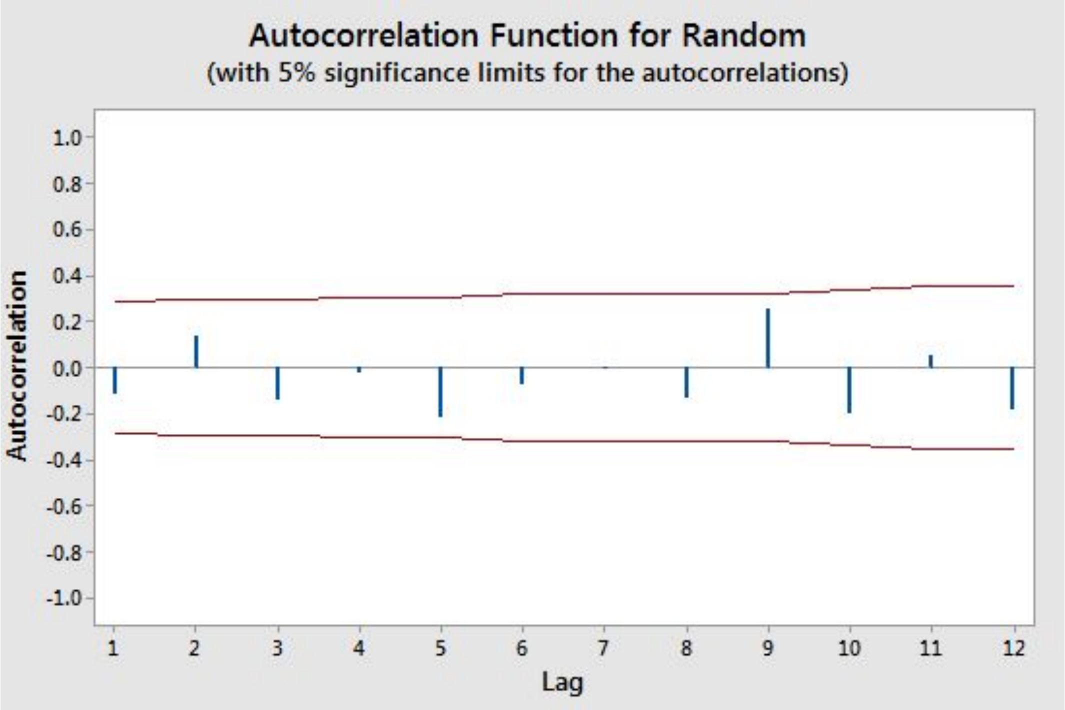
Autocorrelation represents the degree of similarity between a given time series and a lagged version of itself over successive time intervals.

Autocorrelation measures the relationship between a variable's current value and its past values.

A study about autocorrelation, helps to understand trend, seasonality, randomness and stationarity in your time series

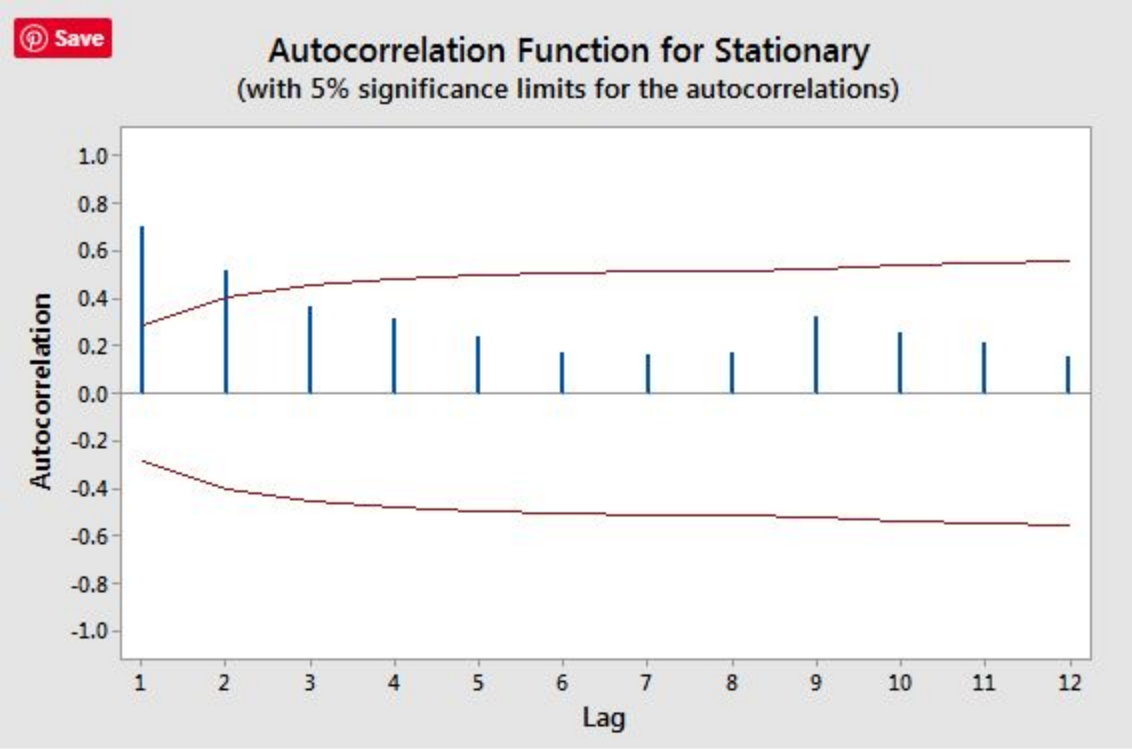
Randomness (White Noice)

For random data, autocorrelations should be near zero for all lags. Analysts also refer to this condition as white noise. Non-random data have at least one significant lag. When the data are not random, it's a good indication that you need to use a time series analysis or incorporate lags into a regression analysis to model the data appropriately.



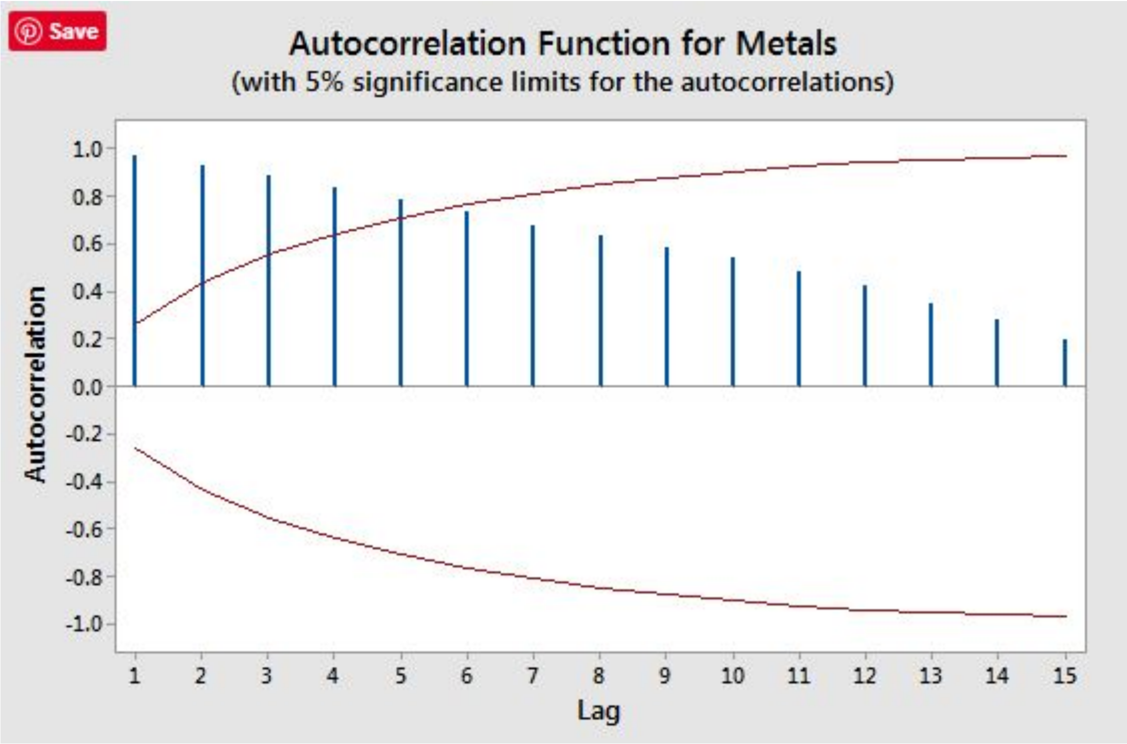
Stationarity

Stationarity means that the time series does not have a trend, has a constant variance, a constant autocorrelation pattern, and no seasonal pattern. The autocorrelation function declines to near zero rapidly for a stationary time series.



Trends

When trends are present in a time series, shorter lags typically have large positive correlations because observations closer in time tend to have similar values. The correlations taper off slowly as the lags increase.

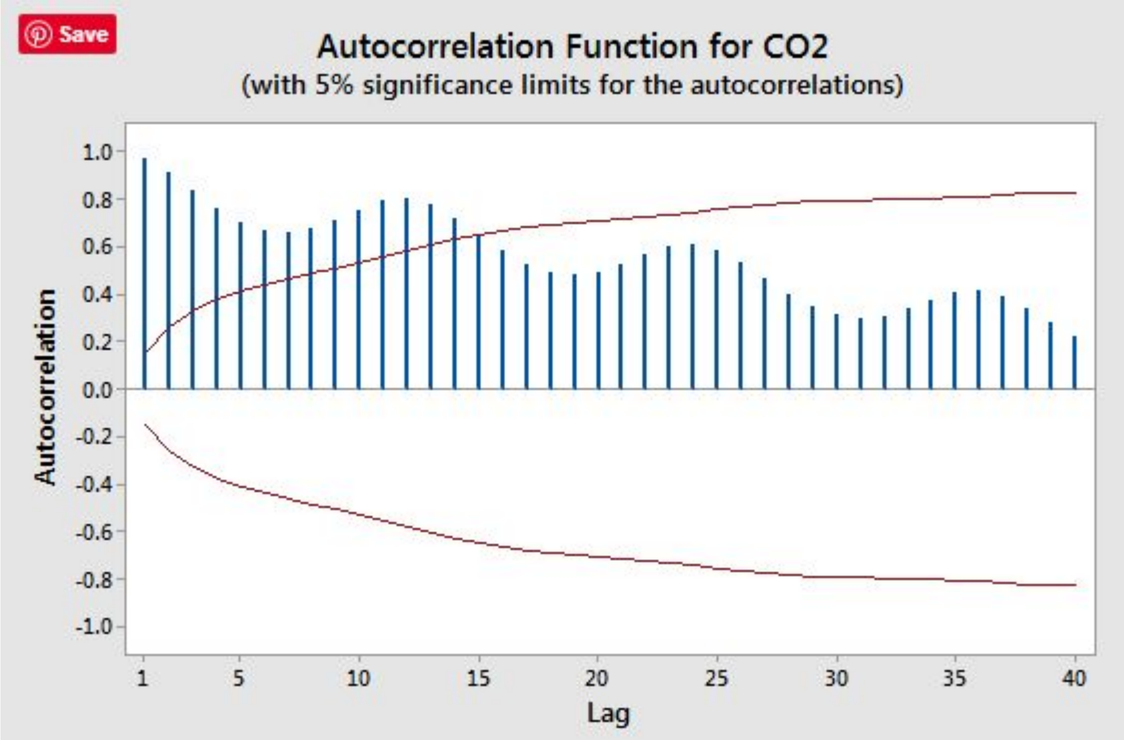


Berkeley Engineering | BerkeleyHaas

Seasonality

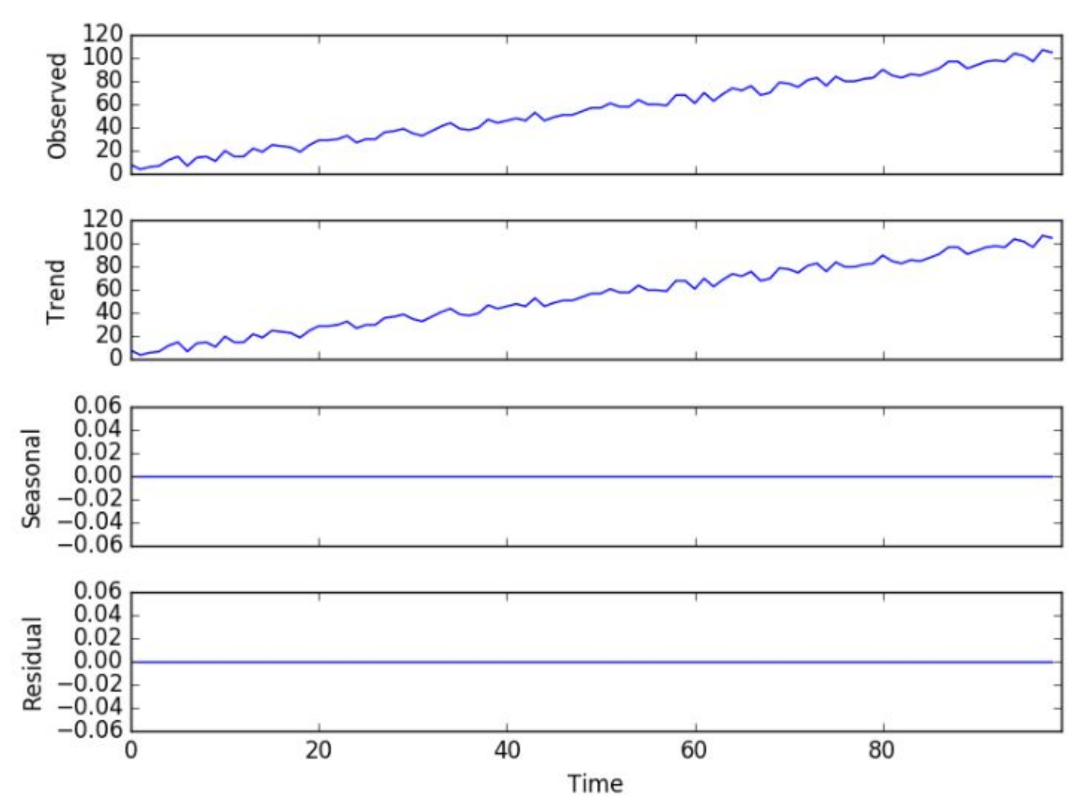
When seasonal patterns are present, the autocorrelations are larger for lags at multiples of the seasonal frequency than for other lags.

You can have both seasonality and trend



Decomposition

Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise (or residual) components



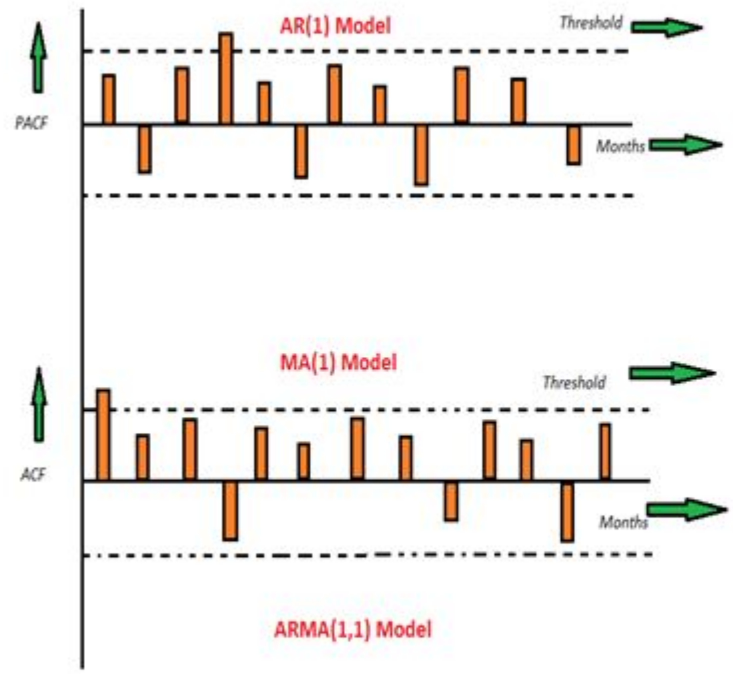
Additive Model Decomposition Plot

Berkeley Engineering | Berkeley Haas

ARMA MODEL

Auto Regressive Moving Average

- **Auto Regressive (AR) Model:** The time period at t is impacted by the observation at various slots t-1, t-2, t-3, ….., t-k. The impact of previous time spots is decided by the coefficient factor at that particular period of time. The price of a share of any particular company X may depend on all the previous share prices in the time series.
- **Moving Average (MA) Model:** The time period at t is impacted by the unexpected external factors at various slots t-1, t-2, t-3, ….., t-k. These unexpected impacts are known as Errors or Residuals. The impact of previous time spots is decided by the coefficient factor α at that particular period of time. The price of a share of any particular company X may depend on some company merger that happened overnight or maybe the company resulted in shutdown due to bankruptcy.
- **Auto Regressive Moving Average (ARMA) Model:** This is a model that is combined from the AR and MA models. In this model, the impact of previous lags along with the residuals is considered for forecasting the future values of the time series.



Berkeley Engineering | BerkeleyHaas

Group Discussion

What could be potential applications of Time Series Analysis?

QUESTIONS?