

Time series regression

Introduction to time series analysis

Yujin Kim, 11-30-2021

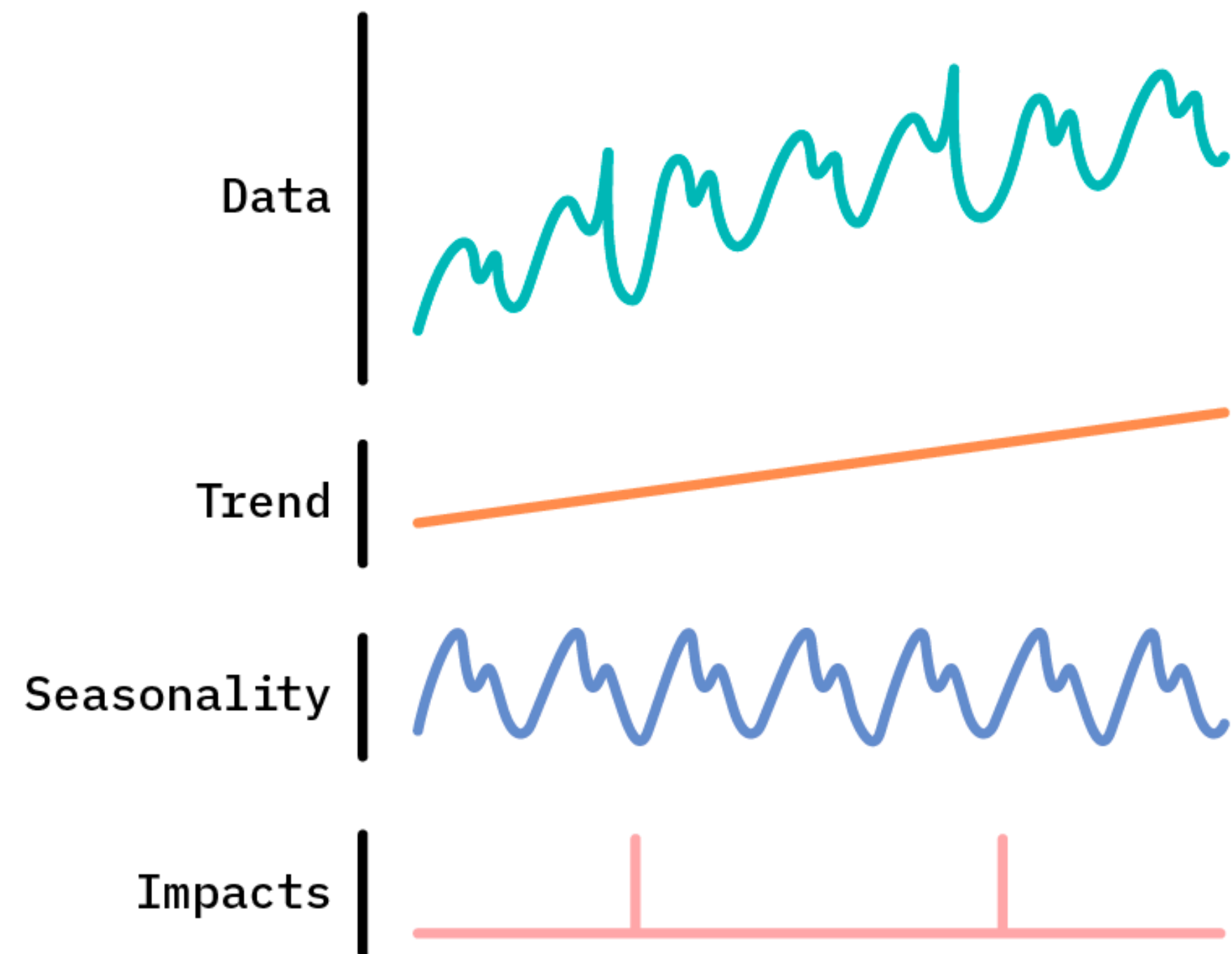
What is time series

- ***Time series analysis*** comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data
 - e.g., stock market data, # of tweets posted everyday about a topic
 - cf. Cross-sectional studies; Data points taken at a single point in time. no natural ordering of the observations. e.g., tweets posted about a topic in a single day
 - A core assumption : (stationarity) each observation in your dataset is independent from one another (i.e. the mean, variance and autocorrelation structure do not change over time)
- Unit of analysis: date (time)
 - lower aggregation: seconds, minutes
 - higher aggregation: weekly, monthly, yearly

Why not just use OLS regression?

- Nay?
 - TS is **NOT** randomly sampled in the same way as cross-sectional; each obs are not independent and identically distributed random variables
 - Leading to many problems: auto-correlation (errors correlated over time), trending (seemingly related to one another; but “spurious”)
- Yay?
 - Some models, such as VAR (Vector AutoRegression), with ordinary least squares is a commonplace, perfectly acceptable practice in finance and economics.
 - If typical VAR assumptions are met (i.e., each equation has the same regressors, the errors are mean independent of the lagged variables), OLS is efficient.

What makes time series analysis different?

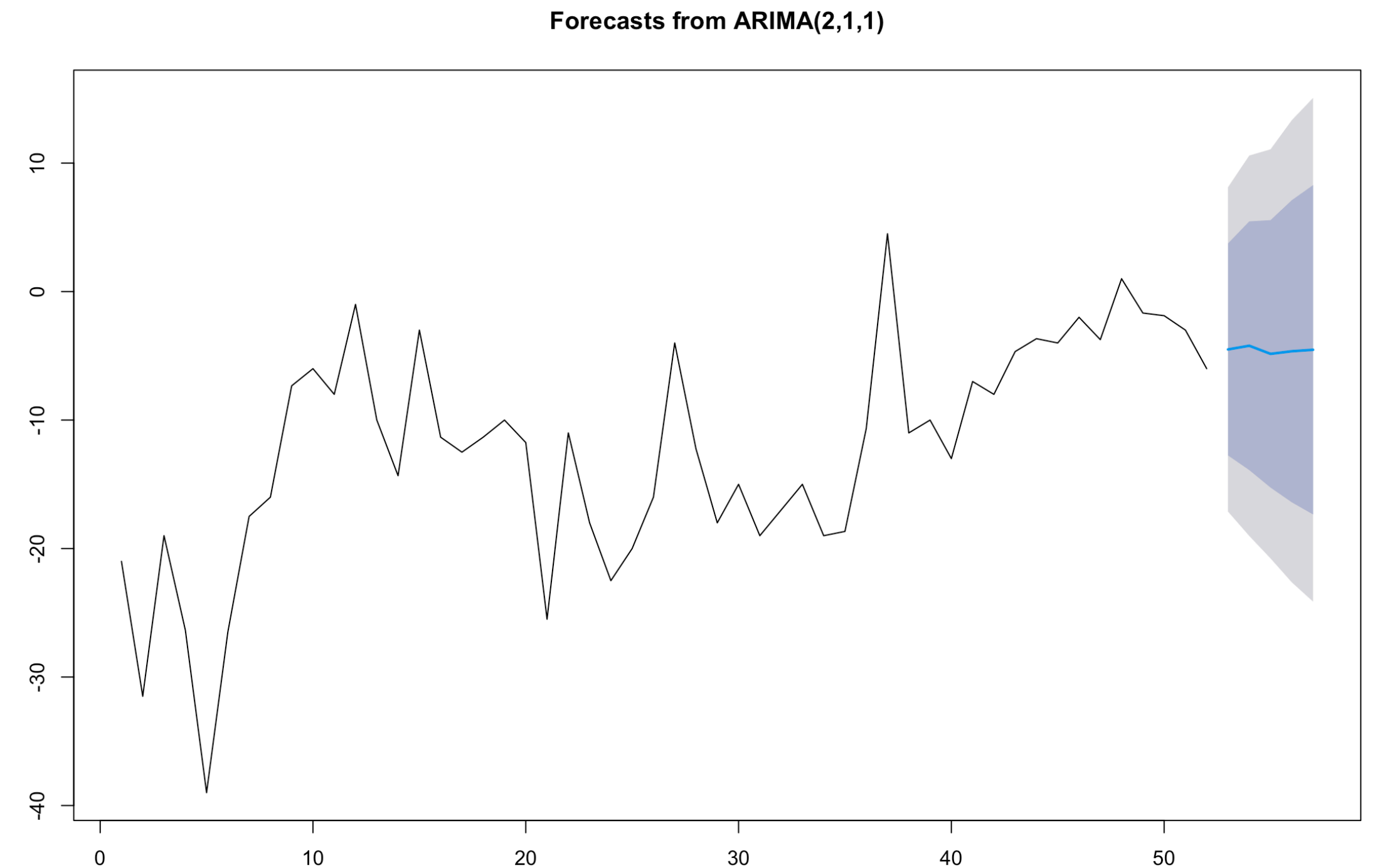


- Time series analysis prompts us to:
 1. Formally disaggregate and identify those temporal properties
 2. Model their functional form and filter them out of your series
 3. Inspect the model residuals and tweak specification as necessary

- The goal is to account for these temporal dependencies within and between the observations, allowing us to more closely model the true data-generating process (DGP)

Univariate time series analysis; ARIMA model

- Model specifications (p, d, q)
- AR: Autoregression. Uses the dependent relationship between an observation and some number of lagged observations (p)
- I: Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series **stationary** (d)
- MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations (q)
- Forecast



Reference

- <https://otexts.com/fpp2/arma.html>
- <https://machinelearningmastery.com/arma-for-time-series-forecasting-with-python/>