Time series analysis

Data to decisions

Definition

- A time series is a sequence $\{\ldots, X_t, X_{t+1}, X_{t+2}, X_{t+3}, \ldots\}$ of random variables
- For instance X may be GDP, sales, EBITDA, … while t denotes quarters or years
- The goal of time series analysis is to use data to find and exploit patterns in the series:
  1. Does the series have a trend?
  2. Does it display seasonality?
  3. Does it revert to some long-term mean?
  4. Does it display persistence?
  5. …

Classical decomposition

- It is useful to think of a time-series as consisting of three parts:

$$X_t = T_t + S_t + Y_t$$

where

1. T_t is the *trend* (a simple, slow-moving, predictable function of time)
2. S_t is a *seasonality component* (a component with known periodicity)
3. Y_t is a *stationary component* (a component with no independent time effect)

- Classical approach to time-series: 1) remove T and S, 2) model Y

Stationarity

- A series is stationary if for $k = 1,2, \ldots$ the distribution of Y_{t+k} given Y_t does not depend on t

- Deep theorems tell us that all such series can be well approximated by:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_k Y_{t-k}$$
$$+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_p \varepsilon_{t-p}$$

where the top-line is the *autoregressive* part and the bottom part is the *moving average* part

Removing the seasonal component

- Two steps:
  1. Look/test for calendar effect (Christmas, Monday, leap year…)
  2. Remove them

- In practice, this is grunt, mechanical work and statistical packages are there for you

- For an example, see the X12 package at the Census Bureau

Modeling stationary series

- Once we are reasonably confident that we are left with a stationary series (there are tests for that, though they are weak), "all" we need to do is choose the length of the AR part and the MA part

- This is a model selection issue so the standard tools apply:
  1. Information criteria
  2. Cross-validation

AR(1) processes

- The simplest model we can write for a persistent time series is:

$$Y_t = a + \rho Y_{t-1} + \varepsilon_t$$

where $0 < \rho < 1$ and $\varepsilon_t \sim Normal(0, \sigma^2)$ is white noise i.e. independent of everything

- If we can estimate a and ρ then an obvious and optimal way to forecast future $Y's$ given our last data point Y_t is:
  1. $\hat{Y}_{t+1} = a + \rho Y_t$
  2. $\hat{Y}_{t+2} = a + \rho \hat{Y}_{t+1} = a + \rho a + \rho^2 Y_t$
  3. $\ldots$
  4. $\hat{Y}_{t+k} = a + \rho a + \cdots \rho^k a + \rho^k Y_t$
  5. $\ldots$
  6. $\hat{Y}_\infty = \frac{a}{1-\rho}$

Confidence intervals for AR(1) forecasts

- If we can estimate σ then the standard error of our $t+k$ forecast is:

$$\hat{\sigma}_{t+k} = \sqrt{\frac{(1-\rho)^{2k}\,\sigma}{1-\rho^2}}$$

- A 95% confidence interval for our $t+k$ forecast is:

$$\left[\hat{Y}_{t+k} - 1.96\hat{\sigma}_{t+k}, \hat{Y}_{t+k} + 1.96\hat{\sigma}_{t+k}\right]$$

But enough chit-chat, it’s time to looks at some examples