Time Series Analysis

Basics of Time Series Analysis

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Decomposition: Trend Estimation

Georgia Tech

About This Lesson

Time Series: Basics

Data: Y_t, where t indexes time, e.g. minute, hour, day, month

Model: Y_t = m_t+s_t+X_t

- m_t is a trend component;
- s_t is a seasonality component with known periodicity d ($s_t = s_{t+d}$) such that $\sum_{j=1}^{d} s_j = 0$
- X_t is a stationary component, i.e. its probability distribution does not change when shifted in time

Approach: m_t and s_t are first estimated and subtracted from Y_t to have left the stationary process X_t to be model using time series modeling approaches.

Exploratory Analysis

```
# Load BTC data
databtc = read.csv('BTC-USD.csv',header = TRUE)
pricebtc = databtc[,c(5)]
mydates=as.Date(databtc[, 1], "%m/%d/%Y")
tsbtc=xts(pricebtc,mydates)
dlbtc=diff(log(tsbtc))[-c(1),]

# Display BTC data
plot(tsbtc,main='BTC-USD')
acf(tsbtc,main='ACF of BTC’)

# Display BTC log differenced data
plot(dlbtc,main='Diff_log_BTC')
acf(dlbtc[-1],main='ACF of Diff_log_BTC')
```

Georgia Tech

Trend: Moving Average

Estimate the trend for t with a width of the moving window *d*:

If the width is $d = 2q$, use

$$\widehat{m}_t = \frac{1}{d}\left[\frac{x_{t-q}}{2} + x_{t-q+1} + x_{t-q+2} + \ldots + x_{t+q-1} + \frac{x_{t+q}}{2}\right]$$

If the width is $d = 2q + 1$, use

$$\widehat{m}_t = \frac{1}{d}\sum_{j=-q}^{q} x_{t+j}$$

The width selection reflects the bias-variance trade-off:

- If width large, then the trend is smooth (i.e. low variability)
- If width small, then the trend is not smooth (i.e. low bias)

Trend: Parametric Regression

Estimate the trend m_t assuming a polynomial in t:

$$m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p$$

- Commonly use small order polynomial (p=1 or 2)
- Estimation approach: Fit a linear regression model where the predicting variables are $(t, t^2, \ldots, t^p)$
- Which terms to keep? Use model selection to select among the predicting variables. **Cautious!** Strong correlation among the predicting variables.

Trend: Non- Parametric Regression

Estimate the trend m_t with t in $\{t_1, t_2, \ldots, t_n\}$:

1. **Kernel Regression**

   $m_t = m(t) = \sum_{i=1}^{n} l_i(t) X_{t_i}$ where $l_i(t)$ a weight function depending on a kernel function.

2. **Local Polynomial Regression**
   - An extension of the kernel regression and the polynomial regression: fit a local polynomial within a width of a data point

3. **Other Approaches**
   - Splines regression
   - Wavelets
   - Orthogonal basis function decomposition

Georgia Tech

Trend: Non- Parametric Regression

Which one to choose?

- Local polynomial regression is preferred over kernel regression since it overcomes boundary problems and its performance is not dependent on the design of the time points
- Other methods are to be selected depending on the level of smoothness of the function to be estimated
- For estimating the trend in time series, local polynomial or splines regression will perform well in most cases

Summary