Modeling Time-Series Data

Collected by Prof. Seungchul Lee
iSystems Design Lab
UNIST
http://isystems.unist.ac.kr/

Table of Contents

1. (Determinstic) Sequences and Difference Equations

We will focus on linear difference equations (LDE), a surprisingly rich topic both theoretically and practivally.

For example,

$$y[0] = 1, \quad y[1] = \frac{1}{2}, \quad y[2] = \frac{1}{4}, \quad \cdots$$

or by closed-form expression,

$$y[n] = \left(\frac{1}{2}\right)^n, \quad n \geq 0$$

or with a difference equation and an initial condition,

$$y[n] = \frac{1}{2}y[n-1], \quad y[0] = 1$$

1.1. First order homogeneous LDE

Stanard form

$$y[n] = \alpha_1 y[n-1]$$

Solution

$$\begin{aligned}
y[n] &= \alpha_1 y[n-1] \\
&= \alpha_1(\alpha_1 y[n-2]) \\
&= \alpha_1(\alpha_1(\alpha_1 y[n-3])) \\
&\vdots \\
&= \alpha_1 \cdots (\alpha_1(\alpha_1 \cdots y[0])) \\
&= y[0]\alpha_1^n
\end{aligned}$$

- If $\alpha_1 < -1$
  - oscillate, the magnitude of their values grow without bound, and the first-order LDE is unstable,
- If $-1 < \alpha_1 < 0$
  - oscillate, their values decay to zero, and the first-order LDE is stable,
- If $\alpha_1 = -1$
  - oscillate, the magnitude of their values neither decay nor grow, and the first-order LDE is neither stable or unstable.

1.2. Second order homogeneous LDE

Standard form

$$y[n] = \alpha_1 y[n-1] + \alpha_2 y[n-2]$$

Assume LDE has solutions of the form

$$y[n] = c\lambda^n$$

Results in a quadratic equation

$$\lambda^2 - \alpha_1\lambda - \alpha_2 = 0,$$

And roots are

$$(\lambda_1, \lambda_2) = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

Finally solutions with two inintial conditions are

$$y[n] = c_1\lambda_1^n + c_2\lambda_2^n$$

Note that λ and c can be complex numbers.

1.3. High order homogeneous LDE

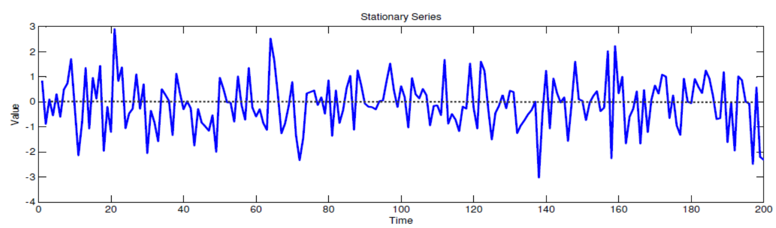
Standard form

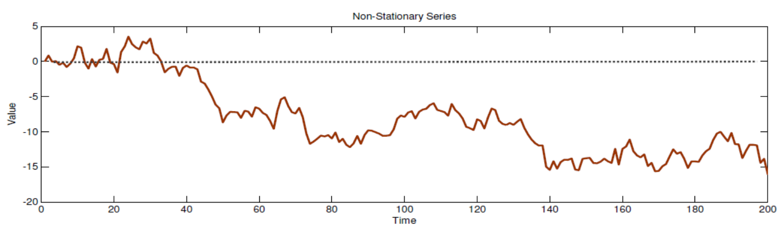
$$y[n] = \alpha_1 y[n-1] + \alpha_2 y[n-2] + \cdots + \alpha_k y[n-k]$$

2. (Stochastic) Time Series Analysis

2.1. Stationarity and Non-Stationary Series

- A series is *stationary* if there is no systematic change in mean and variance over time
  - Example: radio static
- A series is *non-stationary* if mean and variance change over time
  - Example: GDP, population, weather, etc.





2.2. Testing for Non-Stationarity

Formally

- Augmented DickeyFuller test

Informally

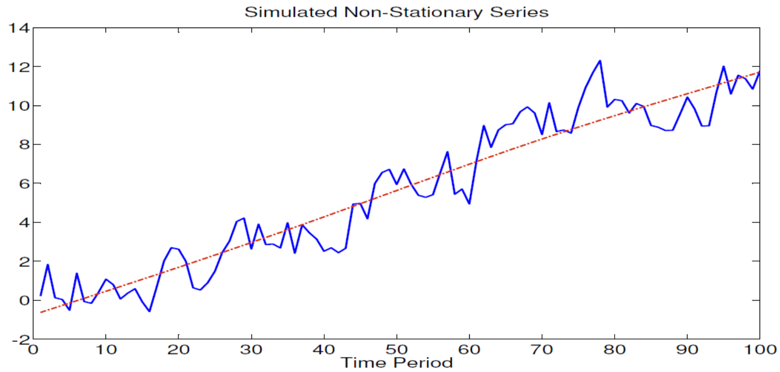
- Auto-Correlation Function (ACF)
- Normal Quantile Plot (Q-Q plot)

Q-Q Plot

- Compare distribution of the residuals to normal
- Scatter plot of residual quantiles against normal
  - Stationary data: quantiles match normal (45^o line)
  - Non-stationary data: quantiles do not match (points off 45^o line)

2.3. Dealing with Non-Stationarity

Linear trends



- One way of dealing with trend is to difference the series

$$\Delta Y_t = Y_t - Y_{t-1}$$

- Then, estimate the model using first differences (*i.e.*, AR(1) model) as

$$\Delta Y_t = \beta_1 + \beta_2 \Delta Y_{t-1} + u_t$$

- If first differences are non-stationary, use second difference

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$$
$$\Delta^2 Y_t = \beta_1 + \beta_2 \Delta^2 Y_{t-1} + u_t$$

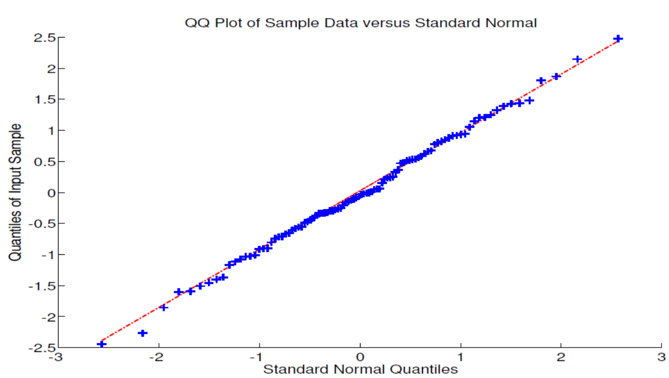
- Another way to deal with linear trend is to include a trend term
- Instead of estimating a plain AR(1) model

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t$$

- Include time t into the regression and estimate

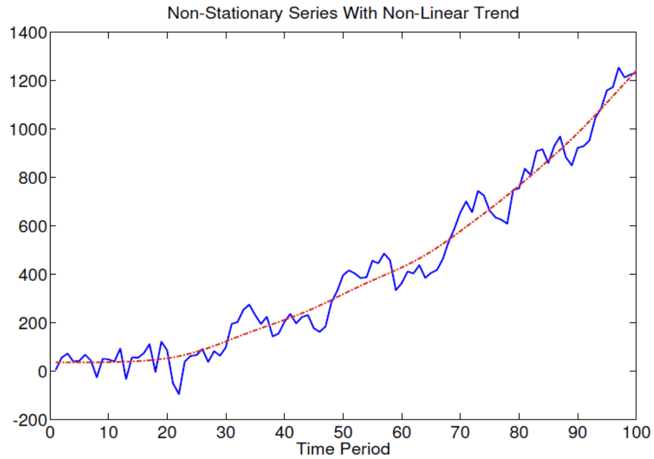
$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 t + u_t$$

- Q-Q residuals plot (stationary data)
  - normal residuals once trend is successfully netted $\leftrightarrow$ see if u_t is normally distributed



Non-linear trends

For example, population may grow exponentially



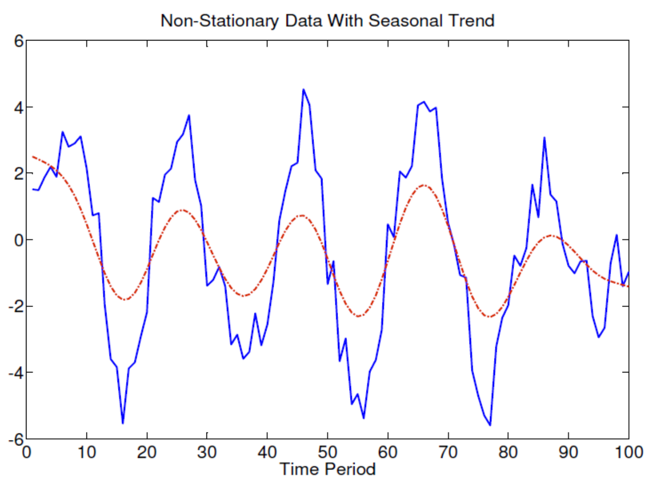
- Non-linear trends can be dealt with by differencing
- Alternatively, include an exponential time term

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 t^{\beta_4} + u_t$$

- Residual Q-Q plot can be used to check model fit

Seasonal trends

- Some series may exhibit seasonal trends
- For example, weather pattern, employment, inflation, etc.



- Including linear or quadratic trend may be insufficient
- Several approaches to accounting for seasonal trends
  - differencing
  - modeling cyclical trends

1. Seasonal differences
   - Suppose trend cycle is repeated with frequency s period
   - For example, for monthly data
     - annual cycles $s = 12$
     - quarterly cycles $s = 3$
   - Solution: work with seasonal differences $\Delta_t^s Y_t$

   $$\Delta_t^s Y_t = Y_t - Y_{t-s}$$

   - Examine the resiudal Q-Q plot to check model fit
   - Choice of s may be challenging (experimentation)

2. Seasonal trend models
   - As with linear or exponential trends, can explicitly include seasonal trend term into the model
   - A common approach is to include cyclical trend term based on sine wave

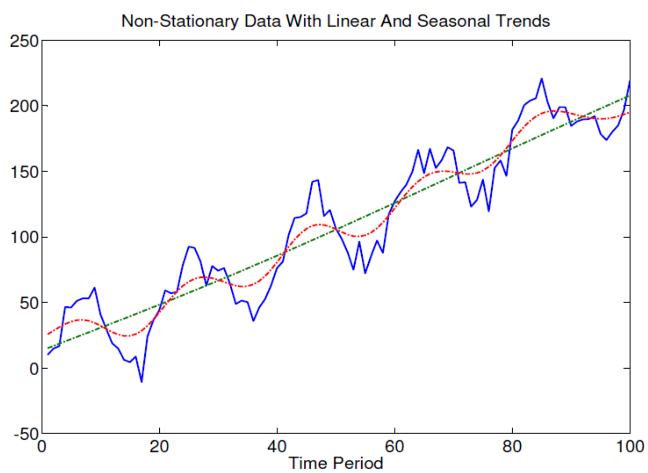
   $$\beta \sin(\omega t + \theta)$$

- Include cyclical trend term into the model by estimating
$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \left(\beta_3 \sin\frac{2\pi}{s}t + \beta_4 \cos\frac{2\pi}{s}t\right) + u_t$$

- Quarterly trend example (monthly data and want to include quarterly trend)
$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \left(\beta_3 \sin\frac{2\pi}{3}t + \beta_4 \cos\frac{2\pi}{3}t\right) + u_t$$
- Annual trend example (monthly data and want to include annual trend)
$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \left(\beta_3 \sin\frac{2\pi}{12}t + \beta_4 \cos\frac{2\pi}{12}t\right) + u_t$$
- Monthly trend example (daily data and want to include monthly trend)
$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \left(\beta_3 \sin\frac{2\pi}{30}t + \beta_4 \cos\frac{2\pi}{30}t\right) + u_t$$

Combining Linear, Quadratic, and Seasonal Trends

- Some data may have a combintation of trends



- One solution is to apply repeated differencing to the series
- For example, first remove seasonal trend. Then remove linear trend
- Inspect model fit by examining residuals Q-Q plot
- Anternatively, include both linear and cyclical trend terms into the model
$$\begin{aligned} Y_t &= \beta_1 + \beta_2 Y_{t-1} \\ &+ \beta_3 t + \beta_4 t^{\beta_5} \\ &+ \beta_6 \sin\frac{2\pi}{s}t + \beta_7 \cos\frac{2\pi}{s}t \\ &+ u_t \end{aligned}$$

2.4. Time-Series Forecasting

Suppose $\hat\beta_1, \hat\beta_2, \cdots, \hat\beta_7$ are estimated from an AR(1) model with linear, exponential, and cyclical trends

Then at some time T we can predict

- Linear trend as
$$\hat L_T = \hat\beta_3 T$$
- Exponential trend as
$$\hat E_T = \hat\beta_4 T^{\beta_5}$$
- Cyclical trend as
$$\hat C_T = \hat\beta_6 \sin\frac{2\pi}{s}T + \hat\beta_7 \cos\frac{2\pi}{s}T$$

Seasonal ajustments and "De-trending"

- Data are often available in seasonally adjusted and/or "de-trending"
- Objective is to remove all trends
- Approach is to estimate a model with trend components only
- For example, suppose data have exponential and cyclical trend components

- Estimate the trend-only model

$$Y_t = \alpha_1 t^{\alpha_2} + \alpha_3 \sin \frac{2\pi}{s} t + \alpha_4 \cos \frac{2\pi}{s} t + u_t$$

- Calcuate the trend estimates
  - Exponential trend component $\hat{E}_t = \alpha_1 t^{\alpha_2}$
  - Cyclical trend component $\hat{C}_t = \alpha_3 \sin \frac{2\pi}{s} t + \alpha_4 \cos \frac{2\pi}{s} t$
- De-trended data: $\bar{Y}_t = Y_t - \hat{E}_t$
- Seasonally-adjusted data: $\tilde{Y} = Y_t - \hat{C}_t$

Forecasting Series

- Given series value at time t, predict future value as

$$\hat{Y}_{t+1} = \hat{\beta}_1 + \hat{\beta}_2 Y_t + \hat{L}_{t+1} + \hat{E}_{t+1} + \hat{C}_{t+1}$$

In [1]:

```
%%html
<iframe src="https://www.youtube.com/embed/xBP4cQetoNM"
width="560" height="315" frameborder="0" allowfullscreen></iframe>
```

- Slides#1 (./files/Medovikov-TimeSeries2011.pdf), Slides#2 (./files/Medovikov-AppliedTimeSeries.pdf), from Prof. Ivan Medovikov (http://medovikov.me/teaching.html) at Brock University

In [2]:

```
%%javascript
$.getScript('https://kmahelona.github.io/ipython_notebook_goodies/ipython_notebook_toc.js')
```