Time Series Analysis

What is time series

A collection of stochastic random variables **indexed by time**, their stochastic **distributions are similar but not the same**, the data are **correlated across time**
In plain words, time series refers to **a series of data points**, for which they follow some kind of distribution.
The goal for studying time series, is really to investigate in those data points, and try to figure out its **underlying distribution, and internal correlation**

Important Characteristics

1. Trend
2. Seasonality
3. Periodicity
4. Cyclical Trend
5. Heteroskedasticity
6. Dependence

Presentation & Basic Decomposition

Y_t, in which t indexes time: $Y_t = m_t + s_t + x_t$ $, with:$
m_t — trend component, trend means that over the time scope, the values are going up or going down,
s_t — seasonal component, seasonality means that it is a periodicity movement, it has a circle of appearance, within the circle, the overall trend should be zero .
x_t — stationary component, stationary means still, it pictures the characteristic that the component is not changing its value

Estimate the trend

1. Moving Average:
   estimate the trend with a moving window
2. Parametric Regression
   fit in a polynomial regression to estimate
3. Non-Parametric Approach
   - Kernel Regression
   - Local Polynomial Regression
   - Other Approaches

```
Example: Temperature Data
1. Read the data file in R
2. Visualize the data (Avg Temp)
3. Estimate the trend (Moving Average)
4. Estimate the trend (Parametric)
5. Estimate the trend (Non-Parametric)
6. Comparison
```

Estimate the Seaonality

General Approach: estimate and subtract m_t and s_t

1. Seasonal Average
   $\hat{s}_k = w_k - \frac{1}{d}\sum_{j=1}^{d} w_j$

   k: the seasonal group, let's say it has d groups. If the seasonality is monthly, then d = 12
   w_k: the average of all seasonal group value
2. Parametric Regression

```
Example: Temperature Data
1. Read the data file in R
2. Visualize the data (Avg Temp)
3. Estimate the trend (Moving Average)
4. Estimate the trend (Parametric)
5. Estimate the trend (Non-Parametric)
6. Comparison
```

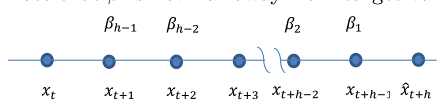
Estimate the Stationary

Important Concepts in Time Series

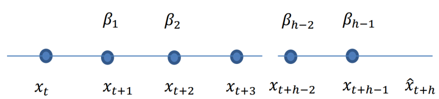
1. **Auto Correlation Function (ACF)**
Like correlation or covariance, the auto correlation between 2 variables can be expressed as:
$\gamma(s,t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$
$\gamma(t,t) = Var[X_t] = \sigma^2$
then we have the auto correlation defined as:
$\gamma_k = \gamma(t, t+k) \approx c_k \rightarrow \frac{\sum(x_t - \bar{x})(x_{t+k} - \bar{x})}{N}$
The underlying assumption is that the time series is stationary, therefore, across all k time horizon, this auto correlation relationship exist

2. **Auto Correlation Coefficient**
Following the auto correlation, the auto correlation coefficient can be defined as: $\rho_K = \frac{\gamma_k}{\gamma_0} \approx r_k \rightarrow \frac{c_k}{c_0}$

3. **Partial Auto Correlation Function (PACF)**
For a series of data, we can do things like:
We first start at a point and look backwards:
$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + ... + \beta_{h-1} x_{t+1}$
Note that β is how far away from target variables



Then we try looking forwards:
$x_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + ... + \beta_{h-1} x_{t+h-1}$
Note that β is how far away from target variables



Finally we remove the predict and see residual:
$Corr[(x_{t+h} - \hat{x}_{t+h}), (x_t - \hat{x}_t)]$
Therefore, this model measure the correlation with a lag, without accounting for the effect of variables in between.

Take away: Partial Out a variable, what PACF measures

4. **Random Walk**
$X_t = X_{t-1} + Z_t$
Z as the white noise, X as stock price
$X_t = \sum_{i=1}^{t} Z_i$
we can interpret X as accumulated white noise
$E[X_t] = \mu t\ ,\ Var[X_t] = \sigma^2 t$

5. **Moving Average Process: q**
How long back need to trace, to have weighted average:
The longer we trace back, distribution be smoother:
For instance, the MA(2) process is correlated back in 2 steps:
$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Zt - 2$
$Z_t \sim N(0,1)$
Moving Average is Weakly Stationary:
$cov(X_t, X_{t+k}) = \sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k}$

6. **Autoregressive Process: p**
How long back need to trace, to have weighted average:
The longer we trace back, distribution be smoother:
Example: random walk: build on historical value
$X_t = Z_t + \phi_1 X_{t-1} + ... + +\phi_p X_{t-p}$
$Z_t = (1 - \phi_1 B + ... + +\phi_p B^p) X_t = \Phi(B) X_t$
$X_t = \frac{1}{1-(1-\phi_1 B + ... + +\phi_p B^p)} Z_t = (1 + \theta_1 B + \theta_2 B^2 + ...) Z_t$

Then we have: $E[X_t] = 0\ ;\ Var[X_t] = \sigma_z^2 \sum_{i=0}^{\infty} \theta_i^2$
$\gamma(k) = \sigma_z^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+k}\ ;\ \rho(k) = \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+k}}{\sum_{i=0}^{\infty} \theta_i \theta_i}$

7. **ARMA (q+p) Models**
$X_t Noise + AR + MA$
$X_t = Z_t + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$
In terms of back function, we have:
$AR : \phi(B) X_t = X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p}$
$MA : \theta(B) Z_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$
$ARMA : \theta(B) Z_t = \phi(B) X_t$

The goal of backward shift function is to transform a mix process to MA process:
$\frac{\theta(B)}{\phi(B)} Z_t = X_t\ ;\ or\ :\ \frac{\phi(B)}{\theta(B)} Z_t = Z_t$

We can solve by geometric Series Transformation

8. **ARIMA (q+p) Models**
From ARMA
$ARMA$
$X_t = Z_t + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$
$\theta(B) Z_t = \phi(B) X_t$
we expect $\beta(z)$ and $\phi(z)$ have all roots lie outside unit circle, z is a complex number. Such that ARMA(q,p) is stationary and invertible

We loose the stationary constraint, introduce difference operator ∇:
$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t$
for random walk: $X_t = X_{t-1} + Z_t \equiv \nabla X_t = Z_t$

Then we define ARIMA(p,d,q) process:
$X_t \rightarrow s.t. \rightarrow Y_t = \nabla^d X_t = (1 - B)^d X_t$ is ARMA(p,q)

$ARIMA : \phi(B) \nabla^d X_t = \beta(B) Z_t$
$or : \phi(B)(1 - B)^d X_t = \beta(B) Z_t$

ARIMA is a process, when trend is removed, be ARMA
d: order of differencing, usually =1 or =2
over differencing may introduce dependence, and we decide whether differencing is needed by looking at ACF (when there is a slow decay)

9. **SARIMA Processes: Seasonality**
First we look at the seasonal ARMA process:
$ARMA(P,Q)_s : \Phi_P(B^s) X_t = \Theta_Q(B^s) Z, with :$
$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps}$
$\Theta_Q(B^s) = 1 + \Theta_1 B^s + +\Theta_2 B^{2s} + ... + +\Theta_Q B^{Qs}$

Then we look at the seasonal ARIMA Process:
$SARIMA(p,d,q,P,D,Q)_s :$
$\Phi_P(B^S) \phi_p(B)(1 - B^S)^D (1 - B)^d X_t = \Theta_Q(B^S) \theta_q(B) Z_t$

$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps}$
$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_P B^p$
$\Theta_Q(B^s) = 1 + \Theta_1 B^s + +\Theta_2 B^{2s} + ... + +\Theta_Q B^{Qs}$
$\theta_q(B) = 1 + \theta_1 B + +\theta_2 B^2 + ... + +\theta_Q B^q$
In which the capital terms are seasonal part

Now we look at the ACF of SARIMA:
Ex: we have seasonality of 12, then we also have some acf at 11

10. **Strict Stationary**
Definition: shifted distribution is the same:
$P\{X(t_1), X(t_2) ... X(t_k)\} = P\{X(t_{1+\tau}), X(t_{2+\tau}) ... X(t_{k+\tau})\}$
$\mu(t) = \mu\ ;\ \sigma^2(t) = \sigma^2\ ; \gamma(t_1, t_2) = \gamma(t_2 - t_1) = \gamma(\tau)$
Which leads to the properties:
1. random variables are **independently** distributed
2. the mean and variance function is identical
3. the covariance (ACF) depends only on **lag spaces**

11. **Weak Stationary**
Definition: constant mean and ACF

12. **Backward Shift Operator**
$BX_t = X_{t-1}\ ;\ B^2 X_t = X_{t-2}\ ;\ B^k X_t = X_{t-k}$
Further decompose the equation, we can have relationship between X and Z:
For $MA(q)$:
$X_t - \mu = \beta(B) Z_t$
$\beta(B) = \phi_0 + \phi_1 B + ... + \phi_q B^q$
For $AR(p)$:
$\phi(B) X_t = Z_t$
$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$

13. **Invertibility**
Definition:
Stochastic process: $\{X_t\}$,
Innovation (random process): $\{Z_t\}$
if $Z_t = \sum_{k=0}^{\infty} \pi_k X_{t-k}$ (AR), then $\{X_t\}$ invertible

---

We try to see if we can make the invert:
Invert a MA process to AR process:
1. First try to express Z_t using X_t, X_{t-1}
2. Change the Z_t in MA definition using X_{t-1}
3. Therefore we change MA to a AR

14. **Duality**
a). Invertibility Condition for MA(q)
$\beta(B) = \beta_0 + \beta_1 B + ... + \beta_q B^q$
its roots all lies **outside** in unit circle, $\rightarrow$, invertible
b). Stationarity Condition for AR(p)
$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - -\phi_p B^p$
its roots all lies **outside** in unit circle, $\rightarrow$, invertible
c). How MA and AR are related
if Invertible: $MA(q) \longrightarrow AR(\infty)$
if Stationary: $AR(p) \longrightarrow MA(\infty)$

15. **Mean-Square Convergence**

16. **Difference Equation**
Recall the way we solve recursive series for general form
Difference equation is dealing with the same problems.

17. **Yule-Walker Equations**
A set of Difference Equations, that governs AR's ACF
It is useful in estimating AR's parameters:
1. AR(p) :
$X_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + Z_t$

2. Expectation Form :
$\mu = \phi_0 + \phi_1 \mu + \phi_2 \mu + ... + \phi_p \mu + 0$

3. Subtract and redefine variable :
$X_t - \mu = 0 + \phi_1(x_{t-1} - \mu) + ... + \phi_p(x_{t-p} - \mu) + Z_t$
$\sim x_t = x_t - \mu$; $E[\sim x_t = 0]$
$\sim x_t = \phi_1 \sim x_{t-1} + \phi_2 \sim x_{t-2} + ... + \phi_p \sim x_{t-p} + Z_t$

4. Since the new process is stationary
its ACF follows the Yule-Walker Equation:
$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + ... + \phi_p \rho(k-p)$

5. fitting in k = 1,2,...p and since $\rho(k) = \rho(-k)$
this gives us a matrix form,
we can use sample to solve and estimate all the Φ

18. **Akaike Information Criterion (AIC)**
This is measuring whether a model is good enough, we can see it as some sort of scores for deciding which model to explain the data

19. **Q-Statistics**
Original Q statistics:
$Q^*(m) = T \sum l = 1^m r_l^2$
for null hypothesis: $H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$
alternative hypothesis: $H_1 : some \rho_i \neq 0$
when r_t is i.i.d:
$Q^*(m) \sim \chi^2(df = m)$

Ljung and Box Q statistics:
$Q(m) = T(T+2) \sum_{l=1}^{m} \frac{r_l^2}{T-l}$ reject null hypothesis, when:
$Q(m) > \chi^2_\alpha$

usually: we use $m \approx log(T)$

20. **Putting it all together:**
1. Trend: there is a trend, we may want differencing
2. Variation of variance: transformation is needed, such as log transformation
3. ACF: tells us the order of MA
4. PACF: tells us the order of AR
5. AIC: tells us which model seems to be better, but remember to consider whether it is worthy to introduce the complexity
6. SSE:
7. Ljung-Box Q-statistics
8. Estimation

21. **Prediction: Exponential Smoothing:**
1. Simple Exponential Smoothing
2. Second Exponential Smoothing
3. Third Exponential Smoothing
Take Away: explain SES