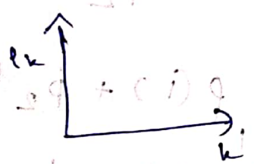
Correlogram or ACF Plot

The graph of ρ_k against k is called as an correlogram. The ACF plays a very important role in model-identification.



Partial Autocorrelation Function (PACF)

PACF of order k (denoted by α_k), is the partial correlation between y_t, y_{t-k} conditional on the intermediate values of the process.

Thus, α_k is the autocorrelation between y_t, y_{t-k} removing their linear dependence on $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$.

$$\text{Corr}\left(y_t, y_{t-k} \mid y_{t-1}, y_{t-2}, \dots y_{t-k+1}\right)$$

General AR(p) process

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \quad (1)$$

where $\{e_t\}$ is an iid sequence with mean zero and variance σ_e^2.

Thus $\phi_1, \phi_2, \dots, \phi_p, c, \sigma_e^2$ are the parameters of the process.

If $E(y_t) = E(y_{t-1}) = \dots E(y_{t-p}) = \mu$

Upon taking expection on (1)

$$E(y_t) = c + \phi E(y_{t-1}) + \dots + \phi_p E(y_{t-p})$$

$$\mu = c + \phi \mu + \dots + \phi_p \mu$$

$$\Rightarrow \mu = \frac{c}{(1 - \phi_1 - \dots - \phi_p)}$$

Non-Stationary Time Series

Transformation → to correct nonstationarity

The Backshift operator (B) is an important operator when it comes to defining time series lags.

$$BY_t = Y_{t-1}$$

$$B^2 Y_t = Y_{t-2} \qquad B(BY_t) = B(Y_{t-1}) = Y_{t-2}$$

$\vdots$

$$B^d Y_t = Y_{t-d}$$

we back shift the time span by d units.

Differencing Operator

By suitable differencing of such a process we are able to come to a stationary process. This technique is due to Box and Jenkins.

inverted ∇ ← $\nabla^d Y_t$ is the d^{th} difference of Y_t for all t, where

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

Nabla

$$\nabla^2 Y_t = \nabla(\nabla Y_t)$$
$$= \nabla(Y_t - Y_{t-1})$$
$$= Y_t - Y_{t-1} - Y_{t-1} + Y_{t-2}$$
$$= Y_t - 2Y_{t-1} + Y_{t-2}$$

∇Y_t is hence called as the differenced process.

What is the relation between ∇ and B?

$$BY_t = Y_{t-1} \qquad \nabla Y_t = Y_t - BY_t = (1-B)Y_t$$

$$\nabla Y_t = (1-B) Y_t$$

Ex Linear Trend

$Y_t = bt + S_t$, where S_t is the stationary process.

Is Y_t stationary?

No.

How would you make it ~~statio~~ stationary.

$$W_t = \nabla Y_t = bt + S_t - (b(t-1) + S_{t-1})$$

$$= S_t + b - S_{t-1} \longrightarrow \text{stationary}$$

Ex Quadratic Trend :

$$Y_t = bt^2 + S_t$$

where S_t is a stationary process.

Is Y_t stationary ?

$$\nabla Y_t = bt^2 + S_t - [b(t-1)^2 + S_{t-1}]$$

$$\cancel{\nabla^2 Y_t =}$$

Let $W_t = \nabla^2 Y_t$,

Then $W_t = Y_t - 2Y_{t-1} + Y_{t-2}$

Thus

$$W_t = bt^2 + S_t - 2[b(t-1)^2 + S_{t-1}]$$
$$+ b(t-2)^2 + S_{t-2}$$

$$= bt^2 + S_t - 2b[t^2 - 2t + 1]$$
$$- 2S_{t-1} + b\, t^2 - 4tb + 4b + S_{t-2}$$

$$= \cancel{bt^2} + S_t - 2\cancel{bt^2} + 4\cancel{bt} - 2b$$
$$- 2S_{t-1} + \cancel{bt^2} - 4\cancel{tb} + 4b + S_{t-2}$$

$$= 2b + S_t - 2S_{t-1} + S_{t-2}$$

which is stationary

Thus if the trend is a polynomial of order k, we can difference the series k times for stationarity.

Random Walk Model

used to describe non-stationary data such as income, stock price etc.

$$Y_t = Y_{t-1} + \epsilon_t$$

why is it non-stationary.

Think about an AR(1) model structure

$$Y_t = c + \phi_1 Y_{t-1} + \epsilon$$

$\Rightarrow c = 0$

$\phi_1 = 1$

AR(1) is stationary only when $|\phi_1| < 1$

But here $\phi = 1$, so non-stationary.

What about ∇Y_t

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

$$= \epsilon_t \rightarrow \text{stationary}$$

Seasonal Model

Example of seasonal data : airline passenger sales data, accident data etc.

Classical decomposition model :

$$Y_t = m_t + S_t + S_t$$

where

m_t : trend component

S_t : seasonal component with period d

S_t : stationary process.

Differencing at lag d

We apply the lag d difference as

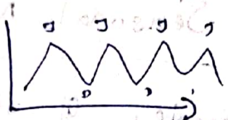
→ Y_t and d^{th} i th la

$$W_t = \nabla_d Y_t = Y_t - Y_{t-d}$$

$$= (m_t - m_{t-d}) + (S_t - S_{t-d})$$

So a lag d difference removes seasonality of period d i.e. $S_t - S_{t-d} = 0$

to reduce seasonality.



Autoregressive Process AR(p)

We forecast the variable of interest using a linear combination of past values of the variable. If the model uses last p values, it's called an AR(p) model.

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

Y_t is regressed on its own past values.

Autoregressive Moving Average Process : ARMA(p,q)

It is the combination of AR(p) and MA(q) model,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} \quad \text{AR term}$$

$$+ e_t + \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad \text{MA(q)}$$

Seasonality of period 4 ⇒ something is repeated in every quarter.

Depending on how the repetitions are occurring in a seasonal time series one can actually define the period of seasonality.

Depending on what period on seasonality you have, one has to apply lag d difference to remove the seasonality.

ARIMA(p, d, q) Process [Autoregressive Integrated Moving Average Process]

A non-stationary time series process Y_t follows ARIMA(p,d,q) model if

d th difference

$W_t = \nabla^d Y_t = (1-B)^d Y_t$, produced by differencing Y_t "d" times, is a stationary ARMA(p, q) process.

Initial model $Y_t \rightarrow$ non stationary, it has very some strong trend component, then

It has to difference d times, the resulting series $W_t \sim$ stationary ARMA(p, q)

If you are starting with some ARIMA model, and if you differencing the ARIMA process 'd' times, so it has become a stationary ARMA process with order p, q. Then the initial

$$Y_t \sim \text{ARIMA}(\underset{\text{AR part}}{p}, d, \underset{\text{MA}}{q})$$

$$Y_t \xrightarrow[\nabla^2 \to 2\text{times}]{} W_t \sim \text{ARMA}(2, 1)$$

Then $Y_t \sim \text{ARIMA}(2, 2, 1)$

Operators

- Lag 1 difference : $\nabla Y_t = Y_t - Y_{t-1}$
- seasonality ← lag d difference $\nabla_d Y_t = (1 - B^d) Y_t$
$= Y_t - Y_{t-d}$
- Trend remove ← d^{th} difference : $\nabla^d Y_t = (1-B)^d Y_t$
- If the trend m_t is a polynomial of order d, then $\nabla^d Y_t$ is stationary.
- If the data is a seasonal with period d, then a lag d difference will remove the seasonality.

Causes of Non-Stationary

Unit Roots

Consider a model

$$Y_t - 1.9 Y_{t-1} + 0.9 Y_{t-2} = \epsilon_t - 0.5 \epsilon_{t-1}$$

This is an ARMA model. (ARMA (2, 1))

The LHS can be written down as -

$$(1 - 1.9B + 0.9B^2) Y_t$$

B is backshift operator $BY_t = Y_{t-1}$

Thus, the roots of $1 - 1.9B + 0.9B^2 = 0$ are of interest to us. The roots happen to be 1 and 10/9. We hence have a 'unit root', which means whenever one of the roots is one, we call it unit root.

Unit roots makes the process non-stationary. However, if we assume, $W_t = \nabla Y_t$ we get

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

$$= 1.9 Y_{t-1} - 0.9 Y_{t-2} + \epsilon_t - 0.5\epsilon_{t-1}$$

$$- 1.9 Y_{t-2} + 0.9 Y_{t-3} + \epsilon_t + 0.5\epsilon_{t-1}$$

$$= 1.9(Y_{t-1} - Y_{t-2}) - 0.9(Y_{t-2} - Y_{t-3})$$

2

$W_t = \nabla Y_t \Rightarrow Y_t - Y_{t-1}$

$W_{t-1} = Y_{t-1} - Y_{t-2}$

We want to rewrite the left hand side of original equation using W_t and W_{t-1}

Start with

$$Y_t - 1.9Y_{t-1} + 0.9Y_{t-2}$$

~~Want~~ $$Y_t - Y_{t-1} + Y_{t-1} - 1.9Y_{t-1} + 0.9Y_{t-2}$$

$$(Y_t - Y_{t-1}) + (1 - 1.9)Y_{t-1} + 0.9Y_{t-2}$$

$$(Y_t - Y_{t-1}) + .9Y_{t-1} + 0.9Y_{t-2}$$

$$Y_t - Y_{t-1} + 0.9(Y_{t-1} - Y_{t-2})$$

$$W_t - 0.9W_{t-1}$$

Original RHS is unchanged

So the final differenced eqn is

$$W_t - 0.9W_{t-1} = \epsilon_t - 0.5\epsilon_{t-1}$$

This is a stationary ARMA(1, 1) model

So the original process before differencing was ARIMA (1, 1, 1)

Why is Unit Root a Problem:

Unit roots imply the following model structure:

$$(1-B)Y_t = Y_t - Y_{t-1} = f(\epsilon_t)$$

where $f(\epsilon_t)$ is some function of error term. This implies that the current value is same as past value plus some white noise structure. Thus, in the long run the process will never converge to its mean and hence is non-stationary.

Causes of Non-Stationarity

A practical time series can be nonstationary due to following reasons:

1. Presence of unit roots
2. Presence of deterministic polynomial trend
3. Presence of a stochastic trend
   ↳ trend depends on time

Tests for Stationarity

Augmented Dickey Fuller Test

It tests if a unit root is present in the time series sample.

H_0: series is non-stationary

H_1: Series is stationary

Kwiatkowski - Phillips - Schmidt - Shin (KPSS) test

It Tests the null hypothesis that ~~the~~ a time series is stationary around a deterministic trend (or mean) against the alternative that it is non-stationary.

H_0: series is stationary

H_1: Series is non-stationary.

Phillips - Perron (PP) test

Similar to the ADF test but it makes different assumptions about the error terms and is more robust in the presence of auto correlation and heteroskedasticity.
↓ changing var

H_0: series is non-stationary

H_1: series is stationary.

Variance Ratio Test

Tests for random walk hypothesis, indicating non-stationarity if the hypothesis is not rejected. Ratio significantly different from one, suggests the series is not a random walk.

H_0: Series is non-stationary

H_1: Series is stationary.

Model Identification

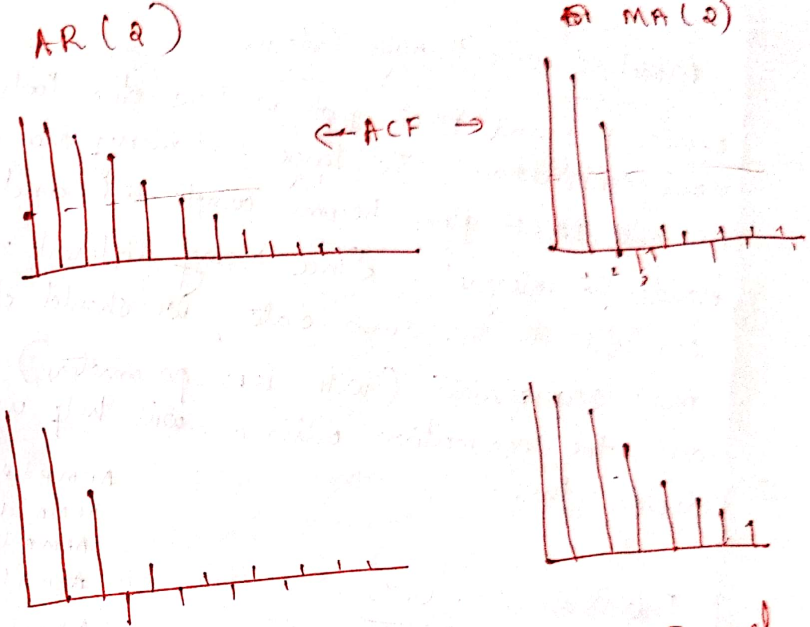
AR, MA, ARMA, ARIMA → discussed

The question now is, how to find the best suitable model, given the practical data, i.e., finding the optimal orders (p,q) of the model.

Using ACF and PACF Plots

As discussed earlier, one can find the optimal orders (p, q) using ACF and PACF plots:

	ACF	PACF
AR(p)	Tails off after lag p	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off after lag q.



AR model show gradual decline in ACF and a single peak at the order in PACF

MA models show gradual decline in PACF and (PF a single peak at the order in ACF.

Model Selection through Criteria

Besides ACF and PACF plots, we have other tools for model identification. For large and many real data, ACF and PACF plots become complicated and harder to interpret. Since many different models can fit to the same data, we should choose most appropriate (with less parameters) one and the information criteria will help us to decide this.

ARMA(2, 2)
ARMA(1, 2)
ARMA(2, 1)
ARMA(1, 1)

Few Information Criteria

- Akaike Information Criteria (Akaike, 1974)
- Schwarz's Bayesian Criteria (SBC) also known as Bayesian Information criteria (Schwarz 1978)
- Hannan-Quinn Criteria (HQIC) (Hannan & Quinn 1979)

Akaike's Information Criteria (AIC)

Assume that a statistical model with M parameters are fitted to data

$$AIC = -2\log(\text{maximum likelihood}_{fn}) + 2M$$

Pick the model (or value of M) with minimum AIC

Thus after minimizing the log-likelihood fun, we get

ARMA (1, 2)
ARMA (2, 1)

~~Schwarz's Bayesian Criteria~~ (SBC)

$$AIC = n\log\hat{\sigma}_e^2 + 2M$$

$\hat{\sigma}_e^2 \rightarrow$ estimation of variance

Pick the model (or value of M) with minimum AIC

~~Shev~~ Schwarz's Bayesian Criteria (SBC)

- SBC is a criteria for selecting from models having different number of parameters.
- When estimating model parameters using maximum likelihood estimation, it is possible to increase the likelihood by adding additional parameters, which may result in overfitting. The BIC resolves this problem by introducing a penality term for the number of parameters in the model.

In SBC, the penality for putting additional parameters is stronger than in AIC. $\rightarrow$ as $n \rightarrow \infty$, $(\log n) \ln n > 2$

$$SBC = n\log\hat{\sigma}_e^2 + M\log n$$

- SBC has superior large sample property
- It is consistent, unbiased and sufficient.

Hannan - Quinn Information Criteria

An alternative to AIC and SBC

$$HQIC = n\log\hat{\sigma}_e^2 + 2M\log(\log n)$$

Other techniques: Sample inverse autocorrelation f

Model Estimation

- After specifying the model orders, we need to estimate the underlying parameters.
- We can assume that the order is known (p, q) and the model has 0 mean.
- If the mean is not zero, we can always substract the sample mean $\bar{Y}$, fit a 0 mean ARMA model (say X_t), and then get back to Y_t by applying $Y_t = X_t + \bar{Y}$.

Estimation Techniques

- Method of Moments (MOME)
- Maximum likelihood estimation (MLE)
- Ordinary least squares estimation (OLS)
- Least ~~sho~~ Squares Estimation (conditional and unconditional)

∴ Method of Moments

- Also called as Yule Walker Estimation
- Works for only AR models for large n
- Not an efficient method.
- Basic idea: Equating the population moments to sample moments and solve for the parameters.

popln mean ← $E(Y_t) = \frac{1}{n}\sum_{t=1}^{n} Y_t$ → sample mean $\Rightarrow \mu = \bar{Y}$

$$E(Y_t\, Y_{t+k}) = \frac{1}{n}\sum^{n} Y_t\, Y_{t+k}$$

$$\Rightarrow \gamma_k = \hat{\gamma}_k$$

Similarly $\rho = \hat{\rho}_k$

Yule walker Estimation for AR(1)

$$Y_t = \phi Y_{t-1} + e_t \qquad \Rightarrow \rho(k) = \phi^k$$

It is know that $\rho_1 = \phi$

$$\text{Thus } \hat{\phi} = \hat{\rho}_1 = \frac{\sum_{t=1}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

Also the error variance (σ_e^2) is unknown,

But we know.

$$\gamma_0 = \frac{\sigma_e^2}{1-\phi^2}$$

Thus

$$\hat{\sigma}_e^2 = (1 - \hat{\phi}^2)\hat{\gamma}_0$$

$$= (1 - \hat{\phi}^2)\frac{1}{n}\sum_{t=1}^{n}(Y_t - \bar{Y})^2$$

$$= (1 - \hat{\rho}_1^2)\frac{1}{n}\sum_{t=1}^{n}(Y_t - \bar{Y})^2$$

Maximum Likelihood Methodology

Assume that errors $e_t \sim N(0, \sigma_e^2)$, and are also independent and identical.

Now the joint distribution $f(y_1 \cdots y_n)$ cannot be broken down into product of marginals.

So we make use of

$$f(e_1, e_2, \dots e_n) = f(e_1) \cdot f(e_2) \cdots f(e_n)$$

$\hookrightarrow$ ∵ all the errors are independent as per our assumption

Diagnostic Checking

After identifying and estimating the model parameters, the goodness of fit of the model and validity of all the assumptions must be checked. This process is called as diagnostic checking.

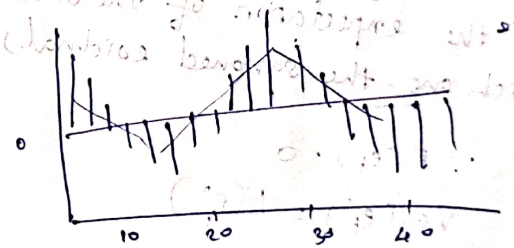
If the model fit is perfect, we can go ahead with forecasting.

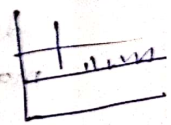
Normality of Errors

- Check the histogram of the standardized residual i.e. $\hat{e}_t / \sigma_e^2$
- Draw Normal Q-Q plots of standardized residual.
- Look at Tukey's five number summary along with skewness and kurtosis or excess kurtosis.
- Conduct formal test such as Shapiro-Wilk, Jarque-Bera etc.

Detection of Serial Correlation:

Generally, residuals are found to be correlated with their own lagged values. This property is called serial autocorrelation.





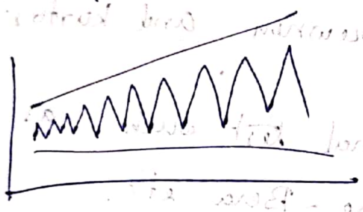
Box-Pierce Test

~~Detection of Serial to~~

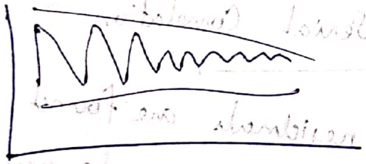
Detecting Heteroscedasticity

Heteroscedasticity implies changing variance. It occurs if the errors are having a non-constant variance structure.

$$Var(\epsilon_t) = \hat{\sigma}_t$$



residual



funnelling in

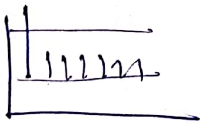
ACF-PACF plot of squared residual

$\because$ ϵ_t's have a zero mean, the variance is defined by the expectation of the second moment, which are the squared residuals.

$$\because E(e_t) = 0$$

$$Var(e_t) = E(e_t^2)$$

Thus if the variance of errors are indeed constant the ACF and PACF of squared residuals should be within 95% limits.



Forecasting Basics

One of the most important objectives in time series is to forecast future values.

- Estimation: Estimating the unknown parameter in a model.
- Prediction: Value of the random process, when we use the estimated values of the parameters.
- Forecasting: Value of any future random process which is not observed by the sample.

Estimation: AR(1) : ϕ_1

$$Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$$

$$\hat{Y}_t = c + \hat{\phi}_1 Y_{t-1} + \epsilon_t$$

Forecasting

$$Y_t = \phi Y_{t-1} + a_t$$

$$\Downarrow$$

$\hat{\phi}$ (estimate of ϕ)

$$\Downarrow$$

$\hat{Y}_t = \hat{\phi} Y_{t-1}$ (Prediction)

$$\Downarrow$$

$\hat{Y}_{t+1}$ (Forecasting).

Forecasting from an ARIMA model

The minimum mean squared error forecasts

Using an observed time series $y_1, y_2, \ldots, y_n$, how do we forecast future values y_{n+1}, y_{n+2} etc. Here n is the forecast origin.

$\hat{Y}_n(1)$: forecast value of Y_{n+1}

$\hat{Y}_n(2)$: forecast value of Y_{n+2}

$y_1, y_2, \ldots y_n \mid \hat{y}_{n+1}, \hat{y}_{n+2}$

$\hat{Y}_n(l)$: forecast value of Y_{n+l} — l-step ahead forecast, obtained using minimum MSE criteria.

Forecasting from an ARMA model

$$\hat{Y}_n(l) = E(Y_{n+l} \mid Y_n, Y_{n-1}, \ldots, Y_1)$$

which is the conditional expectation of Y_{n+l} given the observed sample.

ARMA Models

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} \cdots - \theta_q a_{t-q}$$

which could be written down as

$$\phi_p(B) Y_t = \theta_0 + \theta_q(B) a_t \quad \text{— (1)}$$

The actual l-step ahead value of the process is:

$$Y_{n+l} = \theta_0 + \phi_1 Y_{n+l-1} + \cdots + \phi_p Y_{n+l-p} + a_{n+l} - \theta_1 a_{n+l-1} - \cdots - \theta_q a_{n+l-q}.$$

Considering the random ~~ser~~ shock form of the series

(1) $\Rightarrow$

$$Y_{n+l} = \theta_0 + \frac{\theta_q(B)}{\phi_p(B)} a_{t+l}$$

$$= \theta_0 + \psi(B) a_t$$

$$= \theta_0 + a_{n+l} + \psi_1 a_{n+l-1} + \psi_2 a_{n+l-2} + \cdots + \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots$$

forecast error become

$$e_n(l) = Y_{n+l} - \hat{Y}_n(l)$$

$$E(e_n(l)) = 0$$

$$Var(e_n(l)) = \sigma^2 \sum_{i=0}^{l-1} \psi_i^2$$

$$\hat{Y}_n(l) = E(Y_{n+l} \mid Y_n, Y_{n-1}, \ldots, Y_1) = \theta_0 + \psi_l a_n + \psi_{l+1} a_{n-1} + \ldots$$