Regression and Time Series Model

-Gaurav Suryawanshi

1 SIMPLE LINEAR REGRESSION

Heading	Description or value
Simple Regression Formula	$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ \forall i = 1,2 \dots n$ $\epsilon_i \sim N(0, \sigma^2)$
Regression Coefficients	β_0 and β_1
Expectations and variance	$E(y\|x) = \beta_0 + \beta_1 x$ $var(y\|x) = \sigma^2$
Cost function	$S(\beta_0, \beta_1) = \sum_i^n (y_i - \beta_0 - \beta_1 x_i)^2$
Value of coefficients, beta_0 and beta_1	$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\mathrm{x}$ $\widehat{\beta_1} = \frac{\sum y_i x_i - n\bar{Y}\bar{X}}{\sum x_i^2 - n\bar{X}^2} = \frac{S_{xy}}{S_{xx}}$
Sxx and Sxy	$S_{xx} = \sum_i^n (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\sum x_i^2}{n}$ $S_{xy} = \sum y_i x_i - \frac{\sum y_i \sum x_i}{n} = \sum_i^n y_i(x_i - \bar{X})$
Variance of coefficients	$Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$ $Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$
Residual	$e_i = y_i - \widehat{y_i}$

1.1 BEST ESTIMATORS

Heading	Description or value
SSres	$SS_{res} = \sum e_i^2 = \sum y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy}$ $SS_{res} = SS_T - \hat{\beta}_1 S_{xy}$
SSt	$SS_T = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$
Expected value of SSres	$E(SS_{res}) = \sigma^2$
Unbiased Estimator of Variance or Residual Mean Square Average MSres	$\hat{\sigma}^2 = \frac{SS_{res}}{n-2} = MS_{res}$

Standard Error in Regression	$\sqrt{\hat{\sigma}^2} = \sqrt{MS_{res}}$

1.2 Hypothesis Testing

Heading	Description or value
Standard t-test	$t = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$
t-test, NID stands for Normal Independent	$H_0\!: \beta_1 = \beta_{10}$ $H_1\!: \beta_1 \neq \beta_{10}$ $\epsilon_i \sim iid(0, \sigma^2)$ $y_i = NID(\beta_0 + \beta_1 x_i, \sigma^2)$ $\hat{\beta}_1 = N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ $Z_0 = \frac{(\hat{\beta}_1 - \beta_{10})}{\sqrt{\frac{\sigma^2}{S_{xx}}}}$
Modified t-test as sigma is unknown	$E(\sigma^2) = MS_{res}$ $t_0 = \frac{(\hat{\beta}_1 - \beta_{10})}{\sqrt{\frac{MS_{res}}{S_{xx}}}}$ Follows t_{n-2} distribution Dof=Dof(MS_{res})=n-2 Reject Null Hypothesis if $\lvert t_0 \rvert > t_{\frac{\alpha}{2}, n-2}$
t-test for Intercerpt	$t_0 = \frac{(\hat{\beta}_0 - \beta_{00})}{\sqrt{MS_{res}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}$
Standard error	$se(\hat{\beta}_1) = \sqrt{\frac{MS_{res}}{S_{xx}}}$ $se(\hat{\beta}_0) = \sqrt{MS_{res}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$
ANOVA Test. SS_r stands for Regression	$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$ $SS_T = SS_R + SS_{res}$ $SS_R = \hat{\beta}_1 S_{xy}$ $df_T = n - 1$ $df_R = 1$ $df_{res} = n - 2$ $F_0 = \frac{\left(\frac{SS_R}{df_R}\right)}{\frac{SS_{res}}{df_{res}}}$ Reject null hypothesis if $F_0 > F_{\alpha,1,n-2}$
Expectation of MSres	$E(MS_{res}) = \sigma^2$ $E(MS_R) = \sigma^2 + \beta_1^2 S_{xx}$

1.3 Interval Testing

Heading	Description or value
100(1-α) Confidence Interval of β_i	$d_f = n-2$ $\hat{\beta}_i - t_{\frac{\alpha}{2},n-2}se(\hat{\beta}_i) \le \beta_i \le \hat{\beta}_i + t_{\frac{\alpha}{2},n-2}se(\hat{\beta}_i)$
Confidence Interval in sigma	$\frac{(n-2)MS_{res}}{\chi^2_{\frac{\alpha}{2},n-2}} \le \sigma^2 \le \frac{(n-2)MS_{res}}{\chi^2_{1-\frac{\alpha}{2},n-2}}$
Estimation of mean Response	$E(y\|x_0) = \hat{\mu}_{y\|x_0} = \hat{\beta}_0 + \hat{\beta}x_0$ $Var(\hat{\mu}_{y\|x_0}) = \sigma^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)$
Sampling Distribution with $dof = n-2$	$\frac{\hat{\mu}_{y\|x_0} - E(y\|x_0)}{\sqrt{MS_{res}\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}}$
Confidence Interval of mean Response	$\hat{\mu}_{y\|x_0} - t_{\frac{\alpha}{2},n-2}\sqrt{MS_{res}\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \le E(y\|x_0)$ $\le \hat{\mu}_{y\|x_0}$ $+ t_{\frac{\alpha}{2},n-2}\sqrt{MS_{res}\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$
Prediction of new values interval	$\varphi = y_0 - \hat{y}_0 \sim N(0,\sigma^2)$ $Var(\varphi) = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)$ $\hat{y}_0 - t_{\frac{\alpha}{2},n-2}\sqrt{MS_{res}\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \le y_0$ $\le \hat{y}_0$ $+ t_{\frac{\alpha}{2},n-2}\sqrt{MS_{res}\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$
Coefficient of Determination	$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{res}}{SS_T}$

1.4 Maximum Likelihood Estimators

Heading	Description or value
Generic Function	$e_i = \widehat{-y_i} + y_i = y_i - \beta_0 - \beta_1 x_i$ $L = \prod_i e_i$ $L = \prod_i^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\beta_0-\beta_1x_i)^2}$ $L = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_i^n (y_i-\beta_0-\beta_1x_i)^2}$
Estimators Here Variance estimator is biased.	$\tilde{\beta}_0 = \bar{y} - \beta_1\bar{x}$

	$\tilde{\sigma}^2 = \frac{\sum(y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}$ $\tilde{\sigma}^2 = \frac{(n-1)}{n}\hat{\sigma}^2$

1.5 JOINTLY DISTRIBUTED MODEL

Heading	**Description or value**
Generic Function	$f(y,x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(y-\mu_1)^2}{\sigma_1} + \frac{(x-\mu_2)^2}{\sigma_2} - \frac{2\rho(y-\mu_1)(x-\mu_2)}{\mu_1\mu_2}\right\}}$ $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ is correlation coefficient
Maximum Likelihood Parameters r is for ρ is the measure of linear association b/w y & x	$\tilde{\beta}_0 = \bar{y} - \beta_1\bar{x}$ $\tilde{\beta}_1 = \frac{\sum y_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$ $r = \frac{S_{xy}}{\sqrt{S_{xx}SS_T}}$
Hypothesis Testing for correlation	$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$

2 MULTIPLE LINEAR REGRESSION

Heading	**Description or value**
Generic Function	$y = B_0 + B_1x_1 + \cdots + B_ix_i + \cdots + B_kx_k + \epsilon$
Vectors	$\boldsymbol{y} = \boldsymbol{XB} + \boldsymbol{\epsilon}$ $y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}_{n\times 1}$ $X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n\times k}$ $B = \begin{bmatrix} B_1 \\ B_2 \\ \dots \\ B_k \end{bmatrix}_{k\times 1}$ $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix}_{n\times 1}$
Method Of least Squares.	$S(B) = e'e = (y - XB)'(y - XB)$ $\hat{B} = (X'X)^{-1}X'y = \frac{S_{xy}}{S}$

	$\hat{y} = X\hat{B} = X(X'X)^{-1}X'y = Hy$ $H = X(X'X)^{-1}X'$ $e = y - \hat{y} = (1 - H)y$
Sxx	$S_{xx} = (X'X)^{-1}$ $S_{xy} = X'y$
Covariance Defination	$cov(X,Y) = E((X - E[X])(Y - E[Y]))$ $= E[XY] - E[X]E[Y]$
Properties of LS Operators	$E(\hat{B}) = B$ $cov(\hat{B}) = var(\hat{B}) = var((X'X)^{-1}X'Y)) = \sigma^2(X'X)^{-1}$
Variance property in vector where A is const	$Var(AX) = A'Var(X)A$
04Estimation of σ^2	$SS_{res} = y'y - \hat{B}'X'y$ $MS_{res} = \frac{SS_{res}}{n-k-1}$

2.1 Maximum Likelihood Estimators

Heading	Description or value
Generic Function	$\epsilon = NID(0, \sigma^2 I)$ $L = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\epsilon'\epsilon\right\}$ $L = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(y - XB)'(y - XB)\right\}$
Variance Estimate	$\hat{\sigma}^2 = \frac{(y - X\hat{B})'(y - X\hat{B})}{n}$

2.2 Hypothesis Testing

Heading	Description or value
Best Estimates	$SS_{res} = y'y - \hat{B}'X'y = SS_T + SS_R$ $SS_T = y'y - \frac{(\sum y_i)^2}{n}$ $SS_R = \hat{B}'X'y - \frac{(\sum y_i)^2}{n}$
Hypothesis (ANOVA)	$E(\epsilon_i) = 0$ $Var(\epsilon_i) = \sigma^2$ $H_0: B_0 = B_1 = B_2 = \cdots = B_k = 0$ $H_1: B_i \neq 0$ for at least one i $SS_T = SS_R + SS_{res}$ $F_0 = \frac{\left(\frac{SS_R}{k}\right)}{\frac{SS_{res}}{n-k-1}} = \frac{MS_r}{MS_{res}} \sim F_{k,n-k-1}$ If values of F_0 is large then it's likely that H_1 is true Reject if $F_0 > F_{\alpha,k,n-k-1}$
Variance Estimate	$\hat{\sigma}^2 = \frac{(y - X\hat{B})'(y - X\hat{B})}{n}$

2.3 Hypothesis Testing on Individual Variables

Heading	Description or value
Basic Hypothesis	$H_0: B_j = 0$ $H_1: B_j \neq 0$ If H_0 is not rejected, we can delete the regression variable x_j $\hat{B} = N(B, (X'X)^{-1}\sigma^2)$ $t = \frac{(\hat{B_j} - B_j)}{\hat{\sigma}\sqrt{c_{j+1,j+1}}}$ Where $c_{j+1,j+1}$ is the diagonal element of $(X'X)^{-1}$ and $\hat{\sigma} = MS_{res}$ Reject if $\lvert t_0 \rvert > t_{\frac{\alpha}{2},n-k-1}$
Testing for a set of Regressor Variables	$y_{n\times1} = X_{n\times p}B_{p\times1} + \epsilon_{n\times1}$ $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ where B_1 is $(p-r)\times 1$ and B_2 is $r \times 1$ $H_0: B_2 = 0 \quad H_1: B_2 \neq 0$ $y = X_1B_1 + X_2B_2 + \epsilon$ is the full model $y = X_1B_1 + \epsilon$ is the reduced model $SS_R(B_2\lvert B_1) = SS_R(B) - SS_R(B_1) = \hat{B}'X'y - \hat{B}_1'X_1'y$ $dof(SS_R(B)) = p \;\; dof(SS_R(B_1)) = p - r$ $$F_0 = \frac{\frac{SS_R(B_2\lvert B_1)}{r}}{MS_{res}}$$ Here Fo follows non central F distribution with non-centrality parameter $\lambda = \frac{1}{\sigma^2}B_2'X_2'[I - X_1(X_1'X_1)^{-1}X_1']X_2B_2$
Testing General Linear Hypothesis	$$H_0: T_{m\times p}B = 0$$ Here only r equations out of m in TB=0 are independent. $$SS_{res}(FM_{dof=n-p}) = y'y - \hat{B}'X'y$$ Let we have a reduced model using r equations $$y = Z_{n\times(p-r)}\gamma_{(p-r)\times1} + \epsilon$$ $$SS_{res}(RM_{dof=n-p+r}) = y'y - \hat{\gamma}'Z'y$$ Get $$SS_H = SS_{res}(RM) - SS_{res}(FM)$$ Then $$F_0 = \frac{\frac{SS_H}{r}}{\frac{SS_{res}(FM)}{n-k-1}}$$ Reject H_0 if $F_0 > F_{\alpha,r,n-p}$ Or $$F_0 = \frac{(T\hat{B} - C)'[T(X'X)^{-1}T']^{-1}(T\hat{B} - C)/r}{SS_{res}(FM)/(n-k-1)}$$
Testing Equality of Regression model	Let $y = XB + \epsilon$, we want to check if $$TB = C$$ Then obtain a reduced model such that $$y = Z\gamma + \epsilon$$ Then use $$t_0 = \frac{TB - C}{se(TB - C)}$$

	And apply t-test

3 Residual Analysis

Heading	Description or value
R^2 test	$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{res}}{SS_T}$ $R^2_{adj} = 1 - \frac{SS_{res}}{n-k-1} * \frac{1}{\frac{SS_T}{n-1}}$
Basic Residuals	$e_i = y_i - \hat{y}_i \; \forall i$
Standardized Residuals	$\sigma^2 \sim MS_{res}$ $d_i = \frac{e_i}{\sqrt{MS_{res}}} \sim \mu = 0, \sigma^2 = 1$ Large values tend to be outliers
Studentized Residuals	$e = (1-H)y$ $e = (1-H)\epsilon$ $Var(e) = \sigma^2(1-H)$ $r_i = \frac{e_i}{\sqrt{MS_{res}(1-h_{ii})}}$
PRESS/Jack Knife Residuals	Use $y_i - \hat{y}_{(i)}$ where it is fitted value of response based on all obs but (i) $e_{(i)} = y_i - \hat{y}_{(i)} \; \forall i$ $e_{(i)} = \frac{e_i}{1-h_{ii}}$ $Var\left(e_{(i)}\right) = \frac{\sigma^2}{1-h_{ii}}$ $Std\ Press\ Residual = \frac{e_i}{\sqrt{\sigma^2(1-h_{ii})}}$
R student	Estimate σ^2 with i^{th} data removed $S^2_{(i)} = \frac{(n-k-1)MS_{res} - \frac{e_i^2}{1-h_{ii}}}{n-k-2}$ $t_i = \frac{e_i}{\sqrt{S^2_{(i)}(1-h_{ii})}}$

4 Time Series

4.1 Definitions And Basics

Heading	Description or value
Definition	A time series is generated from uncorrelated variables with 0 mean and fixed variance called white noise $W_t \sim WN(0, \sigma_w^2)$
Implementation	$$X_t = T_t + W_t + S_t$$ Where X_t=Time series at time t=t W_t=White noise, random error added at deterministic point S_t=Seasonal or repetition trend
Mean **Auto Covariance Function**	$\mu_t = E(X_t) = E(T_t) + E(S_t)$ $ACVF = \nu_x(s,t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$
Auto Covariance (ACVF)	Auto Covariance (t1, t2) = $E[(X_{t1} - \mu_{t1})(X_{t2} - \mu_{t2})]$ $ACF = \nu_x(h) = cov(X_{t+h}, X_t)$ Where h is the time period of seasonality
Auto Correlation (ACF)	Auto Correlation (t1,t2)=$E[X_{t1}.X_{t2}]$ $\rho(X_t, X_{t+h}) = \frac{\nu_x(h)}{\sqrt{\nu_x(X_{t+h})\nu_x(X_t)}} = \frac{\nu_x(h)}{\nu_x(0)}$ here $\nu_x(0) = \sigma^2$
Properties of ACF	$\rho(h) = \rho(-h) = \rho(\lvert h \rvert)$ $R = \rho_{ij} = \rho\lvert i-j \rvert = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ R is a p.s.d
Positive Semidefinite Function	$$\sum_i^n \sum_j^n f(\lvert i-j \rvert) a_i a_j \geq 0$$

4.1.1 Weakly and Strongly Stationary

Heading	Description or value
Weakly Stationary	i) $\mu_x(t)$ is independent of t ii) $\nu_x(t+h, t)$ is independent of each h Usually implies there is no trend in the series
Strongly Stationary	If joint distribution of $(X_1, \ldots X_n)$ and $(X_{1+h}, \ldots, X_{n+h})$ are same i.e $P\left(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\right) = P\left(X_{t_{1+h}} \leq x_1, \ldots, X_{t_{n+h}} \leq x_n\right)$

4.2 Different Types of Time Series

4.2.1 Random Walk

Heading	Description or value
Formulae	$S_t = \sum_i^t X_i \ (or) \sum_i^t W_i$

	$X_i \sim^{iid} N(0, \sigma^2)$
Mean and ACVF	$E(S_t) = E(\sum X_i) = 0$ $ACVF(S_t, S_u) = \nu_s(s,t) = var(S_t) = \sigma^2 t$ here t<s

4.2.2 Linear Process

Heading	Description or value
Formulae [WN process] **Weakly stationary** **If normally distributed -> Strongly Stationary**	$X_t = \mu + \sum_{j=-\infty}^{j=+\infty} \varphi_j w_{t-j}$ $w_t \sim WN(0, \sigma_w^2), \mu \in \Re$ $\sum_{-\infty} \lvert\varphi_j\rvert < \infty$
Mean and ACVF	$E(X_t) = \mu$ $Var(X_t) = \sigma_w^2(\sum \varphi_i^2)$ $\nu_x(h) = \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \varphi_j \times \varphi_{j-h} \right) < \infty$
Generic Linear Process	$Y_t = \mu + \sum_{j=-\infty}^{j=+\infty} \varphi_j X_{t-j}$ $\mu \in \Re$ $\sum_{-\infty} \lvert\varphi_j\rvert < \infty$
Mean and ACVF	$E(Y) = \mu$ $Var(X_t) = \sigma_w^2(\sum \varphi_i^2)$ $\nu_x(h) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \varphi_j \times \varphi_{j-h} \times \nu_x(h-j+k) < \infty$
Barlett's Formula about $\hat{\rho}(h)$	If $X_t = \mu + \sum_{j=-\infty}^{\infty} \varphi_j W_{t-j}$ Then $\begin{pmatrix} \hat{\rho}(i) \\ \hat{\rho}(j) \end{pmatrix} \sim N\left(\begin{pmatrix} \rho_{(i)} \\ \rho_{(j)} \end{pmatrix}, \frac{1}{n} \begin{bmatrix} \gamma_{ii} & \gamma_{ij} \\ \gamma_{ji} & \gamma_{jj} \end{bmatrix} \right)$ $\gamma_{ij} = \sum_{h=1}^{\infty} [\rho(h-i) + \rho(h+i) - 2\rho(i)\rho(h)] \times [\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h)]$

4.2.3 Auto Regression

Heading	Description or value
Formulae [AR(1)]	$X_t = \phi X_{t-1} + W_t$ Where $W_t \sim N(0, \sigma_w^2)$ & $\lvert\phi\rvert < 1$ $X_t = \sum_{j=0}^{t} \phi^j w_{t-j}$
Mean and ACVF	$E(X_t) = 0$ $Var(X_t) = \frac{\sigma_w^2}{1-\phi^2}$

	$\nu_x(h) = \frac{\sigma_w^2 \times \phi^{\lvert h \rvert}}{1-\phi^2}$
Formulae [AR(p)]	$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + W_t$ Where $W_t \sim N(0, \sigma_w^2)$ & $\lvert\phi\rvert < 1$

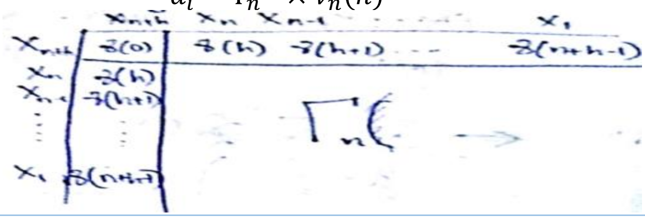
4.2.4 Moving Average (MA) Process

Heading	Description or value
Formulae (MA (1)) **Stationary in nature**	$X_t = W_t + \theta W_{t-1}$
Mean and other stats	$E(X_t) = 0$ $\nu_x(h) = \begin{cases} 0 & \lvert h\rvert \geq 2 \\ \sigma_w^2\theta & h = \pm 1 \\ \sigma_w^2(1+\theta^2) & h = 0 \end{cases}$
ACF	$\rho_x(h) = \frac{\nu_x(h)}{\nu_x(0)} = \begin{cases} 0 & \lvert h\rvert \geq 2 \\ \theta/(1+\theta^2) & \lvert h\rvert = 1 \end{cases}$
Formulae (MA (q))	$X_t = W_t + \sum_{i=1}^{q} \phi_i W_{t-i}$

4.2.5 Auto-Regressive Moving Average (ARMA) Process

Heading	Description or value
Formulae (ARMA (p,q)) **Stationary in nature**	$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$
ARMA(1,1)	$X_t = W_t + (\theta+\phi)\sum_{j}^{\infty} \phi^{j-1} W_{t-j}$ $W_t = X_t + (\theta+\phi)\sum_{i}^{\infty} \theta^{i-1} X_{t-i}$
Mean and Variance	$E(X_t) = 0$ $Var(X_t) = \frac{\sigma_w^2(\theta^2 + 2\phi\theta + 1)}{1-\phi^2}$
ACVF ARMA(1,1)	$\nu_x(h) = \begin{cases} \sigma_w^2\left(1 + \frac{(\theta+\phi)^2}{1-\phi^2}\right) & h = 0 \\ \sigma_w^2\left(\theta + \phi + \frac{\phi(\theta+\phi)^2}{1-\phi^2}\right) & h = 1 \\ \nu(1) \times \phi^{h-1} & h \geq 2 \end{cases}$
ARIMA Process	$\{X_t\}$ is ARIMA(p,d,q) if $Y_t = (I-B)^d X_t \sim ARMA(p,q)$ $\Phi_p(B)\nabla^d X_t = \Theta_q(B) W_t$
Correlation	$Corr(X_s, X_t) = \frac{\min\{s,t\}}{\sqrt{st}}$

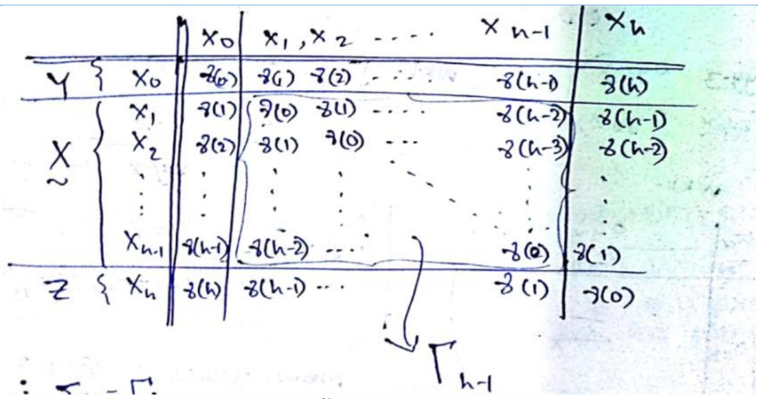
4.3 TREND ESTIMATION

Heading	Description or value
Estimation of Trend in Absence of Seasonality	$X_t = m_t + Y_t$ $E(Y_t) = 0$
Moving average method	$\widehat{m}_t = (2q+1)^{-1} \times \sum_{i=-q}^{q} X_{t+i}$
Exponential Smoothing Method	$m_1 = X_1$ $m_2 = \alpha X_2 + (1-\alpha)m_1$ $m_t = \alpha X_t + (1-\alpha)m_{t-1}$ $= \alpha X_t + \sum \alpha(1-\alpha)^k m_{t-k} + (1-\alpha)^2 X_1$
Poly Fit	$m_t = \sum_{k=0}^{\infty} a_k t^k$
Estimation of Trend & Seasonality	$X_t = S_t + T_t + W_t$ $E(W_t) = 0$ $S_{t+d} = S_t$ $\sum_{t=1}^{d} S_t = 0$
Estimation of Seasonality	First Estimate T_t then, use $w_k = x_{k+jd} - \widehat{m}_{k+jd}$ Where jth period and k=1 to d Then $\hat{S}_k = w_k - d^{-1} \sum_{i=1}^{d} w_i = \hat{S}_{k-d}$
Sample Auto Covariance Function (h is called lag)	$v(h) = \frac{1}{n} \sum_{i=1}^{n-\|h\|} (x_i - \bar{x})(x_{i+\|h\|} - \bar{x})$
Sample Variance	$Var = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$
Sample Mean	$E(\bar{x}) = \mu$ $Var(\bar{x}) = \frac{1}{n}\left[\sum_{h=-n}^{n} \left(1 - \frac{\|h\|}{n}\right) v_x(h)\right]$ $\lim_{n\to\infty} \frac{\sqrt{n}(x-\mu)}{\sigma} \to N(0,1)$
Estimation of Future values	$\hat{x}_{n+h} = \mu + \sum_{i=1}^{N} a_i (X_{h-i+1} - \mu)$ $a_i = \Gamma_n^{-1} \times v_n(h)$ 

Durban Levinson Algorithm	$\hat{X}_{h+n}^{n} = \sum_{i=1}^{n} a_i X_i$ $\hat{\alpha} = \Gamma_n^{-1} \nu_n(h)$ $E\left(X_{n+h} - \hat{X}_{n+h}\right)^2 = \nu_x(0) - \nu_n'(h) \times \Gamma_n^{-1} \times \nu_n(h)$ $a_n = \frac{\nu(\mathrm{n}) - \tilde{\nu}_{n-1}'(1) \times a_{n-1}^{old}}{\nu(0) - \tilde{\nu}_{n-1}'(1) \times a_{n-1}^{old}}$ $a_{n-1}^{new} = a_{n-1}^{old} - a_n \times \Gamma_{n-1}^{-1} \times \tilde{\nu}_{n-1}(1)$
Innovation algorithm	$\vec{U}_n = \vec{X}_n - \vec{\hat{X}}_n = A_n \vec{X}_n$ $\hat{X}_n = \Theta_n U_n = \Theta_n\left(X_n - \hat{X}_n\right)$ $\Theta_n = \begin{bmatrix} 0 & 0 & 0 \\ \theta_{ii} & 0 & 0 \\ \theta_{n-1,n-1} & \theta_{i,i-1} & 0 \end{bmatrix}$

4.4 Causality, Invertibility, Pacf And Model Accuracy

Heading	Description or value
Causality (A process is casual if _)	$X_t = \left(1 + \sum_{i}^{\infty} \phi_i B^i\right) W_t$ Here B is the backshift operator $B^h X_t = X_{t-h}$ and $\sum_{-\infty} \lvert \phi_j \rvert < \infty$
Invertibility (A process is invertible if _)	$W_t = \left(1 + \sum_{i}^{\infty} \theta_i B^i\right) X_t$ Here B is the backshift operator $B^h X_t = X_{t-h}$ and $\sum_{-\infty} \lvert \theta_j \rvert < \infty$
Generic PACF	$\rho_{yz,\vec{x}} = \frac{Cov\left(Y - E(Y \lvert \vec{X}), Z - E(Z \lvert \vec{X})\right)}{\sqrt{Var\left(Y - E(Y \lvert \vec{X})\right) \times Var\left(Z - E(Z \lvert \vec{X})\right)}}$ If normally distributed then $\left(\begin{smallmatrix} y \\ z \end{smallmatrix} \middle\vert \vec{X} = \vec{x}\right) \sim N\left(\begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} + \Sigma_{yz.x}' \times \Sigma_x^{-1} \times (\vec{x} - \overrightarrow{\mu_x}),\ \Sigma_{\mathrm{yz}} - \Sigma_{yz.x}' \times \Sigma_x^{-1} \times \Sigma_{yz.x}\right)$ $\rho_{yz,\vec{x}} = \frac{\sigma_{yz} - \sigma'_{y.\vec{x}} \times \Sigma_{\vec{x}}^{-1} \times \sigma_{z.\vec{x}}}{\sqrt{(\sigma_{yy} - \sigma'_{y.\vec{x}} \times \Sigma_{\vec{x}}^{-1} \times \sigma_{y.\vec{x}}) \times (\sigma_{zz} - \sigma'_{z.\vec{x}} \times \Sigma_{\vec{x}}^{-1} \times \sigma_{z.\vec{x}})}}$

PACF in Time series	 $\alpha(h) = \dfrac{\vartheta(h) - \tilde{\vartheta}'_{h-1}(1) \times \Gamma_{h-1}^{-1} \times \vartheta_{h-1}(1)}{\vartheta(0) - \vartheta'_{h-1}(1) \times \Gamma_{h-1}^{-1} \times \vartheta_{h-1}(1)}$ Where $\tilde{\vartheta}'(1) = \big(\vartheta(h-1) \ldots \vartheta(1)\big)$ $\vartheta'_{h-1}(1) = \big(\vartheta(1) \ldots \vartheta(h-1)\big)'$ $\Gamma_{h-1} = (\vartheta\lvert i-j \rvert)_{ij}$
Akaike Information Criterion	$AIC = 2k - 2\log_e \hat{L}$ Where k is the number of estimated parameters $\hat{L}$ is the maximum likelihood Minimum is the best
Bayesian Information Criterion	$BIC = \ln(n) \times k - 2\log_e \hat{L}$ Lower is prefered