Forecasting weakly stationary time series.

Assumption: (i) $\{X_t\}$ is weakly stationary

(ii) $\{x_1, x_2 \cdots x_n\}$ = data set given.

(iii) $E(X_t) = \mu$ known.

(iv) $\Gamma = ((\gamma|i-j|))$ is known.

GOAL: Given $\{x_1, x_2 \cdots x_n\}$ we want to predict the value of X_{n+h} for $h \in \mathbb{N}$ such that the least square error is minimized.

Notation. $Y = X_{n+h}$. $\underset{\sim}{X} = (X_n, X_{n-1} \cdots X_1)^T$.

Assume W.L.G. $E(Y) = 0$ $E(\underset{\sim}{x}) = \underset{\sim}{0}$.

Let $(Y, \underset{\sim}{X})$ has ~~a~~ joint distribution. $V(Y) = \sigma_{YY}$ $D(\underset{\sim}{X}) = \Sigma_x$ $\operatorname{cov}(Y, \underset{\sim}{X}) = \underset{\sim}{\sigma}_{yx}$. Then the best linear predictor of Y given $\underset{\sim}{X} = \underset{\sim}{x}$, is $\beta^T X$ which minimizes

$$S^2 = E(Y - \underset{\sim}{\beta}^T \underset{\sim}{X})^2 \Rightarrow \frac{\partial S^2}{\partial \beta}\Big|_{\beta} = \underset{\sim}{0} \Rightarrow E(Y, X) = E(X^T X)\hat{\beta}$$

$$= \underset{\sim}{\sigma}_{yx} = \Sigma_x \hat{\beta} \Rightarrow \boxed{\hat{\beta} = \Sigma_x^{-1} \underset{\sim}{\sigma}_{yx}}$$

	ACF	PACF
MA(q)	sharp cut after q lag	exponential decay
AR(p)	exponential decay	sharp cut after p - lag.
ARMA (p, q)	exponential decay	exponential decay.

General results from Multiple linear regression.

Let $(Y, \underset{\sim}{X})$ has some joint density.

Assumptions: $E(Y) = 0$, $E(\underset{\sim}{X}) = \underset{\sim}{0}$ $\quad cov(Y, \underset{\sim}{X}) = \underset{\sim}{\sigma}_{yx}$

$V(Y) = \sigma_{yy}$ $\quad D(\underset{\sim}{X}) = \Sigma_x$ $\quad = E(Y\underset{\sim}{X})$

The best linear predictor of Y given $\underset{\sim}{X} = \underset{\sim}{x}$ is $\underset{\sim}{\beta}^T \underset{\sim}{x}$ which minimizes the least-square (LS) condition

$$S^2 = E\left[\left(Y - \beta^T \underset{\sim}{X}\right)^2\right]$$

Normal equation.

Hence $\left.\dfrac{\partial S^2}{\partial \underset{\sim}{\beta}}\right|_{\hat{\beta}} = 0 \Rightarrow \cancel{E(Y/X)}\ E(Y\underset{\sim}{X}) = E(XX^T)\hat{\beta}$

$$\Rightarrow \underset{\sim}{\sigma}_{yx} = \Sigma_x \hat{\underset{\sim}{\beta}}.$$

$$\Rightarrow \hat{\underset{\sim}{\beta}} = \Sigma_x^{-1} \underset{\sim}{\sigma}_{yx}$$

$\dfrac{\partial}{\partial \underset{\sim}{x}}(A\underset{\sim}{x})$

$\dfrac{\partial}{\partial x}(x^T A x)$

<u>Result 1</u>: In multiple linear regression the best linear predictor is unique.

<u>Result 2</u>: Define the residual variance as.

$$\sigma_{y\cdot x} = Var\,(Y - \hat{\beta}\,\underset{\sim}{X})$$

(i) $$\sigma_{y\cdot x} = \sigma_{yy} - \underset{\sim}{\sigma}_{yx}^{T}\,\Sigma_x^{-1}\,\underset{\sim}{\sigma}_{yx}$$

$$= V(Y) - Var\,(\hat{\beta}^T \underline{X}).$$

(ii) $$\sigma_{y\cdot x}^{-1} = (1,1) \text{ element of } \Sigma^{-1}$$

when $\hat{\beta}$ is the best linear predictor.

$$\hat{\beta} = \Sigma_x^{-1}\,\underset{\sim}{\sigma}_{yx}$$

$$\Sigma = D\begin{pmatrix} Y \\ X \end{pmatrix} = \left(\begin{array}{c|c} \sigma_{yy} & \underset{\sim}{\sigma}_{yx}^{T} \\ \hline \sigma_{yx} & \Sigma_x \end{array}\right)$$

$$V(y - \hat{\beta}\underline{x})$$

$$= V(y) - 2\,\mathrm{cov}(y, \hat{\beta}^T \underline{x}) + V(\hat{\beta}^T \underline{x})$$

$$= \sigma_{yy} - 2\,\hat{\beta}^T \sigma_{yx} + \hat{\beta}^T \Sigma_x \hat{\beta}$$

$$= \sigma_{yy} - 2\,\sigma_{yx}^T \Sigma_x^{-1} \sigma_{yx} + \sigma_{yx}^T \Sigma_x^{-1} \Sigma_x \Sigma_x^{-1} \sigma_{yx}$$

$$= \sigma_{yy} - \sigma_{yx}^T \Sigma_x^{-1} \sigma_{yx} = \sigma_{y.\underline{x}}$$

(1) $$D\begin{pmatrix} y \\ \underline{x} \end{pmatrix} = \left(\begin{array}{c|c} \sigma_{yy} & \sigma_{yx}^T \\ \hline \sigma_{yx} & \Sigma_x \end{array}\right) = \Sigma$$

$$|\Sigma| = |\Sigma_x| \left(\sigma_{yy} - \sigma_{yx}^T \Sigma_x^{-1} \sigma_{yx} \right).$$

$$\sigma_{y.\underline{x}} = \frac{|\Sigma|}{|\Sigma_x|} = \frac{1}{(|\Sigma_x|/|\Sigma|)}$$

$$= \frac{1}{(1,1) \text{ of } \Sigma^{-1}}$$

$$\Rightarrow \sigma_{y.x}^{-1} = (1,1) \text{ element of } \underline{\Sigma^{-1}}.$$

$$\mathrm{cov}(y, \hat{\beta}^T \underline{x})$$

$$= \mathrm{cov}\left(y, \sum_{i=1}^{n} \hat{\beta}_i x_i\right)$$

$$= \cancel{\mathrm{cov}} \sum_i \hat{\beta}_i \left(\mathrm{cov}(y, x_i)\right).$$

$$= \sum_i \hat{\beta}_i \sigma_{yxi}$$

$$= \hat{\beta}^T \sigma_{yx}$$

If

$$Z \sim N(\mu, \Sigma)$$

$$AZ \sim N(A\mu, A\Sigma A^T)$$

$$E(AZ) = A\mu$$

$$D(AZ) = A\Sigma A^T$$

~~Exer~~

Remark 3. $$\left[\mathrm{Corr}\left(Y, \hat{\underset{\sim}{\beta}}\,\underset{\sim}{X}\right)\right]^2 = \frac{\mathrm{Var}\left(\hat{\underset{\sim}{\beta}}^T \underset{\sim}{X}\right)}{\mathrm{Var}(Y)}$$ (HW).

Remark 4 $$\mathrm{cov}\left(Y - \hat{\underset{\sim}{\beta}}^T X,\ X_j\right) = 0 \quad \forall\, j = 1, 2, \ldots n.$$

Best prediction error and predictor has covariance **zero**.

$$X_j = \underset{\sim}{e}_j^T \underset{\sim}{X} \qquad \text{where } \underset{\sim}{e}_j^T = (0\,0 \cdots \underset{j\text{th}}{\underset{\uparrow}{1}}\, 0 \cdots 0).$$

$$\underset{\sim}{X}^T = (X_1\, X_2 \cdots X_j \cdots X_n)$$

$$\mathrm{Cov}\left(Y - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{X},\ \underset{\sim}{e}_j^T \underset{\sim}{X}\right)$$

$$= \mathrm{cov}\left(Y, \underset{\sim}{e}_j^T \underset{\sim}{X}\right) - \mathrm{cov}\left(\hat{\underset{\sim}{\beta}}^T \underset{\sim}{X}, \underset{\sim}{e}_j^T \underset{\sim}{X}\right).$$

$$\boxed{\begin{aligned}&\mathrm{Cov}(AZ, BZ)\\ &= A\,\Sigma_Z B^T\end{aligned}}$$

$$= \left(\underset{\sim}{e}_j^T \underset{\sim}{\sigma}_{YX}\right) - \hat{\underset{\sim}{\beta}}^T \Sigma_X \underset{\sim}{e}_j$$

$$= \left(\underset{\sim}{e}_j^T \underset{\sim}{\sigma}_{YX}\right) - \left(\underset{\sim}{\sigma}_{YX}^T \Sigma_X^{-1} \Sigma_X \underset{\sim}{e}_j\right)$$

$$= \left(\underset{\sim}{e}_j^T \underset{\sim}{\sigma}_{YX}\right) - \left(\underset{\sim}{\sigma}_{YX}^T \underset{\sim}{e}_j\right) = 0$$

Consider a weakly stationary time series $\{X_t\}$ with known $E(X_t) = \mu \ \forall t$ and $\gamma_X(h)$ for lag h

We want to predict $X_{n+h} \equiv Y$ for given

$$\underset{\sim}{X} = (X_n, X_{n-1}, X_{n-2}, \cdots, X_1)$$

$$D\begin{pmatrix} Y \\ \underset{\sim}{X} \end{pmatrix} = D\begin{pmatrix} X_{n+h} \\ X_n \\ X_{n-1} \\ \vdots \\ X_1 \end{pmatrix} = \left(\begin{array}{c|cccc} \gamma(0) & \gamma(h) & \gamma(h+1) & \cdots & \gamma(n+h-1) \\ \hline \gamma(h) & & & & \\ \gamma(h+1) & & & \Gamma_n & \\ \vdots & & & & \\ \gamma(n+h-1) & & & & \end{array}\right)$$

$E(Y) = E(X_{n+h}) = \mu$

$E(\underset{\sim}{X}) = \underset{\sim}{1}\,\mu = \underset{\sim}{\mu}$

$\sigma_{YY} = \gamma(0)$

$$\underset{\sim}{\sigma}_{Yx} = \begin{pmatrix} \gamma(h) \\ \gamma(h+1) \\ \vdots \\ \gamma(h+n-1) \end{pmatrix} = \underset{\sim}{\gamma}_n(h)$$

$\Sigma_X = \Gamma_n$

$$(X_{n+h} - \mu) \triangleq \underset{\sim}{\hat{a}}^T (\underset{\sim}{X} - \mu) = \sum_{i=1}^{n} \hat{a}_i (X_{n+1-i} - \mu)$$

$$\underset{\sim}{\hat{a}}^T = (\hat{a}_1 \ \hat{a}_2 \ \cdots \ \hat{a}_n)$$

$$\underset{\sim}{X}^T = (X_n \ X_{n-1} \ \cdots X_1)$$

① $\hat{\underset{\sim}{a}} = \underline{\Gamma_n^{-1} \underset{\sim}{\gamma}_n(h)}$ became. $\hat{\underset{\sim}{\beta}} = \Sigma_x^{-1} \underset{\sim}{\sigma}_{yx}$

② Prediction error variance. $E\left((X_{n+h}) - \hat{\underset{\sim}{a}}^T(\underset{\sim}{X})\right)^2 = \gamma(0) - \underset{\sim}{\gamma}_n^T(h) \Gamma_n^{-1} \underset{\sim}{\gamma}_n(h)$

became $\sigma_{y \cdot x} = \sigma_{yy} - \underset{\sim}{\sigma}_{yx}^T \Sigma_x^{-1} \underset{\sim}{\sigma}_{yx}$.

③ Prediction error has mean zero.

$$E\left[(X_{n+h} - \mu) - \left[\hat{\underset{\sim}{a}}^T(\underset{\sim}{X} - \mu)\right]\right] = 0$$

④ Predictor variables and prediction error are uncorrelated.

$$\mathrm{Cov}\left((X_{n+h} - \mu) - \hat{\underset{\sim}{a}}^T(\underset{\sim}{X} - \underset{\sim}{\mu}),\ (X_j - \mu)\right) = 0$$

$j = 1, 2, \ldots n.$

became. $\underline{\mathrm{Cov}\left(Y - \hat{\beta}\underset{\sim}{x},\ x_j\right) = 0.}$

Ex1. Let $X_t \sim AR(1)$ process $|\phi| < 1$ and $\phi \neq 0$.

Show that: $E\left(X_{n+1} - \underset{\sim}{\hat{a}}^T \underset{\sim}{x}\right)^2 = \sigma^2 < \frac{\sigma^2}{1-\phi^2}$

Ex2 Let $X_1 = x_1$, $X_3 = x_3$ known for AR(1) process. but X_2 is missing. Find the best linear plugin value of X_2.

Ex3 Predict the value of X_{n+h} given $(X_n\, X_{n-1} \ldots X_1)$ for AR(1) sequence.

Durbin Levinson Algorithm.

Suppose we know the solution for the normal equation.

$$\Gamma_{n-1}\, \underset{\sim}{a}^{old}_{n-1} = \underset{\sim}{\gamma}^{(1)}_{n-1}$$

for $(X_n, X_{n-1} \cdots X_2)$. But now we want to use $(X_n, X_{n-1}, \cdots, X_2, X_1)$ as the predictor.

⊛ We need to estimate $\underline{a_n}$ and update the $\underline{\underset{\sim}{a}_{n-1}}$

$$\begin{pmatrix} \Gamma_{n-1} & \tilde{\underset{\sim}{\gamma}}^{(1)}_{n-1} \\ \tilde{\underset{\sim}{\gamma}}^{(1)\,T}_{n-1} & \gamma(0) \end{pmatrix} \begin{pmatrix} \underset{\sim}{a}^{new}_{n-1} \\ a_n \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\gamma}^{(1)}_{n-1} \\ \gamma(n) \end{pmatrix}$$

$$(\gamma(n-1), \gamma(n-2) \cdots\cdots, \gamma(1)) = \tilde{\underset{\sim}{\gamma}}^{(1)}_{n-1} = \underset{\sim}{\gamma}^{r\,(1)}_{n-1} = \underset{(n-1)\times(n-1)}{R}\, \underset{\sim}{\gamma}^{(1)}_{n-1}$$

$$\hat{a}_n = \frac{\gamma(n) - \left[\underset{\sim}{\gamma}^{(1)}_{n-1}\right]^T \hat{\underset{\sim}{a}}^{old}_{n-1}}{\gamma(0) - \left[\gamma^{(1)}_{n-1}\right]^T \hat{\underset{\sim}{a}}^{old}_{n-1}}$$

$$\hat{a}^{old}_{n-1} = \Gamma^{-1}_{n-1} \underset{\sim}{\gamma}^{(1)}_{n-1}$$

$$= \frac{\gamma(n) - \underline{\gamma}^{(1)}_{n-1} \Gamma^{-1}_{n-1} \underline{\gamma}^{(1)}_{n-1}}{\gamma(0) - \underline{\gamma}^{(1)}_{n-1} \Gamma^{-1}_{n-1} \underline{\gamma}^{(1)}_{n-1}}$$

PACF $\underline{\alpha(n)}$

Note $|a_n| < 1$

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}^{old}$$

$$\hat{\underset{\sim}{a}}^{new}_{n-1} = \hat{\underset{\sim}{a}}^{old}_{n-1} - \hat{a}_n \left[\hat{\underset{\sim}{a}}^{old}_{n-1}\right]^{r}$$

$$\hat{\underset{\sim}{a}}_n = \begin{pmatrix} \hat{\underset{\sim}{a}}^{new}_{n-1} \\ \cdots\cdots \\ \hat{a}_n \end{pmatrix}$$

Prediction ~~sqm.~~ error using $(n-1)$ terms.

$$E\left(X_{n+1} - \underset{\sim}{\hat{a}}_{n-1}^{old\,T} \underset{\sim}{X}_{n-1}\right)^2 = \left(\gamma(0) - \underset{\sim}{\gamma}_{n-1}^{(1)T}\, \Gamma_{n-1}^{-1}\, \underset{\sim}{\gamma}_{n-1}^{(1)}\right)$$

Prediction ~~squa~~ error using n terms

$$E\left(X_{n+1} - \underset{\sim}{a}_n^{new\,T} \underset{\sim}{X}_n\right)^2 = \gamma(0) - \underset{\sim}{\gamma}_n^{(1)T}\, \Gamma_{n-1}^{-1}\, \gamma_n^{(1)}$$

$$= \gamma(0) - \underset{\sim}{a}_n^T \underset{\sim}{\gamma}_n^{(1)}$$

$$= \left(\gamma(0) - \underset{\sim}{\gamma}_{n-1}^{(1)T}\, \Gamma_{n-1}\, \underset{\sim}{\gamma}_{n-1}^{(1)}\right)\left(1 - a_n^2\right).$$

Proportion of ~~diff~~ error reduction is

$$\frac{E\left(X_{n+1} - \underset{\sim}{\hat{a}}_n^T \underline{X}_n\right)^2}{E\left(X_{n+1} - \underset{\sim}{\hat{a}}_{n-1}^T X_{n-1}\right)^2} = \left(1 - a_n^2\right).$$

15

To predict X_{n+1} we were using $(X_n \cdots : X_2)$.

then we have added X_1

then we have added X_0

$\underset{\sim}{X}_{n+1} = (X_n \cdots\cdots X_0)$.

$$\frac{E\left(X_{n+1} - \underset{\sim}{a}^T_{n+1}\,\underset{\sim}{X}_{n+1}\right)^2}{E\left(X_{n+1} - \underset{\sim}{a}^T_{n-1}\,\underset{\sim}{X}_{n-1}\right)^2} = (1 - a_n^2)(1 - a_{n+1}^2)$$

X_1 ______ X_3 $AR(1)$ σ^2, φ known.

$$\frac{\sigma^2}{1-\varphi^2}\left(\begin{array}{c:c} A & B \\ \hline B^T & C \end{array}\right)\hat{\underset{\sim}{\beta}} = \begin{pmatrix} \varphi^{k+1} \\ \vdots \\ \varphi \\ \hdashline \varphi \\ \varphi \\ \vdots \\ \varphi^{n-k} \end{pmatrix} \frac{\sigma^2}{1-\varphi^2}.$$

$\downarrow$

X_1 ______ X_3, X_4

$$\frac{\sigma^2}{1-\varphi^2}\begin{pmatrix} 1 & \varphi^2 \\ \varphi^2 & 1 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \varphi \\ \varphi \end{pmatrix}\frac{\sigma^2}{1-\varphi^2}.$$

$$\Rightarrow \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 & \varphi^2 \\ \varphi^2 & 1 \end{pmatrix}^{-1}\begin{pmatrix} \varphi \\ \varphi \end{pmatrix} = \frac{\varphi}{1+\varphi^2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$\hat{X}_2 = \frac{\varphi}{1+\varphi^2}(X_1 + X_3). \neq \frac{X_1 + X_3}{2}.$$

In practice φ values are not known. Hence use $\hat{\varphi}$ or $\hat{\gamma}(h)$ for the estimation.