Chapter 1

Introduction

$\boxed{\textbf{Def.}}$ **1.0.1** *a Time Series (TS)* [1] X_T *is a sequence of r.v. that ordered in time T. i.e.,*

$$X_T = \{X_t, t \in T\}.$$

1.1 Stationary

$\boxed{\textbf{Def.}}$ **1.1.1** X_T *is (weakly) stationary iff*

- $\mu_i = \mu$, $\forall$ *i, for some const.* $\mu \in R$.
- $\gamma_{i,i+h} = f(|h|), \forall\ i, h$, *for some fnc.* $f : \mathbb{R} \to \mathbb{R}$. [2]

$\boxed{\textbf{Def.}}$ **1.1.2** X_T *is strictly stationary iff the joint distribution*

$$f_{X_i, X_{i+1}, \cdots, X_{i+h}} = f_{X_{i+k}, X_{i+k+1}, \cdots, X_{i+k+h}}, \forall\ i, k, h$$

Thm 1.1.1 *If* $\{X_t\}$ *is strictly stationary &* $\mu_i, \gamma_{i,j}$ *exists,* $\forall\ i, j$, *then* X_T *is weakly stationary.*

E.X. 1.1.1 (strictly stationary $\not\Rightarrow$ weakly stationary) *Cauthy distribution* $F(x|x_o, \gamma) = \frac{1}{\pi}\arctan(\frac{x-x_o}{\gamma}) + \frac{1}{2}$.

pf:

Let k = 2b, then

$$EX = \int_{-\infty}^{\infty} xf(x)dx = \lim_{a,b\to\infty} \int_{-b}^{a} xf(x)dx$$
$$= \lim_{k=a-b, b\to\infty} \int_{k}^{k+b} xf(x)dx + \underbrace{\cancel{\int_{-b}^{b} xf(x)dx}}_{0}.$$

That is, EX DNE.

Similarly, $EX^2 \propto \int_{-\infty}^{\infty} \frac{x^2}{1+x^2}dx = \int_{-\infty}^{\infty} 1dx - \int_{-\infty}^{\infty} \frac{1}{1+x^2}dx = DNE.$

By Holder Inequality, $\forall$ moments of it is DNE. ■

[1] In this chapter, T is always discrete and equally spaced.

[2] From now on, $\gamma_X(h) := f(|h|)$, and $\gamma(h)$ and $\rho(h)$ represent the Short cut of "ACVF" and "ACF" respectively.

1.2 Inference

Note:

$$\boldsymbol{Cov(x,x)} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n) \\ \gamma(-1) & \gamma(0) & \cdots & \gamma(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(-n) & \gamma(-(n-1)) & \cdots & \gamma(0) \end{bmatrix}$$

1.2.1 Mean μ

$$\begin{aligned} Var(\hat{\mu}_n) &= E(\hat{\mu}_n - \mu)^2 = \frac{1}{n^2}E(X_t-\mu)^2 = \frac{1}{n^2}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\gamma(i-j) \\ &= \frac{1}{n^2}(n\gamma(0) + (n-1)\cdot[\gamma(1)+\gamma(-1)] + (n-2)\cdot[\gamma(2)+\gamma(-2)] + \cdots) \end{aligned}$$

Thm 1.2.1 (CLT[1]) *If $\{X_t\}$ is stationary & $\sum_{j=1}^{\infty}|\gamma(j)| < \infty$, then*

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sum_{h=-n}^{n}(1-\frac{|h|}{n})\gamma(h)),$$

where $\gamma(h)$ is assumed to be known. [2]

1.2.2 ACVF $\gamma(h)$ & ACF $\rho(h)$

$\hat{\boldsymbol{\rho}}_k := (\hat{\rho}(1), \hat{\rho}(2), \cdots, \hat{\rho}(k))^T$.

Thm 1.2.2 *If $\{X_t\}$ is stationary & $\sum_{j=0}^{\infty}\gamma(j) < \infty$, then $\hat{\boldsymbol{\rho}}_k \xrightarrow{d} \mathcal{N}(\boldsymbol{\rho}_k, \frac{1}{n}\boldsymbol{W}_k)$, where $\boldsymbol{W}_k$ is a $k \times k$ matrix, s.t.,*

$$w_{ij} = \sum_{k=1}^{\infty}[\rho(k+i)+\rho(k-i)-2\rho(i)\rho(k)]\cdot[\rho(k+j)+\rho(k-j)-2\rho(j)\rho(k)],$$

thus, $CI_{95\%} = \hat{\rho}(h) \pm \frac{1.96}{\sqrt{n}}\sqrt{w_{hh}}$.

[2] If $\gamma(h)$ is unknown, then

$$CI_{95\%} = \hat{\mu}_n \pm \frac{1.96}{\sqrt{n}}\sqrt{\hat{V}}, \quad w.l \quad \hat{V} = \sum_{h=-\sqrt{n}}^{\sqrt{n}}(1-\frac{|h|}{n})\hat{\gamma}(h).$$

Chapter 2

Practical Operation

2.1 Hypothesis tests

i.i,d ?

Target:

- *before*: If true, the history is useless for forecasting.
- *after*: If false, $\hat{\epsilon}$ is not good enough.

1. **Non-rigorous test**

   XXX

   * *dis-adv*: Multiple testing problem [1].

2. **Ljung Box test**[2]

   $H_0 : (i.i.d \Rightarrow \rho = 0 \Rightarrow \rho^2 = 0)$.

$$\hat{Q}_{LB} = n \sum_{j=1}^{h} \hat{\rho}^2(j) \sim \chi_h^2.$$

$$modified - \hat{Q}_{LB} = n(n+2) \sum_{j=1}^{h} \frac{\hat{\rho}^2(j)}{n-j} \sim \chi_h^2.$$

   * *dis-adv*:
   When h is large, the # of Q is too small;
   When h is small, the lag h is not large enough.

   Rule of thumb: h $= 2\lfloor \ln n \rfloor$.

3. **McLeod & Li test**[3]

   $H_0 : (i.i.d \Rightarrow \rho = 0 \Rightarrow \rho^2 = 0)$.

$$\hat{Q}_{ML} = n(n+2) \sum_{j=1}^{h} \frac{\hat{\rho}_{\boldsymbol{x}^2}^2(j)}{n-j} \sim \chi_h^2.$$

[1] t-test v.s F-test.

[2] modified Q_{LB} is better when $n \leq 100$, same when n is large.

[3] The ACF of x_t is not large enough, thus making it to be x^2.

Normal ?

Target:
XXX

1. **Rough test**: Q-Q plot.[1]

2.2 Remove the trend

2.2.1 Remove the trend only

1. * Linear regression $X_t = \beta_0 + \beta_1 t + \epsilon_t$.

   $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{X}_t \perp\!\!\!\perp i$.

2. * Homonic regression $X_t = \sum_{j=1}^{k}[a_j(\cos\lambda_j t) + b_j(\sin\lambda_j t)] + \epsilon_t$.

   - λ_j are chosen manually as the potential frequency $\frac{2\pi}{T}$ (usually).
   - K is large enough.

   $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{X}_t \perp\!\!\!\perp i$.

3. * Exponential Smoothing $X_t = \mu_t + \epsilon_t$.

   $$\widehat{\mu}_t{}^{2} = \sum_{j=0}^{t-2}[\alpha(1-\alpha)^j X_{t-j}] + (1-\alpha)^{t-1}X_1.$$

   $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{\mu}_t$.

4. Moving average $X_t = \mu_t + \epsilon_t$.

   $$\widehat{\mu}_t = \frac{1}{2q+1}\sum_{j=t-q}^{t+q} X_j{}^{3},$$

   where q is manually[4] picked, usually $\lfloor n^{\frac{1}{3}} \rfloor$. $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{\mu}_t$.

5. Differencing[5] $X_t = \mu_t + s_t + \epsilon_t$.

   Def. 2.2.1 *Backward shift operator B is a operator s.t., B: $X_T \to X_T$, w.l. $x_t \to x_t$.*

   $\Rightarrow \Delta X_t{}^{6} = (I - B)X_t$.

[1] for STA457, this is enough.
[2] Esitmation of the trend. Weighted average of $\{X_t, X_{t-1}, \dots, X_1\}$, where ω decaying exponentially & decaying speed $\Uparrow$ as $\alpha \Uparrow$.
[3] if $t-q<0$ or $t+q>n+1$, then let $x_j = x_1$, $x_j = x_n$ respectively.
[4] Non-trivial.
[5] dis: i) # data $\Downarrow$; ii) loss the trend information.
[6] Assume that ΔX_t has no trend. Sometimes it still has, then do it again.

2.2.2 Remove trend & Seasonality

1. Estimation & Removal $X_t = \mu_t + s_t$(Seasonality) + ϵ_t.

$$\widehat{trend}\begin{cases}\widehat{\mu}_t = \frac{1}{2q+1}\sum_{j=t-q}^{t+q} x_j, \quad d = 2q+1(odd)\\ \\ \widehat{\mu}_t = \frac{1}{2q}[0.5x_{t-q} + \sum_{j=t-q}^{t+q} x_j + 0.5x_{t+q}], \quad d = 2q(even)\end{cases}$$

$\Rightarrow \widehat{X_t} = X_t - \widehat{\mu}_t \rightarrow$ Only seasonal component left.

$\widehat{seasonal}$: $\forall\ k \in \mathbb{N}, 0 \le k \le d$, $A_k := \{\widehat{X}_{k+d\cdot j} | j \in \mathbb{N}, k + d \cdot j \le n\} = \{\widehat{x}_k, \widehat{x}_{k+d}, \widehat{x}_{k+2d} \cdots\}$. $\widehat{s}_t :=$ average of obs. in A_k.

Thus, $\forall\ i < d$, $\widehat{s}_t = \widehat{s}_{i-md}$, for some $m \in \mathbb{N}$.

$\Rightarrow \widehat{\epsilon}_t = \widehat{X}_t - \widehat{s}_t$.

2. Differencing $X_t = \mu_t + s_t + \epsilon_t$.

$$\begin{aligned} Y_t &= (I - B_X^d)X_t = X_t - X_{t-d} = (\mu - \mu_{t-d} + \cancelto{0}{(s_t - s_{t-d})} + (\epsilon_t + \epsilon_{t-d}) \\ &= \widetilde{\mu}_t + \widetilde{\epsilon}_t. \end{aligned}$$

$\Rightarrow \Delta Y_t = (I - B_Y)Y_t$.

2.3 ACVF

$\boxed{\textbf{Def.}}$ **2.3.1 (sample ACVF)** *sample auto-covariance function*

$$\widehat{\gamma}_X(h) := \frac{1}{n}\sum_{i=t}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X}), \quad w.l \quad |h| < n.$$

$\boxed{\textbf{Def.}}$ **2.3.2 (sample ACF)** *sample auto-correlation function*

$$\widehat{\rho}_X(h) := \frac{\widehat{\gamma}_X(h)}{\widehat{\gamma}_X(0)}$$

Lemma 2.3.1 *If $\gamma_{t,t+h} = \gamma_X(h), \forall\ t \in T$, then $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$.*
Notice that $\{X_t\}$ do not need to be weakly stationary.

$\boxed{\textbf{Def.}}$ **2.3.3** *$f : \mathbb{Z} \to \mathbb{R}$ is non-negative definite iff*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i f(i-j) a_j \ge 0, \forall\ n \in \mathbb{N}, \boldsymbol{a} \in \mathbb{R}^n.$$

prop. (Basic property of $\gamma(h)$)

1. $|\gamma(h)| \le \gamma(0), \forall\ h$.
2. $\gamma(h) = \gamma(-h), \forall\ h$.

3. $\gamma(h)$ is non-negative definite.

   pf:

   $var(\sum a_i X_i) \geq 0 \Rightarrow \sum_i \sum_j a_i Cov(X_i, X_j) a_j \geq 0.$ ■

Thm 2.3.1 *$\gamma : \mathbb{Z} \to \mathbb{R}$ is ACVF for some X_T iff symmetric & non-negative definite.*

$\boxed{\textbf{Def.}}$ **2.3.4 (sample PACF)** *sample partial auto-correlation function* [1]

$$\widehat{\alpha}(h) := \phi_{hh},$$

where ϕ_{hh} is given by Durbin-Levinson Algorithm 4.1.4.

prop.

1. $\alpha(0) = 1$ and $\alpha(1) = \rho(1)$. ***pf:***

   By definition, $\alpha(1) = \Gamma_1^{-1}\gamma_1 = \frac{1}{\gamma(0)}\gamma(1) = \rho(1)$. ■

2. $|\alpha(h)| \leq 1$, $\forall\ h$.

   pf:

   Recall from Durbin-Levinson Algorithm 4.1.4 that $v(h)$ is the MSE of the best linear forecast, then

$$v(h) = v(h-1)(1-\phi_{hh}^2) \Rightarrow (1-\phi_{hh}^2) \geq 0.$$

■

[1]"Partial" comes from the statistical terminology "partial correlation". XXX

Chapter 3

Processes

$\boxed{\textbf{Def.}}$ **3.0.1** *$\{W_t\}$ is i.i.d noise iff $\{W_t\}$ is i.i.d & $E(W_t) = 0, \forall\ t$.*

$\boxed{\textbf{Def.}}$ **3.0.2** *$\{Z_t\}$ is White noise iff $E(Z_t) = 0$ & $Cov(Z_t, Z_s) = \begin{cases} \sigma^2 & , t = s \\ 0 & , t \neq s \end{cases}$.*

noted as $WN(0, \sigma^2)$.

prop.

1. independent $\Rightarrow \gamma_\epsilon(h) = \begin{cases} 0 & , t \neq s \\ \sigma_i^2 & , t = s \end{cases}$.
2. $i.i.d \Rightarrow$ independent + Homoscedasticity $\Rightarrow WN(0, \sigma^2)$.

   E.X. 3.0.1 ($WN(0, \sigma^2) \nRightarrow$ i.i.d.) *$Y_t = Z_t Z_{t-1}$, where $Z_t \sim N(0, \sigma^2)$.*

$\boxed{\textbf{Def.}}$ **3.0.3** *$\{X_t\}$ is called random walk [1]. iff $X_t = \sum_{j=1}^{t} W_j$, w.l. $\{W_j\}$ be i.i.d noise.*

$\boxed{\textbf{Def.}}$ **3.0.4** *$\{X_t\}$ is called simple random walk iff $(W_j + \frac{1}{2}) \overset{i.i.d}{\sim} bern(\frac{1}{2})$.*

prop.

$Var(X_t) = i\sigma^2$, i.e., not stationary.

3.1 Linear process

$\boxed{\textbf{Def.}}$ **3.1.1** *$\{X_t\}$ is linear process iff*

- *$X_t = \sum_{j=-\infty}^{\infty} a_j Z_j \quad w.l \quad Z_j \sim WN(0, \sigma^2)$.*
- *$\sum_{j=-\infty}^{\infty} |a_j| < \infty$, i.e., converge.*

That is, X_t is a L.C. of $\{Z_j\}$. Notice that a_j could be 0.

prop.

[1] origin: Brownian motion (布朗运动).

Let $\{Y_t\}$ is stationary, and $\{X_t\}$ is a L.C. of $\{Y_t\}$, and $\sum_{j=-\infty}^{\infty}|a_j|<\infty$. Then,

$$\gamma_X(h)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}a_j a_k \gamma_Y(h+k-j).$$

If $Y_t \sim WN(0,\sigma^2)$, then $\gamma_X(h)=\sigma^2\sum_{j=-\infty}^{\infty}a_j a_{j+h}$.

3.1.1 MA(q) process.

Def. **3.1.2** *$\{X_t\}$ is MA(1) process w.l coefficient a, σ^2 iff $X_t=Z_t+aZ_{t-1}$ w.l. $Z_t\sim WN(0,\sigma^2)$.*

prop.

$$\gamma(h)=\begin{cases}(1+a^2)\cdot\sigma^2 & ,|h|=0\\ a\cdot\sigma^2 & ,|h|=1\\ 0 & ,|h|>2\end{cases}$$

Def. **3.1.3** *$\{X_t\}$ is MA(q) process iff*

$$X_t=\sum_{j=t-q}^{t}a_jZ_j \quad w.l \quad Z_j\sim WN(0,\sigma^2).$$

Def. **3.1.4** *$\{X_t\}$ is q-dependent iff $\forall\ k>q$, $X_t \perp\!\!\!\perp X_{t+k}$.*

prop.

$\{e_t\}$ is q-dependent $\Rightarrow$ for any fnc. $G:\mathbb{R}^q\to\mathbb{R}^q$, s.t., $X_t=G(e_t,\cdots,e_{t+q})$, $\{X_t\}$ is q-dependent.

Def. **3.1.5** *$\{X_t\}$ is q-correlated iff $\forall\ k>q$, $\rho_{X_t,X_{t+k}}=0$.*

Thm 3.1.1 *$\{X_t\}$ is exactly a MA(q) process $\Leftrightarrow$ $\{X_t\}$ is q-correlated & stationary.*

pf:

($\Leftarrow$) Trivial, 略.
($\Rightarrow$) Too non-trivial, 略. ■

Note that in practice, we usually plot the sample ACF and check if it cuts off at some p to test whether MA(p) process is a good candidate for $\{X_t\}$.
XXX

Def. **3.1.6** *$\{X_t\}$ is MA(∞) process w.l coefficient $\boldsymbol{a}$, σ^2 iff $X_t=\sum_{j=-\infty}^{t}a_jZ_j$ w.l. $Z_t\sim WN(0,\sigma^2)$ and $\sum_{j=-\infty}^{\infty}|a_j|<\infty$.*

3.1.2 AR(p) process

Def. **3.1.7** *$\{X_t\}$ is a AR(1) process w.l coefficient a, σ^2 iff $X_t = Z_t + aX_{t-1}, \quad Z_t \sim WN(0,\sigma^2)$.*

prop.

1. $\gamma(h) = \frac{a^{|h|}}{1-a^2} \cdot \sigma^2$.
2. $Cov(X_t, Z_{t+1}) = 0, \forall\ t.$ [1]

Def. **3.1.8** *$\{X_t\}$ is causal process iff $X_t = \sum_{j=-\infty}^{t} a_j Z_j$, w.l. $Z_t \sim WN(0,\sigma^2)$.*
Notice that X_t is a L.C. of historical pts. only.

Def. **3.1.9** *$\{X_t\}$ is invertable process iff Z_t can be written as a L.C. of X_s, where $s \le t$.*

Lemma 3.1.1 *AR(1) is casual $\Leftrightarrow$ AR(1) is MA(∞) $\Leftrightarrow |a| \le 1 \Leftrightarrow \sum |a_j| < \infty$.*

pf:

$X_t = \frac{1}{1-aB} Z_t = \sum_{j=-\infty}^{t} a^{t-j} Z_j$ is a kind of MA(∞), and MA(∞) is causal. ■

Thm 3.1.2 *If $\{X_t\}$ is a AR(p) process, then the PACF $\alpha(h) = 0,\ \forall\ |h| \ge p+1$.*

Note that in practice, we usually plot the sample PACF and check if it cuts off at some p to test whether AP(p) process is a good candidate for $\{X_t\}$.
XXX

3.1.3 ARMA(p,q) process

Def. **3.1.10** *$\{X_t\}$ is ARMA(1,1) process iff $(1-aB)X_t = (1+\theta B)Z_t$, w.l. $\theta + a \neq 0$ [2] and $Z_t \sim WN(0,\sigma^2)$.*

prop.

|a|>1, $\exists$ a stationary non-causal solution to $(1-aB)X_t = (1+\theta B)Z_t$.
XXX. (Unfinished)

Lemma 3.1.2 *|a|< 1 $\Rightarrow$ ARMA(1,1) is MA(∞).*

pf:

SInce |a| < 1, $X_t = \frac{1+\theta B}{1-aB} Z_t = (1 + \frac{a+\theta}{1-aB}) Z_t = Z_t + \sum_{j=1}^{\infty} (a+\theta) a^{j-1} Z_j$. [3] ■

Lemma 3.1.3 *$|\theta| < 1 \Rightarrow$ ARMA(1,1) is invertable.*

pf:

[1] Because X_t is a L.C. of $Z_t, Z_{t-1}, \cdots, Z_2, Z_1$.
[2] if $a = -\theta$, then $X_t = Z_t$, $WN(0,\sigma^2)$ は興味がありません！
[3] It is incorrect to write $(1 + \frac{a+\theta}{1-aB})$, but it is fine here.

$Z_t = \frac{1-aB}{1+\theta B} X_t = \cdots B^j X_t.$ ■

$\boxed{\textbf{Def.}}$ **3.1.11** *$\{X_t\}$ is* ARMA(p,q) process *iff $\phi(B)X_t = \theta(B)Z_t$, w.l. $\phi(\cdot), \theta(\cdot)$ be some polynomial function and $Z_t \sim WN(0, \sigma^2)$.*

Thm 3.1.3 *There exists a stationary solution $\{x_t\}$ for an ARMA(p,q) process $\Leftrightarrow \forall\ x_0 \in \mathbb{C}$ as the roots of $\phi(x) = 1 - \phi_1 x - \cdots \phi_p x^p = 0$, $||x_0|| \neq 1$.*

pf:

Given $\phi(B)$ and , $X_t = \frac{\theta(B)}{\phi(B)} Z_t$ if and only if $\forall$ the roots of $\Phi(B)$, $||B_0|| \neq 1$. (???) wtf is that
■

Thm 3.1.4 *ARMA(p,q) process is causal $\Leftrightarrow \forall\ x_0 \in \mathbb{C}$ as the roots of $\phi(x)$, $||x_0|| > 1$.*

pf:

Given p-th degree of polynomial fnc. $\phi(x)$, w.l. $x \in \mathbb{C}$, we have

$$\phi(x) = (1 - \alpha_1 x)(1 - \alpha_2 x) \cdots (1 - \alpha_p x).$$

Thus, the i^{th} root $x^i = \frac{1}{\alpha_i}$. ($\Leftarrow$) If $||x_i|| > 1$, then $||\alpha_i|| < 1$, $\forall\ i$. Therefore,

$$\frac{1}{\phi(B)} = \prod_{i=1}^{p} \frac{1}{1 - \alpha_i B} = \sum_{i=1}^{p} \frac{c_i}{1 - \alpha_i B} = \sum_{i=1}^{p} c_i \sum_{j=1}^{\infty} (\alpha_i B)^j.$$

■

Thm 3.1.5 *ARMA(p,q) process is invertable $\Leftrightarrow \forall\ x_0 \in \mathbb{C}$ as the roots of $\theta(x)$, $||x_0|| > 1$.*

E.X. 3.1.1 $X_t - 2X_{t-1} + 3X_{t-2} = Z_t$.
Clearly it is invertable, and $\theta(x) = 1 = 0 \Rightarrow x_0 \in \emptyset$.

Chapter 4

Linear Forecast

4.1 Stationary Linear Forecast

$\boxed{\textbf{Def.}}$ **4.1.1** *MSE of forecast is defined as* $E[X_{t+h}-f(X_h)]^2$*, where* $f(X_h)$ *is the forecast of* X_{t+h}.

Thm 4.1.1 $E(X_{t+h}|X_t)$ *minimize MSE.*

pf:

Let $f^*(X_t) = E(X_{t+h}|X_t)$, then

$$\begin{aligned} E(X_{t+h}-f)^2 &= E(X_{t+h}-f^*+f^*-f)^2 \\ &= E(X_{t+h}-f)^2 + 2E(X_{t+h}-f^*)(f^*-f) + E(f^*-f)^2. \end{aligned}$$

Recall that $E(Y) = E[E(Y|X)]$, then

$$\begin{aligned} 2E(X_{t+h}-f^*)(f^*-f) &= 2EE[(X_{t+h}-f^*)(f^*-f)|X_t] \\ &= E[(f^*-f)(E(X_{t+h}|X_t)-f^*)] = 0 \end{aligned}$$

■

Thm 4.1.2 *The best linear forecast of* Y *based on a stationary TS* $\{X_t\}_{t=1}^n$ *with known* μ *and* $\gamma(\cdot)$ *is*

$$P(Y|\boldsymbol{x}) = a_0 + \boldsymbol{a}\cdot\boldsymbol{x},$$

where $a_0 = \mu - \boldsymbol{a}\cdot\boldsymbol{x}$ *and* $\boldsymbol{\gamma} = \boldsymbol{\Gamma}\boldsymbol{a}$.

pf:

Criterion:

Let $\boldsymbol{x} = (X_1,\cdots,X_n)^T$, $E(Y) = \mu$ and $E(X_t) = \mu_t$ for notation convenience. Then

$$\min\{\text{MSE of forecast}\} = \min_{(a_0,\boldsymbol{a})} E(Y-a_0-a_1X_1-\cdots-a_nX_n)^2 = \min_{(a_0,\boldsymbol{a})} E(Y-a_0-\boldsymbol{a}\cdot\boldsymbol{x})^2.$$

and view $X_0 = 1$, we have

$$\frac{\partial\mathcal{L}}{\partial a_t} = -2E(Y-a_0-\boldsymbol{a}\cdot\boldsymbol{x})X_t = 0, \forall\ t \quad\Rightarrow\quad E(YX_t) = E(a_0+\boldsymbol{a}\cdot\boldsymbol{x})X_t, \forall\ t.$$

Notice that when t = 0, define $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$, we obtain

$$\mu = a_0 + \boldsymbol{a} \cdot \boldsymbol{\mu} \quad \Leftrightarrow \quad a_0 = \mu - \sum_{j=1}^{n} a_j \mu_j$$

Thus,

$$E(YX_t) = (\mu - \boldsymbol{a} \cdot \boldsymbol{\mu})\mu_t + E(\boldsymbol{a} \cdot \boldsymbol{x})X_t \quad \Rightarrow \quad E(YX_t) - \mu\mu_t = E(\boldsymbol{a} \cdot \boldsymbol{x})X_t - (\boldsymbol{a} \cdot \boldsymbol{x})\mu_t.$$

That is, $Cov(Y, X_t) = Cov(\boldsymbol{x}, X_t), \forall\ t$, or,

$$Cov(Y, \boldsymbol{x}) = \boldsymbol{Cov}(\boldsymbol{x}, \boldsymbol{x})\boldsymbol{a} \Leftrightarrow \gamma = \boldsymbol{\Gamma}\boldsymbol{a}.$$

If $\boldsymbol{\Gamma}$ is not singular, i.e., $det(\boldsymbol{\Gamma}) \neq 0$, then $\boldsymbol{a} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}$ is not only solvable, but also unique. ■

Thm 4.1.3 *The MSE of forecast when predicting Y from $\{X_t\}_{t=1}^n$ is $Var(Y^*) - \boldsymbol{a} \cdot \gamma$, where $Y^* = Y - \mu$.*

pf:

Let $Y^* = Y - \mu$. Then

$$\begin{aligned} \text{MSE of forecast} &= E(Y - \mu + \boldsymbol{a} - \boldsymbol{a} \cdot \boldsymbol{x})^2 \\ &= E(Y - \mu)^2 - 2\boldsymbol{a} \cdot E(\boldsymbol{x})Y + E(\boldsymbol{a} \cdot \boldsymbol{x})^2 \\ &= Var(Y^*) - 2\boldsymbol{a} \cdot \gamma + \boldsymbol{a} \cdot \boldsymbol{\Gamma}\boldsymbol{a} = Var(Y^*) - \boldsymbol{a} \cdot \gamma. \end{aligned}$$

■

prop.

1. $E(Y - P(Y|\boldsymbol{x}) = 0$, i.e., expected forecast error is 0.
2. $E(Y - P(Y|\boldsymbol{x})\boldsymbol{x} = \boldsymbol{0}$, i.e., forecast error is uncorrelated w.l predictors.
3. (Even if $\boldsymbol{\Gamma}$ is singular) $P(Y|\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$ is still unique.

   pf:

   Assume that we have 2 best linear forecasters Q_1 and Q_2, s.t.,

   $$\begin{aligned} Q_1 &= a_0^1 + \boldsymbol{a}_n^1 \cdot \boldsymbol{x} \\ Q_2 &= a_0^2 + \boldsymbol{a}_n^2 \cdot \boldsymbol{x} \end{aligned}$$

   Then, by **prop.** 2, we have

   $$\begin{cases} E(Y - Q_1)\boldsymbol{x} = \boldsymbol{0} \\ E(Y - Q_2)\boldsymbol{x} = \boldsymbol{0} \end{cases} \Rightarrow E(Q_1 - Q_2)\boldsymbol{x} = \boldsymbol{0}.$$

   Therefore,

   $$E(Q_1 - Q_2)^2 = (a_0^1 - a_0^2)\cancelto{0}{E(Q_1 - Q_2)} + \boldsymbol{a} \cdot \cancelto{\boldsymbol{0}}{E(Q_1 - Q_2)\boldsymbol{x}} = 0.$$

   That is, $Q_1 - Q_2 = 0$, w.l $Pr = 1$. ■
4. $P(*|\boldsymbol{x}) : \Omega \to \Omega$ is a linear operator [1] over field $\mathbb{R}$, s.t.,

   $$Cov(Y, \boldsymbol{x}) = \boldsymbol{0} \Rightarrow P(Y|\boldsymbol{x}) = \mu \quad \& \quad P(X_t|\boldsymbol{x}) = \mu_t, \forall\ t.$$
5. (*Tower Law of Predictor*) $P(Y|\boldsymbol{u}_n) = P(P(Y|\boldsymbol{u}_n, \boldsymbol{v}_n)|\boldsymbol{u}_n)$, where $\boldsymbol{v}$ is r.v. s.t., $Cov(\boldsymbol{u}_n, \boldsymbol{v}_n) < \infty$. ***pf:***

   XXX ■

[1] Cov is a linear operator.

4.1.1 Recursive Forecasting Algorithms

Thm 4.1.4 (Durbin-Levinson Algorithm) *Let* $P_n(\boldsymbol{x}_{n+1}) = \boldsymbol{a}_n^T \boldsymbol{x}_n$ *w.l.* $\boldsymbol{\Gamma} \boldsymbol{a}_n = \boldsymbol{\gamma}_n$, *where* $\boldsymbol{\Gamma}_{n\times n}$ *is non-singular* [1] *and* $v_n = MSE_n = \gamma(0) - \boldsymbol{a}_n^T \boldsymbol{\gamma}_n$.

If $\{X_t\}_{t=1}^n$ *is stationary, then for any* $n \in \mathbb{N}$, $P_n(\boldsymbol{x}_{n+1})$ *can be computed by*

$$a_{n,n} = v_{n-1}^{-1}[\gamma(n) - \sum_{j=1}^{n-1} a_{n-1,j}\gamma(n-j)]$$

$$\begin{bmatrix} a_{n,1} \\ \vdots \\ a_{n,n-1} \end{bmatrix} = \begin{bmatrix} a_{n-1,1} \\ \vdots \\ a_{n-1,n-1} \end{bmatrix} - a_{n,n} \begin{bmatrix} a_{n-1,n-1} \\ \vdots \\ a_{n-1,1} \end{bmatrix}$$

$$v_n = v_{n-1}(1 - a_{n,n}^2)$$

where $a_{1,1} = \frac{\gamma(1)}{\gamma(0)}$ *and* $v_0 = \gamma(0)$.

Thm 4.1.5 (The Innovation Algorithm) *Let* $u_n = x_n - P_{n-1}(x_n)$ *be defined as the forecast error, which is usually referred as "innovation". Let* $v_n = MSE_n = \gamma(0) - \boldsymbol{a}_n^T \boldsymbol{\gamma}_n$.

If $\{x\}_{t=1}^n$, *not necessarily stationary, has* $EX_t = 0$, *then* $P_n(\boldsymbol{x}_{n+1}) = \begin{cases} 0, & \text{if } n = 1 \\ \sum_{j=1}^n \theta_{n,j} u_{n-j}, & \text{if } n \geq 1 \end{cases}$, *s.t.,*

$$\theta_{n,n-k} = v_k^{-1}[\gamma_{n+1,k+1} - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j], \quad 0 \leq k < n$$

$$v_n = \gamma_{n+1,n+1} - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j$$

where $v_0 = \gamma_{1,1}$.

Lemma 4.1.1 *If* $\{x_t\}$ *is AR(p) process, then* $a_{n,j} = 0$, $\forall\ j \geq p+1$. *That is,, only* $a_{n1}, \cdots, a_{np} \neq 0$.

pf:

Note that the rigorous proof is not even required for Grads. Thus, we will only prove why $P_n(\boldsymbol{x}_{n+1}) = \begin{cases} 0, & \text{if } n = 1 \\ \sum_{j=1}^n \theta_{n,j} u_{n-j}, & \text{if } n \geq 1 \end{cases}$. Given

$$\boldsymbol{u}_n = \boldsymbol{A}\boldsymbol{x}_n \quad w.l. \quad \boldsymbol{A}_{n\times n} = \begin{bmatrix} 1 & & & \\ a_{11} & 1 & & \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & 1 \end{bmatrix},$$

we know that $det(\boldsymbol{A}) = 1$, s.t.,

$$\boldsymbol{A}_{n\times n}^{-1} = \begin{bmatrix} 1 & & & \\ \theta_{11} & 1 & & \\ \vdots & \vdots & \ddots & \\ \theta_{n1} & \theta_{n2} & \cdots & 1 \end{bmatrix},$$

[1] In practical, we barely have the case where $\boldsymbol{\Gamma}_{n\times n}$ is singular, thus it is a mild assumption.

Thus,

$$\boldsymbol{P}_n = x_n - u_n = (\boldsymbol{A}^{-1} - \boldsymbol{I})\boldsymbol{u}_n = \begin{bmatrix} 0 & & & \\ \theta_{11} & 0 & & \\ \vdots & \vdots & \ddots & \\ \theta_{n1} & \theta_{n2} & \cdots & 0 \end{bmatrix} \boldsymbol{u}_n.$$

■

Lemma 4.1.2 *If $\{x_t\}$ is MA(q) process, then $\theta_{n,j} = 0$, $\forall\ j \geq q+1$. That is,, only $\theta_{n1}, \cdots, \theta_{nq} \neq 0$.*

4.1.2 The World Decomposition

Def. 4.1.2 $\{X_t\}$ *is deterministic iff X_t can be written as a L.C. of $\{X_{t-1}, X_{t-2}, \cdots\}$, $\forall\ t$.*

E.X. 4.1.1 $X_t = A\cos\omega t + \sin\omega t = (2\cos\omega)X_{t-1} - X_{t-2}$, $\forall\ t$. *Notice that X_t is even not a r.v..*

Thm 4.1.6 (World Decomposition) [1] *If $\{X_t\}$ is non-deterministic & stationary, then*

$$X_t = \sum_{j=0}^{\infty} \phi_j Z_{t-j} + V_t,$$

where

1. $\phi_0 = 1$, $\sum \phi_j^2 < \infty$;
2. $Z_t \sim WN(0, \sigma^2)$ *and* $Cov(Z_t, V_s) = 0$, $\forall\ t, s$;
3. $\{V_t\}$ *is a deterministic TS.*

Note that for most process in practice, $V_t = 0$, $\forall\ t$. I.e., $\{x_t\}$ is purely non-deterministic process, s.t.,

$$X_t = \sum_{j=0}^{\infty} \phi_j Z_{t-j}.$$

4.2 Forecast ARMA(p,q) process

Given $\Phi(B)X_t = \Theta(B)Z_t$, $\mathbb{Z} \sim WN(0, \sigma^2)$, where $\{\phi_1, \cdots \phi_p\}$, $\{\theta_1, \cdots \theta_q\}$ and σ^2 are known.

Define

$$W_t = \begin{cases} \sigma^{-1}X_t, & \text{if } t \leq m \\ \sigma^{-1}\Phi(B)X_t, & \text{if } t > m \end{cases},$$

where $m = \max\{p, q\}$. Then W_t is a MA(q) process when $t > m$. Then we have

$$\gamma_W(i,j) = \begin{cases} \sigma^{-2}\gamma_X(|i-j|), & \text{if } \max\{i,j\} \leq m \\ \sigma^{-2}\sum_{r=1}^{q} \theta_r \theta_{r+|i-j|}, & \text{if } i > m\ \&\ j > m \\ \sigma^{-2}[\gamma_X(|i-j|) - \sum_{r=1}^{q} \gamma_X(r + |i-j|)], & \text{if } \min\{i,j\} < m < \max\{i,j\} \leq 2m \\ 0, & \text{otherwise} \end{cases}.$$

[1] Operate theory needed to prove this theorem. Thus not required for STA 457.

Thus, we can use $\gamma_W(i,j)$ to forecast W_{n+1} via $\{W_t\}_{t=1}^n$ using the innovation process. Then

$$P_n(W_{n+1}) = \begin{cases} \sum_{j=1}^n \theta_{nj}(W_{n+1-j} - P_n(W_{n+1+j}), & \text{if } 1 \leq n < m \\ \sum_{j=1}^q \theta_{nj}(W_{n+1-j} - P_n(W_{n+1+j}), & \text{if } n \geq m \end{cases}$$

XXX(?) Notice that when $n \geq m$, there is a significant reduction of computation time since we only sum to q.

Thus, given

$$P_n(W_{n+1}) = \begin{cases} \sigma^{-1} P_n(X_{n+1}), & \text{if } t \leq m \\ \sigma^{-1} P_n[\Phi(B) X_t], & \text{if } t > m \end{cases}$$

we have

$$W_{n+1} - P_n(W_{n+1}) = \sigma^{-1}[X_{n+1} - P_n(W_{n+1})], \quad \forall\ t$$

???

Chapter 5

Modeling and Forecasting

5.1 Modeling and Forecasting with ARMA process

Set up:

Preliminary estimation of the parameters $\phi = (\phi_1, \phi_2, \cdots, \phi_p)^T$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_q)^T$ and σ^2 from n observations $\{x_1, x_2, \cdots, x_n\}$ of the causal ARMA(p,q) process defined by

$$\phi(B)X_t = \Theta(B)Z_t \quad w.l \quad Z_t \sim WN(0, \sigma^2)$$

where p and q are assumed to be known.

5.1.1 Yule-Walker Algorithm

* May related to Durbin-Levinson algorithm

Def. **5.1.1** *The fitted Yule-Walker AR(p) model is*

$$X_t - \hat{\phi}_{p1}X_{t-1} - \cdots - \hat{\phi}_{pp}X_{t-p} = Z_t, \quad w.l. \quad Z_t \sim WN(0, \hat{v}_p).$$

Thm 5.1.1 (Sample Yule-Walker Equation) *If $\{X_t\}$ is a AR(p) process for some unknown $\{\phi_1, \cdots, \phi_p\}$ and σ^2, then the Yule-Walker equation will estimate them as*

$$\hat{\phi}_p = \widehat{\boldsymbol{R}}_p^{-1}\hat{\boldsymbol{\rho}}_p$$
$$\hat{\sigma}_p^2 = \hat{\gamma}(0)[1 - \hat{\boldsymbol{\rho}}_p^T\widehat{\boldsymbol{R}}_p^{-1}\hat{\boldsymbol{\rho}}_p]$$

where $\widehat{\boldsymbol{R}}_p = \hat{\gamma}(0)^{-1}\widehat{\boldsymbol{\Gamma}}_p$ represents the sample ACF matrix.

pf:

Given a casual AR(p) process $X_t - \Phi(B)X_t = Z_t$, where $Z \sim WN(0, \sigma^2)$, we have

$$X_{t-i}X_t - \sum_{j=1}^{p}\phi_j X_{t-i}X_{t-j} = Z_t X_{t-i}, \quad \forall\ i \in \{1, 2, \cdots, p\}.$$

for any t. Therefore,

$$
\begin{aligned}
& E[X_{t-i}X_t]-\sum_{j=1}^{p}\phi_j E[X_{t-i}X_{t-j}]=E[Z_tX_{t-i}], \quad \forall\, i\in\{1,2,\cdots,p\} \\
\Rightarrow\quad & \gamma(i)-\sum_{j=1}^{p}\phi_j\gamma(i-j)=0, \quad \forall\, i\in\{1,2,\cdots,p\} \\
\Rightarrow\quad & \mathbf{\Gamma}_p\phi_p=\gamma_p.
\end{aligned}
$$

If $\mathbf{\Gamma}_p$ is non-singular, then $\phi_p=\mathbf{\Gamma}_p^{-1}\gamma_p$, which is called Yule-Walker Equation (Method of Moment). ■

Thm 5.1.2 *If $\{X_t\}$ is an AR(p) process, and $\hat{\phi}_p$ is estimated by Yule-Walker equation, then*

$$
\hat{\phi}_p \xrightarrow{d} \mathcal{N}(\phi, \frac{1}{n}\sigma^2\mathbf{\Gamma}_p^{-1}).
$$

pf:

Since $\bar{X}\longrightarrow EX=0$, we have $\hat{\gamma}(i)\approx\frac{1}{n}\sum_{i=1}^{n-|j|}X_iX_{i+|j|}$, $\forall\, j$. Therefore, since $X_{i-j}\perp\!\!\!\perp Z_i$,

$$
\hat{\gamma}_p\approx\gamma_p+\frac{1}{n}\sum_{i=1}^{n}Z_i\begin{bmatrix}X_{i-1}\\ \vdots\\ X_{i-p}\end{bmatrix}, \qquad Z_i\begin{bmatrix}X_{i-1}\\ \vdots\\ X_{i-p}\end{bmatrix}\sim \boldsymbol{WN}(\mathbf{0},\sigma^2\mathbf{\Gamma}_p).
$$

Note that $\hat{\phi}_p=\widehat{\mathbf{\Gamma}}_p^{-1}\hat{\gamma}_p\approx\mathbf{\Gamma}_p^{-1}\hat{\gamma}_p$. Therefore,

$$
\hat{\phi}_p\approx\phi_p+\frac{1}{n}\mathbf{\Gamma}_p^{-1}\sum_{i=1}^{n}Z_i\begin{bmatrix}X_{i-1}\\ \vdots\\ X_{i-p}\end{bmatrix}.
$$

Therefore, given the symmetric Γ_p, by CLT, we have $\sqrt{n}(\hat{\phi}_p-\phi_p)\xrightarrow{d}\mathcal{N}(0,\sigma^2\Gamma_p^{-1}\Gamma_p\Gamma_p^{-1})$. Thus, we have [1]

$$
CI_{95\%}=\{\phi\in\mathbb{R}^p|(\hat{\phi}-\phi)^T\widehat{\mathbf{\Gamma}}_p(\hat{\phi}-\phi)\leq\frac{1}{n}\sigma^2\chi^2_{95\%,p}\}
$$

■

Lemma 5.1.1 *Let $\Phi_{0.95}$ denotes the 0.95 quantile of the standard normal distribution and $\hat{v}_{jj}$ denotes the jth diagnoal element of $\hat{v}_p\hat{\Gamma}_p^{-1}$, then for large n,*

$$
CI_{95\%}=\hat{\phi}_{pj}\pm\Phi_{0.975}(\frac{1}{n}\hat{v}_{jj})^{\frac{1}{2}},
$$

contains ϕ_j with 95% confidence level.

[1] If $\boldsymbol{Y}\sim\mathcal{N}(0,\boldsymbol{A})$, then $\boldsymbol{DY}\sim\mathcal{N}(0,\boldsymbol{DAD}^T)$.

5.1.2 Innovations Algorithm

Def. **5.1.2** *The fitted innovations MA(q) model is*

$$X_t = Z_t + \theta_{q1} Z_{t-1} + \cdots + \theta_{qq} Z_{t-q}, \quad w.l. \quad Z_t \sim WN(0, \hat{v}_q),$$

where $\hat{\boldsymbol{\theta}}_q$ and $\hat{v}_q$ are obtained from the innovation algorithm with ACVF replaced by sample ACVF.

Thm 5.1.3 *If $\{X_t\}$ is an MA(q) process, and $\boldsymbol{\theta}_q$ is estimated by Innovations algorithm, then*

$$\boldsymbol{\theta}_p \xrightarrow{d} \mathcal{N}(\boldsymbol{\theta}, \frac{1}{n} \boldsymbol{A}_{q\times q}),$$

where $a_{ij} = \sum_{r=1}^{min(i,j)} \theta_{i-r}\theta_{j-r}$.

Lemma 5.1.2 *In particular,*

$$\sqrt{n}(\hat{\theta}_{qj} - \theta_j) \sim N(0, \sum_{r=1}^{j} \theta_{j-r}^2).$$

In practice,

$$CI_{95\%} = \hat{\theta}_{pj} + \frac{1.96}{\sqrt{n}} \sqrt{\sum_{r=1}^{j} \hat{\theta}_{j-r}^2}).$$

5.1.3 Hannan-Rissanen Algorithm

5.1.4 Maximum Likelihood Algorithm

Denote

5.2 Problem Sets

1. Let $\{Z_t\}$ be a sequence of independent normal r.v. with $E(Z_t) = 0$ and $Var(Z_t) = \sigma^2$, $\forall\ t \in T$. Let a, b, c be constants. Then is the following processes
$$X_t = Z_t \cos ct + Z_{t-1} \sin ct$$
stationary? Specify the mean and auto-variance function if stationary.

2. Let $\{x_1, \cdots, x_n\}$ be observed values of a TS at times $1, \cdots, n$, and let $\widehat{\rho}(h)$ be the sample ACF at lag h. If $x_t = c\cos(\omega t)$, where c and ω are constants s.t., $c \neq 0$ and $\omega \in [-\pi, \pi]$. Then show that
$$\widehat{\rho}(h) \xrightarrow{p} cos(\omega h).$$

3. Consider the AR(1) process $X_t = 0.4X_{t-1} + Z_t$, where $Z_t \sim WN(0, \sigma^2)$. Define the sub-sequences v_k recursively as
$$v_1 = 1, \quad v_{k+1} = v_k + (P_k + 1), \quad \text{for } k \in \{1, 2, \cdots\},$$
where $\{P_t\} \overset{i.i.d}{\sim} Poisson(1)$. Further assume that $\{P_t\}$ and $\{X_t\}$ are independent. Define $Y_k = X_{v_k}$, $k \in \{1, 2, \cdots\}$. Is $\{Z_k\}$ a weakly stationary TS? Prove your conjecture.

4. Suppose that we have the following non-stationary TS model as
$$Y_t = 2\frac{t}{n} + X_t, \quad \text{for } t \in \{1, 2, , \cdots, n\},$$
where $\{X_t\}$ is a AR(1) process s.t., $X_t = 0.5X_{t-1} + Z_t$, where $\{Z_t\}$ are i.i.d. standard normal r.v..
   (a) Calculate $Cov(Y_1, Y_2)$ and $Corr(Y_{n/2}, Y_{n/2+1})$. Are they equal?
   (b) Calculate the first order sample ACF $\widehat{\rho}(1)$ for $\{Z_t\}$ by assuming that n goes to infinity. IS t the same as the first order ACF of $\{X_t\}$?
   (c) If we want to remove the non-stationary trend $2\frac{t}{n}$ from the TS, we can run the following linear regression
$$Y_t = a + b\frac{t}{n} + e_t.$$

Find the CLT of $\widehat{b}$. I.e., find σ^2 s.t.,

$$\sqrt{n}(\widehat{b}-2) \longrightarrow \mathcal{N}(0,\sigma^2).$$

5. Suppose that we have the following non-stationary TS model as

$$Y_t = \omega_1 \frac{t}{n} + (\omega_2 + \frac{t}{n})X_t, \quad \text{for } t \in \{1,2,,\cdots,n\},$$

where $\{X_t\}$ is a AR(1) process s.t., $X_t = 0.5X_{t-1} + Z_t$, where $\{Z_t\}$ are i.i.d. standard normal r.v..

   (a) Let $Z_t = Y_t - Y_{t-1}$. Show that $\{Z_t\}$ is not weakly stationary.

   (b) Given unknown ω_1, ω_2, find a way to transform Y_t into a stationary TS.

6. Suppose that

7. (a) Consider the following ARMA(1,1) process

$$X_t - 0.5X_{t-1} = Z_t + 3Z_{t_1}, \quad \text{where } Z_t \sim WN(0,\sigma^2).$$

   Is the process X_t causal and/or invertible? [1]

   (b) Find the MA(∞) representation of the process $\{X_t\}$ in (a).

   (c) Find the first two ACF's $\rho(1)$ and $\rho(2)$ for X_t defined in (a).

8. Suppose that $\boldsymbol{W}$ is a random vector and Y is a random variable. Suppose that $Var(Y) < \infty$ and matrix $\boldsymbol{\Gamma} = Cov(\boldsymbol{W},\boldsymbol{W})$ is finite. Let $P(Y|\boldsymbol{W})$ be the best linear forecast of Y based on $\boldsymbol{W}$. Denote $\gamma = Cov(Y,\boldsymbol{W})$. Prove that

$$E[Y - P(Y|\boldsymbol{W})]^2 \le Var(Y),$$

and show that $E[Y - P(Y|\boldsymbol{W})]^2 = Var(Y)$ if and only if $\gamma = 0$.

9. Suppose the historical data support that the monthly return of a security (in percentage) follows a stationary AR(1) model

$$X_t = 1 + 0.5X_t + Z_t.$$

Suppose that $X_1 = 2$ and $X_4 = 3$ and the values of X_2 and X_3 are missing. Based on the values of X_1 and X_4, find the best linear guess of $(X_2+X_3)/2$ and find the mean squared error of the guess.

[1] You may refer to Lemma 3.1.1 and 3.1.3.