SF2943 Time Series Analysis: Lecture 13

May 25, 2022

In this lecture, we will review some topics which I consider the most essential in the whole course. This does not mean that the topics not covered here are not important.

1 Time Series Models

In this course, we focus on building a time series model of the following form:

$$X_t = m_t + s_t + Y_t,$$

where m_t is called the trend component, s_t is the seasonal component, and Y_t is a "stationary" noise. To define stationarity, we have to first introduce the definitions of the mean function and the covariance function.

Definition 1.1 *Let $\{X_t\}$ be a time series with $EX_t^2 < \infty$*[1]*. The mean function of $\{X_t\}$ is*

$$\mu_X(t) = EX_t \textit{ for all } t.$$

The covariance function of $\{X_t\}$ is

$$\gamma_X(s,t) = Cov(X_s, X_t) \textit{ for all } s, t.$$

Definition 1.2 *A time series $\{X_t\}$ is (weakly) stationary if*

- *$\mu_X(t)$ is invariant with time.*
- *For any given $h \geq 0$, $\gamma_X(t+h,t)$ is also invariant with time.* [2]

Since for a stationary time series $\{X_t\}$, the covariance function $\gamma_X(t+h,t)$ is invariant with time. Abuse of notation, we denote $\gamma_X(t+h,t)$ by $\gamma_X(h)$ in this case, and call $\gamma_X(h)$ the autocovariance function of $\{X_t\}$ at lag h.

[1]Otherwise, the covariance might be ∞.
[2]Here we implicitly assume that $EX_t^2 < \infty$ for all t.

Definition 1.3 *Let $\{X_t\}$ be a stationary time series. The autocovariance function (acvf) of $\{X_t\}$ at lag h is defined as*

$$\gamma_X(h) = Cov(X_{t+h}, X_t).$$

In particular, $\gamma_X(0) = Cov(X_t, X_t) = Var(X_t) = EX_t^2$. Moreover, the autocorrelation function (acf) of $\{X_t\}$ at lag h is defined as

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

Proposition 1.4 *Let $\{X_t\}$ be a stationary time series with acvf γ. Then γ satisfies*

- $\gamma(0) \geq 0$.
- $|\gamma(h)| \leq \gamma(0)$ *for all h.*
- γ *is a even function, i.e., $\gamma(h) = \gamma(-h)$ for all h.*

Another fundamental property of acvf is nonnegative definite.

Definition 1.5 *A function $\kappa : \mathbb{Z} \to \mathbb{R}$ is nonnegative definite (also known as positive semi-definite) if for any $n \in \mathbb{N}$, and for any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, we have*

$$\sum_{i,j=1}^{n} a_i \kappa(i-j) a_j \geq 0.$$

Remark 1.6 *Equivalently, we can say a function $\kappa : \mathbb{Z} \to \mathbb{R}$ is nonnegative definite if for any $n \in \mathbb{N}$, the associated n-by-n matrix*

$$K_n = [\kappa(i-j)]_{i,j=1}^n = \begin{pmatrix} \kappa(0) & \kappa(1) & \cdots & \kappa(n-1) \\ \kappa(-1) & \kappa(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(-n+1) & \cdots & \cdots & \kappa(0) \end{pmatrix}$$

is a nonnegative definite matrix (in the sense defined in linear algebra). Namely, for any $\mathbf{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$,

$$\mathbf{a}' K_n \mathbf{a} = \sum_{i,j=1}^{n} a_i (K_n)_{ij} a_j \geq 0.$$

This is equivalent to saying that every eigenvalue of K_n is nonnegative.

Theorem 1.7 (Characterization of Autocovariance Functions) *A function $\gamma : \mathbb{Z} \to \mathbb{R}$ is the autocovariance function of a stationary time series if and only if it is even and nonnegative definite.*

Definition 1.8 *The partial autocorrelation function (pacf) $\alpha(\cdot)$ of a stationary time series $\{X_t\}$ is defined by $\alpha(0) = 1$ and*

$$\alpha(h) = Corr(X_h - P(X_h | X_1, \ldots, X_{h-1}), X_0 - P(X_0 | X_1, \ldots, X_{h-1})).$$

An equivalent way to define the pacf α is as follows.

Definition 1.9 *The partial autocorrelation function (pacf) $\alpha(\cdot)$ of a stationary time series $\{X_t\}$ is defined by $\alpha(0) = 1$ and $\alpha(h) = \phi_{hh}$ which is the last component of $\boldsymbol{\phi}_h = (\phi_{h1}, \phi_{h2}, \dots, \phi_{hh})$ satisfying*

$$\Gamma_h \boldsymbol{\phi}_h = \boldsymbol{\gamma}_h \tag{1.1}$$

with $\Gamma_h = (\gamma(i-j))_{i,j=1}^h$ and $\boldsymbol{\gamma}_h = (\gamma(1), \gamma(2), \dots, \gamma(h))'$.

1.1 Stationary Time Series Models

Definition 1.10 (White noise) *A time series $\{X_t\}_t$ is a white noise (with mean 0 and variance σ^2) if*

- $EX_t = 0$,
- *X_s, X_t are uncorrelated for any s, t such that $s \neq t$, i.e., $Cov(X_s, X_t) = 0$ for $s \neq t$.*
- *$EX_t^2 = \sigma^2$ for all t.*

We use the notation $\{X_t\}_t \sim WN(0, \sigma^2)$.

1.1.1 Moving-average processes

Definition 1.11 *A time series $\{X_t\}$ is a moving-average process of order q (aka MA(q) process or q-th order moving average process) for some $q \in \mathbb{N}$ if*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta_1, \dots, \theta_q \in \mathbb{R}$.

Definition 1.12 *A time series $\{X_t\}$ is q-correlated for some $q \in \mathbb{N} \cup \{0\}$ if X_t and X_s are uncorrelated when $|s-t| > q$.*

Remark 1.13 *For a stationary time series $\{X_t\}$, it is q-correlated for some $q \in \mathbb{N} \cup \{0\}$ if its acvf γ_X satisfying $\gamma_X(h) = 0$ for all $|h| > q$.*

Proposition 1.14 *If $\{X_t\}$ is a stationary and q-correlated time series with mean 0, then it can be represented as a MA(q) process, i.e., there exists some $\{Z_t\} \sim WN(0, \sigma^2)$ such taht*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

for all t.

Example 1.1 (First-order moving average or MA(1) process) *Consider a time series $\{X_t\}$ defiend by*

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \ldots$$

where $Z_t \sim WN(0, \sigma^2)$ and θ is some real constant. Its acvf is

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\sigma^2, & \text{if } h = 0 \\ \theta\sigma^2, & \text{if } h = 1, -1 \\ 0, & \text{otherwise} \end{cases}$$

and acf is

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0 \\ \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2}, & \text{if } h = 1, -1 \\ 0, & \text{otherwise.} \end{cases}$$

1.1.2 Autoregressive processes

Definition 1.15 *A time series $\{X_t\}$ is an autoregressive process of order p (aka AR(p) process or p-th autoregressive process) for some $p \in \mathbb{N}$ if*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q \in \mathbb{R}$.

Example 1.2 (First-order autoregression or AR(1) process) *Assume that $\{X_t\}$ is a stationary time series satisfying*

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \ldots$$

where $Z_t \sim WN(0, \sigma^2)$, $|\phi| < 1$. Its acvf is

$$\gamma_X(h) = \begin{cases} \frac{\sigma^2}{1-\phi^2}, & \text{if } h = 0 \\ \phi^{|h|}\frac{\sigma^2}{1-\phi^2}, & \text{otherwise,} \end{cases}$$

and the acf is

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0 \\ \phi^{|h|}, & \text{otherwise.} \end{cases}$$

1.1.3 ARMA(p, q) process

Definition 1.16 *A time series $\{X_t\}$ is an ARMA(p, q) process if $\{X_t\}$ is stationary and*

$$\underbrace{X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}}_{\text{p-th order autoregressive part}} = \underbrace{Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}}_{\text{q-th order moving average part}} \text{ for all } t, \tag{1.2}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and the polynomials

$$\text{AR polynomial: } \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

and

$$\text{MA polynomial: } \theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$$

have no common factors.

Remark 1.17 *If we use the backward shift operator notation B, then (1.2) is equivalent to*

$$\phi(B)X_t = \theta(B)Z_t \qquad \text{for all } t \tag{1.3}$$

with $\phi(z)$ is the AR polynomial and $\theta(z)$ is the MA polynomial.

Definition 1.18 *We say that a time series $\{X_t\}$ is a causal (or future-independent) function of $\{Z_t\}$ if X_t has a representation in terms of $\{Z_s, s \leq t\}$. If $\{X_t\}$ is not causal, we call it noncausal.*

Definition 1.19 *We say that a time series $\{X_t\}$ is a invertible function of $\{Z_t\}$ if Z_t has a representation in terms of $\{X_s, s \leq t\}$. If $\{X_t\}$ is not invertible, we call it noninvertible.*

Properties of an ARMA(p, q) process:

1. (Existence & Uniqueness) If $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| = 1$, then there exists an unique stationary process satisfying (1.3).

2. If $\phi(z) = 0$ for some $z \in \mathbb{C}$ with $|z| = 1$, then there does not exist any stationary process satisfying (1.3).

3. (Causality) The ARMA(p, q) process $\{X_t\}$ is causal if and only if $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.

4. (Invertibility) The ARMA(p, q) process $\{X_t\}$ is invertible if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.

An important class of time series models is called linear processes. This class of models contains causal ARMA(p, q) processes.

Definition 1.20 *A time series $\{X_t\}$ is a linear process if it has the representation*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \text{ for all } t \in \mathbb{Z},$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\{\psi_j\} \subset \mathbb{R}$ with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Proposition 1.21 *Let $\{Y_t\}$ be a stationary time series with mean function $\mu_Y \equiv 0$ and acvf γ_Y. If $\{\psi_j\} \subset \mathbb{R}$ satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series $\{X_t\}$ defined as*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}$$

is stationary with mean function $\mu_X \equiv 0$ and acvf

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

Corollary 1.22 *Every linear process $\{X_t\}$ is stationary with mean function $\mu_x(h) = 0$ for all h and acvf*

$$\gamma_X(h) = \sum_{\ell=-\infty}^{\infty} \psi_\ell \psi_{\ell+h} \sigma^2.$$

Corollary 1.23 *For a causal ARMA(p, q) process $\{X_t\}$, its acvf is*

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|},$$

where

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}.$$

1.1.4 Spectral density

Definition 1.24 *Let $\{X_t\}$ be a zero-mean stationary process with acvf γ satisfying absolute summability, i.e.,*

$$\sum_{h=\infty}^{\infty} |\gamma(h)| < \infty.$$

The spectral density of $\{X_t\}$ is defined by

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h), \qquad \text{for all } \lambda \in \mathbb{R}. \tag{1.4}$$

Lemma 1.25 (Basic Properties of Spectral Density)

1. *f is even, i.e., $f(\lambda) = f(-\lambda)$.*
2. *$f(\lambda) \geq 0$, for all $\lambda \in [-\pi, \pi]$.*
3. *$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(k\lambda) f(\lambda) d\lambda$.*

Definition 1.26 *A function $f : [-\pi, \pi] \to \mathbb{R}$ is the spectral density of a stationary time series $\{X_t\}$ with acvf γ if*

- *$f(\lambda) \geq 0$ for all $\lambda \in [-\pi, \pi]$ and*
- *$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$ for all $h \in \mathbb{Z}$.*

The following proposition characterizes spectral densities.

Proposition 1.27 *A real-valued function f defined on $[-\pi, \pi]$ is the spectral density of a real-valued stationary process if and only if*

1. *$f(\lambda) = f(-\lambda)$.*

2. $f(\lambda) \geq 0$ *for all* $f(\lambda)$.

3. $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$.

Corollary 1.28 *An absolutely summable* $\gamma : \mathbb{Z} \to \mathbb{R}$ *is the acvf of a stationary process if and only if it is even and*

$$f(\lambda) \doteq \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) \geq 0 \text{ for all } \lambda \in [-\pi, \pi],$$

i.e., the spectral density f *is nonnegative everywhere.*

Proposition 1.29 *Let* $\{X_t\}$ *be a stationary time series with mean zero and spectral density* $f_X(\lambda)$*. Suppose* $\{\psi_j\}$ *is absolutely summable, and thus forms a linear filter. Then the time series*

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

is stationary with mean zero and spectral density

$$f_Y(\lambda) = |\psi(e^{-i\lambda})|^2 f_X(\lambda)$$

where

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j.$$

(The function $\psi(e^{-i\lambda})$ *is called the transfer function of the filter, and* $|\psi(e^{-i\lambda})|^2$ *is called the power transfer function of the filter.)*

Example 1.3 (White noise) *Consider* $\{Z_t\} \sim WN(0, \sigma^2)$*. The spectral density is*

$$f_Z(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma_Z(h) = e^{-i \cdot 0 \cdot \lambda} \gamma_Z(0) = \frac{\sigma^2}{2\pi}.$$

Proposition 1.30 *Let* $\{X_t\}$ *be an ARMA*(p, q) *process, i.e.,*

$$\phi(B) X_t = \theta(B) Z_t$$

with $\{Z_t\} \sim WN(0, \sigma^2)$*. The spectral density of* $\{X_t\}$ *is*

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}.$$

(This spectral density is also known as a rational spectral density.)

1.2 Financial Time Seires Models

Definition 1.31 *A process $\{Z_t\}$ is an autoregressive conditional heteroscedasticity process of order p, denoted as ARCH(p), for $p \in \mathbb{N}$ if it is a stationary process and satisfying*

$$\begin{cases} Z_t = \sqrt{h_t} e_t, & \text{with } \{e_t\} \sim IID\ N(0,1) \\ h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 \end{cases}, \tag{1.5}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, \ldots, p$. Here h_t is the conditional variance, i.e., $h_t = Var(Z_t | Z_s, s < t)$, and it is also known as the volatility.

Solution of the ARCH(1) Equations:
If $\alpha_1 < 1$, the unique causal stationary solution of the ARCH(1) equations (1.5) is given by

$$Z_t = e_t \sqrt{\alpha_0 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j e_{t-1}^2 \cdots e_{t-j}^2\right)}.$$

In addition, $EZ_t = 0$,

$$\mathrm{Var}(Z_t) = \frac{\alpha_0}{1 - \alpha_1} \text{ and } E[Z_{t+h} Z_t] = 0 \text{ for } h \neq 0.$$

Definition 1.32 *A process $\{Z_t\}$ is a generalized autoregressive conditional heteroscedasticity process of order (p, q), denoted as GARCH(p, q), for $p, q \in \mathbb{N}$ if it is a stationary process and satisfying*

$$\begin{cases} Z_t = \sqrt{h_t} e_t, & \text{with } \{e_t\} \sim IID\ (0,1) \\ h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \end{cases}, \tag{1.6}$$

where $\alpha_0 > 0$ and $\alpha_i, \beta_j \geq 0$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$.

1.3 Nonstationary Time Series Models

Definition 1.33 *Let $p, d, q \in \mathbb{N} \cup \{0\}$. A time series $\{X_t\}$ is an autoregressive integreted moving average process of order (p, d, q) (or an ARIMA(p, d, q) process) if*

$$Y_t \doteq (I - B)^d X_t$$

is a causal ARMA(p, q) process, where B is the backward shift operator, i.e., $BX_t = X_t - X_{t-1}$.

The definition means that $\{Y_t\}$ satisfies

$$\phi(B) Y_t = \theta(B) Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

where $\phi(z)$ and $\theta(z)$ are polynomials of degrees p and q, respectively, and $\phi(z) \neq 0$ for $|z| \leq 1$ (due to the causality). Because $Y_t \doteq (I - B)^d X_t$, we find that $\{X_t\}$ satisfies

$$\phi^*(B) X_t = \theta(B) Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2), \tag{1.7}$$

where $\phi^*(z) \doteq \phi(z)(1 - z)^d$.

Proposition 1.34 *If $\{X_t\}$ is ARIMA(p,d,q) with $d \geq 1$, and consider a process $\{\tilde{X}_t\}$ satisfying*

$$\tilde{X}_t = X_t + g(t),$$

where $g(t)$ is a polynomial in t of degree $d-1$, then $\{\tilde{X}_t\}$ is also an ARIMA process with the same order (p,d,q). Moreover, if we define $Y_t = (I-B)^d X_t$ and $\tilde{Y}_t = (I-B)^d \tilde{X}_t$, then $Y_t = \tilde{Y}_t$.

Definition 1.35 *Consider $d, D \in \mathbb{N} \cup \{0\}$. The time series $\{X_t\}$ is a seasonal ARIMA process of order $(p,d,q) \times (P,D,Q)_s$, also known as SARIMA$(p,d,q) \times (P,D,Q)_s$, if the differenced series*

$$Y_t \doteq (I-B)^d (I-B^s)^D X_t$$

is a causal ARMA process defined by

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad \{Z_t\} \sim WN(0,\sigma^2), \tag{1.8}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, and $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$.

2 Parameters Estimation

Definition 2.1 *Let $x_1, x_2, \ldots, x_n$ be observations of a time series. The sample mean of $x_1, x_2, \ldots, x_n$ is*

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The sample autocovariance function (sample acvf) is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \ \text{for } -h < n < h.$$

The sample autocorrelation function (sample acf) is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \ \text{for } -h < n < h.$$

Proposition 2.2 *If $\{X_t\}$ is a stationary time series with mean μ and acvf γ, then as $n \to \infty$,*

$$Var(\bar{X}_n) \to 0 \quad \text{if} \quad \gamma(n) \to 0$$

and

$$n\,Var(\bar{X}_n) \to \sum_{h=-\infty}^{\infty}\gamma(h) \quad \text{if} \quad \sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty,$$

where

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Theorem 2.3 *If $\{X_t\}$ is a stationary process satisfying*

$$X_t = \mu + \sum_{j=\infty}^{\infty} \psi_j Z_{t-j}$$

where $\{Z_j\} \sim IID(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, then

$$\frac{1}{\sqrt{n}}(\bar{X}_n - \mu) \xrightarrow{d} N(0, v), \text{ as } n \to \infty$$

with

$$v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right).$$

Theorem 2.4 *If $\{X_t\}$ is a stationary process satisfying*

$$X_t = \mu + \sum_{j=\infty}^{\infty} \psi_j Z_{t-j}$$

where $\{Z_j\} \sim IID(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and either $EZ_t^4 < \infty$ or $\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty$, then for each $h = 1, 2, \ldots,$

$$\frac{1}{\sqrt{n}}(\hat{\boldsymbol{\rho}}(h) - \boldsymbol{\rho}(h)) \xrightarrow{d} N(0, W), \text{ as } n \to \infty$$

with $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(h))'$, $\hat{\boldsymbol{\rho}} = (\hat{\rho}(1), \ldots, \hat{\rho}(h))'$ and $W = (w_{ij})_{i,j=1}^{h}$ given by Bartlett's formula

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

Yule-Walker Equations: Consider fitting a causal AR(p) model (with a given p)

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t.$$

By multiplying both sides of the equation with X_{t-j}, $j = 0, 1, \ldots, p$ and taking expectation, we find

$$\begin{cases} \Gamma_p \boldsymbol{\phi} = \gamma_p \\ \sigma^2 = \gamma(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p \end{cases}, \tag{2.1}$$

where $\Gamma_p = (\gamma(i-j))_{i,j=1}^{p}$, $\boldsymbol{\gamma}_p = (\gamma(1), \ldots, \gamma(p))'$, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$. The equations (2.1) are called Yule-Walker equations.

Sample Yule-Walker Equations: Estimate $\boldsymbol{\phi}$ and σ^2 for an AR(p) model by

$$\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)' = \hat{R}_p^{-1} \hat{\boldsymbol{\rho}}_p$$

and

$$\hat{\sigma}^2 = \hat{\gamma}(0)[1 - \hat{\boldsymbol{\phi}}' \hat{\boldsymbol{\rho}}_p] = \hat{\gamma}(0) \left[1 - \hat{\boldsymbol{\rho}}_p' \hat{R}_p^{-1} \hat{\boldsymbol{\rho}}_p \right],$$

where $\hat{R}_p = (\hat{\rho}(i-j))_{i,j=1}^{p}$, $\hat{\boldsymbol{\rho}}_p = (\hat{\rho}(1), \ldots, \hat{\rho}(p))' = \hat{\boldsymbol{\gamma}}_p / \hat{\gamma}(0)$.

3 Forecasting

Definition 3.1 *The best linear predictor (in the mean squared error) for X_{n+h} given $\{X_1, \cdots, X_n\}$ is $a_0 + a_1 X_n + \cdots + a_n X_1$ where $(a_0, a_1, \ldots, a_n)$ satisfies*

$$\min_{a_0, a_1, \ldots, a_n \in \mathbb{R}} E\left[(X_{n+h} - (a_0 + a_1 X_n + \cdots + a_n X_1))^2 \right].$$

Proposition 3.2 *For a time series $\{X_t\}$ with mean 0, the best linear predictor of X_{n+h} given $X_1, \ldots, X_n$, denoted by $P_n X_{n+h}$, is*

$$P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1,$$

with $a_0 = 0$ and $\mathbf{a}_n = (a_1, \ldots, a_n)'$ solves

$$\Gamma_n \mathbf{a}_n = \gamma_n(h), \tag{3.1}$$

where $\mathbf{a}_n = (a_1, \ldots, a_n)'$, $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$, and $\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$.

Properties of $P_n X_{n+h}$:

1. $P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$, where $\mathbf{a}_n = (a_1, \ldots, a_n)'$ satisfies (3.1).
2. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \mathbf{a}_n' \gamma_n(h)$, where $\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$.
3. $E(X_{n+h} - P_n X_{n+h}) = 0$. Namely, $P_n X_{n+h}$ is an unbiased estimator of X_{n+h}.
4. $E[(X_{n+h} - P_n X_{n+h}) X_j] = 0$ for $j = 1, \ldots, n$.

Let Y and $W_1, \ldots, W_n$ be random variables with finite second moments and assume that the means $\mu = EY$, $\mu_i = EW_i$ and covariances $\mathrm{Cov}(Y, Y)$, $\mathrm{Cov}(Y, W_i)$, and $\mathrm{Cov}(W_i, W_j)$ for any $i, j \in \{1, \ldots, n\}$ are all known. Denote the random vector $\mathbf{W} = (W_n, \ldots, W_1)'$, the vector of means $\boldsymbol{\mu}_W = (\mu_n, \ldots, \mu_1)'$, the vector of covariances

$$\gamma = \mathrm{Cov}(Y, \mathbf{W}) \doteq (\mathrm{Cov}(Y, W_i), \mathrm{Cov}(Y, W_i), \ldots, \mathrm{Cov}(Y, W_i))',$$

and the covariance matrix

$$\Gamma = \mathrm{Cov}(\mathbf{W}, \mathbf{W}) \doteq [\mathrm{Cov}(W_i, W_j)]_{i,j=1}^n.$$

Then by the same arguments used in the calculation of $P_n X_{n+h}$, the best linear predictor of Y in terms of $\{1, W_n, \ldots, W_1\}$ is

$$P(Y|\mathbf{W}) = \mu_Y + \mathbf{a}'(\mathbf{W} - \boldsymbol{\mu}_W),$$

where $\mathbf{a} = (a_1, \ldots, a_n)'$ is any solution of

$$\Gamma \mathbf{a} = \gamma.$$

Properties of the Prediction Operator $P(\cdot|\mathbf{W})$:
Suppose $EU^2 < \infty$, $EV^2 < \infty$, $\Gamma = \mathrm{Cov}(\mathbf{W}, \mathbf{W})$, and $\beta, \alpha_1, \ldots, \alpha_n$ are constant.

1. $P(U|\mathbf{W}) = EU + \mathbf{a}'(\mathbf{W} - E\mathbf{W})$, where $\Gamma\mathbf{a} = \text{Cov}(U, \mathbf{W})$.
2. $E[(U - P(U|\mathbf{W}))\mathbf{W}] = \mathbf{0}$ and $E[U - P(U|\mathbf{W})] = 0$.
3. $E[(U - P(U|\mathbf{W}))^2] = \text{Var}(U) - \mathbf{a}'\text{Cov}(U, \mathbf{W})$.
4. $P(\alpha_1 U + \alpha_2 V + \beta|\mathbf{W}) = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta$.
5. $P(U|\mathbf{W}) = EU$ if $\text{Cov}(U, \mathbf{W}) = \mathbf{0}$.
6. If $U = g(\mathbf{W})$ for some function $g : \mathbb{R}^n \to \mathbb{R}$, then $P(U|\mathbf{W}) = U$.

Example 3.1 (One-Step Prediction of an AR(p) Series) *Consider an AR(p) process $\{X_t\}$, i.e., X_t satisfies*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t.$$

The best linear predictor is

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}.$$