Time Series Analysis

Mohamed Shakir

January 7, 2024

Contents

1 Introduction **2**
- 1.1 Examples of Time Series Models . . . 2
- 1.2 Data Pre-processing and Eliminating Trend . . . 2
- 1.3 Stationary Models and the Autocorrelation Function . . . 3
  - 1.3.1 Properties of Covariances . . . 3
  - 1.3.2 Properties of Autocorrelation and Autocovariance Functions . . . 3
- 1.4 Partial Autocorrelation Function . . . 5

2 Sample Mean, Sample Autocorrelation, Sample PACF **7**

3 Autoregression Moving-Average (ARMA) Models **8**
- 3.1 Linear Processes . . . 8
- 3.2 ARMA Models . . . 9
  - 3.2.1 Difference and Backward Operators . . . 9
  - 3.2.2 Solutions to ARMA Models . . . 10
- 3.3 Linear Representation for ARMA Models . . . 11
- 3.4 Autocovariance Function for ARMA Model . . . 12
  - 3.4.1 Partial Autocorrelation for ARMA Models . . . 14

4 Forecasting of Stationary Time Serires **15**
- 4.1 Yule-Walker Procedure . . . 15
  - 4.1.1 Mean Square Prediction Error (MSPE) . . . 16
- 4.2 Durbin-Levinson Algorithm . . . 17
  - 4.2.1 Partial Autocovariance Functions . . . 18
- 4.3 Forecast Limits, Prediction Intervals . . . 18

5 Estimation in ARMA Models **19**
- 5.1 Estimation of the Mean . . . 19
- 5.2 Yuke-Walker Estimator . . . 20

6 Diagnostic Measures **22**

7 Maximum Likelihood Estimation **23**
- 7.1 Maximum Likelihood Estimation for Time Series Models . . . 23

Chapter 1

Introduction

A **time series** is a set of observations x_t, each one being recorded at a specific time t. A *discrete-time time series* is one in which the set T_0 of times at which observations are made is a discrete set. *Continuous-time time series* are obtained when observations are recorded continuously over some time interval.

Definition 1.0.1. *A* time series model *for the observed data $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$, of which $\{x_t\}$ is postulated to be a realization.*

1.1 Examples of Time Series Models

A typical structure of a time series is

$$X_t = m_t + Y_t + S_t$$

where m_t is a trend, S_t is a seasonal part, and Y_t is a stationary part.

Example 1.1.1 (White Noise)**.** The simplest model for a time series is one with no trend or seasonal component, the observations are independent and identically distributed (iid) random variables with zero mean. Let $\{Z_t\}$ be a sequence of iid random variables with mean zero and variance 1. This sequence is called **white noise**.

Example 1.1.2 (Random Walk)**.** The random walk $\{S_t, t = 0, 1, 2, \ldots\}$ is obtained by cumulatively summing iid random variables. Thus a random walk with zero mean is obtained by defining $S_0 = 0$, and

$$S_t = X_1 + X_2 + \cdots + X_t,\ t = 1, 2, \ldots$$

Let $\{Z_t\}$ be a sequence of iid random variables with mena zero and variance σ_Z^2. We define $X_t = \sum_{i=1}^{t} Z_i$. Then $\{X_t\}$ is a random walk with zero mean.

1.2 Data Pre-processing and Eliminating Trend

In order to analyse the time series $X_t = m_t + Y_t + S_t$, we need to eliminate the trend and seasonal component, leaving us with the *stationary part* Y_t which will be defined below. The methods for eliminating trends are as follows:

- **Differencing:** For the time series $\{X_t,\ t = 1, \ldots, n\}$, calculate

$$\nabla X_t = X_t - X_{t-1},\ t = 2, \ldots, n$$

- **Polynomial Fitting:** If the trend follows some polynomial, say $m_t = a + bt$, we can estimate the parameters a, b using least squares, i.e minimize

$$\sum_{t=1}^{n} (X_t - a - bt)^2$$

  This becomes a simple regression problem with the independent variable being the time t, and the dependent variable is the time series X_t. So we obtain an estimate for the trend

$$\hat{m}_t = \hat{a} + \hat{b}t$$

  and the detrended series is

$$\hat{Y}_t = X_t - \hat{m}_t$$

- **Exponential Smoothing:** Let $\alpha \in (0,1)$. Then the trend is estimated by

$$\hat{m}_1 = X_1,\ \hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1},\ t = 2, \ldots, n$$

- **Moving Average Smoothing:** Let q be a positive integer. Then

$$\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t+j},\ q+1 \leq t \leq n-q$$

The detrended time series becomes

$$\hat{Y}_t = X_t - \hat{m}_t,\ t = q+1, \ldots, n-q$$

Recall that our time series X_t is compromised of a trend m_t and stationary part Y_t, $X_t = m_t + Y_t$. We are assuming the mean $E(Y_t) = 0$, so

$$(2q+1)^{-1} \sum_{j=-q}^{q} X_{t+j} = (2q+1)^{-1} \sum_{j=-q}^{q} m_{t+j} + (2q+1)^{-1} \sum_{j=-q}^{q} Y_{t+j}$$

We assume that the mean is 0, so this simplifies to the above expression.

1.3 Stationary Models and the Autocorrelation Function

Definition 1.3.1. *Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The* mean function *of $\{X_t\}$ is defined by*

$$\mu_X(t) = E(X_t)$$

The (auto) covariance function *of $\{X_t\}$ is defined by*

$$\gamma_X(r,s) = \mathrm{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

Note that we denote $\mu_X(t)$ by $\mu(t)$ and $\gamma_X(t,s)$ as $\gamma(t,s)$.

1.3.1 Properties of Covariances

- For $a \in \mathbb{R}$,

$$\mathrm{Cov}(X, a) = 0$$

- For $a, b \in \mathbb{R}$,

$$\mathrm{Cov}(X, aU + bV) = a\,\mathrm{Cov}(X,U) + b\,\mathrm{Cov}(X,V)$$

- $\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$

Definition 1.3.2. *A time series $\{X_t\}$ is* (weakly) stationary *if*

1. *$\mu(t)$ is independent of t, and*
2. *$\gamma(t+h,t) = \gamma(h)$, in other words γ is independent of t and is a function that depends only on 1 variable $h = t - s$.*

Definition 1.3.3. *Let $\{X_t\}$ be a stationary time series. The* autocorrelation function *is defined as*

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\mathrm{Cov}(X_0, X_h)}{\mathrm{Var}(X_0)}$$

1.3.2 Properties of Autocorrelation and Autocovariance Functions

(i) $|\rho(h)| \leq 1$, i.e $|\gamma(h)| \leq \gamma(0)$.

(ii) $\gamma(-h) = \gamma(h)$

(iii) $\gamma(h)$ is a non-negative definite funtcion, that is for all non-negative integers n and real numbers $a_1, \ldots, a_n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(|i-j|) \geq 0$$

Proof. (i) This result is obvious.

(ii) This result follows from the fact $\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_t, X_{t-h})$.

(iii) This follows from the properties of covariance,

$$0 \leq \text{Var}\left(\sum_{j=1}^{n} a_j X_j\right) = \text{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\text{Cov}(X_i, X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j)$$

Then from (ii), $\gamma(i-j) = \gamma(|i-j|)$.

□

Example 1.3.1 (White Noise)**.** Let $\{Z_t\}$ be a sequence of independent random variables with mean 0 and variance 1. Then, clearly $\{Z_t\}$ is stationary with the covariance function

$$\gamma_Z(t+h, t) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

Example 1.3.2 (Random Walk)**.** Let $\{Z_t\}$ be a sequence of iid random variables with mean 0 and variance σ_Z^2. We will define $S_t = \sum_{i=1}^{t} Z_i$, with $E(S_t) = 0$, then

$$\begin{aligned} \gamma_S(t, t+h) &= \text{Cov}(S_t, +S_{t+h}) \\ &= \text{Cov}(S_t, S_t + Z_{t+1} + \cdots + Z_{t+h}) \\ &= \text{Cov}(S_t, S_t) + \text{Cov}(S_t, Z_{t+1} + \cdots + Z_{t+h}) \\ &= \text{Cov}(S_t, S_t) + \text{Cov}(Z_1 + \cdots Z_t, Z_{t+1} + \cdots + Z_{t+h}) \\ &= \text{Cov}(S_t, S_t) + 0 = \text{Var}(S_t) \\ &= t\sigma_Z^2 \end{aligned}$$

We see that the variance of S_t depends on t, hence the covariance function depends on t. Thus, the model is not stationary.

Example 1.3.3 (Model with Trend)**.** Let $\{Z_t\}$ be a sequence of iid random variables with mean $\mu_Z = E(Z_t)$, and define

$$X_t = 1 + 2tZ_t, \ t = 1, 2, \ldots$$

Then,

$$E(X_t) = 1 + 2t + \mu_Z$$

Thus, the mean function depends on t, so the model is not stationary.

Example 1.3.4. Let $\{Z_t\}$ be a sequence of iid random variables with mean 0 and variance σ_Z^2. Define

$$X_t = Z_t Z_{t-1} Z_{t-2}, \ t \geq 3$$

Then, we have

$$E(X_t) = E(Z_t Z_{t-1} Z_{t-2}) = E(Z_t)E(Z_{t-1})E(Z_{t-2}) = 0$$

$$\text{Var}(X_t) = E(X_t^2) = E(Z_t^2 Z_{t-1}^2 Z_{t-2}^2) = \sigma_Z^6$$

$$\text{Cov}(X_t, X_{t+1}) = E(X_t X_{t+1}) = E(Z_t Z_{t-1} Z_{t-2} Z_{t+1} Z_t Z_{t-1}) = E(Z_{t+1})E(Z_t^2)E(Z_{t-1}^2)E(Z_{t-2}) = 0$$

Therefore, the model is stationary.

Example 1.3.5 (First-order Moving Average MA(1))**.** Consider the series defined by

$$X_t = Z_t + \theta Z_{t-1}$$

with $\{Z_t\}$ being a sequence of iid random variables with mean 0 and variance σ_Z^2, and $\theta \in \mathbb{R}$. Then, we have $E(X_t) = 0$, and

$$\begin{aligned} \text{Var}(X_t) = E(X_t^2) &= E(\{Z_t + \theta Z_{t-1}\}^2) \\ &= E(Z_t^2) + \theta^2 E(Z_{t-1}^2) + 2\theta E(Z_t Z_{t-1}) \\ &= \sigma_Z^2 + \theta^2 \sigma_Z^2 + 0 \\ &= (1 + \theta^2)\sigma_Z^2 \end{aligned}$$

So the covariance function is

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \theta\sigma_Z^2 & h = \pm 1 \\ 0 & |h| > 1 \end{cases}$$

Thus, the model is stationary.

1.4 Partial Autocorrelation Function

The correlation coefficient between 2 two or more variables can be a consequence of multilinearity. We use the partial correlation function to remove multilinearity. Let $\{X_t\}$ be a stationary time series with mean 0. The *partial autocorrelation function* between X_t and X_{t+k} is defined as the correlation between X_t and X_{t+k} after conditioning out in the time series $X_{t+1}, \ldots, X_{t+k-1}$.

Suppose we want to find the partial autocorrelation function of X_1 and X_2 from the time series $\{X_t\}$ when the effect of X_3 is removed, then we follow the following procedure:

(i) We first regress X_1 on X_3, so we compute the regression coefficient β_{13},

$$X_1 = \beta_{13} X_3 + Z$$

where Z has mean zero and is independent of X_3. Then

$$X_1 X_3 = \beta_{13} X_3^2 + Z X_3$$
$$E(X_1 X_3) = \beta_{13} E(X_3^2) + E(Z X_3)$$
$$\gamma_X(2) = \beta_{13}\gamma_X(0) + E(Z X_3) = \beta_{13}\gamma_X(0)$$

Thus,

$$\beta_{13} = \frac{\gamma_X(2)}{\gamma_X(0)} = \rho_X(2)$$

(ii) Then similarily, we regress X_2 on X_3,

$$X_2 = \beta_{23} X_3 + V$$

and find

$$\beta_{23} = \frac{\gamma_X(1)}{\gamma_X(0)} = \rho_X(2)$$

(iii) Finally, we define the partial autocorrelation function (PACF) between X_1 and X_2 when removing the effect of X_3 as

$$\rho_{12.3} = \mathrm{Cor}(X_1 - \beta_{13}X_3, X_2 - \beta_{23}X_3) = \frac{\mathrm{Cov}(X_1 - \beta_{13}X_3, X_2 - \beta_{23}X_3)}{\sqrt{\mathrm{Var}(X_1 - \beta_{13}X_3)}\sqrt{\mathrm{Var}(X_2 - \beta_{23}X_3)}}$$

We can rewrite the covariance as

$$\begin{aligned}
\mathrm{Cov}(X_1 - \beta_{13}X_3, X_2 - \beta_{23}X_3) &= \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(\beta_{13}X_3, \beta_{23}X_3) - \mathrm{Cov}(X_1, \beta_{23}X_3) - \mathrm{Cov}(\beta_{13}X_3, X_2) \\
&= \gamma_X(1) + \beta_{13}\beta_{23}\,\mathrm{Cov}(X_3, X_3) - \beta_{23}\,\mathrm{Cov}(X_1, X_3) - \beta_{13}\,\mathrm{Cov}(X_3, X_1) \\
&= \gamma_X(1) + \beta_{13}\beta_{23}\gamma_X(0) - \beta_{23}\gamma_X(2) - \beta_{13}\gamma_X(1) \\
&= \gamma_X(1) + \rho_X(2)\rho_X(1)\gamma_X(0) - \rho_X(1)\gamma_X(2) - \rho_X(2)\gamma_X(1) \\
&= \gamma_X(1) - \rho_X(2)\gamma_X(1) - \frac{\gamma_X(1)}{\gamma_X(0)}\gamma_X(2) - \rho_X(2)\gamma_X(1) \\
&= \gamma_X(1) + \rho_X(2)\gamma_X(1) - \rho_X(2)\gamma_X(1) - \rho_X(2)\gamma_X(1) \\
&= \gamma_X(1)(1 - \rho_X(2))
\end{aligned}$$

We can also compute the variances

$$\begin{aligned}
\mathrm{Var}(X_1 - \beta_{13}X_3) &= \mathrm{Cov}(X_1 - \beta_{13}X_3, X_1 - \beta_{13}Y_3) \\
&= \mathrm{Cov}(X_1, X_1) + \mathrm{Cov}(\beta_{13}X_3, \beta_{13}X_3) - \mathrm{Cov}(X_1, \beta_{13}X_3) - \mathrm{Cov}(\beta_{13}X_3, X_1) \\
&= \gamma_X(0) + \beta_{13}^2\gamma_X(0) - \beta_{13}\gamma_X(1) - \beta_{13}\gamma_X(1) \\
&= \gamma_X(0) + \beta_{13}^2\gamma_X(0) - 2\beta_{13}\gamma_X(1) \\
&= \gamma_X(0)(1 - \rho_X^2(2))
\end{aligned}$$

Similarily,

$$\mathrm{Var}(X_2 - \beta_{23}X_3) = \gamma_X(0)(1 - \rho_X^2(1))$$

Thus, the partial autocorrelation function is

$$\rho_{12.3} = \frac{\gamma_X(1)(1 - \rho_X(2))}{\gamma_X(0)\sqrt{(1 - \rho_X^2(2))(1 - \rho_X^2(1))}} = \frac{\mathrm{Cor}(X_1, X_2) - \mathrm{Cor}(X_2, X_3)\,\mathrm{Cor}(X_1, X_3)}{\sqrt{(1 - \mathrm{Cor}^2(X_1, X_3))(1 - \mathrm{Cor}^2(X_2, X_3))}}$$

Note. Since the time series $\{X_t\}$ is stationary, we can write

$$\gamma_X(1) = \text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_3)$$

So we can also say $\text{Cor}(X_1, X_2) = \text{Cor}(X_2, X_3)$. The PACF between X_1 and X_3 when removing the effect of X_2 is

$$\rho_{13.2} = \frac{\text{Cor}(X_1, X_3) - \text{Cor}(X_1, X_2)\,\text{Cor}(X_2, X_3)}{\sqrt{(1 - \text{Cor}^2(X_1, X_2))(1 - \text{Cor}^2(X_2, X_3))}}$$

Definition 1.4.1. *Given a time series $\{X_t\}$, the partial autocorrelation at lag h, denoted $\alpha(h)$, is the autocorrelation between X_t and X_{t+h} with the linear dependence of X_t on $X_{t+1}, \ldots, X_{t+h-1}$ removed.*

Note. $\alpha(1) = \rho_X(1)$, and $\alpha(2) = \rho_{13.2}$.

Example 1.4.1 (MA(1))**.** Recall the model $X_t = Z_t + \theta Z_{t-1}$, where $\{Z_t\}$ is a sequence of iid random variables with mean 0 and variance σ_Z^2. We have

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta}{1+\theta^2} & h = \pm 1 \\ 0 & |h| > 1 \end{cases}$$

Then,

$$\begin{aligned} \alpha(2) &= \frac{\text{Cor}(X_1, X_3) - \text{Cor}(X_1, X_2)\,\text{Cor}(X_2, X_3)}{\sqrt{(1 - \text{Cor}^2(X_1, X_2))(1 - \text{Cor}^2(X_2, X_3))}} \\ &= \frac{\rho_X(2) - \rho_X^2(1)}{\sqrt{1 - \rho_X^2(1)}\sqrt{1 - \rho_X^2(1)}} \\ &= \frac{0 - \frac{\theta^2}{(1+\theta^2)^2}}{1 - \frac{\theta^2}{(1+\theta^2)^2}} = \frac{-\theta^2}{1 + \theta^2 + \theta^4} \end{aligned}$$

Chapter 2

Sample Mean, Sample Autocorrelation, Sample PACF

Let $\{X_t\}$ be a stationary time series, then

- The mean $\mu = E(X_0) = E(X_t)$ can be estimated with the sample mean

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- The variance $\sigma_X^2 = \text{Var}(X) = E[(X-\mu)^2]$ can be estimated with the sample variance

$$\hat{\sigma}_X^2 = \hat{\gamma}_X(0) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- The (auto)covariance $\gamma_X(h) = E[(X_0-\mu)(X_h-\mu)]$ can be estimated with the sample (auto)covariance

$$\hat{\gamma}_X(h) = \frac{1}{n-1}\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})$$

- The (auto)correlation $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$ can be estimated with the sample (auto)correlation

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

- The partial autocorrelation function (PACF) is estimated by the sample PACF. For example,

$$\alpha(2) = \frac{\rho_X(2) - \rho_X^2(1)}{1 - \rho_X^2(1)} \implies \hat{\alpha}(2) = \frac{\hat{\rho}_X(2) - \hat{\rho}_X^2(1)}{1 - \hat{\rho}_X^2(1)}$$

Chapter 3

Autoregression Moving-Average (ARMA) Models

3.1 Linear Processes

Definition 3.1.1. *Let $\{Z_t\}$ be a sequence of independent random variables with mean zero and variance $\mathrm{Var}(Z_t) = E(Z_t^2) = \sigma_Z^2$. Let $\psi_j,\ j \geq 0$, be a sequence of constants such that $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Then,*

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

is called a linear process (also known as a moving average, or causal moving average, or one-sided moving average).

The condition $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ is required to ensure that the series converges, so that

$$E(|X_t|) \leq \sum_{j=0}^{\infty} |\psi_j| E(|Z_{t-j}|) = E(|Z_0|) \sum_{j=0}^{\infty} |\psi_j| < \infty$$

Lemma 3.1.1. *The linear process is stationary with $E(X_t) = 0$, and*

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

Proof. Since the mean is 0, we have

$$\gamma_X(h) = E(X_t X_{t+h}) = E\left(\sum_{j=0}^{\infty} \psi_j Z_{t-j} \sum_{i=0}^{\infty} \psi_i Z_{t+h-i} \right) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_j E(Z_{t-j} Z_{t+h-j})$$

Since the noise variables Z_t are independent with mean 0, the only terms that contribute to the sum are those with $i = j + h$, so this simplies to a single sum.

$$\gamma_X(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} E(Z_{t-j} Z_{t+h-j}) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$$

□

Note: The notation $AR(p)$ indicates an autoregressive model of order p.

Example 3.1.1 (AR(1))**.** Let $|\phi| < 1$, and consider the linear process

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

Here we have $\psi_j = \phi^j$, then applying the above lemma,

$$\gamma_X(h) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} = \sigma_Z^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma_Z^2 \frac{\phi^h}{1 - \phi^2}$$

Example 3.1.2 (MA(q))**.** Let

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

This is a linear processes with $\psi_0 = 1, \psi_1 = \theta_1, \ldots, \psi_q = \theta_q$, and $\psi_j = 0$ for $j > q$. Then, for $h = 0, \ldots, q$,

$$\gamma_X(h) = \sigma_Z^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+h} = \sigma_Z^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}$$

3.2 ARMA Models

3.2.1 Difference and Backward Operators

We define the *difference operator* ∇ as

$$\nabla X_t = X_t - X_{t-1}$$

and the backward shift operator as

$$BX_t = X_{t-1} \implies B^k X_t = X_{t-k}$$

Example 3.2.1 (AR(1))**.** Let $\{Z_t\}$ be a sequence of independent random variables with mean 0 and variacne $\text{Var}(Z_t) = E(Z_t^2) = \sigma_Z^2$. Define the linear process

$$X_t = \phi X_{t-1} + Z_t$$

This can be equivalently written using the backshift operator B with $BX_t = X_{t-1}$, so

$$X_t = \phi B X_t + Z_t \implies Z_t = (1 - \phi B) X_t = \phi(B) X_t$$

where $\phi(B) = 1 - \phi B$.

Example 3.2.2 (MA(1))**.** Define the linear process

$$X_t = Z_t + \theta Z_{t-1} \implies X_t = Z_t + \theta B Z_t = \theta(B) Z_t$$

where $\theta(B) = 1 + \theta B$.

Example 3.2.3 (ARMA(1,1))**.** The model is defined as

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

Which we can write as

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

where $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$.

Definition 3.2.1. *Let $\{Z_t\}$ be a sequence of independent random variable with mean zero and variance $\text{Var}(Z_t) = E(Z_t^2) = \sigma_Z^2$. A time series $\{X_t\}$ is called an autoregressive moving average process of order (p, q), denoted by $ARMA(p, q)$, if it solves the equation*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

Equivalently,

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

where

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$
$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

are autoregressive and moving average polynomials, respectively.

3.2.2 Solutions to ARMA Models

(i) A stationary solution for ARMA(p,q) exists whenever the autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for all $|z| = 1$. In other words, for ARMA to be stationary, if z solves $\phi(z) = 0$, then $|z| \neq 1$. For example, if we have the model

$$X_t - 1.1X_{t-1} = Z_t$$

The autoregressive polynomial is $\phi(z) = 1 - 1.1z$, and the only solution to $\phi(z) = 0$ is $z = 1/1.1 \neq 1$. Therefore there exists a stationary solution for this model.

(ii) A stationary and causal solution for ARMA(p,q) exists whenever the autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for all $|z| \leq 1$. In other words, for ARMA to be stationary and causal, if z solves $\phi(z) = 0$, then $|z| > 1$. For example, consider the previous model,

$$X_t - 1.1X_{t-1} = Z_t$$

Then the autoregressive polynomial is $\phi(z) = 1 - 1.1z$, and the only solution to $\phi(z) = 0$ is $z = 1/1.1 < 1$. Therefore there is no causal solution for this model.

Example 3.2.4. Consider the AR(2) process

$$X_t - 0.1X_{t-1} - 0.4X_{t-2} = Z_t$$

Equivalently, we can write this using the backshift operator as

$$X_t - 0.1BX_t - 0.4B^2X_t = Z_t$$

So, the autoregressive polynomial is

$$\phi(z) = 1 - 0.1z - 0.4z^2$$

The roots of this polynomial are $1.46, -1.71$. Both roots are greater than 1 in absolute value. Therefore, there is a causal and stationary solution for this model.

Example 3.2.5. Consider the AR(2) process

$$(1 - B - B^2)X_t = Z_t$$

With the autoregressive polynomial $\phi(z) = 1 - z - z^2$. The roots of this polynomial are complex,

$$z = \frac{-1 \pm i\sqrt{3}}{2}$$

Then, the modulus of the roots is

$$|z| = \sqrt{1+3} = 2$$

Therefore, there exists a stationary and causal solution for this model.

Example 3.2.6. Consider the AR(2) process

$$X_t - \phi X_{t-1}\phi X_{t-2} = Z_t$$

The autoregressive polynomial is

$$\phi(z) = 1 - \phi z - \phi z^2$$

The roots of the polynomial are

$$\frac{\phi \pm \sqrt{\phi^2 + 4\phi}}{2\phi}$$

We have to consider the cases where the roots are real and complex.

- If $\phi \in (-4, 0)$, the roots are complex. Then, the modulus of the roots is

$$\sqrt{\frac{\phi^2 + \phi^2 + 4\phi}{4\phi^2}}$$

This is greater than 1 when $2\phi^2 + 4\phi > 4\phi^2$, so

$$2\phi^2 + 4\phi > 4\phi^2 \iff 0 > 2\phi^2 - 4\phi = 2\phi(\phi - 2) \iff 0 < \phi < 2$$

We had the condition that $\phi \in (-4, 0)$, so the only solution is $\phi \in (-4, 0) \cap (0, 2) = \emptyset$. Therefore, there is no causal solution for this model if the roots are complex.

- If $\phi < -4$ or $\phi > 0$, the roots are real. Then, we want to solve

$$\left|\frac{\phi \pm \sqrt{\phi^2+4\phi}}{2\phi}\right| > 1 \iff \frac{\phi \pm \sqrt{\phi^2+4\phi}}{2\phi} > 1 \text{ or } \frac{\phi \pm \sqrt{\phi^2+4\phi}}{2\phi} < -1$$

We will start with the first case when $\phi > 0$,

$$\frac{\phi \pm \sqrt{\phi^2+4\phi}}{2\phi} > 1 \iff \phi \pm \sqrt{\phi^2+4\phi} > 2\phi \iff \pm\sqrt{\phi^2+4\phi} > \phi$$

$$\frac{\phi \pm \sqrt{\phi^2+4\phi}}{2\phi} < -1 \iff \phi \pm \sqrt{\phi^2+4\phi} < -2\phi \iff \pm\sqrt{\phi^2+4\phi} < -3\phi$$

We can see for the positive root this always holds when $\phi > 0$ since $\sqrt{\phi^2+4\phi} > \sqrt{\phi^2} = \phi$. When $\phi > 0$, $\sqrt{\phi^2+4\phi} > 0$, so the second inequality never holds when $\phi > 0$. Then, for the negative root,

$$-\sqrt{\phi^2+4\phi} > \phi \iff \sqrt{\phi^2+4\phi} < -\phi$$

$$-\sqrt{\phi^2+4\phi} < -3\phi \iff \sqrt{\phi^2+4\phi} > 3\phi \iff \phi^2+4\phi > 9\phi^2 \iff 0 > 2\phi(\phi-2)$$

We can see that the first inequality never holds since $\phi > 0$, and the second inequality holds when $\phi \in (0,2)$. Now, when $\phi < -4$,

$$\sqrt{\phi^2+4\phi} > \phi$$

always holds since the square root will be positive, and

$$\sqrt{\phi^2+4\phi} < -3\phi$$

always holds since $\sqrt{\phi^2+4\phi} < \sqrt{\phi^2} = \phi < -3\phi$ when $\phi < -4$. Then for the second root,

$$\sqrt{\phi^2+4\phi} < -\phi \text{ or } 0 > 2\phi(\phi-2)$$

The first inequality always holds for the same reason as the previous root, $\sqrt{\phi^2+4\phi} < \sqrt{\phi^2} = \phi$, and the second inequality never holds since $\phi < -4 < 0$, thus $2\phi(\phi-2) > 0$. Therefore, there is a causal solution for this model when $\phi < -4$ or $\phi \in (0,2)$.

3.3 Linear Representation for ARMA Models

Given an ARMA(p,q) model, our objective is to write a linear process as given in Definition 3.1.1. In general, this task is not easy and we will look at some basic models.

Example 3.3.1 (MA(q))**.** When $p = 0$, ARMA$(0,q)$ becomes MA(q), so it's linear representation is trivial.

Example 3.3.2 (AR(1))**.** AR(1) is obtained by setting $p = 1$ and $q = 0$, so

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

where $\phi(z) = 1 - \phi z$. Then, we define

$$\chi(z) = \frac{1}{\phi(z)}$$

This function has the following power series expansion when $|\phi| < 1$,

$$\chi(z) = \frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \phi^j z^j$$

Multiplying this with our equation for Z_t,

$$\chi(B)\phi(B)X_t = \chi(B)Z_t \iff X_t = \chi(B)Z_t$$

Thus, our linear representation of AR(1) is

$$X_t = \chi(B)Z_t = \sum_{j=0}^{\infty} \phi^j B^j Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

Here we have our linear representation from the definition with $\psi_j = \phi^j$. Notice that this works under the condition that there exists a causal linear representation. If $|\phi| > 1$, then we can still write a linear representation, however it will not be causal.

Example 3.3.3 (ARMA(1,1))**.** Similarly to our previous examples, here we have $p = 1$ and $q = 1$, so

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

where $\phi(z) = 1 - \phi z$ and $\theta(z) = 1 + \theta z$. Then, we define again

$$\chi(z) = \frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \phi^j z^j$$

Multyipling this gives us

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

So, we can expand this to get

$$X_t = \chi(B)\theta(B)Z_t = \sum_{j=0}^{\infty} \phi^j B^j (1 + \theta B) Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} + \theta \sum_{j=0}^{\infty} \phi^j Z_{t-(j+1)}$$

Now, we want to write X_t in the form $\sum\limits_{j=0}^{\infty} \psi_j Z_{t-j}$, so

$$\sum_{j=0}^{\infty} \psi_j Z_{t-j} = \sum_{j=0}^{\infty} \phi^j Z_{t-j} + \theta \sum_{j=0}^{\infty} Z_{t-j-1}$$

We can rewrite this as

$$\psi_0 Z_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j} = \phi^0 Z_t + \sum_{j=1}^{\infty} (\phi^j + \theta\phi^{j-1}) Z_{j-1}$$

Hence the linear representation of ARMA$(1, 1)$ is

$$\psi_0 = 1,\ \psi_j = \phi^{j-1}(\theta + \psi),\ j \geq 1$$

Example 3.3.4 (ARMA(1,q))**.** ARMA(1,q) works in the same way as ARMA(1,1).

Example 3.3.5 (AR(p))**.** The procedure for AR(p), for $p \geq 2$ does not generalize the same way and is far more complicated.

3.4 Autocovariance Function for ARMA Model

Our goal to find the autocovariance function for an ARMA(p, q) model.

Example 3.4.1 (MA(q))**.** For an MA(q) model, the linear representation is trivial and can be used to evaluate the autocovariance function directly. See Example 3.1.2.

Example 3.4.2 (AR(1))**.** See Example 3.1.1, we have

$$\gamma_X(h) = \sigma_Z^2 \frac{\phi^h}{1 - \phi^2}$$

Example 3.4.3 (ARMA(1,1))**.** For ARMA$(1, 1)$ (this can be generalized for ARMA$(1, q)$), we use the linear representation from Example 3.3.3, we have

$$\psi_0 = 1,\ \psi_j = \phi^{j-1}(\phi + \theta),\ j \geq 1$$

Then, using Lemma 3.1.1, we have

$$\begin{aligned}
\gamma_X(0) &= \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma_Z^2 \psi_0^2 + \sigma_Z^2 \sum_{j=1}^{\infty} \psi_j^2 \\
&= \sigma_Z^2 \left(1 + (\psi + \theta)^2 \sum_{j=0}^{\infty} \psi^{2(j-1)} \right) \\
&= \sigma_Z^2 \left(1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right)
\end{aligned}$$

Similarly for $\gamma_X(1)$,

$$\gamma_X(1) = \sigma_Z^2 \left((\phi + \theta) + \phi \frac{(\theta + \phi)^2}{1 - \phi^2} \right)$$

In general, $\gamma_X(h) = \phi^{h-1}\gamma_X(1)$ for $h \geq 1$, so the covariance function of ARMA$(1,1)$ is

$$\gamma_X(h) = \sigma_Z^2 \phi^{h-1} \left((\theta + \phi) + \phi \frac{(\phi + \theta)^2}{1 - \phi^2} \right), \ h \geq 1$$

Example 3.4.4 (AR(1))**.** For AR(p) models, or ARMA(p,q) models in general with $p \geq 2$, we apply a recursive method. For example, consider AR(1),

$$X_t = \phi X_{t-1} + Z_t$$

Then multiply both sides by X_{t-h} and take the expected value,

$$E(X_t X_{t-h}) = \phi E(X_{t-1} X_{t-h}) + E(Z_t X_{t-h})$$

Since $E(X_t) = 0$, then $E(X_t X_{t-h}) = \gamma_X(h)$, and $E(X_{t-1} X_{t-h}) = \gamma_X(h-1)$. Then, we also have that Z_t is independent of X_{t-h}, so $E(Z_t X_{t-h}) = 0$. Therefore, the recursive formula for AR(1) is

$$\gamma_X(h) = \phi \gamma_X(h-1)$$

Or in general,

$$\gamma_X(h) = \phi^{h-1}\gamma_X(0), \ h \geq 1$$

We need to compute $\gamma_X(0) = \text{Var}(X_t) = \sigma_X^2$. We have

$$\text{Var}(X_t) = \phi^2 \text{Var}(X_{t-1}) + \text{Var}(Z_t)$$

Since X_t is stationary, $\text{Var}(X_t) = \text{Var}(X_{t-1}) = \sigma_X^2$, so

$$\sigma_X^2 = \phi^2 \sigma_X^2 + \sigma_Z^2 \implies \sigma_X^2 = \sigma_Z^2 \frac{1}{1 - \phi^2}$$

Therefore,

$$\gamma_X(h) = \phi^h \gamma_X(0) = \phi^h \sigma_Z^2 \frac{1}{1 - \phi^2}$$

We can see this is the same as the result from Example 3.1.1.

Example 3.4.5 (AR(2))**.** Using the same method as the previous example, we have

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

Then multiplying both sides by X_{t-h} and taking the expected value,

$$E(X_t X_{t-h}) = \phi_1 E(X_{t-1} X_{t-h}) + \phi_2 E(X_{t-2} X_{t-h}) + E(Z_t X_{t-h})$$

This gives us the recursive formula for the covariance of AR(2) as

$$\gamma_X(h) = \phi_1 \gamma_X(h-1) + \phi_2 \gamma_X(h-2)$$

Again, we need to compute $\gamma_X(0) = \text{Var}(X_t) = \sigma_X^2$ and $\gamma_X(1)$,

$$\gamma_X(h) = \phi_1 \gamma_X(h-1) + \phi_2 \gamma_X(h-2), \ h \geq 2$$

$$\gamma_X(1) = \sigma_Z^2 \frac{\phi_1}{(1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2)}$$

$$\gamma_X(0) = \sigma_Z^2 \frac{1 - \phi_2}{(1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2)}$$

Example 3.4.6 (AR(3))**.** The recursive formula for the covariance of AR(3) is

$$\gamma_X(h) = \phi_1 \gamma_X(h-1) + \phi_2 \gamma_X(h-2) + \phi_3 \gamma_X(h-3)$$

3.4.1 Partial Autocorrelation for ARMA Models

Recall the definition for the partial autocorrelation function (PACF). Given a time series $\{X_t\}$, the partial autocorrelation at lag h, denoted by $\alpha(h)$, is the autocorrelation between X_t and X_{t+h} with the linear dependence of X_t on $X_{t+1}, \ldots, X_{t+h-1}$ removed.

Example 3.4.7 (PACF for MA(1))**.** For MA(1) models, we have already calculated $\alpha(2)$. The PACF in general is

$$\alpha(h) = \frac{-(-\theta)^h}{1 + \theta^2 + \cdots + \theta^{2h}}$$

Example 3.4.8 (PACF for AR(1))**.** Consider the AR(1) model $X_t = \phi X_{t-1} + Z_t$. Then, the PACF is

$$\alpha(1) = \rho_X(1) = \phi$$

$$\alpha(2) = \text{Cor}(X_0, X_2 - \phi X_1) = \text{Cor}(X_0, \phi X_1 + Z_2 - \phi X_1) = \text{Cor}(X_0, Z_2) = 0$$

Theorem 3.4.1. *Consider the stationary* AR(p) *time series. Then*

$$\alpha(h) = 0, \;\; h = p+1, p+2, \ldots$$

Chapter 4

Forecasting of Stationary Time Serires

Our objective is to forecast or predict X_{n+k}, for some $k \geq 1$, having observed $\{X_1, \ldots, X_n\}$ from a time series with a known mean μ and known autocovariance function $\gamma_X(k)$.

Consider a stationary sequence with mean $\mu = E(X_t)$ and covariance $\gamma_X(h)$. Denote $P_n X_{n+k}$ as the predicted value of X_{n+k} given that we have n observations $X_1, \ldots, X_n$. Then, the best linear predictor of X_{n+k} is given by

$$P_n X_{n+k} = a_0 + a_1 X_n + \cdots + a_n X_1 = a_0 \sum_{i=1}^{n} a_i X_{n+1-i}$$

where $a_0, \ldots, a_n$ are constants. We want to find optimal values for $a_0, \ldots, a_n$, in this case we will minimize the mean squared error (MSE) of the prediction

$$E\left[(X_{n+k} - P_n X_{n+k})^2\right]$$

4.1 Yule-Walker Procedure

Define

$$S(a_0, a_1, \ldots, a_n) = E\left[(X_{n+k} - P_n X_{n+k})^2\right] = E\left[\left(X_{n+k} - a_0 - \sum_{i=1}^{n} a_1 X_{n+1-i}\right)^2\right]$$

Then, we take the derivative of S with respect to $a_0, \ldots, a_n$ and set them equal to zero. This gives us

$$a_0 = \mu\left(1 - \sum_{i=1}^{n} a_i\right)$$

We assume $\mu = 0$, so $a_0 = 0$. Then we take derivatives with respect to $a_1, \ldots, a_n$ to get

$$E\left[\left(X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}\right) X_{n+1-j}\right] = 0,\ j = 1, \ldots, n$$

Equivenlently, we can write this as

$$E(X_{n+k} X_{n+1-j}) - \sum_{i=1}^{n} a_i E(X_{n+1-i} X_{n+1-j}) = 0,\ j = 1, \ldots, n$$

The above expectations are the covariances at lags $n+k-(n+1-j) = k-1+j$ and $n+1-i-(n+1-j) = i-j$, so we can write this a system of equations

$$\gamma_X(k - i + j) = \sum_{i=1}^{n} a_i \gamma_X(i - j),\ j = 1, \ldots, n$$

We can define this using matrix and vector notation, define

$$\Gamma_n = \left[\gamma_X(i-j)\right]_{i,j=1}^{n}$$

$$\gamma(n;k) = \begin{pmatrix} \gamma_X(k) & \cdots & \gamma_X(k+n-1) \end{pmatrix}^T,\ \vec{a}_n = \begin{pmatrix} a_1, \ldots, a_n \end{pmatrix}^T$$

The matrix Γ_n is called the variance-covariance matrix of $(X_1, \ldots, X_n)$. Then, we can write the system of equations as

$$\Gamma_n a_n = \gamma(n;k)$$

Hence, this gives us the Yule-Walker formula for forecasting

$$\vec{a}_n = \Gamma_n^{-1}\gamma(n;k)$$

This depends on Γ_n being invertible, and that the time series model is stationary. We require the covariances only, and these can be computed if we can verify the data comes from a particular model, otherwise we use the sample covariances.

4.1.1 Mean Square Prediction Error (MSPE)

The Yule-Walker procedure minimizes the mean squared prediction error (MSPE)

$$\text{MSPE}_n(k) = E\left[\left(X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}\right)^2\right]$$

We can rewrite this as

$$E\left[\left(X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}\right)^2\right] = \gamma_X(0) - \vec{a}_n^T\gamma(n;k)$$

The mean square prediction error depends on k, meaning that predictions made further in the future (a larger k value) may have a larger prediction error.

Example 4.1.1 (AR(1))**.** Consider the AR(1) model $X_t = \phi X_{t-1} + Z_t$, where Z_t are iid centered with mean 0 and variance σ_Z^2, and $|\phi| < 1$. Recall that

$$\gamma_X(h) = \phi^h \frac{\sigma_Z^2}{1-\phi^2},\ h \geq 0$$

Then,

$$\gamma(n;k) = \gamma(n;1) = \begin{pmatrix} \gamma_X(1) & \cdots & \gamma_X(n) \end{pmatrix}^T = \frac{\sigma_Z^2}{1-\phi^2}\begin{pmatrix} \phi & \cdots & \phi^n \end{pmatrix}^T$$

Then $\Gamma_n a_n = \gamma(n;1)$ becomes

$$\frac{\sigma_Z^2}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix}\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \frac{\sigma_Z^2}{1-\phi^2}\begin{bmatrix} \phi \\ \vdots \\ \phi^n \end{bmatrix}$$

Here we either inverst the matrix, or assume the solution $\vec{a}_n = (\phi, \cdots, 0)^T$. So, the AR(1) prediction is

$$P_n X_{n+1} = \phi X_n$$

The MSPE for AR(1) is

$$\text{MSPE}_n(1) = \gamma_X(0) - \vec{a}_n^t\gamma(n;1) = \gamma_X(0) - \phi\gamma_X(1) = \frac{\sigma_Z^2}{1-\phi^2} - \phi^2\frac{\sigma_Z^2}{1-\phi^2} = \sigma_Z^2$$

4.2 Durbin-Levinson Algorithm

The Durbin-Levinson algorithm avoids computing the inverse of Γ_n, however the trade of is that we will only predict one step ahead, so only P_nX_{n+1}.

Assume that $\mu = E(X_t) = 0$, so that $a_0 = 0$, we will write the predictors as

$$P_nX_{n+1} = \phi_{n1}X_n + \cdots + \phi_{nn}X_1$$

So, $a_1 = \phi_{n1}, \ldots, a_n = \phi_{nn}$. If we take $n = 1$, we can still use the Yule-Walker formula to get

$$\phi_{11} = \gamma_X^{-1}(0)\gamma_X(1) = \rho_X(1)$$

Taking $n = 2$, we want to find coefficients ϕ_{21}, ϕ_{22} such that

$$P_2X_3 = \phi_{21}X_2 + \phi_{22}X_2$$

Similarly to the Yule-Walker procedure, we want to minimize

$$E[(X_3 - \phi_{21}X_2 - \phi_{22}X_1)^2]$$

Taking the derivatives with respect to ϕ_{21}, ϕ_{22},

$$\begin{cases} E[X_2(X_3 - \phi_{21}X_2 - \phi_{22}X_1)] = 0 \\ E[X_1(X_3 - \phi_{21}X_2 - \phi_{22}X_1)] = 0 \end{cases} \rightarrow \begin{cases} \gamma_X(1) - \phi_{21}\gamma_X(0) - \phi_{22}\gamma_X(1) = 0 \\ \gamma_X(2) - \phi_{21}\gamma_X(1) - \phi_{22}\gamma_X(0) = 0 \end{cases}$$

Dividing each equation by $\gamma_X(0)$ gives us

$$\begin{cases} \phi_{21} = \rho_X(1) - \phi_{22}\rho_X(1) = \rho_X(1) - \phi_{22}\phi_{11} \\ \rho_X(2) - \phi_{21}\rho_X(1) - \phi_{22} = 0 \end{cases}$$

Solving these equations gives us

$$\begin{cases} \phi_{22} = \frac{\rho_X(2) - \phi_{11}\rho_X(1)}{1 - \phi_{11}\rho_X(1)} \\ \phi_{21} = \rho_X(1) - \phi_{22}\phi_{11} \end{cases}$$

Note that ϕ_{11} and $\rho_X(1)$ is used interchangeablly.

Theorem 4.2.1 (Durbin-Levinson Algorithm)**.** *The coefficients $\phi_{n1}, \ldots, \phi_{nn}$ can be computed recursively as*

$$\phi_{nn} = \left[\gamma_X(n) = \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma_X(n-j)\right] v_{n-1}^{-1}$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

and

$$v_n = v_{n-1}[1 - \phi_{nn}^2], \ v_0 = \gamma_X(0), \ \phi_{11} = \rho_X(1)$$

Note that the Durbin-Levinson and Yule-Walker procedures produce the same results since they both compute the coefficients in the linear prediction P_nX_{n+1} using the mean squared error criterion.

Example 4.2.1. Consider the AR(1) model $X_t = \phi X_{t-1} + Z_t$, where Z_t are iid with mean 0 and variance σ_Z^2. The covariance function is given by

$$\gamma_X(h) = \phi^h \frac{\sigma_Z^2}{1 - \phi^2}$$

The correlation is

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

Using the Durbin-Levinson algorith, we find the coefficients and predictors

$$\phi_{11} = \phi, \ P_1X_2 = \phi X_1$$

$$\phi_{22} = 0, \ P_2X_3 = \phi X_2$$

Then in general,

$$\phi_{n1} = \phi, \phi_{n2} = \cdots = \phi_{nn} = 0, \ P_nX_{n+1} = \phi X_n$$

4.2.1 Partial Autocovariance Functions

The Durbin-Levinson procedure also gives us the partial autocovariance function defined as

$$\alpha(0) = 1,\ \alpha(k) = \phi_{kk},\ k \geq 1$$

Example 4.2.2. Consider the AR(1) model $X_t = \phi X_{t-1} + Z_t$, where Z_t are iid with mean 0 and variance σ_Z^2. The partial autocovariance function is given by

$$\alpha(0) = 1,\ \alpha(1) = \phi,\ \alpha(k) = 0,\ k \geq 2$$

The partial autocorrelation function vaniches after lag 1. In general, the partial autocorrelation function of an AR(p) model vanishes after lag p.

Example 4.2.3. For the MA(1) model, the PACF is

$$\alpha(h) = \frac{-(-\theta)^h}{1 + \theta^2 + \cdots + \theta^{2h}}$$

4.3 Forecast Limits, Prediction Intervals

We discussed above formulas for predicting values that depend only on the covariances. To assess the accuracy of these predictions, we need to impose some assumptions on the model.

Assume that $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ is a linear process with mean $E(Z_t) = 0$ and variance $\text{Var}(Z_t) = \sigma_Z^2$. The mean squared prediction error is

$$\text{MSPE}_n(k) \coloneqq E[(X_{n+k} - P_n X_{n+k})^2]$$

Hence, the theoretical forecase limits are

$$P_n X_{n+k} \pm z_{\alpha/2}\sqrt{\text{MSPE}_n(k)} = P_n X_{n+k} \pm z_{\alpha/2}\sigma\sqrt{\sum_{j=0}^{k-1} \psi_j^2}$$

Chapter 5

Estimation in ARMA Models

Suppose we have observations $\{X_1, \ldots, X_n\}$ from a time series with a known model, say ARMA(p, q), our goal is to estimate the parameters in $\phi_1, \ldots, \phi_p$, and $\theta_1, \ldots, \theta_q$.

5.1 Estimation of the Mean

We assume the mean of our stationary process is 0, however in practice the mean is not known and is not typically 0. We will estimate the mean using the sample mean, $\bar{X}$.

Assume that $X_1, \ldots, X_n$ are iid, then

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}n\mu = \mu$$

From independence, we also have

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{\gamma_X(0)}{n}$$

Lemma 5.1.1. *Assume that $X_1, \ldots, X_n$ are independent and identically distributed with mean μ and variance $\gamma_X(0) = \sigma^2$. Then*

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

That is,

$$\lim_{n\to\infty} P\left(\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \leq x\right) = \Phi(x)$$

where Φ is the cdf of the standard normal distribution.

This allows us to construct confidence intervals for the mean

$$\left(\bar{X} - z_{\alpha/2}\frac{\sqrt{\gamma_X(0)}}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sqrt{\gamma_X(0)}}{\sqrt{n}}\right)$$

When dealing with time series, we are not using iid random variables. Suppose we have observations $\{X_1, \ldots, X_n\}$ from a stationary time series. The sample mean still remains an unbiased estimator, however the variance of the sample mean becomes

$$\begin{aligned}
\text{Var}(\bar{X}) = \text{Cov}(\bar{X}, \bar{X}) &= \text{Cov}\left(\frac{X_1 + \cdots + X_n}{n}, \frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(X_i, X_j) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_X(i-j) \\
&= \frac{1}{n^2}\sum_{h=-(n-1)}^{n-1}(n - |h|)\gamma_X(h) = \frac{1}{n^2}\sum_{h=-n}^{n}(n - |h|)\gamma_X(h) \\
&= \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma_X(|h|)
\end{aligned}$$

Lemma 5.1.2. *Assume that $X_1, \ldots, X_n$ is a stationary time series with mean μ and variance $\gamma_X(0)$ and the covariance function $\gamma_X(h)$. Then*

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{\nu}\right) \xrightarrow{d} N(0,1)$$

where

$$\nu^2 = \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)$$

This allows us to construct confidence intervals for the mean

$$\left(\bar{X} - z_{\alpha/2}\frac{\nu}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\nu}{\sqrt{n}}\right)$$

Example 5.1.1. Consider the AR(1) model

$$X_t = \phi X_{t-1} + Z_t$$

In order to obtain a linear representation of the model, we need to have that the mean is 0. We can do this by subtracting the mean from the model, so

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

The stationary solution is then

$$X_t = \mu + \sum_{j=0}^{\infty}\phi^j Z_{t-j}$$

The covariance function is

$$\gamma_X(h) = \sigma_Z^2\frac{\phi^h}{1-\phi^2}$$

Thus, the unknown ν is

$$\nu^2 = \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h) = \sigma_Z^2\frac{1}{1-\phi^2} + 2\sigma_Z^2\frac{1}{1-\phi^2}\frac{1}{1-\phi} = \phi_Z^2\frac{1}{(1-\phi)^2}$$

5.2 Yuke-Walker Estimator

This method works well for AR(p) models. We will start with AR(1),

$$X_t = \phi X_{t-1} + Z_t$$

with $|\phi| < 1$, and $E(Z_t) = 0$, $\text{Var}(Z_t) = \sigma_Z^2$. The model is stationary and casual. Multiplying both sides by X_{t-1} and X_t we get

$$X_t X_{t-1} = \phi X_{t-1}X_{t-1} + Z_t X_{t-1}$$
$$X_t^2 = \phi X_t X_{t-1} + X_t Z_t$$

Now, using the fact that the time series is casual, we have $E(X_t) = 0$ and X_{t-1} is independent of Z_t. So, applying the expected value,

$$\gamma_X(1) = \phi\gamma_X(0) + 0$$
$$\gamma_X(0) = \phi\gamma_X(1) + E(X_t Z_t)$$

Note that

$$E(X_t Z_t) = E[(\phi X_{t-1} + Z_t)Z_t] = \phi E[X_{t-1}Z_t] + E(Z_t^2) = \sigma_Z^2$$

We get the system of equations

$$\begin{cases}\gamma_X(0)\phi = \gamma_X(1)\\ \sigma_Z^2 = \gamma_X(0) - \phi\gamma_X(1)\end{cases}$$

Now, for AR(2),

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

Multiplying X_{t-2}, X_{t-1}, X_t, we get

$$X_t X_{t-2} - \phi_1 X_{t-1}X_{t-2} - \phi_2 X_{t-2}^2 = Z_t X_{t-2}$$

$$X_tX_{t-1} - \phi_1 X_{t-2}^2 - \phi_2 X_{t-2}X_{t-1} = Z_tX_{t-1}$$
$$X_t^2 - \phi_1 X_{t-1}X_t - \phi_2 X_{t-1}X_t = X_tZ_t$$

Applying the expected value, we get

$$\gamma_X(1) - \phi_1\gamma_X(1) - \phi_2\gamma_X(0) = 0$$
$$\gamma_X(1) - \phi_1\gamma_X(0) - \phi_2\gamma_X(1) = 0$$
$$\gamma_X(0) - \phi_1\gamma_X(1) - \phi_2\gamma_X(2) = \sigma_Z^2$$

The variance-covariance matrix is

$$\gamma_p = [\gamma_X(i-j)]_{i,j=1}^p$$

With vectors

$$\vec{\phi}_p = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix},\ \gamma_{X,p} = \begin{bmatrix} \gamma_X(1) \\ \vdots \\ \gamma_X(p) \end{bmatrix}$$

Recall that for $p = 1$, $\Gamma_1 = \gamma_X(0)$, and for $p = 2$

$$\Gamma_2 = \begin{bmatrix} \gamma_X(0) & \gamma_X(1) \\ \gamma_X(-1) & \gamma_X(0) \end{bmatrix} = \begin{bmatrix} \gamma_X(0) & \gamma_X(1) \\ \gamma_X(1) & \gamma_X(0) \end{bmatrix}$$

Then, we can rewrite the system of equations as

$$\Gamma_p\vec{\phi}_p = \gamma_{X,p},\ \sigma_Z^2 = \gamma_X(0) - \vec{\phi}_p^t\gamma_{X,p}$$

These equations Yule-Walker equations, and they involve the autocovariances that are unknown. These equations are combined with the method of moments, so the means are replaced my the sample means and and the covariance is replaced with the sample autocovariance. Thus, the Yule-Walker estimators are

$$\hat{\vec{\phi}}_p = \hat{\Gamma}_p^{-1}\hat{\gamma}_{X,p},\ \hat{\sigma}_Z^2 = \hat{\gamma}_X(0) - \hat{\vec{\phi}}_p^t\hat{\gamma}_{X,p}$$

Theorem 5.2.1. *For a large n, the Yule-Walker estimators are approximately normally distributed*

$$\hat{\vec{\phi}}_p \sim N\left(\vec{\phi}_p, \frac{1}{n}\sigma_Z^2\Gamma_p^{-1}\right)$$

For $p = 1$,

$$\hat{\phi} \sim N\left(\phi, \frac{1}{n}\sigma_Z^2\gamma_X^{-1}\right)$$

That is,

$$\text{Var}(\hat{\phi}) \sim \frac{1}{n}\sigma_Z^2\gamma_X^{-1}(0)$$

Example 5.2.1 (Confidence Intervals for AR(1))**.** When $p = 1$, the confidence interval for the parameter ϕ of AR(1) is

$$\hat{\phi} \pm z_{\alpha/2}\frac{1}{\sqrt{n}}\sigma_Z\sqrt{\gamma_X^{-1}(1)}$$

Here we use the sample estimates for σ_Z^2 and $\gamma_X(0)$. Hence, the confidence intervals for ϕ are

$$\hat{\phi} \pm z_{\alpha/2}\frac{1}{\sqrt{n}}\hat{\sigma}_Z\sqrt{\gamma_X(0)^{-1}}$$

with

$$\hat{\phi} = \frac{\hat{\gamma}_X(1)}{\hat{\gamma}_X(0)},\ \hat{\sigma}_Z^2 = \hat{\gamma}_X(0) - \hat{\phi}\hat{\gamma}_X(1)$$

Example 5.2.2. For the AR(2) model,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

The confidence intervals for ϕ_1 and ϕ_2 are

$$\hat{\phi}_1 \pm z_{\alpha/2}\frac{1}{\sqrt{n}}\sqrt{1-\hat{\phi}_1^2},\ \hat{\phi}_2 \pm z_{\alpha/2}\frac{1}{\sqrt{n}}\sqrt{1-\hat{\phi}_2^2}$$

Chapter 6

Diagnostic Measures

Given an AR(1) model where rwe estimate ϕ and σ_Z^2 with $\hat{\phi}$ and $\hat{\sigma}_Z^2$, we can define the **residuals** as

$$\hat{Z}_t = X_t - \hat{\phi} X_{t-1}$$

We will examine the residuals to see if there is any significant autocorrelation so that we need to check for higher order ARMA processes. We want the autocorrelation of the residuals to be near zero.

- If the random variables are iid, then the correlations at any lag $h \neq 0$ is 0. However the sample correlations are not typically 0. The sample correlation at any lag is approximately normally distributed with mean 0 and variance $\frac{1}{n}$ for large n. The confidence intervals for the sample autocorrelation is $\pm z_{\alpha/2} \frac{1}{\sqrt{n}}$. Whenever the sample autocorrelation is within these confidence intervals, we treat it as 0.
- Let h be a positive integer, we define the Ljung-Box statistic as

  $$Q_h = n \sum_{j=1}^{h} \frac{\hat{\gamma}_X(j)}{\hat{\gamma}_X(0)}$$

  Under the iid assumption, $Q_h \sim \chi^2$ with h degrees of freedom. A large value of Q_h suggests that the sample autocorrelations of the data are too large for the data to be a sample from an iid sequence. We reject the iid hypothesis if $Q_h > \chi^2_{\alpha,h}$. If the iid hypothesis is rejcted, then the fitted model is not correct.

Chapter 7

Maximum Likelihood Estimation

Definition 7.0.1. *The likelihood function of a sequence of iid random variables $\{X_n\}$ with pdf $f(X;\theta)$ is*

$$L(\theta) = L(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} f(X_i;\theta)$$

The *log-likelihood function* is defined as

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(X_i;\theta)$$

Definition 7.0.2. *Let*

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \ell(\theta)$$

Example 7.0.1. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(\beta)$. Recall that the pdf of an exponential distribution is

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The likelihood function is

$$L(\beta) = \beta^{-n} \prod_{i=1}^{n} \exp(-X_i/\beta) = \beta^{-n} \exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} X_i\right)$$

The log-likelihood function is

$$\ell(\beta) = -n\log(\beta) - \frac{1}{\beta}\sum_{i=1}^{n} X_i$$

Now we can find the MLE of β by taking the derivative of $\ell(\beta)$ and setting it to 0.

$$\frac{\partial \ell(\beta)}{\partial \beta} = -n\frac{1}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} X_i = 0 \implies \hat{\beta}_{\text{MLE}} = \bar{X}$$

7.1 Maximum Likelihood Estimation for Time Series Models

Let $X_1, \ldots, X_n$ be obserations from a stationary time series, and $f_n(x_1, \ldots, x_n)$ be their joint density. We assume that the time series is Guassian and centered which is an assumption that needs to be checked. We will define the following vectors

$$X_n = (X_1, \ldots, X_n)', \ \hat{X}_n = (\hat{X}_1, \ldots, \hat{X}_n)', \ U_n = (U_1, \ldots, X_n)'$$

where $U_i = X_i - \hat{X}_i$ are known as the **innovations**. The covariance metrix is defined as $\Gamma_n = E(X_n' X_n)$. The likelihood function is

$$L = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Gamma_n|^{1/2}} \exp\left(-\frac{1}{2} X_n' \Gamma_n^{-1} X_n\right)$$