Chapter 3

An Introduction to Time Series Models

3.1 Objectives[1]

The purpose of this introduction to Autoregressive Moving Averages (ARMA) time series is to provide enough background to the reader to understand and appreciate the more advanced models in later chapters. For a more complete discussion on ARMA time series see Ferreira [16].

3.2 Preliminaries

3.2.1 White Noise

A white noise series is often part of a time series in the form of an "error", an unpredictable randomness.

Definition 3.2.1 *A white noise series* (ε_t) *has the following characteristics for every* $t, s \in \mathbb{R}$

1. $E[\varepsilon_t] = 0$
2. $E[\varepsilon_t^2] = \sigma^2$
3. $E[\varepsilon_t \varepsilon_s] = 0$ *for* $s \neq t$

The white noise process is thus stationary.

[1] Suggested reading: [1] and [18]

3.2.2 Linear Time Series

Definition 3.2.2 $(\Omega, \mathcal{F}, P)$. *A linear time series at time t consists of a $\mathcal{F}_{t-1}$ predictable part plus a random part, that is for a time series*

$$Z_t = E[Z_t \mid \mathcal{F}_{t-1}] + \nu_t$$

where the expected value of the white noise process, ν_t where

$$E[\nu_t \mid \mathcal{F}_{t-1}] = 0$$

3.2.3 Lag Operators and Difference Operators

Definition 3.2.3 *A lag operator L is defined by*

$$L^k Z_t = Z_{t-k}$$

for all $k \in \mathbb{R}^+$.

Definition 3.2.4 *A difference operator Δ is defined by*

$$\Delta_k Z_t = Z_t - Z_{t-k}$$

for all $k \in \mathbb{R}^+$.

Example 3.2.5 *The power of a difference operator Δ^k is different from a higher order difference operator Δ_k.*

$$\begin{aligned} \Delta^2 Z_t &= \Delta(Z_t - Z_{t-1}) \\ &= \Delta Z_t - \Delta Z_{t-1} \\ &= Z_t - 2Z_{t-1} + Z_{t-2} \end{aligned}$$

Definition 3.2.6 *Invertibility of a time series: A time series (Z_t) is invertible if it is possible to write it in terms of an infinite combination of lags.*

3.3 Autoregressive Process (AR)

Definition 3.3.1 *For a stochastic process (Z_t) and white noise process (ε_t), the $AR(p)$ process is defined by*

$$\Phi_p(L) Z_t = \varepsilon_t$$

with

$$\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \tag{3.1}$$

L is a lag operator and p the order of the autoregression polynomial 3.1.

The $AR(p)$ process (Z_t) can thus be written as

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t$$

3.4 Moving Averages Process (MA)

Definition 3.4.1 *For a stochastic process* (Z_t) *and white noise process* (ε_t)*, the* $MA(q)$ *process is defined by*

$$Z_t = \Theta_q(L)\varepsilon_t$$

with

$$\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q \tag{3.2}$$

where L *is a lag operator and* q *the order of the moving averages polynomial 3.2.*

The $MA(q)$ process (Z_t) can thus be written as the sum of past errors

$$Z_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + .. + \theta_q\varepsilon_{t-q} \tag{3.3}$$

The lag operator thus acts on the white noise process not on Z_t.

3.5 Autoregressive Moving Averages (ARMA)

Definition 3.5.1 *For a stochastic process* (Z_t) *and white noise process* (ε_t)*, the* $ARMA(p,q)$ *process is defined by*

$$\Phi_p(L) Z_t = \Theta_q(L)\varepsilon_t$$

with

$$\begin{aligned} \Theta_q(L) &= 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q \\ \Phi_p(L) &= 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \end{aligned}$$

where L *is a lag operator,* p *the order of the autoregression polynomial and* q *the order of the moving averages polynomial*

The $ARMA(p,q)$ process (Z_t) is

$$\begin{aligned} Z_t &= \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} \\ &= \sum_{i=1}^{p}\phi_i Z_{t-i} + \sum_{i=1}^{q}\theta_i\varepsilon_{t-i} \end{aligned}$$

where $\theta_0 = 1$. It is clear that the $ARMA(p,q)$ process, is a combination of an $AR(p)$ and an $MA(q)$ process.

3.6 Stationarity of ARMA Processes

The results in this section was proved in Ferreira [16].

An $MA(\infty)$ process

$$Z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots$$

is stable if and only if its weights are square summable

$$\sum_{i=0}^{\infty} \theta_i^2 < \infty$$

The $AR(p)$ process

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t \tag{3.4}$$

can be rewritten in terms of the *Vector Autoregressive process* denoted by $VAR(1)$

$$\begin{bmatrix} Z_t \\ Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-(p-1)} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ Z_{t-2} \\ Z_{t-3} \\ \vdots \\ Z_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

From this equation we can obtain

$$\boldsymbol{\xi}_t = \mathbf{F}^t \boldsymbol{\xi}_0 + \mathbf{F}^{\mathbf{t-1}} \boldsymbol{\varepsilon}_1 + \ldots + \mathbf{F} \boldsymbol{\varepsilon}_{t-1} + \mathbf{v}_t$$

Theorem 3.6.1 *If all eigenvalues of the matrix* $\mathbf{F}$ *lie within the unit circle,* $|\lambda| < 1$, *then*

$$\sum_{j=0}^{\infty} \mathbf{F}^j = (\mathbf{I} - \mathbf{F})^{-1} \tag{3.5}$$

where $\mathbf{I}$ *is the applicable identity matrix and the right-hand side of equation 3.5 is the inverse of* $\mathbf{I} - \mathbf{F}$.

Proof. Ferreira [16]. ∎

Theorem 3.6.2 *If all the eigenvalues of the $p \times p$ matrix* $\mathbf{F}$ *lie within the unit circle, then*

$$(\mathbf{I}_p - \mathbf{F})^{-1}$$

exists and its element $(1,1)$ *is*

$$\frac{1}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

Proof. Ferreira [16]. ■

Corollary 3.6.3 *If all the eigenvalues of* $\mathbf{F}$ *are less than 1 in magnitude then* $\mathbf{F}^j$ *decays to zero as* j *increases to infinite. A time series with such a property is said to be stable.*

Process 3.4 can be rewritten as

$$\Phi(L)\, Z_t = \varepsilon_t$$

where

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \tag{3.6}$$

Definition 3.6.4 *The characteristic function of the process 3.6 is defined by*

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \tag{3.7}$$

We can then combine the ideas of the root of polynomial 3.7 and the eigenvalues of $\mathbf{F}$.

Theorem 3.6.5 *Factoring the characteristic function is equivalent to finding the eigenvalues of the matrix* $\mathbf{F}$

$$1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L)$$

Proof. Ferreira [16]. ■

Corollary 3.6.6 *The process 3.4 is stable if all the eigenvalues of* $\mathbf{F}$ *all lie inside of the unit circle.*

Theorem 3.6.7 *The characteristic function* $\Phi(L)$ *of an* $AR(p)$ *process can be written in terms of a characteristic function of a* $MA(\infty)$ *process, say* $\pi(L)$

$$\Phi(L) = \pi(L)^{-1}$$

Remark 3.6.8 *Note that only* $\Phi(L)$*, the characteristic function of the autoregressive terms influence stability.*

The results of this section is summarized as follows:

Summary 3.6.9 *An* $AR(p)$ *process is stationary if and only if the eigenvalues of the characteristic function of that process lie inside the unit circle.*

3.7 Estimation of ARMA Parameters

This section focusses on the *maximum likelihood estimation* (MLE) of the ARMA regression model. If we assume that the error process

$$\varepsilon_t = Z_t - \left(\phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}\right)$$

is normally distributed. Then the likelihood function of the ARMA process is

$$\begin{aligned} f^*(\boldsymbol{\theta}) &= \prod_{i=p+1}^{n} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\left[\varepsilon_i/\sigma_\varepsilon\right]^2/2\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}\varepsilon_i^2\right) \end{aligned}$$

where σ_ε^2 is the unconditional (stationary) variance of the error process (ε_t). The product is from the $(p+1)^{th}$ observation to the n^{th} since there are p parameters. Define $n' = n - p$.

Define the parameters matrix by

$$\boldsymbol{\theta} = \left(\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \ldots, \theta_q\right)'$$

The loglikelihood function (the ln of $f^*(\boldsymbol{\theta})$) is

$$f(\boldsymbol{\theta}) = -\frac{1}{2\sigma_\varepsilon^2}\sum_{i=p+1}^{n}\varepsilon_i^2.$$

The MLE parameters are those that maximizes $f^*(\boldsymbol{\theta})$ or $f(\boldsymbol{\theta})$ over a number of observations of (ε_t). Since only the error process is variate in terms of the parameters $\boldsymbol{\theta}$, maximizing $f(\boldsymbol{\theta})$ is equivalent to minimizing $\sum_{i=p+1}^{n}\varepsilon_i^2$.

To comment on the *significance* of the MLE parameter fit, define the information matrix

$$\mathbf{I} = -\lim_{n\to\infty} E\left[\frac{1}{n'}\frac{\partial^2 f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]$$

The asymptotic distribution of MLE estimators is

$$\boldsymbol{\theta} \sim N\left(\boldsymbol{\theta}_0, \frac{1}{n'}\mathbf{I}^{-1}\right)$$

with $\mathbf{I}$ positive definite in the region of the optimal $\boldsymbol{\theta}_0$.

For the second derivative of $f(\boldsymbol{\theta})$ define

$$\mathbf{S} = \frac{\partial^2 f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -\frac{1}{2\sigma_\varepsilon^2}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\sum_{i=p+1}^{n}\varepsilon_i^2$$

thus we can approximate the covariance matrix of $\boldsymbol{\theta}$,

$$\begin{aligned} var\left(\boldsymbol{\theta}\right) &= \frac{1}{n'}\mathbf{I}^{-1} \\ &\approx 2\sigma_\varepsilon^2\mathbf{S} \end{aligned}$$