Chapter 4

ARMA AND VECTOR AUTOREGRESSION REPRESENTATIONS

4.1 Autocorrelation

The Wold representation of a univariate process $\{X_t : -\infty < t < \infty\}$ provides us with a description of how future values of X_t depend on its current and past values (in the sense of linear projections). A useful description of this dependence is *autocorrelation*. The j-th autocorrelation of a process (denoted by ρ_j) is defined as the correlation between X_t and X_{t-j}:

$$Corr(X_t, X_{t-j}) = \frac{Cov(X_t, X_{t-j})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t-j})}}.$$

In general, ρ_j depends on t. If the process is covariance stationary, ρ_j does not depend on t, and is equal to its j-th autocovariance divided by its variance:

$$\rho_j = \frac{\gamma_j}{\gamma_0}, \tag{4.1}$$

where $\gamma_j = Cov(X_t, X_{t-j})$ is the j-th autocovariance, and $\gamma_0 = Var(X_t)$. For covariance stationary processes, $\gamma_j = \gamma_{-j}$, hence $\rho_j = \rho_{-j}$. When we view ρ_j as a function of j, it is called the autocorrelation function. Note that $\rho_0 = 1$ for any process by definition. For a white noise process, $\rho_j = 0$ for $j \neq 0$. The autocorrelation function is a population concept, and can be estimated by its sample counterpart as explained in Chapter 5.

4.2 The Lag Operator

In order to study ARMA representations, it is convenient to use the *lag operator*, denoted by the symbol L. When the operator is applied to a sequence $\{X_t : -\infty < t < \infty\}$ of real numbers, it results in a new sequence $\{Y_t : -\infty < t < \infty\}$, where the value of Y at date t is equal to the value X at date $t-1$:

$$Y_t = X_{t-1},$$

and we write

$$LX_t = X_{t-1}. \tag{4.2}$$

When we apply the lag operator to a univariate stochastic process $\{X_t : -\infty < t < \infty\}$, the lag operator is applied to all sequences of real numbers $\{X_t(\omega) : -\infty < t < \infty\}$ given by fixing the state of the world ω to generate a new stochastic process $\{X_t : -\infty < t < \infty\}$ that satisfies $X_{t-1}(\omega) = LX_t(\omega)$ for each ω.

When the lag operator is applied twice to a process $\{X_t : -\infty < t < \infty\}$, we write $L^2X_t = X_{t-2}$. In general, for any integer $k > 0$, $L^kX_t = X_{t-k}$. It is convenient to define $L^0 = 1$ as the identity operator that gives $L^0X_t = X_t$, and to define L^{-k} as the operator that moves the sequence forward: $L^{-k}X_t = X_{t+k}$ for any integer $k > 0$.

We define a *p-th order polynomial in the lag operator* $B(L) = B_0 + B_1L + B_2L^2 + \cdots + B_pL^p$, where $B_1, \cdots, B_p$ are real numbers, as the operator that yields

$$B(L)X_t = (B_0 + B_1L + B_2L^2 + \cdots + B_pL^p)X_t = B_0X_t + B_1X_{t-1} + \cdots + B_pX_{t-p}.$$

When an infinite sum $B_0X_t + B_1X_{t-1} + B_2X_{t-2} + \cdots$ converges in some sense (such as convergence in L^2),*(?????? Need to use other expressions instead of L^2 because we use L for the lag operator in this paragraph)* we write $B(L) = B_0+B_1L+B_2L^2+\cdots$, and

Masao needs to check this!

$$B(L)X_t = (B_0 + B_1L + B_2L^2 + \cdots)X_t = B_0X_t + B_1X_{t-1} + B_2X_{t-2} + \cdots.$$

For a vector stochastic process $\{\mathbf{X}_t : -\infty < t < \infty\}$, a polynomial in the lag operator $\mathbf{B}_0 + \mathbf{B}_1L + \mathbf{B}_2L^2 + \cdots + \mathbf{B}_pL^p$ for matrices $\mathbf{B}_0, \cdots, \mathbf{B}_p$ with real numbers is used in the same way, so that

$$(\mathbf{B}_0 + \mathbf{B}_1L + \mathbf{B}_2L^2 + \cdots + \mathbf{B}_pL^p)\mathbf{X}_t = \mathbf{B}_0\mathbf{X}_t + \mathbf{B}_1\mathbf{X}_{t-1} + \cdots + \mathbf{B}_p\mathbf{X}_{t-p}.$$

Using the lag operator, $\mathbf{X}_t = \mathbf{\Phi}_0\mathbf{e}_t + \mathbf{\Phi}_1\mathbf{e}_{t-1} + \cdots$ can be expressed as

$$\mathbf{X}_t = \mathbf{\Phi}(L)\mathbf{e}_t, \tag{4.3}$$

where $\mathbf{\Phi}(L) = \mathbf{\Phi}_0 + \mathbf{\Phi}_1L + \mathbf{\Phi}_2L^2 + \cdots$.

4.3 Moving Average Representation

If X_t is linearly regular and covariance stationary with mean μ, then it has a Moving Average (MA) representation of the form $X_t = \mu + \Phi(L)e_t$ or

$$X_t = \mu + \Phi_0e_t + \Phi_1e_{t-1} + \Phi_2e_{t-2} + \cdots, \tag{4.4}$$

where $\Phi_0 = 1$. If $\Phi(L)$ is a polynomial of infinite order, X_t is a moving average process of infinite order (denoted MA(∞)). If $\Phi(L)$ is a polynomial of order q, X_t is a moving average process of order q (denoted MA(q)). In this section, we study how some properties of X_t depend on $\Phi(L)$.

An MA(1) process X_t has a representation $X_t = \mu + e_t + \Phi e_{t-1}$ as in Example 2.8, where e_t is a white noise process that satisfies (2.10), and μ and Φ are constants. The mean, variance, and autocovariance of this process are given in Example 2.8, $E(X_t) = \mu$, and its k-th autocorrelation is $\rho_k = \frac{\Phi}{1+\Phi^2}$ if $|k| = 1$, and $\rho_k = 0$ if $|k| > 1$.

An MA(q) process X_t satisfies

$$X_t = \mu + e_t + \Phi_1 e_{t-1} + \cdots + \Phi_q e_{t-q}, \tag{4.5}$$

where e_t is a white noise process that satisfies (2.10), and μ and $\Phi_1, \cdots, \Phi_q$ are real numbers. A moving average process is covariance stationary for any $(\Phi_1, \cdots, \Phi_q)$.[1] Using (2.10), we obtain the mean of an MA(q) process:

$$E(X_t) = \mu, \tag{4.6}$$

its variance:

$$\gamma_0 = E[(X_t - \mu)^2] = \sigma^2(1 + \Phi_1^2 + \cdots + \Phi_q^2), \tag{4.7}$$

and its j-th autocovariance:

$$\begin{aligned} \gamma_j &= E[(X_t - \mu)(X_{t-j} - \mu)] \\ &= \begin{cases} \sigma^2(\Phi_j + \Phi_{j+1}\Phi_1 + \cdots + \Phi_q \Phi_{q-j}) & \text{for} \quad |j| \le q \\ 0 & \text{for} \quad |j| > q \end{cases}. \end{aligned} \tag{4.8}$$

Hence the j-th autocorrelation of an MA(q) process is zero when $|j| > q$.

When a vector stochastic process $\{\cdots, \mathbf{X}_{-2}, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots\}$ can be written as

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathbf{\Phi}_0 \mathbf{e}_t + \mathbf{\Phi}_1 \mathbf{e}_{t-1} + \cdots + \mathbf{\Phi}_q \mathbf{e}_{t-q}, \tag{4.9}$$

[1] We often impose conditions on $(\Phi_1, \cdots, \Phi_q)$ as we will discuss later in this chapter.

for a white noise process $\mathbf{e}_t$, then $\mathbf{X}_t$ has a *q-th order (one-sided) moving average (MA(q)) representation.* For any $\mathbf{\Phi}_0, \cdots, \mathbf{\Phi}_q$, a process with MA($q$) representation is covariance stationary. As q goes to infinity, an MA(∞) representation

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathbf{\Phi}_0\mathbf{e}_t + \mathbf{\Phi}_1\mathbf{e}_{t-1} + \cdots \tag{4.10}$$

is well defined and covariance stationary if $\sum_{j=0}^{\infty} |\mathbf{\Phi}_j^i|^2 < \infty$ for the i-th row of $\mathbf{\Phi}_j$, $\mathbf{\Phi}_j^i$. In this case, $\mathbf{X}_t$ has a moving average representation of infinite order.

4.4 The Wold Representation

Let $\{\cdots, \mathbf{X}_{-2}, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots\}$ be a covariance stationary n-dimensional vector process with mean zero. Let H_t be the linear information set generated by the current and past values of $\mathbf{X}_t$.[2] We use the notation, $\hat{E}(y|\mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \cdots)$ for $\hat{E}(y|\mathrm{H}_t)$. Note that the information set grows larger over time and the sequence $\{\mathrm{H}_t : -\infty < t < \infty\}$ is increasing in the sense that $\mathrm{H}_t \subset \mathrm{H}_{t+1}$ for all t. Let $\mathrm{H}_{-\infty}$ be the set of random variables that are in H_t for all t: $\mathrm{H}_{-\infty} = \bigcap_{n=1}^{\infty} \mathrm{H}_{t-n}$. Then $0 = \mathbf{0}'\mathbf{X}_t$ is a member of H_t. Therefore, the constant zero is always a member of $\mathrm{H}_{-\infty}$. The stochastic process $\mathbf{X}_t$ is *linearly regular* if $\mathrm{H}_{-\infty}$ contains only the constant zero when $\mathrm{H}_{-\infty} = \bigcap_{n=1}^{\infty} \mathrm{H}_{t-n}$, in which H_t is generated by the current and past values of $\mathbf{X}_t$. The stochastic process $\mathbf{X}_t$ is *linearly deterministic* if $\mathrm{H}_t = \mathrm{H}_{-\infty}$ for all t. For example, if $\mathbf{X}_t$ is an n-dimensional vector of constants, then $\mathbf{X}_t$ is linearly deterministic.

We can now state the Wold decomposition theorem, which states that any covariance stationary process can be decomposed into linearly regular and linearly deterministic components:

[2] We only define the linear information set for a finite number of random variables. See Appendix 3.A for further explanation.

Proposition 4.1 *(The Wold Decomposition Theorem)* Let $\{\cdots, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots\}$ be a covariance stationary vector process with mean zero. Then it can be written as

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{e}_{t-j} + \mathbf{g}_t, \tag{4.11}$$

where $\mathbf{\Phi}_0 = \mathbf{I}_n$, $\sum_{j=0}^{\infty} |\mathbf{\Phi}_j^i|^2 < \infty$ for the i-th row of $\mathbf{\Phi}_j$, $\mathbf{\Phi}_j^i$, and

$$\mathbf{e}_t = \mathbf{X}_t - \hat{E}(\mathbf{X}_t|\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \mathbf{X}_{t-3}, \cdots) \tag{4.12}$$

and

$$\mathbf{g}_t = \hat{E}(\mathbf{X}_t|\mathrm{H}_{-\infty}). \tag{4.13}$$

∎

It can be shown that $\sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{e}_{t-j}$ is a linearly regular covariance stationary process and $\mathbf{g}_t$ is linearly deterministic. Hence if $\mathbf{X}_t$ is not linearly regular, it is possible to remove $\mathbf{g}_t$ and work with a linearly regular process as long as $\mathbf{g}_t$ can be estimated.

Proposition 4.2 *(The Wold Representation)* Let $\{\cdots, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots\}$ be a linearly regular covariance stationary vector process with mean zero. Then it can be written as

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{e}_{t-j}, \tag{4.14}$$

where $\mathbf{\Phi}_0 = \mathbf{I}_n$, $\sum_{j=0}^{\infty} |\mathbf{\Phi}_j^i|^2 < \infty$ for the i-th row of $\mathbf{\Phi}_j$, $\mathbf{\Phi}_j^i$, and $\mathbf{e}_t$ is defined by (4.12). ∎

The Wold representation gives a unique MA representation when the MA innovation $\mathbf{e}_t$ is restricted to the form given by Equation (4.12). There may exist infinitely many other MA representations when the MA innovation is not restricted to be given by (4.12) as we will discuss below.

In many macroeconomic models, stochastic processes that we observe (real GDP, interest rates, stock prices, etc.) are considered to be generated from the nonlinear function of underlying shocks. In this sense, the processes in these models are nonlinear, but Proposition 4.1 states that even a nonlinear stochastic process has a linear moving average representation as long as it is linearly regular and covariance stationary.

In order to give a sketch of a proof of the Wold Representation Theorem, consider a linearly regular stochastic process $\{X_t\}_{-\infty}^{\infty}$ that may not be necessarily a linear function of underlying shocks. Define $u_t = X_t - \hat{E}(X_t|\mathrm{H}_{t-1})$, and

Youngsoo needs to check this!

$$\mathrm{U_t} = \{z|z = bu_t \text{ for } b \in \mathrm{R}\} \tag{4.15}$$

where $\mathrm{H_t}$ is the linear information set generated by the current and past values of X_t. Then, we have the following relationship

$$\mathrm{H_t} = \mathrm{H_{t-1}} + \mathrm{U_t}, \tag{4.16}$$

and each element of $\mathrm{H_t}$ is orthogonal to each element of $\mathrm{U_t}$. In this case,

$$\begin{aligned} \hat{E}(h|\mathrm{H_t}) &= \hat{E}(h|\mathrm{H_{t-1}} + \mathrm{U_t}) \\ &= \hat{E}(h|\mathrm{H_{t-1}}) + \hat{E}(h|\mathrm{U_t}) \end{aligned} \tag{4.17}$$

for any h. Because $\mathrm{H_{t-1}} = \mathrm{H_{t-2}} + \mathrm{U_{t-1}}$, we have

$$\mathrm{H_t} = \mathrm{H_{t-2}} + \mathrm{U_t} + \mathrm{U_{t-1}}, \tag{4.18}$$

and by continuing this process, we have

$$\mathrm{H_t} = \sum_{j=0}^{\infty} \mathrm{U_{t-j}}. \tag{4.19}$$

Therefore, X_t can be written as

$$X_t = \hat{E}(X_t|\mathrm{H_t}) = \hat{E}(X_t|\sum_{j=0}^{\infty}\mathrm{U_{t-j}}) = \sum_{j=0}^{\infty}\Phi_j u_{t-j}, \tag{4.20}$$

which is the Wold representation of X_t.

Example 4.1 Suppose that u_t is a Gaussian white noise with variance of 1. Let $X_t = u_t^2 - 1$. Then the Wold representation of X_t is $X_t = e_t$, where $e_t = u_t^2 - 1$. ∎

In this example, X_t is a nonlinear transformation of a Gaussian white noise. The shock that generates X_t, u_t, is normally distributed. However, the innovation in its Wold representation, e_t, is not normally distributed. Thus, the innovation in the Wold representation of a process can have a different distribution from the underlying shock that generates the process.

Even when the underlying shocks that generate processes are i.i.d., the innovations in the Wold representation may not be i.i.d. as in the next example.

Example 4.2 Suppose that u_t is an i.i.d Gaussian white noise with variance of 1, so that $E(u_t^3) = 0$. Let X_t be generated by $X_t = u_t + \Phi(u_{t-1}^2 - 1)$. Then $E(X_t X_{t-1}) = E[u_t u_{t-1} + \Phi u_{t-1}^3 - \Phi u_{t-1} + \Phi u_t u_{t-2}^2 - \Phi u_t + \Phi^2(u_{t-1}^2 - 1)(u_{t-2}^2 - 1)] = 0$. Hence the Wold representation of X_t is $X_t = e_t$, where $e_t = u_t + \Phi(u_{t-1}^2 - 1)$. ∎

Note that the Wold representation innovation e_t in this example is serially uncorrelated, but not i.i.d. because $e_t(= u_t + \Phi u_{t-1}^2)$ and $e_{t-1}(= u_{t-1} + \Phi u_{t-2}^2)$ are related nonlinearly through the Φu_{t-1}^2 and u_{t-1} terms.

The Wold representation states that any linearly regular covariance stationary process has an MA representation. Therefore, it is useful to estimate an MA representation in order to study how linear projections of future variables depend on their current and past values. Higher order MA representations and vector MA representations are hard to estimate, however, and it is often convenient to consider AR representations and ARMA representations, which are closely related to MA representations.

4.5 Autoregression Representation

A process X_t, which satisfies $B(L)X_t = \delta + e_t$ with $B_0 = 1$ or

$$X_t + B_1 X_{t-1} + B_2 X_{t-2} + \cdots = \delta + e_t$$

for a white noise process e_t, is an autoregression. If $B(L)$ is a polynomial of infinite order, X_t is an autoregression of infinite order (denoted AR(∞)). If $B(L)$ is a polynomial of order p, X_t is an autoregression of order p (denoted AR(p)). In this section, we study how some properties of X_t depend on $B(L)$.

4.5.1 Autoregression of Order One

Consider a process X_t that satisfies

$$X_t = \delta + BX_{t-1} + e_t \quad \text{for} \quad t \geq 1, \tag{4.21}$$

where e_t is a white noise process with variance σ^2 and X_0 is a random variable that gives an initial condition for (4.21). Such a process is called an *autoregression of order 1*, denoted by AR(1). It is often convenient to consider (4.21) in a deviation-from-the-mean form:

$$X_t - \mu = B(X_{t-1} - \mu) + e_t \quad \text{for} \quad t \geq 1, \tag{4.22}$$

where $\mu = \frac{\delta}{1-B}$. Substituting (4.22) recursively, we obtain $X_1 - \mu = B(X_0 - \mu) + e_1$ and $X_2 - \mu = B(X_1 - \mu) + e_2 = B^2(X_0 - \mu) + Be_1 + e_2$, so that

$$X_t - \mu = B^t(X_0 - \mu) + B^{t-1}e_1 + B^{t-2}e_2 + \cdots + Be_{t-1} + e_t \qquad \text{for } t \geq 1. \tag{4.23}$$

In this way, X_t is defined for any real number B.

Suppose that X_0 is uncorrelated with $e_1, e_2, \cdots$. When the absolute value of B is greater than or equal to one, then the variance of X_t increases over time. Hence X_t cannot be covariance stationary. In macroeconomics, the case in which $B = 1$ is of importance, and will be discussed in detail in Chapter 13.

Consider the case where the absolute value of B is less than one. In this case, $B^t X_0(\omega)$ becomes negligible as t goes to infinity for a fixed ω. As seen in Example 2.9, however, the process X_t is not covariance stationary in general. Whether or not X_t is stationary depends upon the initial condition X_0. In order to choose X_0, consider an MA process

$$X_t = \mu + e_t + Be_{t-1} + B^2e_{t-2} + \cdots , \tag{4.24}$$

and choose the initial condition for the process X_t in (4.21) by

$$X_0 = \mu + e_0 + Be_{-1} + B^2e_{-2} + \cdots . \tag{4.25}$$

When this particular initial condition is chosen, X_t is covariance stationary.

With the lag operator, (4.22) can be written as

$$(1 - BL)(X_t - \mu) = e_t. \tag{4.26}$$

We define the inverse of $(1 - BL)$ as

$$(1 - BL)^{-1} = 1 + BL + B^2L^2 + B^3L^3 + \cdots , \tag{4.27}$$

when the absolute value of B is less than one. When a process X_t has an MA representation of the form (4.24), we write

$$X_t = \mu + (1 - BL)^{-1} e_t, \tag{4.28}$$

which is the MA(∞) representation of an AR(1) process.

4.5.2 The p-th Order Autoregression

A p-th order autoregression satisfies

$$X_t = \delta + B_1 X_{t-1} + B_2 X_{t-2} + \cdots + B_p X_{t-p} + e_t \qquad \text{for} \ \ t \geq 1. \tag{4.29}$$

The stability condition is that all the roots of

$$1 - B_1 z - B_2 z^2 - \cdots - B_p z^p = 0 \tag{4.30}$$

are larger than one in absolute value, or equivalently, all the roots of

$$z^p - B_1 z^{p-1} - B_2 z^{p-2} - \cdots - B_p = 0 \tag{4.31}$$

are smaller than one in absolute value.

Consider, for instance, the special case of a AR(1) process with $B_1 = 1$ and $X_0 = 0$:

$$\begin{aligned} X_t &= X_{t-1} + e_t & (4.32)\\ &= e_1 + e_2 + \cdots + e_{t-1} + e_t \qquad \text{for} \ \ t \geq 1, & (4.33) \end{aligned}$$

where $E(X_t) = 0$ and $E(X_{t-i} X_{t-j}) = \sigma^2$ for $i = j$. Note that $Var(X_1) = \sigma^2$, $Var(X_2) = 2\sigma^2$, $\cdots$, $Var(X_t) = t\sigma^2$. Since the variance of X_t varies over time, X_t is nonstationary. Note also that its first difference is stationary since $e_t (= X_t - X_{t-1})$

is stationary. Such a process is called *difference stationary*. When a (possibly infinite order) polynomial in the lag operator $\Phi(L) = \Phi_0 + \Phi_1 L + \Phi_2 L^2 + \cdots$ is given, we consider a complex valued function $\Phi(z^{-1}) = \Phi_0 + \Phi_1 z^{-1} + \Phi_2 z^{-2} + \cdots$ by replacing the lag operator L by a complex number z. Consider a condition

$$\Phi(z) = \Phi_0 + \Phi_1 z + \Phi_2 z^2 + \cdots = 0. \tag{4.34}$$

If a complex number z_i satisfies the condition (4.34), then z_i is a *zero* of $\Phi(z)$. We also say that z_i is a root of the equation $\Phi(z) = 0$.

4.6 ARMA

An ARMA(p, q) process satisfies

$$X_t = \delta + B_1 X_{t-1} + B_2 X_{t-2} + \cdots + B_p X_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots . \tag{4.35}$$

If $B(1) = 1 - B_1 - \cdots - B_p \neq 1$, we have the deviation-from-the-mean form

$$B(L)(X_t - \mu) = \theta(L) e_t, \tag{4.36}$$

Kyungho needs to check this!

where $\mu = \frac{\delta}{B(1)}$. We define the inverse of $B(L) = B_0 + B_1 L + \ldots + B_p L^p$ as the lag polynomial $B(L)^{-1}$ such that

$$B(L)^{-1} B(L) = 1. \tag{4.37}$$

As long as $B_0 \neq 0$, $B(L)^{-1}$ exists always. However, $B(L)^{-1}\varepsilon_t$ may or may not be defined. Provided that the p-th order polynomial $B(z)$ satisfies stability conditions, the ARMA(p, q) process yields the MA(∞) representation

$$X_t = \mu + \Phi(L) e_t, \tag{4.38}$$

where $\Phi(L) = B(L)^{-1}\theta(L) = \Phi_0 + \Phi_1 L + \theta_2 L^2 + \cdots$ and $\sum_{j=0}^{\infty} |\theta_j|^2 \leq \infty$.

On the other hand, if $\theta(z)$ satisfies stability conditions that all roots of $\theta(z) = 0$ lie outsize the unit circle, then $\theta(L)$ is invertible and the ARMA(p, q) process yields the AR(∞) representation[3]

$$\theta(L)^{-1}B(L)X_t = \delta^* + e_t, \tag{4.39}$$

where $\delta^* = \frac{\delta}{\theta(1)}$. Therefore, if both $B(z)$ and $\theta(z)$ satisfy stability conditions, then the ARMA(p, q) process has both the MA(∞) and AR(∞) representations.

4.7 Fundamental Innovations

Let $\mathbf{X}_t$ be a covariance stationary vector process with mean zero that is linearly regular. Then the Wold representation in (4.14) gives an MA representation. There are infinitely many other MA representations.

Example 4.3 let u_t be a white noise, and $X_t = u_t$. Then $X_t = u_t$ is an MA representation. Let $u_t^* = u_{t+1}$. Then $X_t = u_{t-1}^*$ is another MA representation. ∎

In this example, another MA representation is obtained by adopting a different dating procedure for the innovation.

It is often convenient to restrict our attention to the MA representations for which the information content of the current and past values of the innovations is the same as that of the current and past values of $\mathbf{X}_t$. Let

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{u}_{t-j} = \mathbf{\Phi}(L)\mathbf{u}_t \tag{4.40}$$

[3]Without any loss of generality, we assume that there are no common roots of $B(z) = 0$ and $\theta(z) = 0$. In such a case, we can write the ARMA(p, q) process by the ARMA$(p - m, q - m)$ process that has no common roots, where m is the number of common roots. See Hayashi (2000, p. 382) for further discussion.

be an MA representation for $\mathbf{X}_t$. Let H_t be the linear information set generated by the current and past values of $\mathbf{X}_t$, and H_t^u be the linear information set generated by the current and past values of $\mathbf{u}_t$. Then $\mathrm{H}_t \subset \mathrm{H}_t^u$ because of (4.40). The innovation process $\mathbf{u}_t$ is said to be *fundamental* if $\mathrm{H}_t = \mathrm{H}_t^u$. The innovation in the Wold representation is fundamental.

In Example 4.3, $X_t = u_t$ is a fundamental MA representation while $X_t = u_{t-1}^*$ is not. As a result of the dating procedure used for $X_t = u_{t-1}^*$, the information set generated by the current and past values of $u_t^* : \{u_t^*, u_{t-1}^*, \cdots\}$ is equal to H_{t+1}, and is strictly larger than the information set generated by H_t.

The concept of fundamental innovations is closely related to the concept of invertibility. If the MA representation (4.40) is invertible, then $\mathbf{u}_t = \mathbf{\Phi}(L)^{-1}\mathbf{X}_t$. Therefore, $\mathrm{H}_t^u \subset \mathrm{H}_t$. Since (4.40) implies $\mathrm{H}_t \subset \mathrm{H}_t^u$, $\mathrm{H}_t = \mathrm{H}_t^u$. Thus if the MA representation (4.40) is invertible, then $\mathbf{u}_t$ is fundamental.

If all the roots of $det[\mathbf{\Phi}(z)] = 0$ lie outside the unit circle, then $\mathbf{\Phi}(L)$ is invertible, and $\mathbf{u}_t$ is fundamental. If all the roots of $det[\mathbf{\Phi}(z)] = 0$ lie on or outside the unit circle, then $\mathbf{\Phi}(L)$ may not be invertible, but $\mathbf{u}_t$ is fundamental. Thus for fundamentalness, we can allow some roots of $det[\mathbf{\Phi}(z)] = 0$ to lie on the unit circle.

In the univariate case, if $X_t = \Phi(L)u_t$ and all the roots of $\Phi(z) = 0$ lie on or outside the unit circle, then u_t is fundamental. For example, let $X_t = u_t + \Phi u_{t-1}$. If $|\Phi| < 1$, then this MA representation is invertible, and u_t is fundamental. If $\Phi = 1$ or if $\Phi = -1$, then this MA representation is not invertible, but u_t is fundamental. If $|\Phi| > 1$, then u_t is not fundamental.

The MA representations with fundamental innovations are useful; it is easier to express projections of variables onto H_t with them than if they had non-fundamental innovations. For example, let X_t be a univariate process with an MA(1) representation: $X_t = u_t + \Phi u_{t-1}$. It is natural to assume that economic agents observe X_t, but not u_t. Therefore, the economic agents' forecast for X_{t+1} can be modeled as $\hat{E}(X_{t+1}|\mathrm{H}_t)$ rather than $\hat{E}(X_{t+1}|\mathrm{H}_t^u)$. If $|\Phi| \leq 1$, u_t is fundamental, and $\hat{E}(X_{t+1}|\mathrm{H}_t) = \hat{E}(X_{t+1}|\mathrm{H}_t^u) = \Phi u_t$. On the other hand, if $|\Phi| > 1$, u_t is not fundamental, and $\hat{E}(X_{t+1}|\mathrm{H}_t) \neq \hat{E}(X_{t+1}|\mathrm{H}_t^u) = \Phi u_t$, and there is no easy way to express $\hat{E}(X_{t+1}|\mathrm{H}_t)$.

4.8 The Spectral Density

Consider a covariance stationary process Y_t such that $Y_t - E(Y_t)$ is linearly regular. Then $Y_t - E(Y_t) = b(L)e_t = \sum_{j=0}^{\infty} b_j e_{t-j}$ for a square summable $\{b_j\}$ and a white noise process e_t such that $E(e_t^2) = 1$ and $E(e_t e_s) = 0$ for $t \neq s$. Its k-th *autocovariance* $\Phi(k) = E[(Y_t - E(Y_t))(Y_{t-k} - E(Y_{t-k})']$ does not depend on date t. For a real number r, define

$$\exp(ir) = \cos(r) + i\sin(r), \tag{4.41}$$

where $i = \sqrt{-1}$. The spectral density of Y_t, $f(\lambda)$ is defined by

$$f(\lambda) = (\sum_{j=0}^{\infty} b_j \exp(-i\lambda j))(\sum_{j=0}^{\infty} b_j \exp(i\lambda j)). \tag{4.42}$$

Then

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Phi(k) \exp(i\lambda k) \tag{4.43}$$

for a real number λ $(-\pi < \lambda < \pi)$ when the autocovariances are absolutely summable. The spectral density is a function of λ, which is called the frequency. Using the properties of the *cos* and *sin* functions and the fact that $\Phi(k) = \Phi(-k)$, it can be shown that

$$f(\lambda) = \frac{1}{2\pi}\Phi(0) + 2\sum_{k=1}^{\infty}\Phi(k)\cos(\lambda k), \tag{4.44}$$

where $f(\lambda) = f(-\lambda)$ and $f(\lambda)$ is nonnegative for all λ.

Equation (4.43) gives the spectral density from the autocovariances. When the spectral density is given, the autocovariances can be calculated form the following formula:

$$\int_{-\pi}^{\pi} f(\lambda)\exp(i\lambda k)d\lambda = \Phi(k). \tag{4.45}$$

Thus the spectral density and the autocovariances contain the same information about the process. In some applications, it is more convenient to examine the spectral density than the autocovariances. For example, it requires infinite space to plot the autocovariance for $k = 0, 1, 2, \cdots$, whereas the spectral density can be concisely plotted.

An interpretation of the spectral density is given by the special case of (4.45) in which $k = 0$:

$$\int_{-\pi}^{\pi} f(\lambda)d\lambda = \Phi(0). \tag{4.46}$$

This relationship suggests an intuitive interpretation that $f(\lambda)$ is the contribution of the frequency λ to the variance of Y_t.

This intuition can be formalized by the *spectral representation theorem* which states that any covariance stationary process Y_t with absolutely summable autocovariances can be expressed in the form

$$Y_t = \mu + \int_{0}^{\pi}[\alpha(\lambda)\cos(\lambda t) + \delta(\lambda)\sin(\lambda t)]d\lambda, \tag{4.47}$$

where $\alpha(\lambda)$ and $\delta(\lambda)$ are random variables with mean zero for any λ in $[0, \pi]$. These variables have the further properties that for any frequencies $0 < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 < \pi$, the variable $\int_{\lambda_1}^{\lambda_2} \alpha(\lambda)$ is uncorrelated with $\int_{\lambda_3}^{\lambda_4} \alpha(\lambda)$, and the variable $\int_{\lambda_1}^{\lambda_2} \delta(\lambda)$ is uncorrelated with $\int_{\lambda_3}^{\lambda_4} \delta(\lambda)$. For any $0 < \lambda_1 < \lambda_2 < \pi$ and $0 < \lambda_3 < \lambda_4 < \pi$, the variable $\int_{\lambda_1}^{\lambda_2} \alpha(\lambda)$ is uncorrelated with $\int_{\lambda_3}^{\lambda_4} \delta(\lambda)$. For such a process, the portion of the variance due to cycles with frequency less than or equal to λ_1 is given by

$$2 \int_0^{\lambda_1} f(\lambda) d\lambda. \tag{4.48}$$

Exercises

4.1 Let u_t be a white noise, and $x_t = u_t + 0.8u_{t-1}$. Is x_t covariance stationary? Is u_t fundamental for x_t? Give an expression for $\hat{E}(x_t|u_{t-1}, u_{t-2}, \cdots)$ in terms of past u_t's. Is it possible to give an expression for $\hat{E}(x_t|x_{t-1}, x_{t-2}, ...)$ in terms of past u_t's? If so, give an expression. Explain your answers.

4.2 Let u_t be a white noise, and $x_t = u_t + 1.2u_{t-1}$. Is x_t covariance stationary? Is u_t fundamental for x_t? Give an expression for $\hat{E}(x_t|u_{t-1}, u_{t-2}, \cdots)$ in terms of past u_t's. Is it possible to give an expression for $\hat{E}(x_t|x_{t-1}, x_{t-2}, \cdots)$ in terms of past u_t's? If so, give an expression. Explain your answers.

4.3 Let u_t be a white noise, and $x_t = u_t + u_{t-1}$. Is x_t covariance stationary? Is u_t fundamental for x_t? Give an expression for $\hat{E}(x_t|u_{t-1}, u_{t-2}, \cdots)$ in terms of past u_t's. Is it possible to give an expression for $\hat{E}(x_t|x_{t-1}, x_{t-2}, \cdots)$ in terms of past u_t's? If so, give an expression. Explain your answers.

References

HAYASHI, F. (2000): *Econometrics.* Princeton University Press, Princeton.