Lecture 2: ARMA Models*

1 ARMA Process

As we have remarked, dependence is very common in time series observations. To model this time series dependence, we start with univariate ARMA models. To motivate the model, basically we can track two lines of thinking. First, for a series x_t, we can model that the level of its current observations depends on the level of its lagged observations. For example, if we observe a high GDP realization this quarter, we would expect that the GDP in the next few quarters are good as well. This way of thinking can be represented by an AR model. The AR(1) (autoregressive of order one) can be written as:

$$x_t = \phi x_{t-1} + \epsilon_t$$

where $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ and we keep this assumption through this lecture. Similarly, AR(p) (autoregressive of order p) can be written as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t.$$

In a second way of thinking, we can model that the observations of a random variable at time t are not only affected by the shock at time t, but also the shocks that have taken place before time t. For example, if we observe a negative shock to the economy, say, a catastrophic earthquake, then we would expect that this negative effect affects the economy not only for the time it takes place, but also for the near future. This kind of thinking can be represented by an MA model. The MA(1) (moving average of order one) and MA(q) (moving average of order q) can be written as

$$x_t = \epsilon_t + \theta \epsilon_{t-1}$$

and

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}.$$

If we combine these two models, we get a general ARMA(p, q) model,

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}.$$

ARMA model provides one of the basic tools in time series modeling. In the next few sections, we will discuss how to draw inferences using a univariate ARMA model.

2 Lag Operators

Lag operators enable us to present an ARMA in a much concise way. Applying lag operator (denoted L) once, we move the index back one time unit; and applying it k times, we move the index back k units.

$$\begin{aligned} Lx_t &= x_{t-1} \\ L^2x_t &= x_{t-2} \\ &\vdots \\ L^kx_t &= x_{t-k} \end{aligned}$$

The lag operator is distributive over the addition operator, i.e.

$$L(x_t + y_t) = x_{t-1} + y_{t-1}$$

Using lag operators, we can rewrite the ARMA models as:

$$\begin{aligned} \text{AR}(1): &\quad (1 - \phi L)x_t = \epsilon_t \\ \text{AR}(p): &\quad (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)x_t = \epsilon_t \\ \text{MA}(1): &\quad x_t = (1 + \theta L)\epsilon_t \\ \text{MA}(q): &\quad x_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\epsilon_t \end{aligned}$$

Let $\phi_0 = 1, \theta_0 = 1$ and define log polynomials

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

With lag polynomials, we can rewrite an ARMA process in a more compact way:

$$\begin{aligned} \text{AR}: &\quad \phi(L)x_t = \epsilon_t \\ \text{MA}: &\quad x_t = \theta(L)\epsilon_t \\ \text{ARMA}: &\quad \phi(L)x_t = \theta(L)\epsilon_t \end{aligned}$$

3 Invertibility

Given a time series probability model, usually we can find multiple ways to represent it. Which representation to choose depends on our problem. For example, to study the impulse-response functions (section 4), MA representations maybe more convenient; while to estimate an ARMA model, AR representations maybe more convenient as usually x_t is observable while ϵ_t is not. However, not all ARMA processes can be inverted. In this section, we will consider under what conditions can we invert an AR model to an MA model and invert an MA model to an AR model. It turns out that *invertibility*, which means that the process can be inverted, is an important property of the model.

If we let 1 denotes the identity operator, i.e., $1y_t = y_t$, then the inversion operator $(1 - \phi L)^{-1}$ is defined to be the operator so that

$$(1 - \phi L)^{-1}(1 - \phi L) = 1$$

For the AR(1) process, if we premulitply $(1-\phi L)^{-1}$ to both sides of the equation, we get

$$x_t = (1-\phi L)^{-1}\epsilon_t$$

Is there any explicit way to rewrite $(1-\phi L)^{-1}$? Yes, and the answer just turns out to be $\theta(L)$ with $\theta_k = \phi^k$ for $|\phi| < 1$. To show this,

$$\begin{aligned}
& (1-\phi L)\theta(L) \\
= \; & (1-\phi L)(1+\theta_1 L + \theta_2 L^2 + \ldots) \\
= \; & (1-\phi L)(1+\phi L + \phi^2 L^2 + \ldots) \\
= \; & 1 - \phi L + \phi L - \phi^2 L^2 + \phi^2 L^2 - \phi^3 L^3 + \ldots \\
= \; & 1 - \lim_{k\to\infty} \phi^k L^k \\
= \; & 1 \quad \text{for} \quad |\phi| < 1
\end{aligned}$$

We can also verify this result by recursive substitution,

$$\begin{aligned}
x_t &= \phi x_{t-1} + \epsilon_t \\
&= \phi^2 x_{t-2} + \epsilon_t + \phi\epsilon_{t-1} \\
&\vdots \\
&= \phi^k x_{t-k} + \epsilon_t + \phi\epsilon_{t-1} + \ldots + \phi^{k-1}\epsilon_{t-k+1} \\
&= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j \epsilon_{t-j}
\end{aligned}$$

With $|\phi| < 1$, we have that $\lim_{k\to\infty} \phi^k x_{t-k} = 0$, so again, we get the moving average representation with MA coefficient equal to ϕ^k. So the condition that $|\phi| < 1$ enables us to invert an AR(1) process to an MA(∞) process,

$$\begin{aligned}
\text{AR}(1): & \quad (1-\phi L)x_t = \epsilon_t \\
\text{MA}(\infty): & \quad x_t = \theta(L)\epsilon_t \quad \text{with} \quad \theta_k = \phi^k
\end{aligned}$$

We have got some nice results in inverting an AR(1) process to a MA(∞) process. Then, how to invert a general AR(p) process? We need to factorize a lag polynomial and then make use of the result that $(1-\phi L)^{-1} = \theta(L)$. For example, let $p = 2$, we have

$$(1 - \phi_1 L - \phi_2 L^2)x_t = \epsilon_t \tag{1}$$

To factorize this polynomial, we need to find roots λ_1 and λ_2 such that

$$(1 - \phi_1 L - \phi_2 L^2) = (1-\lambda_1 L)(1-\lambda_2 L)$$

Given that both $|\lambda_1| < 1$ and $|\lambda_2| < 1$ (or when they are complex number, they lie within the unit circle. Keep this in mind as I may not mention this again in the remaining of the lecture), we could write

$$\begin{aligned}
(1-\lambda_1 L)^{-1} &= \theta_1(L) \\
(1-\lambda_2 L)^{-1} &= \theta_2(L)
\end{aligned}$$

and so to invert (1), we have

$$\begin{aligned} x_t &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\epsilon_t \\ &= \theta_1(L)\theta_2(L)\epsilon_t \end{aligned}$$

Solving $\theta_1(L)\theta_2(L)$ is straightforward,

$$\begin{aligned} \theta_1(L)\theta_2(L) &= (1+\lambda_1 L+\lambda_1^2 L^2+\ldots)(1+\lambda_2 L+\lambda_2^2 L^2+\ldots) \\ &= 1+(\lambda_1+\lambda_2)L+(\lambda_1^2+\lambda_1\lambda_2+\lambda_2^2)L^2+\ldots \\ &= \sum_{k=0}^{\infty}(\sum_{j=0}^{k}\lambda_1^j\lambda_2^{k-j})L^k \\ &= \psi(L), \quad say, \end{aligned}$$

with $\psi_k = \sum_{j=0}^k \lambda_1^j \lambda_2^{k-j}$. Similarly, we can also invert the general AR(p) process given that all roots λ_i has less than one absolute value. An alternative way to represent this MA process (to express ψ) is to make use of partial fractions. Let c_1, c_2 be two constants, and their values are determined by

$$\frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)} = \frac{c_1}{1-\lambda_1 L}+\frac{c_2}{1-\lambda_2 L} = \frac{c_1(1-\lambda_2 L)+c_2(1-\lambda_1 L)}{(1-\lambda_1 L)(1-\lambda_2 L)}$$

We must have

$$\begin{aligned} 1 &= c_1(1-\lambda_2 L)+c_2(1-\lambda_1 L) \\ &= (c_1+c_2)-(c_1\lambda_2+c_2\lambda_1)L \end{aligned}$$

which gives

$$c_1+c_2=1 \quad \text{and} \quad c_1\lambda_2+c_2\lambda_1=0.$$

Solving these two equations we get

$$c_1=\frac{\lambda_1}{\lambda_1-\lambda_2}, \quad c_2=\frac{\lambda_2}{\lambda_2-\lambda_1}.$$

Then we can express x_t as

$$\begin{aligned} x_t &= [(1-\lambda_1 L)(1-\lambda_2 L)]^{-1}\epsilon_t \\ &= c_1(1-\lambda_1 L)^{-1}\epsilon_t+c_2(1-\lambda_2 L)^{-1}\epsilon_t \\ &= c_1\sum_{k=0}^{\infty}\lambda_1^k\epsilon_{t-k}+c_2\sum_{k=0}^{\infty}\lambda_2^k\epsilon_{t-k} \\ &= \sum_{k=0}^{\infty}\psi_k\epsilon_{t-k} \end{aligned}$$

where $\psi_k = c_1\lambda_1^k + c_2\lambda_2^k$.

Similarly, an MA process,

$$x_t = \theta(L)\epsilon_t,$$

is invertible if $\theta(L)^{-1}$ exists. An MA(1) process is invertible if $|\theta| < 1$, and an MA(q) process is invertible if all roots of

$$1 + \theta_1 z + \theta_2 z^2 + \ldots \theta_q z^q = 0$$

lie outside of the unit circle. Note that for any invertible MA process, we can find a noninvertible MA process which is the same as the invertible process up to the second moment. The converse is also true. We will give an example in section 5.

Finally, given an invertible ARMA(p, q) process,

$$\begin{aligned}
\phi(L)x_t &= \theta(L)\epsilon_t \\
x_t &= \phi^{-1}(L)\theta(L)\epsilon_t \\
x_t &= \psi(L)\epsilon_t
\end{aligned}$$

then what is the series ψ_k? Note that since

$$\phi^{-1}(L)\theta(L)\epsilon_t = \psi(L)\epsilon_t,$$

we have $\theta(L) = \phi(L)\psi(L)$. So the elements of ψ can be computed recursively by equating the coefficients of L^k.

Example 1 For a ARMA$(1,1)$ process, we have

$$\begin{aligned}
1 + \theta L &= (1 - \phi L)(\psi_0 + \psi_1 L + \psi_2 L^2 + \ldots) \\
&= \psi_0 + (\psi_1 - \phi\psi_0)L + (\psi_2 - \phi\psi_1)L^2 + \ldots
\end{aligned}$$

Matching coefficients on L^k, we get

$$\begin{aligned}
1 &= \psi_0 \\
\theta &= \psi_1 - \phi\psi_0 \\
0 &= \psi_j - \phi\psi_{j-1} \quad \text{for} \quad j \geq 2
\end{aligned}$$

Solving those equation, we can easily get

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \phi + \theta \\
\psi_j &= \phi^{j-1}(\phi + \theta) \quad \text{for} \quad j \geq 2
\end{aligned}$$

4 Impulse-Response Functions

Given an ARMA model, $\phi(L)x_t = \theta(L)\epsilon_t$, it is natural to ask: what is the effect on x_t given a unit shock at time s (for $s < t$)?

4.1 MA process

For an MA(1) process,

$$x_t = \epsilon_t + \theta\epsilon_{t-1}$$

the effects of ϵ on x are:

$$\begin{array}{lccccc} \epsilon: & 0 & 1 & 0 & 0 & 0 \\ x: & 0 & 1 & \theta & 0 & 0 \end{array}$$

For a MA(q) process,

$$x_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q},$$

the effects on ϵ on x are:

$$\begin{array}{lcccccccc} \epsilon: & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ x: & 0 & 1 & \theta_1 & \theta_2 & \ldots & \theta_q & 0 \end{array}$$

The left figure in Figure 1 plots the impulse-response function of an MA(3) process. Similarly, we can write down the effects for an MA(∞) process. As you can see, we can get impulse-response function immediately from an MA process.

4.2 AR process

For a AR(1) process $x_t = \phi x_{t-1} + \epsilon_t$ with $|\phi| < 1$, we can invert it to a MA process and the effects of ϵ on x are:

$$\begin{array}{lccccc} \epsilon: & 0 & 1 & 0 & 0 & \ldots \\ x: & 0 & 1 & \phi & \phi^2 & \ldots \end{array}$$

As can be seen from above, the impulse-response dynamics is quite clear from a MA representation. For example, let $t > s > 0$, given one unit increase in ϵ_s, the effect on x_t would be ϕ^{t-s}, if there are no other shocks. If there are shocks that take place at time other than s and has nonzero effect on x_t, then we can add these effects, since this is a linear model.

The dynamics is a bit complicated for higher order AR process. But applying our old trick of inverting them to a MA process, then the following analysis will be straightforward. Take an AR(2) process as example.

Example 2

$$x_t = 0.6x_{t-1} + 0.2x_{t-2} + \epsilon_t$$

or

$$(1 - 0.6L - 0.2L^2)x_t = \epsilon_t$$

We first solve the polynomial:

$$y^2 + 3y - 5 = 0$$

and get two roots[1] $y_1 = 1.2926$ and $y_2 = -4.1925$. Recall that $\lambda_1 = 1/y_1 = 0.84$ and $\lambda_2 = 1/y_2 = -0.24$. So we can factorize the lag polynomial to be:

$$\begin{aligned} (1 - 0.6L - 0.2L^2)x_t &= (1 - 0.84L)(1 + 0.24L)x_t \\ x_t &= (1 - 0.84L)^{-1}(1 + 0.24L)^{-1}\epsilon_t \\ &= \psi(L)\epsilon_t \end{aligned}$$

[1] Recall that the roots for polynomial $ay^2 + by + c = 0$ is $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$.

where $\psi_k = \sum_{j=0}^{k} \lambda_1^j \lambda_2^{k-j}$. In this example, the series of ψ is $\{1, 0.6,\ 0.5616, 0.4579,\ 0.3880, \ldots\}$. So the effects of ϵ on x can be described as:

$$\begin{array}{lllllll} \epsilon: & 0 & 1 & 0 & 0 & 0 & \ldots \\ x: & 0 & 1 & 0.6 & 0.5616 & 0.4579 & \ldots \end{array}$$

The right figure in Figure 1 plots this impulse-response function. So after we invert an AR(p) process to an MA process, given $t > s > 0$, the effect of one unit increase in ϵ_s on x_t is just ψ_{t-s}.

We can see that given a linear process, AR or ARMA, if we could represent them as a MA process, we will find impulse-response dynamics immediately. In fact, MA representation is the same thing as the impulse-response function.

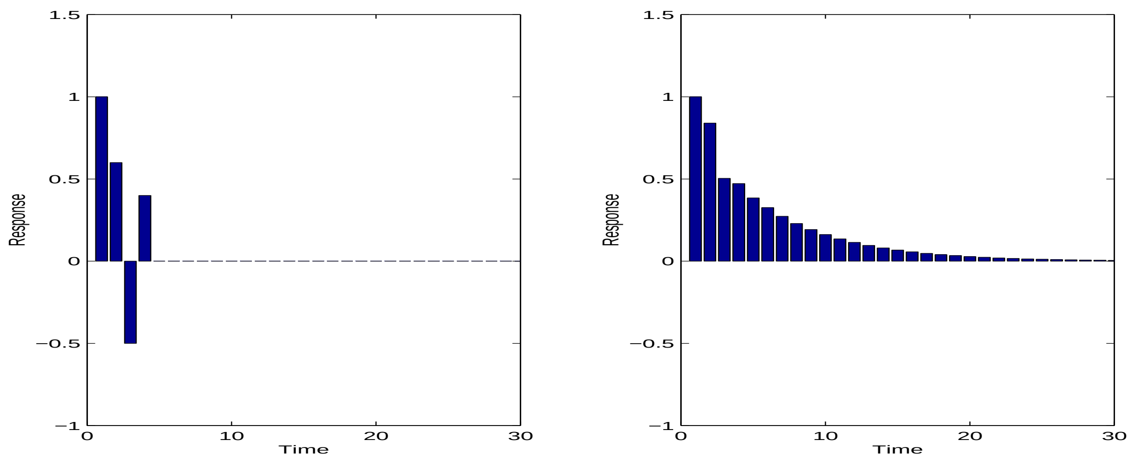

Figure 1: The impulse-response functions of an MA(3) process ($\theta_1 = 0.6, \theta_2 = -0.5, \theta_3 = 0.4$) and an AR(2) process ($\phi_1 = 0.6, \phi_2 = 0.2$), with unit shock at time zero

5 Autocovariance Functions and Stationarity of ARMA models

5.1 MA(1)

$$x_t = \epsilon_t + \theta\epsilon_{t-1},$$

where $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$. It is easy to calculate the first two moments of x_t:

$$E(x_t) = E(\epsilon_t + \theta\epsilon_{t-1}) = 0$$
$$E(x_t^2) = (1 + \theta^2)\sigma_\epsilon^2$$

and

$$\begin{aligned} \gamma_x(t, t+h) &= E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t+h} + \theta\epsilon_{t+h-1})] \\ &= \begin{cases} \theta\sigma_\epsilon^2 & \text{for} \quad h = 1 \\ 0 & \text{for} \quad h > 1 \end{cases} \end{aligned}$$

So, for a MA(1) process, we have a fixed mean and a covariance function which does not depend on time t: $\gamma(0) = (1+\theta^2)\sigma_\epsilon^2$, $\gamma(1) = \theta\sigma_\epsilon^2$, and $\gamma(h) = 0$ for $h > 1$. So we know MA(1) is stationary given any finite value of θ.

The autocorrelation can be computed as $\rho_x(h) = \gamma_x(h)/\gamma_x(0)$, so

$$\rho_x(0) = 1, \quad \rho_x(1) = \frac{\theta}{1+\theta^2}, \quad \rho_x(h) = 0 \quad \text{for} \quad h > 1$$

We have proposed in the section on invertability that for an invertible (noninvertible) MA process, there always exists a noninvertible (invertible) process which is the same as the original process up to the second moment. We use the following MA(1) process as an example.

Example 3 The process

$$x_t = \epsilon_t + \theta\epsilon_{t-1}, \quad \epsilon_t \sim WN(0, \sigma^2) \quad |\theta| > 1$$

is noninvertible. Consider an invertible MA process defined as

$$\tilde{x}_t = \tilde{\epsilon}_t + 1/\theta\tilde{\epsilon}_{t-1}, \quad \tilde{\epsilon}_t \sim WN(0, \theta^2\sigma^2)$$

.

Then we can compute that $E(x_t) = E(\tilde{x}_t) = 0$, $E(x_t^2) = E(\tilde{x}_t^2) = (1+\theta^2)\sigma^2$, $\gamma_x(1) = \gamma_{\tilde{x}}(1) = \theta\sigma^2$, and $\gamma_x(h) = \gamma_{\tilde{x}}(h) = 0$ for $h > 1$. Therefore, these two processes are equivalent up to the second moments. To be more concrete, we plug in some numbers.

Let $\theta = 2$, and we know that the process

$$x_t = \epsilon_t + 2\epsilon_{t-1}, \quad \epsilon_t \sim WN(0, 1)$$

is noninvertible. Consider the invertible process

$$\tilde{x}_t = \tilde{\epsilon}_t + (1/2)\tilde{\epsilon}_{t-1}, \quad \tilde{\epsilon}_t \sim WN(0, 4)$$

.

Note that $E(x_t) = E(\tilde{x}_t) = 0$, $E(x_t^2) = E(\tilde{x}_t)^2 = 5$, $\gamma_x(1) = \gamma_{\tilde{x}}(1) = 2$, and $\gamma_x(h) = \gamma_{\tilde{x}}(h) = 0$ for $h > 1$.

Although these two representations, noninvertible MA and invertible MA, could generate the same process up to the second moment, we prefer the invertible presentations in practice because if we can invert an MA process to an AR process, we can find the value of ϵ_t (non-observable) based on all past values of x (observable). If a process is noninvertible, then, in order to find the value of ϵ_t, we have to know all future values of x.

5.2 MA(q)

$$x_t = \theta(L)\epsilon_t = \sum_{k=0}^{q}(\theta_k L^k)\epsilon_t$$

The first two moments are:

$$\begin{aligned} E(x_t) &= 0 \\ E(x_t^2) &= \sum_{k=0}^{q} \theta_k^2 \sigma_\epsilon^2 \end{aligned}$$

and

$$\gamma_x(h) = \begin{cases} \sum_{k=0}^{q-h} \theta_k \theta_{k+h} \sigma_\epsilon^2 & \text{for} \quad h = 1, 2, \ldots, q \\ 0 & \text{for} \quad h > q \end{cases}$$

Again, a MA(q) is stationary for any finite values of $\theta_1, \ldots, \theta_q$.

5.3 MA(∞)

$$x_t = \theta(L)\epsilon_t = \sum_{k=0}^{\infty} (\theta_k L^k)\epsilon_t$$

Before we compute moments and discuss the stationarity of x_t, we should first make sure that $\{x_t\}$ converges.

Proposition 1 *If $\{\epsilon_t\}$ is a sequence of white noise with $\sigma_\epsilon^2 < \infty$, and if $\sum_{k=0}^{\infty} \theta_k^2 < \infty$, then the series*

$$x_t = \theta(L)\epsilon_t = \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}$$

converges in mean square.

Proof (See Appendix 3.A. in Hamilton): Recall the Cauchy criterion: a sequence $\{y_n\}$ converges in mean square if and only if $\|y_n - y_m\| \to 0$ as $n, m \to \infty$. In this problem, for $n > m > 0$, we want to show that

$$\begin{aligned} & E\left[\sum_{k=1}^{n} \theta_k \epsilon_{t-k} - \sum_{k=1}^{m} \theta_k \epsilon_{t-k}\right]^2 \\ = & \sum_{m \leq k \leq n} \theta_k^2 \sigma_\epsilon^2 \\ = & \left[\sum_{k=0}^{n} \theta_k^2 - \sum_{k=0}^{m} \theta_k^2\right] \sigma_\epsilon^2 \\ \to & \; 0 \quad \text{as} \quad m, n \to \infty \end{aligned}$$

The result holds since $\{\theta_k\}$ is square summable. It is often more convenient to work with a slightly stronger condition – absolutely summability:

$$\sum_{k=0}^{\infty} |\theta_k| < \infty.$$

It is easy to show that absolutely summable implies square summable. A MA(∞) process with absolutely summable coefficients is stationary with moments:

$$\begin{aligned} E(x_t) &= 0 \\ E(x_t^2) &= \sum_{k=0}^{\infty} \theta_k^2 \sigma_\epsilon^2 \\ \gamma_x(h) &= \sum_{k=0}^{\infty} \theta_k \theta_{k+h} \sigma_\epsilon^2 \end{aligned}$$

5.4 AR(1)

$$(1 - \phi L)x_t = \epsilon_t \tag{2}$$

Recall that an AR(1) process with $|\phi| < 1$ can be inverted to an MA(∞) process

$$x_t = \theta(L)\epsilon_t \quad \text{with} \quad \theta_k = \phi^k.$$

With $|\phi| < 1$, it is easy to check that the absolute summability holds:

$$\sum_{k=0}^{\infty} |\theta_k| = \sum_{k=0}^{\infty} |\phi^k| < \infty.$$

Using the results for MA(∞), the moments for x_t in (2) can be computed:

$$\begin{aligned} E(x_t) &= 0 \\ E(x_t^2) &= \sum_{k=0}^{\infty} \phi^{2k} \sigma_\epsilon^2 \\ &= \sigma_\epsilon^2/(1 - \phi^2) \\ \gamma_x(h) &= \sum_{k=0}^{\infty} \phi^{2k+h} \sigma_\epsilon^2 \\ &= \phi^h \sigma_\epsilon^2/(1 - \phi^2) \end{aligned}$$

So, an AR(1) process with $|\phi| < 1$ is stationary.

5.5 AR(p)

Recall that an AR(p) process

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)x_t = \epsilon_t$$

can be inverted to an MA process $x_t = \theta(L)\epsilon_t$ if all λ_i in

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L) \tag{3}$$

have less than one absolute value. It also turns out that with $|\lambda_i| < 1$, the absolute summability $\sum_{k=0}^{\infty} |\psi_k| < \infty$ is also satisfied. (The proof can be found on page 770 of Hamilton and the proof uses the result that $\psi_k = c_1\lambda_1^k + c_2\lambda_2^k$.)

When we solve the polynomial in:

$$(L - y_1)(L - y_2)\dots(L - y_p) = 0 \tag{4}$$

the requirement that $|\lambda_i| < 1$ is equivalent to that all roots in (4) lie outside of the unit circle, i.e., $|y_i| > 1$ for all i.

First calculate the expectation for x_t, $E(x_t) = 0$. To compute the second moments, one method is to invert it into a MA process and using the formula of autocovariance function for MA(∞). This method requires finding the moving average coefficients ψ, and an alternative method which is known as *Yule-Walker* method maybe more convenient in finding the autocovariance functions. To illustrate this method, take an AR(2) process as an example:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t$$

Multiply $x_t, x_{t-1}, x_{t-2}, \dots$ to both sides of the equation, take expectation and and then divide by $\gamma(0)$, we get the following equations:

$$\begin{aligned}
1 &= \phi_1\rho(1) + \phi_2\rho(2) + \sigma_\epsilon^2/\gamma(0) \\
\rho(1) &= \phi_1 + \phi_2\rho(1) \\
\rho(2) &= \phi_1\rho(1) + \phi_2 \\
\rho(k) &= \phi_1\rho(k-1) + \phi_2\rho(k-2) \quad \text{for} \quad k \geq 3
\end{aligned}$$

$\rho(1)$ can be first solved from the second equation: $\rho(1) = \phi_1/(1 - \phi_2)$, $\rho(2)$ can then be solved from the third equation. $\rho(k)$ can be solved recursively using $\rho(1)$ and $\rho(2)$ and finally, $\gamma(0)$ can be solved from the first equation. Using $\gamma(0)$ and $\rho(k)$, $\gamma(k)$ can computed using $\gamma(k) = \rho(k)\gamma(0)$. Figure 2 plots this autocorrelation for $k = 0, \dots, 50$ and the parameters are set to be $\phi_1 = 0.5$ and $\phi_2 = 0.3$. As is clear from the graph, the autocorrelation is very close to zero when $k > 40$.

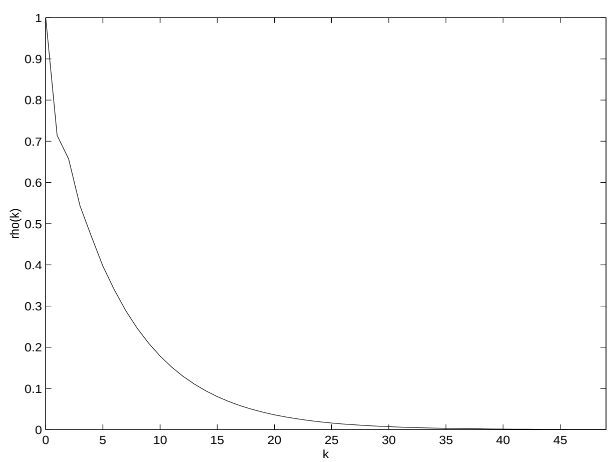

Figure 2: Plot of the autocorrelation of AR(2) process, with $\phi_1 = 0.5$ and $\phi_2 = 0.3$

5.6 ARMA(p, q)

Given an invertible ARMA(p, q) process, we have shown that

$$\phi(L)x_t = \theta(L)\epsilon_t,$$

invert $\phi(L)$ we obtain

$$x_t = \phi(L)^{-1}\theta(L)\epsilon_t = \psi(L)\epsilon_t.$$

Therefore, an ARMA(p, q) process is stationary as long as $\phi(L)$ is invertible. In other words, the stationarity of the ARMA process only depends on the autoregressive parameters, and not on the moving average parameters (assuming that all parameters are finite).

The expectation of this process $E(x_t) = 0$. To find the autocovariance function, first we can invert it to MA process and find the MA coefficients $\psi(L) = \phi(L)^{-1}\theta(L)$. We have shown an example of finding ψ in ARMA(1, 1) process, where we have

$$(1 - \phi L)x_t = (1 + \theta L)\epsilon_t$$

$$x_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

where $\psi_0 = 1$ and $\psi_j = \phi^{j-1}(\phi + \theta)$ for $j \geq 1$. Now, using the autocovariance functions for MA(∞) process we have

$$\begin{aligned}
\gamma_x(0) &= \sum_{k=0}^{\infty} \psi_k^2 \sigma_\epsilon^2 \\
&= \left(1 + \sum_{k=1}^{\infty} \phi^{2(k-1)}(\phi + \theta)^2\right)\sigma_\epsilon^2 \\
&= \left(1 + \frac{(\phi + \theta)^2}{1 - \phi^2}\right)\sigma_\epsilon^2
\end{aligned}$$

If we plug in some numbers, say, $\phi = 0.5$ and $\theta = 0.5$, so the original process is $x_t = 0.5x_{t-1} + \epsilon_t + 0.5\epsilon_{t-1}$, then $\gamma_x(0) = (7/3)\sigma_\epsilon^2$. For $h \geq 1$,

$$\begin{aligned}
\gamma_x(h) &= \sum_{k=0}^{\infty} \psi_k \psi_{k+h} \sigma_\epsilon^2 \\
&= \left(\phi^{h-1}(\phi + \theta) + \phi^{h-2}(\phi + \theta)^2 \sum_{k=1}^{\infty} \phi^{2k}\right)\sigma_\epsilon^2 \\
&= \phi^{h-1}(\phi + \theta)\left(1 + \frac{(\phi + \theta)\phi}{1 - \phi^2}\right)\sigma_\epsilon^2
\end{aligned}$$

Plug in $\phi = \theta = 0.5$ we have for $h \geq 1$,

$$\gamma_x(h) = \frac{5 \cdot 2^{1-h}}{3}\sigma_\epsilon^2.$$

An alternative to compute the autocovariance function is to multiply each side of $\phi(L)x_t = \theta(L)\epsilon_t$ with $x_t, x_{t-1}, \ldots$ and take expectations. In our ARMA(1, 1) example, this gives

$$\begin{aligned}
\gamma_x(0) - \phi\gamma_x(1) &= [1 + \theta(\theta + \phi)]\sigma_\epsilon^2 \\
\gamma_x(1) - \phi\gamma_x(0) &= \theta\sigma_\epsilon^2 \\
\gamma_x(2) - \phi\gamma_x(1) &= 0 \\
&\vdots \\
\gamma_x(h) - \phi\gamma_x(h-1) &= 0 \quad \text{for} \quad h > 2
\end{aligned}$$

where we use that $x_t = \psi(L)\epsilon_t$ in taking expectation on the right side, for instance, $E(x_t\epsilon_t) = E((\epsilon_t + \psi_1\epsilon_{t-1} + \ldots)\epsilon_t) = \sigma_\epsilon^2$. Plug in $\theta = \phi = 0.5$ and solving those equations, we have $\gamma_x(0) = (7/3)\sigma_\epsilon^2$, $\gamma_x(1) = (5/3)\sigma_\epsilon^2$, and $\gamma_x(h) = \gamma_x(h-1)/2$ for $h \geq 2$. This is the same results as we got using the first method.

Summary: A MA process is stationary if and only if the coefficients $\{\theta_k\}$ are square summable (absolute summable), i.e., $\sum_{k=0}^{\infty} \theta_k^2 < \infty$ or $\sum_{k=0}^{\infty} |\theta_k| < \infty$. Therefore, MA with finite number of MA coefficients are always stationary. Note that stationarity does not require MA to be invertible.

An AR process is stationary if it is invertible, i.e. $|\lambda_i| < 1$ or $|y_i| > 1$, as defined in (3) and (4) respectively. An ARMA(p, q) process is stationary if its autoregressive lag polynomial is invertible.

5.7 Autocovariance generating function of stationary ARMA process

For covariance stationary process, we see that autocovariance function is very useful in describing the process. One way to summarize absolutely summable autocovariance functions ($\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$) is to use the autocovariance-generating function:

$$g_x(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h.$$

where z could be a complex number.

For white noise, the autocovriance-generating function (AGF) is just a constant, i.e, for $\epsilon \sim WN(0, \sigma_\epsilon^2)$, $g_\epsilon(z) = \sigma_\epsilon^2$.

For MA(1) process,

$$x_t = (1 + \theta L)\epsilon_t, \quad \epsilon \sim WN(0, \sigma_\epsilon^2),$$

we can compute that

$$g_x(z) = \sigma_\epsilon^2[\theta z^{-1} + (1 + \theta^2) + \theta z] = \sigma_\epsilon^2(1 + \theta z)(1 + \theta z^{-1}).$$

For a MA(q) process,

$$x_t = (1 + \theta_1 L + \ldots + \theta_q L^q)\epsilon_t,$$

we know that $\gamma_x(h) = \sum_{k=0}^{q-h} \theta_k\theta_{k+h}\sigma_\epsilon^2$ for $h = 1, \ldots, q$ and $\gamma_x(h) = 0$ for $h > q$. we have

$$g_x(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h$$

$$
\begin{aligned}
&= \sigma_\epsilon^2 \left(\sum_{k=0}^{q} \theta_k^2 + \sum_{h=1}^{q} \sum_{k=0}^{q-h} (\theta_k \theta_{k-h} z^{-h} + \theta_k \theta_{k+h} z^{h}) \right) \\
&= \sigma_\epsilon^2 \left(\sum_{k=0}^{q} \theta_k z^k \right) \left(\sum_{k=0}^{q} \theta_k z^{-k} \right)
\end{aligned}
$$

For a MA(∞) process $x_t = \theta(L)\epsilon_t$ where $\sum_{k=0}^{\infty} |\theta_k| < \infty$, we can naturally let q be replaced by ∞ in the AGF for MA(q) to get AGF for MA(∞),

$$
g_x(z) = \sigma_\epsilon^2 \left(\sum_{k=0}^{\infty} \theta_k z^k \right) \left(\sum_{k=0}^{\infty} \theta_k z^{-k} \right) = \sigma_\epsilon^2 \theta(z)\theta(z^{-1}).
$$

Next, for a stationary AR or ARMA process, we can invert them to a MA process. For instance, an AR(1) process, $(1-\phi L)x_t = \epsilon_t$, invert it to

$$
x_t = \frac{1}{1-\phi L}\epsilon_t,
$$

and its AGF is

$$
g_x(z) = \frac{\sigma_\epsilon^2}{(1-\phi z)(1-\phi z^{-1})},
$$

which equal to

$$
\sigma_\epsilon^2 \left(\sum_{k=0}^{\infty} \theta_k z^k \right) \left(\sum_{k=0}^{\infty} \theta_k z^{-k} \right) = \sigma^2 \theta(z)\theta(z^{-1}),
$$

where $\theta_k = \phi^k$. In general, the AGF for an ARMA(p, q) process is

$$
\begin{aligned}
g_x(z) &= \frac{\sigma_\epsilon^2 (1 + \theta_1 z + \ldots + \theta_q z^q)(1 + \theta_1 z^{-1} + \ldots + \theta_q z^{-q})}{(1 - \phi_1 z - \ldots - \phi_p z^p)(1 - \phi_1 z^{-1} - \ldots - \phi_p z^{-p})} \\
&= \sigma_\epsilon^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}
\end{aligned}
$$

6 Simulated ARMA process

In this section, we plot a few simulated ARMA processes. In the simulations, the errors are Gaussian white noise $i.i.d.N(0,1)$. As a comparison, we first plot a Gaussian white noise (or AR(1) with $\phi = 0$) in Figure 3. Then, we plot AR(1) with $\phi = 0.4$ and $\phi = 0.9$ in Figure 4 and Figure 5. As you can see, the white noise process is very choppy and patternless. When $\phi = 0.4$, it becomes a bit smoother, and when $\phi = 0.9$, the departures from the mean (zero) is very prolonged. Figure 6 plots an AR(2) process and the coefficients are set to numbers as in our example in this lecture. Finally, Figure 7 plots a MA(3) process. Compare this MA(3) process with the white noise, we could see an increase of volatilities (the volatility of the white noise is 1 and the volatility of the MA(3) process is 1.77).

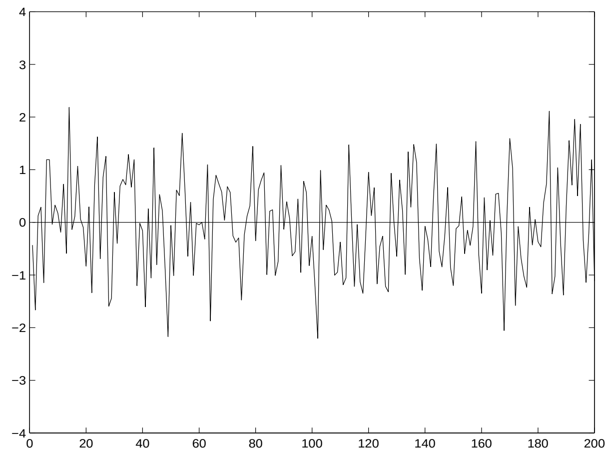

Figure 3: A Gaussian white noise time series

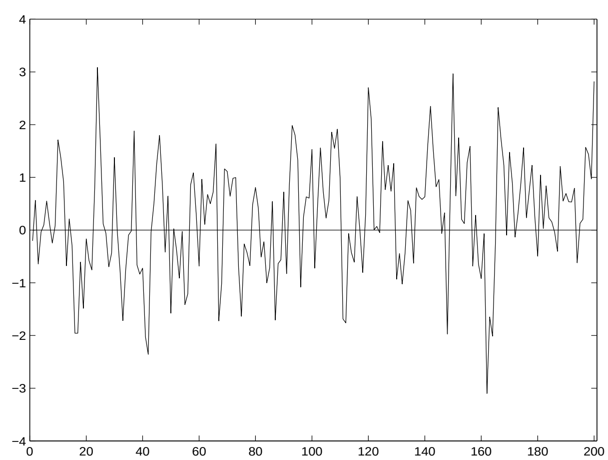

Figure 4: A simulated AR(1) process, with $\phi = 0.4$

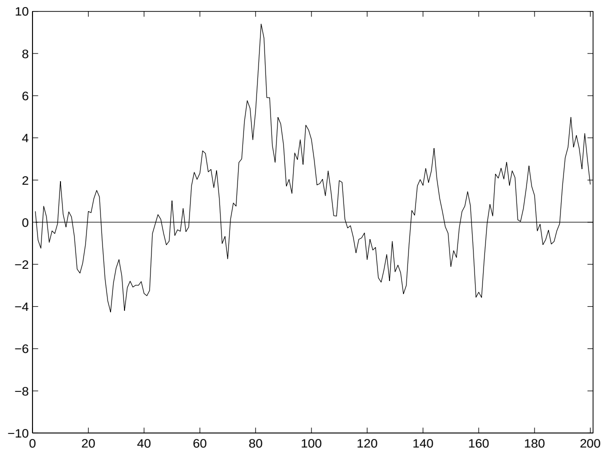

Figure 5: A simulated AR(1) process, with $\phi = 0.9$

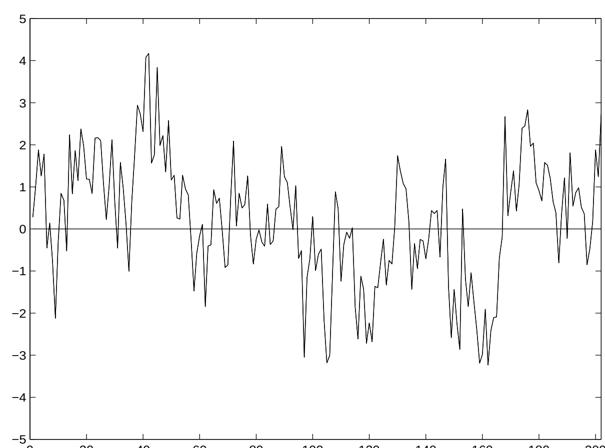

Figure 6: A simulated AR(2) process, with $\phi_1 = 0.6$, $\phi_2 = 0.2$

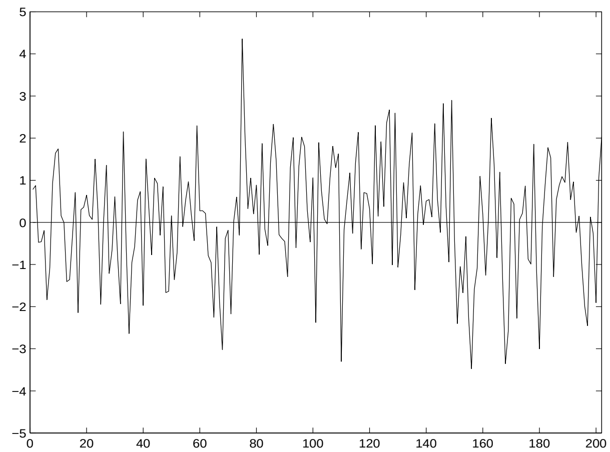

Figure 7: A simulated MA(3) process, with $\theta_1 = 0.6$, $\theta_2 = -0.5$, and $\theta_3 = 0.4$

7 Forecastings of ARMA Models

7.1 Principles of forecasting

If we are interested in forecasting a random variable y_{t+h} based on the observations of x up to time t (denoted by X) we can have different candidates, denoted by $g(X)$. If our criterion in picking the best forecast is to minimize the mean squared error (MSE), then the best forecast is the conditional expectation, $g(X) = E_X(y_{t+h})$. The proof can be found on page 73 in Hamilton. In our following discussion, we assume that the data generating process is known (so parameters are known), so we can compute the conditional moments.

7.2 AR models

Let's start from an AR(1) process:

$$x_t = \phi x_{t-1} + \epsilon_t$$

where we continue to assume that ϵ_t is a white noise with mean zero and variance σ_ϵ^2, then we can compute

$$\begin{aligned}
E_t(x_{t+1}) &= E_t(\phi x_t + \epsilon_{t+1}) = \phi x_t \\
E_t(x_{t+2}) &= E_t(\phi^2 x_t + \phi\epsilon_{t+1} + \epsilon_{t+2}) = \phi^2 x_t \\
\ldots &= \ldots \\
E_t(x_{t+k}) &= E_t(\phi^k x_t + \phi^{k-1}\epsilon_{t+1} + \ldots + \epsilon_{t+k}) = \phi^k x_t
\end{aligned}$$

and the variance

$$\begin{aligned}
\mathrm{Var}_t(x_{t+1}) &= \mathrm{Var}_t(\phi x_t + \epsilon_{t+1}) = \sigma_\epsilon^2 \\
\mathrm{Var}_t(x_{t+2}) &= \mathrm{Var}_t(\phi^2 x_t + \phi\epsilon_{t+1} + \epsilon_{t+2}) = (1+\phi^2)\sigma_\epsilon^2 \\
\ldots &= \ldots \\
\mathrm{Var}_t(x_{t+k}) &= \mathrm{Var}_t(\phi^k x_t + \phi^{k-1}\epsilon_{t+1} + \ldots + \epsilon_{t+k}) = \sum_{j=0}^{k-1} \phi^{2j}\sigma_\epsilon^2
\end{aligned}$$

Note that as $k \to \infty$,

$$E_t(x_{t+k}) \to 0$$

which is the unconditional expectation of x_t, and

$$\mathrm{Var}_t(x_{t+k}) \to \sigma_\epsilon^2/(1-\phi^2)$$

which is the unconditional variance of x_t.

Similarly, for an AR(p) process, we can forecast recursively.

7.3 MA Models

For a MA(1) process,

$$x_t = \epsilon_t + \theta\epsilon_{t-1},$$

if we know ϵ_t, then

$$\begin{aligned}
E_t(x_{t+1}) &= E_t(\epsilon_{t+1} + \theta\epsilon_t) = \theta\epsilon_t \\
E_t(x_{t+2}) &= E_t(\epsilon_{t+2} + \theta\epsilon_{t+1}) = 0 \\
\ldots &= \ldots \\
E_t(x_{t+k}) &= E_t(\epsilon_{t+k} + \theta\epsilon_{t+k-1}) = 0
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}_t(x_{t+1}) &= \mathrm{Var}_t(\epsilon_{t+1} + \theta\epsilon_t) = \sigma_\epsilon^2 \\
\mathrm{Var}_t(x_{t+2}) &= \mathrm{Var}_t(\epsilon_{t+2} + \theta\epsilon_{t+1}) = (1+\theta^2)\sigma_\epsilon^2 \\
\ldots &= \ldots \\
\mathrm{Var}_t(x_{t+k}) &= \mathrm{Var}_t(\epsilon_{t+k} + \theta\epsilon_{t+k-1}) = (1+\theta^2)\sigma_\epsilon^2
\end{aligned}$$

It is easy to see that for an MA(1) process, the conditional expectation for two step ahead and higher is the same as unconditional expectation, so is the variance. Next, for a MA(q) model,

$$x_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q} = \sum_{j=0}^{q} \theta_j\epsilon_{t-j},$$

if we know $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$, then

$$\begin{aligned}
E_t(x_{t+1}) &= E_t(\sum_{j=0}^{q} \theta_j\epsilon_{t+1-j}) = \sum_{j=1}^{q} \theta_j\epsilon_{t+1-j} \\
E_t(x_{t+2}) &= E_t(\sum_{j=0}^{q} \theta_j\epsilon_{t+2-j}) = \sum_{j=2}^{q} \theta_j\epsilon_{t+2-j} \\
\ldots &= \ldots \\
E_t(x_{t+k}) &= E_t(\sum_{j=0}^{q} \theta_j\epsilon_{t+k-j}) = \sum_{j=k}^{q} \theta_j\epsilon_{t+k-j} \quad \text{for} \quad k \leq q \\
E_t(x_{t+k}) &= E_t(\sum_{j=0}^{q} \theta_j\epsilon_{t+k-j}) = 0 \quad \text{for} \quad k > q
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}_t(x_{t+1}) &= \mathrm{Var}_t(\sum_{j=0}^{q}\theta_j\epsilon_{t+1-j}) = \sigma_\epsilon^2 \\
\mathrm{Var}_t(x_{t+2}) &= \mathrm{Var}_t(\sum_{j=0}^{q}\theta_j\epsilon_{t+2-j}) = 1+\theta_1^2\sigma_\epsilon^2 \\
\ldots &= \ldots \\
\mathrm{Var}_t(x_{t+k}) &= \mathrm{Var}_t(\sum_{j=0}^{q}\theta_j\epsilon_{t+k-j}) = \sum_{j=0}^{k}\theta_j^2\sigma_\epsilon^2 \quad \forall \quad k>0
\end{aligned}$$

We could see that for an MA(q) process, the conditional expectation and variance of forecast for $q+1$ and higher is the same as unconditional expectations and variance.

8 Wold Decomposition

So far we have focused on ARMA models, which are linear time series models. Is there any relationship between a general covariance stationary process (maybe nonlinear) to linear representations? The answer is given by the Wold decomposition theorem:

Proposition 2 *(Wold Decomposition) Any zero-mean covariance stationary process x_t can be represented in the form*

$$x_t = \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j} + V_t$$

where

(i) $\psi_0 = 1$ and $\sum_{j=0}^{\infty}\psi_j^2 < \infty$

(ii) $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$

(iii) $E(\epsilon_t V_s) = 0 \quad \forall \quad s, t > 0$

(iv) ϵ_t *is the error in forecasting x_t on the basis of a linear function of lagged x:*

$$\epsilon_t = x_t - E(x_t|x_{t-1}, x_{t-2}, \ldots)$$

(v) V_t *is a deterministic process and it can be predicted from a linear function of lagged x.*

Remarks: Wold decomposition says that any covariance stationary process has a linear representation: a linear deterministic component (V_t) and a linear indeterministic components (ϵ_t). If $V_t = 0$, then the process is said to be purely-non-deterministic, and the process can be represented as a MA(∞) process. Basically, ϵ_t is the error from the projection of x_t on lagged x, therefore it is uniquely determined and it is orthogonal to lagged x and lagged ϵ. Since this error ϵ is the residual from the projections, it may not be the true errors in the DGP of x_t. Also note that the error term (ϵ) is a white noise process, and does not need to be *iid*.

Readings:
Hamilton Ch. 1-4
Brockwell and Davis Ch. 3
Hayashi Ch 6.1, 6.2