# Ch. 15 Forecasting

(March 25, 2019)

Having considered in Chapter 14 some of the properties of $ARMA$ models, we now show how they may be used to forecast future values of an observed time series. For the present we proceed as if the model were known *exactly*.[1]

One of the important problems in time series analysis is the following: Given $T$ observations on a realization, predict the $(T+s)$th observation in the realization, where $s$ is a positive integer. The prediction is sometime called the $forecast$ of the $(T+s)$th observation.

Forecasting is an important concept for the studies of time series analysis. In the scope of regression model we usually has an existing economic theory model for us to estimate their parameters. The estimated coefficients have already a role to play such as to confirm some economic theories. Therefore, to forecast or not from this estimated model depends on researcher's own interest. However, the estimated coefficients from a time series model have no significant meaning to economic theory. An important role that a time series analysis is therefore to be able to forecast precisely from this pure mechanical model.

# 1 Principle of Forecasting

## 1.1 Forecasts Based on Conditional Expectations (With a Known Model)

Suppose we are interested in forecasting the value of a variables $Y_{t+1}$ based on a set of variables $\mathbf{x}_t$ observed at date $t$. For example, we might want to forecast $Y_{t+1}$ based on

[1] That is, the model we assume is a correct specification and its parameters are known.

its $m$ most recent values. In this case, $\mathbf{x}_t = [Y_t, Y_{t-1}, ...., Y_{t-m+1}]'$.

Let $Y_{t+1|t}$ denote a forecast of $Y_{t+1}$ based on $\mathbf{x}_t$ (a function of $\mathbf{x}_t$, depending on how they are realized). To evaluate the usefulness of this forecast, we need to specify a loss function. A quadratic loss function means choosing the forecast $Y^*_{t+1|t}$ so as to minimize

$$MSE(Y_{t+1|t}) = E(Y_{t+1} - Y_{t+1|t})^2,$$

which is known as the *mean squared error*.

**Theorem.**

The smallest mean squared error of in the forecast $Y_{t+1|t}$ is the (statistical) expectation of $Y_{t+1}$ conditional on $\mathbf{x}_t$:

$$Y^*_{t+1|t} = E(Y_{t+1}|\mathbf{x}_t).$$

**Proof**:

Let $g(\mathbf{x}_t)$ be a forecasting function of $Y_{t+1}$ other then the conditional expectation $E(Y_{t+1}|\mathbf{x}_t)$. Then the MSE associated with $g(\mathbf{x}_t)$ would be

$$\begin{aligned} E[Y_{t+1} - g(\mathbf{x}_t)]^2 &= E[Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t) + E(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)]^2 \\ &= E[Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t)]^2 \\ &\quad +2E\{[Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t)][E(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)]\} \\ &\quad +E\{E[(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)]^2\}. \end{aligned}$$

Denote $\eta_{t+1} \equiv \{[Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t)][E(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)]\}$ we have

$$\begin{aligned} E(\eta_{t+1}|\mathbf{x}_t) &= [E(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)] \times E([Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t)]|\mathbf{x}_t) \\ &= [E(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)] \times 0 \\ &= 0. \end{aligned}$$

By laws of iterated expectation, it follows that

$$E(\eta_{t+1}) = E_{\mathbf{x}_t} E(E[\eta_{t+1}|\mathbf{x}_t]) = 0.$$

Therefore we have

$$E[Y_{t+1} - g(\mathbf{x}_t)]^2 = E[Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t)]^2 + E\{E[(Y_{t+1}|\mathbf{x}_t) - g(\mathbf{x}_t)]^2\}. \tag{15-1}$$

The second term on the right hand side of (15-1) cannot be made smaller than zero and the first term does not depend on $g(\mathbf{x}_t)$. The function $g(\mathbf{x}_t)$ that can makes the mean square error (15-1) as small as possible is the function that sets the second term in (15-1) to zero:

$$g(\mathbf{x}_t) = E(Y_{t+1}|\mathbf{x}_t).$$

The MSE of this optimal forecast is thus

$$E[Y_{t+1} - g(\mathbf{x}_t)]^2 = E[Y_{t+1} - E(Y_{t+1}|\mathbf{x}_t)]^2.$$

■

## 1.2 Forecasts Based on Linear Projection (Without a Known Model)

Suppose we now consider only the class of forecast $Y_{t+1}$ that is a linear function of $\mathbf{x}_t$:

$$Y_{t+1|t} = \boldsymbol{\alpha}'\mathbf{x}_t.$$

**Definition**:

The forecast $\boldsymbol{\alpha}'\mathbf{x}_t$ is called the *linear projection* of $Y_{t+1}$ on $\mathbf{x}_t$ if the forecast error $(Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t)$ is uncorrelated with $\mathbf{x}_t$, i.e.

$$E[(Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t)\mathbf{x}_t'] = \mathbf{0}'. \tag{15-2}$$

■

**Theorem.**

The linear projection produces the smallest mean squared error among the class of linear forecasting rule.

**Proof**:

Let $\mathbf{g}'\mathbf{x}_t$ be any arbitrary linear forecasting function of $Y_{t+1}$. Then the MSE associated with $\mathbf{g}'\mathbf{x}_t$ would be

$$\begin{aligned}
E[Y_{t+1} - \mathbf{g}'\mathbf{x}_t]^2 &= E[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\alpha}'\mathbf{x}_t - \mathbf{g}'\mathbf{x}_t]^2 \\
&= E[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t]^2 \\
&\quad + 2E\{[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t][\boldsymbol{\alpha}'\mathbf{x}_t - \mathbf{g}'\mathbf{x}_t]\} \\
&\quad + E[\boldsymbol{\alpha}'\mathbf{x}_t - \mathbf{g}'\mathbf{x}_t]^2.
\end{aligned}$$

Denote $\eta_{t+1} \equiv \{[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t][\boldsymbol{\alpha}'\mathbf{x}_t - \mathbf{g}'\mathbf{x}_t]\}$ we have

$$\begin{aligned}
E(\eta_{t+1}) &= E\{[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t][(\boldsymbol{\alpha} - \mathbf{g})'\mathbf{x}_t]\} \\
&= (E[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t]\mathbf{x}_t')[\boldsymbol{\alpha} - \mathbf{g}] \\
&= \mathbf{0}'[\boldsymbol{\alpha} - \mathbf{g}] \\
&= 0.
\end{aligned}$$

Therefore we have

$$E[Y_{t+1} - \mathbf{g}'\mathbf{x}_t]^2 = E[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t)]^2 + E[\boldsymbol{\alpha}'\mathbf{x}_t - \mathbf{g}'\mathbf{x}_t]^2. \tag{15-3}$$

The second term on the right hand side of (15-3) cannot be made smaller than zero and the first term does not depend on $\mathbf{g}'\mathbf{x}_t$. The function $\mathbf{g}'\mathbf{x}_t$ that can makes the mean square error (15-3) as small as possible is the function that sets the second term in (15-3) to zero:

$$\mathbf{g}'\mathbf{x}_t = \boldsymbol{\alpha}'\mathbf{x}_t.$$

The MSE of this optimal forecast is

$$E[Y_{t+1} - \mathbf{g}'\mathbf{x}_t]^2 = E[Y_{t+1} - \boldsymbol{\alpha}'\mathbf{x}_t]^2.$$

■

**Notation.**

For $\boldsymbol{\alpha}'\mathbf{x}_t$ is a linear projection of $Y_{t+1}$ on $\mathbf{x}_t$, we will use the notation

$$\hat{P}(Y_{t+1}|\mathbf{x}_t) = \boldsymbol{\alpha}'\mathbf{x}_t,$$

to indicate the linear projection of $Y_{t+1}$ on $\mathbf{x}_t$. ■

Notice that

$$MSE[\hat{P}(Y_{t+1}|\mathbf{x}_t)] \geq MSE[E(Y_{t+1}|\mathbf{x}_t)],$$

since the conditional expectation offers the best possible forecast.

For most applications a constant term will be included in the projection. We will use the symbol $\hat{E}$ to indicate a linear projection on a vector of random variables $\mathbf{x}_t$ along a constant term:

$$\hat{E}(Y_{t+1}|\mathbf{x}_t) \equiv \hat{P}(Y_{t+1}|1, \mathbf{x}_t).$$

# 2 Forecast ARMA Model with Infinite Observations

Recall that a general stationary and invertible $ARMA(p,q)$ process is written in this form:

$$\phi(L)(Y_t-\mu)=\theta(L)\varepsilon_t, \tag{15-4}$$

where $\phi(L)=1-\phi_1L-\phi_2L^2-...-\phi_pL^p$, $\theta(L)=1+\theta_1L+\theta_2L^2+...+\theta_qL^q$ and all the roots of $\phi(L)=0$ and $\theta(L)=0$ lie outside the unit circle.

## 2.1 Forecasting Based on Lagged $\varepsilon's$, $MA(\infty)$ form

Assume that the lagged $\varepsilon_t$ is observed directly, then the following principle is useful in the manipulating the forecasting in the subsequence. Consider an $MA(\infty)$ form of (15-4):

$$Y_t-\mu=\varphi(L)\varepsilon_t \tag{15-5}$$

with $\varepsilon_t$ white noise and

$$\varphi(L) \;=\; \theta(L)\phi^{-1}(L)=\sum_{j=0}^{\infty}\varphi_jL^j,$$

where $\varphi_0=1$ and $\sum_{j=0}^{\infty}|\varphi_j|<\infty$.

Suppose that we have an infinite number of observations on $\varepsilon$ through date $t$, that is $\{\varepsilon_t,\varepsilon_{t-1},\varepsilon_{t-2},....\}$, and further know the value of $\mu$ and $\{\varphi_1,\varphi_2,...\}$. Say we want to forecast the value of $Y_{t+s}$ from now. Note that (15-5) implies

$$\begin{aligned}Y_{t+s} &= \mu+\varepsilon_{t+s}+\varphi_1\varepsilon_{t+s-1}+...+\varphi_{s-1}\varepsilon_{t+1}+\varphi_s\varepsilon_t\\ &\quad+\varphi_{s+1}\varepsilon_{t-1}+....\end{aligned}$$

We consider the best linear forecast since the $ARMA$ model we considered so far is covariance-stationary, no assumption is made on their distribution so we cannot form conditional expectation. The best linear forecast takes the form[2]

$$\begin{aligned}\hat{E}(Y_{t+s}|\varepsilon_t,\varepsilon_{t-1},...) &= \mu+\varphi_s\varepsilon_t+\varphi_{s+1}\varepsilon_{t-1}+....\\ &= [\mu,\varphi_s,\varphi_{s+1},...][1,\varepsilon_t,\varepsilon_{t-1},...]'\\ &= \boldsymbol{\alpha}'\mathbf{x}_t.\end{aligned} \tag{15-6}$$

[2]You cannot use conditional expectation that $E(\varepsilon_{t+i}|\varepsilon_t,\varepsilon_{t-1},..)=0$ for $i=1,2,\cdots,s$ since we only assume that $\varepsilon_t$ is uncorrelated which do not imply that they are independent.

That is, the unknown future $\varepsilon$'s are set to their expected value of zero. The error associated with this forecast is uncorrelated with $\mathbf{x}_t = [1, \varepsilon_t, \varepsilon_{t-1}, ...]'$, or

$$\begin{aligned}
& E\{\mathbf{x}_t \cdot [Y_{t+s} - \hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...)]\} \\
&= E\left\{ \begin{bmatrix} 1 \\ \varepsilon_t \\ \varepsilon_{t-1} \\ . \\ . \\ . \end{bmatrix} (\varepsilon_{t+s} + \varphi_1 \varepsilon_{t+s-1} + ... + \varphi_{s-1}\varepsilon_{t+1}) \right\} \\
&= \mathbf{0}.
\end{aligned}$$

The mean squared error associated with this forecast is

$$E[Y_{t+s} - \hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...)]^2 = (1 + \varphi_1^2 + \varphi_2^2 + ... + \varphi_{s-1}^2)\sigma^2.$$

**Example**.
For an $MA(q)$ process, the optimal linear forecast is

$$\hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...] = \begin{cases} \mu + \theta_s \varepsilon_t + \theta_{s+1}\varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q+s} & for \quad s = \quad 1, 2, ..., q \\ \mu & for \quad s = \quad q+1, q+2, ... \end{cases}$$

The MSE is

$$\begin{cases} \sigma^2, & for \quad s = \quad 1 \\ (1 + \theta_1^2 + \theta_2^2 + ... + \theta_{s-1}^2)\sigma^2, & for \quad s = \quad 2, 3, ..., q \\ ((1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2, & for \quad s = \quad q+1, q+2, .... \end{cases}$$

The MSE increase with the forecast horizon $s$ up until $s = q$. If we try to forecast an $MA(q)$ farther than $q$ periods into the future, the forecast is simply the unconditional mean of the series ($E(Y_{t+s}) = \mu$)) and the MSE is the unconditional variance of the series ($Var(Y_{t+s}) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2$). ■

A compact lag operator expression for the forecast in (15-6) is sometimes used. Rewrite $Y_{t+s}$ as in (15-5) as

$$\begin{aligned} Y_{t+s} &= \mu + \varphi(L)\varepsilon_{t+s} \\ &= \mu + \varphi(L)L^{-s}\varepsilon_t. \end{aligned}$$

Consider polynomial that $\varphi(L)$ are divided by $L^s$:

$$\begin{aligned} \frac{\varphi(L)}{L^s} &= L^{-s} + \varphi_1 L^{1-s} + \varphi_2 L^{2-s} + ... + \varphi_{s-1}L^{-1} + \varphi_s L^0 \\ &\quad + \varphi_{s+1}L^1 + \varphi_{s+2}L^2 + .... \end{aligned}$$

The *annihilation operator* is to replace negative powers of $L$ in $\frac{\varphi(L)}{L^s}$ by zero. For example,

$$\left[\frac{\varphi(L)}{L^s}\right]_+ = \varphi_s L^0 + \varphi_{s+1} L^1 + \varphi_{s+2} L^2 + ....$$

Therefore the optimal forecast (15-6) could be written in lag operator notation as

$$\begin{aligned} \hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...] &= \mu + \varphi_s \varepsilon_t + \varphi_{s+1}\varepsilon_{t-1} + .... \\ &= \mu + (\varphi_s L^0 + \varphi_{s+1} L^1 + \varphi_{s+2} L^2 + ....)\varepsilon_t \\ &= \mu + \left[\frac{\varphi(L)}{L^s}\right]_+ \varepsilon_t. \end{aligned} \tag{15-7}$$

## 2.2 Forecasting Based on Lagged $Y's$

The previous forecasts were based on the assumption that $\varepsilon_t$ is observed directly. In the usual forecasting situation, we actually have observation on lagged $Y's$, not lagged $\varepsilon's$. Suppose that the general $ARMA(p, q)$ has an $AR(\infty)$ representation given by

$$\eta(L)(Y_t - \mu) = \varepsilon_t \tag{15-8}$$

with $\varepsilon_t$ white noise and

$$\eta(L) \;=\; \theta^{-1}(L)\phi(L) = \sum_{j=0}^{\infty} \eta_j L^j = \varphi^{-1}(L),$$

where $\eta_0 = 1$ and $\sum_{j=0}^{\infty} |\eta_j| < \infty$.

Under these conditions, we can substitute (15-8) into (15-7) to obtain the forecast of $Y_{t+s}$ as a function of lagged $Y's$:

$$\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + \left[\frac{\varphi(L)}{L^s}\right]_+ \eta(L)(Y_t - \mu) \tag{15-9}$$

or

$$\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + \left[\frac{\varphi(L)}{L^s}\right]_+ \frac{1}{\varphi(L)}(Y_t - \mu). \tag{15-10}$$

Equation (15-10) is known as the *Wiener-Kolmogorov prediction formula.*

### 2.2.1 Forecasting an $AR(1)$ Process

Therefore are two ways to forecast a $AR(1)$ process. They are:

**(a).** By *Wiener-Kolmogorov prediction formula*:
For the covariance stationary $AR(1)$ process, we have

$$\varphi(L) = \frac{1}{1-\phi L} = 1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + ...$$

and

$$\left[\frac{\varphi(L)}{L^s}\right]_+ = \phi^s + \phi^{s+1}L^1 + \phi^{s+2}L^2 + ... = \frac{\phi^s}{1-\phi L}.$$

The optimal linear $s$-period ahead forecast for a stationary $AR(1)$ process is therefore:

$$\begin{aligned}\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...] &= \mu + \frac{\phi^s}{1-\phi L}(1-\phi L)(Y_t - \mu) \\ &= \mu + \phi^s(Y_t - \mu).\end{aligned}$$

The forecast decays geometrically from $(Y_t - \mu)$ toward $\mu$ as the forecast horizon $s$ increase.

**(b).** By *Recursive Substitution and Lag operator*:
The $AR(1)$ process can be represented as (using (13-3) on Ch. 13)

$$Y_{t+s} - \mu = \phi^s(Y_t - \mu) + \phi^{s-1}\varepsilon_{t+1} + \phi^{s-2}\varepsilon_{t+2} + ... + \phi\varepsilon_{t+s-1} + \varepsilon_{t+s},$$

Setting $E(\varepsilon_{t+h}) = 0, h = 1, 2, ..., s$, the optimal linear $s$-period ahead forecast for a stationary $AR(1)$ process is therefore:

$$\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + \phi^s(Y_t - \mu),$$

with forecast error $\phi^{s-1}\varepsilon_{t+1} + \phi^{s-2}\varepsilon_{t+2} + ... + \phi\varepsilon_{t+s-1} + \varepsilon_{t+s}$, which is uncorrelated with the forecast $Y_t(= \mu + \varepsilon_t + \phi^1\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ....)$. □

The MSE of this forecast is

$$E(\phi^{s-1}\varepsilon_{t+1} + \phi^{s-2}\varepsilon_{t+2} + ... + \phi\varepsilon_{t+s-1} + \varepsilon_{t+s})^2 = (1 + \phi^2 + \phi^4 + ... + \phi^{2(s-1)})\sigma^2.$$

Notice that this grows with $s$ and asymptotically approach $\sigma^2/(1-\phi^2)$, the unconditional variance of $Y$.[3]

### 2.2.2 Forecasting an $AR(p)$ Process

**(a).** *By Recursive substitution and Lag operator*:

Following (13-11) in Chapter 13, the value of $Y$ at $t+s$ of an $AR(p)$ process can be represented as

$$\begin{aligned} Y_{t+s} - \mu &= f_{11}^s(Y_t - \mu) + f_{12}^s(Y_{t-1} - \mu) + ... + f_{1p}^s(Y_{t-p+1} - \mu) \\ &\quad + f_{11}^{s-1}\varepsilon_{t+1} + f_{11}^{s-2}\varepsilon_{t+2} + ... + f_{11}^{1}\varepsilon_{t+s-1} + \varepsilon_{t+s}, \end{aligned}$$

where $f_{11}^j$ is the $(1,1)$ elements of $\mathbf{F}^j$, in which

$$\mathbf{F} \equiv \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & . & . & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & . & . & 0 & 0 \\ 0 & 1 & 0 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 & 0 \end{bmatrix}.$$

Setting $E(\varepsilon_{t+h}) = 0, h = 1, 2, ..., s$, the optimal linear $s$-period ahead forecast for a stationary $AR(p)$ process is therefore:

$$\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + f_{11}^s(Y_t - \mu) + f_{12}^s(Y_{t-1} - \mu) + ... + f_{1p}^s(Y_{t-p+1} - \mu). \tag{15-11}$$

The associated forecast error is

$$Y_{t+s} - \hat{E}(Y_{t+s}) = f_{11}^{s-1}\varepsilon_{t+1} + f_{11}^{s-2}\varepsilon_{t+2} + ... + f_{11}^{1}\varepsilon_{t+s-1} + \varepsilon_{t+s}.$$

[3] That is, with the information of $Y_t$ to forecast $Y_{t+s}$, the MSE could never be larger than the unconditional variance of $Y_{t+s}$.

**(b).** *By Law of Iterated Projection*:

The easiest way to calculate the forecast in (15-11) is through a simple recursion called *law of iterated projection*. Suppose at date $t$ we wanted to make a one-period-ahead forecast of $Y_{t+1}$. The optimal (by setting $s = 1$ in (15-11)) is clearly

$$\hat{E}(Y_{t+1} - \mu | Y_t, Y_{t-1}, ...] = \phi_1(Y_t - \mu) + \phi_2(Y_{t-1} - \mu) + ... + \phi_p(Y_{t-p+1} - \mu). \tag{15-12}$$

Next consider a two-period-ahead forecast. Suppose that at date $t + 1$ we were to make a one-period-ahead forecast of $Y_{t+2}$. Replacing $t$ with $t + 1$ in (15-12) give s the optimal forecast as

$$\hat{E}(Y_{t+2} - \mu | Y_{t+1}, Y_t, ...] = \phi_1(Y_{t+1} - \mu) + \phi_2(Y_t - \mu) + ... + \phi_p(Y_{t-p+2} - \mu). \tag{15-13}$$

The law of iterated projections asserts that if this date $t + 1$ forecast of $Y_{t+2}$ is projected on date $t$ information, the results is the data $t$ forecast of $Y_{t+2}$. At date $t$ the values $Y_t, Y_{t-1}, ...., Y_{t-p+2}$ in (15-13) are known . Thus

$$\begin{aligned}\hat{E}(Y_{t+2} - \mu | Y_t, Y_{t-1}, ...] &= \phi_1 \hat{E}(Y_{t+1} - \mu | Y_t, Y_{t-1}, ...] + \phi_2(Y_t - \mu) + ... + \\ &\quad \phi_p(Y_{t-p+2} - \mu).\end{aligned} \tag{15-14}$$

The $s$-period-ahead forecast of an $AR(p)$ process can therefore be obtained by iterating on

$$\begin{aligned}\hat{E}(Y_{t+j} - \mu | Y_t, Y_{t-1}, ...] &= \phi_1 \hat{E}(Y_{t+j-1} - \mu | Y_t, Y_{t-1}, ...] + \phi_2 \hat{E}(Y_{t+j-2} - \\ &\quad \mu | Y_t, Y_{t-1}, ...] + ... + \phi_p \hat{E}(Y_{t+j-p} - \mu | Y_t, Y_{t-1}, ...]\end{aligned} \tag{15-15}$$

for $j = 1, 2, ..., s$ where

$$\hat{E}(Y_\tau - \mu | Y_t, Y_{t-1}, ...] = Y_\tau, \text{ for } \tau \leq t. \qquad \square$$

It is important to note that to forecast an $AR(p)$ process, an optimal $s$-period-ahead linear forecast based on an infinite number of observations $\{Y_t, Y_{t-1}, ...\}$ in fact make use of *only the $p$ most recent value* $\{Y_t, Y_{t-1}, .., Y_{t-p+1}\}$.

### 2.2.3 Forecasting an $MA(1)$ Process

An invertible $MA(1)$ process:

$$Y_t - \mu = (1 + \theta L)\varepsilon_t, \quad \text{with } |\theta| < 1.$$

**(a).** *By the Wiener-Kolmogorov formula*:

Using the Wiener-Kolmogorov formula we have

$$\hat{Y}_{t+s|t} = \mu + \left[\frac{1+\theta L}{L^s}\right]_+ \frac{1}{1+\theta L}(Y_t - \mu).$$

To forecast an $MA(1)$ process one-period-ahead ($s = 1$),

$$\left[\frac{1+\theta L}{L^1}\right]_+ = \theta,$$

and so

$$\begin{aligned}
\hat{Y}_{t+1|t} &= \mu + \frac{\theta}{1+\theta L}(Y_t - \mu) \\
&= \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \theta^3(Y_{t-2} - \mu) - ....
\end{aligned}$$

To forecast an $MA(1)$ process for $s = 2, 3, ...$ periods into the future,

$$\left[\frac{1+\theta L}{L^s}\right]_+ = 0 \quad for\ s = 2, 3, ...,$$

an so

$$\hat{Y}_{t+s|t} = \mu \quad for\ s = 2, 3, ....$$

**(b).** *By Recursive Substitution*:

An $MA(1)$ process at period $t + 1$ is

$$Y_{t+1} - \mu = \varepsilon_{t+1} + \theta\varepsilon_t.$$

At period $t$, $E(\varepsilon_{t+s}) = 0, s = 1, 2, ....$ The optimal linear 1-period-ahead forecast for a stationary $MA(1)$ process is therefore:

$$\begin{aligned}
\hat{Y}_{t+1|t} &= \mu + \theta\varepsilon_t \\
&= \mu + \theta(1 + \theta L)^{-1}(Y_t - \mu) \\
&= \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \theta^3(Y_{t-2} - \mu) - ....
\end{aligned}$$

An $MA(1)$ process at period $t+s$ is

$$Y_{t+s} - \mu = \varepsilon_{t+s} + \theta\varepsilon_{t+s-1},$$

To forecast an $MA(1)$ process for $s = 2, 3, ...$ periods into the future therefore is

$$\hat{Y}_{t+s|t} = \mu \quad for\ s = 2, 3, .... \qquad \square$$

### 2.2.4 Forecasting an $MA(q)$ Process

For an invertible $MA(q)$ process,

$$(Y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\varepsilon_t,$$

the forecast becomes

$$\hat{Y}_{t+s|t} = \mu + \left[\frac{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q}{L^s}\right]_+ \frac{1}{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q}(Y_t - \mu).$$

Now

$$\left[\frac{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q}{L^s}\right]_+ = \begin{cases} \theta_s + \theta_{s+1} L + \theta_{s+2} L^2 + ... + \theta_q L^{q-s}, & for\ s = \quad 1, 2, ..., q \\ 0, & for\ s = \quad q+1, q+2, ... \end{cases}$$

Thus for horizons of $s = 1, 2, ..., q$, the forecast is given by

$$\hat{Y}_{t+s|t} = \mu + (\theta_s + \theta_{s+1} L + \theta_{s+2} L^2 + ... + \theta_q L^{q-s})\frac{1}{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q}(Y_t - \mu).$$

A forecast farther then $q$ periods into the future is simply the unconditional mean $\mu$.

It is important to note that to forecast an $MA(q)$ process, an optimal $s$ ($s \leq q$)-period-ahead linear forecast would in principle *requires all of the historical value* of $Y$ $\{Y_t, Y_{t-1}, ...\}$.

### 2.2.5 Forecasting an $ARMA(1,1)$ Process

For an $ARMA(1,1)$ process

$$(1-\phi L)(Y_t-\mu)=(1+\theta L)\varepsilon_t, \tag{15-16}$$

that is stationary and invertible. Denote $(1+\theta L)\varepsilon_t=\eta_t$, this $ARMA(1,1)$ process can be represented as (using (13-3) on Ch. 13)

$$\begin{aligned}
Y_{t+s}-\mu &= \phi^s(Y_t-\mu)+\phi^{s-1}\underline{\eta_{t+1}}+\phi^{s-2}\eta_{t+2}+\ldots+\phi\eta_{t+s-1}+\eta_{t+s}\\
&= \phi^s(Y_t-\mu)+\phi^{s-1}\underline{(1+\theta L)\varepsilon_{t+1}}+\phi^{s-2}(1+\theta L)\varepsilon_{t+2}+\ldots\\
&\quad +\phi(1+\theta L)\varepsilon_{t+s-1}+(1+\theta L)\varepsilon_{t+s}\\
&= \phi^s(Y_t-\mu)+(\phi^{s-1}\varepsilon_{t+1}+\phi^{s-1}\underline{\theta\varepsilon_t)}+(\phi^{s-2}\varepsilon_{t+2}+\phi^{s-2}\theta\varepsilon_{t+1})+\ldots\\
&\quad +(\phi\varepsilon_{t+s-1}+\phi\theta\varepsilon_{t+s-2})+(\varepsilon_{t+s}+\theta\varepsilon_{t+s-1}).
\end{aligned}$$

Setting $E(\varepsilon_{t+h})=0, h=1,2,\ldots,s$, the optimal linear $s$-period ahead forecast for a stationary $ARMA(1,1)$ process is therefore:

$$\begin{aligned}
\hat{E}(Y_{t+s}|Y_t,Y_{t-1},\ldots] &= \mu+\phi^s(Y_t-\mu)+\phi^{s-1}\theta\varepsilon_t\\
&= \mu+\phi^s(Y_t-\mu)+\phi^{s-1}\theta\frac{1-\phi L}{1+\theta L}(Y_t-\mu) \quad (from\ (15-16))\\
&= \mu+\frac{(1+\theta L)\phi^s+\phi^{s-1}\theta(1-\phi L)}{1+\theta L}(Y_t-\mu)\\
&= \mu+\frac{\phi^s+\theta L\phi^s+\phi^{s-1}\theta-\theta L\phi^s}{1+\theta L}(Y_t-\mu)\\
&= \mu+\frac{\phi^s+\phi^{s-1}\theta}{1+\theta L}(Y_t-\mu).
\end{aligned}$$

It is important to note that to forecast an $ARMA(1,1)$ process, an optimal $s$-period-ahead linear forecast would in principle *requires all of the historical value* of $Y$ $\{Y_t,Y_{t-1},\ldots\}$.

**Exercise 1.**

Suppose that the U.S. quarterly seasonally adjusted unemployment rate for the period $1948-1972$ behaves much like the time series

$$Y_t-4.77=1.54(Y_{t-1}-4.77)-0.67(Y_{t-2}-4.77)+\varepsilon_t,$$

where the $\varepsilon_t$ are uncorrelated random $(0,1)$ variables. Let us assume that we know that this is the proper representation. The four observations updated to 1972 are $Y_{1972}=5.30, Y_{1971}=5.53, Y_{1970}=5.77$, and $Y_{1969}=5.83$. Please make a best linear forecast of $Y_{1972+s}, s=1,2,3,4$ based on the information available. ■

# 3 Forecast Based on Finite $m$ Observations

The section continues to assume that population parameters are known with certainty, but develops forecasts based on a finite $m$ observations, $\{Y_t, Y_{t-1}, ..., Y_{t-m+1}\}$

## 3.1 Approximations to Optimal Forecast In ARMA Process

For forecasting an $AR(p)$ process, an optimal $s$-period-ahead linear forecast based on an infinite number of observations $\{Y_t, Y_{t-1}, \cdots\}$ in fact make use of only the $p$ most recent value $\{Y_t, Y_{t-1}, \cdots, Y_{t-p+1}\}$. For an $MA$ or $ARMA$ process, however, we would require in principle all of the historical value of $Y$ in order to implement the formula of the proceeding section.

In reality we do not have infinite number of observations for forecasting. One approach to forecasting based on a finite number of observations is to act as if pre-sample $Y$'s were all equal to its mean value $\mu$ (or $\varepsilon = 0$). This idea is thus to use the approximation

$$\hat{E}(Y_{t+s}|1, Y_t, Y_{t-1}, ...) \cong \hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ..., Y_{t-m+1}; Y_{t-m} = \mu, Y_{t-m-1} = \mu, ...).$$

For example in the forecasting of an $MA(1)$ process, the 1 period ahead forecast

$$\hat{Y}_{t+1|t} = \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \theta^3(Y_{t-2} - \mu) - ....$$

is approximated by

$$\hat{Y}_{t+1|t} = \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \theta^3(Y_{t-2} - \mu) - .... + (-1)^{m-1}\theta^m(Y_{t-m+1} - \mu). \tag{15-17}$$

For $m$ large and $|\theta|$ small (then the real difference between $Y$ and $\mu$ will multiply a small and smaller number), this clearly gives an excellent approximation. For $|\theta|$ closer to unity, the approximation may be poor.

$\boxed{\text{Example}}$.
To forecast a MA(1) process, most software package is doing this way: because $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, therefore

$$Y_{t+1|t} = \theta\tilde{\varepsilon}_t,$$

where the realization of $\varepsilon_t$, $\tilde{\varepsilon}_t$, is recurved from

$$\tilde{\varepsilon}_t = Y_t - \theta\tilde{\varepsilon}_{t-1}, \quad t = 1, 2, ...,$$

with the assumption that $\varepsilon_0 = 0$. For example,

$$\begin{aligned} Y_{4|3} &= \theta\tilde{\varepsilon}_3 \\ &= \theta(Y_3 - \theta\tilde{\varepsilon}_2) \\ &= \theta Y_3 - \theta^2(Y_2 - \theta\tilde{\varepsilon}_1) \\ &= \theta Y_3 - \theta^2 Y_2 + \theta^3\tilde{\varepsilon}_1 \\ &= \theta Y_3 - \theta^2 Y_2 + \theta^3 Y_1. \end{aligned} \tag{15-18}$$

Substitute $m = t = 3$ and $\mu = 0$ into (15-17), we obtain (15-18). ■

### 3.2 Exact Finite-Sample Forecast by Linear Projection

An alternative approach is to calculate the exact projection of $Y_{t+1}$ on its $m$ most recent values.[4] Let

$$\mathbf{x}_t = \begin{bmatrix} 1 \\ Y_t \\ Y_{t-1} \\ . \\ . \\ . \\ Y_{t-m+1} \end{bmatrix}$$

We thus seek a linear forecast of the form

$$\hat{E}(Y_{t+1}|\mathbf{x}_t) = \boldsymbol{\alpha}'^{(m)}\mathbf{x}_t = \alpha_0^{(m)} + \alpha_1^{(m)}Y_t + \alpha_2^{(m)}Y_{t-1} + ... + \alpha_m^{(m)}Y_{t-m+1}. \tag{15-19}$$

The coefficient relating $Y_{t+1}$ to $Y_t$ in a projection of $Y_{t+1}$ on the $m$ most recent value of $Y$ is denoted $\alpha_1^{(m)}$ in (15-19). This will in general be different from the coefficient relating $Y_{t+1}$ to $Y_t$ in a projection of $Y_{t+1}$ on the $(m+1)$ most recent value of $Y$; the latter coefficient would be denoted $\alpha_1^{(m+1)}$.

From (15-2) we know that

$$E(Y_{t+1}\mathbf{x}_t') = \boldsymbol{\alpha}'E(\mathbf{x}_t\mathbf{x}_t'),$$

[4]So we do not need to assume any model such as $ARMA$ here.

or

$$\boldsymbol{\alpha}' = E(Y_{t+1}\mathbf{x}_t')[E(\mathbf{x}_t\mathbf{x}_t')]^{-1}, \tag{15-20}$$

assuming that $E(\mathbf{x}_t\mathbf{x}_t')$ is a nonsingular matrix.

Since for a covariance stationary process $Y_t$, $\gamma_j = E(Y_{t+j}-\mu)(Y_t-\mu) = E(Y_{t+j}Y_t) - \mu^2$, therefore

$$\begin{aligned} E(Y_{t+1}\mathbf{x}_t') &= E(Y_{t+1})[1 \;\; Y_t \;\; Y_{t-1} \;......\; Y_{t-m+1}] \\ &= [\mu \;\; (\gamma_1+\mu^2) \;\; (\gamma_2+\mu^2) \;.....\; (\gamma_m+\mu^2)] \end{aligned}$$

and

$$E(\mathbf{x}_t\mathbf{x}_t') = E\begin{bmatrix} 1 \\ Y_t \\ Y_{t-1} \\ . \\ . \\ . \\ Y_{t-m+1} \end{bmatrix} \begin{bmatrix} 1 & Y_t & Y_{t-1} & . & . & . & Y_{t-m+1} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \mu & \mu & . & . & . & \mu \\ \mu & \gamma_0+\mu^2 & \gamma_1+\mu^2 & . & . & . & \gamma_{m-1}+\mu^2 \\ \mu & \gamma_1+\mu^2 & \gamma_0+\mu^2 & . & . & . & \gamma_{m-2}+\mu^2 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \mu & \gamma_{m-1}+\mu^2 & \gamma_{m-2}+\mu^2 & . & . & . & \gamma_0+\mu^2 \end{bmatrix}.$$

Hence

$$\begin{aligned} \boldsymbol{\alpha}'^{(m)} &= [\mu \;\; (\gamma_1+\mu^2) \;\; (\gamma_2+\mu^2) \;.....\; (\gamma_m+\mu^2)] \times \\ & \begin{bmatrix} 1 & \mu & \mu & . & . & . & \mu \\ \mu & \gamma_0+\mu^2 & \gamma_1+\mu^2 & . & . & . & \gamma_{m-1}+\mu^2 \\ \mu & \gamma_1+\mu^2 & \gamma_0+\mu^2 & . & . & . & \gamma_{m-2}+\mu^2 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \mu & \gamma_{m-1}+\mu^2 & \gamma_{m-2}+\mu^2 & . & . & . & \gamma_0+\mu^2 \end{bmatrix}^{-1}. \end{aligned}$$

When a constant term is included in $\mathbf{x}_t$, it is more convenient to express variables in deviations from the mean. Then we could calculate the projection of $(Y_{t+1}-\mu)$ on $\mathbf{x}_t = [(Y_t-\mu),\; (Y_{t-1}-\mu),\; ......,\; (Y_{t-m+1}-\mu)]'$:

$$\hat{Y}_{t+1|t} - \mu = \alpha_1^{(m)}(Y_t-\mu) + \alpha_2^{(m)}(Y_{t-1}-\mu) + ... + \alpha_m^{(m)}(Y_{t-m+1}-\mu).$$

For this definition of $\mathbf{x}_t$ the coefficients can be calculated from (15-20) to be

$$\boldsymbol{\alpha}^{(m)} = \begin{bmatrix} \alpha_1^{(m)} \\ \alpha_2^{(m)} \\ . \\ . \\ . \\ \alpha_m^{(m)} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & . & . & . & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & . & . & . & \gamma_{m-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \gamma_{m-1} & \gamma_{m-2} & . & . & . & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ . \\ . \\ . \\ \gamma_m \end{bmatrix}. \tag{15-21}$$

To generate an $s$-period-ahead forecast $\hat{Y}_{t+s|t}$ we would use

$$\hat{Y}_{t+s|t} - \mu = \alpha_1^{(m,s)}(Y_t - \mu) + \alpha_2^{(m,s)}(Y_{t-1} - \mu) + ... + \alpha_m^{(m,s)}(Y_{t-m+1} - \mu),$$

where

$$\begin{bmatrix} \alpha_1^{(m,s)} \\ \alpha_2^{(m,s)} \\ . \\ . \\ . \\ \alpha_m^{(m,s)} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & . & . & . & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & . & . & . & \gamma_{m-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \gamma_{m-1} & \gamma_{m-2} & . & . & . & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_s \\ \gamma_{s+1} \\ . \\ . \\ . \\ \gamma_{s+m-1} \end{bmatrix}.$$

# 4 Sums of $ARMA$ Process

This section explores the nature of series that result from adding two different $ARMA$ process together.

## 4.1 Sum of an $MA(1)$ Process and White Noise

Suppose that a series $X_t$ follows a zero-mean $MA(1)$ process:

$$X_t = u_t + \delta u_{t-1},$$

where $u_t$ is white noise with variance $\sigma_u^2$. The autocovariance of $X_t$ are thus

$$E(X_t X_{t-j}) = \begin{cases} (1+\delta^2)\sigma_u^2 & for\ j = 0 \\ \delta\sigma_u^2 & for\ j = \pm 1 \\ 0 & otherwise. \end{cases} \tag{15-22}$$

Let $v_t$ indicate a separate white noise series with variance $\sigma_v^2$. Suppose, furthermore, that $v$ and $u$ are uncorrelated at all leads and lags, i.e.,

$$E(u_t v_{t-j}) = 0 \ \ for\ all\ j,$$

implying that

$$E(X_t v_{t-j}) = o\ for\ all\ j. \tag{15-23}$$

Let an observed series $Y_t$ represent the sum of the $MA(1)$ and the white noise process:

$$\begin{aligned} Y_t &= X_t + v_t \\ &= u_t + \delta u_{t-1} + v_t. \end{aligned} \tag{15-24}$$

The question now posed is: What are the time series properties of $Y$ ?

Clearly, $Y_t$ has mean zero, and its autocovariances can be deduced from (15-22) through (15-23):

$$\begin{aligned} E(Y_t Y_{t-j}) &= E(X_t + v_t)(X_{t-j} + v_{t-j}) \\ &= E(X_t X_{t-j}) + E(v_t v_{t-j}) \\ &= \begin{cases} (1+\delta^2)\sigma_u^2 + \sigma_v^2, & for\ j = 0; \\ \delta\sigma_u^2, & for\ j = \pm 1; \\ 0, & otherwise. \end{cases} \end{aligned} \tag{15-25}$$

Thus, the sum, $X_t + v_t$, is covariance-stationary, and its autocovariances are zero beyond one lags, as those for an $MA(1)$. We might naturally then ask whether there exist a zero-mean $MA(1)$ represent for $Y$,

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \tag{15-26}$$

with

$$E(\varepsilon_t\varepsilon_{t-j}) = \begin{cases} \sigma^2, & for\ j = 0; \\ 0, & otherwise, \end{cases} \tag{15-27}$$

whose autocovariance match those implied by (15-25). The autocovariance of (15-26) would be given by

$$E(Y_tY_{t-j}) = \begin{cases} (1+\theta^2)\sigma^2 & for\ j = 0 \\ \theta\sigma^2 & for\ j = \pm 1 \\ 0 & otherwise. \end{cases}$$

In order to be consistent with (15-25), it would be the case that

$$(1+\theta^2)\sigma^2 = (1+\delta^2)\sigma_u^2 + \sigma_v^2 \tag{15-28}$$

and

$$\theta\sigma^2 = \delta\sigma_u^2. \tag{15-29}$$

Taking the value[5] associated with the invertible representation $(\theta^*, \sigma^{2*})$ which is solved from (15-28) and (15-29) simultaneously. Let us consider whether (15-26) could indeed characterize the data $Y_t$ generated by (15-24). This would require

$$(1+\theta^* L)\varepsilon_t = (1+\delta L)u_t + v_t,$$

or

$$\begin{aligned} \varepsilon_t &= (1+\theta^* L)^{-1}[(1+\delta L)u_t + v_t] \\ &= (u_t - \theta^* u_{t-1} + \theta^{*2}u_{t-2} - \theta^{*3}u_{t-3} + \cdots) \\ &\quad +\delta(u_{t-1} - \theta^* u_{t-2} + \theta^{*2}u_{t-3} - \theta^{*3}u_{t-4} + \cdots) \\ &\quad +(v_t - \theta^* v_{t-1} + \theta^{*2}v_{t-2} - \theta^{*3}v_{t-3} + \cdots). \end{aligned} \tag{15-30}$$

[5]Two value of $\theta$ that satisfy (15-28) and (15-29) can be found from the quadratic formula:

$$\theta = \frac{[(1+\delta^2) + (\sigma_v^2/\sigma_u^2)] \pm \sqrt{[(1+\delta^2) + (\sigma_v^2/\sigma_u^2)]^2 - 4\delta^2}}{2\delta}$$

The series $\varepsilon_t$ defined in (15-30) is a distributed lag on past value of $u$ and $v$, so it might seem to possess a rich autocorrelation structure. In fact, it turns out to be **white noise** ! To see this, note from (15-25) that the autocovariance-generating function of $Y$ can be written

$$\begin{aligned} g_Y(z) &= \delta\sigma_u^2 z^{-1} + [(1+\delta^2)\sigma_u^2 + \sigma_v^2]z^0 + \delta\sigma_u^2 z^1 \\ &= \sigma_u^2(1+\delta z)(1+\delta z^{-1}) + \sigma_v^2, \end{aligned} \tag{15-31}$$

so the autocovariance-generating function of $\varepsilon_t = (1+\theta^* L)^{-1} Y_t$ is

$$g_\varepsilon(z) = \frac{\sigma_u^2(1+\delta z)(1+\delta z^{-1}) + \sigma_v^2}{(1+\theta^* z)(1+\theta^* z^{-1})}. \tag{15-32}$$

But $\theta^*$ and $\sigma^{*2}$ were chosen so as to make the autocovariance-generating function of $(1+\theta^* L)\varepsilon_t$, namely,

$$(1+\theta^* z)\sigma^{*2}(1+\theta^* z^{-1}), \tag{15-33}$$

identical to the right side of (15-31). Thus, (15-32) is simply equal to

$$g_\varepsilon(z) = \sigma^{2*},$$

a white noise series.

To summarize, adding an $MA(1)$ process to a white noise series with which it is uncorrelated at all leads and lags produce a new $MA(1)$ process characterized by (15-26).

### 4.2 Sum of Two Independent Moving Average Process

As a necessary preliminary to what follows, consider a stochastic process $W_t$, which is the sum of two independent moving average processes of order $q_1$ and $q_2$, respectively. That is,

$$W_t = \theta_1(L)u_t + \theta_2(L)v_t,$$

where $\theta_1(L)$ and $\theta_2(L)$ are polynomials in $L$, of order $q_1$ and $q_2$, and the white noise processes $u_t$ and $v_t$ have zero means and are mutually uncorrelated at all lead and lags. Suppose that $q = \max(q_1, q_2)$; then it is clear that the autocovariance function $\gamma_j$ for

$W_t$ must be zero for $j > q$.[6] It follows that there exists a representation of $W_t$ as a single moving average process of order $q$:

$$W_t = \theta_3(L)\varepsilon_t, \tag{15-34}$$

where $\varepsilon_t$ is a white noise process with mean zero. Thus the sum of two uncorrelated moving average processes is another moving average process, whose order is the same as that of the component process of higher order.

### 4.3 Sum of an $ARMA(p, q)$ Plus White Noise

Consider the general stationary and invertible $ARMA(p, q)$ model

$$\phi(L)X_t = \theta(L)u_t,$$

where $u_t$ is a white noise process with variance $\sigma_u^2$. Let $v_t$ indicate a separate white noise series with variance $\sigma_v^2$. Suppose, furthermore, that $v$ and $u$ are uncorrelated at all leads and lags.

Let an observed series $Y_t$ represent the sum of the $ARMA(p, q)$ and the white noise process:

$$Y_t = X_t + v_t. \tag{15-35}$$

In general we have

$$\begin{aligned} \phi(L)Y_t &= \phi(L)X_t + \phi(L)v_t \\ &= \theta(L)u_t + \phi(L)v_t. \end{aligned}$$

Let $q* = \max(q, p)$ and from (15-34) we have that $Y_t$ is an $ARMA(p, q*)$ process.

[6] To see this, let $W_t = X_t + Y_t$, then

$$\begin{aligned} \gamma_j^W = E(W_t W_{t-j}) &= E(X_t + Y_t)(X_{t-j} + Y_{t-j}) \\ &= E(X_t X_{t-j})(Y_t Y_{t-j}) \\ &= \begin{cases} \gamma_j^X + \gamma_j^Y & for\ j = 0, \pm 1, \pm 2, ..., \ \pm q, \\ 0 & otherwise. \end{cases} \end{aligned}$$

### 4.4 Adding Two Autoregressive Processes

Suppose now that $X_t$ are $AR(p_1)$ and $W_t$ are $AR(p_2)$ process:

$$\begin{aligned} \phi_1(L)X_t &= u_t \\ \phi_2(L)W_t &= v_t, \end{aligned} \tag{15-36}$$

where $u_t$ and $v_t$ are uncorrelated white noise at all lead and lags. Again suppose that we observe

$$Y_t = X_t + W_t,$$

we wish to determine the nature of the observed process $Y_t$.

From (15-36) we have

$$\begin{aligned} \phi_2(L)\phi_1(L)X_t &= \phi_2(L)u_t \\ \phi_1(L)\phi_2(L)W_t &= \phi_1(L)v_t, \end{aligned}$$

then

$$\phi_1(L)\phi_2(L)(X_t + W_t) = \phi_1(L)v_t + \phi_2(L)u_t$$

or

$$\phi_1(L)\phi_2(L)Y_t = \phi_1(L)v_t + \phi_2(L)u_t.$$

Therefore $Y_t$ is an $ARMA(p_1 + p_2, \max\{p_1, p_2\})$ process from (15-34).

# 5 Wold's Decomposition

All of the covariance stationary *ARMA* process considered in Chapter 14 can be written in the form

$$Y_t = \mu + \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}, \tag{15-37}$$

where $\varepsilon_t$ is a white noise process and $\sum_{j=0}^{\infty} \varphi_j^2 < \infty$ with $\varphi_0 = 1$.

One might think that we are able to write all these processes in the form of (15-37) because the discussion was restricted to a convenient class of models (parametric *ARMA* model). However, the following result establishes that the representation (15-37) is in fact *fundamental for any covariance-stationary* time series.

**Theorem** (Wold's Decomposition):
Let $Y_t$ be **any** covariance stationary stochastic process with $EY_t = 0$. Then it can be written as

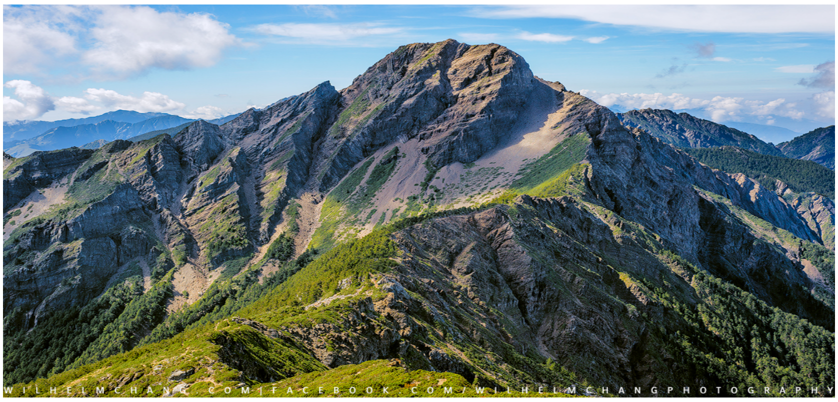
$$Y_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j} + \eta_t, \tag{15-38}$$

where $\varphi_0 = 1$ and where $\sum_{j=0}^{\infty} \varphi_j^2 < \infty$, $E\varepsilon_t^2 = \sigma^2 \geq 0$, $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$, $E\varepsilon_t = 0$ and $E\varepsilon_t \eta_s = 0$ for all $t$ and $s$; and $\eta_t$ is a process that can be predicted arbitrary well by a linear function of only past value of $Y_t$, i.e., $\eta_t$ is linearly deterministic.[7] Furthermore, $\varepsilon_t = Y_t - \hat{E}(Y_t | Y_{t-1}, Y_{t-2}, \cdots)$, i.e. the term $\varepsilon_t$ is a white noise and represents the error made in forecasting $Y_t$ on the basis of a linear function of lagged $Y$.

**Proof**:
See Sargent (1987, pp. 286-90) or Fuller (1996, pp. 94-98), or Brockwell and Davis (1991, pp. 187-89). ∎

Wold's decomposition is important for us because it provides an explanation of the sense in which *ARMA* model (stochastic difference equation) provide a general model for the indeterministic part of any univariate stationary stochastic process, and also the sense in which there exist a white-noise process $\varepsilon_t$ (which is in fact the forecast error from one-period ahead linear projection) that is the building block for the indeterministic part of $Y_t$.

[7] When $\eta_t = 0$, then the process (15-38) is called purely linear indeterministic.



Jade Mountain (3952m). The highest peak in Taiwan

*End of this Chapter*