Time Series Econometrics, 2ST111

Lecture 3. Stationary ARMA Processes and Forecasting

Yukai Yang

Department of Statistics, Uppsala University

Outline of Today's Lecture

- Stationary ARMA Processes (pp.43-71 in Hamilton)
  - Expectations, Stationarity & Ergodicity
  - MA, AR & ARMA Processes
  - Invertibility
- Forecasting (pp.72-116 in Hamilton)
  - Based on Conditional Expectation
  - Based on Linear Projection
  - Based on an Infinite Number of Observations
  - Based on a Finite Number of Observations

Stochastic Processes

- Consider the elements of an observed time series as being realizations (outcomes) of a stochastic (random) process. (Recall the graph in Lecture 1)
- In modeling such a process, we attempt to capture the characteristics.
- The univariate ARMA processes provide a very useful class of models for describing the dynamics of an individual time series.

Stochastic Processes

Suppose that we have observed a sample

$$\{y_t\}_{t=1}^{T} = \{y_1, y_2, ..., y_T\}$$

of size T of some random variables $\{Y_t\}_{t=1}^{T}$.

If we could observe, which is not possible, the process for an infinite period, then the full sample is

$$\{y_t\}_{t=-\infty}^{\infty} = \{..., y_{-1}, y_0 y_1, y_2, ..., y_T, y_{T+1}, ...\},$$

from the random variables $\{Y_t\}_{t=-\infty}^{\infty}$.

They are both one realization of the underlying data generating process but with different sample sizes!

Stochastic Processes

If we could independently repeat the data generating process at time t for I times, then we can collect

$$\{y_t^{(i)}\}_{i=1}^{I} = \{y_t^{(1)}, y_t^{(2)}, ..., y_t^{(I)}\}.$$

We obtain the cross-sectional data. Very often, this is impossible in time series.

Denote $f_{Y_t}(y)$ the unconditional density function of the random number Y_t at time t. We have

$$f_{Y_t}(y) \geq 0, \quad \text{and} \quad \int_{-\infty}^{\infty} f_{Y_t}(y)\,\mathrm{d}y = 1$$

Then $f_{Y_t}(y_t)$ is the value of the density function when the argument y equals the observation y_t, to be precise.

Expectation

The expectation of the tth observation of a time series refers to the following integral, provided it exists:

$$\mathrm{E}(Y_t) = \int_{-\infty}^{\infty} y \, f_{Y_t}(y) \, \mathrm{d}y. \tag{1}$$

You will see or have seen the notation in the literature like $y \cdot f_{Y_t}$, which implies the integral above (most probably, it is $y \cdot \mu$ where $\mu = f_{Y_t}$).

This existence is also termed integrable.

The ensemble average or ensemble mean of the observations at time t

$$I^{-1} \sum_{i=1}^{I} y_t^{(i)} \xrightarrow{p} \mathrm{E}(Y_t) \tag{2}$$

Strong LLN.

So far, Y_t is the implied random variable.

Expectation

Some expectations

- $Y_t = \mu + \varepsilon_t$ implies $\mathrm{E}(Y_t) = \mu$.
- $Y_t = \beta t + \varepsilon_t$ implies $\mathrm{E}(Y_t) = \beta t$.
- If the expectation is time-varying, for example, a function of the date like above, we denote $\mathrm{E}(Y_t) = \mu_t$.

Variance

The unconditional variance of the random variable Y_t is defined as follows

$$\gamma_{0t} = \mathrm{Var}(Y_t) = \mathrm{E}(Y_t - \mu_t)^2 = \int_{-\infty}^{\infty} (y - \mu_t)^2 f_{Y_t}(y)\, \mathrm{d}y \tag{3}$$

Note that $\mathrm{E}(Y_t - \mu_t)^2 = \mathrm{E}\left((Y_t - \mu_t)^2\right)$, which differs from $\mathrm{E}^2(Y_t - \mu_t) = (\mathrm{E}(Y_t - \mu_t))^2$.

If $Y_t = \beta t + \varepsilon_t$, and $\varepsilon_t \sim (0, \sigma^2)$, then $\gamma_{0t} = \mathrm{E}(Y_t - \beta t)^2 = \mathrm{E}(\varepsilon_t^2) = \sigma^2$.

Autocovariance

The jth autocovariance

$$\begin{aligned}\gamma_{jt} &= \mathrm{Cov}(Y_t, Y_{t-j}) = \mathrm{E}(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - \mu_t)(x - \mu_{t-j}) f_{Y_t, Y_{t-j}}(y, x)\,\mathrm{d}y\mathrm{d}x \qquad (4)\end{aligned}$$

where $f_{Y_t, Y_{t-j}}(y, x)$ is the joint density function of the random variables Y_t and Y_{t-j}.

Note that
$\int_{-\infty}^{\infty} f_{Y_t, Y_{t-j}}(y, x)\,\mathrm{d}x = f_{Y_t}(y)$ and $\int_{-\infty}^{\infty} f_{Y_t, Y_{t-j}}(y, x)\,\mathrm{d}y = f_{Y_{t-j}}(x)$.

This is telling the same story as [3.1.10] on pp.45 in Hamilton, but they look so different. Do you know why?

Autocovariance

- The autocovariance is the covariance between Y_t and its own lag.
- The 0th autocovariance, γ_{0t}, is the variance of Y_t.
- We have the ensemble average, if the pair of the observations $(y_t^{(i)}, y_{t-j}^{(i)})$ at time t and $t-j$ can be repeatedly independently sampled:

$$I^{-1} \sum_{i=1}^{I} (y_t^{(i)} - \mu_t)(y_{t-j}^{(i)} - \mu_{t-j}) \xrightarrow{p} \gamma_{jt}.$$

Stationarity

If neither the expectation μ_t nor the autocovariances γ_{jt} depend on the time t, the the process for Y_t is said to be covariance-stationary or weakly stationary.

$$\begin{aligned} E(Y_t) &= \mu, & \text{for all } t; \\ E(Y_t - \mu)(Y_{t-j} - \mu) &= \gamma_j < \infty, & \text{for all } t \text{ and any } j. \end{aligned} \tag{5}$$

Example:
$Y_t = \mu + \varepsilon_t$, where $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$,
with $E(Y_t) = \mu$, $\gamma_{0t} = \sigma^2$ and $\gamma_{jt} = 0$ for $j \neq 0$.

If a process is covariance-stationary, then it follows that

$$\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu) = E(Y_{t-j} - \mu)(Y_t - \mu) = \gamma_{-j}. \tag{6}$$

Stationarity

A process is said to be strictly stationary if, for any (integer) values of $j_1, j_2, ..., j_n$, the joint distribution of $(Y_t, Y_{t+j_1}, Y_{t+j_2}, ..., Y_{t+j_n})$ depends not on the time t, but only on $j_1, j_2, ..., j_n$.

Remarks:

- If a strictly stationary process has finite autocovariances, then it is covariance-stationary.
- A covariance-stationary process may not be strictly stationary, as some higher moments (> 2) can be time dependent.
- The assumption of strict stationarity is too strong to verify in most cases in practice.
- By default, "stationary" means "covariance-stationary".

Gaussian Process

A process $\{Y_t\}$ is said to be Gaussian, if the joint density

$$f_{Y_t, Y_{t+j_1}, Y_{t+j_2}, \dots, Y_{t+j_n}}(y_0, y_1, y_2, \dots, y_n)$$

is multivariate Gaussian for any $j_1, j_2, \dots, j_n$.

A covariance-stationary Gaussian process is strictly stationary.

Defn: $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \iff \forall v \neq 0, \; v^T X \sim N(\dots).$

Ergodicity

Motivation:
Since we are not dealing with cross-sectional data, it is not realistic in practice to have $y_t^{(1)}, y_t^{(2)}, ..., y_t^{(I)}$ at time t. We only have one single realization y_t. Can we still infer something from the following time average?

$$\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$$

Whether the time average as such eventually converge to the ensemble $\mathrm{E}(Y_t)$ for a stationary process has to do with ergodicity.

Ergodicity

A stationary process is said to be ergodic for the mean, if the time average converges in probability to $\mathrm{E}(Y_t)$ as $T \to \infty$.

$$T^{-1} \sum_{t=1}^{T} y_t \xrightarrow{p} \mathrm{E}(Y_t) \tag{7}$$

Remarks

- A process is ergodic for the mean provided that the autocovariance γ_j goes to zero sufficiently fast as $j \to \infty$.
- We will see (chapter 7 in Hamilton) that if $\sum_{j=0}^{\infty} |\gamma_j| < \infty$ (absolute summability) holds for a stationary process Y_t, then Y_t is ergodic for the mean.

Ergodicity

A stationary process is said to be ergodic for second moments, if

$$(T-j)^{-1} \sum_{t=j+1}^{T} (y_t - \mu)(y_{t-j} - \mu) \xrightarrow{p} \gamma_j \tag{8}$$

for all j.

Remarks

- If Y_t is a stationary Gaussian process, the absolute summability $\sum_{j=0}^{\infty} |\gamma_j| < \infty$ is sufficient for ergodicity for all moments.
- The Gaussian assumption offers great convenience.
- Sufficient conditions for more general cases can be found in chapter 7 in Hamilton.

Example: Stationary but Not Ergodic

Suppose that

$$Y_t = U_t + Z$$

where $U_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_u^2)$, $Z \overset{iid}{\sim} \mathcal{N}(0, 1)$, and U_t and Z are independent to each other.

We have $\mathrm{E}(Y_t) = 0$, and

$$\begin{aligned} \mathrm{Cov}(Y_t, Y_{t-j}) &= \mathrm{E}(U_t + Z)(U_{t-j} + Z) \\ &= \mathrm{Var}(Z) = 1. \end{aligned}$$

Thus, the process Y_t is stationary with $\gamma_j = 1$ for all j.

Example: Stationary but Not Ergodic

However, when you observe the sample y_t (note that you cannot see u_t and z), the time average

$$\bar{y} = T^{-1} \sum_{t=1}^{T} y_t = T^{-1} \sum_{t=1}^{T} u_t + z \xrightarrow{p} z,$$

and $z \neq 0$ almost surely.

Even worse, when you resample it for I times, you will find that $\bar{y}^{(i)}$ are distinct, for $i = 1, ..., I$, as $\bar{y}^{(i)} \xrightarrow{p} z^{(i)}$ and $z^{(i)}$ are distinct.

In reality, normally the data generating cannot be repeated. Most probably, you will regard z as a constant, due to the conditioning like $\mathrm{E}(Y_t|z) = z$. This process, therefore, becomes ergodic for the mean.

White Noise

A white noise process is a sequence $\{\varepsilon_t\}_{-\infty}^{\infty}$ whose elements satisfy

$$\mathrm{E}(\varepsilon_t) = 0 \tag{9}$$
$$\mathrm{E}(\varepsilon_t^2) = \sigma^2 \tag{10}$$
$$\mathrm{E}(\varepsilon_t \varepsilon_{t-j}) = 0 \quad \text{if } j \neq 0 \tag{11}$$

for all integers t and j.

A stronger version of the white noise process is to replace (11) by

$$\varepsilon_t \text{ and } \varepsilon_{t-j} \text{ are independent if } j \neq 0, \tag{12}$$

which is said to be the independent white noise process.

White Noise

Remarks:

- The white noise process is the basic building block for the ARMA processes.
- The white noise process, by construction, is stationary.
- The white noise process is called Gaussian white noise process if any joint distribution of $\varepsilon_t, \varepsilon_{t+j_1}, ..., \varepsilon_{t+j_n}$ is Gaussian distributed.
- A Gaussian white noise process is strictly stationary.

Autocorrelation

The jth autocorrelation of a stationary process is defined as

$$\rho_j = \gamma_j / \gamma_0. \tag{13}$$

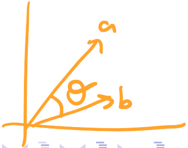
Remarks

- Autocorrelation comes from the correlation between Y_t and Y_{t-j}

$$\text{Corr}(Y_t, Y_{t-j}) = \frac{\text{Cov}(Y_t, Y_{t-j})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-j})}} = \frac{\gamma_j}{\gamma_0} = \rho_j \tag{14}$$

- By the Cauchy-Schwarz inequality, $|\rho_j| \leq 1$ for all j.
- $\rho_0 = 1$.

$$a \cdot b = \|a\| \cdot \|b\| \cdot \cos\theta$$
$$\leq \|a\| \cdot \|b\|$$



Moving Average Processes

Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be a white noise process. The qth-order moving average process or MA(q) is given by

$$Y_t = \mu + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \tag{15}$$

where $\theta_0 = 1$ and $\theta_i \in \mathbb{R}$. It can be shown that

- The expectation $\mathrm{E}(Y_t) = \mu$.
- $\{Y_t\}_{t=-\infty}^{\infty}$ is stationary for all $\theta_i \in \mathbb{R}$, with

$$\gamma_j = \begin{cases} 0 & \text{for} \quad j > q \\ \sigma^2 \sum_{i=0}^{q-j} \theta_i \theta_{i+j} & \text{for} \quad j = 0, ..., q \\ \gamma_{-j} & \text{for} \quad j < 0 \end{cases} \tag{16}$$

  hence, $\sum_{j=0}^{\infty} |\gamma_j| < \infty$ (absolutely summable) and $\rho_j = 0$ for $j > q$.
- If $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a Gaussian white noise process, then $\{Y_t\}_{t=-\infty}^{\infty}$ is ergodic for all moments.

Moving Average Processes

Likewise, the infinite-order moving average process or MA(∞) is given by

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{17}$$

where $\psi_0 = 1$ and $\psi_i \in \mathbb{R}$. $\mathrm{E}(Y_t) = \mu$.

Recall the lag operator.

- The MA(q) can be written as

$$Y_t = \mu + \theta(L)\varepsilon_t \tag{18}$$

  where $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$.
- The MA(∞) can be written as

$$Y_t = \mu + \psi(L)\varepsilon_t \tag{19}$$

  where $\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + ...$

Appendix 3.A on pp.69-70 in Hamilton shows that MA(∞) is a well defined stationary process provided that

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty \quad \text{square-summability}, \tag{20}$$

or, stronger and more often used,

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \quad \text{absolute summability}, \tag{21}$$

We have $\sum_{i=0}^{\infty} |\psi_i| < \infty \implies \sum_{i=0}^{\infty} \psi_i^2 < \infty$.

Moving Average Processes

Remarks for MA(∞)

- The variance is $\text{Var}(Y_t) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$.
- The autocovariance is

$$\gamma_j = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j} < \infty, \quad j = 0, 1, 2, ... \tag{22}$$

- If the coefficients ψ_i are absolutely summable, the corresponding autocovariance is absolutely summable

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty. \tag{23}$$

  See pp.70 in Hamilton.
- Recall (pp.15 in the slides) that if the autocovariance is absolutely summable, the MA(∞) is ergodic for the mean.
- If in addition ε_t is Gaussian, then you know... ergodic for all moments.

Autoregressive Processes

Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be a white noise process. The *p*th-order autoregressive process or AR(p) is given by

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t, \tag{24}$$

where $c, \phi_i \in \mathbb{R}$. Alternatively we can write

$$\phi(L) Y_t = c + \varepsilon_t, \tag{25}$$

where $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$.

Let $w_t = c + \varepsilon_t$. From Lecturer 2, we know that this difference equation is stable when the roots of $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p = 0$ lie outside the unit disk, or equivalent by denoting $\lambda = 1/z$, the roots (eigenvalues of the companion matrix) of $\lambda^p - \phi_1 \lambda^{p-1} - ... - \phi_p = 0$ lie inside the unit disk.

Proposition: The AR(p) process is stationary, if the corresponding difference equation is stable.

If the difference equation is stable, then Y_t has the MA(∞) representation:

$$Y_t = \mu + \psi(L)\varepsilon_t \tag{26}$$

where

$$\mu = c\phi^{-1}(L) = c\phi^{-1}(1) = \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} \tag{27}$$

$$\begin{aligned}
\psi(L) &= \phi^{-1}(L) = (1 - \phi_1 L - \ldots - \phi_p L^p)^{-1} \\
&= (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1} \ldots (1 - \lambda_p L)^{-1} \quad (\text{fact: } |\lambda_i| < 1) \\
&= \left(\sum_{i=0}^{\infty} \lambda_1^i L^i\right)\left(\sum_{i=0}^{\infty} \lambda_2^i L^i\right) \ldots \left(\sum_{i=0}^{\infty} \lambda_p^i L^i\right) \\
&= 1 + \psi_1 L + \psi_2 L^2 + \ldots
\end{aligned} \tag{28}$$

Stability and Absolute Summability

- A Cauchy sequence α_j, $j = 1, \ldots$ is a sequence satisfying that, for any small positive number ϵ, there exists a N such that $|\alpha_n - \alpha_m| < \epsilon$ for any $n, m > N$.
- A sequence is convergent iff it is a Cauchy sequence.
- Given $|\lambda| < 1$, $1 - \lambda L$ is stable and has the inverse $\sum_{i=0}^{\infty} \lambda^i L^i$ which has absolutely summable coefficients $\sum_{i=0}^{\infty} |\lambda|^i < \infty$.

  To see this, define

$$\alpha_j = \sum_{i=0}^{j} |\lambda|^i.$$

  Assuming $n > m$ without loss of generality,
  $|\alpha_n - \alpha_m| = \sum_{i=m+1}^{n} |\lambda|^i = |\lambda|^{m+1}(1 - |\lambda|^{n-m})/(1 - |\lambda|)$ goes to zero. Thus, α_j is Cauchy and then it is convergent (absolute summability).

Stability and Absolute Summability

- If two lag polynomials are both absolutely summable, its product is absolutely summable as well.

$$\sum_{i=0}^{\infty} \phi_i L^i \quad \text{with} \quad \sum_{i=0}^{\infty} |\phi_i| < \infty$$
$$\sum_{i=0}^{\infty} \psi_i L^i \quad \text{with} \quad \sum_{i=0}^{\infty} |\psi_i| < \infty$$

We need to check whether the lag polynomial $\left(\sum_{i=0}^{\infty} \phi_i L^i\right) \left(\sum_{i=0}^{\infty} \psi_i L^i\right)$ has absolutely summable coefficients.

Stability and Absolute Summability

The product $\left(\sum_{i=0}^{\infty} \phi_i L^i\right)\left(\sum_{i=0}^{\infty} \psi_i L^i\right)$ has the terms

$$
\begin{array}{cccccc}
 & L^0 & L^1 & L^2 & L^3 & \dots \\
\phi_0 \cdot & \psi_0 & \psi_1 & \psi_2 & \psi_3 & \dots \\
\phi_1 \cdot & & \psi_0 & \psi_1 & \psi_2 & \dots \\
\phi_2 \cdot & & & \psi_0 & \psi_1 & \dots \\
\vdots & & & & \vdots & \vdots
\end{array}
$$

They are

$$
\begin{aligned}
& \phi_0\psi_0 + (\phi_0\psi_1 + \phi_1\psi_0)L + (\phi_0\psi_2 + \phi_1\psi_1 + \phi_2\psi_0)L^2 + \dots \\
= & \sum_{k=0}^{\infty} \left(\sum_{i+j=k} \phi_i \psi_j \right) L^k.
\end{aligned}
$$

Then we need to check $\sum_{k=0}^{\infty} \left| \sum_{i+j=k} \phi_i \psi_j \right| < \infty$.

Stability and Absolute Summability

We have the inequality

$$\sum_{k=0}^{\infty} \left| \sum_{i+j=k} \phi_i \psi_j \right| \leq \sum_{k=0}^{\infty} \sum_{i+j=k} |\phi_i||\psi_j|.$$

And

$$\sum_{k=0}^{\infty} \sum_{i+j=k} |\phi_i||\psi_j| = \left(\sum_{i=0}^{\infty} |\phi_i| \right) \left(\sum_{i=0}^{\infty} |\psi_i| \right) < \infty \quad Q.E.D.$$

Conclusion : If a p-order lag polynomial is stable, then its inverse polynomial has absolutely summable coefficients.

$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} - \varepsilon_t$

$\phi(L) Y_t = c + \varepsilon_t$

Now let us go back to AP(p).

Proposition: The AR(p) process is stationary, if the corresponding difference equation is stable.

$Y_t = \phi^{-1}(L)(c + \varepsilon_t)$

If the difference equation is stable, then Y_t has the MA(∞)

$$Y_t = \mu + \psi(L)\varepsilon_t$$

$= \psi(L)(c + \varepsilon_t)$

$= \psi(L)c + \psi(L)\varepsilon_t$

$= \underbrace{\psi(1)c}_{\mu} + \psi(L)\varepsilon_t$

where

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

$Y_t - \mu = \psi(L)\varepsilon_t$

$$\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \ldots$$

$\Leftrightarrow \phi(L)\underbrace{(Y_t - \mu)}_{Z_t} = \varepsilon_t$

The coefficients ψ_i are absolutely summable, definitely.

$Y_t - \mu = Z_t = \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p}$

Autoregressive Processes

The autocovariances of an AR(p) process are given by

$$\gamma_j = \begin{cases} \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + ... + \phi_p\gamma_{j-p} & \text{for} \quad j = 1, 2, ... \\ \phi_1\gamma_1 + \phi_2\gamma_2 + ... + \phi_p\gamma_p + \sigma^2 & \text{for} \quad j = 0 \\ \gamma_{-j} & \text{for} \quad j < 0 \end{cases} \tag{29}$$

Remarks

- Actually the system of equations (29) for $j = 0, 1, ..., p$ can be solved for $\gamma_0, \gamma_1, ..., \gamma_p$, by using $\gamma_j = \gamma_{-j}$, as functions of $\sigma^2, \phi_1, ..., \phi_p$.
- Recall the autocovariances of the stationary MA(∞) process. The same result (absolutely summable autocovariances) applies here if the AR(p) is stable.

Autoregressive Processes

The autocorrelations of an AR(p) process are given by

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + ... + \phi_p \rho_{j-p}, \quad j = 1, 2, ... \tag{30}$$

the so-called Yule-Walker equations.

Note that the autocovariances and the autocorrelations follow the same pth-order difference equation as the AR(p) process itself.

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t \iff {}^{*}\phi(L) Z_t = \varepsilon_t$$

$$Z_t = Y_t - \mu, \quad \mu = \frac{c}{\phi(1)}$$

$$(*) \; Z_t = \sum_{t=1}^{p} \phi_i Z_{t-i} + \varepsilon_t$$

$$E(Z_t Z_{t-1}) = E\left(\phi_1 Z_{t-1}^2 + \phi_2 Z_{t-2} Z_{t-1} + \ldots + \varepsilon_t Z_t\right]$$

$$E(Y_t - \mu)(Y_{t-1} - \mu) = \gamma_1$$

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1 + \ldots + \phi_p \gamma_p$$

$$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0 + \phi_3 \gamma_1 , \ldots$$

$$\gamma_3 = \phi_1 \gamma_2 + \phi_2 \gamma_1 + \phi_4 \gamma_0 \ldots$$

Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be a white noise process. The ARMA(p, q) is given by

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}, \tag{31}$$

where $c, \phi_i, \theta_i \in \mathbb{R}$ and $\theta_0 = 1$. Alternatively we can write

$$\phi(L) Y_t = c + \theta(L) \varepsilon_t, \tag{32}$$

where $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$, and $\theta(L) = 1 + \theta_1 L + \ldots + \theta_p L^p$.

Autoregressive Moving Average Processes

Assuming that the roots of $\phi(z) = 0$ lie outside the unit disk, both sides of (32) can be divided by $\phi(L)$ to obtain

$$Y_t = \mu + \psi(L)\varepsilon_t \tag{33}$$

where where

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - ... - \phi_p}$$

$$\psi(L) \quad = \quad \phi^{-1}(L)\theta(L) = 1 + \psi_1 L + \psi_2 L^2 + ...$$

The coefficients ψ_i are absolutely summable. Note that $\theta(L)$ has finite number of coefficients, and hence it is absolutely summable.

Autoregressive Moving Average Processes

The autocovariances of an ARMA(p, q) process can be computed using standard methods, for $j > q$ they are given by

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + ... + \phi_p \gamma_{j-p}, \quad j = q+1, q+2, ...$$

Remarks:

- The stationarity of the ARMA(p, q) process, where p, q are finite, depends entirely on the stability of $\phi(L)$, not on $\theta(L)$.
- An ARMA(p, q) process will have more complicated autocovariances γ_j for $j = 1, ..., q$ than would the corresponding AR(p) process.
- There is a potential for redundant parameterization with ARMA processes, see pp.60-61 in Hamilton.

Inveritibility

Consider the MA(1) process

$$Y_t - \mu = (1 + \theta L)\varepsilon_t.$$

Provided that $|\theta| < 1$, both sides of the equation can be multiplied by $(1 + \theta L)^{-1}$, where

$$(1 + \theta L)^{-1} = (1 - (-\theta)L)^{-1} = 1 + (-\theta)L + (-\theta)^2 L^2 + (-\theta)^3 L^3 + \ldots$$

Then we have

$$(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \ldots)(Y_t - \mu) = \varepsilon_t$$

which could be viewed as an AR(∞) representation.

Inveritibility

Remarks:

- If an MA(1) representation can be rewritten as an AR(∞) representation by inverting $(1 + \theta L)$, then the MA(1) representation is said to be invertible.
- For an MA(1) process invertibility requires $|\theta| < 1$.
- For any invertible MA(1) representation, there is a noninvertible MA(1) representation with the same first and second moments as the invertible representation. See pp.65 in Hamilton for details.

Inveritibility

Remarks:

- Either representation could be used as an equally valid description of any given MA(1) process.
- For estimation and forecasting purposes, we prefer to work with the invertible representation.
- The innovation (noise term, error term) associated with the invertible representation is sometimes called the fundamental innovation.
- The concept of invertibility can be extended to the general MA(q) process.

Forecasts Based on Conditional Expectation

Suppose we are interested in forecasting the random variable Y_{t+s} based on a set of variable $\mathbf{x}_t$ available at time t.

$$\mathbf{x}_t = (1, y_t, y_{t-1}, ..., y_{t-m+1})' \tag{34}$$

Let $Y^*_{t+s|t}$ denote such a s-step ahead forecast ($s = 1, 2, ...$). Actually it is a function of $\mathbf{x}_t$.

The performance of the forecast $Y^*_{t+s|t}$ is evaluated in terms of some loss function $g : \mathbb{R} \to \mathbb{R}$.

Consider the quadratic loss function $g(x) = (Y_{t+s} - x)^2$. We choose the forecast $Y^*_{t+s|t}$ to minimize

$$\mathrm{E}_t\, g(x).$$

Forecasts Based on Conditional Expectation

The mean squared error (MSE) associated with $Y^*_{t+s|t}$ is given by

$$\mathrm{E}(Y_{t+s} - Y^*_{t+s|t})^2 \tag{35}$$

We are actually finding a functional form for $Y^*_{t+s|t}$ with the argument $\mathbf{x}_t$ in order to minimize the expected loss function $\mathrm{E}g$.

Suppose that $Y^*_{t+s|t} = h(\mathbf{x}_t)$ for some function $h(\cdot)$, then the forecast that minimizes

$$\mathrm{E}(Y_{t+s} - Y^*_{t+s|t})^2 = \mathrm{E}(Y_{t+s} - h(\mathbf{x}_t))^2 \tag{36}$$

is given by $h(\mathbf{x}_t) = \mathrm{E}(Y_{t+s}|\mathbf{x}_t)$.

Forecasts Based on Linear Projection

Let $\mathbf{h}'\mathbf{x}_t$ denote any arbitrary linear forecasting rule, then the forecast that minimizes

$$\mathrm{E}(Y_{t+s} - \mathbf{h}'\mathbf{x}_t)^2 \tag{37}$$

is given by $Y^*_{t+s|t} = \hat{\mathbf{h}}'\mathbf{x}_t$, where $\hat{\mathbf{h}}'\mathbf{x}_t$ satisfies

$$\mathrm{E}(Y_{t+s} - \hat{\mathbf{h}}'\mathbf{x}_t)\mathbf{x}_t' = \mathbf{0}' \tag{38}$$

Forecasts Based on Linear Projection

Remarks:

- $\hat{\mathbf{h}}'\mathbf{x}_t$ is called the linear projection of $Y + t + s$ on $\mathbf{x}_t$ and is the optimal linear forecast.
- Since $\mathrm{E}(Y_{t+s}|\mathbf{x}_t)$ offers the best possible forecast (in terms of MSE), we have that

$$\mathrm{E}(Y_{t+s} - \hat{\mathbf{h}}'\mathbf{x}_t)^2 \geq \mathrm{E}(Y_{t+s} - \mathrm{E}(Y_{t+s}|\mathbf{x}_t)) \tag{39}$$

- By (38),

$$\hat{\mathbf{h}} = [\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')]^{-1}\mathrm{E}(\mathbf{x}_t Y_{t+s}) \tag{40}$$

- Hamilton uses the symbol $\hat{E}$ to indicate a linear projection on a vector of random variables along with a constant term.
- Linear projection is closely related to OLS regression.

Forecasts Based on $\varepsilon_t, \varepsilon_{t-1}, \ldots$

Consider a process with MA(∞) representation

$$Y_t - \mu = \psi(L)\varepsilon_t \tag{41}$$

where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a white noise process, and $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$ with $\psi_0 = 1$ and is absolutely summable.

In addition, assume that, for simplicity, $\varepsilon_t, \varepsilon_{t-1}, \ldots$ are observed and the parameters μ and $\psi_1, \psi_2, \ldots$ are known.

We are going to forecast Y_{t+s}

$$Y_{t+s} = \mu + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \ldots + \psi_{s-1}\varepsilon_{t+1} + \psi_s \varepsilon_t + \ldots$$

The optimal linear forecast is

$$\hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots) = \mu + \psi_s \varepsilon_t + \ldots$$

Forecasts Based on $\varepsilon_t, \varepsilon_{t-1}, ...$

The accompanying forecast error

$$Y_{t+s} - \hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...) = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + ... + \psi_{s-1}\varepsilon_{t+1}$$

And MSE

$$\mathrm{E}(Y_{t+s} - \hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...))^2 = (1 + \psi_1^2 + ... + \psi_{s-1}^2)\sigma^2$$

In particular, if Y_t follows an MA(q) process with $\psi(L) = 1 + \theta_1 L + ... + \theta_q L^q$, then the MSE increases with the increasing of s until $s = q$.

The forecast for $s > q$ is just μ and the MSE is always $(1 + \theta_1^2 + ... + \theta_q^2)\sigma^2$.

Forecasts Based on $\varepsilon_t, \varepsilon_{t-1}, ...$

It is convenient to introduce the compact lag operator expression of the s-step ahead forecast $\hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...)$.

First consider dividing $\psi(L)$ by L^s

$$\frac{\psi(L)}{L^s} = L^{-s} + \psi_1 L^{1-s} + \psi_2 L^{2-s} + ... + \psi_{s-1} L^{-1} + \psi_s L^0 + \psi_{s+1} L^1 + ...$$

and let $[\cdot]_+$ denote the annihilation operator, which replaces negative powers of L by zero,

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \psi_s L^0 + \psi_{s+1} L^1 + \psi_{s+2} L^2 + ...$$

Hence, we have the compact form

$$\hat{E}(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...) = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \varepsilon_t \tag{42}$$

Forecasts Based on $y_t, y_{t-1}, \ldots$

In practice, we observe $y_t, y_{t-1}, \ldots$, but not $\varepsilon_t, \varepsilon_{t-1}, \ldots$.

Suppose that $Y_t - \mu = \psi(L)\varepsilon_t$ has an AR(∞) representation given by $\eta(L)(Y_t - \mu) = \varepsilon_t$, where $\eta(L) = \psi^{-1}(L)$ with $\eta(L) = \sum_{i=0}^{\infty} \eta_i L^i$, $\eta_0 = 1$ and the absolute summability.

We can construct $\varepsilon_t, \varepsilon_{t-1}, \ldots$ based on $y_t, y_{t-1}, \ldots$.

Examples:

- AR(1) with $\eta(L) = 1 - \phi L$, then $(1 - \phi L)(y_t - \mu) = \varepsilon_t$, or

$$\varepsilon_t = (y_t - \mu) - \phi(y_{t-1} - \mu)$$

- MA(1) with $\eta(L) = (1 + \theta L)^{-1}$, then $(1 + \theta L)^{-1}(y_t - \mu) = \varepsilon_t$, or

$$\varepsilon_t = (y_t - \mu) - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) - \theta^3(y_{t-3} - \mu) + \ldots$$

Forecasts Based on $y_t, y_{t-1}, \ldots$

The $\varepsilon_t, \varepsilon_{t-1}, \ldots$ constructed from $y_t, y_{t-1}, \ldots$ can be plugged into the compact form (42)

$$\hat{E}(Y_{t+s}|y_t, y_{t-1}, \ldots) = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \eta(L)(y_t - \mu). \tag{43}$$

This is called the Wiener-Kolmogorov prediction formula.

Consider once again the AR(1) process with $\eta(L) = 1 - \phi L$ and $|\phi| < 1$. We have

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

Therefore, by the Wiener-Kolmogorov prediction formula, the optimal linear s-step ahead forecast is

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

Forecasts Based on a Finite Number of Observations

Consider forecasting a stationary AR(p) process with known parameters μ and $\phi_1, \phi_2, ..., \phi_p$. From Lecture 2 we know that

$$\begin{aligned} Y_{t+s} - \mu &= f_{11}^{(s)}(Y_t - \mu) + f_{12}^{(s)}(Y_{t-1} - \mu) + ... + f_{1p}^{(s)}(Y_{t-p+1} - \mu) \\ &\quad + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s} + \psi_2 \varepsilon_{t+s} + ... + \psi_{s-1} \varepsilon_{t+s} \end{aligned}$$

where $\psi_i = f_{11}^{(i)}$.

The optimal s-step ahead forecast is

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

Forecasts Based on a Finite Number of Observations

Remarks:

- For forecasting the AR(p) process, we only need its p most recent observations, $y_t, y_{t-1}, \ldots y_{t-p+1}$.
- However, for MA or ARMA, we generally need infinite observations, $y_t, y_{t-1}, \ldots$.

Approximations to Optimal Forecasts

One approach to forecasting based on a finite number of values $y_t, y_{t-1}, ...y_{t-m+1}$ is to replace all presample ε's with zero.

Precisely speaking, the idea is to replace $\hat{E}(Y_{t+s}|y_t, y_{t-1}, ...)$ by

$$\hat{E}(y_t, y_{t-1}, ...y_{t-m+1}, \varepsilon_{t-m} = 0, \varepsilon_{t-m-1} = 0, ...). \quad (44)$$

Exact Finite Sample Forecasts

An alternative approach is to calculate the linear projection of $Y_{t+s} - \mu$ on its m most recent values. To this end, let

$$\mathbf{x}_t = ((y_t - \mu), (y_{t-1} - \mu), ..., (y_{t-m+1} - \mu))'$$

Then we look for a linear forecast of the form

$$\begin{aligned} Y^*_{t+s|t} - \mu &= \boldsymbol{\alpha}' \mathbf{x}_t \\ &= \alpha_1 (y_t - \mu) + \alpha_2 (Y_{t-1} - \mu) + ... + \alpha_m (y_{t-m+1} - \mu) \end{aligned}$$

Exact Finite Sample Forecasts

Under the assumption of stationarity, the coefficients α_i can be calculated directly from (40)

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_s \\ \gamma_{s+1} \\ \vdots \\ \gamma_{s+m-1} \end{pmatrix}$$

To be continued! Thank you!