AR, MA and ARMA models

Linear Time Series Analysis and Its Applications[1]

For basic concepts of linear time series analysis see

- Box, Jenkins, and Reinsel (1994, Chapters 2-3), and
- Brockwell and Davis (1996, Chapters 1-3)

The theories of linear time series discussed include

- stationarity
- dynamic dependence
- autocorrelation function
- modeling
- forecasting

[1] Tsay (2010), Chapter 2.

The econometric models introduced include

(a) simple autoregressive models,
(b) simple moving-average models,
(b) mixed autoregressive moving-average models,
(c) seasonal models,
(d) unit-root nonstationarity,
(e) regression models with time series errors, and
(f) fractionally differenced models for long-range dependence.

Strict stationarity

The foundation of time series analysis is stationarity.

A time series $\{r_t\}$ is said to be *strictly stationary* if the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \ldots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers.

The joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is invariant under time shift.

This is a very strong condition that is hard to verify empirically.

Weak stationarity

A time series $\{r_t\}$ is *weakly stationary* if both the mean of r_t and the covariance between r_t and r_{t-l} are time invariant, where l is an arbitrary integer.

More specifically, $\{r_t\}$ is weakly stationary if
(a) $E(r_t) = \mu$, which is a constant, and
(b) $Cov(r_t, r_{t-l}) = \gamma_l$, which only depends on l.

In practice, suppose that we have observed T data points $\{r_t | t = 1, \ldots, T\}$. The weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a fixed level.

In applications, weak stationarity enables one to make inference concerning future observations (e.g., prediction).

Properties

The covariance

$$\gamma = Cov(r_t, r_{t-l})$$

is called the lag-l autocovariance of r_t.

It has two important properties:
(a) $\gamma_0 = Var(r_t)$, and (b) $\gamma_{-l} = \gamma_l$.

The second property holds because

$$\begin{aligned} Cov(r_t, r_{t-(-l)}) &= Cov(r_{t-(-l)}, r_t) \\ &= Cov(r_{t+l}, r_t) \\ &= Cov(r_{t_1}, r_{t_1} - l), \end{aligned}$$

where $t_1 = t + l$.

AUTOCORRELATION FUNCTION

The autocorrelation function of lag l is

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}} = \frac{Cov(r_t, r_{t-l})}{Var(r_t)} = \frac{\gamma_l}{\gamma_0}$$

where the property $Var(r_t) = Var(r_{t-l})$ for a weakly stationary series is used.

In general, the lag-l sample autocorrelation of r_t is defined as

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2} \qquad 0 \leq l < T - 1.$$

Portmanteau test

Box and Pierce (1970) propose the Portmanteau statistic

$$Q^*(m) = T \sum_{l=1}^{m} \hat{\rho}_l^2$$

as a test statistic for the null hypothesis

$$H_0 : \rho_1 = \cdots = \rho_m = 0$$

against the alternative hypothesis

$$H_a : \rho_i \neq 0 \text{ for some } i \in \{1, \ldots, m\}.$$

Under the assumption that $\{r_t\}$ is an iid sequence with certain moment conditions, $Q^*(m)$ is asymptotically χ^2_m.

Ljung and Box (1978)

Ljung and Box (1978) modify the $Q^*(m)$ statistic as below to increase the power of the test in finite samples,

$$Q(m) = T(T+2)\sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l}.$$

The decision rule is to reject H_0 if $Q(m) > q_\alpha^2$, where q_α^2 denotes the $100(1-\alpha)th$ percentile of a χ_m^2 distribution.

```
# Load data
da = read.table("http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/m-ibm3dx2608.txt",
header=TRUE)

# IBM simple returns and squared returns
sibm=da[,2]
sibm2 = sibm^2

# ACF
par(mfrow=c(1,2))
acf(sibm)
acf(sibm2)

# Ljung-Box statistic Q(30)
Box.test(sibm,lag=30,type="Ljung")
Box.test(sibm2,lag=30,type="Ljung")

> Box.test(sibm,lag=30,type="Ljung")

Box-Ljung test

data:  sibm
X-squared = 38.241, df = 30, p-value = 0.1437

> Box.test(sibm2,lag=30,type="Ljung")

Box-Ljung test

data:  sibm2
X-squared = 182.12, df = 30, p-value < 2.2e-16
```

White noise

A time series r_t is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance.

All the ACFs are zero.

If r_t is normally distributed with mean zero and variance σ^2, the series is called a *Gaussian white noise*.

Linear Time Series

A time series r_t is said to be linear if it can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

where μ is the mean of r_t, $\psi_0 = 1$, and $\{a_t\}$ is white noise.

a_t denotes the new information at time t of the time series and is often referred to as the *innovation* or *shock* at time t.

If r_t is weakly stationary, we can obtain its mean and variance easily by using the independence of $\{a_t\}$ as

$$E(r_t) = \mu, \qquad V(r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2,$$

where σ_a^2 is the variance of a_t.

The lag-l aucovariance of r_t is

$$
\begin{aligned}
\gamma_l &= Cov(r_t, r_{t-l}) \\
&= E\left[\left(\sum_{i=0}^{\infty}\psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty}\psi_j a_{t-l-j}\right)\right] \\
&= E\left(\sum_{i,j=0}^{\infty}\psi_i\psi_j a_{t-i}a_{t-l-j}\right) \\
&= \sum_{i=0}^{\infty}\psi_{j+l}\psi_j E(a_{t-l-j}^2) = \sigma_a^2\sum_{j=0}^{\infty}\psi_j\psi_{j+l},
\end{aligned}
$$

so

$$\rho_l = \frac{\gamma_l}{\gamma_0} = \frac{\sum_{i=0}^{\infty}\psi_i\psi_{i+l}}{1+\sum_{i=1}^{\infty}\psi_i^2}$$

AR(1)

Linear time series models are econometric and statistical models used to describe the pattern of the ψ weights of r_t. For instance, an stationary AR(1) model can be written as

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t$$

where $\{a_t\}$ is white noise. It is easy to see that

$$r_t - \mu = \sum_{i=0}^{\infty} \phi_1^i a_{t-i},$$

and

$$V(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2},$$

provided that $\phi_1^2 < 1$. In other words, the weak stationarity of an AR(1) model implies that $|\phi_1| < 1$.

Using $\phi_0 = (1 - \phi_1)\mu$, one can rewrite a stationary AR(1) model as

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t,$$

such that ϕ_1 measures the persistence of the dynamic dependence of an AR(1) time series.

The ACF of the AR(1) is

$$\gamma_l = \phi_1 \gamma_{l-1} \qquad l > 0,$$

where $\gamma_0 = \phi_1 \gamma_1 + \sigma_a^2$ and $\gamma_l = \gamma_{-l}$.

Also,

$$\rho_l = \phi_1^l,$$

i.e., the ACF of a weakly stationary AR(1) series decays exponentially with rate ϕ_1 and starting value $\rho_0 = 1$.

AR(2)

An AR(2) model assumes the form

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t,$$

where

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2},$$

provided that $\phi_1 + \phi_2 \neq 1$.

It is easy to see that

$$\gamma_l = \phi_1 \gamma_{l-1} + \phi_2 \gamma_{l-2}, \qquad \text{for } l > 0,$$

and that

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2}, \qquad l \geq 2,$$

with $\rho_1 = \phi_1/(1 - \phi_2)$.

US real GNP

As an illustration, consider the quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991.

Here we simply employ an AR(3) model for the data. Denoting the growth rate by r_t the fitted model is

$$r_t = 0.0047 + 0.348 r_{t-1} + 0.179 r_{t-2} - 0.142 r_{t-3} + a_t,$$

with $\hat{\sigma}_a = 0.0097$.

Alternatively,

$$r_t - 0.348r_{t-1} - 0.179r_{t-2} + 0.142r_{t-3} = 0.0047 + a_t,$$

with the corresponding third-order difference equation

$$(1 - 0.348B - 0.179B^2 + 0.142B^3) = 0$$

or

$$(1 + 0.521B)(1 - 0.869B + 0.274B^2) = 0$$

The first factor

$$(1 + 0.521B)$$

shows an exponentially decaying feature of the GNP.

Business cycles

The second factor $(1 - 0.869B + 0.274B^2)$ confirms the existence of stochastic business cycles. For an AR(2) model with a pair of complex characteristic roots, the average length of the stochastic cycles is

$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]}$$

or $k = 10.62$ quarters, which is about 3 years.

Fact: If one uses a nonlinear model to separate U.S. economy into "expansion" and "contraction" periods, the data show that the average duration of contraction periods is about 3 quarters and that of expansion periods is about 12 quarters.

The average duration of 10.62 quarters is a compromise between the two separate durations. The periodic feature obtained here is common among growth rates of national economies. For example, similar features can be found for many OECD countries.

R code

```
gnp=scan(file="http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/dgnp82.txt")

# To create a time-series object
gnp1=ts(gnp,frequency=4,start=c(1947,2))

par(mfrow=c(1,1))
plot(gnp1)
points(gnp1,pch="*")

# Find the AR order
m1=ar(gnp,method="mle")
m1$order
m2=arima(gnp,order=c(3,0,0))
m2

# In R, intercept denotes the mean of the series.
# Therefore, the constant term is obtained below:
(1-.348-.1793+.1423)*0.0077

# Residual standard error
sqrt(m2$sigma2)

# Characteristic equation and solutions
p1=c(1,-m2$coef[1:3])
roots = polyroot(p1)

# Compute the absolute values of the solutions
Mod(roots)
[1] 1.913308 1.920152 1.913308

# To compute average length of business cycles:
k=2*pi/acos(1.590253/1.913308)
```

AR(p)

The results of the AR(1) and AR(2) models can readily be generalized to the general AR(p) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t,$$

where p is a nonnegative integer and $\{a_t\}$ is white noise.

The mean of a stationary series is

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

provided that the denominator is not zero.

The associated characteristic equation of the model is

$$1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p = 0.$$

If all the solutions of this equation are greater than 1 in modulus, then the series r_t is stationary.

Partial ACF

The PACF of a stationary time series is a function of its ACF and is a useful tool for determining the order p of an AR model. A simple, yet effective way to introduce PACF is to consider the following AR models in consecutive orders:

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1,t}, \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2,t}, \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3,t}, \\
&\vdots
\end{aligned}
$$

The estimate $\hat{\phi}_{i,i}$ the ith equation is called the lag-i sample PACF of r_t.

For a stationary Gaussian $AR(p)$ model, it can be shown that the sample PACF has the following properties:

- $\hat{\phi}_{p,p} \to \phi_p$ as $T \to \infty$.
- $\hat{\phi}_{l,l} \to 0$ for all $l > p$.
- $V(\hat{\phi}_{l,l}) \to 1/T$ for $l > p$.

These results say that, for an $AR(p)$ series, the sample PACF cuts off at lag p.

AIC

The well-known Akaike information criterion (AIC) (Akaike, 1973) is defined as

$$AIC = \underbrace{-\frac{2}{T}\log(\text{likelihood})}_{\text{goodness of fit}} + \underbrace{\frac{2}{T}(\text{number of parameters})}_{\text{penalty function}},$$

where the likelihood function is evaluated at the maximum-likelihood estimates and T is the sample size.

For a Gaussian $AR(l)$ model, AIC reduces to

$$AIC(l) = \log(\tilde{\sigma}_l^2) + \frac{2l}{T}$$

where $\tilde{\sigma}_l^2$ is the maximum-likelihood estimate of σ_a^2, which is the variance of a_t and T is the sample size.

BIC

Another commonly used criterion function is the SchwarzBayesian information criterion (BIC).

For a Gaussian $AR(l)$ model, the criterion is

$$BIC(l) = \log(\tilde{\sigma}_l^2) + \frac{l \log(T)}{T}$$

The penalty for each parameter used is 2 for AIC and $\log(T)$ for BIC.

Thus, BIC tends to select a lower AR model when the sample size is moderate or large.

Forecasting

For the $AR(p)$ model, suppose that we are at the time index h and are interested in forecasting r_{h+l} where $l \geq 1$.

The time index h is called the forecast origin and the positive integer l is the forecast horizon.

Let $\hat{r}_h(l)$ be the forecast of r_{h+l} using the minimum squared error loss function, i.e.

$$E\{[r_{h+l} - \hat{r}_h(l)]^2 | F_h\} \leq \min_g E[(r_{h+l} - g)^2 | F_h],$$

where g is a function of the information available at time h (inclusive), that is, a function of F_h.

We referred to $\hat{r}_h(l)$ as the l-step ahead forecast of r_t at the forecast origin h.

1-step-ahead forecast

It is easy to see that

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i},$$

and the associated forecast error is

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1},$$

and

$$V(e_h(1)) = V(a_{h+1}) = \sigma_a^2.$$

2-step-ahead forecast

Similarly,

$$\hat{r}_h(2) = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \cdots + \phi_p r_{h+2-p},$$

with

$$e_h(2) = a_{h+2} + \phi_1 a_{h+1}$$

and

$$V(e_h(2)) = (1 + \phi_1^2)\sigma_a^2.$$

Multistep-ahead forecast

In general,

$$r_{h+l} = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+l-i} + a_{h+l},$$

and

$$\hat{r}_h(l) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(l-i),$$

where $\hat{r}_h(i) = r_{h+i}$ if $i \leq 0$.

Mean reversion

It can be shown that for a stationary $AR(p)$ model,

$$\hat{r}_h(l) \to E(r_t) \qquad mboxas\ l \to \infty,$$

meaning that for such a series long-term point forecast approaches its unconditional mean.

This property is referred to as the mean reversion in the finance literature.

For an AR(1) model, the speed of mean reversion is measured by the half-life defined as

$$\text{half-life} = \frac{\log(0.5)}{\log(|\phi_1|)}.$$

MA(1) model

There are several ways to introduce MA models.

One approach is to treat the model as a simple extension of white noise series.

Another approach is to treat the model as an infinite-order AR model with some parameter constraints.

We adopt the second approach.

We may entertain, at least in theory, an AR model with infinite order as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + a_t.$$

However, such an AR model is not realistic because it has infinite many parameters.

One way to make the model practical is to assume that the coefficients ϕ_i's satisfy some constraints so that they are determined by a finite number of parameters.

A special case of this idea is

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t.$$

where the coefficients depend on a single parameter θ_1 via $\phi_i = -\theta_1^i$ for $i \geq 1$.

Obviously,

$$\begin{aligned} r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \theta_1^3 r_{t-3} + \cdots &= \phi_0 + a_t \\ \theta_1(r_{t-1} + \theta_1 r_{t-2} + \theta_1^2 r_{t-3} + \theta_1^3 r_{t-4} + \cdots) &= \theta_1(\phi_0 + a_{t-1}) \end{aligned}$$

so

$$r_t = \phi_0(1-\theta_1) + a_t - \theta_1 a_{t-1},$$

i.e., r_t is a weighted average of shocks a_t and a_{t-1}.

Therefore, the model is called an MA model of order 1 or $MA(1)$ model for short.

MA(q)

The general form of an MA(q) model is

$$r_t = c_0 - \sum_{i=1}^{q} \theta_i a_{t-i},$$

or

$$r_t = c_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t,$$

where $q > 0$.

Moving-average models are **always weakly stationary** because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant.

$$\begin{aligned} E(r_t) &= c_0 \\ V(r_t) &= (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_a^2. \end{aligned}$$

ACF of an MA(1)

Assume that $c_0 = 0$ for simplicity. Then,

$$r_{t-l} r_t = r_{t-l} a_t - \theta_1 r_{t-l} a_{t-1}.$$

Taking expectation, we obtain

$$\gamma_1 = -\theta_1 \sigma_a^2 \quad \text{and} \quad \gamma_l = 0, \quad \text{for} \;\; l > 1.$$

Since $V(r_t) = (1 + \theta_1^2)\sigma_a^2$, it follows that

$$\rho_0 = 1, \quad \rho_1 = -\frac{\theta_1}{1 + \theta_1^2}, \quad \rho_l = 0, \quad \text{for} \;\; l > 1.$$

ACF of an MA(2)

For the MA(2) model, the autocorrelation coefficients are

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_l = 0, \quad \text{for} \;\; l > 2.$$

Here the ACF cuts off at lag 2.

This property generalizes to other MA models.

For an MA(q) model, the lag-q ACF is not zero, but $\rho_l = 0$ for $l > q$.

an MA(q) series is only linearly related to its first q-lagged values and hence is a “finite-memory” model.

Estimation

Maximum-likelihood estimation is commonly used to estimate MA models. There are two approaches for evaluating the likelihood function of an MA model.

The first approach assumes that $a_t = 0$ for $t \leq 0$, so $a_1 = r_1 - c_0$, $a_2 = r_2 - c_0 + \theta_1 a_1$, etc. This approach is referred to as the *conditional-likelihood method*.

The second approach treats $a_t = 0$ for $t \leq 0$, as additional parameters of the model and estimate them jointly with other parameters. This approach is referred to as the *exact-likelihood method*.

Forecasting an MA(1)

For the 1-step-ahead forecast of an MA(1) process, the model says

$$r_{h+1} = c_0 + a_{h+1} - \theta_1 a_h.$$

Taking the conditional expectation, we have

$$\begin{aligned}
\hat{r}_h(1) &= E(r_{h+1}|F_h) = c_0 - \theta_1 a_h, \\
e_h(1) &= r_{h+1} - \hat{r}_h(1) = a_{h+1}
\end{aligned}$$

with $V[e_h(1)] = \sigma_a^2$.

Similarly,

$$\begin{aligned}
\hat{r}_h(2) &= E(r_{h+1}|F_h) = c_0 \\
e_h(2) &= r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}
\end{aligned}$$

with $V[e_h(2)] = (1 + \theta_1^2)\sigma_a^2$.

Forecasting an MA(2)

Similarly, for an MA(2) model, we have

$$r_{h+l} = c_0 + a_{h+l} - \theta_1 a_{h+l-1} - \theta_2 a_{h+l-2},$$

from which we obtain

$$\begin{aligned}
\hat{r}_h(1) &= c_0 - \theta_1 a_h - \theta_2 a_{h-1}, \\
\hat{r}_h(2) &= c_0 - \theta_2 a_h, \\
\hat{r}_h(l) &= c_0, \quad \text{for } l > 2.
\end{aligned}$$

Summary

A brief summary of AR and MA models is in order. We have discussed the following properties:

- For MA models, ACF is useful in specifying the order because ACF cuts off at lag q for an MA(q) series.
- For AR models, PACF is useful in order determination because PACF cuts off at lag p for an AR(p) process.
- An MA series is always stationary, but for an AR series to be stationary, all of its characteristic roots must be less than 1 in modulus.

Carefully read Section 2.6 of Tsay (2010) about ARMA models.