ARMA processes

AR $\equiv$ auto-regressive MA $\equiv$ moving average

ARMA processes are simple models for time series.

An ARMA process is constructed from a sequence of random shocks (also called white noise) which are independent $N(0, \sigma_a^2)$ random variables denoted:

$$\ldots, a_{-3}, a_{-2}, a_{-1}, a_0, a_1, a_2, a_3, \ldots$$

Other letters are used besides 'a' such as 'ε' or 'N'.

We have written the sequence as starting in the remote past (time $-\infty$) and continuing into the remote future (time $+\infty$), but there can also be finite beginning and ending times (e.g., 0 to 100).

Autoregressive Processes

$$\begin{aligned}
\text{AR(1)} : z_t &= C + \phi_1 z_{t-1} + a_t \\
\text{AR(2)} : z_t &= C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \\
\text{AR(p)} : z_t &= C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t
\end{aligned}$$

The values of the parameters $C, \phi_1, \ldots, \phi_p$, σ_a^2 determine the behavior of the process.

To generate a realization $z_1, z_2, \ldots, z_n$ from an AR(1) process: Choose a starting value z_1. Then

$$\begin{aligned}
z_2 &= C + \phi_1 z_1 + a_2 \\
z_3 &= C + \phi_1 z_2 + a_3 \\
&\vdots \\
z_n &= C + \phi_1 z_{n-1} + a_n
\end{aligned}$$

with the computer supplying the random shocks $a_2, a_3, \ldots$

To generate a realization from an AR(2) process, we need two starting values z_1, z_2. Then

$$\begin{aligned} z_3 &= C + \phi_1 z_2 + \phi_2 z_1 + a_3 \\ z_4 &= C + \phi_1 z_3 + \phi_2 z_2 + a_4 \\ &\vdots \\ z_n &= C + \phi_1 z_{n-1} + \phi_2 z_{n-2} + a_n \end{aligned}$$

To generate an AR(p), we need p starting values.

What do these processes look like? (Display some examples.)

Moving Average Processes

$$\begin{aligned}
\text{MA(1)} : z_t &= C + a_t - \theta_1 a_{t-1} \\
\text{MA(2)} : z_t &= C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\
\text{MA(q)} : z_t &= C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}
\end{aligned}$$

Note the sign convention: the MA(2) process

$$\begin{aligned}
z_t &= 10 + a_t + .7a_{t-1} + .4a_{t-2} \qquad \text{is written as} \\
z_t &= 10 + a_t - (-.7)a_{t-1} - (-.4)a_{t-2}
\end{aligned}$$

and so has $\theta_1 = -.7$ and $\theta_2 = -.4$.

To generate $z_1, z_2, \ldots, z_n$:

- for an MA(1) process, you use $a_0, a_1, \ldots, a_n$.
- for an MA(2) process, you use $a_{-1}, a_0, a_1, \ldots, a_n$.
- for an MA(q) process, you use $a_{1-q}, a_{2-q}, \ldots, a_n$.

What do MA processes look like? (Display some examples.)

ARMA Processes

$$\begin{aligned}
\text{ARMA(1,1)} : z_t =& C + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} \\
\text{ARMA(p,q)} : z_t =& C + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} \\
& + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}
\end{aligned}$$

$\text{ARMA(p,0)} \equiv \text{AR(p)}$

$\text{ARMA(0,q)} \equiv \text{MA(q)}$

Remark: For any ARMA process, we can generate a huge (essentially infinite) number of realizations, and think of them as forming a population of time series. For any time t, we can define quantities like Ez_t, Ez_t^2, $Ez_t z_{t-1}$ as averages over this population. Similarly, we can define $\text{Var}(z_t)$, $\text{Cov}(z_t, z_{t-1})$, $\text{Corr}(z_t, z_{t-1})$, etc.

Behavior of AR Processes (Stationarity)

AR(1): If $|\phi_1| < 1$, then, for any choice of initial value z_1, the AR(1) process converges to a stationary process after an intial transient phase.

Stationary (very roughly) means stable over time, not changing in behavior with time.

We are not usually interested in the transient initial behavior of the AR(1) process, but only in its long-run stable behavior. Whenever

we use the AR(1) process in applications, we assume it started far enough in the past so that we are observing its stationary (stable) behavior. (The same remark applies to general AR(p) and ARMA(p,q) processes.)

AR(2): If all of the following hold:

$$|\phi_2| < 1\,, \quad \phi_2 + \phi_1 < 1\,, \quad \phi_2 - \phi_1 < 1$$

then, for any choice of initial values z_1, z_2, the AR(2) process converges to a stationary process. If any of these conditions fails, the process is non-stationary.

AR(p): Whether or not an AR(p) process is stationary depends only on the values of $\phi_1, \ldots, \phi_p$. There are simple easy-to-check conditions (stated earlier) for $p = 1$ and $p = 2$. There is a general condition valid for all p (stated later) which is more difficult to check.

Behavior of MA(q) Processes

MA(q) processes are always stationary (regardless of the values of their parameters $\theta_1, \ldots, \theta_q$, C, σ_a^2). There is no transient initial phase; they reach their stationary behavior immediately.

Behavior of ARMA(p,q) Processes

An ARMA(p,q) process is stationary if and only if $\phi_1, \ldots, \phi_p$ satisfy the same conditions required for an AR(p) to be stationary.

Definition: A process $z_1, z_2, z_3, \ldots$ is called a (weakly or second order) **stationary** process if:

- $Ez_t = \mu_z$ for all t,
- $\text{Var}(z_t) = \sigma_z^2$ for all t,
- For any positive integer k,

$$\frac{\text{Cov}(z_t, z_{t-k})}{\sigma_z^2} = \text{Corr}(z_t, z_{t-k}) = \rho_k \quad \text{for all } t.$$

The process is called **strictly** stationary if, in addition:

z_t has the same distribution for all t,

(z_t, z_{t+1}) has the same joint distribution for all t

(z_t, z_{t+1}, z_{t+2}) has the same joint distn for all t, etc.

The value ρ_k is called the autocorrelation at lag k. The sequence $\rho_0, \rho_1, \rho_2, \rho_3, \ldots$ is called the autocorrelation function (ACF).

The Partial Autocorrelation Function (PACF)

Let z_t, $t = 1, 2, 3, \ldots$, be a stationary ARMA process.

The PACF is a sequence of values $\phi_{11}, \phi_{22}, \phi_{33}, \ldots$ defined (informally) as follows:

Consider a series of regressions:

$$
\begin{aligned}
z_t &= C_1 + \phi_{11} z_{t-1} + \varepsilon_{1,t} \\
z_t &= C_2 + \phi_{21} z_{t-1} + \phi_{22} z_{t-2} + \varepsilon_{2,t} \\
z_t &= C_3 + \phi_{31} z_{t-1} + \phi_{32} z_{t-2} + \phi_{33} z_{t-3} + \varepsilon_{3,t} \\
&\vdots \\
z_t &= C_k + \phi_{k1} z_{t-1} + \phi_{k2} z_{t-2} + \cdots + \phi_{kk} z_{t-k} + \varepsilon_{k,t} \\
&\vdots
\end{aligned}
$$

In the k-th regression, we are trying to predict z_t using the previous k values $z_{t-1}, \ldots, z_{t-k}$.

Suppose we have an infinite amount of data and we fit the regressions by least squares. Then we obtain the "true" values of the regression parameters, the values giving the best predictions in the population. (The "true" values are also referred to as the "exact" or "theoretical" or "population" values.)

The PACF is the sequence of estimated regression coefficients: $\phi_{11}, \phi_{22}, \phi_{33}, \ldots$ The value ϕ_{kk} is the true lag k coefficient in the regression of z_t on $z_{t-1}, \ldots, z_{t-k}$.

A formal definition: If you choose the values C, $\phi_{k1}, \ldots, \phi_{kk}$ to minimize

$$E(z_t - C - \phi_{k1} z_{t-1} - \phi_{k2} z_{t-2} - \cdots - \phi_{kk} z_{t-k})^2 ,$$

then ϕ_{kk} is the value of the PACF at lag k.

ϕ_{kk} can be expressed as a function of $\rho_1, \rho_2, \ldots, \rho_k$:

$$\phi_{kk} = g_k(\rho_1, \rho_2, \ldots, \rho_k)\,.$$

This function has a simple form in matrix notation.

Identifying ARMA Processes Using the ACF and PACF

For an AR(p) process:

- The theoretical ACF decays to zero, either exponentially or with a damped sine wave pattern or with both of these patterns.
- The theoretical PACF has a cutoff to zero after lag p; the last nonzero value is at lag p, and it is exactly zero for lags greater than p:

$$\phi_{pp} \neq 0 \quad \text{and} \quad \phi_{kk} = 0 \text{ for } k > p;$$

For an MA(q) process:

- The theoretical ACF has a cutoff to zero after lag q; the last nonzero value is at lag q, and it is exactly zero for lags greater than q:

$$\rho_q \neq 0 \quad \text{and} \quad \rho_k = 0 \text{ for } k > q\,.$$

- The theoretical PACF decays to zero.

For an ARMA(p, q) process with both $p > 0$ and $q > 0$, the theoretical ACF and PACF both decay to zero; neither of them has a cutoff.

For a **random shock process**, the ACF and PACF are exactly zero for all nonzero lags:

$$\rho_k = 0 \quad \text{and} \quad \phi_{kk} = 0 \quad \text{for all } k > 0.$$

Estimating the ACF

Suppose we have data $z_1, z_2, \ldots, z_n$ which is a realization of a stationary process.

The theoretical ACF of this process is estimated by the sample ACF (SACF):

$$\rho_k \quad \text{is estimated by} \quad r_k = \frac{\sum_{t=1}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2}$$

which is approximately the sample correlation in the scatterplot of z_t versus z_{t-k}.

The sequence $r_1, r_2, r_3, \ldots$ is the SACF.

$Er_k \approx \rho_k$ unless n is fairly small. (r_k is approximately unbiased.)

$r_k \to \rho_k$ as $n \to \infty$.

Estimating the PACF

The theoretical ACF and PACF are related by:

$$\phi_{kk} = g_k(\rho_1, \rho_2, \ldots, \rho_k)\,.$$

Suppose we have data $z_1, z_2, \ldots, z_n$.

The sample PACF is obtained from the sample ACF in the same way:

$$\hat{\phi}_{kk} = g_k(r_1, r_2, \ldots, r_k)\,.$$

The sample PACF value $\hat{\phi}_{kk}$ may also be approximated by the estimated coefficient for z_{t-k} obtained from fitting a regression of z_t on $z_{t-1}, \ldots, z_{t-k}$ (using least squares). This approximation will be good when n is large.