smfl9.tex

Lecture 9. 4.2.2011

5. CONDITION FOR STATIONARITY

We return to the general case. Just as in the $AR(2)$ example above, if the $AR(p)$ process has a moving-average representation

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

then if $\sigma^2 = var \epsilon_t$,

$$var X_t = \sigma^2 \sum_{i=0}^{\infty} \phi_i^2.$$

The condition

$$\sum_{i=0}^{\infty} \phi_i^2 < \infty$$

((ϕ_i) is *square-summable*, or is *in* L_2) is necessary and sufficient for
(i) $var X_t < \infty$;
(ii) the series $\sum \phi_i \epsilon_{t-i}$ in the moving-average representation to be convergent in mean square – or, in L_2. So if we interpret convergence in the mean-square sense, $\sum \phi_i^2 < \infty$ is the necessary and sufficient condition (NASC) for the moving-average representation of X_t to *exist*. Since $\sum \phi_i \epsilon_{t-i}$ is (when convergent) stationary (because (ϵ_t) is stationary):
if $\sum \phi_i^2 < \infty$, then (X_t) is stationary. The converse is also true, giving:

THEOREM (Condition for Stationarity). The following are equivalent:
(i) The parameters $\phi_1, \cdots, \phi_p$ in the $AR(p)$ model

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t, \qquad (\epsilon_t) \quad WN(\sigma^2) \qquad (*)$$

define a stationary process (X_t);
(ii) The roots of the polynomial

$$\phi(\lambda) := \phi_p \lambda^p + \cdots + \phi_1 \lambda - 1 = 0$$

lie outside the unit disc in the complex λ-plane;
(iii) X_t has the moving-average representation

$$X_t = \sum_{i=0}^{\infty} \phi_i \epsilon_{t-i}$$

with

$$\sum_{i=0}^{\infty} \phi_i^2 < \infty.$$

Proof. Substituting the moving-average representation into $(*)$,

$$\begin{aligned}
\sum_{i=0}^{\infty}\phi_i\epsilon_{t-i} &= \sum_{k=1}^{p}\phi_k\sum_{i=0}^{\infty}\psi_i\epsilon_{t-k-i}+\epsilon_t \\
&= \sum_{k=1}^{p}\phi_k\sum_{i=k}^{\infty}\phi_{i-k}\epsilon_{t-i}+\epsilon_t \\
&= \sum_{i=1}^{\infty}\Big(\sum_{k=1}^{\min(i,p)}\phi_k\phi_{i-k}\Big)\epsilon_{t-i}+\epsilon_t.
\end{aligned}$$

Equating coefficients of ϵ_{t-i}, we obtain the difference equation

$$\phi_i = \sum_{k=1}^{p}\phi_k\phi_{i-k} \qquad (i \geq p)$$

(with similar equations for $i = 0, 1, \cdots, p-1$, which provide starting-values for the difference equation above). The difference equation, of order p, has general solution

$$\phi_i = \sum_{k=1}^{p} c_k\lambda_k^i,$$

where $\lambda_1, \cdots, \lambda_p$ are the roots of the characteristic polynomial

$$\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_{p-1}\lambda - \phi_p = 0$$

(with appropriate modifications in the case of repeated roots, as before). [Check: if $\phi_i = \lambda^i$ is a trial solution of the difference equation, $\lambda^i = \sum_1^p \phi_k\lambda^{i-k}$. Multiply through by λ^{p-i}: $\lambda^p = \sum_1^p \phi_k\lambda^{p-k}$.] Now as $\phi_i = \sum_1^p c_k\lambda_k^i$ and $|\lambda_k^i| \to \infty, = 1$ or $\to 0$ as $i \to \infty$ according as $|\lambda_k| > 1, = 1$ or < 1, $\sum \phi_i^2 < \infty$ iff each $|\lambda_k| < 1$, i.e. each root of $\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p = 0$ is inside the unit disk, i.e. each root of

$$\phi(\lambda) = \phi_p\lambda^p + \phi_{p-1}\lambda^{p-1} + \cdots + \phi_1\lambda - 1 = 0$$

is outside the unit disk. This is all that remained to be proved. //

In the stationary case, we thus have

$$\gamma_t = cov(X_t, X_{t+\tau}) = \sigma^2\sum_{i=0}^{\infty}\phi_i\phi_{i+\tau},$$

with $\sum \phi_i^2 < \infty$ and $\phi_i = \sum_{k=1}^p c_k\lambda_k^i$, $|\lambda_i| < 1$. If λ_1 (say) is the root of largest modulus, $\phi_i \sim c_1\lambda_1^i$ for large i, and $\phi_i\phi_{i+\tau} \sim c_1^2\lambda_1^{\tau+2i}$. So for large τ, we can expect

$$\gamma_\tau \sim \sigma^2\sum c_1^2\lambda_1^{\tau+2i} \sim const.\lambda_1^\tau, \qquad \rho_\tau \sim \gamma_\tau/\gamma_0 \sim \lambda_1^\tau.$$

Thus for a stationary $AR(p)$ model, we expect that the autocorrelation decreases geometrically to zero for large lag τ (the decay rate being the characteristic root of largest modulus).
Note. For $AR(1)$, the autocorrelation function *is* geometrically decreasing: $\rho_\tau = \rho^\tau$. This holds exactly, even for small τ. Since the sample autocorrelation (correlogram) r_τ approximates the population autocorrelation $\rho_\tau = \rho^\tau$: for $AR(1)$,

$$r_\tau \sim \rho^\tau :$$

the sample ACF is *approximately* geometrically decreasing (i.e., geometrically decreasing plus sampling error), even for small lags τ. We can look for this pattern at the beginning of a plot of the ACF, and this is the signature of an $AR(1)$ process. For $AR(p)$, $p > 1$, matters are not so simple. The approximation above only holds for large τ, by which time r_τ will be small (it approximates ρ_τ, which tends to zero as τ increases), and the pattern of geometric decrease will tend to be swamped by sampling error. Consequently, it is *much harder* to interpret the correlogram of an $AR(p)$ for $p > 1$ than for an $AR(1)$.

By contrast, the moving average – $MA(q)$ – models considered below have autocorrelations that *cut off* - they are zero beyond lag q, apart from sampling error. This is the signature of the ACF of an $MA(q)$, and is easy to interpret; an $AR(1)$ signature is easy to interpret; that of an $AR(p)$ for $p > 1$ is (usually) not.

6. MOVING AVERAGE PROCESSES, MA(q).

Suppose we have a system in which new information arrives at regular intervals, and new information affects the system's response for a limited period. The new information might be economic, financial etc., and the system might involve the price of some commodity, for example.

The simplest possible model for the new information process, or *innovation process*, is white noise, $WN(\sigma^2)$, so we assume this. The simplest possible model for a response with such a limited time-influence is

$$X_t = \epsilon_t + \sum\nolimits_{j=1}^{q} \theta_j \epsilon_{t-j}, \qquad (\epsilon_t) \quad WN(\sigma^2).$$

This is called a *moving average process* or *order* q, $MA(q)$.

In terms of the lag operator B, $\epsilon_{t-j} = B^j \epsilon_t$, so if

$$\theta(B) := 1 + \sum\nolimits_{j=1}^{q} \theta_j B^j,$$

we can write

$$X_t = \theta(B)\epsilon_t.$$

Autocovariance. Since $E\epsilon_t = 0$, $EX_t = 0$ also. So writing $\theta_0 = 1$,

$$\begin{aligned}\gamma_k = cov(X_t, X_{t+k}) = E[X_t X_{t+k}] &= E[\textstyle\sum_{i=0}^q \theta_i \epsilon_{t-i} \sum_{j=0}^q \theta_j \epsilon_{t-k-j}] \\ &= \displaystyle\sum_{i,j=0}^q \theta_i \theta_j E[\epsilon_{t-i}\epsilon_{t-k-j}].\end{aligned}$$

Now $E[.] = 0$ unless $i = j + k$, when it is σ^2. It suffices to take $k \geq 0$ (as $\gamma(-k) = \gamma(k)$). If also $k \leq q$, we can take $j = i - k$, and then the limits on j are $0 \leq j \leq q - k$, as $0 \leq i \leq q$. If however $k > q$, there are no non-zero terms as there are no $i = k + j$ with $0 \leq i, j \leq q$. So

$$\gamma(k) = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}, & \text{if } k = 0, 1, \cdots, q, \\ 0 & \text{if } k > q, \end{cases}$$

$$\gamma_0 = \sigma^2 \textstyle\sum_{j=0}^q \theta_j^2,$$

so the autocorrelation is

$$\rho_k = \begin{cases} \sum_{i=0}^{q-k} \theta_i \theta_{i+k} / \sum_{i=0}^q \theta_i^2 & \text{if } k = 0, 1, \cdots, q, \\ 0 & \text{if } k > q. \end{cases}$$

This sudden cut-off of the autocorrelation after lag $k = q$ is the signature of an $MA(q)$ process.

First-order case: $MA(1)$.

The model equation is

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

By above,

$$\rho_0 = 1, \qquad \rho_1 = \theta/(1 + \theta^2), \qquad \rho_k = 0 \quad (k \geq 2).$$

In terms of the lag (backward shift) operator B:

$$X_t = (1 + \theta B)\epsilon_t.$$

Hence formally

$$\epsilon_t = (1 + \theta B)^{-1} X_t = \textstyle\sum_0^\infty (-\theta)^k B^k X_t = X_t + \sum_1^\infty (-\theta)^k X_{t-k} :$$

$$X_t = \epsilon_t - \sum_1^\infty (-\theta)^k X_{t-k}.$$

This is an *infinite-order autoregressive* representation of (X_t). For (mean-square) convergence on RHS, as in the AR theory above, we need

$$|\theta| < 1.$$

The $MA(1)$ model is then said to be *invertible*: the passage from the $MA(1)$ representation using $(1+\theta B)$ to the $AR(\infty)$ representation using $(1+\theta B)^{-1}$ is called *inversion*.
Note. If we replace θ by $1/\theta$, ρ_1 goes from $\theta/(1+\theta^2)$ to

$$(1/\theta)/[1+(1/\theta)^2] = \theta/(1+\theta^2)$$

– the same as before. So for $\theta \neq 1$, two *different* $MA(1)$ processes have the *same* ACF:we cannot hope to identify the process from the ACF, or its sample version, the correlogram. However, for $|\theta| \neq 1$, exactly *one* of these processes is invertible. So if we restrict attention to invertible MA processes, *identifiability* is restored in general ($|\theta| \neq 1$), but not in the exceptional case $|\theta| = 1$, $\theta \neq 1$.
General case: $MA(q)$. As above,

$$X_t = \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} = \theta(B)\epsilon_t, \quad \text{where} \quad \theta(\lambda) = 1 + \sum_{j=1}^q \theta_j \lambda^j.$$

So formally, if we can invert this to obtain

$$\epsilon_t = \theta(B)^{-1} X_t,$$

and as $\theta(\lambda) = 1 + \theta_1\lambda + \cdots$, $1/\theta(\lambda) = 1 + c.\lambda + \cdots$. So

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_i X_{t-i} + \cdots + \epsilon_t,$$

for some constants ϕ_i. This expresses the new value X_t at the current time t as a sum of two components:
(i) an (infinite) linear combination of previous values X_{t-i}, and
(ii) the new white-noise term ϵ_t, thought of as the *innovation* at time t. It is thus plausible that it should be possible to forecast future values of such a process given knowledge of its history.

Proceeding as in the proof of the Condition for Stationarity in Section 4, we find that ϕ_i is of the form

$$\phi_i = \sum_{k=1}^q c_k \lambda_k^i,$$

where the λ_k are the roots of the polynomial

$$\lambda^p + \theta_1 \lambda^{p-1} + \cdots + \theta_p = 0.$$

For $\phi_i \to 0$ as $i \to \infty$ – that is, for the influence of the remote past of the process to damp out to zero – we need all $|\lambda_i| < 1$. That is, all roots of the above polynomial (which is $\theta(1/\lambda)$) should lie *inside* the unit disc in the complex λ-plane. Equivalently, all roots of $\theta(\lambda) = 0$ lie *outside* the unit disc. Then as before, $\sum \phi_i^2 < \infty$ and the series $\sum \phi_i X_{t-i}$ converges in mean square. To summarise, we have:

THEOREM (Condition for Invertibility). For the $MA(q)$ model

$$X_t = \theta(B)\epsilon_t, \qquad (\epsilon_t) \quad WN$$

to be invertible as

$$\epsilon_t = \theta(B)^{-1} X_t,$$

it is necessary and sufficient that all roots λ_i of the polynomial equation

$$\lambda^p + \theta_1 \lambda^{p-1} + \cdots + \theta_p = 0$$

should lie outside the unit disc. Then

$$\epsilon_t = \sum\nolimits_1^\infty \phi_i X_{t-i}$$

with $\sum \phi_i^2 < \infty$ and the series convergent in mean square.

Note. 1. The Condition for Stationarity for $AR(p)$ processes and the Condition for Invertibility for $MA(q)$ processes exhibit a *duality*, in which the roles of X_t and ϵ_t are interchanged.
2. We shall confine ourselves in what follows to the invertible case. Then the parameters θ_j are uniquely determined by the autocorrelation function ρ_τ.
3. In the $MA(1)$ case, the above characteristic equation is

$$\lambda + \theta_1 = 0,$$

with root $\lambda = -\theta_1$. For invertibility, we need $|\theta_1| < 1$, as before. Invertibility avoids the ambiguity of both θ_1 and $1/\theta_1$ giving the same ACF

$$\rho_0 = 1, \qquad \rho_1 = \theta_1/(1+\theta_1^2), \qquad \rho_k = 0 \qquad (k \geq 2).$$