smfd11.tex
Day 11 21.11.2014.

4. General autoregressive processes, AR(p).

Again working with the zero-mean case for simplicity, the extension of the above to p parameters is the model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t, \qquad (*)$$

with (ϵ_t) WN as before. Since $X_{t-i} = B^i X_t$, we may re-write this as

$$X_t - \phi_1 B X_t - \cdots - \phi_p B^p X_t = \epsilon_t.$$

Write

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

for the pth order polynomial here. Then formally,

$$\phi(B) X_t = \epsilon_t : \qquad X_t = \phi(B)^{-1} \epsilon_t,$$

so if we expand $1/\phi(\lambda)$ in a power series as

$$1/\phi(\lambda) \equiv 1 + \beta_1 \lambda + \cdots + \beta_n \lambda^n + \cdots,$$

$$X_t = \sum\nolimits_{i=0}^{\infty} \beta_i B^i \epsilon_t = \sum\nolimits_0^{\infty} \beta_i \epsilon_{t-i}.$$

This is the analogue of $X_t = \sum_0^\infty \phi^i \epsilon_{t-i}$ for $AR(1)$, and shows that X_t can again be represented as an infinite moving-average process – or *linear process* (X_t is an (infinite) *linear combination* of the ϵ_{t-i}).

Multiply $(*)$ through by X_{t-k} and take expectations. Since $E[X_{t-k} X_{t-i}] = \rho(|k-i|) = \rho(k-i)$, this gives

$$\rho(k) = \phi_1 \rho(k-1) + \cdots + \phi_p \rho(k-p) \qquad (k > 0). \qquad (YW)$$

These are the *Yule-Walker equations*, due to G. Udny YULE (1871-1951) in 1926 and Sir Gilbert WALKER (1868-1958) in 1931.

The Yule-Walker equations (YW) have the form of a *difference equation* of *order p*. The *characteristic polynomial* of this difference equation is

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

which by above is

$$\phi(1/\lambda) = 0.$$

If the roots are $\lambda_1, \cdots, \lambda_p$, the trial solution $\rho(k) = \lambda^k$ is a solution iff λ is one of the roots λ_i. Since the equation is linear,

$$\rho(k) = c_1 \lambda_1^k + \cdots + c_p \lambda_p^k$$

(for $k \geq 0$, and use $\rho(-k) = \rho(k)$ for $k < 0$) is a solution for all choices of constants $c_1, \cdots, c_p$. This is the *general solution* of (YW) if all the roots λ_i are distinct, with appropriate modifications for repeated roots (if $\lambda_1 = \lambda_2$, use $c_1 \lambda_1^k + c_2 k \lambda_1^k$, etc.).

Now $|\rho(k)| \leq 1$ for all k (as $\rho(.)$ is a correlation coefficient), and this is *only possible* if

$$|\lambda_i| \leq 1 \qquad (i = 1, \cdots, p)$$

– that is, all the roots lie inside (or on) the unit circle. This happens (as our polynomial is $\phi(1/\lambda)$) if and only if *all the roots of the polynomial $\phi(\lambda)$ lie outside (or on) the unit circle.* Then $|\rho(k)| \leq 1$ for all k, and when there are no roots of unit modulus, also $\rho(k) \to 0$ as $k \to \infty$ – that is, the influence of the remote past tends to zero, as it should. We shall see below that this is also the condition for the $AR(p)$ process above to be *stationary.*

Example of an $AR(2)$ process.

$$X_t = \frac{1}{3} X_{t-1} + \frac{2}{9} X_{t-2} + \epsilon_t, \qquad (\epsilon_t) \quad WN. \tag{1}$$

Moving-average representation. Let the MA representation of (X_t) be

$$X_t = \sum\nolimits_{i=0}^{\infty} \psi_i \epsilon_{t-i}. \tag{2}$$

Substitute (2) into (1):

$$\begin{aligned} \sum\nolimits_0^\infty \psi_i \epsilon_{t-i} &= \frac{1}{3} \sum\nolimits_0^\infty \psi_i \epsilon_{t-i-1} + \frac{2}{9} \sum\nolimits_0^\infty \psi_i \epsilon_{t-2-i} + \epsilon_t \\ &= \frac{1}{3} \sum\nolimits_1^\infty \psi_{i-1} \epsilon_{t-i} + \frac{2}{9} \sum\nolimits_2^\infty \psi_{i-2} \epsilon_{t-i} + \epsilon_t. \end{aligned}$$

Equate coefficients of ϵ_{t-i}:
$i = 0$ gives $\psi_0 = 1$; $i = 1$ gives $\psi_1 = \frac{1}{3}\psi_0 = 1/3$; $i \geq 2$ gives

$$\psi_i = \frac{1}{3} \psi_{i-1} + \frac{2}{9} \psi_{i-2}.$$

Proceed as above. The characteristic polynomial is

$$\lambda^2 - \frac{1}{3}\lambda - \frac{2}{9} = 0, \qquad \text{or} \qquad (\lambda - \frac{2}{3})(\lambda + \frac{1}{3}) = 0,$$

with roots $\lambda_1 = 2/3$ and $\lambda_2 = -1/3$. The general solution of the difference equation is thus $\psi_i = c_1\lambda_1^i + c_2\lambda_2^i = c_1(2/3)^i + c_2(-1/3)^i$. We can find c_1, c_2 from the values of ψ_0, ψ_1, found above:
$i = 0$ gives $c_1 + c_2 = 0$, or $c_2 = 1 - c_1$.
$i = 1$ gives $c_1.(2/3)+(1-c_1)(-1/3) = \psi_1 = 1/3$: $2c_1-(1-c_1) = 1$: $c_1 = 2/3$, $c_2 = 1/3$. So

$$\psi_i = \frac{2}{3}(\frac{2}{3})^i + \frac{1}{3}(\frac{-1}{3})^i = (\frac{2}{3})^{i+1} - (\frac{-1}{3})^{i+1},$$

and

$$X_t = \sum_0^\infty [(\frac{2}{3})^{i+1} - (\frac{-1}{3})^{i+1}]\epsilon_{t-i},$$

giving the moving-average representation, as required.
Autocovariance. Recall the Yule-Walker equations

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2)$$

for $AR(2)$. As before,

$$\rho(k) = a\lambda_1^k + b\lambda_2^k$$

for some constants a, b. Taking $k = 0$ and using $\rho(0) = 1$ gives $a + b = 1$: $b = 1 - a$. So here,

$$\rho(k) = a(2/3)^k + (1-a)(-1/3)^k.$$

Taking $k = 1$ in the Yule-Walker equations gives

$$\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1),$$

which as $\rho(0) = 1$ and $\rho(-1) = \rho(1)$ gives

$$\rho(1) = \phi_1/(1-\phi_2).$$

As here $\phi_1 = 1/3$ and $\phi_2 = 2/9$, this gives $\rho(1) = 3/7$. We can now use this and the above expression for $\rho(k)$ to find a: taking $k = 1$ and equating,

$$\rho(1) = 3/7 = a.(2/3) + (1-a).(-1/3).$$

That is,

$$(\frac{3}{7}+\frac{1}{3}) = a.(\frac{2}{3}+\frac{1}{3}) = a:$$

$a = (9+7)/21 = 16/21$. Thus

$$\rho(k) = \frac{16}{21}(\frac{2}{3})^k + \frac{5}{21}(\frac{-1}{3})^k.$$

Note. For large k, the first term dominates, and

$$\rho^k \sim \frac{16}{21}.(\frac{2}{3})^k \qquad (k \to \infty).$$

AR(p) processes (continued). 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 $\sigma^2 = var \epsilon_t$,

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

The condition

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

(in words: (ψ_i) is *square-summable*, or is *in* ℓ_2) is necessary and sufficient for
(i) $var X_t < \infty$;
(ii) the series $\sum \psi_i \epsilon_{t-i}$ in the moving-average representation to be convergent in mean square – or, in ℓ_2.
So for convergence in ℓ_2, $\sum \psi_i^2 < \infty$ is the necessary and sufficient condition (NASC) for the moving-average representation of X_t to *exist*. Since $\sum \psi_i \epsilon_{t-i}$ is (when convergent) stationary (because (ϵ_t) is stationary: if $\sum \psi_i^2 < \infty$, then X_t is stationary. The converse is also true; see Section 5 below.

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

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

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 $(*)$,

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

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

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

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

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

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 *is* geometrically decreasing: $\rho_\tau = \rho^\tau$ – 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. But for $AR(p)$, $p > 1$, 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

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

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

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

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

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 $0 \le j \le q-k$, as $0 \le i \le q$. If however $k > q$, there are no non-zero terms as there are no $i = k + j$ with $0 \le i, j \le q$. So

$$\gamma_0 = \sigma^2 \sum_{j=0}^{q} \theta_j^2, \quad \gamma_k = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}, & \text{if } k = 0, 1, \cdots, q, \\ 0 & \text{if } k > q, \end{cases}$$

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,

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

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

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

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

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

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 $MA(1)$ using $(1+\theta B)$ to $AR(\infty)$ 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 $\theta \ne 1$, 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. But for $|\theta| \ne 1$, exactly *one* of these processes is invertible. So if we restrict attention to invertible MA processes, *identifiability* is restored in general ($|\theta| \ne 1$), but not when $|\theta| = 1$, $\theta \ne 1$.