Multivariate ARMA, Kalman filter

October 1, 2015

The finite MA class of models

$$y_t = \sum_{s=0}^{k} a_s \varepsilon_{t-s} = a(L)\varepsilon_t \,.$$

y may be $m \times 1$, in which case a_s is $m \times m$. $\varepsilon \sim N(0, \Sigma)$, i.i.d. Or sometimes just mean 0, variance Σ, not serially correlated.

Properties:

- Dense in the space of LR stationary processes.
- Closed under taking linear combinations.
- Closed under taking subvectors.

- To keep uniqueness, must restrict parameter space to fundamental MA's. This restriction (on roots) is quite nonlinear. But the fundamental MA's form a closed set with open interior, since roots are continuous functions of parameters.

Is the set of fundamental MA operators convex?

This is an important point, since if it is convex, then iterative methods for maximizing likelihood subject to the constraint are likely to work well, because concave functions over a convex set have a unique maximum, while with a non-convex constraint set there can be multiple boundary maxima, even for concave functions.

For second-order operators, the set is convex. There is a famous diagram showing the set of values of ρ_1 and ρ_2 for which the roots of $1 + \rho_1 z + \rho_2 z^2$ all lie on or outside the unit circle, shown in Figure 1.

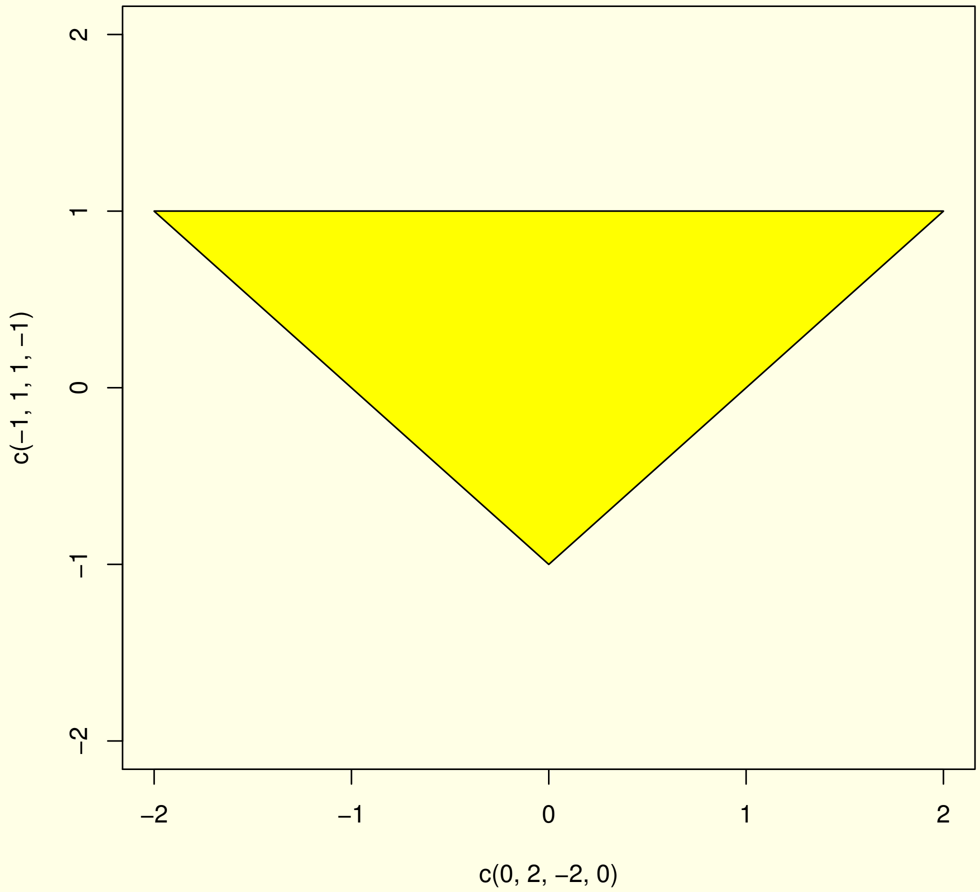

Figure 1: ρ_1, ρ_2 values yielding invertible $1 + \rho_1 L + \rho_2 L^2$

This is obviously a convex region.

However, beyond two dimensions the region is no longer convex. For example, consider $P(L) = 1 - 3L + 3L^2 - L^3$ and $Q(L) = 1 + L$. The first has three roots, all 1. The second has one root of -1. An equal weighted linear combination of the two, $1 - L + 1.5L^2 - .5L^3$ has a pair of complex roots with absolute value .89, and one real root of 2.5. So the region is not convex for third order polynomials.

The finite AR class of models

$$y_t = \sum_{s=1}^{k} b_s y_{t-s} + \varepsilon_t, \qquad \text{or } b(L)y_t = \varepsilon_t \; .$$

$\varepsilon \sim N(0, \Sigma)$, or sometimes just mean 0, variance Σ, not correlated with past y's, and therefore not serially correlated.

Properties:

- Dense in the space of LR stationary processes, plus includes some types of non-stationary processes
- *Not* closed under taking linear combinations

- *Not* closed under taking subvectors.

- No uniqueness problem. Every set of real numbers used to populate b_s, $s = 1, \ldots, k$ results in a distinct model. Restrictions like that to obtain fundamental MA's if we want to consider only stationary models. But this restriction is not needed to prevent redundancy.

Finite-order ARMA models

$$B(L)y_t = A(L)\varepsilon_t \ ,$$

where $\varepsilon_t \perp \{y_s,\, s < t\}$ (and ε_t is therefore the innovation in y at t) and B and A are finite-order polynomials in L, perhaps with matrix-valued coefficients.

Properties:

- Contains MA and AR models, so is also dense in the LR class of models.
- Closed under taking linear combinations.
- Like the finite-order AR class, contains non-stationary as well as stationary models.

- Has the same problems as the MA class with possible redundancy in the $A(L)$ parameter space.

- Has the same problem as the AR class with the restrictions on $B(L)$ needed if we want to restrict to stationary models.

- Has its own special, severe problem of non-uniqueness, because of possible cancellation between AR and MA roots.

AR and MA representations

- When B is a finite-order polynomial and there is no root z of $|B(z)| = 0$ with $|z| = 1$, and when ε_t are i.i.d. with finite variance,

$$y_t = B^{-1}(L)\varepsilon_t \Rightarrow B(L)y_t = \varepsilon_t$$

$$y_t = B(L)\varepsilon_t \Rightarrow B^{-1}(L)y_t = \varepsilon_t$$

- In fact these relations are more general. $B(z)$ can be defined even for infinite-order z, so long as the coefficients go to zero sufficiently fast, and therefore the finite-order requirement can be dropped.

AR and MA representations, cont.

- When B is a finite-order polynomial in non-negative powers of L and there is no root z of $|B(z)| = 0$ with $|z| = 1$, ε_t are i.i.d. with finite variance, and ε_t is uncorrelated with y_{t-s}, all $s > 0$,

$$B(L)y_t = \varepsilon_t \Rightarrow y_t = B^{-1}(L)\varepsilon_t$$

MAR to R_X

If $X(t) = A(L)\varepsilon_t$, with ε_t i.i.d. and $\mathrm{Var}(\varepsilon_t) = \Omega$, the matrix valued function $R_X(t)$ is the list of coefficients of the matrix polynomial

$$R_X(L) = A(L)\Omega A'(L)\,,$$

Where we interpret the “prime” as implying both transposition of the coefficients and replacement of the argument by its inverse, so

$$A'(L) = \sum A'_s L^{-s}\,.$$

Distinguishing fundamental from non-fundamental MAR's

- Generally there are many MA representations for the same stochastic process, but only one that is fundamental.

- If $y_t = A(L)\varepsilon_t$ and $y_t = B(L)\eta_t$, with both ε_t and η_t i.i.d. $N(0, I)$, then since there is only one $R_y(t)$, we must have $R_y(L) = A(L)A'(L) = B(L)B'(L)$.

- If $A(L)$ is finite order, $B(L)$ is finite-order of the same order.

- The roots of $A'(L)$ are the inverses of the roots of $A(L)$, and same for $B(L), B'(L)$.

- Therefore the roots of $A(L)$ are either roots of $B(L)$ or inverses of roots of $B(L)$.

Flipping roots

Since in the univariate case the roots of a polynomial fully characterize it (up to a scale factor), we can in that case convert a finite-order non-fundamental moving average operator $A(L)$ to its fundamental counterpart by "flipping" all its roots that lie inside the unit circle, replacing them with their inverses. We then generally need to rescale to get variances right. E.g.:

$$y_t = \varepsilon_t + 2\varepsilon_{t-1} + .75\varepsilon_{t-2} = (1 + 2L + .75L^2)\varepsilon_t = (1 + 1.5L)(1 + .5L)\varepsilon_t \,.$$

$$y_t = (1 + \tfrac{2}{3}L)(1 + .5L)\eta_t = (1 + 1.16667L + .3333L^2)\eta_t \,.$$

The rescaling comes in because to make R_y the same for these two representations, we need $\mathrm{Var}(\eta_t) = 2.25\,\mathrm{Var}(\varepsilon_t)$. If we normalize instead by making $\mathrm{Var}(\varepsilon_t) = \mathrm{Var}(\eta_t) = 1$, then the coefficients in the fundamental polynomial have to be multiplied by 1.5.

Multivariate flipping

In the multivariate case it is still true that any finite order polynomial can be written as a product of monic polynomials:

$$A(L) = A_0(I - M_1 L)(I - M_2 L) \ldots (I - M_k L) ,$$

where the M_i are rank one matrices. Every root of the characteristic polynomial $|A(z)| = 0$ is the root of one of the characteristic polynomials $|I - M_i z| = 0$. The collection of roots and inverses of roots of $|A(z)|$ will be the collection of roots and inverses of roots of the fundamental MA polynomial $|B(L)|$ as in the univariate case. But now there is no operation that corresponds to simply "flipping" a root. In particular, simply multiplying M_i by a scale factor α so that $|I - \alpha M_i|$ has the inverse of the root of $|I - M_i|$ does not work. (If this is done to each factor, so that no roots are inside the unit circle, it does produce a matrix polynomial that would be the fundamental MA operator for *some* process, but the resulting R_y would not match that implied by the original $A(L)$.)

Some qualitative rules

- If one of $A(L)$ is fundamental and $B(L)$ is not, and if they are normalized so disturbance variance is the identity, then $A_0A_0' - B_0B_0'$ is positive semi-definite.

- This could be a problem for fitting, since it implies that the innovation variance is *larger* for the fundamental representation. (So non-fundamental representations give better forecasts?)

Some qualitative rules

- If one of $A(L)$ is fundamental and $B(L)$ is not, and if they are normalized so disturbance variance is the identity, then $A_0A_0' - B_0B_0'$ is positive semi-definite.

- This could be a problem for fitting, since it implies that the innovation variance is *larger* for the fundamental representation. (So non-fundamental representations give better forecasts?)

- No. A non-fundamental MA disturbance η_t cannot be constructed from $\{X_s, s < t\}$.

- Therefore if $A_0A_0' \succ B_0B_0'$, B is not fundamental. If neither $A_0A_0' \succ B_0B_0'$ nor $B_0B_0' \succ A_0A_0'$, neither is fundamental.

Putting an ARMA into canonical KF form

$$\sum_{s=0}^{k} A_s y_{t-s} = \sum_{s=0}^{\ell} B_s \varepsilon_{t-s} \quad A_0 = I, \quad B_0 = I\ .$$

$$Y_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-k} \\ \varepsilon_t \\ \vdots \\ \varepsilon_{t-\ell} \end{bmatrix} \quad \eta_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad Y_t = GY_{t-1} + M\eta_t \quad Z_t = \begin{bmatrix} I & 0 & \dots & 0 \end{bmatrix} Y_t$$

$$G = \begin{bmatrix} A_1 & \dots & A_k & 0 & \dots & 0 \\ & I & & 0 & 0 \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & & I & 0 \end{bmatrix} \qquad M = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Flipping for a vector MA

(This is here for reference. You don't need to actually be able to do this on an exam, for example.) $y_t = A(L)\varepsilon_t$, with ε_t i.i.d. $N(0, I)$. $A(L)$ is a matrix polynomial in non-negative powers of L. It has at least one root z_0 inside the unit circle. That is, $|z_0| < 1$ and $|A(z_0)| = 0$. This means it is not a fundamental MA operator. If $M(L)$ satisfies $M(L)M'(L) = I$, then, because

$$R_y(L) = A(L)A'(L) = A(L)M(L)M'(L)A'(L)\,,$$
$$y_t = A(L)M(L)M'(L)\varepsilon_t = A(L)M(L)\eta_t$$

is a new MA for y_t that defines the same R_y function and hence the same process. Note also that $M(L)M'(L) = I \Rightarrow M'(L)M(L) = I$. Therefore

$\eta_t = M'(L)\varepsilon_t$ is itself i.i.d. $N(0, I)$ if ε_t is. If $A(L)$ is scalar, i.e. a 1×1 matrix, then we know that we can take

$$M(L) = \frac{z_0 L - 1}{L - z_0}$$

and satisfy both the condition that $M(L)M'(L) = 1$ and the requirement that the z_0 root of $A(L)$ is replaced in $A(L)M(L)$ by a z_0^{-1} root. This is "root flipping". In the scalar case it can be described more simply: factor $A(L)$, replace any factor $1 - aL$ with $|a| > 1$ by $a - L^{-1}$. But this simple explanation only can work because in the scalar case the ordering of the monic factors $1 - aL$ does not matter. In the multivariate case, we proceed as follows.

If $|z_0| < 1$ and $|A(z_0)| = 0$, find the singular value decomposition $UDV' = A(z_0)$. Because $A(z_0)$ is singular, at least one element of the diagonal of D is zero, say the one at the n'th position. Let $Q(z)$ be a diagonal matrix with all ones on the diagonal, but with $(z_0 z-1)/(z-z_0)$ in the $n'th$ position. Then $Q(z)t(Q(z^{-1}))=I$ and therefore $Q(L)Q'(L)$=I. (Here $t()$ is the pure transposition operator, without complex conjugation.) Form $M(L)=VQ(L)V'$. This matrix satisfies $M(L)M'(L)=I$. But then, because $A(L)$ is a finite order polynomial, it is possible to verify that it must be that $A(L)M(L)$ also has the property that its original root at z_0 is now replaced by a root at z_0^{-1}. Perhaps more surprisingly, the special factor in the n'th position, whose own convergent expansion is a two-sided polynomial in powers of both L and L^{-1}, induces no non-zero coefficients on L^{-1} in $A(L)M(L)$.

If you want to see for yourself that this works, a simple example that can be solved by hand is $A(L)=I-\underset{2\times 2}{\mathbf{1}}L$. There is just one non-zero eigenvalue, 2, for A_1 in this case, so $|A(.5)|=0$. Try to calculate the fundamental MA operator corresponding to this $A(L)$, and also the (possibly infinitely long) polynomial in the lag operator that relates η_t to ε_t in this case. Even in this simplest possible case the calculations are somewhat tedious.

Impulse responses

- If $y_t = A(L)\varepsilon_t$ is a MAR of y_t, with $A(L)$ involving only non-negative powers of L (i.e., a one-sided, but not necessarily fundamental, MAR), the coefficients a_{ijs} in the i'th row, j'th column, of A_s, can be regarded as "impulse responses" when treated as a function of s for fixed i, j.

- They represent the time path of $y_{i,t+s}$ starting at time t, if ε_u=0, all u, except $\varepsilon_{jt} = 1$. I.e., the effect on the time path of y_i of a one-time unit disturbance in ε_j.

- This "impulse response" interpretation of the $a_{ij\cdot}$ is valid even if $A(L)$ is generated as the non-convergent one-sided inverse of the $B(L)$ in $B(L)y_t = \varepsilon_t$, when y is a non-stationary autoregressive process.

Finding unconditional covariance matrices for AR models

Simplest approach:

- First get the system into first-order form, by stacking variables:

$$y_t = By_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma)\,.$$

- Since $\mathrm{Cov}(\varepsilon_t, y_{t-1}) = 0$ by assumption, if the process is stationary

$$\Omega = \mathrm{Var}(y_t) = B\Omega B' + \Sigma$$

This is sometimes called a Sylvester equation, but also Lyapunov and Ricatti equations are closely related forms.

- For a stationary model, the solution is, by recursive substitution,

$$\Omega = \sum_{s=0}^{\infty} B^s \Sigma (B')^s \,.$$

Doubling algorithm

I. $\Omega \leftarrow \Sigma$

II. $C \leftarrow B$

III. $d\Omega \leftarrow C\Omega C'$

IV. $\Omega \leftarrow \Omega + d\Omega$

V. If $d\Omega$ very small, return Ω.

VI. $C \leftarrow CC$

VII. go to III

Practice exercise

Suppose $y_t = .7y_{t-1} + \varepsilon_t$, $z_t = .9z_{t-1} + \nu_t$, where ε_t and ν_t are i.i.d. $N(0, I)$ and are uncorrelated with all y_s, z_s for $s < t$. (This makes them the innovations in the joint y, z process, of course.) Find the fundamental univariate MA representation for $x_t = y_t + z_t$. [Hint: This will take the form $(P(L)/Q(L))\eta_t$, where η_t is the univariate innovation in x_t. Form the acf of x, as a ratio of polynomials in the lag operator, and then factor to get expressions in positive powers of L with no roots inside the unit circle.]