Innovations*

These notes cover the lecture of 10/26 and proceed slightly beyond it.

1. Linear Projection

For a random variable y and a collection of random variables $\{x_1, \ldots, x_n\}$, we define the **linear projection operator** $\mathcal{E}$ by

$$\mathcal{E}[y \,|\, \{x_1, \ldots, x_n\}] = \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_n x_n \,, \tag{1}$$

where $\hat{\beta}_1, \ldots, \hat{\beta}_n$ are the solution to

$$\min_{\beta_1,\ldots,\beta_n} \mathrm{Var}(y - \beta_1 x_1 - \ldots - \beta_n x_n) \,. \tag{2}$$

In other words, $\mathcal{E}[y \,|\, \{x_1, \ldots, x_n\}]$ is the best linear predictor of y based on $\{x_1, \ldots, x_n\}$. In order for this projection to be well defined, y must have finite variance. The definition obviously can be interpreted as applying to random vectors y and x_j also.

Here are some useful properties of $\mathcal{E}$.

(i) $\mathcal{E}$ is linear, meaning $\mathcal{E}[aX + bY \,|\, Z] = a\mathcal{E}[X \,|\, Z] + b\mathcal{E}[Y \,|\, Z]$, where a and b are numbers and X, Y, and Z are vectors of random variables.

(ii) The prediction error $Y - \mathcal{E}[Y \,|\, X]$ is uncorrelated with any linear function of X.

(iii) There is a law of iterated projections, formally similar to the law of iterated expectations:

$$\mathcal{E}\left[\mathcal{E}[Y \,|\, X, Z] \,|\, X\right] = \mathcal{E}[Y \,|\, X] \,. \tag{3}$$

Since $Y = (Y - \mathcal{E}[Y \,|\, X, Z]) + \mathcal{E}[Y \,|\, X, Z]$, the law of iterated projections then follows from (ii) and (i).

(iv) If $\mathrm{Cov}(X, Z) = 0$, then $\mathcal{E}[Y \,|\, X, Z] = \mathcal{E}[Y \,|\, X] + \mathcal{E}[Y \,|\, Z]$. This can be derived from (i) and (iii).

For our purposes it will be important that we can use $\mathcal{E}$ to project on countably infinite collections of random variables, not just the finite lists of random variables for which the definition (1) works. In order to do this we have to define a notion of convergence for a sequence of random variables, and to do that we have to define a notion of distance between random variables.

The definition of distance we will use here is the root-mean-square metric. That is, we are taking the distance $\rho(a, b)$ between two random variables to be $\sqrt{\mathrm{Var}(a - b)}$. A sequence $\{a_n\}$ of random variables then converges to a limit b iff $\mathrm{Var}(b - a_n) \to 0$ as $n \to \infty$. If $\rho(a_n - a_m) \to 0$ as $n, m \to \infty$, then $\{a_n\}$ is a Cauchy sequence, and we can take it as defining a random variable which is its limit.

*Copyright 1999 by Christopher A. Sims. This document may be reproduced for educational and research purposes, so long as the copies contain this notice and are retained for personal use or distributed free.

We can now define linear projection onto an infinite sequence of random variables as follows:

$$\mathcal{E}[y \,|\, \{x_s,\, s = 1, \ldots, \infty\}] = \lim_{k \to \infty} \mathcal{E}[y \,|\, \{x_s,\, s = 1, \ldots, k\}] \,. \tag{4}$$

Of course there is a question as to whether the limit on the right-hand side of (4) exists, but it turns out that the answer is always yes. Let us use $\hat{y}_k$ to refer to $\mathcal{E}[y \,|\, \{x_1, \ldots, x_k\}]$ and ζ_k to refer to $\hat{y}_k - \hat{y}_{k-1}$. By the law of iterated projections,

$$\hat{y}_k = \mathcal{E}[\hat{y}_{k+1} \,|\, \{x_1, \ldots, x_k\}] \,. \tag{5}$$

(5) means that ζ_{k+1}, being the error in prediction of $\hat{y}_{k+1}$ based on $\{x_1, \ldots, x_k\}$, is uncorrelated with those x's or any linear combinations of them. But ζ_j for any $j \leq k$ is a linear combination of $\{x_1, \ldots, x_k\}$. So we conclude that the sequence ζ_k of projection revisions as we add more x's is serially uncorrelated.

It is easy to see that

$$\hat{y}_k = \sum_{j=1}^{k} \zeta_j \,. \tag{6}$$

(Here we use the fact that $\hat{y}_0$, the best linear predictor of y based on nothing, is just zero.) Since for every k it must be true that

$$\mathrm{Var}(\hat{y}_k) \leq \mathrm{Var}(y) \,, \tag{7}$$

We know that $\mathrm{Var}(\hat{y}_k)$, being an increasing sequence bounded above, converges, and thus that

$$\lim_{k \to \infty} \mathrm{Var}(\hat{y}_k) = \sum_{j=1}^{\infty} \mathrm{Var}\, \zeta_j \leq \mathrm{Var}(y) < \infty \,. \tag{8}$$

But then

$$\mathrm{Var}(\hat{y}_k - \hat{y}_m) = \sum_{j=m+1}^{k} \mathrm{Var}(\zeta_j) \xrightarrow[m,k \to \infty]{} 0 \,, \tag{9}$$

meaning that $\hat{y}_k$ is a Cauchy sequence with (therefore) a well-defined limit.

So we have proved

Proposition 1. *The limit on the right-hand side of* (4) *exists for any sequence of random variables* $\{x_s\}_{s=1}^{\infty}$.

It is then not hard to verify that properties (i)-(iv) for $\mathcal{E}$ carry over to the case where the conditioning variables are countably infinite collections of random variables instead of just finite vectors. In the case of (ii), the property becomes

ii′. $y - \mathcal{E}[y \,|\, X]$ is uncorrelated with any linear combination of elements of X and any limit of such linear combinations.

2. Linear Prediction

We now turn our attention to a stochastic process y (possibly vector-valued) and define the notation $\hat{y}_k(t)$ for the best k-step-ahead linear predictor of $y(t)$ based on past y. That is

$$\hat{y}_k(t) = \mathcal{E}[y(t) \,|\, \{y(s),\ s < t-k\}]\,. \tag{10}$$

The **innovation** in y at t is $\tilde{y}(t) = y(t) - \hat{y}_1(t)$. $\tilde{y}(t)$ forms an uncorrelated sequence, since from its definition we know that $\tilde{y}(t)$ is uncorrelated with any linear combination of $\{y(t-s), s>0\}$ or any limit of such linear combinations, and $\tilde{y}(t-s)$ is, for any $s>0$, a limit of such linear combinations.

We can form the projection of $y(t)$ on $\{\tilde{y}(t), \tilde{y}(t-1), \ldots, \tilde{y}(t-k)\}$, and it will take the form

$$\mathcal{E}[y(t) \,|\, \{\tilde{y}(t), \ldots, \tilde{y}(t-k)\}] = \sum_{s=0}^{k} A_{st}\, \tilde{y}(t-s)\,. \tag{11}$$

It is important to note here that we write the coefficient matrices A_{st} with no k subscript. This is because of the absence of serial correlation in $\tilde{y}(t)$ and because of property (iv) of the $\mathcal{E}$ operator. By the same kind of argument we went through in proving Proposition 1, the projections in (11) have a well-defined limit as $k \to \infty$. Thus we can write

$$y(t) = \sum_{s=0}^{\infty} A_{st}\, \tilde{y}(t-s) + \bar{y}(t)\,. \tag{12}$$

This equation, known as the **Wold Decomposition**, writes $y(t)$ as the sum of a linear combination of its own past innovations and another component, $\bar{y}(t)$, which is defined by this equation. This $\bar{y}(t)$ component has the property that

$$\bar{y}(t) = \mathcal{E}[\bar{y}(t) \,|\, \{y(t-s), s>0\}] \tag{13}$$

(If this were not true, $\tilde{y}(t)$ would not be the innovation in $y(t)$.) This means that $\bar{y}(t)$ is perfectly predictable one step ahead. It also must be true, from the way that it is constructed, that $\bar{y}(t)$ is uncorrelated with $\tilde{y}(t-s)$ for any $s \geq 0$. And since $\bar{y}(t)$ is a limit of linear combinations of y's dated t and earlier, it is automatically uncorrelated with $\tilde{y}(t+s)$ for $s>0$.

It turns out that not only is $\bar{y}(t)$ perfectly predictable one step ahead, but

Proposition 2. *$\bar{y}(t)$ is perfectly linearly predictable from past y's, arbitrarily far ahead.*

To see this we first note that, because $\tilde{y}(t)$ is for each t a limit of linear combinations of y's dated t and earlier, projection on $\{y(t-s), s \geq 0\}$ is equivalent to projection on $\{\tilde{y}(t-s), s=0,\ldots,k-1; y(t-u), u \geq k\}$. Since

$$\mathcal{E}[\bar{y}(t) \,|\, \{\tilde{y}(t-s), s=0,\ldots,k-1\}] = 0 \tag{14}$$

and

$$\bar{y}(t) = \mathcal{E}[\bar{y}(t) \,|\, \{y(t-s),\ s \geq 0\}]\,, \tag{15}$$

we can apply (iv)) to arrive at the conclusion that

$$\bar{y}(t) = \mathcal{E}[\bar{y}(t) \,|\, \{y(t-s), s \geq k\} \ , \tag{16}$$

and this is true for any k. This is the result we stated in Proposition 2.

If $\bar{y}(t) = 0$, we say that y is **linearly regular** at t. If $\bar{y}(t) = y(t)$, we say that y is **linearly deterministic**. A non-stationary process could be linearly regular at some values of t and not linearly regular at others. For a stationary process, linear regularity or linear determinism holds for all t or none, and the coefficients A_{st} in the Wold Decomposition (12) do not need the t subscript and are written simply as A_s.

In economic modeling, we generally take it to be unrealistic to assume that processes have components that can be forecast perfectly arbitrarily far into the future, except possibly for trend or seasonality components. So most models we use imply that processes are either linearly regular or the sum of a linearly regular component and a predictable trend or seasonal component.

The representation of a stationary linearly regular process y as

$$y(t) = \sum_{s=0}^{\infty} A_s \tilde{y}(t-s) \tag{17}$$

is known as the Wold or **fundamental** moving average representation of y.

3. Examples of Linearly Deterministic Processes

Two kinds of stationary, linearly deterministic processes are easy to understand intuitively. One is a process whose paths are all constant. That is, a process in which $x(t)$ has some finite variance distribution, but $x(t) = x(t+1)$ for all t with probability one. The second (which includes a constant process as a special case) is a process whose paths are perfectly periodic, so that with probability one $x(t) = x(t-S)$ for all t and some fixed periodicity S. We sometimes work with such processes as representations of components of seasonality in economic time series.

A third kind of linearly deterministic process, though, is less obvious. Any process whose spectral density is zero over an interval of non-zero length is linearly deterministic.[1] Such a process cannot be forecast perfectly from any finite span of past data, so that $\mathrm{Var}(y(t) - \mathcal{E}[y(t) \,|\, y(t-1), \ldots, y(t-k)]) > 0$ for every k, but nonetheless $\hat{y}_k(t) = y(t)$. If we attempt to fit autoregressive equations to such a process, the fit will, in big enough samples, get better and better as we add more lags, but the sequence of coefficients will be unstable as we add lags, no matter how big the sample. Though we do not expect to encounter such processes as data, it is worth knowing that they exist as a limiting case. Also, trend removal and seasonal adjustment procedures may come fairly close to making S_x zero over an interval, and they may therefore make the adjusted data behave almost as if the process were perfectly predictable. With

[1] The precise result is that a process is linearly deterministic if its spectral density exists and $\int_0^{2\pi} \log(S_x(\omega))\, d\omega$ is not finite.

seasonally adjusted data, for example, we may find that if we add enough lags to the autoregressive model the fitted model starts to "unravel" the seasonal adjustment to achieve accuracy in forecasts that would not have been possible with the unadjusted data.

4. Fundamental and Other MA Representations

For autoregressive models of the form

$$B(L)y(t) = \varepsilon(t) \tag{18}$$

we can always write formally

$$y(t) = B^{-1}(L)\varepsilon(t)\ . \tag{19}$$

Assuming as usual that $B_0 = I$ and the maximum lag in the model is k, we can always find $B^{-1}(L) = A(L)$'s infinite sequence of coefficients A_s on L^s, $s \geq 0$ by polynomial long division. Of course it could be that this sequence does not converge, in which case (19), expressed as

$$y(t) = \sum_{s=0}^{\infty} A_s\, \varepsilon(t-s)\ , \tag{20}$$

does not have any meaning when ε is i.i.d. (It has no meaning since the limit of its partial sums is not a well-defined random variable.) However, if the A_s sequence does converge, in particular if $\sum_0^\infty A_s A_s' < \infty$, then the right-hand side of (20) is well-defined and in that case y, if it is linearly regular for all t, satisfies (20) and $\varepsilon(t)$ is the innovation in y. We formalize this as

Proposition 3. *If*

(a) *ε is a finite-variance, stationary, serially uncorrelated process,*
(b) *$B(L)y(t) = \varepsilon(t)$, where $B(L)$ is a polynomial in non-negative powers of L with $B_0 = I$,*
(c) *$B^{-1}(L) = A(L)$, where A is a polynomial in positive powers of L with $\sum_0^\infty A_s A_s' < \infty$, and*
(d) *y_t is linearly regular for all t,*

then y is stationary and $\varepsilon(t)$ is the innovation in $y(t)$.

Note that it is not assumed in this proposition that B is of finite order, just that it is in positive powers of L with $B_0 = I$ and that $B(L)y(t)$ is a well-defined random variable.

To prove this proposition, we note first that (c) means that $y(t) = A(L)\varepsilon(t)$ is one solution to $B(L)y(t) = \varepsilon(t)$. There are other solutions, in general, however. For any time series (stochastic or not) $z(t)$ that satisfies $B(L)z(t) \equiv 0$, $y(t) = A(L)\varepsilon(t) + z(t)$

also solves (18). However, any such z is obviously linearly deterministic, and we have assumed linear regularity for y and thus that it has no such component.[2]

Then (b) implies that $\varepsilon(t)$ is a limit of linear combinations of $\{y(t-s), s \geq 0\}$, so that

$$\text{Var}\left(y(t) - \mathcal{E}[y(t) \,|\, \{\varepsilon(t-s),\, s>0\}]\right) \geq \text{Var}\left(y(t) - \mathcal{E}[y(t) \,|\, \{y(t-s),\, s>0\}]\right) \ . \quad (21)$$

This follows because the left-hand side is the error from a projection on limits of certain linear combinations of past y's, can be no better a predictor than a projection on all possible linear combinations of past y's. At the same time, the inequality holds in reverse because past y's are linear combinations of past ε's. But then it must be that the two error variances are equal, and thus in turn that the two forecast errors are the same random variable.[3] But from (20) it is clear that $y(t) - \mathcal{E}[y(t) \,|\, \varepsilon(t-s),\, s>0] = \varepsilon(t)$, and since we now know that this is the same as $y(t) - \mathcal{E}[y(t) \,|\, y(t-s),\, s>0] = \tilde{y}(t)$, we have finished our argument.

From this proposition we can conclude that if we are presented with an autoregressive model $B(L)y(t) = \varepsilon(t)$ and the assertion that y is linearly regular and $\varepsilon(t)$ is stationary and not serially correlated, a sufficient condition that $\varepsilon(t)$ be the innovation in y with y stationary is assumption (c) of Proposition 3. Similarly, if we are presented with a model in the form (19) or (20), with the same properties for ε, a sufficient condition that $\varepsilon(t)$ be the innovation in $y(t)$ is that there be a $B(L) = A^{-1}(L)$ that is in non-negative powers of L, has $B_0 = I$ and makes $B(L)y(t)$ a well-defined random variable.

From our discussion of difference equation solutions, we can conclude that when $B(L)$ is of finite order and satisfies $|B(z)| > 0$ for all $|z| \leq 1$, $B(L)$ has an inverse in non-negative powers of L that has a finite sum of squares. So this, with linear regularity, is a sufficient condition that a finite order autoregressive model imply a stationary process for y with ε as its innovation.[4] The proposition also implies that if we are presented with a finite-order $A(L)$ in the model $y(t) = A(L)\varepsilon(t)$, $\varepsilon(t)$ i.i.d., a sufficient condition that $\varepsilon(t)$ be the innovation in $y(t)$ is that $|A(z)| > 0$ for all $|z| \leq 1$.

Note, though, that these are only *sufficient* conditions. We have discussed in class the example of $A(L) = 1 - L$, for which A^{-1} does not exist as a polynomial in positive powers of L with convergent coefficients, but for which nonetheless $y(t) = A(L)\varepsilon(t)$ is

[2] An example of a non-stationary solution to (18) that is not linearly regular, with $B(L)$ invertible: $y(t) = \rho y(t-1) + \varepsilon(t)$, $z(t) = \varepsilon(100)\rho^t$, $y(t) = \sum \rho^s \varepsilon(t-s) + z(t)$. This y is nonstationary, has linearly deterministic component z, and does not have $\varepsilon(t)$ as its innovation for $t = 100$.

[3] Since both forecasts are limits of linear combinations of $\{y(t-s),\, s>0\}$, their average is a predictor of $y(t)$ based on limits of linear combinations of past y's. If both have the same error variance, their average has an even smaller error variance, unless the two errors are perfectly correlated, which would imply the two predictors are the same. But no limit of linear combinations of past y's can produce a better predictor of $y(t)$ than $\mathcal{E}[y(t) \,|\, y(t-s),\, s>0]$, by definition.

[4] Note that if the root condition on $|B(z)|$ fails, the conclusion is *either* that y is non-stationary *or* that ε is not $\tilde{y}$. Usually in econometric modeling we think of $\{B(z)\} = 0$ on or inside the unit circle as implying non-stationarity, but we occasionally encounter situations where instead y is stationary and the failure of the root condition only implies that ε is not its innovation.

the fundamental MA representation. The necessary and sufficient versions of Proposition 3, which we state without proof, are

Proposition 4. *If*

(a) ε *is a finite-variance, stationary, serially uncorrelated process,,*
(b) $B(L)y(t) = \varepsilon(t)$*, where* $B(L)$ *is a square-summable polynomial in non-negative powers of* L *with* $B_0 = I$*, and*
(c) y *is linearly regular,*

then

(i) *if* y *is stationary,* $\varepsilon(t)$ *is the innovation in* $y(t)$ *if and only if* $|B(z)| > 0$ *for all* $|z| < 1$*, and*
(ii) *if* ε *is the innovation in* y*,* y *is stationary if and only if* $|B(z)| > 0$ *for all* $|z| \leq 1$*.*

Proposition 5. *If*

(a) ε *is a finite-variance, stationary, serially uncorrelated process, and*
(b) $y(t) = A(L)\varepsilon(t)$*, where* $A(L)$ *is a polynomial in non-negative powers of* L *with* $A_0 = I$ *and* $\sum A_s' A_s < \infty$*,*

then $\varepsilon(t)$ *is the innovation in* $y(t)$ *if and only if* $|A(z)| > 0$ *for all* $|z| < 1$*.*

5. Autocovariance Functions from MA Representations

If $y(t) = (I + AL)\varepsilon(t)$, with $\mathrm{Var}(\varepsilon(t)) = \Sigma$ and ε i.i.d., then

$$\begin{aligned} R_y(0) = \mathrm{Var}(y(t)) &= E[y(t)y(t)'] = \\ &E[\varepsilon(t)\varepsilon(t)' + A\varepsilon(t-1)\varepsilon(t)' + \varepsilon(t)\varepsilon(t-1)'A' + A\varepsilon(t-1)\varepsilon(t-1)'A'] \\ &= \Sigma + A\Sigma A' \end{aligned} \tag{22}$$

$$\begin{aligned} R_y(1) = &\\ &E[\varepsilon(t)\varepsilon(t-1)' + A\varepsilon(t-1)\varepsilon(t-1)' + \varepsilon(t)\varepsilon(t-2)'A' + A\varepsilon(t-1)\varepsilon(t-2)'A'] \\ &= A\Sigma \end{aligned} \tag{23}$$

$$R_y(-1) = \Sigma A' \tag{24}$$

$$R_y(s) = 0,\ |s| > 1\ . \tag{25}$$

These results can be summarized as

$$\sum_{t=-\infty}^{\infty} R_y(t)L^t = (I + AL)\Sigma(I + A'L^{-1})\ . \tag{26}$$

This formula can be generalized. Whenever $y = A(L)\varepsilon$, with A square-summable and ε i.i.d. and of finite variance Σ,

$$\sum_{t=-\infty}^{\infty} R_y(t)L^t = R_y(L) = A(L)\Sigma A'(L)\,, \tag{27}$$

where we interpret $A'(L)$ as inverting L as well as transposing the coefficients in A, i.e.

$$A'(L) = \sum_{s=-\infty}^{\infty} A'_s L^{-s}\,. \tag{28}$$

Note that (27) involves questionable, though common, notation: one has to see from its L argument $R_y(L)$ that R_y here is an operator, not a function of time with matrix values.

6. MA Representations from Autocovariance Functions

For any linearly regular process y, we have the representation (17), from which we know we can write R_y in the form (27), with $A(L)$ taken from the Wold representation and Σ the covariance matrix of innovations. If A is of finite order, then $R_y(L)$ is also a finite order polynomial, and it is easily seen to have the property that if $|R_y(z)| = 0$, it must be that $R_y(z^{-1})$ is also zero. I.e., every root inside the unit circle corresponds to another one outside the unit circle that is its inverse. Any R_y such that $R_y(L)$ has its roots arranged this way and has $R_y(t) = 0$ for $|t| > k$ for some finite k is the autocovariance function of a linearly regular process, and the corresponding polynomial in L can be factored into the form (27). Directly computing the roots of the polynomial to carry out the factorization is feasible for scalar models of moderate order. For multivariate models, factoring the matrix polynomials is a much messier computation and is seldom carried out directly in practice.

Another approach for finding an MA representation from a given autocovariance function is to use R_y to form the projection of $y(t)$ on $y(t-s)$, $s = 1, \ldots, k$ for some fairly large k. This approximation to the AR representation of the process can then be inverted to obtain an approximate MA representation. For example, if $R_y(t) = 0, 1, 2, 1, 0$ for $t = -2, -1, 0, 1, 2$, and $R_y(t) = 0$ for $|t| > 1$, the coefficients in $\mathcal{E}[y(t) \,|\, \{y(t-s),\, s = 1, \ldots, 5\}$ can be found from the usual OLS formula:

$$\begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ -\frac{2}{3} \\ \frac{1}{2} \\ -\frac{1}{3} \\ \frac{1}{6} \end{bmatrix}\,. \tag{29}$$

If we now invert $B(L) = 1 - (5/6)L + (2/3)L^2 - (1/2)L^3 + (1/3)L^4 + (1/6)L^5$ by polynomial long division (matlab's `filter` command will do this), we get as the first 10 terms:

1.0000 0.8333 0.0278 -0.0324 0.0378 -0.0441 0.0515 0.1066 0.0145 -0.0123 .

In this case we know the true MA representation: $y(t) = (1 + L)\varepsilon(t)$. One can see that inverting the fifth-order approximate AR has rather roughly approximated this.

7. Wold MA from Non-Wold MA

It can happen that we are presented with a model in MA form

$$y(t) = C(L)\eta(t) \tag{30}$$

with $\eta(t)$ i.i.d. and serially uncorrelated with finite variance Σ_η, but with C having its roots in the wrong place, so $\eta(t) \neq \tilde{y}(t)$. For example, such a representation can easily emerge if the model is derived as a linearization of a macroeconomic stochastic equilibrium model. For analyzing the properties of the model or for fitting the model, it may be important to be able to find the Wold representation.

Since the formula (27) relating MA polynomial to autocovariance function is valid whether or not the MA polynomial is invertible, one brute force way to proceed is to use C to construct R_y, then apply the ideas of the previous section to extract the Wold representation polynomial $A(L)$ from R_y. Sometimes, though, it makes sense to proceed more directly. Both A and C have to generate the same R_y, i.e.

$$A(L)\Sigma_\varepsilon A'(L) = R_y(L) = C(L)\Sigma_\eta C'(L) \; . \tag{31}$$

Every root of $A(L)$ must lie on or outside the unit circle. This means that every root of $C(L)$ inside the unit circle must be the inverse of a root of $A(L)$ outside the unit circle. So we can find A from C by finding all the roots of C and "flipping" the roots that are inside the unit circle. That is, we can write $C(L)$ as a product of monomial polynomials, and replace the terms in this product corresponding to roots inside the unit circle with terms that correspond to inverses of those roots. This is straightforward to do for scalar y. For non-scalar y, the fact that matrix multiplication is not commutative makes this approach many times more complicated, and we will not discuss here how to do it explicitly. An explicit treatment appears in Rozanov (1967, p.47), though that discussion includes an algebra slip.

Here's an example for a simple univariate case. Suppose $C(L) = 1-3.5L+1.5L^2$ and $\sigma_\eta^2 = 1$. The roots of C are 2 and 1/3. The 1/3 root tells us that C is not invertible. Writing C as the product of monomials, it is $C = (1-3L)(1-.5L)$. "Flipping the bad root", we arrive at $F(L) = (3-L)(1-.5L)$. By construction, this F has the property that $F(L)\Sigma_\eta F'(L) = C(L)\Sigma_\eta C'(L)$, so it implies the same autocovariance function as C, and all F's roots are outside the unit circle. However, the lead coefficient in F is $F_0 = 3$, not one. To normalize it to 1, we have to multiply F by 1/3 and replace Σ_η by $\Sigma_\varepsilon = 9\Sigma_\eta$. Then with $A = (1/3)F$, we satisfy (31) with an A that is invertible and that is therefore the Wold MA polynomial. It turns out as $1 - (5/6)L + (1/2)L^2$.

8. Shortcuts for Detecting Non-Invertibility

Some facts about invertibility that can be useful as checks:

- If $C(L)\Sigma_\eta C'(L) = A(L)\Sigma_\varepsilon A'(L)$ and A is the Wold MA polynomial, then $\Sigma_\varepsilon - \Sigma_\eta$ is positive semi-definite. I.e., the Wold MA implies higher variance for the disturbance than any other MA consistent with the same autocovariance function. (This is easy to prove, and we went through the argument in class. You should be able to reproduce it.)
- If C is of order k and is a Wold MA polynomial, $\Sigma - C_k \Sigma C_k'$ is positive semi-definite. That is, the lead coefficient in the MA must be "bigger" (in a certain metric) than the last coefficient. This follows because C "reversed" must imply the same R_y as C itself, i.e. $G(L) = \sum C_{k-s} s L^s$ implies the same R_y as does C_s. Once we see this, (8) implies the result. In the univariate case this shortcut is particularly quick to apply, as it becomes simply $|C_k| < 1$.
- If C is a k'th order matrix polynomial with the smallest eigenvalue of C_0 exceeding in absolute value the largest eigenvalue of $\sum_{s=1}^{k} |C_s|$, all roots of $|C(z)|$ are outside the unit circle.

9. Exercise due Friday, 11/19/99

(i) In each of the following models, determine whether it is possible that ε_t is the innovation in y_t and (at the same time) y_t is stationary:
  (a) $y_t = 1.8y_{t-1} - .9y_{t-2} + \varepsilon_t$
  (b) $y_t = 2y_{t-1} - .9y_{t-2} + \varepsilon_t$
  (c) $y_t = .975y_{t-1} - .45y_{t-2} + \varepsilon_t$
  (d) $y_t = \varepsilon_t - 1.9\varepsilon_{t-1} + .9\varepsilon_{t-2}$
  (e) $y_t = \varepsilon_t - 2.11\varepsilon_{t-1} + 1.11\varepsilon_{t-2}$

(ii) Suppose

$$y_t = \varepsilon_t + \begin{bmatrix} 1.3700 & -0.6300 \\ -0.5700 & 1.4300 \end{bmatrix} \varepsilon_{t-1} \,, \tag{32}$$

with ε_t i.i.d. $N(0, \Sigma)$ and

$$\Sigma = \begin{bmatrix} 0.6250 & 0.3750 \\ 0.3750 & 0.6250 \end{bmatrix} .$$

Show that if y is a stationary process, then it can also be represented as

$$y_t = \nu_t + \begin{bmatrix} 0.7585 & 0.3277 \\ 0.3877 & 0.9915 \end{bmatrix} \nu_{t-1} \,, \tag{33}$$

with $\nu(t)$ i.i.d. $N(0, \Omega)$ and

$$\Omega = \begin{bmatrix} 0.8892 & -0.1662 \\ -0.1662 & 0.7508 \end{bmatrix} .$$

Because these processes are all normally distributed, you can show that they are the same processes by showing that they imply the same $R_y(s) = \text{Cov}(y_t, y_{t-s})$ function.

(iii) Show that neither ν nor ε in ii is the innovation in y.

(iv) Calculate, by hand using polynomial long division or by using matlab's `filter` command, the first 6 (if by hand) to 15 (if by matlab) coefficients in the impulse responses for the AR models in (ia)-(ic) above.

(v) Find the fundamental MAR for the y process in (ii). [Extra credit. Or maybe a better way to put it is say that this is a puzzle for your amusement, which you are not actually expected to solve. I think it cannot be solved directly with any method we have yet studied.]

References

Rozanov, Y. A. (1967): *Stationary Random Processes*. Holden-Day, San Francisco, Cambridge, London, Amsterdam, translated by A. Feinstein.