Lecture Notes 3. Moving Average and Autoregressive Representations of Time Series

In these notes, some important ways of expressing linear time series relationships are developed.

1. Lag Operators

The representation and analysis of time series is greatly facilitated by lag operators.

Definition 3.1. Lag operator

$L^j$ is a linear operator which maps $H_i(x)$ onto $H_{i-j}(x)$ such that $L^j x_i = x_{i-j}$.

To see how concept simplifies time series notation, using lag operator notation, the moving averaging representation of a time series may be expressed as $x_i = \alpha(L) \varepsilon_i$, where $\alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j$. $\alpha(L)$ is an example of a lag polynomial.

Lag polynomials fulfill the basic rules of algebra, in the sense that there is an isometric isomorphism\(^1\) between the Hilbert space of lag polynomials (whose metric is the squared sum of coefficients) and the space of algebraic polynomials.

---

\(^1\)Two spaces are isometrically isomorphic if there exists a one to one mapping between them that preserves the distances between elements so that the distance between any pair of elements in one space equals the distance between the mappings of the two elements in the other space.
As a result, restrictions on operations with lag polynomials are straightforward to identify.

This equivalence between the properties of lag operators and complex polynomials is a consequence of the fact that the Hilbert space $l^2$, which is defined as the Hilbert space generated by all square summable sequences of complex numbers $m = (\ldots m_{-1}, m_0, m_1, \ldots)$, endowed with the inner product $\langle m, n \rangle = \sum_{i=-\infty}^{\infty} m_i \overline{n_i}$, is isometrically isomorphic to the Hilbert space of $L^2$ functions, generated by complex valued functions $f(\omega)$ defined on the interval $\omega \in [-\pi, \pi]^2$ and endowed with the inner product $\langle f, g \rangle = \int_{-\pi}^{\pi} f(\omega) \overline{g(\omega)} d\omega$. The key idea is that each element of $l^2$ maps to an element of $L^2$ in a way that preserves distances is formalized in the Riesz-Fischer Theorem. Notice that I state the theorem in terms of an orthonormal basis of $L^2$, the normalized complex exponentials. This orthonormal basis is the foundation of the theory of Fourier analysis, which I will not develop further here.

**Theorem 3.1. Riesz-Fischer Theorem**

Define the sequence of orthogonal polynomials (relative to the $L^2$ norm)

$$\phi_j(\omega) = \frac{e^{-ij\omega}}{\sqrt{2\pi}}$$

and denote $c$ as any square summable sequence $\{\ldots, c_{-1}, c_0, c_1, \ldots\}$. Then there exists a function $f(\omega) \in L^2$ such that

$$i. \quad c_j = \langle \phi_j(\omega), f(\omega) \rangle,$$

$\omega$ is sometimes known as a frequency.
ii. the infinite sum $\hat{f}_c(\omega) = \sum_{j=0}^{\infty} c_j \phi_j(\omega)$ arbitrarily well approximates $f(\omega)$ in the sense that $|f - \hat{f}_c| = 0$.

iii. Any two series, $c$ and $d$, are distinct if and only if the infinite sums $\hat{f}_c$ and $\hat{f}_d$ are also distinct, i.e. $|\hat{f}_c - \hat{f}_d| = 0$.

iv. $|f|^2 = \sum_{j=-\infty}^{\infty} c_j^2$.

As a result, restrictions on operations with lag polynomials are straightforward to identify. For example, the product of $\alpha(L)$ and $\beta(L)$ corresponds to the product of $\alpha(e^{-i\omega})$ and $\beta(e^{-i\omega})$. Of particular importance, one can take the inverse of $\alpha(L)$ if and only if $\alpha(e^{-i\omega})$ is invertible. When will this polynomial have an inverse? To answer this question, recall that by the fundamental theorem of algebra, $\alpha(e^{-i\omega})$ can always be factored as the product of simple polynomials, i.e. there exists a representation such that $\alpha(e^{-i\omega}) = \prod_{k} (1 - \lambda_k e^{-i\omega})$. A simple polynomial $1 - \lambda e^{-i\omega}$ has an inverse if $|\lambda| < 1$. Hence $\alpha(e^{-i\omega})$ possesses an inverse if $|\lambda_k| < 1 \forall k$. I will return to the question of when lag polynomials may be inverted.

2. z-transforms

Working with infinite sequences is cumbersome, so it is valuable to be able work with functions that correspond to them. For a given sequence, one such function is its z-transform. The z-transform of any sequence $\pi_j$ is defined as

$$\pi(z) \triangleq \sum_{j=0}^{\infty} \pi_j z^j$$ (3.2)
where \( z = e^{-i\omega} \). The \( z \)-transform is simply another way of describing elements of \( L^2 \) but is particularly easy to work with. Notice that the \( z \)'s are orthogonal but not orthonormal. In order to recover the original sequence of coefficients from the \( z \)-transform, one can employ the formula

\[
\pi_j = \frac{1}{2\pi} \int_{|z|=1} \pi(z) z^j \, dz = \frac{1}{2\pi} \int_{-\pi}^{\pi} \pi(e^{-i\omega}) e^{j\omega} \, d\omega
\]

(3.3)

The \( z \)-transform of the autocovariances,

\[
\sigma_s(z) = \sum_{j=-\infty}^{\infty} \sigma_s(j) z^j
\]

(3.4)

summarizes all second moment information in the time series. Notice that this transform may not exist for all \( \omega \in [-\pi, \pi] \), i.e. the function may be unbounded for some frequencies. The \( z \)-transform \( \alpha(z) \) similarly fully characterizes the Wold moving average representation.

The relationship between \( \sigma_s(z) \) and \( \alpha(z) \) is important from the perspective of inference as it is only the autocovariances which are observable from the data. Time series data are simply realizations of a stochastic process. One can compute sample autocovariances which, under mild regularity conditions, will represent consistent estimates of the population autocovariances. Our question is how to use these estimates to identify the moving average parameters of the process.

As a preliminary to establishing this relationship, the following theorem establishes a relationship between any set of MA coefficients\(^3\) and the associated autocovariances of a time series.

---

\(^3\)This means that the theorem does not assume that the moving average representation is the fundamental one.
Theorem 3.2. Relationship between autocovariances and moving average coefficients.

If \( x_t = \sum_{j=-\infty}^{\infty} \gamma_j \eta_{t-j} \), then \( \sigma_x(z) = \sigma_\eta^2 \gamma(z) \gamma(z^{-1}) \).

pf. Clearly

\[
E(\gamma(L) \eta_t, \gamma(L) \eta_{t-k}) = \sigma_\eta^2 \sum_{j=-\infty}^{\infty} \gamma_j \gamma_{j-k}
\]

(3.5)

Therefore the \( z \)-transform of the autocovariance function is

\[
\sum_{k=-\infty}^{\infty} \sigma_\eta^2 \sum_{j=-\infty}^{\infty} \gamma_j \gamma_{j-k} z^k = \sigma_\eta^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \gamma_j z^j \gamma_{j-k} z^{k-j}
\]

(3.6)

\[
= \sigma_\eta^2 \sum_{k=-\infty}^{\infty} \gamma_{j-k} z^{k-j} \sum_{j=-\infty}^{\infty} \gamma_j z^j = \sigma_\eta^2 \gamma(z) \gamma(z^{-1}).
\]

This factorization, as noted before, did not require that the \( \eta_t \)'s were fundamental for the \( x_t \) process. In general, the autocovariance function of a stochastic process is consistent with an infinity of MA representations. Intuitively, this is obvious, since there is no unique orthogonal basis for \( H_t(x) \). For a specific example, suppose that the autocovariance function of \( x_t \) equals \( 4z^{-1} + 17 + 4z \). It is straightforward to verify that this autocovariance function could have been generated by a moving average representation of \( x \) described by

\[
 x_t = \varepsilon_t + \frac{1}{4} \varepsilon_{t-1}
\]

(3.7)

with \( \sigma_\eta^2 = 16 \), or by a moving average representation
\[ x_t = \eta_t + 4\eta_{t-1} \quad (3.8) \]

where \( \sigma^2_\eta = 1 \). Both are consistent with the autocovariances embedded in the \( z \)-transform. How can we tell whether either one is the fundamental representation?

### 3. Fundamental moving average and autoregressive representations

The relationship between the \( z \)-transform of the autocovariances and the fundamental MA coefficients, can be best understood by considering the relationship between the moving average and autoregressive (AR) representation of \( x_t \) defined as

\[ \beta(L)x_t = \epsilon_t \quad (3.9) \]

where \( \beta_0 = 1 \). If \( \alpha(L)^{-1} \) is well defined, then the AR polynomial is immediately generated by the inversion of the MA polynomial, whose existence and uniqueness is ensured by the second Wold theorem. Consequently, if we can identify the fundamental MA polynomial and it is invertible, we can identify the AR polynomial. This makes intuitive sense, since if an AR representation exists, this will define the projection \( x_{\phi t-1} \).

Repeating an earlier argument, the Fundamental Theorem of Algebra, any MA polynomial \( c(z) \) may be factored such that

\[ c(z) = \prod_k (1 - \lambda_k z) \quad (3.10) \]

which in turn means that the autocovariance generating function can always be written as

\[ \sigma(z) = \sigma^2_\eta \prod_k (1 - \lambda_k z) \prod_k (1 - \lambda_k z^{-1}) \quad (3.11) \]
We assume that $|\lambda_k| \neq 1 \forall k$. Now suppose that $|\lambda_k| < 1 \forall k$. This means that the polynomial $\Pi_k(1-\lambda_kz)$ is invertible. Hence this must be the fundamental MA representation, as it is the MA representation associated with the unique AR representation. Conversely, suppose that some of the roots are greater than one in magnitude. Without loss of generality, suppose that only $\lambda_i$ is greater then one. Rewrite the $\sigma_s(z)$ as

$$
\sigma_s(z) = \sigma_n^2 (1-\lambda_i z)(1-\lambda_i^{-1}z) \Pi_{k=1}^k (1-\lambda_k z) \Pi_{k=1}^k (1-\lambda_k^{-1}z) 
$$

(3.12)

Now consider the following trick. Since

$$
(1-\lambda_i z)(1-\lambda_i^{-1}z) = \lambda_i^2 (1-\lambda_i^{-1}z)(1-\lambda_i^{-1}z^{-1}) 
$$

(3.13)

$\sigma_s(z)$ may be rewritten as

$$
\sigma_s(z) = \sigma_n^2 \lambda_i^2 (1-\lambda_i^{-1}z)(1-\lambda_i^{-1}z^{-1}) \Pi_{k=1}^k (1-\lambda_k z) \Pi_{k=1}^k (1-\lambda_k^{-1}z) 
$$

(3.14)

The polynomial $\Pi_{k=1}^k (1-\lambda_k z)$ now possesses all roots inside the unit circle and therefore is invertible. Hence this is the fundamental MA representation. Notice that the nonfundamental innovation variance is $\sigma_n^2$ whereas the fundamental innovation variance is $\sigma_n^2 \lambda_i^2$. Therefore, by flipping the roots of any MA representation inside the unit circle, one can generate the fundamental polynomial structure.

Another way to interpret the transformation of the nonfundamental MA polynomial is that it is multiplied by terms of the form

$$
\frac{\lambda_k (1-\lambda_k^{-1}z)}{(1-\lambda_k z)}
$$

(3.15)
whenever $|\lambda_k| > 1$.

What happens when $|\lambda_k| = 1$? This occurs, for example, when

$$x_t = \varepsilon_t - \varepsilon_{t-1}. \quad (3.16)$$

In this case, the representation is fundamental, yet there does not exist an autoregressive representation with square summable coefficients. Intuitively, the projection of $x_t$ onto $H_{t-1}(x)$ lies in the closure of the linear combinations lagged $x$’s used to generate the Hilbert space. Therefore, if the fundamental MA representation of a process is such that (for at one) $|\lambda_k| = 1$, then an autoregressive representation does not exist.

Taking these arguments together, one can conclude that a given MA representation is fundamental if $|\lambda_k| \leq 1 \ \forall k$. 