

## 1 Short Introduction to Time Series

A time series is a collection of stochastic variables  $x_1, \dots, x_t, \dots, x_T$  indexed by an integer value  $t$ . The interpretation is that the series represent a vector of stochastic variables observed at equal-spaced time intervals. The series is also some times called a stochastic process.

The distinguishing feature of time series is that of temporal dependence: the distribution of  $x_t$  *conditional* on previous value of the series depends on the outcome of those previous observations, i.e., the outcomes are not independent. For the purpose of analyzing a time series we will usually model the time series over all the non-negative integers:  $x_t ; t = \{0, 1, \dots, \infty\}$  or  $x_t ; t = \{-\infty, \dots, 0, 1, \dots, \infty\}$ . Time 1 *or* time 0 will be the first period that you observe the series. In a specific model you will have to be explicit about the initial value, as will be clear from the following.

### 1.1 Stationarity

**Definition** A time series is called *stationary* (more precisely *covariance stationary*) if

$$\begin{aligned} E(x_t) &= \mu , \\ E[(x_t - \mu)^2] &= \gamma(0) , \\ E[(x_t - \mu)(x_{t+k} - \mu)] &= \gamma(k) = \gamma(-k) = E[(x_t - \mu)(x_{t-k} - \mu)]; \quad k = 1, 2, \dots , \end{aligned}$$

where  $\gamma(k) ; k = 0, 1, \dots$  are independent of  $t$  and finite.

There is a quite long tradition in time series to focus on only the first two moments of the process, rather than on the actual distribution of  $x_t$ . If the process is normally distributed all information is contained in the first two moments and most of the statistical theory of time series estimators is asymptotic and more often than not only dependent on the first two moments of the process. The  $\gamma(k)$ 's for  $k \neq 0$  are called **autocovariances** and if we divide by the variance we obtain the **autocorrelations**  $\rho(k) = \gamma(k)/\gamma(0)$ . These are the correlation of  $x_t$  with it own lagged values.

Note that if  $\Sigma_T$  is the matrix of variances and covariance of  $x_1, \dots, x_T$  then

$$\Sigma_T = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(T-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(T-2) & \dots & \dots & \dots & \gamma(1) \\ \gamma(T-1) & \gamma(T-2) & \dots & \gamma(1) & \gamma(0) \end{pmatrix} .$$

So if we let  $\Omega_T$  be the matrix of autocorrelations, i.e.  $\Sigma_T = \gamma(0)\Omega_T$  we will have

$$\Omega_T = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(T-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(T-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(T-1) & \rho(T-2) & \dots & \rho(1) & 1 \end{pmatrix}.$$

Time series models are simple models for the (auto-) correlation of the  $x_t$ 's that allow us to avoid specifying the variance-covariance matrices explicitly.

**Definition:** A stationary process  $e_t$  with mean 0 is called *white noise* if  $\gamma(k) = 0$  for  $k \neq 0$ . Of course this implies that the autocorrelation matrix is just an identity matrix, so the standard OLS assumptions on the error term can also be formulated as “the error term is assumed to be white noise.”

### 1.1.1 The lag-operator, lag-polynomials, and their inverses

The lag operator  $L$  is defined by  $Lx_t = x_{t-1}$ . We will also define the symbol  $L^k$  as  $L^k x_t = x_{t-k}$ . You should think of the lag-operator as moving the whole process  $\{x_t ; t = -\infty, \dots, \infty\}$ . Notice that it is here practical to assume that the series is defined for all integer  $t$ , rather than starting at some period 0—in applications this may not always make sense and in these cases you may have to worry about your starting values. We will define **lag polynomials** as polynomials in the lag operator as follows. Let  $a(L)$  be the lag polynomial

$$a(L) = a_0 + a_1L + \dots + a_pL^p$$

which is defined as an operator such that

$$a(L)x_t = a_0x_t + a_1x_{t-1} + \dots + a_px_{t-p}.$$

This simply means that the equation above defines  $a(L)$  by the way it operates on  $x_t$ .

A key observation is that we can add and multiply lag-polynomials in exactly the same way as we can add and multiply polynomials in complex variables. For example, if  $a(L) = (1 - aL)$  and  $b(L) = (1 - bL)$  then

$$a(L)b(L)x_t = (1 - aL)(1 - bL)x_t = (1 - aL)(x_t - bx_{t-1}) = x_t - bx_{t-1} - aL(x_t - bx_{t-1})$$

which, after simplifying, gives you

$$a(L)b(L)x_t = x_t - (a + b)x_{t-1} + abx_{t-2} = (1 - (a + b)L + abL^2)x_t.$$

Notice, that if we denote  $c(L) = 1 - (a + b)L + abL^2$ , the coefficients of the lag polynomial  $c(L)$  are equal to the coefficients of the polynomial  $c(z) = 1 - (a + b)z + abz^2 = (1 - az)(1 - bz)$  where  $z$  is a real or complex variable. For a given lag-polynomial  $a(L)$  we therefore define the corresponding

*z-transform* (also a label from the engineering literature)  $a(z)$  where  $z$  is a complex number. The point is that if we define  $a(z)$  and  $b(z)$  as the complex polynomials we get from substituting the complex number  $z$  for  $L$  then we know that  $a(z)*b(z)$  is a complex polynomial which we can denote  $c(z)$ . Because the operations are similar, it is also the case that  $c(L) = a(L) * b(L)$ . Conceptually,  $c(L)$  is a function that works on a time series and  $c(z)$  is totally different animal—we just use  $c(z)$  as a quick way to find the coefficients of  $c(L)$ .

One can invert lag-polynomials (although it is not always the case that an inverse exist, as in the case of matrices). We use the notation

$$a(L)^{-1} \text{ or } \frac{1}{a(L)} .$$

Of course  $a(L)^{-1}$  is defined as the operator such that  $a(L)^{-1}a(L)x_t = a(L)a(L)^{-1}x_t = x_t$ .

All problems concerning inversion of lag polynomials can be reduced to inversion of the first order polynomial  $1 - az$ —we will show the details for the first order polynomial. You all know the formula (valid for  $|a| < 1$ )

$$\frac{1}{1 - a} = 1 + a + a^2 + a^3 + \dots$$

(If you are not fully familiar with this equation, then you should take a long look at it, prove it for yourself, and play around with it. It is very important for the following.)

For  $|a| < 1$  we get

$$\frac{1}{1 - az} = 1 + az + (az)^2 + (az)^3 + \dots$$

which converges if  $|z| = 1$ . In order for infinite polynomials like this to converge the  $z$ -transform  $a(z)$  is defined only for complex numbers of length 1, i.e.,  $|z| = 1$ . We can therefore find the inverse of the lag-polynomial  $1 - aL$  as

$$\frac{1}{1 - aL} = I + aL + (aL)^2 + (aL)^3 + \dots$$

so

$$(1 - aL)^{-1}x_t = x_t + ax_{t-1} + a^2x_{t-2} + \dots$$

The **difference operator**  $\Delta$  is defined as  $\Delta = I - L$  giving  $\Delta x_t = x_t - x_{t-1}$ .

## 1.2 AR models:

The most commonly used type of time series models are the auto regressive (AR) models.

$$x_t = \mu + a_1x_{t-1} + \dots + a_kx_{t-k} + u_t ,$$

where the innovation  $u_t$  is white noise. Here  $k$  is a positive integer called the order of the AR-process. Such a process is usually referred to as an AR( $k$ ) process.

Most of the intuition for AR processes can be gained from looking at the AR(1) process, which is also by far the most commonly applied model. Consider the process:

$$x_t = \mu + ax_{t-1} + u_t \quad (*) .$$

If  $|a| < 1$  the process is called *stable*. F.eks. assume that  $x_0$  is a fixed number. Then  $x_1 = \mu + ax_0 + u_1$  using (\*). You can use (\*) again and find  $x_2 = (1+a)\mu + a^2x_0 + au_1 + u_2$ . Continuing this (often called “iterating” equation (\*) or “recursive iteration”) you get

$$x_t = \frac{1-a^t}{1-a}\mu + x_0a^t + a^{t-1}u_1 + \dots + au_{t-1} + u_t .$$

If the  $u_t$  error terms are i.i.d. with variance  $\sigma^2$  then

$$\text{var}(x_t) = \sigma^2[1 + a^2 + \dots + (a^2)^{(t-1)}] = \frac{\sigma^2(1-a^{2t})}{1-a^2} \rightarrow \frac{\sigma^2}{1-a^2} ,$$

for  $t \rightarrow \infty$ .

Notice that this process is not stationary if  $x_0$  is fixed (as the variance varies with  $t$ ). However, for large  $t$  the process is approximately stationary (the variance is approximately  $1/(1-a^2)$  independently of  $t$ ), and if the process was started at  $-\infty$  then  $x_0$  would be stochastic with variance  $\frac{1}{1-a^2}$ . So the stationary AR(1) is a process that we imagine started in the infinite past. Using the lag operator methodology, we have (for  $\mu = 0$  for simplicity)

$$(1-aL)x_t = u_t \Rightarrow x_t = (1-aL)^{-1}u_t \Rightarrow x_t = u_t + au_{t-1} + a^2u_{t-2} + \dots$$

Compare this to what we obtained from iterating the process, and you can see that lag-operator manipulations in essence is nothing but a convenient way of iterating the process all the way back to the infinite past.

Notice what happens if  $a \rightarrow 1$ : the variance tends to infinity and the stationary initial condition can not be defined - at least in typical case as where  $u_t \sim N(0, \sigma^2)$  since a normal distribution with infinite variance is not well defined.

The AR(1) process

$$x_t = x_{t-1} + u_t ,$$

with  $u_t$  iid white noise, is called a *random walk*. Random walk models (and other models with infinite variance) used to be considered somewhat pathological (they also behave “strangely” when used in regression models). Hall’s 1998 demonstration that consumption is a random walk (more precisely a martingale) in a not-too-far-fetched version of the PIH model, was therefore a major surprise for the profession. (The difference between a random walk and a martingale is mainly that in a random walk the variance of the  $u_t$  term is constant, which it need not be for a martingale.)

Consider the

### 1.3 MA models:

The simplest time series models are the moving average (MA) models:

$$x_t = \mu + u_t + b_1 u_{t-1} + \dots + b_l u_{t-l} = \mu + b(L)u_t,$$

where the innovation  $u_t$  is white noise and the lag-polynomial is defined by the equation. The positive integer  $l$  is called the **order** of the MA-process. MA processes are quite easy to analyze because they are given as a sum of independent (or uncorrelated) variables. However, they may not always be easy to estimate: since it is only the  $x_t$ 's that are observed, the  $u_t$ 's are unobserved, i.e., latent variables.

Consider the simple scalar MA(1)-model (I leave out the mean for simplicity)

$$(*) x_t = u_t + b u_{t-1} .$$

If  $u_t$  is an independent series of  $N(0, \sigma_u^2)$  variables, then this model really only says that  $x_t$  has mean zero and autocovariances:  $\gamma(0) = (1 + b^2)\sigma_u^2$ ;  $\gamma(1) = \gamma(-1) = b\sigma_u^2$ ;  $\gamma(k) = 0$ ;  $k \neq -1, 0, 1$ . (Notice that I here figured out what the model says about the distribution of the observed  $x$ 's. In some economic models, the  $u_t$  terms may have an economic interpretation, but in many applications of time series the the MA- (or AR-) model simply serves as a very convenient way of modelling the autocovariances.) The autocorrelations for the MA(1) model are trivially 0, except for  $\rho(-1) = \rho(1) = \frac{b}{1+b^2}$ .

Consider equation (\*) again. In lag-operator notation it reads

$$x_t = (1 + bL)u_t ,$$

which can be inverted to

$$u_t = (1 + bL)^{-1}x_t = x_t - b x_{t-1} + b^2 x_{t-2} + \dots$$

It is quite obvious that this expression is not meaningful if  $|b| \geq 1$  since the power term blows up. In the case where  $|b| < 1$  the right hand side converges to a well defined random variable.

**Definition:** The scalar MA(q) model is called *invertible* if all the roots of the lag-polynomial  $b(L)$  (strictly speaking the corresponding z-transform  $b(z)$ ) are outside the unit circle.

An ARMA model is a model that has both AR- and MA-component. For example, the model

$$x_t = .3x_{t-1} + .1x_{t-2} + u_t + .4u_{t-1}$$

is called an ARMA(2,1) model.

### Inversion of a general lag polynomial—not on exam.

It is well known that if  $a(z) = 1 + a_1z + \dots + a_kz^k$  is a complex polynomial then

$$a(z) = (1 - \alpha_1z)(1 - \alpha_2z)\dots(1 - \alpha_lz)$$

where  $\frac{1}{\alpha_1}, \dots, \frac{1}{\alpha_k}$  are the roots of the polynomial. Recall that the roots will be complex conjugates if the coefficients of  $a(z)$  are real numbers. Of course the lag-polynomial  $a(L)$  factors the same way. This means that

The inverse of the general scalar lag polynomial is now simply defined by inverting the “component” first order lag polynomials one-by-one just as you invert a complex polynomial.

Example: Consider the polynomial  $a(L) = (1 + .3L - .1L^2)$ . This can be factored as  $a(L) = (1 + .5L)(1 - .2L)$ . Since  $(1 + .5L)^{-1} = 1 - .5L + .25L^2\dots$  and  $(1 - .2L)^{-1} = 1 + .2L + .04L^2\dots$ , we have  $(1 + .3L - .1L^2)^{-1} = (1 - .5L + .25L^2\dots)(1 + .2L + .04L^2\dots) = 1 - .3L + .28L^2\dots$

In general an AR(k) model is called stable if all the roots  $r_1, \dots, r_k$  of the lag polynomial  $a(L)$  are larger than 1 in absolute value. Let  $\alpha_i = \frac{1}{r_i}$ ;  $i = 1, \dots, k$ , then in the stable case we get

$$x_t = a(L)^{-1}u_t = (1 - \alpha_kL)^{-1}\dots(1 - \alpha_1L)^{-1}u_t ;$$

which is well defined since we can invert the stable first order lag polynomials one by one.

**Example:** If you are given for example an AR(2) process, like

$$x_t = 1.5x_{t-1} + x_{t-2} + u_t ,$$

you should be able to tell if the process is stable. In the example we find the roots of the polynomial  $1 - 1.5z - z^2$  to be

$$r_i = .5 * (-1.5 \pm \sqrt{2.25 + 4}) ,$$

so the roots are -2 and .5. Since .5 is less than one in absolute value the process is not stable.