TOPIC 3
INTRODUCTION TO TIME SERIES ANALYSIS

1. Introduction

We now cover the [illegible]me series analysis. As we will see, understanding these concepts is [illegible]nding of the time series models that capture time-varying volatility [illegible] will be developed in topic 4.

2. Covariance St[illegible]s

Let y_t be the valu[illegible] financial variable at time *t*. In practice, the data on the variable we actually observe at time *T* is $\{y_1, y_2, \ldots y_T\}$, where it is assumed that values of the variable are recorded from $t = 1$ onward.

In time series analysis, we say that y_t is a covariance stationary process if the following conditions hold;

(a) $E(y_t) = \mu$ for all t. (The mean of the series is constant over time)
(b) $\mathrm{cov}(y_t, y_{t-j}) = \gamma(j)$. (The covariance between y_t and y_{t-j} depends only on the displacement in time between the two y's, which is *j* periods in this case, and not on the time period t). Note that the autocovariance function is symmetric; that is, $\gamma(j) = \gamma(-j)$. Symmetry reflects the fact that the autocovariances of a covariance stationary series depends only on displacement (i.e. on *j*).

(c) $\mathrm{var}(y_t) = \gamma(0)$ must be finite. Note that $\gamma(0)$ is the variance of y_t since $\mathrm{cov}(y_t, y_t) = \mathrm{var}(y_t)$. It can be shown that no autocovariance can be larger in absolute value than $\gamma(0)$, so if $\gamma(0) < \infty$ then so are all the other autocovariances.

The (population) autocorrelation function is

$$\begin{aligned}
\rho(j) &= \frac{\mathrm{cov}(y_t, y_{t-j})}{\sqrt{\mathrm{var}(y_t)}\sqrt{\mathrm{var}(y_{t-j})}} \\
&= \frac{\mathrm{cov}(y_t, y_{t-j})}{\sqrt{\mathrm{var}(y_t)}\sqrt{\mathrm{var}(y_t)}} \quad \text{by covariance stationarity} \\
&= \frac{\gamma(j)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} \\
&= \frac{\gamma(j)}{\gamma(0)}
\end{aligned}$$

Note: Brooks (book) uses τ_s to denote autocorrelation. This is very uncommon notation in the literature and ρ is typically used.

By contrast, the partial autocorrelation at lag j measures the association between y_t and y_{t-j} after controlling for the effects of the intervening values y_{t-1} through $y_{t-(j-1)}$. The partial autocorrelation at lag j, denoted *p*(*j*) is just the coefficient on y_{t-j} in a (population) regression of y_t on a constant and $y_{t-1}, y_{t-2}, \ldots, y_{t-j}$.

For a cova[illegible]cess the autocorrelations and partial autocorrelations a[illegible]isplacement (*j*) becomes large. The estimator of the autocorrelatio[illegible]ample is found by replacing expected values by sample averages i[illegible] autocorrelation function. Thus,

$$\hat{\rho}(j) = \frac{\frac{1}{T}\sum_{t=j+1}^{T}\left[(y_t-\bar{y})(y_{t-j}-\bar{y})\right]}{\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2} = \frac{\sum_{t=j+1}^{T}\left[(y_t-\bar{y})(y_{t-j}-\bar{y})\right]}{\sum_{t=1}^{T}(y_t-\bar{y})^2}$$

The sample partial autocorrelation at displacement *j* is

$$\hat{p}(j) = \hat{\beta}_j$$

where the fitted regression is

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \ldots + \hat{\beta}_j y_{t-j}.$$

The sampling distribution of both the autocorrelation and partial autocorrelation function is $N(0,1/T)$ so that under the null hypothesis of zero correlation a 95% confidence interval is $\pm(2/\sqrt{T})$ for both the sample autocorrelation and partial autocorrelation coefficients.

3. White Noise Process

White noise processes are the building blocks of time series analysis. Suppose ε_t is distributed with mean zero and constant variance and is serially uncorrelated. That is,

$$\varepsilon_t \sim (0, \sigma^2)$$

and $\mathrm{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for all *t* and *j*. In addition, assume that the variance is finite, that is, $\sigma^2 < \infty$. Such a process is called white noise process (with 0 mean) and is denoted as

$$\varepsilon_t \sim WN(0, \sigma^2).$$

Although ε_t is serially uncorrelated, it is not necessarily independent.

Independence is a property that pertains to a conditional distribution. Let $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ be the information set comprising the past history of the process at time *t*.

If the random variable $\varepsilon_t \mid \Omega_{t-1}$, conditional on the information set Ω_{t-1}, has the same distribution [illegible] dom variable ε_t

$$\varepsilon_t \mid \Omega_{t-1} \sim \text{[illegible]}$$

then ε_t is indepen[illegible]y distributed. Moreover, if random variable ε_t is continuous with p[illegible] f_{ε_t}

$$f_{\varepsilon_t|\Omega_{t-1}}(z \mid \Omega_{t-1}) = f_{\varepsilon_t}(z) \text{ for all possible } z \text{ (realizations of } \varepsilon_t)$$

When ε_t is a white noise process and the ε_t's are independently and identically distributed, then the process for ε_t is said to be independent white noise or strong white noise and is denoted as

$$\varepsilon_t \sim iid\,WN(0, \sigma^2)$$

Conditional and unconditional means and variances for an independent white noise process are identical since the conditional and unconditional distributions are the same. As before, the information set upon which we condition contains the past history of the series so that $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. Then the conditional mean is

$$\begin{aligned} E(\varepsilon_t \mid \Omega_{t-1}) &= \int z f_{\varepsilon_t|\Omega_{t-1}}(z \mid \Omega_{t-1}) dz = \qquad \text{by independence} \\ &= \int z f_{\varepsilon_t}(z) dz = E(\varepsilon_t) = 0 \end{aligned}$$

Note: when we condition on $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$, we reveal the *realizations* of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$. That is why $E(\varepsilon_{t-j} \mid \Omega_{t-1}) = \varepsilon_{t-j}$, for all $j \geq 1$.

and, similarly, the conditional variance is

$$\begin{aligned} \operatorname{var}(\varepsilon_t \mid \Omega_{t-1}) &= E\left[(\varepsilon_t - E(\varepsilon_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}\right] \qquad \text{by independence} \\ &= E\left[\left(\varepsilon_t - E(\varepsilon_t)\right)^2\right] = \sigma^2 \end{aligned}$$

Note: $\operatorname{var}(\varepsilon_{t-j} \mid \Omega_{t-1}) = 0$, for all $j \geq 1$, again in this case we know the realization exactly and variance of deterministic variable is 0.

For a white noise process, the autocovariance function is

$$\gamma(j) = \begin{cases} \sigma^2, j = 0 \\ 0 \end{cases}$$

and the autocorre

$$\rho(j) = \begin{cases} 1 \\ 0 \end{cases}$$

The partial autocorrelation function is

$$p(j) = \begin{cases} 1, j = 0 \\ 0, j \geq 1 \end{cases}$$

since in a population regression of ε_t on its lagged values, the regression coefficients are all zero.

4. General Linear Processes

The Wold representation theorem is a very important result in time series analysis. It says that if y_t is a covariance stationary process, it can be represented as

$$y_t = \alpha + \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

where $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$. The latter condition is to ensure that $\mathrm{var}(y_t)$ is finite. This says that any covariance stationary series can be represented by some infinite distributed lag of white noise. The ε_t's are often called innovations or shocks. This representation for y_t is known as the general linear process. It is general since any covariance stationary process can be represented this way and linear because y_t is a linear function of the innovations. Although Wold's theorem says the innovations are serially uncorrelated, we will also make the assumption that the innovations are independent. Thus, we will assume that the innovations are strong or independent white noise.

The unconditional mean and variance of y_t is, respectively,

$$E(y_t) = E\left(\alpha + \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right)$$

$$= \alpha + \sum_{i=0}^{\infty} b_i E\varepsilon_{t-i}$$

$$= \alpha \text{ [illegible]}$$

$$= \alpha \text{ [illegible]}$$

$$\text{var}(y_t) = \text{[illegible]}$$

$$= \text{var}\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right)$$

$$= \sum_{i=0}^{\infty} b_i^2 \ \text{var}(\varepsilon_{t-i}) \text{ since } \varepsilon_t \text{'s are uncorrelated}$$

$$= \sum_{i=0}^{\infty} b_i^2 \sigma^2$$

$$= \sigma^2 \sum_{i=0}^{\infty} b_i^2$$

We will now calculate the conditional moments of the process under the assumption that the innovations are strong white noise. Define the information set as follows: $\Omega_t = \{y_t, y_{t-1}, y_{t-2}, \ldots\}$. Now

$$E(y_{t+1} \mid \Omega_t) = \alpha + E\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t+1-i} \mid \Omega_t\right)$$

$$= \alpha + E\left(\varepsilon_{t+1} \mid \Omega_t\right) + E\left(\sum_{i=1}^{\infty} b_i \varepsilon_{t+1-i} \mid \Omega_t\right)$$

$$= \alpha + 0 + \sum_{i=1}^{\infty} b_i \varepsilon_{t+1-i}$$

$$= \alpha + \sum_{i=1}^{\infty} b_i \varepsilon_{t+1-i}$$

The expression on the last line above is a conditional mean but it is also the optimal one-step ahead forecast of *y*, conditional on information available at time *t*. Similarly, we can calculate

$$E(y_{t+2} \mid \Omega_t) = \alpha + E\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t+2-i} \mid \Omega_t\right)$$
$$= \alpha + E\left(\varepsilon_{t+2} \mid \Omega_t\right) + E\left(b_1 \varepsilon_{t+1} \mid \Omega_t\right) + E\left(\sum_{i=2}^{\infty} b_i \varepsilon_{t+2-i} \mid \Omega_t\right)$$
$$= \alpha + \sum_{i=2}^{\infty} b_i \varepsilon_{t+2-i}$$

and this conditional mean gives the optimal two-step ahead forecast of *y* on the basis of information available at time *t*. The key point is that the conditional mean *moves* over time – it is time-varying as it depends on *t*. Another way of thinking about this is to consider the one-step ahead forecast of *y* conditional on information available at time *t+1*. It is

$$E(y_{t+2} \mid \Omega_{t+1}) = \alpha + \sum_{i=1}^{\infty} b_i \varepsilon_{t+2-i}$$

which incorporates the latest information to arrive, namely, by way of the innovation ε_{t+1}. In other words, the conditional mean moves over time in response to an evolving information set – the conditional mean depends on the conditioning information set.

Now let us calculate the corresponding conditional variances.

$$\begin{aligned}\operatorname{var}(y_{t+1} \mid \Omega_t) &= E\left[(y_{t+1} - E(y_{t+1} \mid \Omega_t))^2 \mid \Omega_t\right] \\ &= E(\varepsilon_{t+1}^2 \mid \Omega_t) \\ &= E(\varepsilon_{t+1}^2) \text{ by independence} \\ &= \sigma^2\end{aligned}$$

$$\begin{aligned}\operatorname{var}(y_{t+2} \mid \Omega_t) &= E\left[(y_{t+2} - E(y_{t+2} \mid \Omega_t))^2 \mid \Omega_t\right] \\ &= E\left[(\varepsilon_{t+2} + b_1 \varepsilon_{t+1})^2 \mid \Omega_t\right] \\ &= E(\varepsilon_{t+2} + b_1 \varepsilon_{t+1})^2 \text{ by independence} \\ &= (1 + b_1^2)\sigma^2 \text{ since the innovations are uncorrelated}\end{aligned}$$

Note that conditional variance is not time varying – it does not depend on *t*. This is the important observation. It does depend on the forecast horizon, however. In the one step ahead case, the conditional variance is given by one term (σ^2) and in the two step ahead case by the sum of two terms $\sigma^2 + b_1^2\sigma^2$. Note also that the conditional variance is always smaller than the unconditional variance. Nevertheless, the important point is that the conditional variance does not evolve over time – it does not depend on the conditioning information set. To see this, note that

$$\operatorname{var}(y_{t+2} \mid [illegible]$$

and

$$\operatorname{var}(y_{t+3} \mid [illegible]$$

so that the arrival [illegible] by way of y_{t+1} (or equivalently ε_{t+1}) does not affect the conditional va[illegible] undesirable feature for the purposes of modeling financial data. In [illegible]nditional variance appears to change over time in response of the arrival of new information and the model as it stands cannot capture this feature. The reason why the conditional variances above are not time-varying is because the innovations are assumed to be independent white noise. Later, we will relax the assumption of independence and assume that the innovations are white noise with a particular dependence structure. This gives rises to the ARCH\GARCH class of models.

Finally, the two standard error confidence interval for the one and two step ahead forecasts is, respectively,

$$\alpha + \sum_{i=1}^{\infty} b_i \varepsilon_{t+1-i} \pm 2\sigma$$

and

$$\alpha + \sum_{i=2}^{\infty} b_i \varepsilon_{t+2-i} \pm 2\sigma\sqrt{1+b_1^2}$$

5. Parsimonious Models

It is impractical to estimate models of the form

$$y_t = \alpha + \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$
$$\varepsilon_t \sim iid\,WN(0,\sigma^2)$$

from a finite sample of data because there is an infinite number of parameters in the sum to estimate. In many applications, however, simpler specifications involving far fewer parameters (parsimonious specifications) provide good approximations to the representation of a covariance stationary series given above. We will consider two such representations.

(a) The MA(1) Model.

The moving average model of order one (the MA(1) model) is

$$y_t = \alpha + \text{[illegible]}$$
$$\varepsilon_t \sim iid\,W\text{[illegible]}$$

In terms of the W[illegible] $= 0$ for $i = 2,3,\ldots$. The MA(1) is a "short memory" process [illegible]nly on the innovation last period and not on innovations earlie[illegible] results above, the unconditional mean and variance of the M[illegible]

$$E(y_t) = \alpha$$
$$\mathrm{var}(y_t) = (1 + b_1^2)\sigma^2$$

The conditional means and variances of the MA(1) process are

$$E(y_{t+j} \mid \Omega_t) = \begin{cases} \alpha + b_1\varepsilon_t & \text{for j = 1} \\ \alpha & \text{for j > 1} \end{cases}$$

and

$$\mathrm{var}(y_{t+j} \mid \Omega_t) = \begin{cases} \sigma^2 & \text{for j = 1} \\ (1 + b_1^2)\sigma^2 & \text{for j > 1} \end{cases}$$

As before, the conditional variance is not time varying. Notice that for *j*>1, the conditional mean is the same as the unconditional mean and the conditional variance is the same as the unconditional variance.

The two-standard error confidence interval for the one-step ahead forecast from the MA(1) model is

$$\alpha + b_1\varepsilon_t \pm 2\sigma$$

The two-standard error confidence interval for the *j*-step ahead forecast where $j > 1$ is

$$\alpha \pm 2\sigma\sqrt{(1 + b_1^2}$$

Estimation of an MA(1) model requires nonlinear methods (e.g. maximum likelihood). Once estimates of the parameters (α, b_1, σ) are found, they can be substituted into the above two equations.

MA model can be easily extended to higher order processes, MA(q), where q is maximum lag:

$$y_t = \alpha + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \ldots + b_q\varepsilon_{t-q}$$

$$0, \sigma^2)$$

The unconditional ... of the MA(q) are

$$E(y_t) = \alpha$$

$$\mathrm{var}(y_t) = (1 + \sum$$

The conditional means and variances of the MA(q) process are

$$E(y_{t+j} \mid \Omega_t) = \begin{cases} \alpha + \sum_{l=j}^{q} b_l\,\varepsilon_{t+l-q} & \text{for } j \le q \\ \alpha & \text{for } j > q \end{cases}$$

and

$$\mathrm{var}(y_{t+j} \mid \Omega_t) = \begin{cases} \sigma^2 & \text{for } j=1 \\ (1 + \sum_{l=1}^{j-1} b_l^2)\sigma^2 & \text{for } 1 < j \le q \\ (1 + \sum_{l=1}^{q} b_l^2)\sigma^2 & \text{for } j > q \end{cases}$$

Note: you do not have to remember the above formulae, but you need to know how to derive them for given lag j.

(b) The AR(1) Model

The first-order autoregressive model (AR(1)) for y_t is

$$y_t = a + b_1 y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim iidWN(0, \sigma^2)$$

Provided $|b_1| < 1$, y_t can be written (use backward substitution) as infinite distributed lag on the ε_t's where the weights are geometrically declining. Specifically,

$$y_t = a\sum_{i=0}^{\infty} b_1^i + \sum_{i=1}^{\infty} b_1^i \varepsilon_{t-i} + \varepsilon_t = \frac{a}{1-b_1} + \sum_{i=1}^{\infty} b_1^i \varepsilon_{t-i} + \varepsilon_t,$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

AR(1) is a parsimonious representation since there is only one b parameter to estimate, namely b_1. The AR(1) model is referred to as a long memory process because the current value of y is a function of all the past innovations.

The uncon[illegible]riance of the AR(1) process are, respectively,

$$E(y_t) = \text{[illegible]}$$

$$\mathrm{var}(y_t) = \text{[illegible]}$$

Notice the requirement that $|b_1| < 1$ ensures the unconditional mean and variance are both finite (which would not be the case if $|b_1| = 1$) and that the conditional variance is positive. In fact, $|b_1| < 1$ is the condition for the AR(1) process to be covariance stationary.

The mean and variance of y_{t+1}, conditional on the information set at time t, is as follows. Note that for an AR(1) process, the only information relevant at time t is y_t since earlier y values don't impact on y_{t+1}.

$$\begin{aligned}
E(y_{t+1} \mid y_t) &= E[a + b_1 y_t + \varepsilon_{t+1} \mid y_t] \\
&= a + E(b_1 y_t \mid y_t) + E(\varepsilon_{t+1} \mid y_t) \\
&= a + b_1 y_t
\end{aligned}$$

$$\begin{aligned}
\mathrm{var}(y_{t+1} \mid y_t) &= \mathrm{var}(a + b_1 y_t + \varepsilon_{t+1} \mid y_t) \\
&= \mathrm{var}(a + b_1 y_t \mid y_t) + \mathrm{var}(\varepsilon_{t+1} \mid y_t) \\
&= \mathrm{var}(\varepsilon_{t+1}) \text{ since the } \varepsilon_t\text{'s are independent} \\
&= \sigma^2
\end{aligned}$$

In general,

$$E(y_{t+j} \mid y_t) = a(1 + b_1 + b_1^2 + \ldots + b_1^{j-1}) + b_1^j y_t$$

$$\mathrm{var}(y_{t+j} \mid y_t) = (1 + b_1^2 + \ldots + b_1^{2(j-1)})\sigma^2$$

The conditional variance is not time varying. Consider the case when $j = 2$. The two standard error confidence interval for the forecast of y_{t+2}, conditional on y_t is

$$\left(a(1 + b_1) + b_1^2 y_t\right) \pm 2\sigma\sqrt{(1 + b_1^2)}$$

Estimates of the parameters of an AR(1) model can easily be obtained by OLS methods. Finally, notice that as $j \to \infty$, the conditional moments converge to the unconditional ones.

Similarly to MA p[rocesses, AR process]es can be extended to higher order p, AR(p).

$$y_t = \alpha + b_1 y_{t-1} + \dots + \varepsilon_t$$
$$\varepsilon_t \sim iidWN(0, \sigma^2)$$

The unconditiona[l mean i]s easy to obtain:

$$E(y_t) = \frac{\alpha}{1 - \sum_{i=1}^{p} b_i}$$

To find unconditional variance $\gamma(0)$ and covariances $\gamma(j)$ one needs to solve a system of p Yule-Walker equations:

$$\gamma(j) = b_1\gamma(j-1) + b_2\gamma(j-2) + b_3\gamma(j-3) + \ldots + b_p\gamma(j-p)$$

Conditional moments are rather cumbersome in their general notation. However, we can apply the same principles to find them.

6. Stationarity of AR(p) models and characteristic equations

Start with AR(p) model:
$y_t = \alpha + b_1 y_{t-1} + b_2 y_{t-2} + \ldots b_p y_{t-p} + \varepsilon_t, \ \varepsilon_t \sim iidWN(0, \sigma^2)$

Introduce lag operator L: $Ly_t = y_{t-1}; L^2(y_t) = L(L(y_t)) = y_{t-2}$

Rewrite the model:
$$y_t = \alpha + b_1 y_t + b_2 L^2 y_t + \ldots + b_p L^p y_t + \varepsilon_t$$

Rearrange:
$$y_t - b_1 L y_t - b_2 L^2 y_t - \ldots - b_p L^p y_t = \alpha + \varepsilon_t$$
$$(1 - b_1 L - b_2 L^2 - \ldots - b_p L^p) y_t = \alpha + \varepsilon_t$$
Note: for any AR(p) model the left-hand side part in () will always have this form $(1 - b_1 L - b_2 L^2 - \ldots - b_p L^p)$. For example, for an AR(1) process: $(1 - b_1 L)$

Form the **characteristic polynomial** on the basis of $(1 - b_1 L - b_2 L^2 - \ldots - b_p L^p)$:
$$1 - b_1 z - b_2 z^2 - \ldots - b_p z^p$$

According to the **Fundamental theorem of Algebra** any polynomial can be factorized as $1-b_1z-b_2z^2-\ldots-b_pz^p=-b_p(z-z_1^*)(z-z_2^*)\ldots(z-z_p^*)$, where z_1^* are the roots of the **characteristic equation:**

$$1-b_1z-b_2z^2-\ldots-b_pz^p=0$$

This is similar to … d at school for a quadratic polynomial:

$$ax^2+bx+c=a(\ldots \quad x_{1,2}^*=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

After the factorization is done we may put L back instead of z, z^* are now specific numbers, roots of the polynomial.

$$1-b_1L-b_2L^2-\ldots-b_pL^p=-b_p(L-z_1^*)(L-z_2^*)\ldots(L-z_p^*)$$

Note: the roots may be complex, but we are able to handle this.

Then, we want to modify each factor so that it looks like a usual AR(1) form $(1-\gamma L)$,

$$1-b_1L-b_2L^2-\ldots-b_pL^p=-b_p(L-z_1^*)(L-z_2^*)\ldots(L-z_p^*)=$$

$$=-b_p(-z_1^*)\left(1-\frac{1}{z_1^*}L\right)(-z_2^*)\left(1-\frac{1}{z_2^*}L\right)\ldots(-z_p^*)\left(1-\frac{1}{z_p^*}L\right)=$$

$$=-b_p(-z_1^*)(-z_2^*)\ldots(-z_p^*)\left(1-\frac{1}{z_1^*}L\right)\left(1-\frac{1}{z_2^*}L\right)\ldots\left(1-\frac{1}{z_p^*}L\right)$$

Part $-b_p(-z_1^*)(-z_2^*)\ldots(-z_p^*)$ is a constant, which equals to 1 (follows from Vieta's Formulae and the fact that the coefficient corresponding to zero lag is 1 for the lag polynomials).

Rewrite our original model $(1-b_1L-b_2L^2-\ldots-L^p)y_t=\alpha+\varepsilon_t$ as

$$\left(1-\frac{1}{z_1^*}L\right)\left(1-\frac{1}{z_2^*}L\right)\ldots\left(1-\frac{1}{z_p^*}L\right)y_t=\alpha+\varepsilon_t$$

We have now multiple AR(1) models operating on top of each other. For this model overall to be stationary we require that each AR(1) part is stationary, that is $\left|\frac{1}{z_i^*}\right|<1$, which gives us condition on the roots of the characteristic equation $\left|z_i^*\right|>1$ for all *i*.

If z_i^* is complex number $\left|z_i^*\right|=\sqrt{\mathrm{Re}^2+\mathrm{Im}^2}>1$. That is where the “unit circle” comes in.

***Additional use: MA representation of an AR(p) process**

The characterizati[illegible]y important if you want to represent an AR(p) process as an MA[illegible]

For AR(1) it is re[illegible]

$(1 - b_1 L) y_t = \alpha +$ [illegible] $(1 - b_1 L)$

$y_t = \frac{1}{1 - b_1 L}(\alpha +$ [illegible]

Unfortunately we do not know how the lag operator works when it is in the denominator, but we may use the following trick.

We all remember from school that the infinite sum of converging geometric progression is

$\sum_{i=0}^{\infty} b^i = \frac{1}{1-b}$ for $|b| < 1$.

We may use the same trick with $b_1 L$ as long as $b_1 L$ will produce *converging* series, i.e.

$\frac{1}{1 - b_1 L} = \sum_{i=0}^{\infty} (b_1 L)^i$, for $|b_1| < 1$ and the process to the left of L is nicely behaved, non-explosive (or if you are really picky, grows over time in a slower pace than b_1^i decays).

$$y_t = \sum_{i=0}^{\infty} (b_1 L)^i (\alpha + \varepsilon_t) = \sum_{i=0}^{\infty} (b_1 L)^i \alpha + \sum_{i=0}^{\infty} (b_1 L)^i \varepsilon_t = \alpha \sum_{i=0}^{\infty} (b_1)^i + \sum_{i=0}^{\infty} (b_1 L)^i \varepsilon_t = \frac{a}{1 - b_1} + \sum_{i=0}^{\infty} (b_1 L)^i \varepsilon_t$$

Note: the lag operator operating on a constant is a constant!

Here we just showed that a stationary AR(1) process can be represented as MA(∞). Now things are very easy for any stationary AR(p) using the factorization we showed before.

We showed that $(1 - b_1 L - b_2 L^2 - \ldots - L^p) y_t = \alpha + \varepsilon_t$ can be written as

$$\left(1 - \frac{1}{z_1^*} L\right)\left(1 - \frac{1}{z_2^*} L\right) \ldots \left(1 - \frac{1}{z_p^*} L\right) y_t = \alpha + \varepsilon_t$$

Invert the process

$$y_t = \frac{1}{\left(1 - \frac{1}{z_1^*} L\right)\left(1 - \frac{1}{z_2^*} L\right) \ldots \left(1 - \frac{1}{z_p^*} L\right)} (\alpha + \varepsilon_t)$$

$$y_t = \sum_{i_1=0}^{\infty}\left(\frac{1}{z_1^*}L\right)^{i_1}\sum_{i_2=0}^{\infty}\left(\frac{1}{z_2^*}L\right)^{i_2}\cdots\sum_{i_p=0}^{\infty}\left(\frac{1}{z_p^*}L\right)^{i_p}(\alpha+\varepsilon_t) =$$

$$= \frac{\alpha}{\left(1-\frac{1}{z_1^*}\right)\left(1-\frac{1}{z_2^*}\right)\cdots} + \sum^{\infty}\left(\frac{1}{z_1^*}L\right)^{i_1}\sum_{i_2=0}^{\infty}\left(\frac{1}{z_2^*}L\right)^{i_2}\cdots\sum_{i_p=0}^{\infty}\left(\frac{1}{z_p^*}L\right)^{i_p}(\varepsilon_t)$$

I agree the process [illegible] ut it is a valid MA(∞) process.

As a reality check [illegible] the mean of an AR(1) process, or AR(2) process and compare it with the mean of the MA(∞) process above. It should work for the variance as well if you have got a spare time.

We can do a similar inversion for an *invertible* MA(q) into an AR(∞) process.

7. Autocorrelation and Partial Autocorrelation Functions of MA(1) and AR(1) Processes

(a) The MA(1) Process

The autocovariance function for the MA(1) process is

$$\begin{aligned}\gamma(j) &= E\left[(y_t - E(y_t))(y_{t-j} - E(y_{t-j}))\right] \\ &= E((\varepsilon_t + b_1\varepsilon_{t-1})(\varepsilon_{t-j} + b_1\varepsilon_{t-j-1})) \\ &= \begin{cases} b_1\sigma^2, & j = 1 \\ 0, & \text{otherwise } (j \neq 0)\end{cases}\end{aligned}$$

The autocorrelation function is just the autocovariance scaled by the variance

$$\begin{aligned}\rho(j) &= \frac{\gamma(j)}{\gamma(0)} \\ &= \begin{cases} \dfrac{b_1}{1+b_1^2}, & j = 1 \\ 0, & \text{otherwise } (j \neq 0)\end{cases}\end{aligned}$$

The key result is that the autocorrelation function cuts off after lag one.
To calculate the partial autocorrelation function, write the MA(1) model as

$$\varepsilon_t = y_t - \alpha - b_1\varepsilon_{t-1}$$

Lagging by successively more periods gives expressions for the innovation at various dates,

$$\varepsilon_{t-1} = y_{t-1} - \alpha - b_1\varepsilon_{t-2}$$
$$\varepsilon_{t-2} = y_{t-2} - \alpha - b_1\varepsilon_{t-3} \text{ etc.}$$

By backward substitution in the MA(1) process, we obtain

$$y_t = \alpha + b_1\varepsilon_{t-1} + \varepsilon_t = \alpha + b_1\left(y_{t-1} - \alpha - b_1\varepsilon_{t-2}\right) + \varepsilon_t = \alpha + b_1\left(y_{t-1} - \alpha - b_1\left(y_{t-2} - \alpha - b_1\left(\ldots\right)\right)\right) =$$
$$= \alpha\sum_{i=0}^{\infty}\left(-b_1\right)^i - \sum_{i=1}^{\infty}\left(-b_1\right)^i y_{t-i} + \varepsilon_t = \frac{\alpha}{1+b_1} - \sum_{i=1}^{\infty}\left(-b_1\right)^i y_{t-i} + \varepsilon_t$$

Normally the assumption $|b_1| < 1$ is made here, since otherwise the sum becomes infinitely large.

Note: We showed above that *invertible* (when $|b_1| < 1$) MA(1) process can be expressed in terms of AR(∞) process.

(b) The AR(1) Process

To calculate the autocorrelation function, first do similar trick the AR(1) model.

Lagging by successively more periods gives expressions for y at various dates,

$$y_t = a + b_1 y_{t-1} + \varepsilon_t$$
$$y_{t-1} = a + b_1 y_{t-2} + \varepsilon_{t-1}$$
$$y_{t-2} = a + b_1 y_{t-3} + \varepsilon_{t-2} \text{ etc}$$

By backward substitution in the AR(1) process, we obtain

$$y_t = a + b_1(a + b_1(a + b_1(\ldots) + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t$$
$$y_t = a\sum_{i=0}^{\infty} b_1^i + \sum_{i=1}^{\infty} b_1^i\varepsilon_{t-i} + \varepsilon_t = \frac{a}{1-b_1} + \sum_{i=1}^{\infty} b_1^i\varepsilon_{t-i} + \varepsilon_t$$

The assumption of stationarity $|b_1| < 1$ is required here, since otherwise the sum becomes infinitely large.

Note: We showed above that *invertible* (when $|b_1| < 1$) AR(1) process can be expressed in terms of MA(∞) process.

Next step is easy, for the AR(1) model the autocovariance function is

$$\begin{aligned}\gamma(j) &= E\left[(y_t - E(y_t))(y_{t-j} - E(y_{t-j}))\right] \\ &= E(y_t - \tfrac{a}{1-b_1})(y_{t-j} - \tfrac{a}{1-b_1}) \\ &= E(\text{[illegible]} \ldots)(\varepsilon_{t-j} + b_1\varepsilon_{t-j-1} + b_1^2\varepsilon_{t-j-2} + \ldots) \\ &= b_1^j(\text{[illegible]}) \quad \text{[illegible]ince the innovations are uncorrelated} \\ &= \frac{\text{[illegible]}}{1-\text{[illegible]}}\end{aligned}$$

Dividing through [illegible] gives the autocorrelations

$$\rho(j) = b_1^j, j = 0, 1, 2, \ldots$$

Recall that for covariance stationarity, $|b_1| < 1$. Thus the autocorrelations are non-zero at all lags and approach zero as the lag length increases.

Finally, the partial autocorrelation function for the AR(1) is simply

$$p(j) = \begin{cases} b_1, j = 1 \\ 0, j > 1 \end{cases}$$

This follows since, in a regression of y_t on its lagged values, the population regression coefficients on y_{t-i} for $i \geq 2$ are zero, if the true model is an AR(1).

8. The Random Walk Model

The series y_t is a random walk with drift if $y_t = a + y_{t-1} + \varepsilon_t$ where ε_t is a white noise process. This specification can be thought of as an AR(1) process with $|b_1| = 1$. By recursive substitution,

$$y_{t+j} = j \cdot a + y_t + \varepsilon_{t+j} + \varepsilon_{t+j-1} + \ldots + \varepsilon_{t+1}$$

It then follows that

$$E(y_{t+j} \mid y_t) = j \cdot a + y_t$$
$$\text{var}(y_{t+j} \mid y_t) = j \cdot \sigma^2$$

As $j \to \infty$, the mean and variance are unbounded. Consequently, a random walk is not a covariance stationary process. When $|b_1| = 1$, we say that the AR(1) process has a unit root.

The AR(1) model: $y_t = a + b_1 y_{t-1} + \varepsilon_t$ can be reparameterised as

$$\Delta y_t = a + \gamma y_{t-1} + \varepsilon_t \text{ where } \gamma = b_1 - 1$$

The Dickey-Fulle[illegible]ypothesis that $\gamma = 0$ (i.e. $b_1 = 1$) The t-ratio or t-statistic associate[illegible]value of γ is used to test this hypothesis by comparison with [illegible]l value. Often the regression above is augmented with lags of Δy_t [illegible]t-hand side to account for serial correlation. In this case, the test is re[illegible]nted Dickey-Fuller test and is performed in exactly the same [illegible]ller test.