Madras School of Economics
PGDM research and analytics

Time Series Fundamentals

Compiled by the PGDM research cell

This series of notes has been compiled from various books regarding Time Series Analysis and modeling. Some of the crucial fundamental concepts have been presented here. Starting from a quick review of limit theorems and an introduction to time series analysis, we move on to building the mathematics behind these core concepts. Some of our references were - Hamilton (Time Series Analysis), Basic Econometrics (D. Gujarati) and Econometrics (Fumio Hayashi).

September 20, 2020

Contents

Chapter 1

A basic introduction to time series concepts

1.1 Reviewing Limit Theorems

In this section we shall consider the limiting behaviour of a sequence of random variables given by $\{z_n\}$. We can discuss the various modes of convergence in a step wise manner.

- A sequence of random variables $\{z_n\}$ **converges in probability** to a constant α if for any $\epsilon > 0$ we have:

$$\lim_{n\to\infty} Prob(|z_n - \alpha| > \epsilon) = 0 \tag{1.1}$$

  Note that the constant α is called the **probability limit** of z_n and can be also written in the following notations as well:

$$plim_{n\to\infty} z_n = \alpha \tag{1.2}$$

$$z_n \xrightarrow{P} \alpha \tag{1.3}$$

- A sequence of random scalars $\{z_n\}$ **converges almost surely** to a constant α if we have:

$$Prob(\lim_{n\to\infty} z_n = \alpha) = 1 \tag{1.4}$$

  This is a stronger condition than the convergence in probability. That is, if a sequence of random variables converges almost surely, then it converges in probability as well.

- A sequence of random variables $\{z_n\}$ **converges in mean square** to α if we have:

$$\lim_{n\to\infty} E[(z_n - \alpha)^2] = 0 \tag{1.5}$$

- Now in all of the above scenarios our sequence was converging to a constant value, however convergence holds for a target random variable as well. We

can say that a sequence of random variables $\{z_n\}$ converges to a random variable z if:

$$\{z_n - z\} \xrightarrow{P} 0 \tag{1.6}$$

$$z_n \xrightarrow{P} z \tag{1.7}$$

- Let $\{z_n\}$ be a sequence of random variables and let F_n be the cumulative frequency distribution of z_n. We can say that the sequence $\{z_n\}$ **converges in distribution** to random variable z if the CDF F_n of z_n converges to the CDF F of z at every continuity point of F. This condition can be written as:

$$z_n \xrightarrow{D} z \tag{1.8}$$

Note additionally that F is known as the **asymptotic distribution** of z_n.

1.2 Fundamentals of Time Series

In time series analysis we will mostly be dealing with stochastic processes. A **Stochastic process** is basically a sequence of random variables. Now if the index for the random variables is interpreted as representing time, then what we have is essentially a **time series**. Further we note that if $\{z_i\}(i = 1, 2, \cdots)$ is a stochastic process, its **realization** is an assignment for each i of a possible value of z_i. Therefore the realizations of $\{z_i\}$ is essentially a sequence of real numbers. A fundamental point to note about time series is that we only observe the realisations of the stochastic process underlying the time series, only once.
As an example, consider the annual inflation rate of some country between 1950 and 2000. This would essentially be a list of 50 numbers or values which would form **one possible outcome** for the stochastic process underlying the inflation rate variable. If history took a different course, we would have had a different sample of realizations of the same stochastic process. Now if we could observe historical values many times, we could assemble many samples, each containing a different list of 50 inflation rate values. Note that in this case the mean inflation rate for say 1950 would be the mean inflation rate for 1950 across all the historical samples. This kind of a population mean is called the **ensemble mean** and is defined as the average across all possible different states of nature at a given time period.
While it is obviously not possible to observe alternate histories, if we make the assumption that the distribution of the inflation remains unchanged, that is the set of 50 values we observe are all assumed to have come from the same distribution, then we are essentially making a **stationarity assumption**. Further we state **ergodicity** as the level of persistence in the process, that is the extent to which each element will contain some information not available in other elements.

1.2.1 Stationary processes

A stochastic process $\{z_i\}(i = 1, 2, \cdots)$ is said to be strictly stationary if for any given finite integer r and for a set of subscripts: $i_1, i_2, \cdots, i_r$, the joint distribution of $(z_i, z_{i_1}, z_{i_2}, \cdots, z_{i_r})$ depends only on: $(i_1 - i, i_2 - i, \cdots, i_r - i)$ and not on i. What

this basically means is that the length of time period lag is what defines the distributional features and not the start or end of the lag. For example, the distribution of (z_1, z_5) is the same as (z_{12}, z_{16}). The distribution of z_i does not depend on the absolute position of i rather on its relative position. We can infer from this statement that the mean, variance and other higher moments remain the same across all i. Now we note some important definitions within this framework.

- A sequence of **independently and identically distributed** random variables is a stationary process that exhibits no serial dependence.
- There are many aggregate time series such as GDP that are **not stationary** because they exhibit **time trends**. Further we note that many time trends can be reduced to stationary processes. A process is called **trend stationary** if it becomes stationary after subtracting from it a linear function of time. Also, if a process is non stationary but its first difference $z_i - z_{i-1}$ is stationary, then the sequence $\{z_i\}$ is called **difference stationary**.
- A stochastic process is said to be **weakly covariance stationary** if:

  $$E(\boldsymbol{z_i}) \textit{ does not depend on } i$$
  $$Cov(\boldsymbol{z_i}, \boldsymbol{z_{i=j}}) \textit{ depends on the index } j \textit{ and not on } i.$$

  The j^{th} order **autocovariance** denoted by Γ_j is defined as:

  $$\Gamma_j = Cov(\boldsymbol{z_i}, \boldsymbol{z_{i-j}}) \tag{1.9}$$

  Further we note that Γ_j does not depend on i due to covariance stationarity. Another condition thus satisfied is:

  $$\Gamma_j = \Gamma_{-j} \tag{1.10}$$

  We can say the the 0^{th} order autocovariance is nothing but the variance given by:

  $$\Gamma_0 = Var(\boldsymbol{z_i}) \tag{1.11}$$

  The corresponding notation for scalar quantities is :

  $$\gamma_j = \gamma_{-j} \tag{1.12}$$

  If we take a string of n successive values of the stochastic process $(z_i, z_{i+1}, \cdots, z_{i+n-1})$ then by the rule of covariance stationarity we can say that the $(n \times n)$ covariance matrix is the same as that of $(z_1, z_2, \cdots, z_n)$ and is given by:

  $$Var(z_i, z_{i+1}, \cdots, z_{i+n-1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \vdots & \dots & \dots & \dots & \vdots \\ \gamma_{n-2} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} \tag{1.13}$$

Finally, the j^{th} order **autocorrelation coefficient** is given by:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{Cov(z_i, z_{i-1})}{Var(z_i)} \tag{1.14}$$

The plot of $\{\rho_j\}$ against the time index j is called a **correlogram**.

- Another important class of weakly stationary processes is the **white noice process** which is a process with zero mean and no serial correlation.

Covariance stationary process $\{z_i\}$ *is white noise if*
$E(z_i) = 0$ *and* $Cov(z_i, z_{i-j}) = 0$

Additionally we note that an independently and identically distributed sequence with zero mean and finite variance is called an **independent white noise process**.

Chapter 2

Working with difference equations

2.1 Difference Equations

Time series analysis is concerned with dynamic consequences of events over time. We will first consider a dynamic equation that relates the value of a variable y on date ty to its value on the previous date and some other value w_t. This give us a **linear first order difference equation**.

$$y_t = \phi y_{y-1} + w_t \tag{2.1}$$

A difference equation is nothing but an expression that relates a variable y_t to its previous values and the above equation is **first order** because only the first lag of the variable (y_{t-1}) appears in the equation. Now we will try to find out that in this dynamic system, how does y change as we change the values of w. Now before solving a difference equation we will assume that the same equation stands for all dates t. Hence we can now write down the difference equation for each date.

$$\begin{gathered}
\textit{Date } 0\text{: } y_0 = \phi y_{-1} + w_0 \\
\textit{Date } 1\text{: } y_1 = \phi y_0 + w_1 \\
\textit{Date } 2\text{: } y_2 = \phi y_1 + w_2 \\
\textit{Date } t\text{: } y_t = \phi y_{t-1} + w_t
\end{gathered}$$

Now if we are aware of the starting value of y that is y_{-1} and we are also aware of all the t values of w then it is possible for us to simulate this system to obtain all values of y. Getting the value of y_1 as follows:

$$y_1 = \phi y_0 + w_1 = \phi(\phi y_{0-1} + w_0) + w_1 \tag{2.2}$$

$$y_1 = \phi^2 y_{-1} + \phi w_0 + w_1 \tag{2.3}$$

Similarly we can calculate for y_2 as follows:

$$y_2 = \phi y_1 + w_2 \tag{2.4}$$

$$y_2 = \phi(\phi^2 y_{-1} + \phi w_0 + w_1) + w_2 \tag{2.5}$$

$$y_2 = \phi^3 y_{-1} + \phi^2 w_0 + \phi w_1 + w_2 \tag{2.6}$$

We can continue in this manner of recursive substitution, we will get the value of y_t in terms of initial value of y and the history of values that w takes on.

$$y_t = \phi^{t+1} y_{-1} + \phi^t w_0 + \phi^{t-1} w_1 + \cdots + \phi w_{t-1} + w_t \tag{2.7}$$

2.2 Lag Operators

A time series is a collection of observations indexed by the date of the observations. We can write this collection of random variables as:

$$(y_1, y_2, \cdots, y_T) \tag{2.8}$$

Now we note that a time series operator typically transforms one type of series into another type of time series and one such popular operator is the **lag operator**. This lag operator basically gives the previous values of a variable at a particular date. It is represented by L and its operation is shown as:

$$Lx_t = x_{t-1} \tag{2.9}$$

$$L(Lx_t) = L(x_{t-1}) = L^2(x_t) = x_{t-2} \tag{2.10}$$

As a general rule we can define the lag operator's functionality as :

$$L^k x_t = x_t - k \tag{2.11}$$

2.2.1 First order difference equation

Let us first start with a first order difference equation given by:

$$y_t = \phi y_{t-1} + w_t \tag{2.12}$$

This can be written using the lag operator as:

$$y_t = \phi L y_t + w_t \tag{2.13}$$

$$y_t - \phi L y_t = w_t \rightarrow (1 - \phi L) y_t = w_t \tag{2.14}$$

Now we will multiply both sides of this equation with the following operator:

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t) \tag{2.15}$$

This would give the main equation as:

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L) y_t = (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t) w_t \tag{2.16}$$

Consider only the LHS compound operator and expand the operator in brackets to get:

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L) \tag{2.17}$$

$$= (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t) - (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)\phi L \tag{2.18}$$

$$= (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t) - (\phi L + \phi^2 L^2 + \phi^3 L^3 + \cdots + \phi^{t+1} L^{t+1}) \tag{2.19}$$

$$= (1 - \phi^{t+1} L^{t+1}) \tag{2.20}$$

Now we can substitute this compound operator back into our main equation given by equation 16 and obtain:

$$(1 - \phi^{t+1} L^{t+1}) y_t = (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t) w_t \tag{2.21}$$

$$y - \phi^{t+1} y_{t-t-1} = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \cdots \phi^t w_0 \tag{2.22}$$

$$y_t = \phi^t y_{-1} + w_t + \phi w_{t-1} + \cdots + \phi^t w_0 \tag{2.23}$$

An interesting point to notice is the behaviour of this compound operator as t becomes large. We already saw by expanding out the LHS:

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L) y_t = y_t - \phi^{t+1} y_{-1} \tag{2.24}$$

Now note that if t becomes very large and if $|\phi| < 1$ then the expression $\phi^{t+1} y_{-1}$ would tend to 0. Hence we can think of the operator: $(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)$ to be an approximation for the inverse of $(1 - \phi L)$, which would inturn satisfy the following condition:

$$(1 - \phi L)^{-1}(1 - \phi L) = 1 \tag{2.25}$$

With this kind of an operation over our difference equation we can essentially write it in the form:

$$y_t = (1 - \phi L)^{-1} w_t \tag{2.26}$$

$$y_t = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \cdots \tag{2.27}$$

2.2.2 Second order difference equation

Now consider a second order difference equation of the form:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t \tag{2.28}$$

Writing this equation in the lag operator format we get:

$$(1 - \phi_1 L - \phi_2 L^2) y_t = w_t \tag{2.29}$$

For a moment, let us only consider the lag operator polynomial in the LHS. We can essentially factorize this polynomial by selecting numbers λ_1 and λ_2 such that:

$$(1 - \phi_1 L - \phi_2 L) = (1 - \lambda_1 L)(1 - \lambda_2 L) = (1 - [\lambda_1 + \lambda_2] L + \lambda_1 \lambda_2 L^2) \tag{2.30}$$

Clearly in our search for these λ values we look to satisfy the following properties:

$$\lambda_1 + \lambda_2 = \phi_1 \tag{2.31}$$

$$\lambda_1 \lambda_2 = -\phi_2 \tag{2.32}$$

Now obviously we need to ensure that the left hand side of the above equation is equal to the right hand side. With this, we can actually write out a corresponding polynomial that fulfils the same criterion.

$$(1 - \phi_1 z - \phi_2 z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z) \tag{2.33}$$

We must understand that by replacing the lag operator by a scalar z, it would allow us to find values of z that would set the RHS to 0. This makes our job easy since the value of z that sets the RHS to 0 is also the value that would set the LHS

to 0. Basically the solutions would be: $z_1 = \lambda_1^{-1}$ and $z_2 = \lambda_2^{-1}$. Therefore we now have:

$$(1 - \phi_1 z - \phi_2 z^2) = 0 \tag{2.34}$$

Finally by the quadtratic formula we can find the value of z that solves the above equation.

$$z_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \tag{2.35}$$

$$z_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \tag{2.36}$$

Note that with this result we make the statement that the difference equation would be stable if the roots of the lag polynomial lie outside the unit circle.

2.3 White noise

The fundamental building block of various time series processes is the white noise process which is a sequence $\{\epsilon_t\}_{t=-\infty}^{\infty}$ where the elements have zero mean and finite variance.

$$E(\epsilon_t) = 0 \tag{2.37}$$

$$E(\epsilon_t^2) = \sigma^2 \tag{2.38}$$

Also the successive values are uncorrelated across time.

$$E(\epsilon_t, \epsilon_\tau) = 0 \tag{2.39}$$

An independent element white noise process following a normal distribution is called a **Gaussian white noise process**.

$$\epsilon \sim N(0, \sigma^2) \tag{2.40}$$

2.4 Moving average process

Let $\{\epsilon_t\}$ be a white noise sequence, then we can define as process as:

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1} \tag{2.41}$$

This is known as a **first order moving average process** denoted as $MA(1)$. We call this a moving average because Y is contructed from a weighted average of the two most recent values of the white noise terms. The expectation of Y is given by:

$$E(Y_t) = E(\mu + \epsilon_t + \theta\epsilon_{t-1}) = \mu \tag{2.42}$$

We note that the constant term included in the process equation is actually the mean. The variance of Y is given by:

$$E(Y_t - \mu)^2 = E(\epsilon_t + \theta\epsilon_{t-1})^2 \tag{2.43}$$

$$= E(\epsilon_t^2 + \theta^2\epsilon_{t-1}^2 + 2\theta\epsilon_t\epsilon_{t-1}) \tag{2.44}$$

$$= (1 + \theta^2)\sigma^2 \tag{2.45}$$

The first **autocovariance** is given by:

$$E(Y_t - \mu)(Y_{t-1} - \mu) = E(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2}) \tag{2.46}$$

$$= E(\epsilon_t\epsilon_{t-1} + \theta\epsilon_t\epsilon_{t-2} + \theta\epsilon_{t-1}^2 + \theta^2\epsilon_{t-1}\epsilon_{t-2}) \tag{2.47}$$

$$= \theta\sigma^2 \tag{2.48}$$

We further note that the higher autocovariances are all zero. Note that since the mean and variance of $MA(1)$ are not dependent on time, they are considered to be **covariance stationary**. The j^{th} autocorrelation of a covariance stationary process is thus given by:

$$\rho_j = \frac{\gamma_j}{\gamma_0} \tag{2.49}$$

$$Corr(Y_t, Y_{t-j}) = \frac{Cov(Y_t, Y_{t-j})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-j})}} = \frac{\gamma_j}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \rho_j \tag{2.50}$$

We can write the first autocorrelation of $MA(1)$ as follows:

$$\rho_1 = \frac{\theta\sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{1 + \theta^2} \tag{2.51}$$

Note that all the higher autocorrelations will be zero in this case.

2.4.1 MA(q)

The q^{th} order moving average process can be described as:

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q} \tag{2.52}$$

Mean of this process is given by:

$$E(Y_t) = E(\mu) + E(\epsilon_t) + \theta_1 E(\epsilon_{t-1}) + \theta_2 E(\epsilon_{t-2}) + \cdots + \theta_q E(\epsilon_{t-q}) = \mu \tag{2.53}$$

Variance of the $MA(q)$ process is given by:

$$\gamma_0 = E(Y_t - \mu)^2 = E(\epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q})^2 \tag{2.54}$$

Now since the white noise terms are uncorrelated, we can write the variance as:

$$\gamma_0 = \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2 + \cdots + \theta_q^2\sigma^2 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2 \tag{2.55}$$

Coming to the computation of the j^{th} lag covariance and dropping out the cross product terms of the white noise terms (since they are uncorrelated and will resolve to zero), we get:

$$\gamma_j = E[(\epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q})(\epsilon_{t-j} + \theta_1\epsilon_{t-j-1} + \cdots + \theta_q\epsilon_{t-j-q})] \tag{2.56}$$

$$= E(\theta_j \epsilon_{t-j}^2 + \theta_{j+1}\theta_1 \epsilon_{t-j-1}^2 + \cdots + \theta_q \theta_{q-j} \epsilon_{t-q}^2) \tag{2.57}$$

Therefore finally we can get the covariance function as:

$$\gamma_j = \begin{cases} [\theta_j + \theta_{j+1}\theta_1 + \cdots + \theta_q \theta_{q-j}]\sigma^2, & \text{if } j \leq q. \\ 0, & \text{if } j > q. \end{cases} \tag{2.58}$$

We can note the corresponding values for the $MA(2)$ process:

$$\gamma_0 = [1 + \theta_1^2 + \theta_2^2]\sigma^2 \tag{2.59}$$

$$\gamma_1 = [\theta_1 + \theta_2\theta_1]\sigma^2 \tag{2.60}$$

$$\gamma_2 = [\theta_2]\sigma^2 \tag{2.61}$$

$$\gamma_3 = \gamma_4 = \cdots = 0 \tag{2.62}$$

In this case, the autocorrelation function is 0 after q lags.

2.5 Infinite order moving average

The previously defined $MA(q)$ process can be written as:

$$Y_t = \mu + \sum_{j=0}^{q} \theta_j \epsilon_{t-j} \tag{2.63}$$

With $\theta_0 = 1$. Now we consider the process that results when $q \to \infty$. this can be shown as:

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \cdots \tag{2.64}$$

This is essentially an $MA(\infty)$ process. We further note that this infinite sequence ensures a covariance stationary process provided that:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \tag{2.65}$$

Note that a sequence of numbers satisfying the above condition is said to be **absolutely summable**. We can now calculate the mean and autocovariances of an $MA(\infty)$ process with absolutely summable coefficients.

$$E(Y_t) = \lim_{T \to \infty} E(\mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \cdots + \psi_T \epsilon_{t-T}) = \mu \tag{2.66}$$

$$\gamma_0 = E(Y_t - \mu)^2 = \lim_{T \to \infty} E(\psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \cdots + \psi_T \epsilon_{t-T})^2 \tag{2.67}$$

$$= \lim_{T \to \infty} (\psi_0^2 + \psi_1^2 + \cdots + \psi_T^2)\sigma^2 \tag{2.68}$$

$$\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu) \tag{2.69}$$

$$= \sigma^2(\psi_j \psi_0 + \psi_{j+1}\psi_1 + \cdots) \tag{2.70}$$

An important point to note is that an absolutely summable $MA(\infty)$ coefficients implies absolutely summable autocovariances as well.

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty \tag{2.71}$$

Chapter 3

The Autoregressive process

3.1 Autoregressive process

A first order autoregression $AR(1)$ is defined by the following equation:

$$Y_t = c + \phi Y_{t-1} + \epsilon_t \tag{3.1}$$

Where again $\{\epsilon_t\}$ is a white noise process. In earlier sections when we looked at the analysis of difference equations, we learnt that if $|\phi| \geq 1$ then the effect of the ϵ terms on Y tend to accumulate rather than die out over time, in which case a covariance stationary process would not exist. If however $|\phi| < 1$ then we would have a covariance stationary process which could be characterized by the following stable equation. Note that this the same equation we obtained after recursively substituting a general difference equation (here $w = (c + \epsilon)$.

$$Y_t = (c + \epsilon_t) + \phi(c + \epsilon_{t-1}) + \phi^2(c + \epsilon_{t-2}) + \cdots \tag{3.2}$$

$$= [c/(1 - \phi)] + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots \tag{3.3}$$

Look closely and we would notice that this is actually an $MA(\infty)$ process with $\psi_j = \phi^j$. Now when we incorporate the stationarity condition that $|\phi| < 1$ then we would satisfy:

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi|^j \tag{3.4}$$

This would end up being equal to $[1/(1 - |\phi|)]$ since it would essentially be a geometric series of partial sums. Mean of the $AR(1)$ process can be represented as:

$$E(Y_t) = \frac{c}{1 - \phi} \tag{3.5}$$

The variance can be represented as:

$$\gamma_0 = E(Y_t - \mu)^2 \tag{3.6}$$

$$= E(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots)^2 \tag{3.7}$$

$$= (1 + \phi^2 + \phi^4 + \cdots)\sigma^2 = \frac{\sigma^2}{1 - \phi^2} \tag{3.8}$$

The j^{th} autocovariance can be shown by:

$$\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu) \tag{3.9}$$

$$= E(\epsilon_t + \phi\epsilon_{t-1} + \cdots + \phi^j\epsilon_{t-j} + \phi^{j+1}\epsilon_{t-j-1} + \cdots)(\epsilon_{t-j} + \phi\epsilon_{t-j-1} + \cdots) \tag{3.10}$$

$$= [\phi^j + \phi^{j+2} + \phi^{j+4} + \cdots]\sigma^2 \tag{3.11}$$

$$= \phi^j(1 + \phi^2 + \phi^4 + \cdots)\sigma^2 \tag{3.12}$$

$$= \frac{(\phi^j)}{(1-\phi^2)}\sigma^2 \tag{3.13}$$

Therefore we can now write the autocorrelation function as:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi^j \tag{3.14}$$

This autocorrelation function follows a pattern of geometric decay. We can make interpretations like - the impact of a unit increase in ϵ_t on Y_{t+j} is basically equal to the correlation beteen Y_t and Y_{t+j}. Note that as seen before we have actually derived the moments of the $AR(1)$ process by essentially viewing it as an $MA(\infty)$ process, however we can directly find those moments out as well.

$$E(Y_t) = c + \phi E(y_{t-1}) + E(\epsilon_t) \tag{3.15}$$

Note that since we have assumed convariance stationarity, we would also have:

$$E(Y_t) = E(Y_{t-1}) = \mu \tag{3.16}$$

Therefore we would have:

$$\mu = \frac{c}{1-\phi} \tag{3.17}$$

In a similar manner we can obtain the second moment in the following way:

$$Y_t = \mu(1-\phi) + \phi Y_{t-1} + \epsilon_t \tag{3.18}$$

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t \tag{3.19}$$

Now we can square both sides and take expectations.

$$E(Y_t - \mu)^2 = \phi^2 E(Y_{t-1} - \mu)^2 + E(\epsilon^2) + 2\phi E(\epsilon(Y_{t-1} - \mu)) \tag{3.20}$$

Note an interesting point that $(Y_{t-1} - \mu)$ is essentially a function of ϵ_{t-1} and so on.

$$Y_{t-1} - \mu = \epsilon_{t-1} + \phi\epsilon_{t-2} + \cdots \tag{3.21}$$

Therefore since ϵ_t is not correlated with those other values of ϵ then it would certainly not be correlated with $(y_{t-1} - \mu)$ either. Hence we can safely that :

$$E[(Y_{t-1} - \mu)\epsilon_t] = 0 \tag{3.22}$$

Also assuming **covariance stationarity** we can also say that:

$$E(y_t - \mu)^2 = E(Y_{t-1} - \mu)^2 = \gamma_0 \tag{3.23}$$

Our main equation would then resolve to:

$$\gamma_0 = \phi^2\gamma_0 + 0 + \sigma^2 \tag{3.24}$$

$$\gamma_0 = \frac{\sigma^2}{1-\phi^2} \tag{3.25}$$

To obtain the j^{th} order autocovariance we would multiply equation 19 with $(Y_{t-j} - \mu)$ and then take expectations.

$$E[(Y_t - \mu)(Y_{t-j} - \mu)] = \phi E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + E[\epsilon_t(Y_{t-j} - \mu)] \tag{3.26}$$

Again by the previous logic, we can say the term $(Y_{t-j} - \mu)$ is actually a function of ϵ_{t-j} onwards and hence is uncorrelated with ϵ_t. We can also write the following:

$$E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] = E[(Y_{t-1} - \mu)(Y_{(t-1)-(j-1)} - \mu)] = \gamma_{j-1} \tag{3.27}$$

Therefore our main equation then becomes:

$$\gamma_j = \phi\gamma_{j-1} \tag{3.28}$$

$$\gamma_j = \phi^j\gamma_0 \tag{3.29}$$

3.1.1 AR(2)

The second order autoregressive function can be written as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \tag{3.30}$$

We can write this equation in the lag format as follows:

$$(1 - \phi_1 L - \phi_2 L^2)Y_t = c + \epsilon_t \tag{3.31}$$

From our earlier discussions regarding difference equations and their stability, we can say that this equation is stable if the roots of the characteristic polynomial lie outside the unit circle. It is only when this condition is satisfied that we can say that the $AR(2)$ process is covariance stationary.

$$(1 - \phi_1 z - \phi_2 z^2) = 0 \tag{3.32}$$

Note that the inverse of this autoregressive operator can be written as:

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1} = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \tag{3.33}$$

Now multiplying both sides of our main equation with this function we get:

$$Y_t = \psi(L)c + \psi(L)\epsilon_t \tag{3.34}$$

Also since c is a constant the operator premultiplied with c would simply become:

$$\psi(L)c = \frac{c}{1 - \phi_1 - \phi_2} \tag{3.35}$$

Additionally we also have the condition that :

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \tag{3.36}$$

Now since we have effectively resolved our $AR(2)$ process to an $MA(\infty)$ process as is evident by equation 34 we can state the mean of the process as:

$$\mu = \frac{c}{1 - \phi_1 - \phi_2} \tag{3.37}$$

Now we can find the second moment by rewriting the main equation as:

$$Y_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \tag{3.38}$$

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t \tag{3.39}$$

Multiplying both sides by $(Y_{t-j} - \mu)$ and taking expectations we would get:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \tag{3.40}$$

Consequently we can find the autocorrelations by dividing throughout with γ_0 and obtain:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} \tag{3.41}$$

Now for $j = 1$ we would get:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \tag{3.42}$$

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \tag{3.43}$$

For $j = 2$ we would get:

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \tag{3.44}$$

For getting the variance we can multiply equation 39 with $(Y_t - \mu)$ and get:

$$E[Y_t - \mu]^2 = \phi_1 E[(Y_t - \mu)(Y_{t-1} - \mu)] + \phi_2 E[(Y_t - \mu)(Y_{t-2} - \mu)] + E[\epsilon_t(Y_t - \mu)] \tag{3.45}$$

Note what the last term in this equation would resolve to:

$$E[\epsilon_t(Y_t - \mu)] = E(\epsilon_t)[\phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t] = \sigma_2 \tag{3.46}$$

The final equation then becomes:

$$\gamma_0 = \phi_1 \rho_1 \gamma_0 + \phi_2 \rho_2 \gamma_0 + \sigma^2 \tag{3.47}$$

Finally we get the variance as follows:

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2^2)[(1 - \phi_2)^2 - \phi_1^2]} \tag{3.48}$$

3.1.2 AR(p)

An autoregressive processes of the p^{th} order can be written as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \tag{3.49}$$

For ensuring stationarity the roots of our characteristic polynomial must lie outside the unit cirlce.

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \tag{3.50}$$

After applying the inverse of the characteristic lag operator polynomial we can obtain the covarince stationary transformation of this process as follows:

$$Y_t = \mu + \psi(L)\epsilon_t \tag{3.51}$$

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)^{-1} \tag{3.52}$$

Note that the stationarity condition is as follows:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \tag{3.53}$$

We can take expectations on the main equation to get the mean as follows:

$$\mu = c + \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu \tag{3.54}$$

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \tag{3.55}$$

Writing the main autoregressive equation in mean deviation form we obtain:

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{y-2} - \mu) + \cdots + \phi_p (Y_{t-p} - \mu) + \epsilon_t \tag{3.56}$$

Now if we multiply both sides by $Y_{t-j} - \mu$ we would essentially obtain the autocovariance functions as:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p} \tag{3.57}$$

Note that for $j = 0$ we will essentially get the variance as:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma^2 \tag{3.58}$$

3.2 ARMA(p,q)

First we note that $ARMA(p, q)$ is a process that includes both autoregressive and moving average terms.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \tag{3.59}$$

We can write the equation in lag operator notation as:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\epsilon_t \tag{3.60}$$

Our precondition for stationarity in the $ARMA$ process is essentially the same condition as the AR process and its stationarity essentially depends on the AR parameters.

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = 0 \tag{3.61}$$

The above equation should ideally have roots that lie outside the unit circle for the equation system to be stable and for stationarity to exist. We would now divide both sides of the main equation by $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$ to get :

$$Y_t = \mu + \psi(L)\epsilon_t \tag{3.62}$$

Where we have:

$$\psi(L) = \frac{1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q}{q - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p} \tag{3.63}$$

We would ultimately obtain the process mean as:

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \tag{3.64}$$

The mean is the same as for the $AR(p)$ process. Now we can write the $ARMA$ in terms of mean deviations to get:

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \cdots + \theta_q \epsilon_{t-q} \tag{3.65}$$

Premultiplying this equation with $(Y_{t-j} - \mu)$ we would get the covariance function as:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p} \tag{3.66}$$

A word of caution here is that the above set of covariances are true only for $j > q$. It is basically after q lags that the autocovariance function of $ARMA$ follows the same autocovariance pattern as the $AR(p)$ process. Further we note that the above autocovariance function does not hold for $j \leq q$ because of the presence of correlations between $\theta_j \epsilon_{t-j}$ and Y_{t-j}.

3.3 Invertibility

We we will define invertibility of an $MA(1)$ process. Consider the $MA(1)$ as follows:

$$Y_t - \mu = (1 + \theta L)\epsilon_t \tag{3.67}$$

We note that the white noise terms are uncorrelated and have constant variance σ^2. Provided that $|\theta| < 1$ we can multiply both sides $(1 - \theta L)^{-1}$ to obtain:

$$(1 - \theta L - \theta^2 L^2 - \cdots)(Y_t - \mu) = \epsilon_t \tag{3.68}$$

We note that the above equation can essentially be viewed as an $AR(\infty)$ representation. If a moving average process can be written as an infinite autoregressive process by simply inverting the moving average operator $(1 + \theta L)$, then the moving average process is said to be **invertible**. In a similar manner we can also define invertibility for an $MA(q)$ process as well.

$$Y_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\epsilon_t \tag{3.69}$$

Now provided that the roots of the characteristic polynomial lie outside the unit circle, the invertibility condition would be valid.

$$1 - \theta_1 z + \theta_2 z^2 + \cdots \theta_q z^q = 0 \tag{3.70}$$

Therefore we can now invert the moving average operator :

$$(1 + \eta_1 L + \eta_2 L^2 + \cdots) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)^{-1} \tag{3.71}$$

Multiplying on both sides we would get:

$$(1 + \eta_1 L + \eta_2 L^2 + \cdots)(Y_t - \mu) = \epsilon_t \tag{3.72}$$

The above equation is essentially an $AR(\infty)$ process.

Chapter 4

Key definitions and processes

4.1 Time Series

Consider that we have collected GDP data for some period as our primary sample data. As a first step the natural log transformation of the GDP value are plotted along time to see the general trend of this variable with time. Further suppose that we see an upward trend, with some fluctuations along the way. The key insight from this is that we might want to predict the trajectory of this curve beyond our sample period and for that we would need to the **statistical, stochastic mechanism** or the **data generating process** that generated these curves.

4.2 Stochastic processes

A random or stochastic process is a collection of random variables indexed with time. In terms of notation, we refer to a continuous random variable as $Y(t)$ and a discrete random variable as Y_t. Letting Y represent GDP, our data would comprise the following collection of random variables: $(Y_1, Y_2, \cdots Y_{243}, Y_{244})$. Here the subscript 1 refers to the first observation that is the GDP of the first quarter of the first year of measurement (say 1950) and the subscripy 244 refers to the last observation in our data. However an explanation is needed as to how exactly we can think of this GDP measure as a random variable - Suppose that in one particular instance the GDP for the first quarter of 1970 was $\$3,758$ billion. Now in theory we say that at that particular time this GDP figure could have taken on many possible values depending on various factors of **states of nature**. Therefore we say that the figure $\$3758$ is one particular **realization** of all such possibilities. Hence we can say that the random variable representing GDP is actually a stochastic process and the data we are looking at are the realizations of that stochastic process during the particular time period.

4.2.1 Stationary stochastic process

A stochastic process is said to be stationary if its mean and variance are constant over time and the covariance between two time periods depends only on the lag

between the time periods and not the actual value of time for which the covariance may be required. This is a classic definition of a **weakly stationary** or **covariance stationary process**. Let Y_t be a stochastic time series with the following properties:

$$E(Y_t) = \mu \tag{4.1}$$

$$Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \tag{4.2}$$

$$\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)] \tag{4.3}$$

We say that γ_k is the autocovariance of lag k, that is the covariance between Y values k periods apart. Note that if $k = 0$ then we simply get the variance of the process, for $k = 1$ we get γ_1 which is the covariance between adjacent values. Now an important point to note is that if we shift the origin of our GDP time series from Y_t to Y_{t+m} (supposing this shift represents shifting the first quarter of 1950 to say the first quarter of 1960), then we will find that if Y_t is stationary then the mean, variance and autocovariances of Y_{t+m} would be the same as that for Y_t . We can therefore state that **if a time series is stationary then it mean, variance and autocovariances at various lags are the same, regardless of what point in time we measure them - these measures are time invariant**.

Another important concept is that of **mean reversion** which essentially means that a stationary time series will always fluctuate around its mean and tend to return to it with a constant finite variance. The speed with which it reverts to the mean value depends on the strength of autocovariances. Another special type of stochastic process to make note of is the **white noise** process which is characterized typically by having a zero mean, constant finite variance and serially uncorralated terms.

4.2.2 Nonstationary stochastic process

A classic example of a nonstationary process is the **random walk model**. Suppose we have a white noise process characterized by u_t having 0 mean and variance σ^2. In this case the series Y_t would be a random walk if:

$$Y_t = Y_{t-1} + u_t \tag{4.4}$$

Thus in this model the value of Y_t is equal to the value of Y_{t-1} plus the effects of a random shock. Using continuous substitution we can write:

$$Y_1 = Y_0 + u_1 \tag{4.5}$$

$$Y_2 = Y_1 + u_2 = Y_0 + u_1 + u_2 \tag{4.6}$$

If the process started at some initial period $t = 0$ then we can the time series as:

$$Y_t = Y_0 + \sum u_t \tag{4.7}$$

We can compute the mean and variance as follows:

$$E(Y_t) = E(Y_0 + \sum u_t) = Y_0 \tag{4.8}$$

$$Var(Y_t) = t\sigma^2 \tag{4.9}$$

We can clearly notice that the variance of this series depends on time and hence violates the stationarity condition. Further we note that a RWM with no intercept is essentially a model **without drift** and here $Y_0 = 0$, therefore $E(Y_t) = 0$. An interesting feature of the RWM is the **persistence of random shocks** as is even evident from equation 7. We can see here that Y_t is the sum of initial value Y_0 plus the sum of various random shocks. We can say that the impact of a particular shock does not die out. For example if we encounter $u_2 = 2$ then every value of Y_t from Y_2 onwards will be 2 units higher persistently hence the effect of this shock does not die out. For this reason a RWM has **infinite memory**. We note that the quantity $\sum u_t$ is known as a **stochastic trend**. Now if we write the RWM equation as:

$$Y_t - Y_{t-1} = \Delta Y_t = u_t \tag{4.10}$$

With Δ being the first difference operator we notice that while the series of Y_t is nonstationary, the series of its first difference is infact stationary.

4.2.3 Random walk with drift

We can write the RWM equation as:

$$Y_t = \delta + Y_{t-1} + u_t \tag{4.11}$$

Here δ is called the **drift parameter**. We can rewrite the equation as:

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \tag{4.12}$$

This basically shows that y_t shifts upwards or downwards depending on the value of δ. Computing the mean and variance for this model:

$$E(Y_t) = Y_0 + t\delta \tag{4.13}$$

$$Var(Y_t) = t\sigma^2 \tag{4.14}$$

Even in this we can see that the variance, as well the mean, depends on time and hence violates the stationarity conditions. The RWN with and without drift are nonstationary series.

4.2.4 Unit root process

We can write the RWM as:

$$Y_t = \rho Y_{t-1} + u_t \tag{4.15}$$

Note that this model resembles an $AR(1)$ model and with $\rho = 1$ we will basically have a RWN without drift. Now in the case that $\rho = 1$ we face what is known as a **unit root problem** that makes the series nonstationary. However if we have $|\rho| < 1$ then the series would be stationary.

4.3 TS and DS processes

Firstly we note that if the time trend (evolution of series with time) is a **deterministic function** of time, then it is a deterministic trend and predictable and otherwise it known as a **stochastic trend**. Consider the following time series:

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \tag{4.16}$$

Where u_t is a white noise process and t is time measured in chronological order. With this we have the following possibilities:

- **Pure random walk** - If $\beta_1 = \beta_2 = 0$ and $\beta_3 = 1$ we will get:

  $$Y_t = Y_{t-1} + u_t \tag{4.17}$$

  This is essentially a nonstationary RWM without drift. But we note that:

  $$\Delta Y_t = Y_t - Y_{t-1} = u_t \tag{4.18}$$

  This is stationary. Hence we can say that a RWM without drift is a **difference stationary process** (DSP).

- **Random walk with drift** - If we set $\beta_1 \neq 0$, $\beta_1 = 0$ and $\beta_3 = 1$ we would get:

  $$Y_t = \beta_1 + Y_{t-1} + u_t \tag{4.19}$$

  This is essentially a RWM with drift and is nonstationary. We can write this as:

  $$Y_t - Y_{t-1} = \Delta Y_t = \beta_1 + u_t \tag{4.20}$$

  This is an example of a **stochastic trend**. This is also a DSP process since the nonstationarity in Y has been effectively eliminated by taking first differences.

- **Deterministic trend** - If we set $\beta_1 \neq 0$, $\beta_2 \neq 0$ and $\beta_3 = 0$ we would get:

  $$Y_t = \beta_1 + \beta_2 t + u_t \tag{4.21}$$

  This is an example of a **trend stationary process** (TSP). The mean of Y_t is not constant $(\beta_1 + \beta_2 t)$ but the variance is constant σ^2. If we subtract the mean from Y_t we would basically get a stationary series and hence this is called **trend stationary**. This process of removing the deterministic trend if called **detrending**.

- **Random walk with drift and deterministic trend** - If $\beta_1 \neq 0$, $\beta_2 \neq 0$ and $\beta_3 = 1$ we get:

  $$Y_t = \beta_1 + \beta_2 t + Y_{t-1} + u_t \tag{4.22}$$

  Here we have a random walk with drift and a deterministic trend which is further evident if we write the same equation as:

  $$\Delta Y_t = \beta_1 + \beta_2 t + u_t \tag{4.23}$$

- **Deterministic trend with stationary** $AR(1)$ **component** - If $\beta_1 \neq 0$, $\beta_2 \neq 0$ and $\beta_3 < 1$ then we have:

  $$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \tag{4.24}$$

  This process is stationary around its deterministic trend.

4.4 Integrated stochastic process

We note that the RWM is a special case of a more general class of stochastic processes called **integrated processes**. For a RWM without drift that is nonstationary but its first difference is stationary, we call it - RWM without drift **integrated of order 1**, denoted as $I(1)$. In this manner, if the take the first difference of the first difference, thereby making a process stationary, it is a time series **integrated of order 2**. Note the following for $I(2)$ (producing a stationary process) before moving further:

$$\Delta\Delta(Y_t) = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} - Y_{t-2} \tag{4.25}$$

In general, a time series integrated of order d is represented as $Y_t \sim I(d)$. Note that if a series is stationary it is effectively integrated with order 0 and is shown as $Y_t \sim I(0)$.

4.4.1 Properties of integrated series

Here are some important properties of integrated time series. We let X_t, Y_t, Z_t represent three time series.

- If $X_t \sim I(0)$ and $Y_t \sim I(1)$ then $Z_t = (X_t + Y_t) \sim I(1)$.
- If $X_t \sim I(d)$ then $Z_t = (a + bX_t) \sim I(d)$.
- If $X_t \sim I(d_1)$ and $Y_t \sim I(d_2)$ then $Z_t = (aX_t + bY_t) \sim I(d_2)$ where $d_1 < d_2$.
- If $X_t \sim I(d)$ and $Y_t \sim I(d)$ then $Z_t = (aX_t + bY_t) \sim I(d*)$. Note that $d*$ can be equal to d or even less than it.

4.5 Spurious regression

Consider the following random walk models:

$$Y_t = Y_{t-1} + u_t \tag{4.26}$$

$$X_t = X_{t-1} + v_t \tag{4.27}$$

Suppose we generated 500 obsevations from $u_t \sim N(0, 1)$ and 500 observations from $v_t \sim N(0, 1)$. Assume that initial values of both X_t and Y_t are zero. Further assume that u_t and v_r are serially and mutually uncorrelated. We know that both these series are nonstationary, are $I(1)$ and exhibit stochastic trends. Now if we regress Y_t on X_t we should expect an R^2 that tends to be 0 and no relation since they are fundamentally uncorrelated processes. But if by chance we obtain a result that gives a statistically significant coefficient, then that would be termed as a **spurious regression**. A good rule of thumb to identify spurious regressions is if $R^2 > d$ where d is the **Durbin Watson statistic**.

4.6 Tests of stationarity: ACF

A popular testing procedure for stationarity concerns the **autocorrelation function**. The ACF at lag k is given as:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{4.28}$$

Now note that if we plot ρ_k against k we obtain a **population correlogram**. Since we are effectively always dealing with realizations or samples of stochastic processes we tend to compute the sample autocorrelation function denoted as $\hat{\rho}_k$. In order to compute this, we first calculate the sample autocovariance $\hat{\gamma}_k$ and the sample variance $\hat{\gamma}_0$.

$$\hat{\gamma}_k = \frac{\sum(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n} \tag{4.29}$$

$$\hat{\gamma}_0 = \frac{\sum(Y_t - \bar{Y})^2}{n} \tag{4.30}$$

Here n is the sample size. Finally we can get the sample ACF at k lag as:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \tag{4.31}$$

A plot of sample ACF with k is called the **sample correlogram**. If the correlogram of a time series hovers around zero and resembles that of a purely white noise correlogram, it is probably a stationary series. On the other hand the correlogram of a nonstationary random walk series will exhibit strong correlations upto large lag lengths. In this case the autocorrelation coefficient starts at a very high value and slowly declines towards 0 as lag length increases. Next we will begin to answer some pertinent questions like - How do we choose the lag length over which to observe ACF patterns. And how to decide if the correlation coefficient at a particular lag is statistically significant or not.

4.6.1 Statistical significance of ACF

We typically judge the statistical significance of $\hat{\rho}$ by its standard error. If a time series is purely random then its sample autocorrelation function is approximately distributed as:

$$\hat{\rho}_k \sim N(0, 1/n) \tag{4.32}$$

Basically in large samples the sample autocorrelation coefficients are normally distributed with mean 0 and variance equal to the inverse of the sample size. Suppose we are given the standard deviation to be 0.0640 then we can obtain the 95% population confidence interval for ρ_k as:

$$\hat{\rho}_k \pm 1.96(0.0640) \tag{4.33}$$

$$Prob(\hat{\rho}_k \pm 1.96(0.0640) < \rho_k < \hat{\rho}_k \pm 1.96(0.0640)) = 0.95 \tag{4.34}$$

We essentially reject the null hypothesis that the true autocorrelation ρ_k is 0 is 0 is not contained in the confidence interval. Now instead of testing individual lag

autocorrelations, we can simultaneously test the joint hypothesis that all the ρ_k values upto a certain lag length are zero. We can do this using the Q statistic test of **Box and Pierce**.

$$Q = n \sum_{k=1}^{m} \hat{\rho}_k^2 \tag{4.35}$$

where n is the sample size the m is the lag length. This test is usually used to check if a given series is a white noise series or not. In large samples this Q test statistic is approximately distributed as a chi square variable with m degrees of freedom. Therefore if the computed Q exceeds the critical Q from the chi square distribution at the chosen level of significance, then we reject the null hypothesis that all the ρ_k are simultaneously zero. A variation of this test is the **Ljung Box test** statistic which is given as:

$$LN = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-K} \sim \chi^2(m) \tag{4.36}$$

Chapter 5

Testing methods

5.1 The unit root test

The unit root test is basically a test of stationarity. Let us start with the following series:

$$Y_t = \rho Y_{t-1} + u_t \tag{5.1}$$

We know that if $\rho = 1$ then the above model would become a RWM without drift which we know is a nonstationary process. Now the central idea behind the unit root test is that we can essentially regress Y_t on its lagged value Y_{t-1} and check the statistical significance of the coefficient ρ. If it is statistically equal to 1 then we conclude that the process is unit root nonstationary. Before actually performing OLS on this, we will manipulate the equation as:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t = Y_{t-1}(\rho - 1) + u_t \tag{5.2}$$

Now setting $(\rho - 1) = \delta$ we can write the above equation as:

$$\Delta Y_t = \delta Y_{t-1} + u_t \tag{5.3}$$

With this manipulation we are effectively testing the hypothesis that $\delta = 0$ against the alternative hypothesis that $\delta < 0$. If $\delta = 0$ therefore $\rho = 1$ we conclude that our process is infact unit root nonstationary. Before proceeding another interesting point to note is that if δ actually turns out to be 0 then our model would effectively collapse to:

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \tag{5.4}$$

which implies our earlier point that the first differences of a RWM is a stationary random process. Now reiterating the problem - we will first regress the first differences of Y_t on Y_{t-1} and check if the estimated slope of the regression $\hat{\delta}$ is statistically equal to 0 or not. It if it zero we conclude the series to be nonstationary and if it is negative we conclude it to be stationary. Note that we cannot use a t test for the null hypothesis $\delta = 0$ here because even in large samples the coefficient does not asymptotically resolve to a normal distribution. Now **Dickey and Fuller** showed that under the null hypothesis, t value of the coefficient of Y_{t-1} follows a **tau statistic**. This Tau statistic test is known as the **DF test**. The various possible scenarios under which this test might be applied are:

- Y_t **is a random walk**: $\Delta Y_t = \delta Y_{t-1} + u_t$.
- Y_t **is a random walk with drift**: $\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$.
- Y_t **is a RWM with drift around a deterministic trend**: $\beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$
- In all cases the null hypothesis is given by: $H_0 : \delta = 0$ - there is a unit root; time series is nonstationary; has a stochastic trend.
- Alternative hypothesis is given by: $H_1 : \delta < 0$ - time series is stationary, possibly around a deterministic trend.

Now we note that if the null hypothesis is rejected it could mean that (1) Y_t is stationary with zero mean in the first case. (2) Y_t is stationary with nonzero mean. We must note that the critical values of the Tau statistic would be different for different model specifications and hence we must try to ensure that we are not committing a specification error in modeling. Now we estimate as follows:

- Estimate either of the specified models by OLS.
- Divide the estimated coefficient of Y_{t-1} by its standard error to get the Tau statistic and refer to the DF table.
- If the computed Tau value $|\tau|$ exceeds the critical DF value then we reject the null hypothesis and conclude stationarity of the process. Alternatively if the computed value is smaller than the DF value we reject null hypothesis. As a demonstration the regression of the above three model specifications is given below:
- $\Delta \hat{GDP}_t = 0.000968 GDP_{t-1}$. Specifications: $t = (12.9)$, $R^2 = 0.0147$. $d = 1.31$.
- $\Delta GDP_t = 0.0221 - 0.00165 GDP_{t-1}$. Specifications: $t = (2.432)(-1.56)$. $R^2 = 0.0096$. $d = 1.34$.
- $\Delta GDP_t = 0.2092 + 0.0002t - 0.0269 GDP_{t-1}$. Specifications: $t = (1.89)(1.70)(-1.89)$. $R^2 = 0.0215$. $d = 1.33$.
- After this we find from the DF table that the 5% signifiance level critical values for the three models are : -1.95 (no intercept, no trend). -2.88 (intercept, no trend). -3.43 (intercept and trend).
- Now first of all we should immediately rule out model 1 because its coefficient is positive. $\delta > 0$ would imply a greater than 1 correlation which would make the model explode.
- In the other two models we can accordingly compute the Tau stat and compare with the critical values.

5.1.1 Augmented DF test

In the above DF test we assumed that the error terms u_t are uncorrelated. In case the error terms are correlated we use the **augmented Dickey Fuller test**. In this test we essentially augment in the above model equations, the lagged values of the dependent variable ΔY_t. In the ADF test we compute the following regression.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \epsilon_t \tag{5.5}$$

Here ϵ_t is a pure white noise error term and $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$, $\Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$ and so on. The point is to specifically include the lagged terms so as to ensure that the error terms are not correlated, thereby giving us unbiased estimates of δ. Note that lag length is usually selected based on **information criteria**.

5.2 Transformation: Difference stationary process

If we have a time series with unit root, then we first differences of such a series would be stationary. Therefore taking first differenes will be our goto transformation. Consider the previous GDP data example. We can take the first differences of this nonstationary GDP series to get $D_t = \Delta GDP_t = (GDP_t - GDP_{t-1})$. We can now consider the following regression:

$$\hat{\Delta D_t} = 0.00557 - 0.6781 D_{t-1} \tag{5.6}$$

$$t = (7.14)(-11.0204), R^2 = 0.33, d = 2.05 \tag{5.7}$$

Further we note that the 1% critical value of DF Tau value for this model is -3.45 and since the computed Tau value -11.02 is significantly lesser than the critical value we safely reject the null hypothesis and conclude that the first differenced series is infact stationary, or it is $I(0)$.

5.2.1 Trend stationary process

A TSP is stationary around the trend line. The simplest way to make a trend time series stationary is to regress the series on time and as a result the residuals of this regression will be stationary.

$$Y_t = \beta_1 + \beta_2 t + u_t \tag{5.8}$$

$$\hat{u}_t = (Y_t - \hat{\beta}_1 - \hat{\beta}_2 t) \tag{5.9}$$

This series will essentially be stationary and $\hat{u}_t$ is known as the **detrended time series**.

5.3 Cointegration: Regressing unit root over unit root

In earlier sections we have seen that if we regress a nonstationary time series on another nonstationary series we might obtain spurious regression results. Consider two series of LPCE and LDPI, we find that both are unit root with $I(1)$ and both contain a stochastic trend. If supposing the two series share the same common trend, then the regression results will not be spurious. Consider running the following regression:

$$LPCE_t = \beta_1 + \beta_2 LDPI_t + u_t \tag{5.10}$$

Note that the L denotes natural logarithm and β_2 represents the elasticity of real personal consumption expenditure to real disposable personal income. We can denote it as **consumption elasticity**.

$$u_t = LPCE_t - \beta_1 - \beta_2 LDPI \tag{5.11}$$

Now suppose we subject u_t to a unit root test and find that it is stationary $I(0)$. The interesting point here is that even though LPCE and LDPI are individually $I(1)$ having stochastic trends, we find that their linear combination is $I(0)$. Basically we can say that the linear combination has the effect of cancelling out the two stochastic trends in the series. Therefore in this scenario this regression of consumption on income would make sense. We say that the two variables are **cointegrated** if their linear combination gives a stationary series $I(0)$. Note that typically variables are cointegrated if they have a long term relationship between them. In short, to test cointegration we see if the residuals of the regression function are $I(0)$.

5.3.1 Testing for Cointegration

The central process in testing for cointegration is to take a typical regression of two nonstationary processes like:

$$LPCE_t = \beta_1 + \beta_2 LDPI + u_t \tag{5.12}$$

After estimating the regression, we will find the residuals and then use the DF and ADF tests of unit root testing. We note that since the estimated residuals of u_t are based on the estimated cointegrating parameter β_2, the DF and ADF critical values are not appropriate. This is because the estimated parameter is found by the original regression function and not by the OLS procedure of the residual based function. Therefore in this context the DF and ADF tests are refered to as **Engle Granger**(EG) and **augmented Engle Granger**(AEG) tests. First we run the first regression and find the coefficient estimates:

$$\hat{LPCE_t} = -0.1942 + 1.0114 LDPI_t \tag{5.13}$$

$$t = (-8.23)(348.94), R^2 = 0.998, d = 0.1558 \tag{5.14}$$

Initially it might seem that this is a spurious regression result since both the series are individually nonstationary. However we will now conduct a unit root test on the residuals to ascertain that.

$$\Delta\hat{u}_t = -0.0764\hat{u}_{t-1} \tag{5.15}$$

$$t = (-3.04), R^2 = 0.0369, d = 2.53 \tag{5.16}$$

Now we note that the EG asymptotic 5% and 10% critical values for the test statistic are -3.34 and -3.04 respectively. Therefore we can say that the residuals are nonstationary at the 5% level. We might be able to do better. Let us try modeling the same regression problem by including the trend component and then see if the residuals are stationary or not.

$$\hat{LPCE_t} = 2.813 + 0.0037t + 0.5844LDPI_t \tag{5.17}$$

$$t = (21.34)(22.93)(31.27), R^2 = 0.99, d = 0.29 \tag{5.18}$$

Now we conduct the unit root test on the residuals to get:

$$\Delta\hat{u}_t = -0.1498\hat{u}_{t-1} \tag{5.19}$$

$$t = (-4.45), R^2 = 0.075, d = 2.39 \tag{5.20}$$

Now the corresponding DF tests show that the residuals are stationary. The key catch here is that the residuals are stationary $I(0)$ but they are stationary around a deterministic trend that is linear. The residuals are $J(0)$ plus a linear trend.

Chapter 6

Forming models

6.1 AR, MA and ARIMA modeling

We can choose to model our GDP variable series as follows:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t \tag{6.1}$$

Further we note that here δ is the mean of Y_t and u_t is a white noise process with mean 0 and variance σ^2. We can say that Y_t follows a **first order autoregressive process** or is $AR(1)$. The value of Y at time t depends on its previous first lag value and a random shock. Note that the Y values are typically expressed as their mean deviations and the model is basically stating that the current value of Y is a proportion of its previous value and an error term. In a more general way, an $AR(p)$ process can be expressed as:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \cdots + \alpha_p(Y_{t-p} - \delta) + u_t \tag{6.2}$$

6.1.1 Moving average process

Suppose we choose to model Y as:

$$Y_t = \mu + \beta_1 u_t + \beta_2 u_{t-1} \tag{6.3}$$

Here μ is a constant and u is a stochastic error term modeled as white noise. We say that Y at time t is equal to a constant plus a moving average of the white noise error terms. The above is an $MA(1)$ process. Similarly, an $MA(q)$ process can be represented as:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q} \tag{6.4}$$

6.1.2 ARMA process

If Y has characteristics of both MA and AR then it can be modeled as an $ARMA$ process. Here is an example of $ARMA(1,1)$:

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_1 u_t + \beta_2 u_{t-1} \tag{6.5}$$

6.1.3 ARIMA process

We know from earlier discussions that for a series to be weakly stationary they have constant mean and variance and their covariances are time invariant. Further we also learned that some nonstationary series are integrated with order $I(1)$ then their first difference is stationary $I(0)$. Generally, if we take a d order difference of an integrated $I(d)$ series then we would get a stationary $I(0)$ series. Now if we have a series for which we have to take d differences and then we apply a stationary $ARMA(p, q)$ model to it, then we say that the original series is $ARIMA(p, d, q)$ or an **autoregressive integrated moving average series**. Therefore an $ARIMA(2, 1, 2)$ series can essentially be differenced once and then be modeled as an $ARMA(2, 2)$ process. We further note that an $ARIMA(p, 0, q) = ARMA(p, q)$, $ARIMA(p, 0, 0) = AR(p)$ and $ARIMA(0, 0, q) = MA(q)$.

6.2 Box Jenkins methodology

The key question for us to answer is that if we are looking at some time series data like the GDP example we have mentioned so frequently, then how would we know if it is a purely $AR(p)$ process, a purely $MA(q)$ process, an $ARMA(p, q)$ process or an $ARIMA(p, d, q)$ and correspondingly how can be figure out the p, d and q values appropriate to model the data. For this reason we use the **Box Jenkins method**.

- **identification**: To find the appropriate values of p, d and q we will see how a **correlogram** and **partial correlogram** can help.
- **Estimation**: After figuring out the appropriate p and q values to form the model, we need to estimate the parameters included in the model which are the AR and MA terms.
- **Diagnostic checking**: Having chosen a particular $ARIMA$ model we will then check if its fits the data reasonable well. If not we might attempt to find another $ARIMA$ model for the same. A simple test is to see if the residuals of the model are white noise or not. If they are then our model is fine. If not, then we need to make it again.
- **Forecasting**: We will then use the model to forecast values.

6.3 Identification

The main tools used in the identification process are the ACF, PACF and their corresponding correlograms which are nothing but plots of ACF/PACF with the corresponding lag lengths. Now the concept of partial autocorrelation is analogous to the concept of a partial regression coefficient in multivariate regression analysis. For example the k^{th} coefficient β_k measures the rate change in the mean value of the regressand for a unit increase in the value of the k^{th} regressor, while holding the effect of all other regressors constant. Similarly, the **partial autocorrelation** ρ_{kk} measures the correlation between observations that are k time periods

apart after controlling for autocorrelations in the intermediate time lags (less than k). For example the correlation exhibited between Y_t and Y_{t-k} might be due to the correlation they exhibit with intermediate lags $Y_{t-1}, Y_{t-2}, \cdots$. The partial correlation ρ_{kk} basically removes their influence.

From the standard figures of ACF and PACF correlograms in the GDP series we will notice that the ACF declines very slowly and the ACF values upto 22 lag values are statistically significantly different from zero, that is they all fall outside the 95% confidence interval bounds. On the other hand, the PACF falls dramatically after the second lag and PACF values after lag 2 are statistically insignificant. Now we must note that this GDP series is nonstationary and before applying the Box Jenkins approach we must convert it to a stationary series. What we have seen before as well is that if we take the first differences in GDP we do not observe any time trends and also after conducting **DF unit root tests** we were able to ascertain that the first differenced series is stationary.

We see that after this the ACF and PACF plots look very different. In this particular example we have the ACF at lags $1, 2, 5$ that are statistically significant and all others are insignificant. In the PACF, we have only the lags $1, 12$ that are statistically significant. Now to see if our data can be modeled as an $ARMA, MA$ or AR process, we need to compare the ACF and PACF plots of standard models of $AR(1), MA(1), ARMA(1,1), MA(2), ARIMA(2,2)$ and so on. All these series have typically characteristic patterns in the ACF and PACF and if our data ACF and PACF patterns resemble any of those, we will select that as our modeling approach. We will then apply diagnostic checks to see if the chosen $ARMA$ model is accurate or not. to do this we keep some ground rules handy.

- $AR(p)$ - ACF Decays exponentially or with a damped sine wave or both. PACF has significant spikes through lags p
- $MA(q)$ - ACF has significant spikes through lags q. PACF delines exponentially.
- $ARMA(p,q)$ - ACF has exponential decay. PACF has exponential decay.

6.4 Estimation

Let $Y*$ denote the first differenced logged GDP figures and let us assume that our data exhibits an $MA(2)$ pattern and so we model it that way.

$$y*_t = \mu + \beta_1 u_{t-1} + \beta_2 u_{t-2} \tag{6.6}$$

We might obtain the following estimates from the data.

$$\hat{Y*}_t = 0.0082 + 0.291 u_{t-1} + 0.202 u_{t-2} \tag{6.7}$$

$$t = (9.32)(4.61)(3.20), R^2 = 0.12, d = 1.97 \tag{6.8}$$

However upon doing some diagnostic checking including checking the ACF, PACF plots (and checking for individual significance) and the Ljung Box statistic to check for collective significance of autocorrelation terms, we may find that we cannot reject the null hypothesis and hence we may need to look for another $ARIMA$ model. Note that forecasting the level value rather than the first differenced value we can seperate out the model as :

$$Y_t - Y_{t-1} = \mu + \beta_1 u_{t-1} + \beta_2 u_{t-2} \tag{6.9}$$

References

[1] Hamilton - Time series analysis

[2] Damodar Gujarati, Dawn Porter - Basic Econometrics

[3] Fumio Hayashi - Econometrics