Chapter 6: Time Series Theory

6.0 Introduction

The time series developed in this chapter are based on the assumption that the series to be forecasted has been generated by a stochastic process. That is we assume that each value $Y_t = \{y_1, y_2, ..., y_n\}$ in the series is drawn randomly from a probability distribution

6.1 Stationary processes

An important class of stochastic processes are those which are stationary.

Strict stationarity
A time series X_t is **strictly stationary** if the joint distribution of $\{Y_1, Y_2 \dots , \ Y_n\}$ is the same as the joint distribution $\{Y_{1+k}, Y_{2+k}, \dots , Y_{n+k}\}$ for all $t = 1,2, \dots , n$. In other words, shifting the time origin by an amount k has no effect on the joint distributions, which must therefore depend only on the intervals between $t_1, t_2 \dots , t_n$. This means that the joint distribution of any moments of any degree (e.g. expected values, variances, third order and higher moments) within the process is constant, i.e. $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$ which do not depend on time.

Note: Such a specification is rather impractical. This definition is too strict to be used for any real-life model. In practice it is often useful to define stationarity in a less restricted way than that described above.

Weak stationarity
In practice it is often useful to define stationarity in a less restricted way than that described above. A process is called second-order stationary (or weakly stationary) if the mean does not depend on t (constant). Instead, we consider first and second order moments of the joint distributions, i.e.,

$$E(Y_t) \quad \text{and} \quad E(Y_{t+k}, Y_t) \quad \text{for} \quad t = 1,2, ..., \ k = 0,1,2, ...$$

and examine properties of the time series which depend on these.

Therefore, a time series $\{X_t\}$ is called **weakly stationary** or just stationary if
1. $E(Y_t) = \mu Y_t = \mu < \infty$, that is the expectation of Y_t is finite and does not depend on t, and
2. $\gamma_{(Y_{t+k}, Y_t)} = \gamma_k$ that is for each k the autocovariance of random variables (Y_{t+k}, Y_t) does not depend on t (it is constant for a given lag k).

6.2 Backshift and differencing operators

A main reason for using a stationary data sequence instead of a non-stationary sequence in time series is that non-stationary sequences, usually, are more complex and take more calculations when forecasting is applied to a data series. To convert a non-stationary time series into stationary, we can apply a difference operator to the data series.

Backshift operator
In time series analysis, the backshift operator (also called the "lag operator"). It shifts a time series back so that the shifted time series lags one-time unit behind, i.e., given some time series Y_t, then

$$BY_t = Y_{t-1}$$

Example:

t	Y_t	BY_t
⋮	⋮	⋮
2013	5.4	⋮
2014	5.5	5.4
2015	5.8	5.5
2016	⋮	5.8
⋮	⋮	⋮

For the two applications of B to Y_t shifts the data back two periods, as follows:

$$B(BY_t) = B^2 Y_t = Y_{t-2}$$

In general,

$$B^k Y_t = Y_{t-k}$$

Example:
For monthly data, if we wish to shift attention to "the same month last year," then B^{12} is used, and $B^{12} Y_t = Y_{t-12}$.

Note:
A (finite or infinite order) polynomial can be defined in B or a filter according to:

$$a(B) = a_0 + a_1 B + a_2 B^2 + \cdots$$

So, one benefit of introducing the backshift operator is that it provides us with a compact notation for writing filters. Second, the algebra of polynomials can be applied to filters. This turns out to provide us with a very useful way to study and manipulate the behaviour of covariance stationary processes.

Differencing
For 1st difference, $Y_t' = Y_t - Y_{t-1} = (1-B)Y_t$

2nd order difference, $Y_t'' = Y_t' - Y_{t-1}'$

$$= Y_t - 2Y_{t-1} + Y_{t-2}$$
$$= (1 - 2B + B^2)Y_t$$
$$= (1-B)^2 Y_t$$

In general, a d^{th} order difference can be written as $(1-B)^d Y_t$.

A seasonal difference followed by a first difference can be written as $(1-B)(1-B^s)Y_t$.
The terms can be multiplied together to see the combined effect. For example

$$(1-B)(1-B^s)Y_t = (1 - B - B^s + B^{s+1})Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$$

6.3 The Autocorrelation Function (ACF)

We have already noted in the last chapter that the sample autocorrelation coefficients of an observed time series are an important set of statistics for describing the time series. Similarly, the (theoretical) ACF of a stationary stochastic process is an important tool for assessing its properties.

Properties of ACF

1. The ACF is an even function of the lag in that $\rho(k) = \rho(-k)$
2. $|\rho(k)| \leq 1$
3. Non-uniqueness

6.4 White noise model

Consider the model $Y_t = \mu + \varepsilon_t$
where μ is the overall mean and $\varepsilon_t \sim NID(0, \sigma^2)$.

A **purely random process** is often called a "*white noise*" model, a terminology which comes from engineering.

The white noise model is fundamental to many techniques in time series analysis. Any good forecasting model should have forecast errors which follow a white noise model.

6.5 The Partial Autocorrelation Coefficient ($PACF$)

In regression analysis, if the forecast variable Y is regressed on explanatory variables X_1 and X_2, then it might be of interest to ask how much explanatory power does X_1 have if the effects of X_2 are somehow *partialled out* first. Typically, this means regressing Y on X_2, getting the residual errors from this analysis, and finding the correlation of the residuals with X_1. In time series analysis there is a similar concept.

Partial autocorrelations are used to measure the degree of association between Y_t and Y_{t-k}, when the effects of other time lags – $1, 2, 3, ..., k-1$ are removed.

The partial autocorrelation at lag k is

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} r_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j}\, r_j} & \text{if } k = 2, 3, ... \end{cases}$$

where $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$ for $j = 1, 2, 3, ..., k-1$

As with the ACF, the partial autocorrelations should all be close to zero for a white noise series.

6.6 A random walk model

In a random walk process each successive change in Y_t is drawn independently from a probability distribution with 0 mean. Thus

$$Y_t - Y_{t-1} = \varepsilon_t \text{ or } Y_t = Y_{t-1} + \varepsilon_t$$

where $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$.

$$V(\varepsilon_t) = \sigma_\varepsilon^2 .$$

Note:
Random walk model is widely used for non-stationary data. Random walks typically have long periods of apparent trends up or down which can suddenly change direction unpredictably.

A point forecast made in time period t for Y_{t+l} is

$$\hat{y}_{t+l} = E(y_{t+l} \mid y_t, \ldots, y_1) = y_t$$

For one period ahead, the forecast error is

$$e_1 = y_{t+1} - \hat{y}_{t+1} = (y_t + \varepsilon_{t+1}) - (y_t) = \varepsilon_{t+1}$$

Thus $V(e_1) = V(\varepsilon_{t+1}) = \sigma_\varepsilon^2$

For two period forecast,

$$\begin{aligned} e_2 &= y_{t+2} - \hat{y}_{t+2} \\ &= (y_{t+1} + \varepsilon_{t+2}) - (y_t) \\ &= (y_t + \varepsilon_{t+1} + \varepsilon_{t+2}) - (y_t) \\ &= \varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

So, $V(e_2) = V(\varepsilon_{t+1} + \varepsilon_{t+2}) = 2\sigma_\varepsilon^2$

Similarly, for the $l-$period forecast,
$V(e_l) = l\sigma_\varepsilon^2$

6.7 Linear processes

A time series Y_t is said to be a **linear process** if it has the representation

$$Y_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

where ε_t is white noise. Y_t is stationary if and only if $\sum_{j=-\infty}^{\infty} |a_j| < \infty$. In practice, the general linear model is useful when the a_j are expressible in terms of a finite number of parameters which can be estimated.

The linear process can be expressed using backward shift operator, B, i.e.

$$Y_t = a(B)\varepsilon_t$$

where

$$a(B) = \sum_{j=-\infty}^{\infty} a_j B^j$$

The operator $a(B)$ is a linear filter, which when applied to a non-stationary process produces a stationary process.

Note:
Moving average (MA) and autoregressive (AR) processes are special cases of the linear process. For example,

1. White noise: $a_0 = 1$
2. $MA(q)$: $a_0 = 1$, $a_1 = \theta_1$, …, $a_j = \theta_q$
3. $AR(p)$: $a_0 = 1$, $a_1 = \phi$, $a_2 = \phi^2$, …, $a_j = \phi^p$

6.8 Probability models for time series

The three popular linear time series models are

1. Autoregressive (AR) model
$$Y_t = \sum_{p=1}^{n} \phi_p Y_{t-p} + \delta + \varepsilon_t = \delta + \sum_{p=0}^{n} \phi^p \varepsilon_{t-p}$$
2. Moving average (MA) model
$$Y_t = \sum_{q=0}^{\infty} \theta_q \varepsilon_{t-q} \quad \text{where} \quad \theta_0 = 1$$
3. Autoregressive moving average ($ARMA$) model
$$Y_t = \delta + \sum_{p=1}^{n} \phi_p Y_{t-p} + \sum_{q=0}^{m} \theta_q \varepsilon_{t-q}$$

6.8.1 Autoregressive Models

In the autoregressive process of order p, the current observation Y_t is generated by a weighted average of past observations going back p periods, together with a random error in the current period. We denote this process as $AR(p)$ and write its equation as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \delta + \varepsilon_t$$

where δ is a constant term which relates to the mean of the stochastic process.

If the autoregressive process is *stationary,* then its mean must be invariant with respect to time; i.e. $E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = \ldots E(Y_{t-p}) = \mu$

Thus, $\quad \mu = \phi_1 \mu + \phi_2 \mu + \ldots + \phi_p \mu + \delta$

or

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} \quad \text{and} \quad \phi_1 + \phi_2 + \ldots + \phi_p < 1$$

Note:

If the process is stationary, the mean μ must be finite (i.e. $\phi_1 + \phi_2 + \ldots + \phi_p < 1$)

Example:

$AR(1)$, $\quad Y_t = \phi_1 Y_{t-1} + \delta + \varepsilon_t$

The process has mean $\mu = \dfrac{\delta}{1 - \phi_1}$ and is stationary if $|\phi_1| < 1$

Let $\delta = 0$ (to scale the process to one that has zero mean, $\mu = 0$),

$$\gamma_0 = E[(Y_t - \mu)^2]$$

$$\gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)] =$$

$$\gamma_2 = E[(Y_t - \mu)(Y_{t-2} - \mu)]$$

Similarly, the covariance for a k-lag displacement is

$$\gamma_k = \phi_1^k \gamma_0 = \frac{\phi_1^k \sigma_\varepsilon^2}{1-\phi_1^2}$$

The autocorrelation function for $AR(1)$ is

$$\rho_0 = 1, \rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k .$$

An example of $AR(1)$ would be $y_t = 0.9y_{t-1} + 2 + \varepsilon_t$.

<u>Example:</u>
$AR(2)$ $\quad\quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \delta + \varepsilon_t$

The process has mean $\mu = \dfrac{\delta}{1-\phi_1-\phi_2}$
and a necessary condition for stationary is that $\phi_1 + \phi_2 < 1$.

Let $\delta = 0$, so $\mu = 0$, variances and covariances of Y_t (when Y_t measured in deviations form).

$$\gamma_0 = E[Y_t(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t)] = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2$$

$$\gamma_1 = E[Y_{t-1}(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t)] = \phi_1\gamma_0 + \phi_2\gamma_1, \ \gamma_1 = \frac{\phi_1\gamma_0}{1-\phi_2} .$$

$$\gamma_2 = E[Y_{t-2}(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t)] = \phi_1\gamma_1 + \phi_2\gamma_0$$

The three equations above can be solve simultaneously to get γ_0 in terms of ϕ_1, ϕ_2, and σ_ε^2, and hence

$$\gamma_0 = \frac{(1-\phi_2)\sigma_\varepsilon^2}{(1+\phi_2)\left[(1-\phi_2)^2 - \phi_1^2\right]}$$

In general, for $k \geq 2$, $\gamma_k = E[Y_{t-k}(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t)] = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}$.

The autocorrelation functions:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1}{1-\phi_2}; \quad\quad \rho_2 = \frac{\gamma_2}{\gamma_0} = \phi_2 + \frac{\phi_1^2}{1-\phi_2}; \ \rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \text{ for } k \geq 2 .$$

An example of $AR(2)$ would be $y_t = 0.9y_{t-1} - 0.7y_{t-2} + 2 + \varepsilon_t$

6.8.2 <u>Moving Average Models</u>

In a moving average process of order q each observation Y_t is generated by a weighted average of random error terms going back q periods. A dependence relationship is set up among the successive error terms and we denote this process as $MA(q)$ and write its equation as

$$Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$$

where $\theta_1, \cdots, \theta_q$ may be positive or negative.

Note:
1. It is a convention to include the minus sign on the coefficients θ_k.
2. The mean of the MA process is independent of time, since $E(X_t) = \mu$.
3. The error term, ε_t is assumed to be independently distributed across time, i.e., generated by a white noise process, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ for $k \neq 0$. (i.e. $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_t \varepsilon_{t-k}) = 0$.)

$$Var(Y_t) = \gamma_0 = E[(Y_t - \mu)^2]$$

$$Cov(Y_t, Y_{t-k}) = \gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)] \text{ for } k = 1, 2, ..., q$$

Note:
$Cov(Y_t, Y_{t-k}) = 0$ for $k > q$. The process $MA(q)$ has a memory of only q periods; any value y_t is correlated with y_{t-q} and with y_{t+q} but with no other time-series values. In effect, the process forgets what happened more than q periods in the past. In general, the limited memory of a MA process is important. It suggests that a MA model provides forecasting information only to a limited number of periods into the future.

The autocorrelation function (ACF) is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2} & k = 1, ..., q \\ 0 & k > q \end{cases}$$

Note:
The ACF can be useful in specifying the order of a MA process. The ACF, ρ_k, for the $MA(q)$ process has q nonzero values and is 0 for $k > q$.

Example:
$MA(1)$, $\quad Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$

$$\gamma_0 = E[(Y_t - \mu)^2]$$

$$\gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)] =$$

For $k > 1$, $\quad \gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)] =$

The autocorrelation function is given by:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1,}{1 + \theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

An example of a first-order moving average process might be given by $y_t = 2 + \varepsilon_t + 0.8\varepsilon_{t-1}$.

Note:
The process $MA(1)$ has a memory of exactly one period, so that the value of x_t is influenced only by events that took place in the current period and one period back.

$MA(2)$, $\quad Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$

$$\gamma_0 = \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2)$$
$$\gamma_1 = E[(Y_t-\mu)(Y_{t-1}-\mu)] = E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2} - \theta_2\varepsilon_{t-3})]$$
$$= -\theta_1(1-\theta_2)\sigma_\varepsilon^2$$
$$\gamma_2 = E[(Y_t-\mu)(Y_{t-2}-\mu)] = E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-2} - \theta_1\varepsilon_{t-3} - \theta_2\varepsilon_{t-4})]$$
$$= -\theta_2\sigma_\varepsilon^2$$

For $k > 2$, $\quad \gamma_k = E[(Y_t-\mu)(Y_{t-k}-\mu)]$
$$= E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \theta_2\varepsilon_{t-k-2})]$$
$$= 0$$

The autocorrelation function is given by:
$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}; \qquad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}; \qquad \rho_k = 0 \text{ for } k > 2.$$

An example of a second-order moving average process might be given by
$$y_t = 2 + \varepsilon_t + 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$$

Note:
The process $MA(2)$ has a memory of exactly two periods, so that the value of x_t is influenced only by events that took place in the current period, one period back, and two periods back.

6.8.3 Mixed Autoregressive-Moving Average Models

Many stationary random processes cannot be modeled as purely moving average or purely autoregressive, since they have the qualities of both types of processes. The logical extension is to mixed the $AR(p)$ and $MA(q)$ models and known as autoregressive-moving average process of order (p, q), $ARMA(p, q)$.
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - -\theta_q\varepsilon_{t-q}$$

If the process is assumed to be stationary, so that its mean is constant over time and is given by
$$\mu = \phi_1\mu + \phi_2\mu + \ldots + \phi_p\mu + \delta \qquad \text{or}$$
$$\mu = \frac{\delta}{1-\phi_1-\phi_2-\ldots-\phi_p} \quad \text{and } \phi_1+\phi_2+\ldots+\phi_p < 1$$

Example
$ARMA(1,1)$, $\quad Y_t = \phi_1 Y_{t-1} + \delta + \varepsilon_t - \theta_1\varepsilon_{t-1}$

Setting $\delta = 0$, so $\mu = 0$,, the variances and covariances of this process are shown as follows:
$$\gamma_0 = E[Y_t(\phi_1 Y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1})]$$
$$= E[(\phi_1 Y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1})^2]$$
$$= \phi_1^2\gamma_0 - 2\phi_1\theta_1 E[Y_{t-1}\varepsilon_{t-1}] + \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 \qquad \text{(note: } \because \text{stationarity, so } E(Y_{t-1}^2) = \gamma_0\text{)}$$
$$= \phi_1^2\gamma_0 - 2\phi_1\theta_1\sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 \qquad \text{(note: } E[Y_{t-1}\varepsilon_{t-1}] = \sigma_\varepsilon^2,\ E[Y_t\varepsilon_{t+k}] = 0, k \neq 0\text{)}$$

Hence $\gamma_0(1-\phi_1^2)=\sigma_\varepsilon^2(1+\theta_1^2-2\phi_1\theta_1)$

$$\gamma_0=\frac{\sigma_\varepsilon^2(1+\theta_1^2-2\phi_1\theta_1)}{1-\phi_1^2} \qquad \text{and } |\phi_1|<1.$$

We can determine the covariance $\gamma_1, \gamma_2, \ldots$ recursively:

$$\begin{aligned}\gamma_1 &= E[Y_{t-1}(\phi_1 Y_{t-1}+\varepsilon_t-\theta_1\varepsilon_{t-1})] \\ &= \phi_1\gamma_0-\theta_1\sigma_\varepsilon^2 \\ &= \frac{\sigma_\varepsilon^2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\end{aligned}$$

$$\gamma_2 = E[Y_{t-2}(\phi_1 Y_{t-1}+\varepsilon_t-\theta_1\varepsilon_{t-1})]=\phi_1\gamma_1 \qquad (\text{note: } E[Y_{t-2}Y_{t-1}]=\gamma_1)$$

and similarly, $\gamma_k=\phi_1\gamma_{k-1}$ for $k\ge 2$.

The autocorrelation function, is then given by

$$\rho_1=\frac{\gamma_1}{\gamma_0}=\frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$$

and $\quad \rho_k=\phi_1\rho_{k-1}$ for $k\ge 2$

For higher order processes, i.e., the general $ARMA(p,q)$ process, the variance, covariances, and autocorrelation function are solutions to difference equation that usually cannot be solved by inspection. It can be shown easily, however, that

$$\gamma_k=\phi_1\gamma_{k-1}+\phi_2\gamma_{k-2}+\ldots+\phi_p\gamma_{k-p} \text{ and } \rho_k=\phi_1\rho_{k-1}+\phi_2\rho_{k-2}+\ldots+\phi_p\rho_{k-p} \text{ for } k\ge q+1.$$

<u>Note:</u>
q is the memory of the MA part of the process, so that for $k\ge q+1$ the ACF (and covariances) exhibits the properties of a purely autoregressive process.

An example of $ARMA(1,1)$ would be $y_t=0.8y_{t-1}+2+\varepsilon_t-0.9\varepsilon_{t-1}$.

6.8.4 *ARIMA* Models

In practice, many of the time series we work with are non-stationary, so that the characteristics of the underlying stochastic process change over time. We can construct the models for those non-stationary series which can be transformed into stationary series by differencing them one or more times. We say that Y_t is homogeneous non-stationary of order d if

$$W_t=\Delta^d Y_t$$

is a stationary series. Here Δ denotes differencing, i.e.,

$$\Delta Y_t=Y_t-Y_{t-1}$$

and

$$\begin{aligned}\Delta^2 Y_t &= \Delta Y_t-\Delta Y_{t-1} \\ &= Y_t-Y_{t-1}-(Y_{t-1}-Y_{t-2}) \\ &= Y_t-2Y_{t-1}+Y_{t-2} \\ &= (1-2B+B^2)Y_t \\ &= (1-B)^2 Y_t\end{aligned}$$

and so forth.

After we have differenced the series Y_t to produce the stationary series W_t, we can model W_t as an $ARMA$ process. If $W_t = \Delta^d Y_t$ and W_t is an $ARMA(p, q)$ process, then we say that Y_t is an integrated autoregressive-moving average process of order (p, d, q) or $ARIMA(p, d, q)$. We can write the equation for the process $ARIMA(p, d, q)$ using backward shift operator, as

$$\phi(B)\Delta^d Y_t = \delta + \theta(B)\varepsilon_t$$

with $\quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \quad$ (AR operator)

and $\quad \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^p \quad$ (MA operator)

The equation for the simplest case, $ARIMA(1, 1, 1)$ is as follow:

$$\underbrace{(1 - \phi_1 B)}_{AR(1)}\underbrace{(1 - B)}_{\text{First Difference}}Y_t = \delta + \underbrace{(1 - \theta_1 B)}_{MA(1)}\varepsilon_t.$$

In practice, it is seldom necessary to deal with values p, d or q that are other than 0, 1, or 2.

Seasonality and ***ARIMA*** **models**

In exactly the same way that *consecutive* data points might exhibit AR, MA, mixed $ARMA$, or mixed $ARIMA$ properties, so data separated by a whole season (i.e., a year) may exhibit the same properties. The $ARIMA$ notation can be extended readily to handle seasonal aspects, and the general shorthand notation is

$$ARIMA\ \underbrace{(p, d, q)}_{\text{(Non-seasonal part of the model)}}\underbrace{(P, D, Q)_s}_{\text{(Seasonal part of the model)}}$$

where s = number of periods per season.

For illustrative purposes consider the following general $ARIMA$ (1, 1, 1)(1, 1, 1)$_4$ model:

$$\underbrace{(1 - \phi_1 B)}_{\text{Non-seasonal AR(1)}}\ \underbrace{(1 - \Phi_1 B^4)}_{\text{seasonal AR(1)}}\underbrace{(1 - B)}_{\text{Non-seasonal difference}}\ \underbrace{(1 - B^4)}_{\text{seasonal difference}}\ Y_t = \underbrace{(1 - \theta_1 B)}_{\text{Non-seasonal MA(1)}}\ \underbrace{(1 - \Theta_1 B^4)}_{\text{seasonal MA(1)}}\varepsilon_t.$$

All the factors can be multiplied out and the general model written as follows:

$$\begin{aligned}Y_t &= (1+\phi_1)Y_{t-1} - \phi_1 Y_{t-2} + (1+\Phi_1)Y_{t-4} - (1+\phi_1+\Phi_1+\phi_1\Phi_1)Y_{t-5} \\ &\quad - (\phi_1+\phi_1\Phi_1)Y_{t-6} - \Phi_1 Y_{t-8} + (\Phi_1+\phi_1\Phi_1)Y_{t-9} - \phi_1\Phi_1 Y_{t-10} \\ &\quad + \varepsilon_t - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-4} + \theta_1\Theta_1\varepsilon_{t-5}\end{aligned}$$

In this form, once the coefficients θ_1, ϕ_1, Φ_1, Θ_1 have been estimated from the data, equation above can be used for forecasting.