Forecasting

Simon Shaw
s.c.shaw@maths.bath.ac.uk

2005/06 Semester II

1 Introduction

A critical aspect of managing any business is planning for the future. Predicting future events is called **forecasting**.

> e.g. A company wishes to forecast future sales so that it can assess how many items to produce. The company may use the judgement of experts (qualitative method) or have an extensive amount of historical sales records and use these to make the forecast (quantitative method).

We shall concern ourselves with the latter scenario. Quantitative methods use historical data to identify a pattern, fit a model to the pattern and then use that model to forecast points in the future.

- There is an assumption that the identified pattern will continue into the future.
- The measurements in the historical data are correlated.
- The model will typically lose accuracy the longer into the future we try to predict and so, in general, the shorter the time frame, the more accurate the forecast.

We focus attention upon TIME SERIES MODELS. Time series models are based on the analysis of a chronological sequence of observations on a particular variable. The level of these observations may thus be plotted against time creating the so-called TIMEPLOT.

> e.g. Economic indices such as the FTSE 100; weekly sales figures; patient's temperature; monthly recorded level of water in a reservoir.

1.1 The components of a time series

A time series may be thought of as consisting of several components. The usual assumption is that four separate components: **trend; cyclical; seasonal and irregular** combine to provide the observed level of the series.

1.1.1 Trend

This is the discernable change in the general level, e.g. a steady growth or decline in values over a noticeable period , which is not attributable to the other components described below. Trends do not generally remain constant and may frequently reverse.

> e.g. Sales of DVD players. This might exhibit an upwards trend as they are introduced into the market at ever cheaper prices. However, there may be a downward trend once most families possess a player.

1.1.2 Seasonality

Many business and economic series are taken quarterly or monthly and exhibit seasonality - a periodic repetition of level changes (typically one year).

> e.g. Ice cream sales, bookings at holiday resorts.

1.1.3 Cyclical Variation

Periodic variation but is not connected with the season. Time series exhibits an alternating sequence of points below and above the trend line.

> e.g. A time series for house prices might have a generally increasing trend line but the boom-bust nature of economic markets may cause prices to alternate above and below this trend line. This was a very typical occurrence in the 1970s and 1980s.

1.1.4 Irregular variation

The random, unpredictable variation that cannot be explained by any of the above effects.

We shall assume that our measurements are made at discrete time points and that X_t denotes our measurement for the quantity X at time t and $\{X_t\}$ denotes the series of measurements on X. $\hat{X}_t$ denotes our forecast for X at time t.

There are many ways that the components described above might combine. For example, we might have an **additive seasonal model with trend**

$$X_t = M + T_t + S_t + I_t$$

or a **multiplicative seasonal model with trend**

$$X_t = (M + T_t)S_t + I_t$$

where M represents the general level of the series and T_t, S_t, and I_t the trend, seasonal and irregular components at time t.

2 Exponential smoothing

Exponential smoothing is a very simple technique which can be of use in its own right and is very widely used in generalised forms.

Suppose that the time series is nonseasonal and has no significant upward or downward trend. In this case, we can think of the model being of the form

$$X_t = \beta_0 + I_t$$

where β_0 represents the general level and I_t the random error. Our object is to estimate the current level (note that we do not observe the level but X_t which includes the random error). Exponential smoothing achieves this by using a weighted average of all of the observations.

Let M_t denote the estimate of the level at time t where

$$M_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \cdots \tag{1}$$

where $0 < \alpha < 1$ is the smoothing constant. At time $t-1$, the estimate of the level is

$$M_{t-1} = \alpha X_{t-1} + \alpha(1-\alpha)X_{t-2} + \alpha(1-\alpha)^2 X_{t-3} + \cdots . \tag{2}$$

Thus, noting that equation (1) may be expressed as

$$M_t \;=\; \alpha X_t + (1-\alpha)\{\alpha X_{t-1} + \alpha(1-\alpha)X_{t-2} + \alpha(1-\alpha)^2 X_{t-3} + \cdots\},$$

substituting (2) into (1) gives

$$M_t = \alpha X_t + (1-\alpha)M_{t-1}. \tag{3}$$

Our new estimate for the level, M_t is a weighted average of our actual data, X_t and our old estimate for the level, M_{t-1}. Starting with an initial estimate of the level, $M_1 = X_1$ say, we may then repeatedly use equation (3) to obtain M_2, $M_3, \ldots$.

Let $\hat{X}_{t+n}$ denote the forecast at time $t+n$ made using the observations $\{X_1, X_2, \ldots, X_t\}$. $\hat{X}_{t+n}$ is called the n-step ahead forecast from X_t. For exponential smoothing, all future forecasts are the same, and equal to the current estimate of the level, that is

$$\hat{X}_{t+n} \;=\; M_t, \text{ for all } n = 1, 2, \ldots$$

In particular, noting that the one-step ahead forecast from X_{t-1}, $\hat{X}_{(t-1)+1}$, is equal to M_{t-1} and M_t is the one-step ahead forecast from X_t, we may express equation (3) as

$$\hat{X}_{t+1} \;=\; \alpha X_t + (1-\alpha)\hat{X}_t. \tag{4}$$

The forecast error in period t is the difference between the actual value of X and the forecast value for X,

$$\text{Forecast error in period } t \;=\; X_t - \hat{X}_t.$$

We may thus reexpress equation (4) as

$$\hat{X}_{t+1} \;=\; \hat{X}_t + \alpha(X_t - \hat{X}_t). \tag{5}$$

Equation (5) thus reveals that the new forecast $\hat{X}_{t+1}$ is equal to the previous forecast $\hat{X}_t$ plus an adjustment which is α times the most recent forecast error.

The objective of smoothing methods is to "smooth out" the random fluctuations caused by the irregular component of the time series (as we observe X_t which includes the irregular component, I_t). If the irregular component in the series is LARGE then a SMALL value of α, the smoothing constant, is preferred as much of the forecast error is due to random variability. If the irregular component is SMALL then a LARGE value of α is preferred and we quickly adjust the forecast when forecasting errors occur.

2.1 Criterion for determining α

We can assess the effectiveness of the forecast by looking at the MEAN SQUARE DEVIATION,

$$MSD \;=\; \text{Average of the sum of squared errors.}$$

So, if our observations are X_1, $X_2, \ldots, X_T$ then

$$MSD \;=\; \frac{1}{T-1}\sum_{t=2}^{T}(X_t - \hat{X}_t)^2.$$

A small MSD suggests a good fit to our data. We could choose α to minimise MSD.

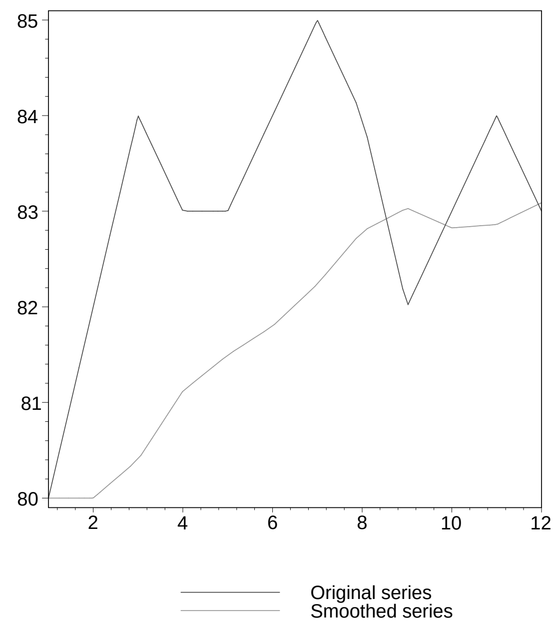

Figure 1: The time series and smoothed series for the shipments.

2.2 Worked example

The data below show the monthly percentage of all shipments that were received on time over the past 12 months.

80 82 84 83 83 84 85 84 82 83 84 83

1. Construct an exponential forecast using $\alpha = 0.2$ and calculate the forecast error for each month. Take $\hat{X}_2 = M_1 = 80$.

Month, t	X_t	$\hat{X}_t = M_{t-1} = \alpha X_{t-1} + (1-\alpha)M_{t-2}$	$X_t - \hat{X}_t$
1	80		
2	82	$M_1 = 80$	$82 - 80 = 2$
3	84	$M_2 = 0.2(82) + 0.8(80) = 80.4$	$84 - 80.4 = 3.6$
4	83	$M_3 = 0.2(84) + 0.8(80.4) = 81.12$	$83 - 81.12 = 1.88$
5	83	$M_4 = 0.2(83) + 0.8(81.12) = 81.496$	$83 - 81.496 = 1.504$
6	84	$M_5 = 0.2(83) + 0.8(81.496) = 81.7968$	$84 - 81.7968 = 2.2032$
7	85	$M_6 = 82.23744$	2.76256
8	84	$M_7 = 82.789952$	1.210048
9	82	$M_8 = 83.0319616$	-1.0319616
10	83	$M_9 = 82.82556928$	0.17443072
11	84	$M_{10} = 82.86045542$	1.13954458
12	83	$M_{11} = 83.08836434$	-0.08836434

The original time series and smoothed series are shown in Figure 1. Notice that the initial starting point was low relative to the remainder of the series and it took some time for the smoothed series to reach what appears to be the underlying level of the series (the mean of the twelve observations is 83.08333).

2. What is the forecast for future months?

$$\hat{X}_{12+n} = M_{12} = 0.2(83) + 0.8(83.08836434) = 83.07069147.$$

3. What is the value of the mean square deviation?

$$\begin{aligned} MSD &= \{2^2 + 3.6^2 + 1.88^2 + 1.504^2 + 2.2032^2 + 2.76256^2 + 1.210048^2 + (-1.0319616)^2 + \\ &\quad 0.17443072^2 + 1.13954458^2 + (-0.08836434)^2\}/11 \\ &= \{4 + 12.96 + 3.5344 + 2.262016 + 4.85409024 + 7.631737754 + 1.464216162 + \\ &\quad 1.064944744 + 0.030426076 + 1.29856185 + 0.007808257\}/11 \\ &= 39.10820108/11 \\ &= 3.55529108. \end{aligned}$$

3 The Holt-Winters forecasting method

So far, we have been considering a model of the form $X_t = \beta_0 + I_t$ where I_t represents random error and β_0 the general level. Thus, the fluctuation in the time series around β_0 comes only from the random error I_t and we try to smooth the series out using M_t as a smoothed estimate for the level at time period t and α the (constant) weighting factor. Our basic equation, see equation (3), was

$$\text{estimate} = (\text{constant})(\text{actual data}) + (1 - \text{constant})(\text{old estimate}) \tag{6}$$

and the exponential smoothing forecast of X_{t+n} based on t observations $X_1, \ldots, X_t$ is $\hat{X}_{t+n} = M_t$, $n = 1, 2, 3, \ldots$. We now exploit this smoothing technique for more complicated models. We shall focus upon two types of model: the first displaying a linear trend and the second displaying a linear trend with a multiplicative seasonal effect.

3.1 Holt's linear trend method

Suppose that the series is nonseasonal but does display trend. Model the trend as linear and write

$$X_t = \beta_0 + \beta_1 t + I_t \tag{7}$$

where β_0 represents the intercept and β_1 the slope. Consider the series at time $t + n$. From equation (7) we have

$$\begin{aligned} X_{t+n} &= \beta_0 + \beta_1(t+n) + I_{t+n} \\ &= (\beta_0 + \beta_1 t) + \beta_1 n + I_{t+n}. \end{aligned} \tag{8}$$

Equation (8) shows that we may express the time series at time $t + n$ as the level at time t plus n times the slope of the linear trend plus the irregular variation. Let M_t and T_t denote the smoothed estimates for the level and slope at time period t. Thus, M_t may be viewed as an estimate of $\beta_0 + \beta_1 t$ and T_t of the estimated slope of the trend line as of time t (i.e. of β_1).

Suppose that we wish to obtain the n-step ahead forecast from time t, so using observations $X_1, \ldots, X_t$. Thus, in equation (8), X_{t+n} is unknown and we may replace it by its forecast $\hat{X}_{t+n}$. M_t is our forecast for $\beta_0 + \beta_1 t$ and T_t our forecast for β_1. The irregular component I_{t+n} may be forecasted by its expectation which is zero. Thus, our n-step ahead forecast from X_t, assuming the level and trend remain at the current values, is

$$\hat{X}_{t+n} \;=\; M_t + T_t n, \;\; n = 1, 2, 3, \ldots$$

We use exponential smoothing to estimate M_t and T_t. Using the general principle of exponential smoothing, see equation (6), we note that the old estimate for the level is given by the one-step ahead forecast from X_{t-1}, namely $\hat{X}_{(t-1)+1} = M_{t-1} + T_{t-1}$. Thus,

$$M_t \;=\; \alpha X_t + (1-\alpha)(M_{t-1} + T_{t-1}) \tag{9}$$

where $0 < \alpha < 1$ is the smoothing constant. Notice that a current observation of the slope may be obtained by subtracting the current and previous estimates of the level. Thus, our smoothed values for T_t are given by

$$T_t \;=\; \gamma(M_t - M_{t-1}) + (1-\gamma)T_{t-1} \tag{10}$$

where $0 < \gamma < 1$ is the smoothing constant. In order to solve the recursions given in equations (9) and (10) we need initial conditions. A natural choice is to set $T_2 = X_2 - X_1$, so that we may view T_t as a trend representing the difference between the current and previous level, and $M_2 = X_2$.

3.2 Holt-Winters multiplicative model

We now extend Holt's linear trend method to take into account seasonality. There are two versions: when the seasonal effect is additive and when it is multiplicative. We shall study the more widely used of these two variants, the multiplicative version.

The basic idea is that a time series shows both trend and seasonality. The seasonal pattern is amplified by the level of the series. Modelling the trend as linear we write

$$X_t \;=\; sn_t(\beta_0 + \beta_1 t) + I_t,$$

where we assume that there are s periods in the seasonal component sn_t. i.e. s periods in a "year" or season.

e.g. $s = 4$ for quarterly data; $s = 12$ for monthly data.

Thus, for forecasting, we will MULTIPLY the updated trend by the updated seasonal factor. To deseasonalise the data, we DIVIDE the appropriate seasonal factor into the data.

Let M_t, T_t, S_t denote the smoothed estimates for the level, slope and season at time period t. Once again, we utilise the general principle for exponential smoothing given by equation (6). The actual deseasonalised data for time period t may be obtained by dividing by the seasonal effect for the corresponding time s periods ago. Thus,

$$M_t = \alpha \frac{X_t}{S_{t-s}} + (1-\alpha)(M_{t-1} + T_{t-1}) \tag{11}$$

where $0 < \alpha < 1$ denotes the smoothing constant. For T_t we have

$$T_t \;=\; \gamma(M_t - M_{t-1}) + (1-\gamma)T_{t-1} \tag{12}$$

where $0 < \gamma < 1$ is the smoothing constant. For the seasonal effect, note that a measure of the actual seasonal variation in the data may be obtained by dividing the new estimate of the level into the actual data point. Our old smoothed estimate of the seasonal component is that obtained at the time period $t - s$. Thus, we use

$$S_t = \delta \frac{X_t}{M_t} + (1-\delta) S_{t-s} \tag{13}$$

where $0 < \delta < 1$ is the smoothing constant. After the level, slope and seasonal estimates have been smoothed then our n-step ahead forecast from X_t is given by:

$$\hat{X}_{t+n} = \begin{cases} (M_t + nT_t) S_{t+n-s} & \text{for } n = 1, 2, \ldots, s \\ (M_t + nT_t) S_{t+n-2s} & \text{for } n = s+1, s+2, \ldots, 2s \end{cases} \quad \text{and so on,}$$

since

$$X_{t+n} = sn_{t+n}\{\beta_0 + \beta_1(t+n)\} + I_{t+n} = sn_{t+n}\{(\beta_0 + \beta_1 t) + \beta_1 n\} + I_{t+n}$$

so that we replace X_{t+n} by its forecast value, $\hat{X}_{t+n}$, estimate $(\beta_0 + \beta_1 t)$ by M_t and β_1 by T_t. I_{t+n} is estimated by its expectation which is zero. For the seasonal component, sn_{t+n} we estimate it by the equivalent seasonal component prior to time t. So, for $n \leq s$, this is S_{t+n-s}, for $s+1 \leq n \leq 2s$ this is S_{t+n-2s} and so on for all possible values of n.

Note that in order to solve the recursions given by equations (11), (12) and (13) we need to give starting values. There are several ways in which we could obtain starting values which depend upon the seasonality of the data. We shall not go into these.

3.2.1 Worked example

The table below shows the quarterly figures for US Retail Sales (1984 - 1987) of a product. Using a season of $s = 4$, a Holt-Winters multiplicative model with $\alpha = 0.11$, $\gamma = 0.01$ and $\delta = 0.01$ was fitted. $\hat{X}_t$ denotes the one-step ahead forecast for X_t.

t	X_t	M_t	T_t	S_t	$\hat{X}_t$
1				0.9040	
2				1.0150	
3				1.0050	
4		3085.02	48.79	1.0750	
5	2881	3139.65	48.85	0.9041	2832.96
6	3249	3189.87	48.86	1.0150	3236.33
7	3180	3230.54	48.78	1.0048	3254.93
8	3505	3277.24	48.76	1.0749	3525.26
9	3020	3327.56	48.78	0.9042	3007.16
10	3449	3378.71	48.80	1.0151	3427.10
11	3472	3430.58	48.83	1.0049	3443.94
12	3715	3476.84	48.80	1.0749	3740.18
13	3184	3525.18	48.80	0.9042	3187.78
14	3576	3568.35	48.74	1.0150	3627.92
15	3657	3619.54	48.77	1.0049	3634.70
16	3941	3668.10	48.77	1.0749	3942.99
17	3319	3711.80	48.71	0.9041	3360.65
18	3850	3764.11	48.75	1.0150	3816.79
19	3883	3818.49	48.81	1.0050	3831.63
20	4159	3867.51	48.81	1.0749	4156.86

1. Explicitly calculate $\hat{X}_5$.

$$\begin{aligned} \hat{X}_5 = \hat{X}_{4+1} &= (M_4 + T_4)S_1 \\ &= (3085.02 + 48.79)0.904 \\ &= 2832.96. \end{aligned}$$

2. Show how M_5, T_5, S_5 were calculated.

$$\begin{aligned} M_5 &= \alpha\frac{X_5}{S_1} + (1-\alpha)(M_4 + T_4) \\ &= 0.11\left(\frac{2881}{0.904}\right) + 0.89(3085.02 + 48.79) \\ &= 3139.65; \\ T_5 &= \gamma(M_5 - M_4) + (1-\gamma)T_4 \\ &= 0.01(3139.65 - 3085.02) + 0.99(48.79) \\ &= 48.85; \\ S_5 &= \delta\frac{X_5}{M_5} + (1-\delta)S_1 \\ &= 0.01\left(\frac{2881}{3139.65}\right) + 0.99(0.904) \\ &= 0.9041. \end{aligned}$$

3. How were M_{20}, T_{20}, S_{20} calculated?

$$\begin{aligned} M_{20} &= \alpha\frac{X_{20}}{S_{16}} + (1-\alpha)(M_{19} + T_{19}) \\ &= 0.11\left(\frac{4159}{1.0749}\right) + 0.89(3818.49 + 48.81) \\ &= 3867.51; \\ T_{20} &= \gamma(M_{20} - M_{19}) + (1-\gamma)T_{19} \\ &= 0.01(3867.51 - 3818.49) + 0.99(48.81) \\ &= 48.81; \\ S_{20} &= \delta\frac{X_{20}}{M_{20}} + (1-\delta)S_{16} \\ &= 0.01\left(\frac{4159}{3867.51}\right) + 0.99(1.0749) \\ &= 1.0749. \end{aligned}$$

4. Find the three-step and five-step ahead forecasts from time $t = 20$.

$$\begin{aligned} \hat{X}_{23} &= (M_{20} + 3T_{20})S_{19} \\ &= (3867.51 + 3(48.81))1.0050 \ = \ 4034.01; \\ \hat{X}_{25} &= (M_{20} + 5T_{20})S_{17} \\ &= (3867.51 + 5(48.81))0.9041 \ = \ 3717.26. \end{aligned}$$

4 Box-Jenkins methodology

This is an iterative model building procedure.

1. Identify a possible model from a general class of linear models.
2. Estimate the parameters of the model.
3. Diagnostic checking: check the model against historical data to see if it accurately describes the underlying process that generates the series.

   The model fits well if the differences between the original data and the forecasts are small, independent and random.
4. If the model doesn't fit well, the process is repeated using another model designed to improve the original one.

The process is carried out on a time series that is STATIONARY

> i.e. the distribution of X_t does not depend on t, implying that the series has a constant mean, μ, a constant variance, σ^2, and a constant autocorrelation function, $\rho(k) = Corr(X_t, X_{t+k})$.

The models employed are called AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) models. We will describe the constituent parts of these models.

4.1 Autoregressive models and autocorrelation

In an autoregressive model, values of the time series are regressed on one or more previous values. We might expect to find a pattern of correlation between values of X_t and X_{t-k} for various values of k, the **autocorrelation at lag** k.

> e.g. Suppose that the correlation between adjacent values is some number ϕ_1, say; that between values two periods apart is ϕ_1^2; that between values three periods is ϕ_1^3, ...,
>
> $$\text{i.e. } Corr(X_t, X_{t-k}) \;=\; \phi_1^k \text{ for } k = 1, 2, 3, \ldots$$
>
> One model giving rise to such an autocorrelation structure is
>
> $$X_t \;=\; \gamma + \phi_1 X_{t-1} + a_t \qquad (14)$$
>
> where γ and ϕ_1 are fixed and $\{a_t\} = \{a_1, a_2, \ldots\}$ is a sequence of uncorrelated random variables with zero mean and constant variance. The model given by (14) is termed a **first order autoregressive model**, written $AR(1)$, as we use one past observation.

Definition 1 *(Autoregressive model of order p)*
The autoregressive model of order p, $AR(p)$, is written

$$X_t \;=\; \gamma + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t \qquad (15)$$

where γ, $\phi_1, \ldots, \phi_p$ are fixed and $\{a_t\} = \{a_1, a_2, \ldots\}$ is a sequence of uncorrelated random variables with zero mean and constant variance. γ is the constant term.

4.2 Moving average models

In a moving average model, the present observation is a linear function, or weighted average, of present and past errors. The number of past error terms used is known as the order.

Definition 2 *(Moving average model of order q)*
The moving average model of order q, $MA(q)$, is written

$$X_t \;=\; \gamma + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \tag{16}$$

where γ, $\theta_1, \ldots, \theta_q$ are fixed and $\{a_t\} = \{a_1, a_2, \ldots\}$ is a sequence of uncorrelated random variables with zero mean and constant variance. γ is the constant term.

Notice that X_t and X_{t+k} are independent if $k > q$ as, in this case,

$$X_{t+k} \;=\; \gamma + a_{t+k} + \theta_1 a_{t+k-1} + \cdots + \theta_q a_{t+k-q}$$

shares no common error terms with those in X_t and hence $\rho(k) = Corr(X_t, X_{t+k}) = 0$.

4.3 $ARMA(p, q)$ models

In an ARMA model, X_t is a combination of an autoregressive and a moving average model, that is, it is a linear function of past observations and present and past forecasting errors.

Definition 3 *($ARMA(p, q)$)*
The $ARMA(p, q)$ model is a mixture of an $AR(p)$ model and a $MA(q)$ model and is written

$$X_t = \gamma + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \tag{17}$$

where γ, $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$ are fixed and $\{a_t\} = \{a_1, a_2, \ldots\}$ is a sequence of uncorrelated random variables with zero mean and constant variance. γ is the constant term.

Note that a $MA(q)$ model is often (equivalently) written as

$$X_t \;=\; \gamma + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

and similarly for an $ARMA(p, q)$ model.

4.4 Achieving stationarity through differencing

The Box-Jenkins methodology assumes that the time series it fits models to are stationary. Unfortunately, most time series are not stationary.

e.g. Series with trend and seasonality are not stationary.

There are many ways to transform series to stationarity. Box-Jenkins use a technique known as **differencing**. Careful use of differencing operators can induce stationarity.

Definition 4 *(Difference)*
The (backward) difference is a finite difference defined by $\nabla_k X_t = X_t - X_{t-k}$. Most commonly,

$$\nabla X_t \;=\; \nabla_1 X_t \;=\; X_t - X_{t-1}.$$

Finite differencing is the discrete analogue to differentiation. Some examples:

1. $\nabla_{12} X_t = X_t - X_{t-12}$.

2. $\nabla_4 X_t = X_t - X_{t-4}$.

3. $\nabla_{12}^2 X_t = \nabla_{12}(\nabla_{12} X_t) = \nabla_{12}(X_t - X_{t-12}) = X_t - 2X_{t-12} + X_{t-24}$.

4. If $X_t = \beta_0 + \beta_1 t + I_t$, so a linear trend, show that ∇X_t is a stationary process.

$$\begin{aligned}
\nabla X_t &= X_t - X_{t-1} \\
&= \{\beta_0 + \beta_1 t + I_t\} - \{\beta_0 + \beta_1(t-1) + I_{t-1}\} \\
&= \beta_1 + I_t - I_{t-1}.
\end{aligned}$$

As the errors I_t and I_{t-1} are assumed to be random with zero mean and constant variance and β_1 is constant, then ∇X_t does not depend upon t and is thus stationary.

∇ denotes the first difference, ∇^2 the second difference, ∇^3 the third difference and so on.

Definition 5 *($ARIMA(p,d,q)$)*
Let X_t be a time series. Suppose that after d simple differencing passes have been made, the series is stationary and a suitable $ARMA(p,q)$ model is fitted. The process is then integrated, that is the differencing operations are reversed. The resulting model is termed an autoregressive integrated moving average model, $ARIMA(p,d,q)$.

An $ARIMA(p,d,q)$ model thus consists of an $AR(p)$ model, d differences and a $MA(q)$ model. A couple of examples:

1. $ARIMA(p,1,q)$. Then $W_t = \nabla X_t$ is stationary and the $ARMA(p,q)$ model is fitted to W_t. That is,

$$W_t = \gamma + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}.$$

2. Suppose that, after first differencing the series $\{X_s, s = 1, 2, \ldots, t\}$ to obtain the series $\{\nabla X_s\}$, you decide to fit the $MA(q)$ model with no constant term, that is

$$\nabla X_s = a_s + \theta_1 a_{s-1} + \theta_2 a_{s-2} + \ldots + \theta_q a_{s-q}.$$

What $ARIMA(p,d,q)$ model has been fitted to $\{X_s\}$?

We have fitted the $ARMA(0,q)$ model to the first difference so we have fitted the $ARIMA(0,1,q)$ model to $\{X_s\}$.

4.5 Forecasting

This is achieved by using the estimated values of the parameters and the residuals from the fit. We use the $ARMA$ equation we have fitted which will involve X_ts and a_ts as well as the estimates of γ, the ϕs and the θs. Suppose we have data up to time t and we wish to forecast ahead from there.

- Use the known values of $\{X_s : s \leq t\}$ when we have them and forecast values $\{\hat{X}_{t+n} : n = 1, 2, \ldots\}$ when we do not.
- Use the residual $\hat{a}_s = X_s - \hat{X}_s$ for $s \leq t$ if a_s appears in the forecast equation.
- Set to zero (its expected value) any a_{t+n} for $n \geq 0$ which occurs in the forecast equation.
- Replace γ, $\{\phi\}$, $\{\theta\}$ by their estimates $\hat{\gamma}$, $\{\hat{\phi}\}$, $\{\hat{\theta}\}$ in the forecast equation.
- Forecasting ahead one step ahead at a time means that everything is available by the time it is needed.

4.5.1 Worked example

Consider the following $ARIMA(0,2,2)$ model,

$$\nabla^2 X_t \quad = \quad a_t - 0.9a_{t-1} + 0.5a_{t-2}.$$

1. Express the model in the form $X_t =$.

   We note that $\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$ so that

   $$X_t - 2X_{t-1} + X_{t-2} \;=\; \nabla^2 X_t \;=\; a_t - 0.9a_{t-1} + 0.5a_{t-2}$$

   whence $X_t = 2X_{t-1} - X_{t-2} + a_t - 0.9a_{t-1} + 0.5a_{t-2}$.

2. Using the usual rules, write down the forecasts $\hat{X}_{t+1}$, $\hat{X}_{t+2}$, $\hat{X}_{t+3}$.

   (a) One-step ahead forecast.

   For X_{t+1}, we have

   $$X_{t+1} \quad = \quad 2X_t - X_{t-1} + a_{t+1} - 0.9a_t + 0.5a_{t-1}.$$

   X_t, X_{t-1} are available but X_{t+1} is replaced by its forecast value $\hat{X}_{t+1}$. a_{t+1} is replaced by its expected value 0 and a_t, a_{t-1} by the their respective residuals from the fit, $\hat{a}_t$, $\hat{a}_{t-1}$. The one-step ahead forecast is

   $$\hat{X}_{t+1} \quad = \quad 2X_t - X_{t-1} - 0.9\hat{a}_t + 0.5\hat{a}_{t-1}.$$

   (b) Two-step ahead forecast.

   We consider the model for X_{t+2}. We find

   $$X_{t+2} \quad = \quad 2X_{t+1} - X_t + a_{t+2} - 0.9a_{t+1} + 0.5a_t.$$

   X_t is available but X_{t+2}, X_{t+1} are replaced by their respective forecast values $\hat{X}_{t+2}$, $\hat{X}_{t+1}$. a_{t+2}, a_{t+1} are replaced by their expected values: 0 and a_t by its residual from the fit, $\hat{a}_t$. The two-step ahead forecast is

   $$\hat{X}_{t+2} \quad = \quad 2\hat{X}_{t+1} - X_t + 0.5\hat{a}_t.$$

   (c) Three-step ahead forecast.

   For the three-step ahead forecast we have

   $$X_{t+3} \quad = \quad 2X_{t+2} - X_{t+1} + a_{t+3} - 0.9a_{t+2} + 0.5a_{t+1}.$$

   X_{t+3}, X_{t+2}, X_{t+1} are replaced by their respective forecast values $\hat{X}_{t+3}$, $\hat{X}_{t+2}$, $\hat{X}_{t+1}$. a_{t+3}, a_{t+2}, a_{t+1} are replaced by their expected values: 0. The three-step ahead forecast is

   $$\hat{X}_{t+3} \quad = \quad 2\hat{X}_{t+2} - \hat{X}_{t+1}.$$

   Notice that all future forecasts will have this form, that is for $k \geq 3$ the k-step ahead forecast from time t is given by

   $$\hat{X}_{t+k} \quad = \quad 2\hat{X}_{t+k-1} - \hat{X}_{t+k-2}.$$

4.6 Box-Jenkins model identification (not in the exam)

As we outlined before, there are four basic steps to the Box-Jenkins methodology.

1. Identification of a model.
2. Estimation of parameters in the model.
3. Diagnostic checking of the model.
4. Forecasting.

In the previous subsection, we have discussed how to forecast from a given $ARIMA(p, d, q)$ model. To determine the appropriate model to fit, it is necessary to study analyse the behaviour of the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

The **autocorrelation coefficient** measures the correlation between a set of observations and a lag set of observations in a time series. Given the time series $\{X_s : s = 1, \ldots, t\}$ the autocorrelation between X_s and X_{s+k}, $\rho(k)$ measures the correlation between the pairs (X_1, X_{1+k}), $(X_2, X_{2+k}), \ldots, (X_{t-k}, X_t)$. Note that the series is assumed to be stationary so that the correlation depends only on k, how far apart the observations are, and not on t. An estimate of $\rho(k)$ is the sample autocorrelation coefficient

$$r_k \quad = \quad \frac{\sum_s (X_s - \overline{X})(X_{s+k} - \overline{X})}{\sum_s (X_s - \overline{X})^2},$$

where X_s is the data from the stationary time series, X_{s+k} the data k time periods ahead and $\overline{X}$ the mean of the stationary time series.

When the sample autocorrelation coefficients are computed for lag 1, lag 2, lag 3, ... and graphed r_k versus k, the resultant plot is called the **sample autocorrelation function (ACF)** or **CORRELOGRAM**.

The **partial autocorrelation coefficient** (ρ_{kk}) is a measure of the relationship between X_s and X_{s+k} when the effect of the intervening variables $X_{s+1}, X_{s+2}, \ldots, X_{s+k-1}$ has been removed. This adjustment is made to see if the correlation between X_s and X_{s+k} is due to the intervening variables or something else. We may estimate the sample partial autocorrelation coefficient as r_{kk}. A plot of r_{kk} versus k is known as the **autocorrelation coefficient function (PACF)**.

The ACT and PACF may be used to aide the identification of p and q in the $ARIMA$ model. The following table provides a brief summary.

Model	ACF	$PACT$
$MA(q)$	Cuts off after lag q (Typically, $q = 1, 2$)	Dies down exponentially and/or sinusoidally
$AR(p)$	Dies down exponentially and/or sinusoidally	Cuts off after lag p (Typically, $p = 1, 2$)
$ARMA(p, q)$	Dies down exponentially and/or sinusoidally	Dies down exponentially and/or sinusoidally

Having identified a model, estimated the parameters (for example, by least squares fitting to the observed data) and checked the residuals to examine the fit, if we are happy we go ahead and forecast.