Forecasting Techniques

13

"Those who have knowledge don't predict. Those Who Predict Don't have Knowledge".

—Lao Tzu

LEARNING OBJECTIVES

LO 13-1 Understand the importance of forecasting and its impact on the effectiveness of the supply chain and overall performance of an organization.

LO 13-2 Learn various components of time-series data such as trend, seasonality, cyclical component, and random component.

LO 13-3 Learn different techniques such as moving average, exponential smoothing, and Croston's method.

LO 13-4 Learn Auto-Regression (AR), Moving Average (MA), and Auto-Regressive Integrated Moving Average models (ARIMA).

LO 13-5 Learn practical challenges associated with forecasting models using case studies.

IMPORTANCE OF FORECASTING

Forecasting is one of the most important and frequently addressed problems in analytics. Inaccurate forecasting can have significant impact on both top line and bottom line of an organization. For example, non-availability of product in the market can result in customer dissatisfaction, whereas, too much inventory can erode the organization's profit. Thus, it becomes necessary to forecast the demand for a product and service as accurately as possible.

13.1 | INTRODUCTION TO FORECASTING

Forecasting is by far the most important and frequently used application of predictive analytics since it has significant impact on both top line and bottom line of an organization. Every organization prepares long-range and short-range planning for the organization and forecasting demand for product and service is an important input for both long-range and short-range planning. Different capacity planning problems such as manpower planning, machine capacity, warehouse capacity, materials requirements planning (MRP) will depend on the forecasted demand for the product/service. Budget allocation for marketing promotions and advertisement are usually made based on forecasted demand for the product. Forecasting can be very challenging due to several factors that influence the demand and scale of business with stock keeping units (SKUs) running into several millions. For example:

1. Boeing 747-400 has more than 6 million parts and several thousand unique parts (Hill, 2011). Forecasting demand for spare parts is important since non-availability of mission critical parts can result in aircraft on ground (AOG) which can be very expensive for airlines.
2. Amazon.com sells more than 350 million products through its E-commerce portal. Amazon itself sells about 13 million SKUs and has more (about 2 million) retailers selling their products through Amazon (Ali, 2017). Predicting demand for these products is important since overstocking can impact the bottom line and under stocking can result in customer dissatisfaction. Amazon.com may not stock all SKUs they sell through their portal since most of them are sold by their suppliers (online marketplace) directly to the customers, but even if they have to predict demand for products directly sold by them, then the number of SKUs is 13 million.
3. Walmart sells more than 142,000 products through their supercenters (*source:* Walmart website[1]). Being a brick-and-mortar retail store, Walmart does not have the advantages of Amazon.com (being also a market place, Amazon do not have to predict demand for all the products sold through their portal). They have to maintain stock for each and every product sold by Walmart and predict demand for the products as accurately as possible.
4. Demand for products and service is not the only application of forecasting, even manpower planning requires the use of sophisticated models. Indian information technology (IT) companies struggle to manage the right level of manpower for each skill required to manage their business. This would involve forecasting business opportunities, skills required to manage current and future projects, and so on.
5. Many products may have intermittent demands, that is, gap between two demands can be long and the gap itself may be random. The modeler has to forecast the next instance of demand and the actual demand quantity when demand occurs, making it much more difficult to forecast.
6. One of the innovative applications of forecasting was the Netflix forecasting contest in which the participants were challenged to forecast the movie rating (on a scale of 1 to 5) likely to be given by a customer for a movie. An accurate customer movie rating forecast can further be used for movie recommendations to customers.

13.2 | TIME-SERIES DATA AND COMPONENTS OF TIME-SERIES DATA

Time-series data is a data on a response variable, Y_t, such as demand for a spare parts of a capital equipment or a product or a service or market share of a brand observed at different time points t. Whenever we have a forecasting problem, we will be using a time-series data. The variable Y_t is a random variable. The data points or measurements are usually collected at regular intervals and are arranged in chronological order. If the time-series data contains observations of just a single variable (such as demand of a product at time t), then it is termed as univariate time series. If the data consists of more than one variable, for example, demand for a product at time t, price at time t, amount of money spent by the company on promotion at time t, competitors price at time t, etc. then it is called multivariate time-series data.

[1] Source: http://corporate. walmart. com/_news_/news-archive/2005/01/07/our-retail-divisions

From a forecasting perspective, a time-series data can be broken into the following components [Figures 13.1(a)–(d)]:

1. **Trend Component (T_t):** Trend is the consistent long-term upward or downward movement of the data over a period of time.
2. **Seasonal Component (S_t):** Seasonal component (measured using seasonality index) is the repetitive upward or downward movement (or fluctuations) from the trend that occurs within a calendar year such as seasons, quarters, months, days of the week, etc. The upward or downward fluctuation may be caused due to festivals, customs within a society, school holidays, business practices within the market such as '*end of season sale*', and so on. For example, in India demand for many products surge during the festival months of October and November. A similar pattern exists during December in many countries due to Christmas. Usually, for a given context seasonal fluctuation occurs at fixed intervals (such as months, quarters) known as periodicity of seasonal variation and repeats over time.
3. **Cyclical Component (C_t):** Cyclical component is fluctuation around the trend line that happens due to macro-economic changes such as recession, unemployment, etc. Cyclical fluctuations have repetitive patterns with a time between repetitions of more than a year. Whereas in case of seasonality, the fluctuations are observed within a calendar year and are driven by factors such as festivals and customs that exist in a society. A major difference between seasonal fluctuation and cyclical fluctuation is that seasonal fluctuation occurs at fixed period within a calendar year, whereas cyclical fluctuations have random time between fluctuations. That is, periodicity of seasonal fluctuations is constant, whereas the periodicity of cyclical fluctuations is not constant.
4. **Irregular Component (I_t):** Irregular component is the white noise or random uncorrelated changes that follow a normal distribution with mean value of 0 and constant variance.

The time-series data can be modelled as an addition of the above components or product of the above components. The additive time-series model is given by

$$Y_t = T_t + S_t + C_t + I_t \tag{13.1}$$

The additive models assume that the seasonal and cyclical components are independent of the trend component. Additive models are not very common since in many cases the seasonal component may not be independent of trend. At the Indian Institute of Management Bangalore (IIMB) there are many weekend programs and the number of students enrolled in these programs is fixed. The demand for food at the canteens of IIMB increases by a fixed quantity on Saturdays. This increase for demand is additive in nature.

The multiplicative time-series model is given by

$$Y_t = T_t \times S_t \times C_t \times I_t \tag{13.2}$$

Multiplicative models are more common and are a better fit for many data sets. In many cases, we will use the form $Y_t = T_t \times S_t$ which is simpler form of Eq. (13.2). To estimate the cyclical component we will need a large data set.

The additive model is appropriate if the seasonal component remains constant about the level (or mean) and does not vary with the level of the series, while the multiplicative model is more appropriate if seasonal variation is correlated with level (local mean).

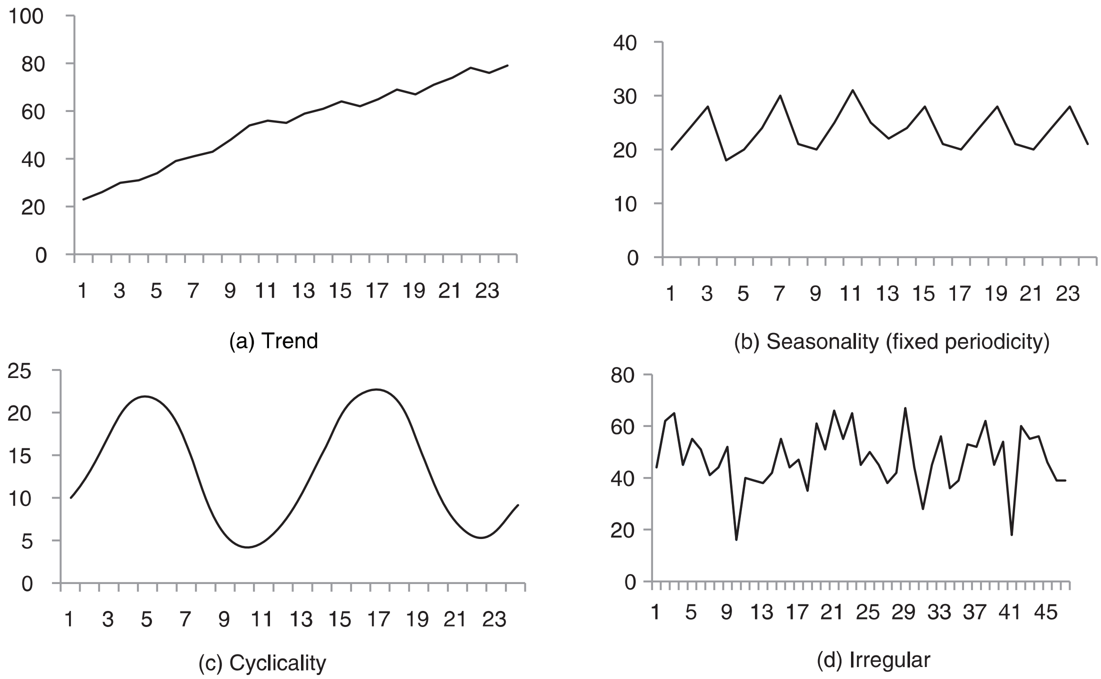


FIGURE 13.1 Trend in time-series data.

13.3 | FORECASTING TECHNIQUES AND FORECASTING ACCURACY

There are many forecasting techniques developed based on different logics. Simple techniques such as moving average and exponential smoothing predict the future value of a time-series data as a function of the past observations. Whereas the regression-based models such as auto-regressive (AR), moving average (MA), auto-regressive and moving average (ARMA), auto-regressive integrated moving average (ARIMA), and auto-regressive integrated moving average with X (ARIMAX) use more sophisticated regression models to forecast the future value of a time-series data. It is important to note that using complex mathematical models does not guarantee more accurate forecast. Simple moving average technique may outperform complex ARIMA models in few cases. In fact, in an editorial in the International Journal of Forecasting, Chatfield (1986) claimed that simple forecasting models sometimes outperform complex models.

Usually, many different forecasting techniques such as moving average, exponential smoothing, and ARIMA are used for forecasting before selecting the best model. The model selection may depend on the chosen forecasting accuracy measure. The following four forecasting accuracy measures are frequently used:

1. Mean absolute error
2. Mean absolute percentage error
3. Mean squared error
4. Root mean square error

In the following subsections, we will discuss these measures.

13.3.1 | Mean Absolute Error (MAE)

Mean absolute error (MAE) is the average absolute error and should be calculated on the validation data set. Assume that the validation data has n observations and forecasting is carried out on these n observations using the model developed. The mean absolute error is given by

$$MAE = \sum_{t=1}^{n} \frac{|Y_t - F_t|}{n} \tag{13.3}$$

In Eq. (13.3), Y_t is the actual value of Y at time t and F_t is the corresponding forecasted value.

13.3.2 | Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error (MAPE) is the average of absolute percentage error. Assume that the validation data has n observations and the forecasting is carried out on these n observations. The mean absolute percentage error is given by

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y_t - F_t|}{Y_t} \times 100\% \tag{13.4}$$

MAPE defined in Eq. (13.4) is one of the popular forecasting accuracy measures used by practitioners since it expresses the average error in percentage terms and is easy to interpret. Since MAPE is dimensionless it can be used for comparing different models with varying scales.

13.3.3 | Mean Square Error (MSE)

Mean square error is the average of squared error calculated over the validation data set. MSE is given by

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2 \tag{13.5}$$

Lower MSE implies better prediction. However, it depends on the range of the time-series data.

13.3.4 | Root Mean Square Error (RMSE)

Root mean square error (RMSE) is the square root of mean square error and is given by

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2} \tag{13.6}$$

RMSE along with MAPE are two most popular accuracy measures of forecasting. RMSE is the standard deviation of errors or residuals. In 2006, Netflix, the movie portal, announced a competition with a prize money worth one million dollars to predict the rating on a 5-point scale likely to be given a customer for a movie[2] (*source*: Wikipedia). The participants were given a target RMSE of 0.8572 to qualify for the prize.

[2] https://en. wikipedia. org/wiki/Netflix_Prize

13.4 | MOVING AVERAGE METHOD

Moving average is one of the simplest forecasting techniques which forecasts the future value of a time-series data using average (or weighted average) of the past 'N' observations. Mathematically, a simple moving average is calculated using the formula

$$F_{t+1} = \frac{1}{N}\sum_{k=t+1-N}^{t} Y_k \tag{13.7}$$

The above formula is called simple moving average (SMA) since 'N' past observations are given equal weights ($1/N$). In a weighted moving average, past observations are given differential weights (usually the weight decrease as the data becomes older). Weighted moving average is given by

$$F_{t+1} = \sum_{k=t+1-N}^{t} W_k \times Y_k \tag{13.8}$$

where W_k is the weight given to value of Y at time k (Y_k) and $\sum_{k=t+1-N}^{t} W_k = 1$.

EXAMPLE 13.1

We Sell Beauty (WSB) is a manufacturer and distributor of health and beauty products. WSB is interested in forecasting demand for 'Kesh', their shampoo brand which is sold in 100 ml bottles. WSB believes that the monthly demand for 'Kesh' depends on the promotion expenditure (in thousands of rupees) and whether the competition was on promotion or not during that month. The data for 48 months (starting from January 2012) is shown in Table 13.1. The table has the quantity of 100 ml bottles sold during the month, promotion expenses (in thousands of rupees) incurred by the company, and whether the competition was on promotion (value of 1 implies that the competition was on promotion and 0 otherwise). Use simple moving average with N = 12 and forecast the demand of Kesh for months 37 to 48. Calculate the values of MAPE and RMSE.

TABLE 13.1 Data on sales of shampoo, promotion expenses (in 1000 of rupees), and dummy variable for promotion by competition

Month	Sale Quantity	Promotion Expenses	Competition Promotion	Month	Sale Quantity	Promotion Expenses	Competition Promotion
1	3002666	105	1	25	4634047	165	0
2	4401553	145	0	26	3772879	129	1
3	3205279	118	1	27	3187110	120	1
4	4245349	130	0	28	3093683	112	1
5	3001940	98	1	29	4557363	162	0

TABLE 13.1 Data on sales of shampoo, promotion expenses (in 1000 of rupees), and dummy variable for promotion by competition—Continued

Month	Sale Quantity	Promotion Expenses	Competition Promotion	Month	Sale Quantity	Promotion Expenses	Competition Promotion
6	4377766	156	0	30	3816956	140	1
7	2798343	98	1	31	4410887	160	0
8	4303668	144	0	32	3694713	139	0
9	2958185	112	1	33	3822669	141	1
10	3623386	120	0	34	3689286	136	0
11	3279115	125	0	35	3728654	130	1
12	2843766	102	1	36	4732677	168	0
13	4447581	160	0	37	3216483	121	1
14	3675305	130	0	38	3453239	128	0
15	3477156	130	0	39	5431651	170	0
16	3720794	140	0	40	4241851	160	0
17	3834086	167	1	41	3909887	151	1
18	3888913	148	1	42	3216438	120	1
19	3871342	150	1	43	4222005	152	0
20	3679862	129	0	44	3621034	125	0
21	3358242	120	0	45	5162201	170	0
22	3361488	122	0	46	4627177	160	0
23	3670362	135	0	47	4623945	168	0
24	3123966	110	1	48	4599368	166	0

Moving average forecast for the period n = 37 to 48 is given by

$$F_{t+1} = \frac{1}{12}\sum_{k=t+1-12}^{t} Y_k, \text{ for } t = 36,\ 37,\ \ldots,\ 47$$

The forecasted values using 12 period moving average and the corresponding RMSE and MAPE calculations are given in Table 13.2.

TABLE 13.2 Simple moving average forecast, RMSE, and MAPE calculations

Month	Actual Demand (Y_t)	Forecasted Demand (F_t)	$(Y_t - F_t)^2$	$\lvert Y_t - F_t\rvert / Y_t$
37	3216483	3928410	5.07E + 11	0.221337
38	3453239	3810280	1.27E + 11	0.103393
39	5431651	3783643	2.72E + 12	0.303408
40	4241851	3970688	7.35E + 10	0.063926

(*Continued*)

TABLE 13.2 Simple moving average forecast, RMSE, and MAPE calculations —Continued

Month	Actual Demand (Y_t)	Forecasted Demand (F_t)	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
41	3909887	4066369	2.45E + 10	0.040022
42	3216438	4012413	6.34E + 11	0.247471
43	4222005	3962370	6.74E + 10	0.061496
44	3621034	3946629	1.06E + 11	0.089918
45	5162201	3940490	1.49E + 12	0.236665
46	4627177	4052117	3.31E + 11	0.124279
47	4623945	4130275	2.44E + 11	0.106764
48	4599368	4204882	1.56E + 11	0.08577

The RMSE using the moving average forecast is given by 734725.8359 and the MAPE value is 0.1403 (or 14.03%). The graph of actual and forecasted demand is shown in Figure 13.2.

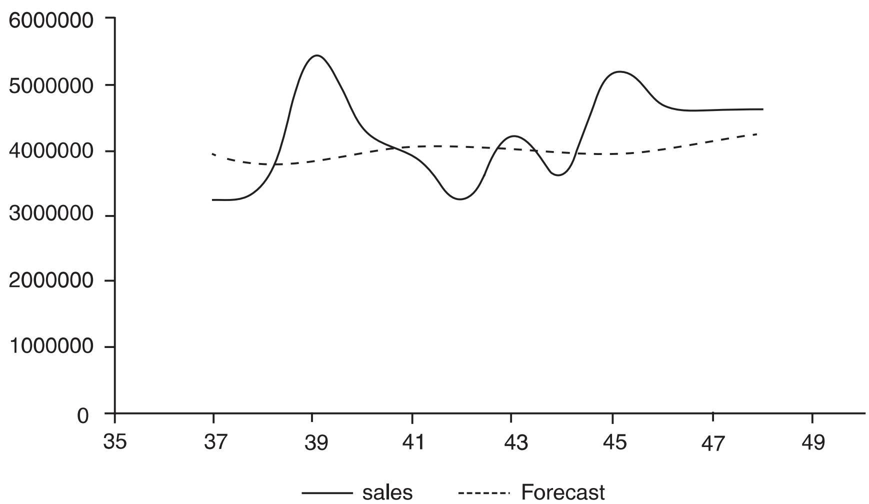


FIGURE 13.2 Plot of actual sales forecasted sales using moving average.

In moving average an important decision that a one has to take is the number of periods, N. The forecast accuracy will depend on the chosen N. If N is small, then the average tends to be more sensitive to recent observations or more responsive to recent trend. So, if responsiveness is important, then relatively few data points may be included. This would enable the moving average to make adjustments with the changes in the data quickly, though at times it would also be responding to just the random noise in the data. On the other hand, if N is large, that is more data points are included, then the forecast will be less sensitive or response to the recent changes in the data. Since the moving average will always be centered around the range of the data points considered, it will lag behind the trend until about $(N + 1)/2$ time periods.

13.5 | SINGLE EXPONENTIAL SMOOTHING (ES)

One of the drawbacks of simple moving average technique is that it gives equal weight to all the previous observations used in forecasting the future value. This can be overcome by assigning differential weights to the past observations [Eq. (13.8)]. One easier way to assign differential weight is achieved by using single exponential smoothing (SES) technique (also known as simple exponential smoothing). Just like the moving average, SES assumes a fairly steady time-series data with no significant trend, seasonal or cyclical component. Here, the weights assigned to past data decline exponentially with the most recent observations assigned higher weights.

In single ES, the forecast at time $(t + 1)$ is given by (Winters, 1960)

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t \tag{13.9}$$

Parameter α in Eq. (13.9) is called the **smoothing constant** and its value lies between 0 and 1. Since the model uses one smoothing constant, it is called **single exponential smoothing**. Substituting for F_t recursively in Eq. (13.9), we get

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t F_1 \tag{13.10}$$

From Eq. (13.10), it is evident that the weights assigned to older observations decrease exponentially. Figure 13.3 shows the rate at which the weight decreases for older observations when $\alpha = 0.4$ and 0.8; the plot resembles the exponential decay curve.

The forecasted values for months 37 to 48 for the data in Table 13.1 using simple exponential smoothing is shown in Table 13.3. Exponential smoothing uses the entire historical data. To begin exponential

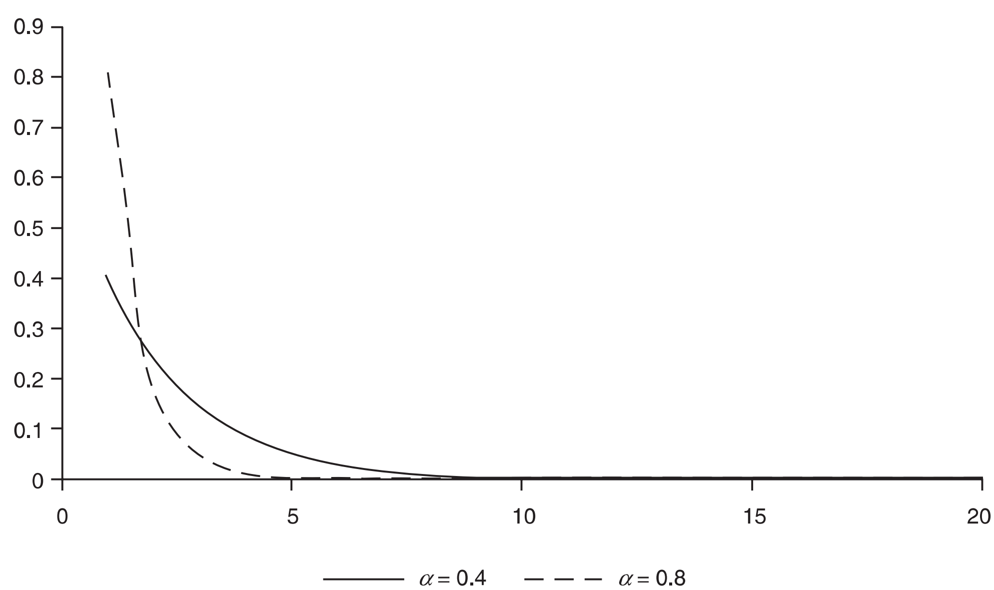


FIGURE 13.3 Exponential decay of weights to older observations.

TABLE 13.3 Forecast for periods 37 to 48 using single exponential smoothing ($\alpha = 0.2$)

Month	Actual Demand	Forecasted Demand $\alpha = 0.2$	$(Y_t - F_t)^2$	$\frac{\lvert Y_t - F_t \rvert}{Y_t}$
37	3216483	3980905	5.8434E + 11	0.237658
38	3453239	3828020	1.4046E + 11	0.10853
39	5431651	3753064	2.8177E + 12	0.309038
40	4241851	4088781	2.343E + 10	0.036086
41	3909887	4119395	4.3894E + 10	0.053584
42	3216438	4077494	7.4142E + 11	0.267705
43	4222005	3905283	1.0031E + 11	0.075017
44	3621034	3968627	1.2082E + 11	0.095993
45	5162201	3899108	1.5954E + 12	0.244681
46	4627177	4151727	2.2605E + 11	0.102752
47	4623945	4246817	1.4223E + 11	0.08156
48	4599368	4322243	7.6799E + 10	0.060253

smoothing we will need the forecast for the F_t in Eq. (13.9). We can use $F_t = Y_t$ or use moving average to forecast the initial forecast F_t. The forecasted value for period 2 is given by

$$F_2 = \alpha Y_1 + (1 - \alpha) F_1$$

We will assume F_1 same as Y_1. Thus the value of F_2 will be same as Y_1, that is 3002666. The forecasted values using single exponential smoothing with $\alpha = 0.2$ are shown in Table 13.3.

The RMSE using the single exponential smoothing with $\alpha = 0.2$ is given by 742339.222 and the MAPE value is 0.1394 (or 13.94%).

In summary, single exponential smoothing technique has the following advantages:

1. It uses all the historic data unlike the moving average where only the past few observations are considered to predict the future value.
2. It assigns progressively decreasing weights to older data.

Some disadvantages of smoothing methods are:

1. Increasing n makes forecast less sensitive to changes in data.
2. It always lags behind trend as it is based on past observations. The longer the time period n, the greater the lag as it is slow to recognize the shifts in the level of the data points.
3. Forecast bias and systematic errors occur when the observations exhibit strong trend or seasonal patterns.

13.5.1 | Optimal Smoothing Constant in a Single Exponential Smoothing (SES)

Choosing optimal smoothing constant α is important for accurate forecast. Whenever the data is smooth (without much fluctuations), we may choose higher value of α. However, when the data is highly fluctuating, then it is better to choose lower value of α. We can find the optimal value of the smoothing constant by solving a non-linear optimization problem. For example, assume that we have to find the optimal α that will give the minimum RMSE. This can be achieved by solving the following optimization problem:

$$\underset{\alpha}{Min}\left[\sqrt{\frac{1}{n}\sum_{t}\left(Y_t - F_t\right)^2}\right] \tag{13.11}$$

subject to the constraint: $0 < \alpha < 1$. For the data in Table 13.1, the optimal value of α that minimizes the RMSE is 0.1574 and the corresponding RMSE is 739399.76. Table 13.4 shows the forecasted value, RMSE, and MAPE calculations for $\alpha = 0.1574$ (rounded to 4 decimals).

13.6 | DOUBLE EXPONENTIAL SMOOTHING – HOLT'S METHOD

One of the drawbacks of single exponential smoothing is that the model does not do well in the presence of trend. This can be improved by introducing an additional equation for capturing the trend in the time-series data. Double exponential smoothing uses two equations to forecast the future values of the time series, one for forecasting the level (short term average value) and another for capturing the trend. The two equations are provided in Eqs. (13.12) and (13.13).

TABLE 13.4 Forecast for periods 37 to 48 using single exponential smoothing ($\alpha = 0.1574$)

Month	Actual Demand	Forecasted Demand $\alpha = 0.1574$	$(Y_t - F_t)^2$	$\frac{\lvert Y_t - F_t \rvert}{Y_t}$
37	3216483	3931892	5.1181E + 11	0.22242
38	3453239	3819242	1.3396E + 11	0.105988
39	5431651	3761610	2.789E + 12	0.307465
40	4241851	4024580	4.7207E + 10	0.051221
41	3909887	4058792	2.2173E + 10	0.038084
42	3216438	4035345	6.7061E + 11	0.254601
43	4222005	3906397	9.9608E + 10	0.074753
44	3621034	3956094	1.1227E + 11	0.092532
45	5162201	3903334	1.5847E + 12	0.243862
46	4627177	4101560	2.7627E + 11	0.113594
47	4623945	4184325	1.9327E + 11	0.095075
48	4599368	4253549	1.1959E + 11	0.075188

Level (or Intercept) equation (L_t):

$$L_t = \alpha \times Y_t + (1 - \alpha) \times F_t \tag{13.12}$$

The trend equation is given by (T_t)

$$T_t = \beta \times (L_t - L_{t-1}) + (1 - \beta) \times T_{t-1} \tag{13.13}$$

α and β are the smoothing constants for level and trend, respectively, and $0 < \alpha < 1$ and $0 < \beta < 1$.

The forecast at time $t + 1$ is given by

$$F_{t+1} = L_t + T_t \tag{13.14}$$

$$F_{t+n} = L_t + nT_t \tag{13.15}$$

where L_t is the level which represents the smoothed value up to and including the last data, T_t is the slope of the line or the rate of increase or decrease at period t, n is the number of time periods into the future.

Initial value of L_t is usually taken same as Y_t (that is, $L_t = Y_t$). The starting value of T_t can be taken as $(Y_t - Y_{t-1})$ or the difference between two previous actual values of observations prior to the period for which forecasting is carried out. Another option for T_t is $(Y_t - Y_1)/(t - 1)$.

The value of

$$L_1 = Y_1 = 3002666$$

and $T_1 = (Y_{36} - Y_1)/35 = (4732677 - 3002666)/35 = 49428.8857$

The value of $F_2 = L_1 + T_1 = 3002666 + 49428.8857 = 3052095$

The forecasted values for periods 37 to 48 are shown in Table 13.5 ($\alpha = 0.0328$ and $\beta = 0.9486$).

The RMSE and MAPE of the forecast using double exponential smoothing is given by 659888.9554 and 0.1135 (11.35%). The values of α and β used in Table 13.5 are optimized values of α and β that minimize the root mean square error.

TABLE 13.5 Forecasted values using double exponential smoothing ($\alpha = 0.0328$ and $\beta = 0.9486$)

Month	Actual Demand	L_t	T_t	$F_t (= L_{t-1} + T_{t-1})$	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
37	3216483	3678293	66894.6916	3693955	2.27979E+11	0.148445
38	3453239	3735612	57810.9617	3745188	85234318285	0.084544
39	5431651	3847157	108782.913	3793423	2.68379E+12	0.301608
40	4241851	3965318	117678.771	3955940	81745109031	0.067402
41	3909887	4077319	112292.624	4082997	29966946139	0.044275
42	3216438	4157691	82013.2329	4189611	9.47066E+11	0.302562
43	4222005	4239124	81462.532	4239704	313269245.9	0.004192
44	3621034	4297641	59696.6025	4320586	4.89374E+11	0.193191
45	5162201	4383737	84739.1839	4357338	6.47805E+11	0.155915
46	4627177	4473682	89677.0074	4468476	25185883916	0.034298
47	4623945	4565346	91562.092	4563359	3670690475	0.013103
48	4599368	4655021	89771.7849	4656908	3310862728	0.01251

13.7 | TRIPLE EXPONENTIAL SMOOTHING (HOLT-WINTER MODEL)

Moving averaging and single and double exponential smoothing techniques discussed so far can handle data as long as the data do not have any seasonal component associated with it. However, when there is seasonality in the time-series data, techniques such as moving average, exponential smoothing, and double exponential smoothing are no longer appropriate. In most cases, the fitted error values (actual demand minus forecast) associated with simple exponential smoothing and Holt's method will indicate systematic error patterns that reflect the existence of seasonality. For example, presence of seasonality may result in all positive errors, except for negative values that occur at fixed intervals. Such pattern in error would imply existence of seasonality. Such time series data require the use of a seasonal method to eliminate the systematic patterns in error.

Triple exponential smoothing is used when the data has trend as well as seasonality. The following three equations which account for level, trend, and seasonality are used for forecasting (for a multiplicative model, Winters 1960):

Level (or Intercept) equation:

$$L_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)\left[L_{t-1} + T_{t-1}\right] \tag{13.16}$$

Trend equation:

$$T_t = \beta \times (L_t - L_{t-1}) + (1 - \beta)\, T_{t-1} \tag{13.17}$$

Seasonal equation:

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-c} \tag{13.18}$$

The forecast F_{t+1} using triple exponential smoothing is given by

$$F_{t+1} = \left[L_t + T_t\right] \times S_{t+1-c} \tag{13.19}$$

where c is the number of seasons (if it is monthly seasonality, then $c = 12$; in case of quarterly seasonality $c = 4$; and in case of daily data $c = 7$). The initial values of L_t and T_t are calculated using the following equations:

$$L_t = Y_t \tag{13.20}$$

Alternatively

$$L_t = \frac{1}{c}\left(Y_1 + Y_2 + \cdots + Y_c\right) \tag{13.21}$$

$$T_t = \frac{1}{c}\left[\frac{Y_t - Y_{t-c}}{12} + \frac{Y_{t-1} - Y_{t-c-1}}{12} + \frac{Y_{t-2} - Y_{t-c-2}}{12} + \ldots + \frac{Y_{t-c+1} - Y_{t-2c+1}}{12}\right] \tag{13.22}$$

Several techniques exist to calculate the initial seasonality index (Winters, 1961; Makridakis *et al.*, 1998; Taylor 2011). The initial seasonality index can be calculated using a technique called method of simple averages which is described in Section 13.7.1. Several variations to the procedure discussed in Section 13.7.1 exist, such as ratio-to-moving average.

13.7.1 | Predicting Seasonality Index Using Method of Averages

The following steps are used for predicting the seasonality index using method of averages:

STEP 1

Calculate the average of value of Y for each season (that is, if the data is monthly data, then we need to calculate the average for each month based on the training data). Let these averages be $\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_c$.

STEP 2

Calculate the average of the seasons' averages calculated in step 1 (say $\bar{\bar{Y}}$).

STEP 3

The seasonality index for season k is given by the ratio $\bar{Y}_k / \bar{\bar{Y}}$. Variation to the procedure explained above is first divide the value of Y_t with its yearly average and calculate the seasonal average. We will use first 3 years of data in Table 13.1 to calculate the seasonality index for various months. The seasonality index based on first 3 years of data using method of averages is shown in Table 13.6.

Seasonality index can be interpreted as percentage change from the trend line. For example, the seasonality index for January is approximately 1.088 or 108.8%. This implies that in January, the demand will be approximately 8.8% more from the trend line. The seasonality index for March is 0.8885 or 88.85%.

TABLE 13.6 Seasonality index using method of averages

Month	Sale Quantity (2012)	Sale Quantity (2013)	Sale Quantity (2014)	Monthly Average $\bar{Y}_K$	Seasonality Index $\bar{Y}_k / \bar{\bar{Y}}$
January	3002666	4447581	4634047	4028098.00	1.087932
February	4401553	3675305	3772879	3949912.33	1.066815
March	3205279	3477156	3187110	3289848.33	0.888541
April	4245349	3720794	3093683	3686608.67	0.9957
May	3001940	3834086	4557363	3797796.33	1.02573
June	4377766	3888913	3816956	4027878.33	1.087872
July	2798343	3871342	4410887	3693524.00	0.997568
August	4303668	3679862	3694713	3892747.67	1.051375
September	2958185	3358242	3822669	3379698.67	0.912808
October	3623386	3361486	3689286	3558053.33	0.960979
November	3279115	3670362	3728654	3559377.00	0.961337
December	2843766	3123966	4732677	3566803.00	0.963342
Average of monthly averages				3702528.22	

This implies that the demand in March will be 11.15% less from the trend line. Note that, multiplicative model is used in this example.

To start the triple exponential smoothing, we need to set the starting values of level and trend.

$$L_{36} = Y_{36}/S_{36} = 4732677/0.9633 = 4912983.494$$

The initial value of trend (T_{36}) can be calculated based on second and third year by using Eq. (13.22):

$$T_{36} = \frac{1}{12}\left[\frac{Y_{36}-Y_{24}}{12}+\frac{Y_{35}-Y_{23}}{12}+\frac{Y_{34}-Y_{22}}{12}+\cdots+\frac{Y_{25}-Y_{13}}{12}\right]$$

$$T_{36} = \frac{1}{12}\left[\frac{4732677-3123966}{12}+\frac{3728654-3670362}{12}+\ldots+\frac{4634047-4447581}{12}\right] = 21054.35$$

The forecast for period 37 using triple exponential smoothing is given by

$$F_{37} = [L_{36} + T_{36}] \times S_{37-12} = [L_{36} + T_{36}] \times S_{25} \tag{13.23}$$

The seasonal index S_{25} (seasonality index for January) is 1.088. Substituting the values of L_{36}, T_{36} and S_{25}, we get

$$F_{37} = [4912983.494 + 21054.35] \times 1.088 = 5368233.2$$

The forecast for the period 37 to 48 for the data in Table 13.1 is given in Table 13.7. Note that the values such as seasonality index are rounded to two decimals, the forecast values will be different if the actual seasonality index values in Table 13.6 are used.

The RMSE and MAPE using triple exponential smoothing are 1228588.29 and 0.2208 (22.08%), respectively. The values of $\alpha = 0.32$, $\beta = 0.5$, and $\gamma = 1$ are used for calculating the level, trend, and seasonal components. It is important to note that the exponential smoothing techniques are very sensitive to initial values of level, trend, and seasonal index.

13.8 | CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

Products such as spare parts may have intermittent demands. Exponential smoothing models discussed so far in the chapter will produce biased estimate when used for intermittent demand. Croston (1972) developed a model that uses two separate exponential smoothing equations for predicting mean time between demands and the magnitude of demand whenever the demand occurs. That is, Croston's method has two components: (a) Predicting time between demand and (b) magnitude of the demand. The primary objective of Croston's method is to forecast mean demand per period. Let

Y_t = Demand at time t (Y_t may take value 0)
F_t = Forecasted demand
TD_t = Time between the latest and the previous non-zero demand in period t
FTD_t = Forecasted time between demand at period t

The following steps are used for forecasting demand:

$$\text{If } Y_t = 0 \text{ then } F_{t+1} = F_t \text{ and } FTD_{t+1} = FTD_t \tag{13.24}$$

$$\text{If } Y_t \neq 0 \text{ the } F_{t+1} = \alpha \times Y_t + (1-\alpha)F_t \text{ and } FTD_{t+1} = \beta \times TD_t + (1-\beta) \times FTD_t \tag{13.25}$$

TABLE 13.7 Forecasting using triple exponential smoothing (values differ for different round off values of parameters)

Month t	Actual Demand	L_{t-1}	T_{t-1}	S_t	F_t	$(Y_t - F_t)^2$	$\|Y_t - F_t\|/Y_t$
37	3216482	4912983.49	21054.35	1.09	5367895.97	4.62858E+12	0.668872
38	3453239	4301229.28	−295349.93	1.07	4273531.48	6.7288E+11	0.237543
39	5431651	3759825.78	−418376.71	0.89	2969014.38	6.06458E+12	0.453386
40	4241851	4228345.39	25071.45	1.00	4235127.90	45200134.5	0.001585
41	3909887	4255577.53	26151.79	1.03	4391900.21	2.32337E+11	0.123281
42	3216437	4131354.31	−49035.71	1.09	4441041.44	1.49966E+12	0.380733
43	4222004	3722098.55	−229145.74	1.00	3484457.63	5.43975E+11	0.174691
44	3621034	3729543.06	−110850.61	1.05	3804603.81	33697874118	0.050695
45	5162201	3562820.55	−138786.56	0.91	3125486.18	4.14821E+12	0.394544
46	4627176	4138038.05	218215.47	0.96	4186269.09	1.94399E+11	0.095286
47	4623945	4503072.73	291625.07	0.96	4609319.06	213918263.4	0.003163
48	4599368	4799566.34	294059.34	0.96	4906905.01	94579010434	0.066865

α and β are smoothing constants for forecasted demand and forecasted time between demands, respectively. Once the forecasted demand and time between demands are known, then the mean demand per period, D_{t+1}, is given by

$$D_{t+1} = \frac{F_{t+1}}{FDT_{t+1}} \tag{13.26}$$

EXAMPLE 13.2

Quarterly demand for spare parts of avionics system of an aircraft is shown in Table 13.8. Use the demand during the quarters 1 to 4 as training data to forecast demand for periods 5 to 16 using Croston's method.

TABLE 13.8 Quarterly demand for avionic system spares

Quarter	1	2	3	4	5	6	7	8
Demand	20	12	0	18	16	0	20	22
Quarter	9	10	11	12	13	14	15	16
Demand	0	28	0	0	30	26	0	34

Procedure used for starting values of F_t and FTD_t is shown in Table 13.9.

TABLE 13.9 Calculating initial values in Croston's method

Quarter	Demand	TD_t	FTD_t	F_t
1	20			
2	12	1		
3	0			
4	18	2	1.5	16.67

In Table 13.9, $TD_4 = 2$ since the elapsed time from the previous demand and current demand period is 2 (4 – 2). The forecasted time between demand is the average TD_t values up to $t = 4$. So, $FTD_4 = (1+2)/2 = 1.5$. The forecasted demand F_4 for $t = 4$ is $(20 + 12 + 18)/3 = 16.67$. Note that the total value is divided by 3 (not 4) since only 3 quarters had non-zero demand. So, the starting values for Croston's method are

$TD_4 = 2$, $FTD_4 = 1.5$, and $F_4 = 16.67$

Let $\alpha = \beta = 0.2$. Then

$$F_5 = 0.2 \times 18 + (1 - 0.2) * 16.67 = 16.936$$

$$FTD_5 = 0.2 \times 2 + (1 - 0.2) \times 1.5 = 1.6$$

The forecasted values for remaining quarters are shown in Table 13.10.

TABLE 13.10 Forecasted demand for periods 5 to 16 using Croston's method

Quarter	Demand	TD_t	FTD_t	F_t	$D_t = (F_t/FTD_t)$
1	20				
2	12	1			
3	0				
4	18	2	1.5000	16.67	11.11333
5	16	1	1.6000	16.936	10.585
6	0		1.4800	16.7488	11.31676
7	20	2	1.4800	16.7488	11.31676
8	22	1	1.5840	17.39904	10.98424
9	0		1.4672	18.31923	12.48585
10	28	2	1.4672	18.31923	12.48585
11	0		1.5738	20.25539	12.8707
12	0		1.5738	20.25539	12.8707
13	30	3	1.5738	20.25539	12.8707
14	26	1	1.8590	22.20431	11.94417
15	0		1.6872	22.96345	13.61034
16	34	2	1.6872	22.96345	13.61034

13.9 | REGRESSION MODEL FOR FORECASTING

Regression is probably more appropriate method for forecasting when the data has values of predictor (explanatory) variables in addition to the dependent variable Y_t. In the data provided in Table 13.1, we also have information such as promotion expenses and whether the competition was on promotion or not. Using the values of these predictor variables is likely to give better forecast than the exponential smoothing techniques discussed in the previous sections. Parker and Segura (1971) claimed that regression method can predict more accurately than less scientific methods such as exponential smoothing. The forecasted value at time t, F_t, can be written as a regression equation as follows:

$$F_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_n X_{nt} + \varepsilon_t \tag{13.27}$$

Here F_t is the forecasted value of Y_t, and X_{1t}, X_{2t}, etc. are the predictor variables measured at time t. The regression equation for Example 13.1 is

$$F_t = \beta_0 + \beta_1 \text{ promotion expenses at time } t + \beta_2 \text{ competition promotion at time } t$$

For the data in Table 13.1, the regression outputs using SPSS are shown in Tables 13.11 and 13.12. The model is developed based on first 36 months data.

TABLE 13.11 Model summary

Model	*R*	*R*-Square	Adjusted *R*-Square	Std. Error of the Estimate	Durbin–Watson
1	0.928	0.862	0.853	207017.359	1.608

The R-square for the model is 0.862 (note that we will need high R-square value for forecasting applications) and the Durbin–Watson statistic value is 1.608. Since this is a time-series data we need to check whether the errors, ε_t, are correlated (auto-correlation). For n = 36 (sample size) and number of predictor variables = 2, the Durbin–Watson critical values are d_L = 1.153 and d_U = 1.376. Since the model Durbin–Watson statistic D = 1.608 (4 – D = 2.392) lies within d_U and (4 – d_U), we can conclude that there is no auto-correlation. Whenever regression model is used for forecasting, it should be checked for auto-correlation among the errors. Presence of auto-correlation may lead to inclusion of a non-significant variable in the model since the standard error of the regression coefficient is underestimated when auto-correlation of errors is present.

TABLE 13.12 Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	808471.843	278944.970		2.898	0.007
	Promotion Expenses	22432.941	1953.674	0.825	11.482	0.000
	Competition Promotion	–212646.036	77012.289	–0.198	–2.761	0.009

The regression model (based on values in Table 13.12) is given by

$$F_t = 808471.843 + 22432.941 X_{1t} - 212646.036 X_{2t} \tag{13.28}$$

where

$$X_{1t} = \text{Promotion expenses at time } t$$

$$X_{2t} = \begin{cases} 1 & \text{Competition is on promotion} \\ 0 & \text{Otherwise} \end{cases}$$

As expected, the sales increases as the promotion expenses increase and the sales decreases whenever the competition is on promotion. The forecasted values for period 37 to 48 using the regression model [Eq. (13.28)] is shown in Table 13.13.

TABLE 13.13 Forecasts using regression model

Period	Y_t	X_{1t}	X_{2t}	F_t	$(Y_t - F_t)^2$	$\frac{\lvert Y_t - F_t \rvert}{Y_t}$
37	3216483	121	1	3310211.67	8785063205	0.02914
38	3453239	128	0	3679888.29	5.137E+10	0.065634
39	5431651	170	0	4622071.81	6.5542E+11	0.149048
40	4241851	160	0	4397742.4	2.4302E+10	0.036751
41	3909887	151	1	3983199.9	5374781013	0.018751
42	3216438	120	1	3287778.73	5089499329	0.02218
43	4222005	152	0	4218278.88	13884007.5	0.000883
44	3621034	125	0	3612589.47	71310120.7	0.002332
45	5162201	170	0	4622071.81	2.9174E+11	0.104632
46	4627177	160	0	4397742.4	5.264E+10	0.049584
47	4623945	168	0	4577205.93	2184540571	0.010108
48	4599368	166	0	4532340.05	4492746215	0.014573

The RMSE and MAPE based on regression model are 302969 and 0.0419 (or 4.19%), respectively. The RMSE and MAPE for regression based forecasting are much smaller than the values that we obtained so far using moving average and exponential smoothing techniques. For Example 13.1, the moving average method resulted in an RMSE of 734725.84 and MAPE is 14.03%. The RMSE and MAPE for single exponential smoothing are 742339.22 and 13.94%, respectively. The plot of actual demand and forecasted demand using regression model is shown in Figure 13.4.

13.9.1 | Forecasting Time-Series Data with Seasonal Variation

As mentioned earlier, one can expect seasonal variation in demand for many products and services. The following steps are used to forecast time-series data with seasonal variations:

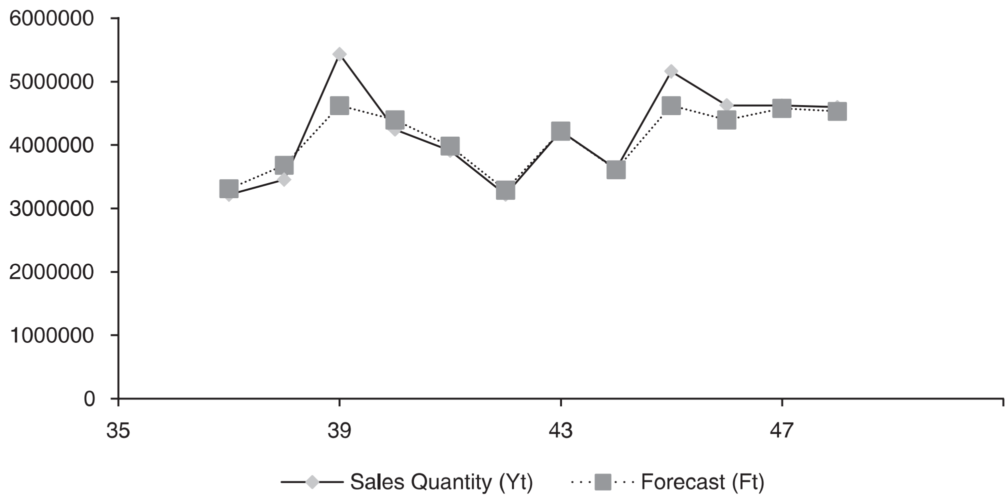


FIGURE 13.4 Actual sales quantity and forecasted sales using regression model.

STEP 1

Estimate the seasonality index (using techniques such as method of averages or ratio to moving average).

STEP 2

De-seasonalize the data using either additive or multiplicative model. For example, in multiplicative model, the de-seasonalized data $Y_{d,t} = Y_t / S_t$, where $Y_{d,t}$ is the de-seasonalized data and S_t is the seasonality index for period t.

STEP 3

Develop a forecasting model on the de-seasonalized data ($F_{d,t}$).

STEP 4

The forecast for period $t + 1$ is $F_{t+1} = F_{d,\, t+1} \times S_{t+1}$.

EXAMPLE 13.3

Hiccup Viking (HV) is The Vice President of Viking Cookies that specialized into chocolate chip cookies (Choco-Chip). Viking Cookies believes that demand for cookies is seasonal and is driven by several factors such as school holidays, festivals, etc. The shelf life of Choco-Chip cookies is 6 months and excess inventory and running out of stock can have financial impact. Hiccup would like to develop a forecasting model that they can use for forecasting the demand. The past monthly demand (quantity of 200 gram packets) for four years (January 2013 to December 2016) along with average price per unit during that month is shown in Table 13.14. Develop a forecasting model using regression to predict demand between months 37 and 48, given that the data is seasonal.

TABLE 13.14 Monthly demand (quantity of 200 gram packets) along with average price per unit

Period	Month	Demand in Units	Average Price	Period	Demand in Units	Average Price
1	January	10500472	37	25	10658309	36
2	February	10123572	34	26	8677622	38
3	March	7372141	36	27	7330354	37
4	April	7764303	38	28	8115471	37
5	May	6904463	40	29	8481936	34
6	June	10068862	34	30	8778999	37
7	July	6436190	40	31	10145039	32
8	August	9898436	34	32	8497839	38
9	September	6803825	39	33	8792138	34
10	October	8333787	36	34	8485358	36
11	November	7541964	39	35	8575904	36
12	December	8540662	37	36	9885156	32
13	January	10229437	37	37	11023467	35
14	February	8453201	38	38	7942451	40
15	March	7997459	35	39	12492798	32
16	April	8557825	35	40	9756258	32
17	May	7818397	36	41	8992741	32
18	June	8944499	37	42	7397807	40
19	July	8904086	36	43	9710611	32
20	August	8463682	39	44	8328379	39
21	September	7723957	37	45	11873063	32
22	October	7731422	39	46	10642507	32
23	November	8441834	35	47	10635075	32
24	December	7485122	40	48	10578547	32

Solution:

Since the demand is seasonal, the first step in forecasting is to estimate the seasonality index. We can use first 36 months data to estimate the seasonality index using method of averages explained in Section 13.7.1. Table 13.15 gives the seasonality index for various months. For example, the seasonality index for January is 1.2251. That is, in January the demand will increase by 22.51% from the trend.

TABLE 13.15 Seasonality index for various months

Month	Demand (2012)	Demand (2013)	Demand (2014)	Average	Seasonality Index
1	10500472	10229437	10658309	10462739	1.2251
2	10123572	8453201	8677622	9084798	1.0637
3	7372141	7997459	7330354	7566651	0.8860
4	7764303	8557825	8115471	8145866	0.9538
5	6904463	7818397	8481936	7734932	0.9057
6	10068862	8944499	8778999	9264120	1.0847
7	6436190	8904086	10145039	8495105	0.9947
8	9898436	8463682	8497839	8953319	1.0483
9	6803825	7723957	8792138	7773307	0.9102
10	8333787	7731422	8485358	8183522	0.9582
11	7541964	8441834	8575904	8186567	0.9585
12	8540662	7485122	9885156	8636980	1.0113
		Average of monthly averages		8540659	

De-seasonalized data is calculated by dividing the value of Y_t with the corresponding seasonality index. The de-seasonalized data for periods 1 to 48 is shown in Table 13.16.

TABLE 13.16 De-seasonalized demand (seasonality index from Table 13.25 is rounded to 2 decimals)

Month	Demand	Seasonality Index	De-seasonalized Demand	Month	Demand	Seasonality Index	De-seasonalized Demand
1	10500472	1.23	8571459.88	25	10658309	1.23	8700301.09
2	10123572	1.06	9517214.68	26	8677622	1.06	8157870.71
3	7372141	0.89	8321110.54	27	7330354	0.89	8273944.56
4	7764303	0.95	8140603.02	28	8115471	0.95	8508790.52
5	6904463	0.91	7623682.26	29	8481936	0.91	9365476.36
6	10068862	1.08	9282556.42	30	8778999	1.08	8093422.43

TABLE 13.16 De-seasonalized demand (seasonality index from Table 13.25 is rounded to 2 decimals)—Continued

Month	Demand	Seasonality Index	De-seasonalized Demand	Month	Demand	Seasonality Index	De-seasonalized Demand
7	6436190	0.99	6470703.29	31	10145039	0.99	10199440.54
8	9898436	1.05	9442215.37	32	8497839	1.05	8106172.12
9	6803825	0.91	7475473.63	33	8792138	0.91	9660065.60
10	8333787	0.96	8697481.33	34	8485358	0.96	8855667.03
11	7541964	0.96	7868174.77	35	8575904	0.96	8946835.52
12	8540662	1.01	8445415.13	36	9885156	1.01	9774915.11
13	10229437	1.23	8350215.95	37	11023467	1.23	8998376.94
14	8453201	1.06	7946891.53	38	7942451	1.06	7466733.21
15	7997459	0.89	9026921.81	39	12492798	0.89	14100917.65
16	8557825	0.95	8972583.38	40	9756258	0.95	10229098.91
17	7818397	0.91	8632818.30	41	8992741	0.91	9929490.54
18	8944499	1.08	8245998.07	42	7397807	1.08	6820091.57
19	8904086	0.99	8951833.08	43	9710611	0.99	9762682.98
20	8463682	1.05	8073589.43	44	8328379	1.05	7944522.57
21	7723957	0.91	8486437.69	45	11873063	0.91	13045128.20
22	7731422	0.96	8068828.55	46	10642507	0.96	11106956.05
23	8441834	0.96	8806966.63	47	10635075	0.96	11095071.36
24	7485122	1.01	7401646.68	48	10578547	1.01	10460573.30

Regression output for the de-seasonalized demand and average price using Microsoft Excel are shown in Table 13.17.

TABLE 13.17 Regression output using SPSS for data in Table 13.16 (based on first 36 cases)

Model		Unstandardized Coefficients		*T*	Sig.
		B	Std. Error		
1	(Constant)	20812014.673	717702.417	28.998	0.000
	Average Price	−335945.859	19616.915	−17.125	0.000

Regression model for demand forecasting based on first 36 months of de-seasonalized data is given by

$$F_{d,t} = 20812014.673 - 335945.859 \times \text{Average Price}$$

The forecasted values are given in Table 13.18.

TABLE 13.18 Forecasted values for the data in Table 13.15

Month	Demand	Seasonality Index (S_t)	De-seasonalized Demand	$F_{d,t}$	$F_t = F_{d,t}{}^* S_t$	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
37	11023467	1.2251	8998377	9053910	11091497	4628131462	0.006171
38	7942451	1.0637	7466733	7374180	7844001	9692313467	0.012395
39	12492798	0.8860	14100918	10061747	8914269	1.2806×10^{13}	0.286447
40	9756258	0.9538	10229099	10061747	9596642	2.5477×10^{10}	0.01636
41	8992741	0.9057	9929491	10061747	9112521	1.4347×10^{10}	0.01332
42	7397807	1.0847	6820092	7374180	7998831	3.6123×10^{11}	0.081244
43	9710611	0.9947	9762683	10061747	10008080	8.8488×10^{10}	0.030633
44	8328379	1.0483	7944523	7710126	8082657	6.0379×10^{10}	0.029504
45	11873063	0.9102	13045128	10061747	9157730	7.373×10^{12}	0.228697
46	10642507	0.9582	11106956	10061747	9641005	1.003×10^{12}	0.094104
47	10635075	0.9585	11095071	10061747	9644592	9.8106×10^{11}	0.093134
48	10578547	1.0113	10460573	10061747	10175223	1.6267×10^{11}	0.038127

RMSE and MAPE values are 1381119.09 and 0.0775 (7.75%), respectively.

13.10 | AUTO-REGRESSIVE (AR), MOVING AVERAGE (MA) AND ARMA MODELS

Auto-regressive (AR) and moving average (MA) models are popular models that are frequently used for forecasting. AR and MA models are combined to create models such as auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) models. The initial ARMA and ARIMA models were developed by Box and Jenkins in 1970 (Box and Jenkins, 1970). ARMA models are basically regression models; auto-regression simply means regression of a variable on itself measured at different time periods. One of the fundamental assumptions of AR model is that the time series is assumed to be a stationary process. If a time-series data, Y_t, is stationary, then it satisfies the following conditions:

1. The mean values of Y_t at different values of t are constant.
2. The variances of Y_t at different time periods are constant (Homoscedasticity).
3. The covariances of Y_t and Y_{t-k} for different lags depend only on k and not on time t.

When the time series data is not stationary (that is, any one of the above conditions are not satisfied), then we have to convert the non-stationary times-series data to stationary data before applying AR models. Another important concept associated with forecasting based on regression-based models is the white noise of residuals. White noise is a process of residuals ε_t that are uncorrelated and follow normal distribution with mean 0 and constant standard deviation. In AR models, one of the important assumptions that we make is that the errors follow a white noise.

13.11 | AUTO-REGRESSIVE (AR) MODELS

Auto-regression is regression of a variable on itself measured at different time points. Auto-regressive model with lag 1, AR(1), is given by

$$Y_{t+1} = \beta Y_t + \varepsilon_{t+1} \tag{13.29}$$

which can be re-written as

$$Y_{t+1} - \mu = \beta \times (Y_t - \mu) + \varepsilon_{t+1} \tag{13.30}$$

where ε_{t+1} is a sequence of uncorrelated residuals that follow normal distribution with zero mean and constant standard deviation. $Y_{t+1} - \mu$ can be interpreted as deviation from mean value μ and is known as mean centered series. Equation (13.30) can be expanded recursively as shown in Eqs. (13.31) and (13.32):

$$Y_{t+1} - \mu = \beta \times \left[\beta \times (Y_{t-1} - \mu) + \varepsilon_t\right] + \varepsilon_{t+1} \tag{13.31}$$

$$Y_{t+1} - \mu = \beta^t (Y_0 - \mu) + \beta^{t-1}\varepsilon_1 + \beta^{t-2}\varepsilon_2 + \cdots + \beta\varepsilon_t + \varepsilon_{t+1} \tag{13.32}$$

Equation (13.32) can be written as

$$Y_{t+1} - \mu = \beta^t (Y_0 - \mu) + \sum_{k=1}^{t-1} \beta^{t-k} \times \varepsilon_k + \varepsilon_{t+1} \tag{13.33}$$

In Eq. (13.33), if $|\beta| > 1$ in the first part on the right-hand side $[\beta^t (Y_0 - \mu)]$ will result in infinitely large value of Y_{t+1} as the value of t increases and is not very useful for practical applications. The value of $|\beta| = 1$ would imply that the future value of Y depends on the entire past (and will lead to non-stationarity). For practical applications, the value of $|\beta|$ should be less than one. The second part of Eq. (13.33), $\sum_{k=1}^{t-1} \beta^{t-k} \times \varepsilon_k$, can also become infinitely large if the errors do not follow a white noise. When the errors are white noise then the expected value of $\sum_{k=1}^{t-1} \beta^{t-k} \times \varepsilon_k$ is zero. The coefficient β in Eq. (13.30) can be estimated using ordinary least squares estimate. The sum of squared errors is given by

$$\sum_{t=2}^{n} \varepsilon_t^2 = \sum_{t=2}^{n} [(Y_t - \mu) - \beta \times (Y_{t-1} - \mu)]^2 \tag{13.34}$$

Taking first-derivative of Eq. (13.34) with respect to β and equating that to zero, the estimate of β is given by

$$\hat{\beta} = \frac{\sum_{t=2}^{n} (Y_t - \mu)(Y_{t-1} - \mu)}{\sum_{t=2}^{n} (Y_{t-1} - \mu)^2} \tag{13.35}$$

We can generalize auto-regressive model with 1 lag, AR(1), to auto-regressive model with p lags, AR(p) process is given by

$$Y_{t+1} = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \cdots + \beta_p Y_{t-p+1} + \varepsilon_{t+1} \tag{13.36}$$

The forecasted value is given by

$$F_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 Y_t + \hat{\beta}_2 Y_{t-1} + \cdots + \hat{\beta}_p Y_{t-p+1} \tag{13.37}$$

where $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, ... are the estimated values $\beta_0, \beta_1, \beta_2$, and so on. Note that software such as SPSS use the mean centered series [that is, equation of the form (13.30)] while estimating the parameters and hence forecasting the future values should be carried out using Eq. (13.30).

13.11.1 | AR Model Identification: ACF and PACF

One of the important tasks in using auto-regressive model in forecasting is the model identification, which is, identifying the value of p (the number of lags). One of the standard approaches used for model identification is auto-correlation function (ACF) and partial auto-correlation function (PACF). Auto-correlation is the correlation between Y_t measured at different time periods (for example, Y_t and Y_{t-1} or Y_t and Y_{t-k}). Auto-correlation indicates the memory of a process, that is, how far in time can it remember what has happened before. The auto-correlation of k-lags (correlation between Y_t and Y_{t-k}) is given by

$$\rho_k = \frac{\sum_{t=k+1}^{n} \left(Y_{t-k} - \bar{Y}\right)\left(Y_t - \bar{Y}\right)}{\sum_{t=1}^{n} \left(Y_t - \bar{Y}\right)^2} \tag{13.38}$$

where n is the number of observations in the sample. A plot of auto-correlation for different values of k is called **auto-correlation function (ACF)** or **correlogram**.

Partial auto-correlation of lag k (ρ_{pk}) is the correlation between Y_t and Y_{t-k} when the influence of all intermediate values (Y_{t-1}, Y_{t-2}, …, Y_{t-k+1}) is removed (partial out) from both Y_t and Y_{t-k}. A plot of partial auto-correlation for different values of k is called **partial auto-correlation function (PACF)**. Hypothesis tests can be carried out to check whether the auto-correlation and partial auto-correlation values are different from zero. The corresponding null and alternative hypotheses are

H_0: $\rho_k = 0$ and H_A: $\rho_k \neq 0$, where ρ_k is the auto-correlation of order k

H_0: $\rho_{pk} = 0$ and H_A: $\rho_{pk} \neq 0$, where ρ_{pk} is the partial auto-correlation of order k

The null hypothesis is rejected when $|\rho_k| > 1.96/\sqrt{n}$ and $|\rho_{pk}| > 1.96/\sqrt{n}$. To select the appropriate p in the auto-regressive model, the following thumb rule may be used. The number of lags is p when (Yaffee and McGee, 2000)

1. The partial auto-correlation, $|\rho_{pk}| > 1.96/\sqrt{n}$ for first p values (first p lags) and cuts off to zero.
2. The auto-correlation function (ACF), ρ_k, decreases exponentially.

Note that the model identification is an iterative process and may require additional inputs. The model identification using ACF and PACF cannot be taken as conclusive evidence for the number of lags in AR process.

EXAMPLE 13.4

Dr Dawai Sundari (DS) is concerned about the amount of food wasted at her Die Another Day (DAD) hospital. DAD hospital prepares food for all their patients which accounted for 4% of the operating cost. Dr DS found that as high as 50% of the food prepared was wasted on few days. Dr DS believed that an accurate forecasting model will help her reduce the food wastage. The demand for continental breakfast at DAD hospital for past 37 days is shown in Table 13.19. Build an auto-regressive model based on the first 30 days of data and forecast the demand for continental breakfast on days 31 to 37. Comment on the accuracy of the forecast.

TABLE 13.19 Demand for continental breakfast at DAD

Day	Demand CB	Day	Demand CB
1	25	20	43
2	25	21	41
3	25	22	46
4	35	23	41
5	41	24	40
6	30	25	32
7	40	26	41
8	40	27	41
9	40	28	40
10	40	29	43
11	40	30	46
12	40	31	45
13	44	32	45
14	49	33	46
15	50	34	43
16	45	35	40
17	40	36	41
18	42	37	41
19	40		

Solution:

The first step in AR model building is the identification of the right value of p using ACF and PACF plots. ACF and PACF based on the first 30 observations are given in Figures 13.5 and 13.6, respectively. The horizontal lines in the plot represent the upper and lower critical values for ρ_k and ρ_{pk}. The correlation values (vertical bars) beyond the critical values will result in rejection of the null hypothesis.

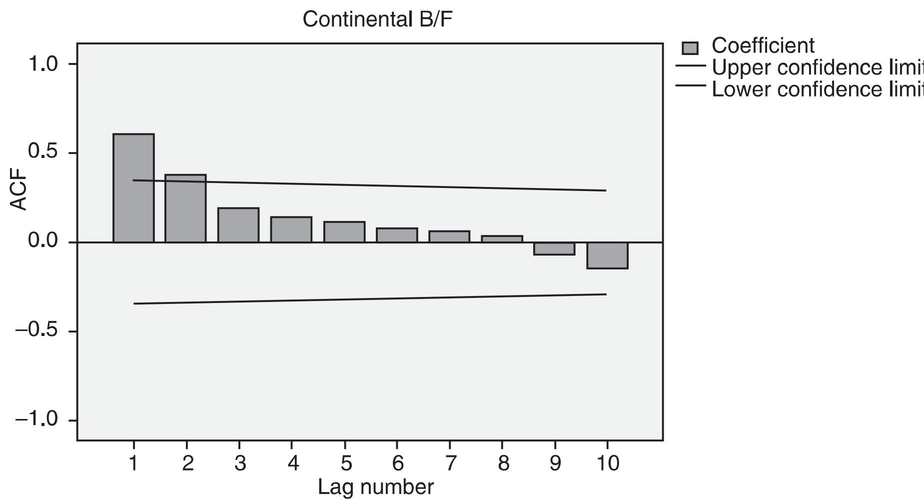


FIGURE 13.5 ACF plot for demand for continental breakfast.

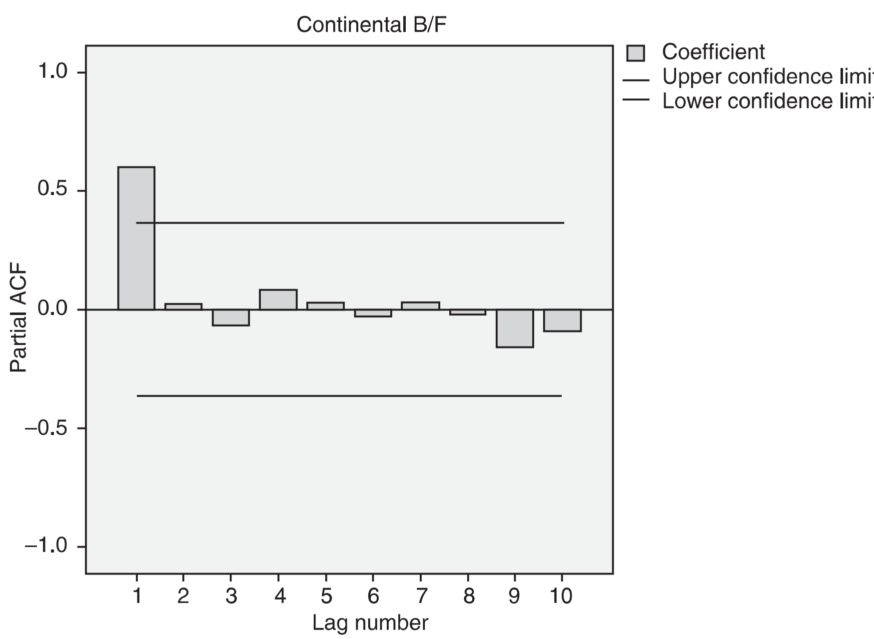


FIGURE 13.6 PACF plot for demand for continental breakfast.

In Figure 13.6, the PACF values cut-off to zero after lag 1 and in Figure 13.5 ACF, the values of auto-correlations are decreasing exponentially. Thus, we can conclude that the value of p in this case is 1 (note that this is a thumb rule, the correct model may be different from what we identify using ACF and PACF plots). Presence of outliers may have series impact on ACF and PACF and thus the identification of lag value. The values of R^2, RMSE, MAPE, and regression parameter estimates of AR(1), using SPSS are shown in Tables 13.20 and 13.21.

TABLE 13.20 AR(1) model statistics

Model	Model Fit Statistics			
	R-Square	RMSE	MAPE	Normalized BIC
Continental B/F-Model_1	0.373	5.133	10.518	3.498

TABLE 13.21 ARIMA model parameters

				Estimate	*SE*	*T*	Sig.
Continental B/F-Model_1	Continental B/F		Constant	38.890	2.995	12.985	0.000
		AR	Lag 1	0.731	0.130	5.616	0.000

The residual ACF and PACF plots are shown in Figure 13.7. It is important that the residual ACF and PACF follow a white noise. Since all the auto- (and partial) correlation values are within the critical values, we can conclude that residuals are uncorrelated. The normal P-P plot is shown in Figure 13.8 showing approximate normal distribution for residuals. Since the residual follows white noise as evident from Figures 13.7 and 13.8, we can accept the AR(1) model. Also, the p-value for lag 1 term in Table 13.21 is less than 0.05. The AR(1) model based on Table 13.21 is given by

$$(F_{t+1} - 38.890) = 0.731(Y_t - 38.890) \tag{13.39}$$

The following equation can be used when we need to forecast several periods in future using forecasted values repetitively.

$$(F_{t+k} - 38.890) = 0.731(F_{t+k-1} - 38.890) \tag{13.40}$$

Using Eq. (13.39), the forecasted value for F_{31} is given by

$$(F_{31} - 38.890) = 0.731\ (Y_{30} - 38.890)$$

$$\Rightarrow F_{31} = 38.890 + 0.731\ (46 - 38.890) = 44.08$$

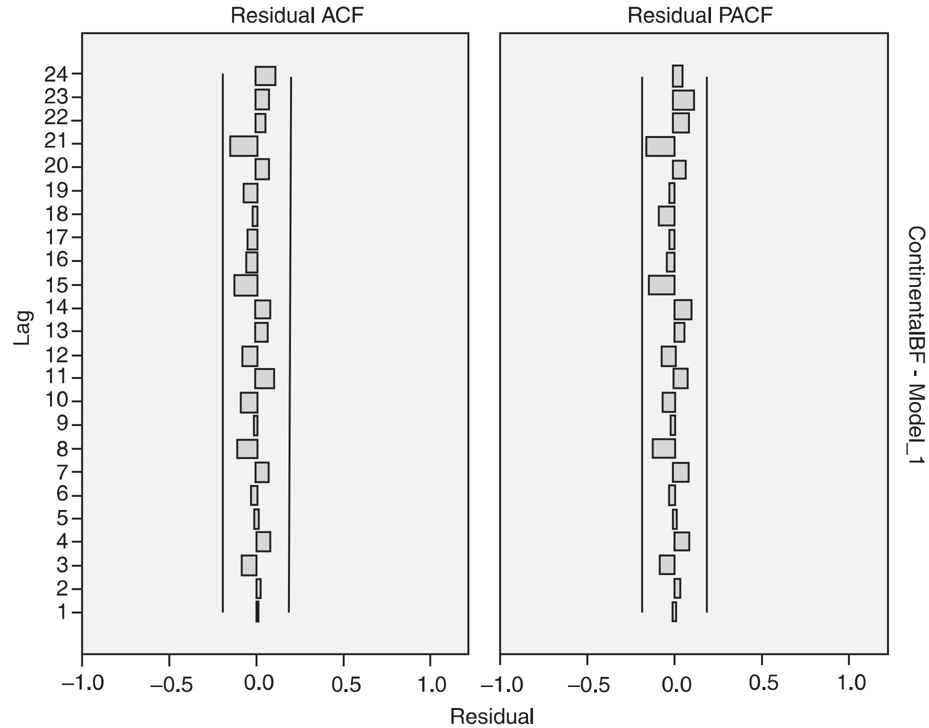


FIGURE 13.7 ACF and PACF plot of residuals.

The forecasted values for periods 31–37 and RMSE and MAPE calculations are shown in Table 13.22.

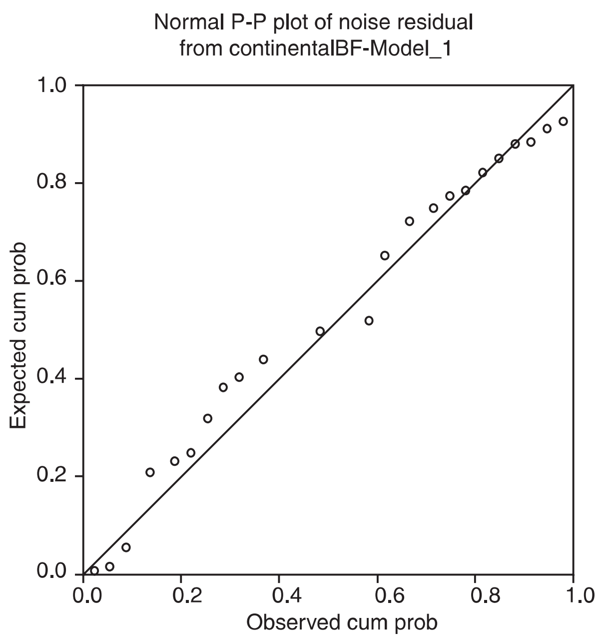


FIGURE 13.8 Normal P-P plot of residuals.

TABLE 13.22 AR(1) model forecast

Day	Y_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
31	45	44.08741	0.832821	0.02028
32	45	43.35641	2.701388	0.036524
33	46	43.35641	6.988568	0.057469
34	43	44.08741	1.182461	0.025289
35	40	41.89441	3.588789	0.04736
36	41	39.70141	1.686336	0.031673
37	41	40.43241	0.322158	0.013844

The RMSE and MAPE for the validation data (days 31 and 37) are 1.5721 and 0.0332 (3.32%), respectively. Applying Eq. (13.40) that is, using the F_t values to forecast F_{t+1} instead of Y_t, we get the values in Table 13.23.

TABLE 13.23 AR(1) model forecast using equation 13.40

Day	Y_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
31	45	44.0874	0.8328	0.0203
32	45	42.6893	5.3393	0.0513
33	46	41.6673	18.7723	0.0942
34	43	40.9202	4.3256	0.0484
35	40	40.3741	0.1399	0.0094
36	41	39.9749	1.0509	0.0250
37	41	39.6830	1.7344	0.0321

The RMSE and MAPE for the validation data (days 31 to 37) using Eq. (13.40) are 2.1446 and 0.04009 (4.009%), respectively.

13.12 | MOVING AVERAGE PROCESS MA(q)

Moving average (MA) processes are regression models in which the past residuals are used for forecasting future values of the time-series data. Moving average process is different from moving average technique discussed in Section 13.4, except that the regression model of MA process can be considered as weighted moving average of past residuals. Moving average process of lag 1, MA(1), is given by

$$Y_{t+1} = \mu + \alpha_1\varepsilon_t + \varepsilon_{t+1} \tag{13.41}$$

Alternatively, a moving average process of lag 1 can be written as

$$Y_{t+1} = \alpha_1\varepsilon_1 + \varepsilon_{t+1} \tag{13.42}$$

MA(1) process uses the previous residual, ε_t, to forecast the next value of the time series. The reasoning behind MA process is that the error (also called shock or innovation) at the current period, ε_t, and the error at the next period, ε_{t+1}, drive the next value of the time series Y_{t+1}. A moving average process with q lags, MA(q) process, is given by

$$Y_{t+1} = \mu + \alpha_1\varepsilon_t + \alpha_2\varepsilon_{t-1} + \cdots + \alpha_q\varepsilon_{t-q+1} + \varepsilon_{t+1} \tag{13.43}$$

The value of q (number of lags) in a moving average process can be identified using the following rule (Yaffee and McGee, 2000):

1. Auto-correlation value, $|\rho_p| > 1.96/\sqrt{n}$ for first q values (first q lags) and cuts off to zero.
2. The partial auto-correlation function, ρ_{pk}, decreases exponentially.

13.13 | AUTO-REGRESSIVE MOVING AVERAGE (ARMA) PROCESS

Auto-regressive moving average (ARMA) is a combination auto-regressive and moving average process. ARMA(p, q) process combines AR(p) and MA(q) processes. ARMA(p, q) model is given by

$$Y_{t+1} = \overbrace{\beta_1 Y_t + \beta_2 Y_{t-1} + \ldots + \beta_p Y_{t-p+1}}^{\text{Auto Regressive Part}} + \overbrace{\alpha_1\varepsilon_t + \alpha_2\varepsilon_{t-1} + \ldots + \alpha_q\varepsilon_{t-q+1}}^{\text{Moving Average Part}} + \varepsilon_{t+1} \tag{13.44}$$

The parameters are estimated using Box–Jenkins methodology. The values of p and q in a ARMA process can be identified using the following thumb rule:

1. Auto-correlation value, $|\rho_p| > 1.96/\sqrt{n}$ for first q values (first q lags) and cuts off to zero.
2. Partial auto-correlation function, $|\rho_{pk}| > 1.96/\sqrt{n}$ for first p values and cuts off to zero.

EXAMPLE 13.5

Monthly demand for avionic system spares used in Vimana 007 aircraft is provided in Table 13.24. Build an ARMA model based on the first 30 months of data and forecast the demand for spares for months 31 to 37. Comment on the accuracy of the forecast.

TABLE 13.24 Demand for avionic spare parts

Month	Demand for Spares	Month	Demand for Spares
1	457	20	516
2	439	21	656
3	404	22	558
4	392	23	647
5	403	24	864
6	371	25	610
7	382	26	677
8	358	27	609
9	594	28	673
10	482	29	400
11	574	30	443
12	704	31	503
13	486	32	688
14	509	33	602
15	537	34	629
16	407	35	823
17	523	36	671
18	363	37	487
19	479		

Solution:

The first step in building ARMA model is the model identification using ACF and PACF plots. The ACF and PACF plots based on the first 30 months demand data are given in Figures 13.9 and 13.10.

In Figure 13.9, the auto-correlations cuts off to zero after 2 lags (note that auto-correlation of lag 3 is just below the critical value). The PACF value cuts off to zero after the first lag. So, the appropriate model could be ARMA(1, 2) process. The SPSS output for ARMA(1, 2) process is given in Tables 13.25 and 13.26.

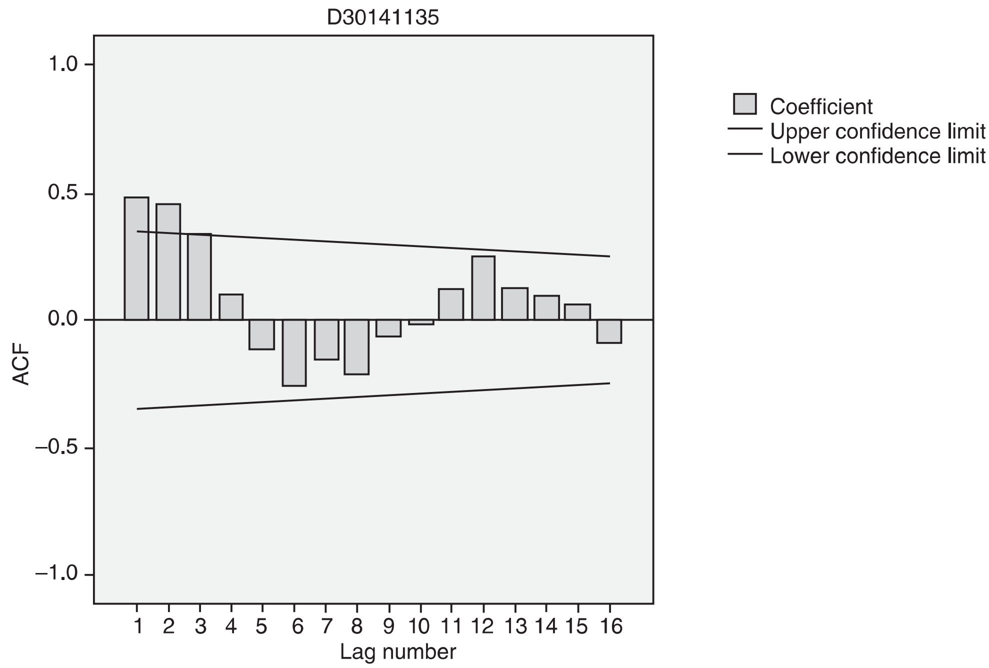


FIGURE 13.9 ACF plot for avionic system spares demand.

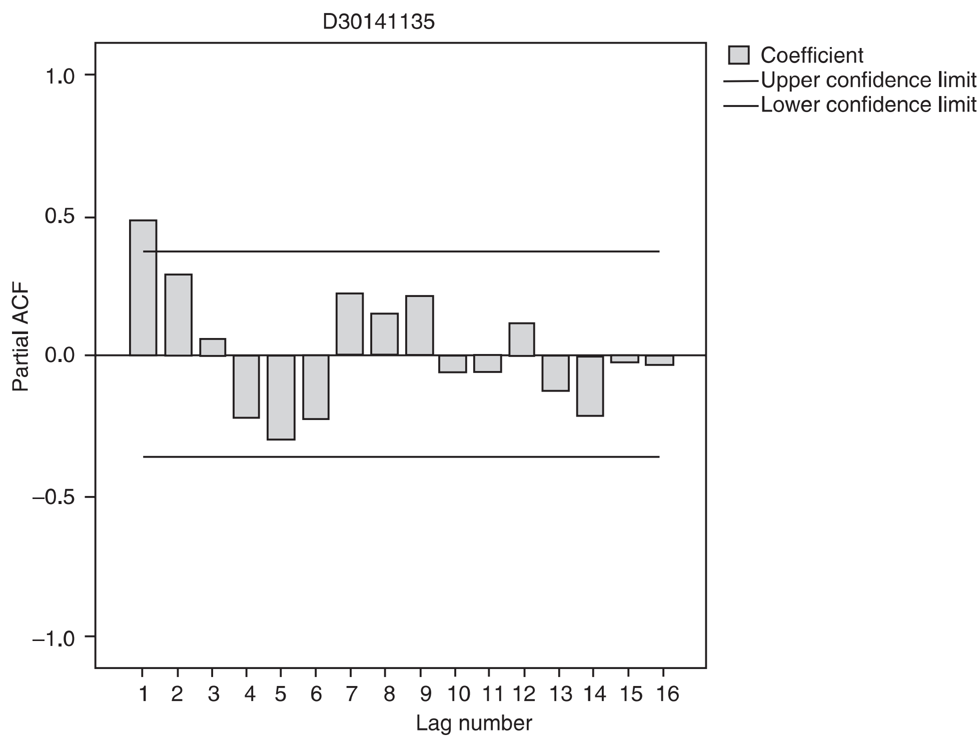


FIGURE 13.10 PACF plot for avionic system spares demand.

TABLE 13.25 Summary statistics

Model	Model Fit Statistics		
	Stationary *R*-Squared	RMSE	MAPE
Avionic Spares	0.429	98.824	14.231

TABLE 13.26 ARIMA model parameters

			Estimate	*SE*	*T*	Sig.
Avionic Spares	Constant		496.699	57.735	8.603	0.000
	AR	Lag 1	0.706	0.170	4.153	0.000
	MA	Lag 1	0.694	0.173	4.006	0.000
		Lag 2	−0.727	0.170	−4.281	0.000

All the three components in the ARMA model (AR lag 1 and MA lags 1 and 2) are statistically significant (Table 13.26). The model equation using SPSS is given by

$$Y_{t+1} - 496.669 = 0.706 \times (Y_t - 496.699) - 0.694 \times \varepsilon_t + 0.727 \times \varepsilon_{t-1} \tag{13.45}$$

The plot of forecasted and actual values for demand for avionic spare parts based on the model in Eq. (13.45) is given in Figure 13.11. The plot of residual ACF and PACF is shown in Figure 13.12 (it is evident that the residual are white noise).

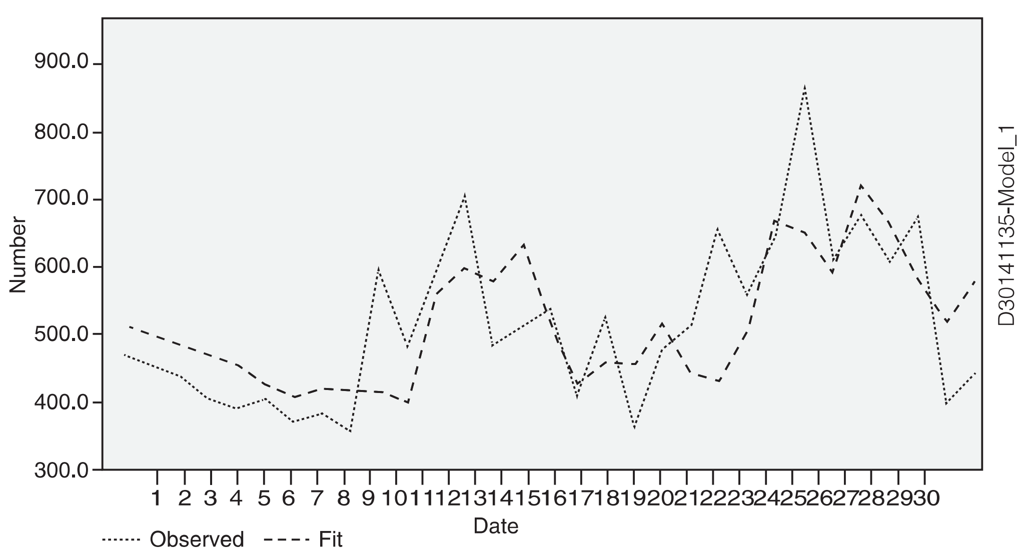


FIGURE 13.11 Observed versus forecasted demand.

TABLE 13.27 ARMA(1, 2) model forecast

Month	Y_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
31	503	464.8107	1458.423	0.075923
32	688	378.5341	95769.15	0.449805
33	602	444.6372	24763.04	0.2614
34	629	685.8851	3235.909	0.090437
35	823	743.5124	6318.281	0.096583
36	671	630.7183	1622.614	0.060032
37	487	649.3491	26357.22	0.333366

The RMSE and MAPE for the validation data (months 31 and 37) are 150.961 and 0.1953 (19.53%), respectively (Table 13.27).

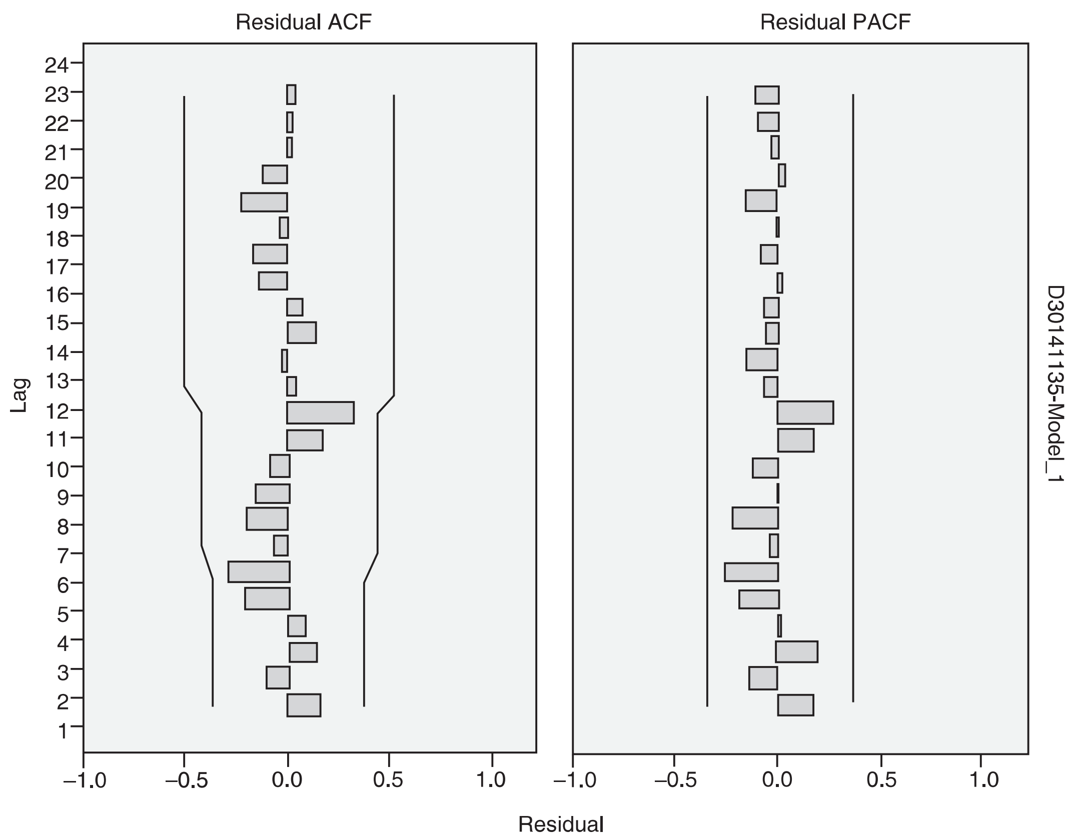


FIGURE 13.12 ACF and PACF of residuals.

The forecasted values using F_t instead of Y_t when forecasting for more than one period ahead in time are shown in Table 13.28.

TABLE 13.28 ARMA (1, 2) forecast

Month	Y_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
31	503	464.4239	1488.1147	0.0767
32	688	377.8374	96200.8258	0.4508
33	602	444.5195	24800.1101	0.2616
34	629	687.2082	3388.1980	0.0925
35	823	744.9583	6090.4998	0.0948
36	671	630.5592	1635.4571	0.0603
37	487	648.3959	26048.6313	0.3314

The RMSE and MAPE for the validation data (months 31 and 37) are 151.02 and 0.1954 (19.54%), respectively.

13.14 | AUTO-REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) PROCESS

ARMA models can be used only when the time-series data is stationary. ARIMA models are used when the time-series data is non-stationary. ARIMA model was proposed by Box and Jenkins (1970) and thus is also known as Box–Jenkins methodology. ARIMA has the following three components and is represented as ARIMA(p, d, q):

1. Auto-regressive component with p lags AR(p).
2. Integration component (d).
3. Moving average with q lags, MA(q).

The main objective of integration component is to convert a non-stationary time-series process to stationary process so that AR and MA processes can be used for forecasting. When the data is non-stationary, the auto-correlation function will not be cut-off to zero quickly; rather ACF may show a very slow decline. When the time series is non-stationary, the model parameter values will be greater than or equal to one resulting in non-convergence of the series and this causes a slow decrease in the values of ACF and PACF. Thus, the presence of non-stationarity can be identified by plotting ACF. A slow decrease in ACF can be diagnosed as non-stationary process. Sample ACF plot for non-stationary data is shown in Figure 13.13.

The non-stationarity could arise from deterministic or stochastic trend; identification of the source of non-stationarity will be useful for using appropriate transformation to make the process stationary. In addition to the visual evidence from ACF plot, Dickey–Fuller or augmented Dickey–Fuller tests are used to check the presence of stationarity.

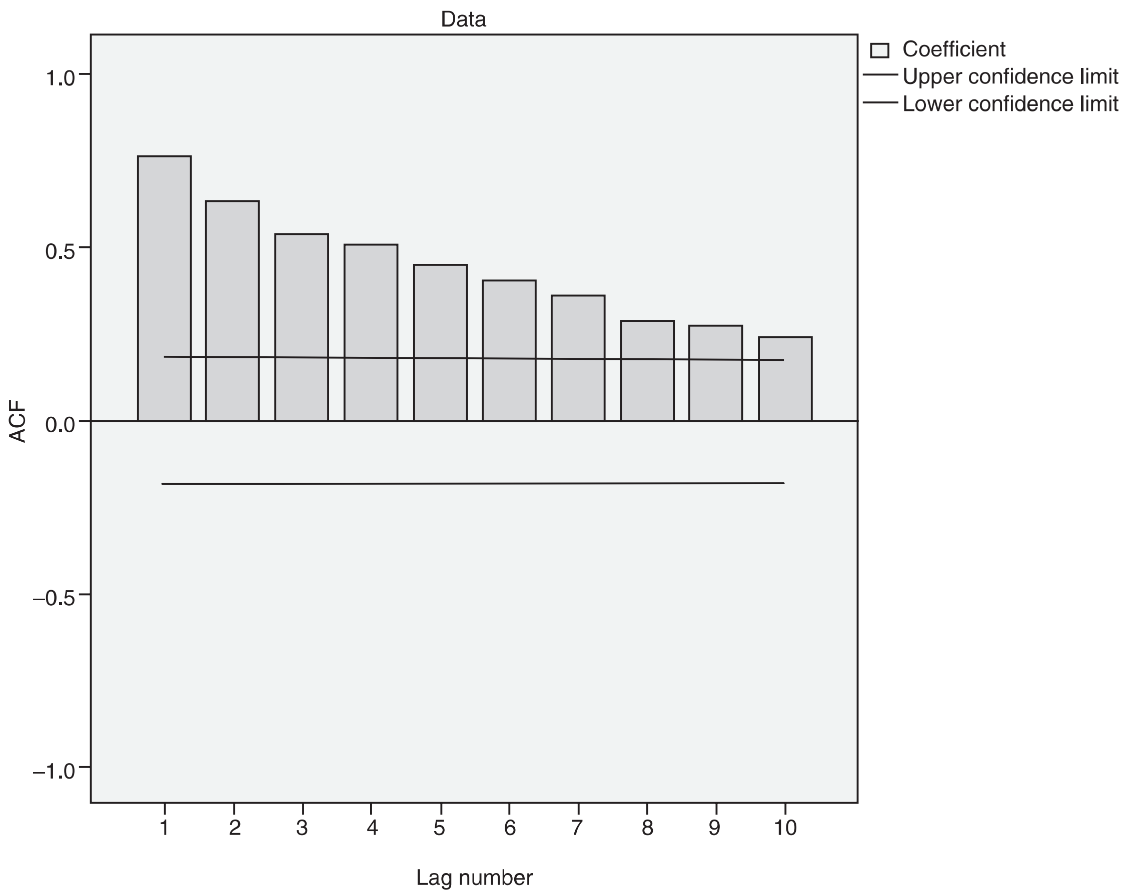


FIGURE 13.13 ACF of a non-stationary process (slowly decreasing auto-correlation values).

13.14.1 | Dickey Fuller Test

Consider AR(1) process defined below:

$$Y_{t+1} = \beta Y_t + \varepsilon_{t+1}$$

In Section 13.11, we proved that the AR(1) process can become very large when $\beta > 1$ and is non-stationary when $|\beta| = 1$. Dickey–Fuller test (Dickey and Fuller, 1979) is a hypothesis test in which the null hypothesis and alternative hypothesis are given by

H_0: $\beta = 1$ (the time series is non-stationary)

H_1: $\beta < 1$ (the time series is stationary)

The AR(1) can be written as

$$Y_{t+1} - Y_t = \Delta Y_t = (\beta - 1)Y_t + \varepsilon_{t+1} = \psi Y_t + \varepsilon_{t+1} \tag{13.46}$$

In Eq. (13.46), $\psi = 0$ is same as $\beta = 1$. So, the Dickey–Fuller test can be written in terms of ψ as

$$H_0: \psi = 0 \text{ (the time series is non-stationary)}$$
$$H_A: \psi < 0 \text{ (the time series is stationary)}$$

The test statistic is given by

$$\text{DF Test Statistic} = \frac{\psi}{S_e(\psi)} \tag{13.47}$$

where S_e is the standard error. Note that DF test statistic is not t-statistic since the null hypothesis is on non-stationary process. Critical values are derived based on simulation (Fuller, 1976).

13.14.2 | Augmented Dickey–Fuller Test

Dickey–Fuller test is valid only when the residual ε_{t+1} follows a white noise. When ε_{t+1} is not white noise, the actual series may not be AR(1); it may have more significant lags. To address this issue, we augment p-lags of the dependent variable Y. The model in Eq. (13.46) can be written as

$$\Delta Y_t = \psi Y_t + \sum_{i=0}^{p} \alpha_i \Delta Y_{t-i} + \varepsilon_{t+1} \tag{13.48}$$

The above equation can be now tested for non-stationarity. Again the null and alternative hypotheses are

H_0: $\psi = 0$ (the time series is non-stationary)
H_A: $\psi < 0$ (the time series is stationary)

13.14.3 | Transforming Non-Stationary Process to Stationary Process Using Differencing

The first step in ARIMA is to identify the order of differencing (d) required to convert a non-stationary process into a stationary process. Many time-series data will be non-stationary due to factors such as trend and seasonality. If the non-stationary behaviour is due to trend, then it can be converted into a stationary process by de-trending the data. De-trending is usually achieved by fitting a trend line and subtracting it from the time series; this is known as **trend stationarity**. When the reason is not due to trend stationarity, then differencing the original time series may be useful for converting the non-stationary process into a stationary process (called **difference stationarity**).

The first difference ($d = 1$) is the difference between consecutive values of the time series (Y_t and Y_{t-1}). That is, the first difference ΔY_t is given by

$$\nabla Y_t = Y_t - Y_{t-1} \tag{13.49}$$

The second difference ($d = 2$) is the difference of the first differences and is given by

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = Y_t - 2Y_{t-1} + Y_{t-2} \tag{13.50}$$

In most cases, $d \le 2$ will be sufficient to convert a non-stationary process to a stationary process.

13.14.4 | ARIMA(*p, d, q*) Model Building

The first step in ARIMA(p, d, q) is the model identification, that is, identifying the values of $p, d,$ and q. Box and Jenkins (1970) proposed the following procedure to build the ARIMA(p, d, q) model.

STAGE 1 *Model Identification*

The main objective of model identification stage is to identify the right values of p (auto-regressive lags), d (order of differencing), and q (moving average lags). The following flow chart can be used during the model identification stage (Figure 13.14). The first step is to plot the ACF and PACF to identify whether the time series is stationary or not. If the time series is stationary then $d = 0$ and the model is ARIMA(p, 0, q) or ARMA(p, q) model. If the time series is non-stationary then it has to be converted into a stationary process by identifying the order of differencing. Once the value of d is known that will make the process stationary, then p and q are identified for the stationary process.

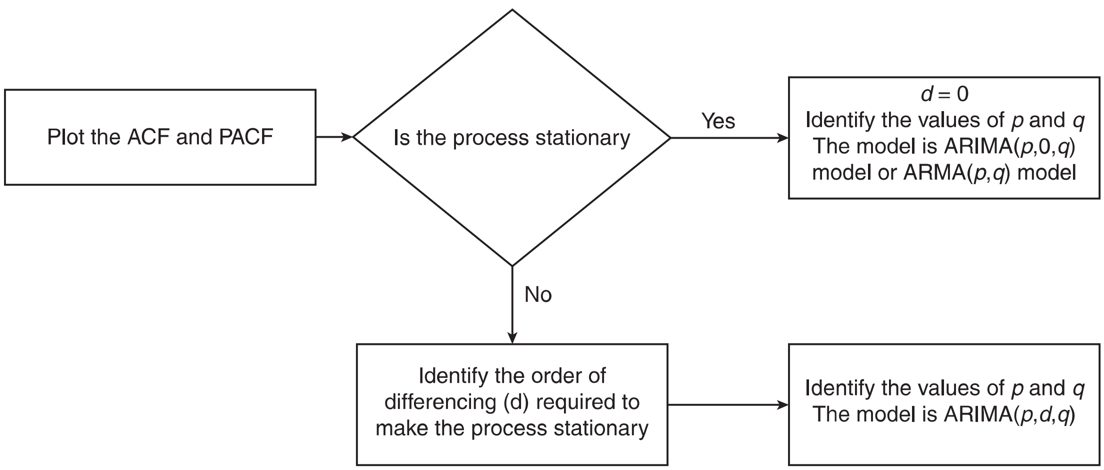


FIGURE 13.14 Model identification in ARIMA model.

STAGE 2 *Parameter Estimation and Model Selection*

Once the model is identified (values of p, d, and q), the next step in ARIMA model building is the parameter estimation. That is, the estimation of coefficients in AR and MA components which are achieved using ordinary least squares. The model selection may be carried using several criteria such as RMSE, MAPE, Akaike Information Criteria (AIC), or Bayesian Information Criteria (AIC).

AIC and BIC are measures of distance from the actual values to the forecasted values. AIC is given by

$$\text{AIC} = -2LL + 2K \tag{13.51}$$

where LL is the log likelihood function and K is the number of parameters estimated (in this case $p + q$). BIC is given by

$$\text{BIC} = -2LL + K\ln(n) \tag{13.52}$$

In BIC equation, n is the number of observations in the sample. BIC assigns higher penalty compared to AIC for every additional variable added to the model. Lower values of AIC and BIC are preferred.

STAGE 3 *Model Validation*

ARIMA model is a regression model and thus has to satisfy all the assumptions of regression. The residual should be white noise. We can also perform a goodness of fit test using Ljung–Box test before accepting the model.

EXAMPLE 13.6

Daily demand for Omelette at Die Another Day (DAD) hospital for the past 115 days is given in the excel sheet Example 13.6.xlsx. Develop an appropriate ARIMA model that DAD hospital can use for forecasting demand for Omelette.

Solution:

The time-series plot of the daily demand for Omelette is shown in Figure 13.15. The corresponding ACF plot is shown in Figure 13.16. From Figure 13.15, it is evident that the mean is not constant for different values of t.

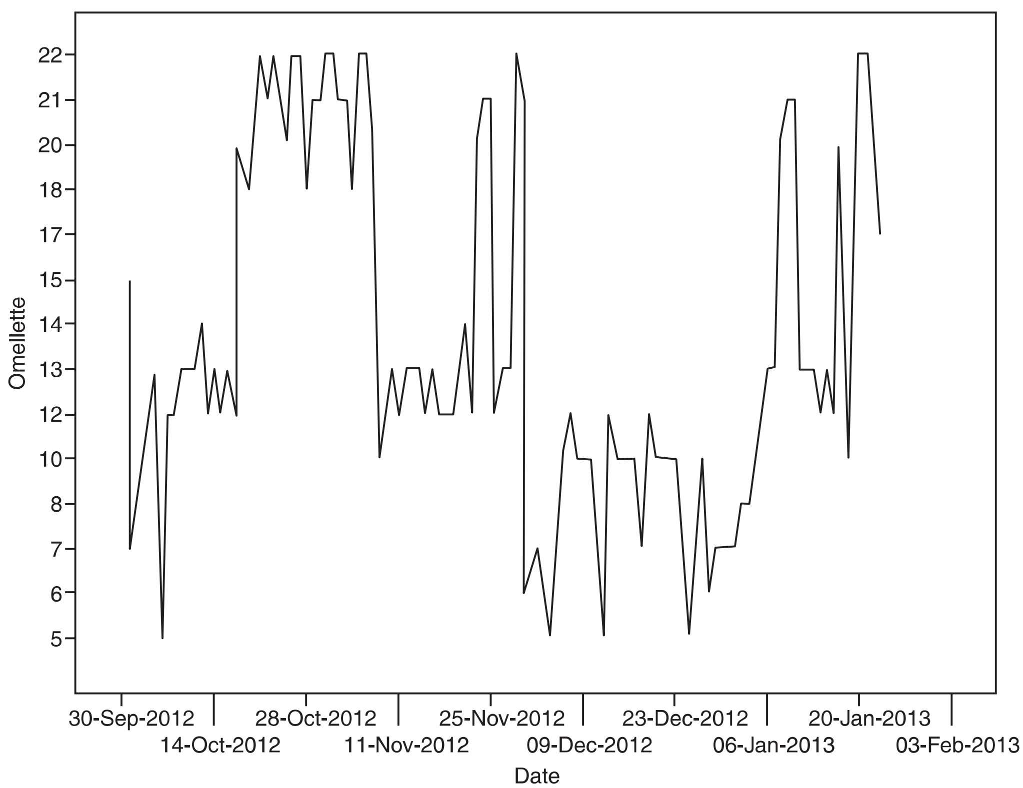


FIGURE 13.15 Time-series plot of demand for Omelette at DAD hospital.

Since the ACF plot shows a very slowly decreasing pattern, we may conclude that the time series is not stationary. We have to convert the process to a stationary process before we can develop a forecasting model. The ACF and PACF plots after differencing ($d = 1$) are shown in Figures 13.17 and 13.18, respectively.

Since both ACF and PACF values are cutting off to zero after the first difference, we may conclude that the appropriate model is ARIMA(1,1,1). Note that subsequent

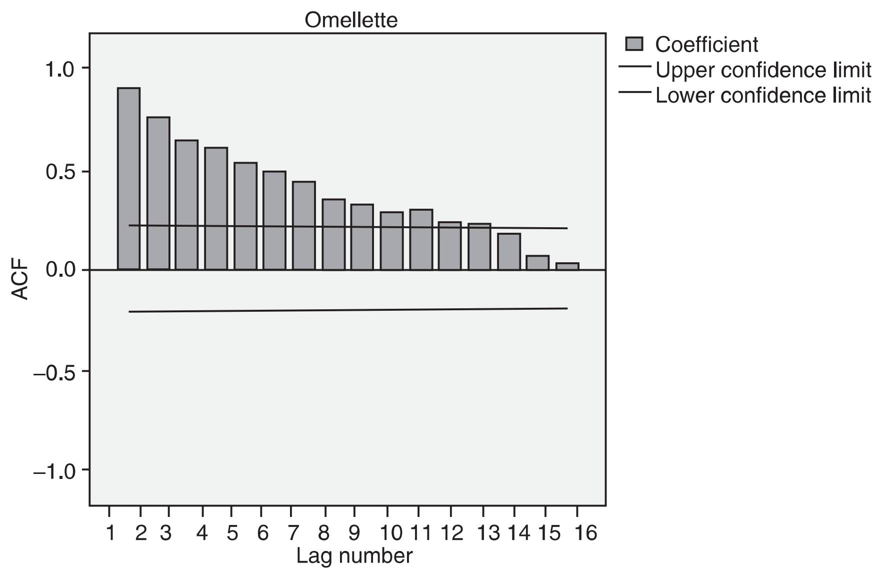


FIGURE 13.16 ACF plot of demand for Omelette at DAD hospital.

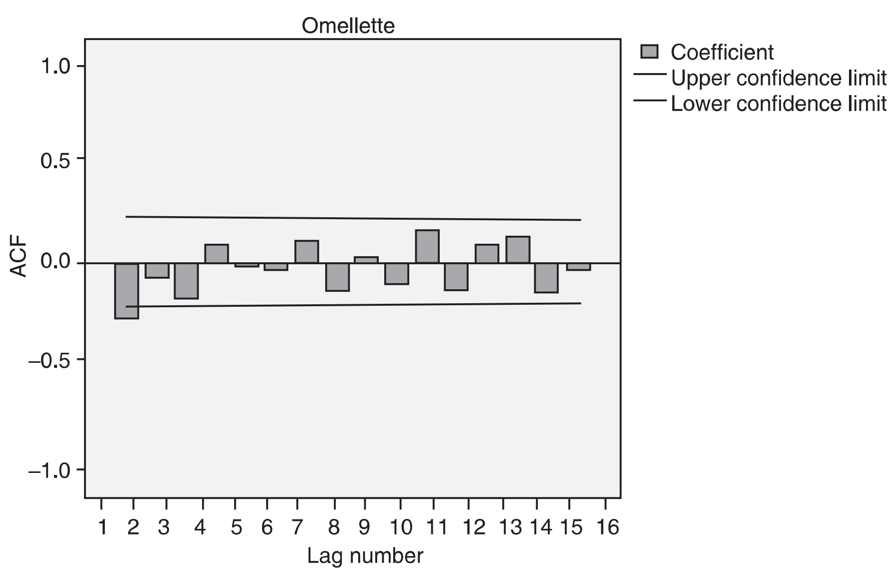


FIGURE 13.17 ACF plot of demand for Omelette after differencing ($d = 1$).

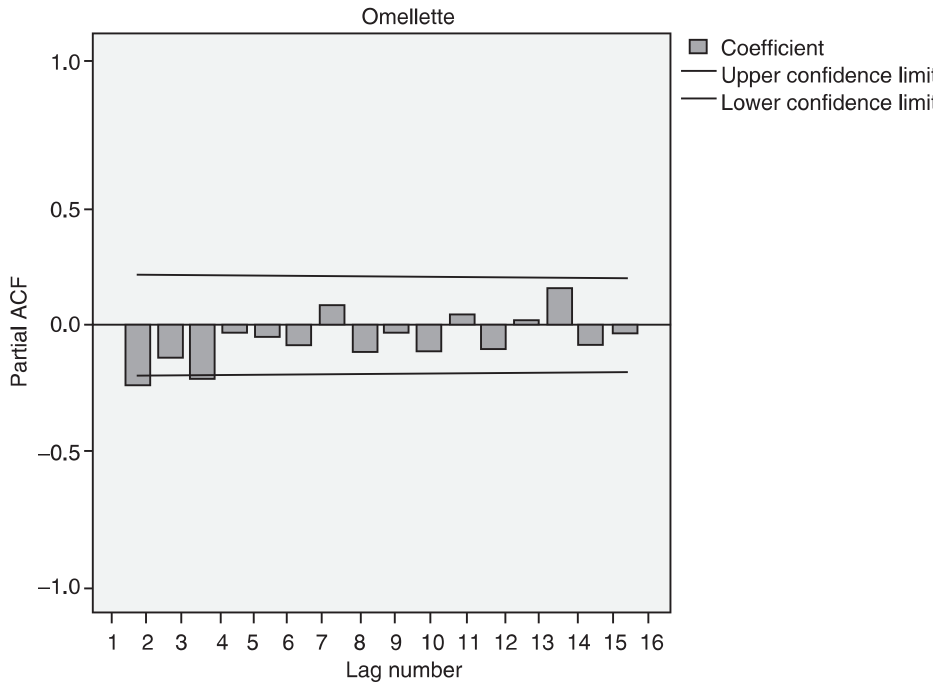


FIGURE 13.18 PACF plot of demand for Omelette after differencing ($d = 1$).

correlations once it cuts off to zero is not useful and we will ignore them (for example, in PACF plot in Figure 13.17, the partial auto-correlation value with lag 3 is beyond the critical line). However, if the data has seasonal fluctuations, then ACF plot may show consistent spikes as per the periodicity of the seasonal variation. The ARIMA(1, 1, 1) model summary and parameter estimates are shown in Tables 13.29 and 13.30.

TABLE 13.29 ARIMA(1, 1, 1) model summary for Omelette demand

Model	Model Fit Statistics			Ljung–Box *Q*(18)		
	R-Squared	RMSE	MAPE	Statistics	*Df*	Sig.
Omellette-Model_1	0.584	3.439	20.830	10.216	16	0.855

TABLE 13.30 ARIMA model parameters

			Estimate	*SE*	*T*	Sig.
Omellette-Model_1	Constant		0.055	0.137	0.402	0.689
	AR	Lag 1	0.439	0.178	2.475	0.015
	Difference		1			
	MA	Lag 1	0.767	0.128	6.004	0.000

AR and MA components in Table 13.25 are statistically significant since the corresponding p-values are less than 0.05. The ACF and PACF of residuals are shown in Figure 13.19 which shows white noise of residuals.

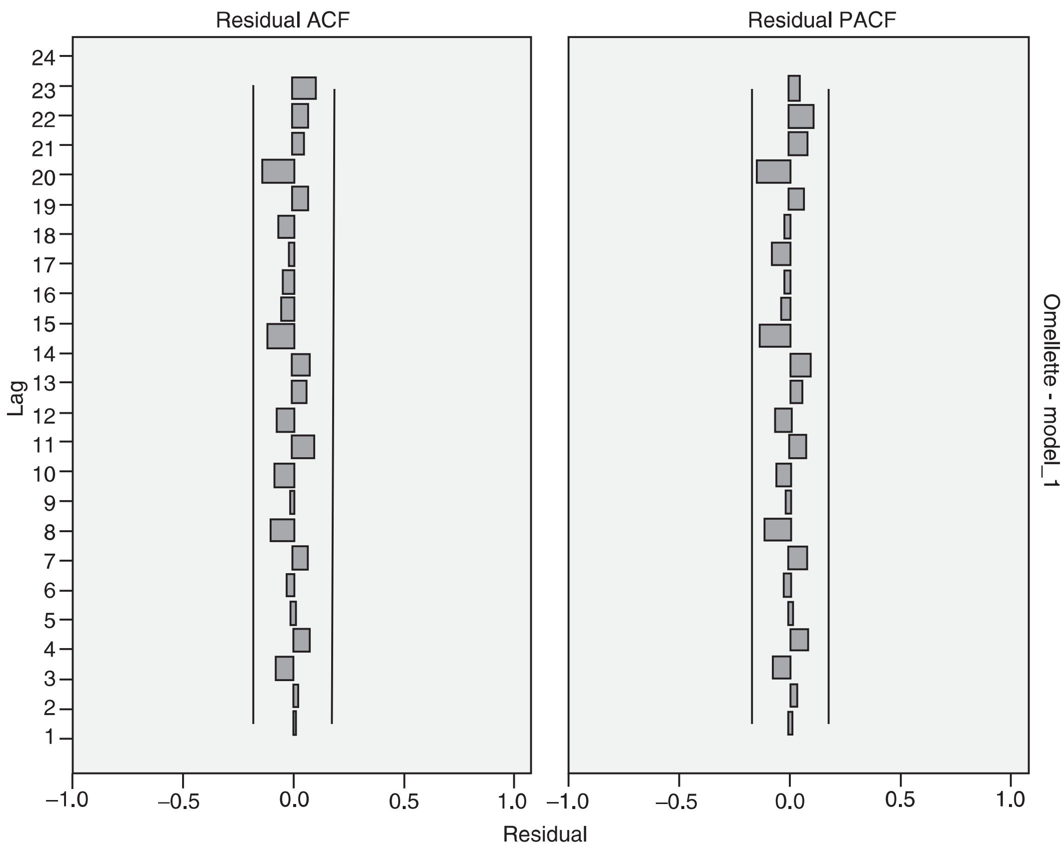


FIGURE 13.19 ACF and PACF of residuals.

Since the residuals follow white noise, we can use ARIMA(1, 1, 1) model for forecasting.

13.14.5 | Ljung–Box Test for Auto-Correlations

Ljung–Box is a test of lack of fit of the forecasting model and checks whether the auto-correlations for the errors are different from zero. The null and alternative hypotheses are given by

H_0: The model does not show lack of fit
H_1: The model exhibits lack of fit

The Ljung–Box statistic (Q-Statistic) is given by (Ljung and Box, 1978)

$$Q(m) = n(n+2)\sum_{k=1}^{m}\frac{\rho_k^2}{n-k} \tag{13.53}$$

where n is the number of observations in the time series, k is the number of lag, ρ_k is the auto-correlation of lag k, and m is the total number of lags. Q-statistic is an approximate chi-square distribution with $m - p - q$ degrees of freedom where p and q are the AR and MA lags. The Q-statistic for ARIMA(1, 1, 1) is 10.216 (Table 13.29) and the corresponding p-value is 0.855 and thus we fail to reject the null hypothesis. $Q(m)$ measures accumulated auto-correlation up to lag m.

13.15 | POWER OF FORECASTING MODEL: THEIL'S COEFFICIENT

The power of forecasting model is a comparison between Naïve forecasting model and the model developed. In the Naïve forecasting model, the forecasted value for the next period is same as the last period's actual value ($F_{t+1} = Y_t$). Theil's coefficient (U-statistic) is given by (Theil, 1965)

$$\mathrm{U} = \frac{\sum_{t=1}^{n}(Y_{t+1} - F_{t+1})^2}{\sum_{t=1}^{n}(Y_{t+1} - Y_t)^2} \tag{13.54}$$

Theil's coefficient is the ratio of the mean squared error of the forecasting model to the MSE of the Naïve model. The value of $U < 1$ indicates that forecasting model is better than the Naïve forecasting model. $U > 1$ indicates that the forecasting model is not better than Naïve model. For the data shown in Table 13.14 (demand for avionic system spares), the U-statistic calculations are shown in Table 13.31.

TABLE 13.31 U-statistic calculation

Day	Y_t	ARMA (1,2) Forecast	$(Y_t - F_t)^2$	Naïve Forecast ($F_{t+1} = Y_t$)	$(Y_t - F_t)^2$
31	503	464.8107	1458.423	443	3600
32	688	378.5341	95769.15	503	34225
33	602	444.6372	24763.04	688	7396
34	629	685.8851	3235.909	602	729
35	823	743.5124	6318.281	629	37636
36	671	630.7183	1622.614	823	23104
37	487	649.3491	26357.22	671	33856
		Total	159524.6	Total	140546

The U-statistic value = 159524.6 / 140546 = 1.1350. That is, ARMA(1, 2) model is not better than Naïve forecasting.

SUMMARY

1. Forecasting is one of the important tasks carried out using analytics by many organizations since accurate forecasting is important for taking several decisions such as man-power planning, materials requirement planning, budgeting, and supply chain related issues.
2. Forecasting is carried out on a time-series data in which the dependent variable Y_t is observed at differ time periods *t*.
3. Several techniques such moving average, exponential smoothing, and auto-regressive models are used for forecasting future value of Y_t.
4. The forecasting models are validated using accuracy measures such as RMSE, MAPE, AIC, and BIC.
5. Simple techniques such as moving average and exponential smoothing may outperform complex regression based models in certain scenarios. Thus, it is important to develop forecasting models using several techniques before selecting the final model.
6. Regression model in the presence of independent variables may outperform other techniques.
7. Auto Regressive (AR) models are regression based models in which dependent variable is Y_t and the independent variables are Y_{t-1}, Y_{t-2}, etc.
8. AR models can be used only when the data is stationary.
9. Moving average (MA) models are regression models in which the independent variables are past error values.
10. Auto-regressive integrated moving average (ARIMA) has 3 components: Auto-regressive component with *p* lags – AR(*p*), moving average component with *q* lags – MA(*q*), and integration which is differencing the original data to make it stationary.
11. One of the necessary conditions of acceptance of ARIMA model is that the residuals should follow white noise.
12. In ARIMA, the model identification, that is identifying the value of *p* in AR and *q* in MA, is achieved through auto-correlation function (ACF) and partial auto-correlation function (PACF).
13. The stationarity of time-series data is usually checked using Dickey–Fuller and Augmented Dickey Fuller test.
14. The overall model accuracy of forecasting model is tested using Ljung–Box test.

MULTIPLE CHOICE QUESTIONS

1. Seasonality in time-series data is caused due to
   (a) Changes in macro-economic factors such as recession, unemployment, and so on
   (b) Festivals and customs in a society
   (c) Random events that occur over a period of time
   (d) Changes in customer behaviour driven by new products and promotions
2. In a simple exponential smoothing method, the low value of smoothing constant α is chosen when
   (a) The data has high fluctuations around the trend line
   (b) There is seasonality in the data
   (c) The data is smooth with low fluctuations
   (d) There are variations in the data due to cyclical component
3. White noise is
   (a) Uncorrelated errors with expected value 0.
   (b) Uncorrelated errors that are constant and do not change with time.
   (c) Uncorrelated errors that follow normal distribution with mean 0 and constant standard deviation
   (d) Errors that follow normal distribution with constant mean and standard deviation

4. A stationary process in a time series is a process for which
   (a) Mean and variance are constant at different time points
   (b) The time series follows normal distribution with zero mean and constant standard deviation
   (c) The covariance of the time series depends only on the lag
   (d) Mean and standard deviation are constant at different time points and the covariance depends only on the lag between the values and is constant for a given lag
5. In a pure auto-regressive process, AR(p), the value of p can be identified using
   (a) Auto-correlation function (b) Partial auto-correlation function
   (c) Auto-correlation and partial auto-correlation function (d) Ljung–Box test
6. Power of a forecasting model is calculated using
   (a) Root mean square error (RMSE) (b) Theil's coefficient
   (c) Mean absolute percentage error (MAPE) (d) Bayesian information criteria (BIC)
7. A necessary condition for accepting a time-series forecasting model is
   (a) The residuals should follow a normal distribution
   (b) The residuals should be white noise
   (c) The residuals should be black noise
   (d) The residuals should follow a normal distribution and the R-square should be high
8. In an ARIMA model, differencing is carried out
   (a) To convert a stationary process to a non-stationary process
   (b) To convert a non-stationary process to a stationary process
   (c) To remove seasonal fluctuations from the data
   (d) To remove cyclical fluctuations from the data
9. Overall fitness of a forecasting model is checked using
   (a) Durbin–Watson Test (b) Theil coefficient (c) Ljung–Box test (d) Dickey–Fuller test
10. Presence of non-stationarity is checked using
   (a) Durbin–Watson Test (b) Theil coefficient (c) Ljung–Box test (d) Dickey–Fuller test

EXERCISES

1. Quarterly demand for certain parts manufactured by Jack and Jill company is shown in Table 13.32.

TABLE 13.32 Quarterly demand

Year	Quarter	Value
2012	Q1	75
	Q2	60
	Q3	54
	Q4	59
2013	Q1	86
	Q2	65
	Q3	63
	Q4	80
2014	Q1	90

(*Continued*)

TABLE 13.32 Quarterly demand—Continued

Year	Quarter	Value
	Q2	72
	Q3	66
	Q4	85
2015	Q1	100
	Q2	78
	Q3	72
	Q4	93

(a) Calculate the seasonality index for different quarters using the first 3 years of data.

(b) Develop forecasting models using moving average, single exponential smoothing, and an appropriate ARMA model after de-seasonalizing the data (assume multiplicative model, $Y_t = T_t \times S_t$).

(c) Forecast the demand for 2015 (all four quarters) using moving average, exponential smoothing, and ARMA. Calculate RMSE, MAPE, and Theil's coefficient.

2. Data on monthly demand for a product over 3 years (between 2013 and 2015) is given in Table 13.33.

TABLE 13.33 Monthly demand

Month	2013	2014	2015
January	15	23	25
February	16	22	25
March	18	28	35
April	18	27	36
May	23	31	36
June	23	28	30
July	20	22	30
August	28	28	34
September	29	32	38
October	33	37	47
November	33	34	41
December	38	44	53

(a) Calculate the seasonality index using methods of averages.

(b) De-seasonalize the data assuming that Y_t is product of trend and seasonality.

(c) Develop the best forecasting model by comparing MAPE of MA, ES, and ARMA models. Compare the models using MAPE and Theil's coefficient.

3. Television rating points of a television program over 30 episodes is shown in Table 13.34.

TABLE 13.34 Television rating points

Episode	1	2	3	4	5	6	7	8	9	10
TRP	7.98	9.8	9.53	7.23	7.34	9.62	9.8	7.9	8.26	8.17
Episode	11	12	13	14	15	16	17	18	19	20
TRP	8.36	8.5	9.03	9.82	9.77	10.77	9.46	9.31	10.32	9.03
Episode	21	22	23	24	25	26	27	28	29	30
TRP	10.22	10.28	11.99	11.21	9.81	9.35	9.93	11.22	10.4	10.94

(a) Develop a forecasting model using regression $Y_t = \beta_0 + \beta_1 t$, where Y_t is the TRP at time t. Is there any trend in the data? Use the regression model developed to answer.

(b) Is there an auto-correlation in the data? Conduct an appropriate hypothesis test to justify your answer.

(c) The television channel would like to replace the program with a new program, the average TRP of new program will be 8 points. Based on the model developed, comment whether they should replace the program with a new program.

(d) Calculate the probability that the TRP for episode 31 will be more than 10.

4. Auto-correlation function and partial auto-correlation functions for a data set are shown in Figures 13.20 and 13.21, respectively.

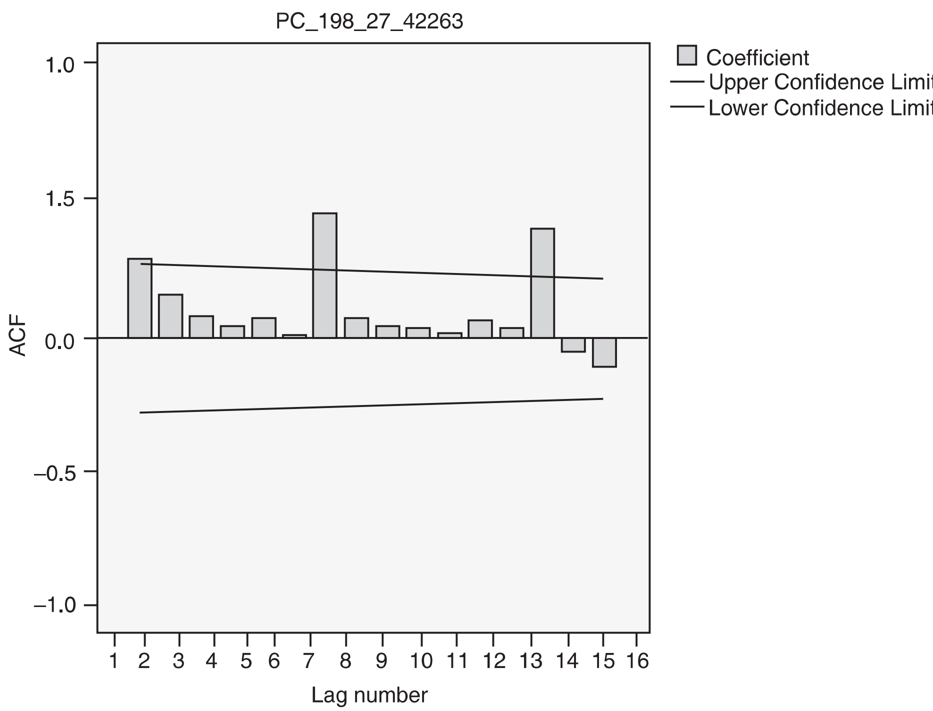


FIGURE13.20 ACF Plot.

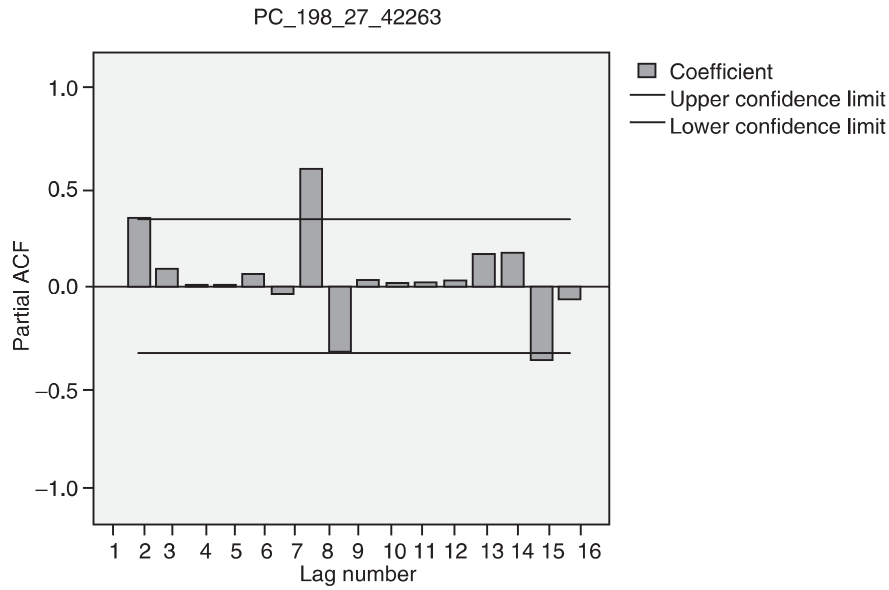


FIGURE 13.21 PACF plot.

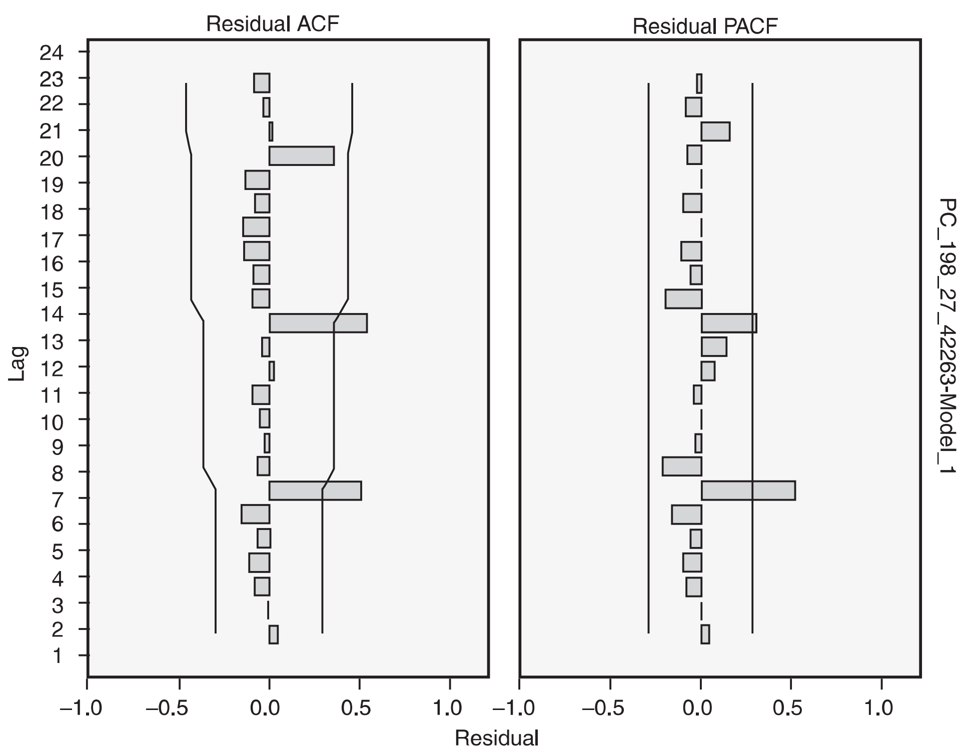


FIGURE 13.22 ACF and PACF of residuals.

(a) Based on ACF and PACF plot, what values of p and q are suitable for auto-regressive and moving average processes?
(b) The ACF and PACF of residuals are shown in Figure 13.22, what can you conclude from Figure 13.22 about the model?

5. Table 13.35 shows the volume of sales in a retail store on a day and snow fall in inches in the region.

TABLE 13.35

Day	1	2	3	4	5	6	7	8	9	10
Sales	72462	607500	816150	973300	744180	255665	1014410	464105	848775	1182225
Snow fall	16.2	20.9	24.84	18.52	5.83	26.74	10.95	22.05	30.35	15.63
Day	11	12	13	14	15	16	17	18	19	20
Sales	646825	656760	597360	559205	159470	225250	1105890	879610	718800	1000875
Snow fall	16.68	15.04	14.11	3.5	5.82	29.18	21.9	18	25.73	18.83
Day	21	22	23	24	25	26	27	28	29	30
Sales	757585	558160	899260	975355	1195165	542225	687160	220905	988275	931930
Snow fall	13.88	23.24	24.77	30.55	12.83	17.64	4.95	26.09	23.46	4.57
Day	31	32	33	34	35	36	37	38	39	40
Sales	218295	387140	289410	448985	1031180	147510	649140	316935	861060	287745
Snow fall	10.08	7.18	11.59	26.88	0.57	17.28	7.53	22.56	6.47	12.77

(a) Develop a forecasting model using moving average, single exponential smoothing, regression and AR(2). Calculate the MAPE for all 4 models. Which model gives the least MAPE?
(b) Construct ACF and PACF plot. Develop an ARIMA model [if the model is different from AR(2) model developed in (a)].
(c) Which model will you recommend to the retail store for forecasting volume of sales?

Larsen and Toubro – Spare Parts Forecasting[3]

Case Study

L&T has been in the construction and mining business for the last three decades. It also provides spare parts and after-sales service to customers to improve utilization and helps them to obtain the best possible value. The spare parts business of construction and mining equipment has a profitability of around 30%, while the equipment business has profitability of around 7%. Spare parts availability is very important to the consumer since this affects the

Case Study

Continued . . .

> availability of the high-capital cost equipment. However, with over 20,000 different types of spare parts, forecasting the demand of each part is a challenge. Technology changes frequently and superseding parts that have a different material, a different design, or improved quality are developed such that these can replace the previous parts. New machines that are developed require additional spare parts. Further, demand is quite seasonal since most of the customers plan their annual overhaul during the monsoon season, when construction is virtually at a standstill.
>
> – *Vijaya Kumar, DGM, L&T Construction and Mining Business (April 2014)*

Monday, April 21, 2014: Vijaya Kumar walked out of his cabin trying to clear his head that was filled with numbers. He was going over the demand trends and forecasts for around 20,000 spare parts of the various construction and mining equipment sold by Larsen and Toubro (L&T). Vijaya Kumar was the Deputy General Manager of the Supply Chain Department of L&T's Construction and Mining Business (CMB). L&T has been India's largest technology, engineering, construction, and manufacturing company. CMB provided heavy construction and mining equipment to its customers, along with support services and spare parts. The supply of spare parts was critical since the customer would face severe losses in the event of equipment unavailability. Forecasting was done *ad hoc*, based on the experience of the planning personnel. The value of each spare part ranged from INR 10 to INR 8 million (USD 1 ≅ INR 60 in April 2014). Maintaining the balance of the spare parts inventories was critical since unavailability would result in loss of revenues, decreased profitability, and increased customer dissatisfaction, and would also give rise to the spurious products industry. Excess inventory would lead to high inventory carrying costs, working capital lock-in, and the possibility of the spare parts becoming obsolete. Kumar had to arrive at an accurate forecasting methodology for the 20,000-odd spare parts that CMB had to supply. In theory, 20,000 spare parts called for 20,000 forecasting models; however, such a large number of models would be very time-consuming as well as expensive to develop and manage. Kumar wanted to build the forecasting model quickly so that he could roll out the forecasting strategy on a pan-India basis within a few weeks.

Construction and Mining Equipment Industry In India

Construction and mining equipment mainly included earthmoving equipment and material-handling equipment, along with a variety of other machineries (**Exhibit 1**). Prior to 1960, India imported all of its construction and mining equipment owing to the lack of indigenous manufacturing facilities. In 1964, Bharat Earth Movers Ltd. (BEML), a public sector enterprise, was established to produce dozers and dumpers under technology license from LeTorneu Westinghouse (USA) and Komatsu (Japan). Another private sector enterprise, Hindustan Motors, forayed into this sector in 1969. Hindustan Motors had technological collaboration with Terex, UK. Several foreign players such as Case, Caterpillar, Ingersoll Rand, Komatsu, JCB, Hitachi, Volvo, and Lieber entered the country after the 1991 economic reforms, either through joint ventures with Indian companies or by setting up wholly owned subsidiaries.

Case Study

Continued · · ·

The industry size was estimated to be around INR 153 billon (USD 2.55 billion); it had grown at a CAGR of around 10% during the fiscal period of 2009–2013 owing to a high growth of 33% during FY 2009–2010. The annual growth rate was around 3.4% during 2010–2013.[4] The growth of the construction and mining equipment industry depended on the construction and mining activities in the country, which had considerably slowed down owing to the economic slowdown, high interest rates, and the slowdown in policy making by the Indian government. The construction and mining sectors reported growth of around 5.9% and 0.4%, respectively, in FY 2012–2013. The construction sector comprised residential and commercial real estate along with infrastructure construction. The demand for residential and commercial estate had decreased. There was a slowdown in the infrastructure sector owing to delays in obtaining government clearances, trouble with land acquisition, and the high cost of capital. The mining sector largely included coal, iron ore, and limestone mining; this sector was affected by regulatory uncertainties. The proposed infrastructure spending of USD 1 trillion under the 12th Five Year Plan (2012–2017) and regulatory clarity over mining licenses were expected to revive the construction and mining industry; the projected growth in the infrastructure sector was estimated to be not more than 5% per annum. Similar growth was expected in the construction and mining equipment industry.

The construction and mining equipment industry mainly involved large players owing to the need for high upfront capital investments and technical expertise. BEML, L&T, Caterpillar, JCB India, Komatsu, Case, Volvo, Telcon, Escorts, Tata-Hitachi, Kobelco, Liebherr, Ingersoll Rand, and Voltas were some of the major players in the industry in India.

In the construction industry, the customers were mainly unorganized players, since construction activities were sub-contracted to smaller players by the large construction companies. However, in the mining industry, the customers were large coal mining companies (such as Coal India Ltd. and its subsidiaries), large steel manufacturers (for iron ore) (such as Steel Authority of India, Tata Steel, and Jindal Steel), and cement manufacturers (such as ACC, Ambuja Cements, etc.). Construction and mining equipment were expensive, and the life of each piece of equipment was around 20,000 hours (around 3.5 years, working in two shifts). The customers required high equipment availability to enable them to recover their capital costs. The maintenance and availability of spare parts were critical for ensuring equipment availability. Further, genuine spare parts from the original equipment manufacturer (OEM) had to be used to avoid damage to the equipment. Some customers purchased spare parts from local vendors for one of the two reasons – unavailability of spare parts with the OEM, or the lower cost of the local/spurious spare parts, which affected the performance of the equipment. Thus, service and spare parts were critical elements of the construction and mining equipment industry.

L&T: The Indian Engineering Giant

L&T was founded in Bombay (Mumbai) in 1938 by two Danish engineers, Henning Holck-Larsen and Soren Kristian Toubro. They started their operations by representing Danish dairy

[4] Source: India's Construction and Mining Equipment Industry, D&B, August 2013.

Continued · · ·

Case Study

equipment manufacturers. World War II stopped Danish supplies, which forced the founders to manufacture equipment indigenously. The war created opportunities for repairing ships, which led to the formation of repair and fabrication shops. During 1944–1946, L&T entered into several foreign collaborations. After India gained independence in 1947, there was a growing demand for equipment across industries. L&T expanded and set up offices across the eastern (Calcutta), southern (Madras), and northern (New Delhi) regions of India. In 1950, L&T became a public company, and it grew rapidly in the 1960s. Toubro and Larsen retired in 1962 and 1978, respectively. The company grew into an engineering major under the guidance of several eminent leaders.

In 2014, L&T established presence in several businesses – infrastructure, defense, aerospace, hydrocarbons, heavy engineering, construction, power, mining and metallurgy, electrical and automation products, machinery and industrial products, information technology, financial services, ship building, and railway projects. L&T had design-to-build capacities for most of its businesses. It had 137 subsidiaries and 16 associated companies; it had not only a pan-India presence but also a global presence across 30 countries. The L&T Group had gross revenues of INR 752 billion (USD 13 billion) in FY 2013.[5]

L&T's Construction and Mining Business

The Construction and Mining Business (CMB) formed a part of the Machinery and Industrial Products (MIP) business at L&T (**Exhibit 2**). The MIP segment earned revenues of INR 23 billion, which formed around 4% of L&T's total revenues. However, the MIP segment had gross profit margins of around 16.3% as against L&T's overall gross profit margin of 13.2%.[6] Thus, even though MIP was a small segment in terms of revenues, it was important from the profitability perspective. CMB was formed in 1998 as a 50:50 joint venture between L&T and Komatsu Asia & Pacific, Singapore, which was a wholly owned subsidiary of Komatsu, Japan. In April 2013, L&T bought out Komatsu's 50% stake.[7]

L&T's CMB sold equipment such as dozer shovels, dozers, dumpers, hydraulic excavators, motor graders, pipe layers, surface miners, tipper trucks, wheel dozers, and wheel loaders (**Exhibit 3**). CMB also provided equipment installation and commissioning services as well as other maintenance services. CMB did not manufacture the equipment that it sold and serviced; it had tie-ups with OEMs such as Komatsu, Scania, and so on for sourcing the equipment. Additionally, CMB sold and serviced construction and mining equipment that was manufactured in other L&T divisions.

CMB involved different verticals such as sales and marketing, services, and supply chain. The sales and marketing vertical was responsible for marketing and completing the sale orders that were

[5] Source: http://www. larsentoubro. com/.

[6] Source: L&T Annual Report, 2012–13.

[7] Source: http://articles. economictimes. indiatimes. com/2013-04-13/news/38511372_1_larsen-toubro-construction-equipment-brand-equity

Case Study

Continued . . .

placed with the OEMs. The OEM invoiced the customer and dispatched the equipment. Subsequently, CMB's service team would install and commission the equipment for the customer. The service team also took care of servicing during the warranty period and dealt with any other kind of servicing-related assistance that the customer required. The supply chain team was responsible for the availability of spare parts and the related internal logistics.

Spare Parts Supply Chain at CMB

CMB supplied parts across the country through its central warehouse at Nagpur. The CMB group had four fully equipped service stations at Chennai, Delhi, Durgapur, and Pune, which catered to the southern, northern, eastern, and western regions of the country, respectively. The facility at Nagpur catered to the central region. There were 58 sub-service centres under the fully equipped service centres. In addition to this, CMB had 26 dealers who had 84 outlets across the country. The spare parts of the construction equipment were generally supplied through the dealers, while the mining equipment spare parts were supplied via a mix of service centres and dealer outlets.

The spare parts were classified into six main categories: filters, engine maintenance parts, seals and hoses, v-belts, undercarriage, and ground engaging tools (**Exhibit 4**). Each category included different types of spare parts, and each type consisted of several varieties with different technical specifications. On average, a major piece of equipment would consist of around 1,100 unique parts. Thus, there were more than 20,000 spare parts across the different types of equipment sold by L&T.

The demand for the spare parts would vary owing to a variety of reasons. Some spare parts would be required because of operational wear and tear, while others would be needed owing to breakdown. Customers followed different maintenance strategies such as preventive or breakdown maintenance, which caused variation in demand. CMB provided warranties for the equipment; in the instance of any breakdown during the warranty period, spare parts would be required. Warranties were generally provided for 3,000 hours or 1 year of equipment operations; the typical life of each piece of equipment was around 20,000 hours. Further, several customers opted for extended warranty, which added to the variation in demand. During the extended warranty period, L&T had to support the machine similar to what was done during the warranty period. Annual Maintenance Contracts (AMC) guaranteed the availability of equipment to customers. Availability of spare parts played a critical role in fulfilling AMC services. Additionally, the sales and marketing team sometimes offered spare parts free of cost. The sales team would offer some free spare parts to the customers as per the prevailing offers made by the competition. Some of the customers would not opt for these offers and would demand equivalent value. In such instances, L&T would offer them equivalent value for the future purchase of any other spare parts from L&T (for this value). This was similar to providing coupons that were equal in value to the free spare parts offered by L&T. These coupons could be redeemed for spare parts purchased from L&T. The customer benefitted from the option of using the value gained from the 'free of cost' offer for future purchases. However, the purchases made by the customers using such coupons were generally one-time purchases, and the probability of the repetition of this demand was low. Thus, the variability in demand increased with such 'free of cost' offers.

Case Study

Continued · · ·

The after-sales business in the machine and plant construction industry generally accounted for approximately 25% of the total sales (with two-thirds being derived from the sales of spare parts and one-third from services), while it accounted for almost 50% of the total profits.[8] At CMB, spare parts contributed 25% of the total revenues; the rest was contributed by equipment sales. The profitability of the spare parts business was around 30%, while the profitability of the equipment sales business was around 7%. The unavailability of spare parts led to not only the loss of a transaction but also the potential loss of a customer, since the customer preferred immediate replacement of the damaged spare parts to avoid losses to his/her business.

Forecasting Demand for Spare Parts

Kumar was looking at the data for the period April 2009–April 2013. He had monthly details of the demand for each of the spare parts. It was difficult and very expensive to forecast the individual demand for 20,000 spare parts. Therefore, the CMB team had categorized the spare parts in several ways. The first categorization was based on the frequency of demand, according to which three categories – fast (F), medium (M), and slow (S) – were formed. This was known as the FMS categorization. The second categorization was based on the value of the product. Three categories were formed according to value: high (H), medium (M), and low (L); this was known as the HML categorization. The third categorization was based on the well-known ABC analysis according to which the spare parts belonging to category A constituted 70% of the sales, the spare parts belonging to category B constituted 25% of the sales, while category C spare parts constituted 5% of the sales. The spare parts were further classified based on combinations of these three categories and were ranked accordingly. For example, a category A spare part that was fast-moving (F) and had high value (H) was assigned the top rank; this combination was termed 'AHF' (**Exhibit 5**). According to Kumar:

> We used several analytical tools such as exponential smoothing, autoregressive integrated moving average (ARIMA) model, and Croston's method for forecasting the different categories of spare parts. We figured that one forecasting model would not be suitable for all the spare parts since the pattern of demand varied across the spare parts.

Kumar's team wanted to develop forecasting models for spare parts that would result in less than 10% error. However, Kumar wondered whether it was possible to develop such a model for all the spare parts and whether this model would work in the face of constantly changing industry trends. Was there any additional data that could be incorporated into the model to make it more robust? What strategies should be followed by the CMB team to ensure better availability and reduced inventory costs?

---

[8] Stephen M. Wagner, Ruben Jonke, and Andreas B. Eisingerich, A Strategic Framework for Spare Parts Logistics, *California Management Review*, 2012, 54(4), 69.

Case Study

Continued···

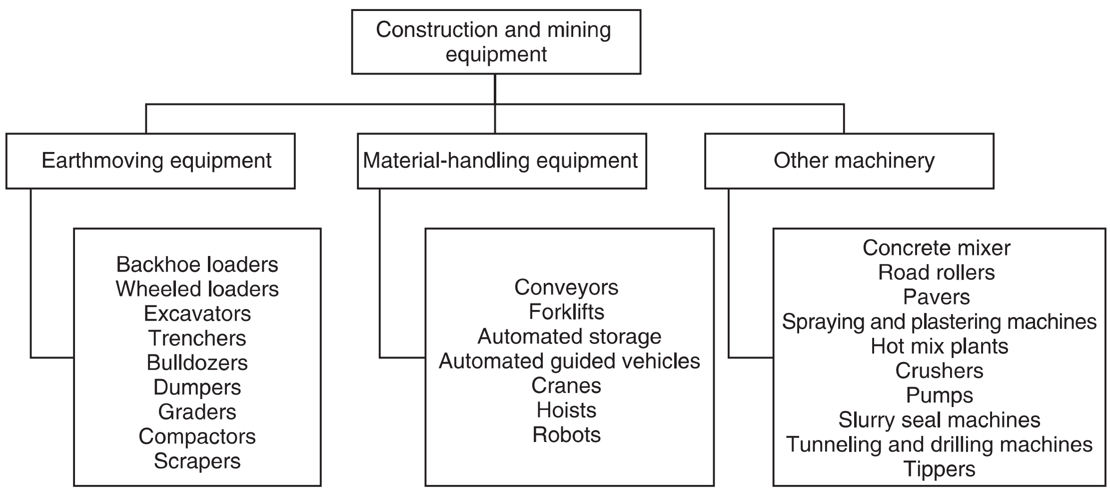


EXHIBIT 1 Classification of construction and mining equipment. Source: India's construction and mining equipment Industry, D&B, August 2013.

Case Study

Continued · · ·

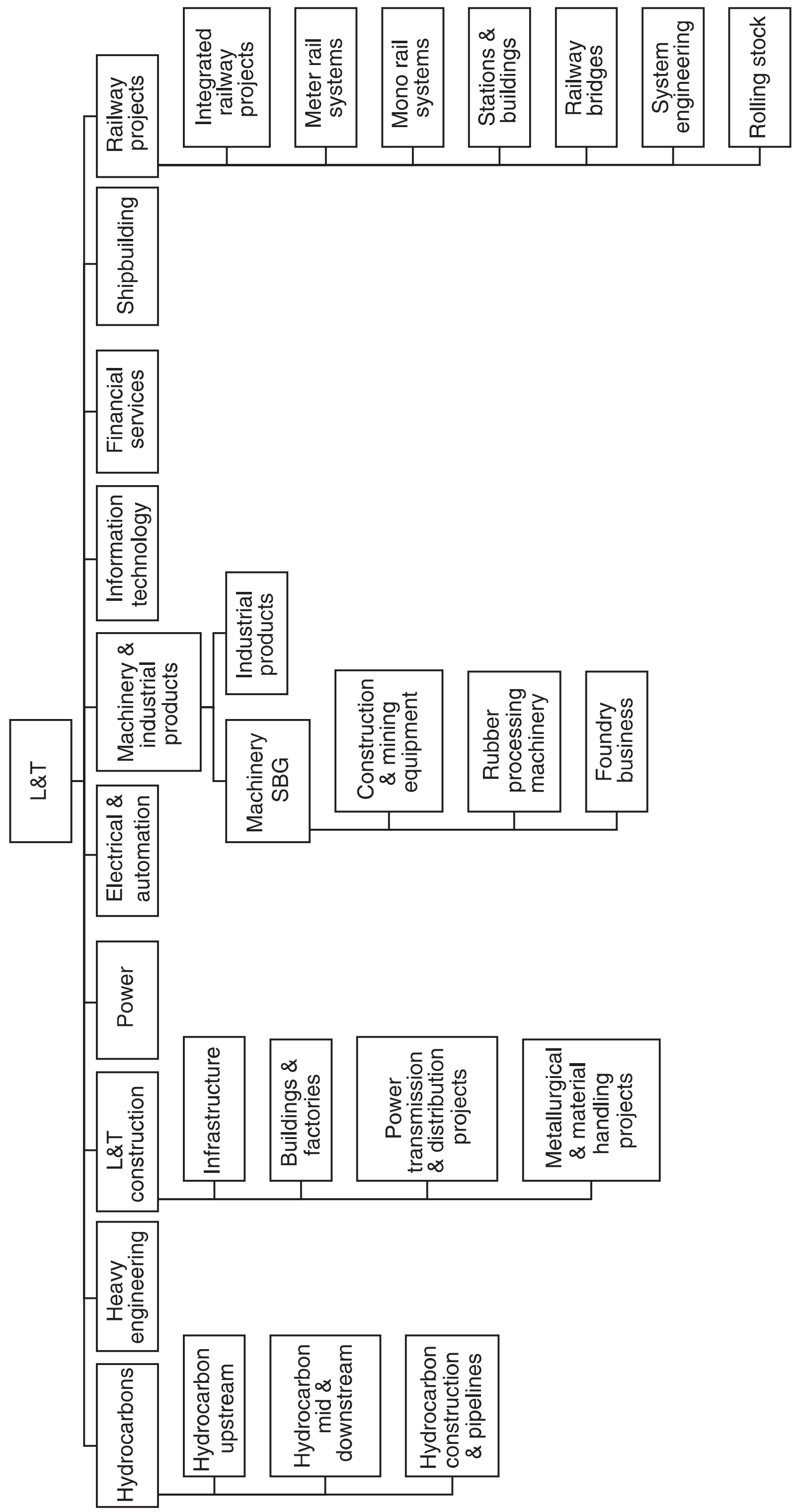


EXHIBIT 2 L&T's organizational structure. Source: L&T.

Case Study

Continued . . .

Dozers

Dumpers

Hydraulic Excavators

Motor Graders

Surface Miners

Tipper Trucker

EXHIBIT 3 L&T CMB's equipment. Source: L&T.

Continued···

EXHIBIT 4 L&T CMB's spare parts

Filters	Engine Maintenance Parts	Seals & Hoses	Undercarriage	Ground Engaging Tools	V-Belts
Engine Filters	Gasket Kits	Hydraulic Hoses	Track-Links	Bucket Teeth	V-Belts
1000-Hour Filters	Rain Caps	O-Rings	Sprocket Teeth	Cutting Edges	
Fuel Filters	Exhaust Pipe	Oil Seals	Track Shoes	Blades	
Oil Filters	Mufflers	Fuel Hoses	Rollers	Side Cutters	
Air Cleaner	Cylinder Liners	Dust Seals	Idlers		
Corrosion Resistor	Pistons	Seal Washers			
Hydraulic Filter	Piston Rings	Low Pressure Heads			
	Thermostat	Water Hoses			
	Fuel Water	Hose Clamps			
	Separators	Back-up Rings			

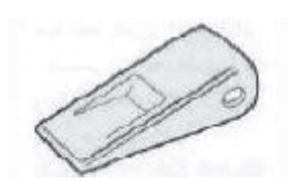
Source: L&T CMB.

Fuel Filters

Gasket Kit

O-Rings

Rollers



Bucket Tooth

Case Study

Continued...

EXHIBIT 5 Categorization of spare parts

FMS Categorization (Demand)	
Fast (F)	Demand > 8 per month
Medium (M)	8 per month ≥ Demand > 4 per month
Slow (S)	Demand ≤ 4 per month

ABC Categorization (Sales)	
A	70% of sales
B	25% of sales
C	15% of sales

HML Categorization (Value)	
High (H)	Value ≥ INR 50,000
Medium (M)	INR 50,000 > Value > INR 10,000
Low (L)	Value ≤ INR 10,000

ABC	HML	FMS	Comb	Rank
A	H	F	AHF	1
A	M	F	AMF	2
A	L	F	ALF	3
B	H	F	BHF	4
B	M	F	BMF	5
B	L	F	BLF	6
C	H	F	CHF	7
C	M	F	CMF	8
C	L	F	CLF	9
A	L	M	ALM	10
B	M	M	BMM	11
B	L	M	BLM	12
C	M	M	CMM	13
C	L	M	CLM	14
A	H	M	AHM	15
A	M	M	AMM	16
B	H	M	BHM	17
C	H	M	CHM	18
A	H	S	AHS	19
A	M	S	AMS	20

ABC	HML	FMS	Comb	Rank
A	L	S	ALS	21
B	H	S	BHS	22
B	M	S	BMS	23
B	L	S	BLS	24
C	H	S	CHS	25
C	M	S	CMS	26
C	L	S	CLS	27

Source: L&T construction and mining business

Continued · · ·

CASE QUESTIONS (USE THE DATA PROVIDED)

1. What strategy should Vijaya Kumar adopt for developing forecasting model for demand estimation of 20,000 spare parts?
2. Develop forecasting models for data provided in the Excel sheet titled "L&T Spare Parts Forecasting" and discuss the choice for using a particular forecasting model.
3. Which forecasting techniques should L&T use to forecast different spare items?

REFERENCES

1. Ali F (2017), "Amazon could Drive 80% of U.S. e-Commerce Growth", *Digital 360 Commerce,* March 8 2017. Available at https://www. digitalcommerce360.com/2017/03/08/amazon-drive-80-u-s-e-commerce-growth-next-year/, accessed on 10 May 2017.
2. Box G E P and Jenkins G M (1970), "*Time Series Analysis, Forecasting and Control*", Holden Day, San Francisco.
3. Chatfield C (1986), "Simple is Best?", Editorial in the *International Journal of Forecasting*, **2**, 401–402.
4. Croston J D (1972), "Forecasting and Stock Control for Intermittent Demands", **23**(3), 289–303.
5. Dickey D A and Fuller W A (1979), "Distribution of Estimation of Auto-Regressive Time Series with a Unit Root", *Journal of the American Statistical Association,* **74**, 427–431.
6. Fuller W. (1979), "Introduction to Statistical Time Series", John Wiley and Sons, New York.
7. Hill K (2011), "Extreme Engineering: The Boeing 747", *Science Based Life – Add Little Reason to Your Day,* 25 July 2011, available at https://sciencebasedlife. wordpress.com/2011/07/25/extreme-engineering-the-boeing-747/ accessed on 10 May 2017.
8. Ljung G M and Box G E P (1978), "On a measure of Lack of fit in Time Series Models", Biometrika, **65**(2), 297–303.
9. Makridakis S, Wheelwright S C, and Hyndman R J (1998), "*Forecasting – Methods and Applications, Third Edition*", John Wiley & Sons, USA.
10. Parker G C and Segura E L (1971), "How to get a Better Forecast", *Harvard Business Review,* March-April 1971, 99–109.
11. Taylor, J W (2011), "Multi-Item Sales Forecasting with Total and Split Exponential Smoothing", *The Journal of the Operational Research Society,* **62**(3), 555–563.
12. Theil H (1965), "*Economic Forecasts and Policy*", North Holland, Amsterdam.
13. Yaffee R A and McGee M (2000), "*An Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*", Academic Press, New York.
14. Winters P R (1960), "Forecasting Sales by Exponentially Weighted Moving Averages", *Management Science,* **6**(3), 324–342.

CLUSTERING 14

"Barn's burnt down now I can see the moon"
— Mizuta Masahide

LEARNING OBJECTIVES

- **LO 14-1** Understand the role of clustering and its importance in analytics.
- **LO 14-2** Learn different types of clustering techniques.
- **LO 14-3** Understand various distance measures such as Euclidean distance, Minkowski distance, Cosine similarity, Jaccard distance, Gower's similarity and its applications in clustering.
- **LO 14-4** Identify cluster characteristics and its importance in designing strategies.
- **LO 14-5** Understand how clustering helps data scientists with customer segmentation and personalized actions.

ESSENCE OF CLUSTERING

Clustering is one of the most frequently used analytics applications. Clustering helps data scientists to create homogeneous group of customers/entities for better management of customers. In many analytics projects, once the data preparation is complete, clustering is usually carried out before applying other analytical models. Clustering is a divide-and-conquer strategy which divides the data set into homogenous groups which can be further used to prescribe right strategy for different groups. In clustering, the objective is to ensure that the variation within a cluster is minimized whereas the variation between clusters is maximized.

IMPORTANT

Clustering is usually one of the first tasks performed in most analytics projects. It helps data scientists to analyze individual clusters further.

14.1 | INTRODUCTION TO CLUSTERING

Clustering in an important task in analytics in which the data (customers or entities) is grouped into finite subsets such that each subset is a homogeneous group of entities. Many analytics projects may start first with clustering after performing descriptive statistics and visualization on the data, since it assists data scientists to apply appropriate strategies for different clusters identified through cluster characteristics.