Chapter 12
Forecasting Analytics

Konstantinos I. Nikolopoulos and Dimitrios D. Thomakos

*Of course, **there is no accurate forecast**, but at times this shifts the focus for ... If **there is no** perfect plan, is **there such thing** as a good enough plan? ...*[1]

1 Introduction

Forecasting analytics (FA) is a subset of predictive analytics focusing only on predictions about the future. This does not necessarily include predicting exercises typically in the likes of regression analysis aiming at the "holy grail" of causality! In forecasting analytics, we do not underestimate the importance of causality, but we can live without it, and as long as we can accurately predict elements of the future, we are good to go.

In a nutshell, *forecasting analytics* is the extensive use of data and quantitative models as well as evidence-based management and judgment so as to produce alternative point and density estimates, paths, predictions, and scenarios for the future.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978- 3-319-68837-4_12) contains supplementary material, which is available to authorized users.

[1]By Kirk D. Zylstra, 2005, Business & Economics, John Wiley & Sons.

K. I. Nikolopoulos (✉)
Bangor Business School, Bangor, Gwynedd, UK
e-mail: k.nikolopoulos@bangor.ac.uk

D. D. Thomakos
University of Peloponnese, Tripoli, Greece

B. Pochiraju, S. Seshadri (eds.), *Essentials of Business Analytics*, International Series in Operations Research & Management Science 264,
https://doi.org/10.1007/978-3-319-68837-4_12

Forecasting analytics is probably the most difficult part of the analytics trio: descriptive, prescriptive, and predictive analytics. More challenging as it is about the future, and although everybody is right once the forecasts are set, only the very few (and brave!) will be right when the future is realized and the forecast accuracy is evaluated.... That is the judgment day for any predictive analytics professional (PAP).

Forecasting analytics is the key for an effective and efficient applied business and industrial forecasting process. *Applied* ... as the focus is primarily on evidence-based practical tools and algorithms for business, industrial and operational forecasting methods and applications, rather than upon problems from economics and finance. The rather more advanced techniques required for the latter are more of the core of a more focused chapter on financial predictive analytics (FPA). Similar is the case and the narrower focus of marketing analytics (MA).

Forecasting analytics is also the next big thing in the employment front, with millions of jobs on demand expected in the next few years.[2]

Forecasting analytics is a crucial function in any twenty-first century company and is the one that can truly give a competitive advantage to nowadays managers and entrepreneurs as a bad forecast can be translated into:

Either

> ... *lost sales,* thus poor service and unsatisfied customers!

Or

> ... *products left in shelves*, thus high inventory and logistics costs!

Wait a minute ... this sounds like a *lose-lose* situation! If you do not get it exactly right, you will lose money—one way or another. What's more, as you might have guessed, *you will not ever get it exactly right!* Even the most advanced forecasting system, only by pure chance, will give you a perfect forecast...

Thus, the angle of this chapter, and our sincere advice to the reader would be to:

> "...make sure you do your best to get an as-accurate-*forecast*-as-possible,[3] and learn to live with the *uncertainty* that will inevitably come with this forecast ...[4]" (exactly as the introductory quote wisely suggests).

In practice,[5] although forecasting is a key function in operations, it is usually very poorly performed. This chapter aims to shed light on the practical aspects of everyday business forecasting analytics, by adopting some well-informed, academically proven, and easily implemented processes, which in most cases are just simple heuristics. Given that this is just a chapter and not an entire book—as it may

[2] Fisher, Anne (May 21, 2013), Big Data could generate millions of new jobs, http://fortune.com/2013/05/21/big-data-could-generate-millions-of-new-jobs [Accessed on Oct 1, 2017].

[3] To take into account all available *Information* that is relevant to the specific forecasting task—usually referred to as *Marketing Intelligence*.

[4] This line is taken from Makridakis et al. (2009).

[5] Armstrong, S. (2001), *Principles of Forecasting*, Kluwer Academic Publishers.

well could be—focus is given on the process (that we abbreviate as AFA—applied forecasting analytics process) and some basic techniques but not all the specific techniques and algorithms used at each stage.

If you had googled ... "forecasting" (back in 2005) here is what you would have got:

> Forecasting is the process of estimation in unknown situations. Prediction is a similar, but more general term, and usually refers to estimation of time series, cross-sectional or longitudinal data. In more recent years, Forecasting has evolved into the practice of Demand Planning in everyday business forecasting for manufacturing companies. The discipline of demand planning, also sometimes referred to as supply chain forecasting, embraces both statistical forecasting and consensus process.[6]

The first sentence of the quote is the most important one: forecasting is more-or-less about *estimating in unknown situations*—thus your only weapon is the past and how much the latter resembles the former. This wiki-quote continues with nicely distinguishing between *forecasting* and *prediction*. In this chapter, we will use them interchangeably.[7] The next sentence fully aligns with the beliefs of the authors of this chapter: *Forecasting has evolved into the practice of Demand Planning in everyday business forecasting for manufacturing companies* ... sometimes referred to as supply chain forecasting.

The world of everyday business forecasting, comes with the assumption that some kind of regularly observed quantitative information will be available for the products under consideration. In other words, *time-series* data will be available.

A time series (Fig. 12.1) is just a series of observations over a long period of time; those observations are usually taken in equally distanced periods (months, week, quarters, etc.). That is typically how data look like in business and operational forecasting; in most cases, observations for more than 3 years per product are available, while these are recorded quite frequently (every month or less).

But this is not always the case as:

- You may have *cross-sectional* data, data referring to the same point of time but for different product/services, etc.—for example, sales for ten different car makes in a given day.
- You may have *no data* at all—so you end up using entirely your judgment as to make some forecasts.

In this chapter, focus is basically given to time-series forecasting[8] and how this integrates efficiently with judgmental adjustments. These adjustments are driven

[6]From Wikipedia, https://en.wikipedia.org/wiki/Forecasting (accessed on Feb 22, 2018).

[7]Among other similar terms like: projecting, extrapolating, foreseeing, etc. In a business context all these terms could be used.

[8]Due to limited space, other forecasting methods such as Bayesian forecasting technique, Artificial Neural Network, and special event forecasting are not included.

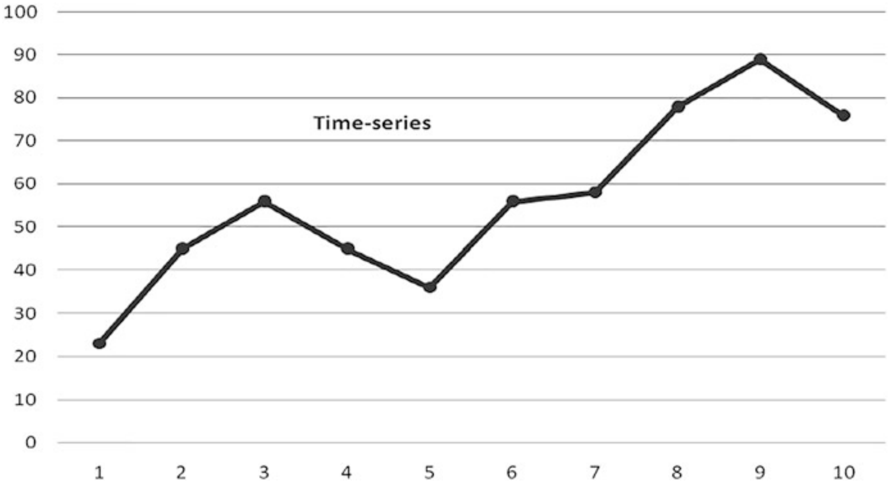

Fig. 12.1 A time series

from all sources of marketing intelligence.[9] Analysis and forecasting of cross-sectional data that is mainly the focus of regression analysis, but still remain out of the core scope of this chapter (please see Chap. 7 of this volume).

In Fig. 12.2, the aforementioned process would be to ...:

... get the *thick-line* right in the first place ...!

This thick-line stands for the history of the specific product you are interested to forecast. As explained in the data collection chapter (Chap. 2), data-related problems such as outliers, missing data, and sudden shifts need to be treated before forecasting. Also, data transformations, such as taking roots, logarithms, and differencing, might be required to conform to the requirements of the model.

The next logical step would be to project this thick-line into the future: forecasting the available time series. Time-series forecasting is based on the assumption that a particular variable will behave in the future in much the same way as it behaves in the past.[10] Thus, the dotted-line should be the "natural" extension of the thick-black-line. Natural ... in the sense that history repeats itself. This is the basic assumption of statistical forecasting; thus *Statistics*—abbreviated Stats—is the second fundamental part of the forecasting process.

We will call this a *point-forecast*. In most of the cases you will usually be interested in many points of time in the future, so forecasts for the full forecasting *horizon* as it is usually termed, and not just a single-point forecast, are of interest.

[9]Marketing intelligence or market intelligence—http://en.wikipedia.org/wiki/Market_Intelligence (accessed on Feb 22, 2018).

[10]Keast, S., & Towler, M. (2009). Rational Decision-making for Managers: An Introduction (Chapter 2). John Wiley & Sons.

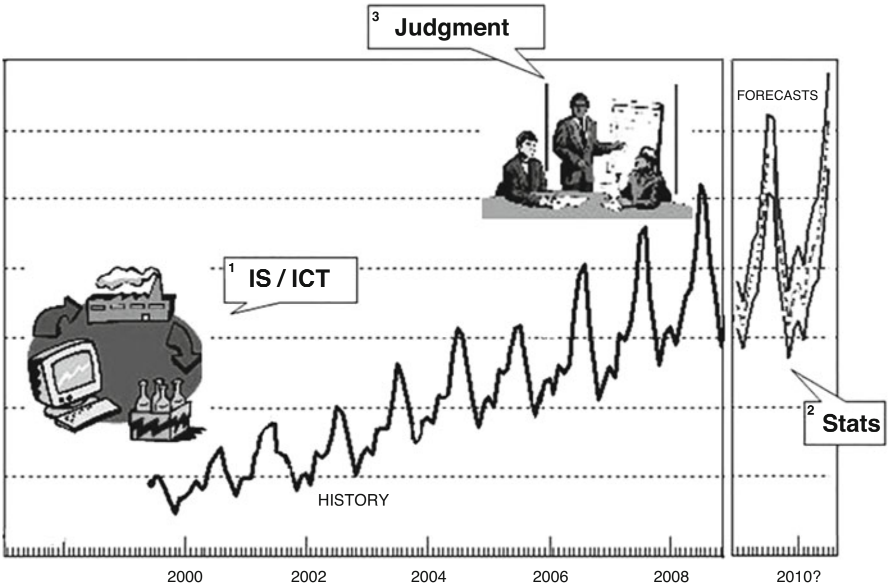

Fig. 12.2 What is needed in order to prepare a good set of forecasts?

In order to live with the aforementioned risk, we would like to have the *black-lines* as well, as shown in Fig. 12.2. Those lines are the forecast/*prediction intervals*—in this case symmetric over and under the point forecasts—and their very reason for existence is to give a sense of the *uncertainty* around point forecasts. In essence they tell you:

> ... if it's not going to be the *dotted-line*, then with great confidence it would be something from the *lower black-line* up to the *upper black-line*!

We would like to set these confidence levels around 95%, thus being 95% certain that the future unknown demand will appear somewhere between those solid-lines. But in real life this results in something that managers totally dislike: solid-lines being far out from the dotted line... And as a result, managers go one step back and require only the point forecasts to be reported to them. This is the reason that most advanced forecasting software—FSS (stands for forecasting support systems) usually do not report the prediction intervals at all.

Another critical part of the forecasting process, as presented in Fig. 12.2, is *Human Judgment*! Humans don't really like machines... They're afraid of them! They think they will get their jobs and eventually they will get fired! As a result, they dislike ready-made solutions that do not require their intervention. They would like to have some ownership of the produced forecast. So ... they *Adjust*!

They basically adjust for two reasons:

(a) Because they think that they are *better forecasters* than the FSS system in front of them! They believe they're better at selecting and/or optimizing, as well as calibrating the available forecasting models provided to them by the FSS … *that is obviously wrong*! FSS use advanced optimizers to select among thousands of values as to initialize, optimize, select, tune, and fine-tune the 100s of models available to their forecasting engine. So when an FSS suggests a model, usually termed as the *Expert* or *Auto* forecast, a very serious optimization procedure has taken place, and a challenge is more often than not futile.
(b) Because they think that they KNOW something that the FSS system in front of them does not! Now if they really know something, they would be correct to act. If the information is reliably sourced and they are confident they are doing the right thing, they should go for it. But, be aware, there are certain rules to how these adjustments should be made.
(c) What are the conceptual differences between forecasting and Statistics? Time-series Forecasting used to be a part of Statistics—nowadays (thanks to Makridakis[11] et al.) it is a far more generic and multidisciplinary scientific field. Multidisciplinary should be already obvious as we have clearly identified (a) Stats (so Math, Statistics), (b) IS/ICT (so Information and Computer Sciences), and (c) Human Judgment (so Psychology), being all essential parts of the discipline.

From a methodological point of view, *forecasters*—in contrast to *Statisticians*, quest for an optimal model in a different way, so:

How do you define the "best" forecasting method/model?

Statistics fundamentally makes the assumption that there is a *true underlying model* under the observed data series, that is, the black-dotted-line under the noisy time series in Fig. 12.3. If we identify that true underlying model, then all we have to do is project it into the future, in our case the grey-dotted line. However, is this the best possible forecast?

[11]The true birth of the Forecasting discipline dates back to late 1970s, early 1980s at the hands of Spyros Makridakis (at INSEAD), Robert Fildes (then at Manchester Business School, now in Lancaster University), and Scott Armstrong (Wharton). Benito Carbone also played a key role in the early stages. The result was to create two journals International Journal of Forecasting—IJF (Elsevier) and Journal of Forecasting—JoF (Wiley), a conference ISF (https://isf.forecasters.org/, accessed on Feb 22, 2018), an Institute IIF (www.forecasters.org, accessed on Feb 22, 2018), in a word ... a DISCIPLINE! Many have followed since then and are part of the forecasting community now, including the authors of these texts, but history was written by those 3–4 men and their close associates. More details can be found in the interview of Spyros for IJF: Fildes, R. and Nikolopoulos, K. (2006) "Spyros Makridakis: An Interview with the International Journal of Forecasting". International Journal of Forecasting, 22(3): 625–636.

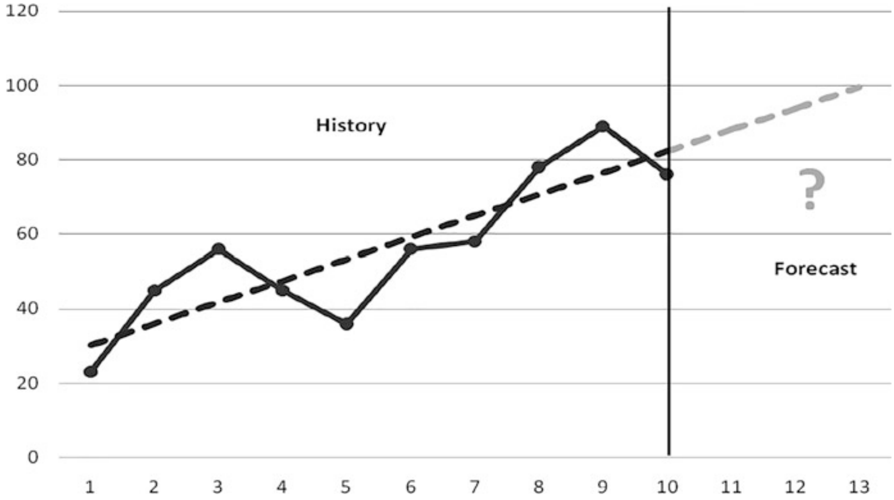

Fig. 12.3 Forecasting, extrapolation and … statistics

In essence, in statistics we try to find the model that *best fits* the data. And since we expect history to repeat itself, we project it and we are happy. Thus, the statistical forecasting recipe is: Find the best fit → Get the job done → Sleep tight!

However, in time-series forecasting, history very rarely repeats itself!

Forecasters instead focus on which model *forecasts best* rather than which model fits best! Hold on, we have an *oxymoron* here? How can we know which model forecasts best since we do not know the future?

To resolve this, we do our first forecasting *trick*: we hide a part of the series, usually the very recent one: A 20% of the most recent part of the series is usually enough. Others suggest we have to hide as much as the forecasting horizon we are interested in—thus if we have to forecast 3 months ahead we should hide the last 3 months of the available data. We call this the *holdout* data (or sample) and we will use it to evaluate which model forecasts best. For example, we hide the last year of our time series, and we use the previous years to forecast this last hidden one, with a variety models, and the one model that goes "closer" to the hidden values is the model that … forecasts "best." And this of course is not necessarily the one that fits the whole available dataset the best (the standard technique used in statistics).

Unfortunately, our approach is not bullet-proof either … as:

There is no guarantee that the model that forecasts best, will keep on forecasting best …

However, it still produces on average better forecasts than the model that fits best! At least, that is what most empirical investigations suggest. The next section discusses time-series techniques, process, and other applications.

2 Methods and Quantitative Approaches of Forecasting

2.1 Data and Statistics

Quantitative techniques mainly rely on data, statistical models, and estimation techniques for forecasting and are the basic elements of forecasting process as an aid to business decision-making, corporate planning, and management. Data can be in several forms: spoken and unspoken expression, alphanumeric written language, and other forms of communications and may consists of numbers, texts, signs, and images. These data can be time series, cross section, or combination of the two, that is, panel data. More often, in business and industry, real-time geographically distributed data on sales, orders, stocks, returns, failures, scheduling, logistics, budget, and information on competitors are very important and used in forecasting. Detecting abnormalities and irregularities, identifying outliers and missing observations, and cleaning and editing datasets for internal consistency are important aspects and one of the basic steps of setting up a quantitative forecasting model. One may in addition need to take into account known variations, such as holidays, calendar days correction (e.g., for leap year), special events, and changes in inflation rate.

Outliers are observations whose values are influenced by external factors and deviate markedly from other observations in the sample. They fall outside the 95% confidence interval around the mean of the dataset and affect forecast accuracy of the quantitative models (Hanke and Wichern 2005). There are numerous ways of identifying outliers and dealing with them including visualization and graphical presentation and newly developed methods like trimming and winsorizing (Jose and Winkler 2008). Trimmed means deleting the k smallest and the k largest observations from the sample, that is, observations are trimmed at each end. In case of winsorized means of N data points, the outer most k-values on either end are replaced with the $(k + 1)$st and $(N - k - 1)$th value at either end. Alternatively, the top and bottom values for trimming and winsorizing are determined by using some fixed percent criteria. For example, using 95% confidence interval criteria, top 2.5% and bottom 2.5% values are trimmed or winsorized.

Quantitative forecasting methods use historical patterns from time series in their prediction of future values (Makridakis et al. 1998). These historical patterns present in the data are broken down into various components using the methods of moving averages and autocorrelation analysis. The process is called decomposition of time series and pattern. The data is usually divided into seasonal, trend, cyclical, and irregular or error components. Each component is analyzed separately and the trend-cycle components are used for forecasting with the application of various statistical techniques and models. The method of moving average can use an additive or a multiplicative approach. The additive approach is used when the seasonal fluctuations do not change with the level of the series while the multiplicative approach is used when fluctuations change with the level.

Seasonal and Cyclical Adjustment: Seasonal fluctuations and changes can occur and repeat more or less regularly and periodically within a year and behavior of the data show predictability. The drivers of seasonal demands and supply are climate or festivals which repeat every year during a particular month. The most widely used tool to test and determine seasonality in time series is plotting the autocorrelation function (ACF).

Analysis of the autocorrelation coefficient or autocovariance function (ACF) which shows the relationship between current and lagged values of a time series is a way to decompose the data and investigate repeating patterns and presence of a periodic signal obscured by noise. The autocorrelation coefficient can be used to detect the presence of stationary, seasonality, trend, and random variability in the data. Specific aspects of autocorrelation processes such as unit root, trend stationary, autoregressive, and moving averages can be computed. The autocorrelation coefficient (r_k) is computed as:

$$r_k = \frac{\sum_{i=k+1}^{n} \left(Y_i - \overline{Y}\right)\left(Y_{i-k} - \overline{Y}\right)}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}.$$

where k = time lag, n is the number of observations, and Y = observed value. Close to zero values of r_k indicate no autocorrelation—the series is not related to each other at any lag k and the variability in the values is random with zero mean and constant variance (Fig. 12.4a). If there is a trend the r_k value is high initially then drops off to zero (Fig. 12.4b). In the case there is a seasonal pattern in the data series, r_k reappears in cycles, for example, of 4 or 12 lags depending on quarterly or yearly series (Fig. 12.4c).

The autocorrelation between two observations at prior time steps, that is, correlations between observations at predetermined or specified time lags, in a data series consists of direct correlations among themselves, as well as indirect correlations with observation at intervening time steps, that is, correlations with observations in between specified time lag. The ACF comprises both direct and indirect correlations among observations and does not control for correlations of a particular observation with observations at other than the specified lag. An alternative to ACF is the partial autocorrelation function (PACF) in which indirect correlations with values at shorter lags are excluded and only direct correlations with its own lagged values are taken. (Under the assumption of stationarity, the jth PACF value is obtained by regressing the present values against the past j values and taking the coefficient of the jth value as the estimate of the PACF coefficient.) The ACF and the PACF both play important roles in identifying the extent of lags in autoregressive model such as the ARIMA model discussed in a subsequent section.

Cyclical fluctuations and data behavior indicate regular changes over a period of more than 1 year, at least 2 years, and can be analyzed and forecasted. The classical decomposition method provides ways to separate cyclical movements in the data.

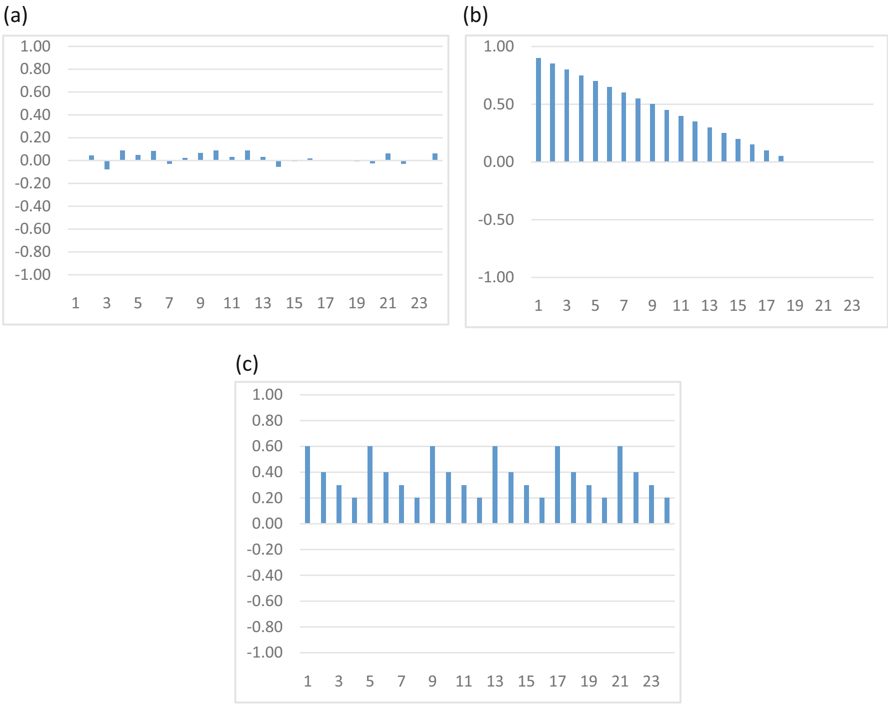

Fig. 12.4 Autocorrelation function

2.2 Time-Series Methods

A time-series data may consist of seasonal fluctuations, a trend, cyclical change, and irregular components. A simple divide and conquer approach could be to remove the seasonal and trend components by directly providing estimates of those using any number of simple techniques and using the smoothing techniques to forecast. After obtaining such a forecast, the trend and seasonal components can be added back. Below, we describe smoothing methods (different from the divide and conquer approach) for handling all three types of series, those without trend and seasonality, those without seasonality, and those with all the three.

2.2.1 NAÏVE Method (NF)

A naïve method assumes no seasonality and no trend-cycle in the data and simply sets the latest available actual observation to be the point forecast for periods in the future. Sometimes seasonally adjusted data is used and the forecasts are re-seasonalized. The naïve model is a kind of a random walk model. The NF is considered a simple benchmark against which the more advanced results may be compared. In some ways this is reasonable: the naïve method measures the volatility inherent in the data and in many systems nothing works better than what happened yesterday.

2.2.2 Average and Moving Average (MA) Methods

A simple average is obviously easy to compute but misses trends and recent changes in the series. Moving averages are a simple way of smoothing out the seasonality and noise in a series to reveal the underlying signal of trend used for forecasting. In the simplest version, the forecast for the future periods is set equal to the average of the observations of the past k periods. One may wish to optimize on the value of k. Variants of the simple moving average are weighted and exponentially weighted moving averages. In weighted moving average, different weights are assigned to various point observations within a seasonal period to be used for averaging, while in exponentially weighting, higher weight is assigned for the latest point observation of a season and lesser weights are assigned in a continuously decreasing manner to the earlier point observations.

One will notice that there will be no forecast for the first k periods unless fewer periods are used to produce the initial forecast. Also, the prediction after the last period of data will be same for every period thereafter.

2.2.3 Simple Exponential Smoothing (SES)

In this method one assumes the absence of trend and seasonality in the data. Brown (1956) is credited with the development of the single exponential smoothing methodology.

The following formula is used to forecast using the SES method:

$$F_{t+1} = \alpha * Y_t + (1 - \alpha) * F_t;$$

where Y_t is the actual observation in the period t and F_t is the forecast value from $(t - 1)$ period.

Also, $e_t = Y_t - F_t$, is the error between the observation and forecast value.

By substitution, one may also write:

$$F_t = \alpha * Y_{t-1} + (1 - \alpha) * F_{t-1}$$

$$= \alpha * Y_{t-1} + \alpha * (1 - \alpha) * Y_{t-2} + \alpha * (1 - \alpha)^2 * Y_{t-3} + \cdots + \alpha * (1 - \alpha)^{n-1} * Y_{t-n} + (1 - \alpha)^n * F_{t-n}.$$

The best α can be found using an optimization approach or simply by trial and error. The forecast can be started in many ways. A popular method is to use the first actual value as the first forecast ($Y_1 = F_1$) or set the average of the first few values as the value of the first forecast.

In sheet "SES" of spreadsheet "FA-Excel Template.xlsx" we have provided sample data (which can be changed) and the value of the smoothing constant, α, that can be changed. As α changes from zero to 1, the forecast will be seen to follow the most recent value more closely. In the Appendix in the section "SES Method" we provide the R command for SES. The data is shown in Table 12.6 and the output in Table 12.10.

In the adaptive-response-rate single exponential smoothing (ARRSES) α can be modified as changes occur:

$$F_{t+1} = \alpha_t Y_t + (1 - \alpha_t)\, F_t$$

$$\alpha_{t+1} = ABS\left(\frac{A_t}{M_t}\right)$$

$$A_t = \beta e_t + (1 - \beta)\, A_{t-1}$$

$$M_t = \beta * ABS\,(e_t) + (1 - \beta)\, M_{t-1}$$

$$e_t = Y_t - F_t.$$

Here, β is a smoothing constant to change α. In this case the smoothing constant α changes over time. The idea behind the approach is that when A and M are close to one another, then the errors have the same sign and this might indicate bias. In that case adjusting the value of α closer to one might restart the forecasting process with the most recent observation.

Starting this method is somewhat more complicated. One may set $Y_1 = F_1$, $A_1 = M_1 = 0, \alpha_2 = \alpha_3 = \alpha_4$ equal to preset value, say 0.3. The last is done so that we have a few values to warm-up before changing the value of α. In sheet "ARRSES" of spreadsheet "FA-Excel Template.xlsx" we have provided sample data (which can be changed) and the value of the smoothing constant, α, set equal to 0.3 for the first three values. This value changes as time progresses. The value of β is set to 0.5. The data is shown in Table 12.6 and the output in Table 12.10.

2.2.4 Holt Exponential Smoothing (HES)

Brown's SES methodology (1956) was extended by Holt (1957) who added a parameter for smoothing the short-term trend. The current value is called the level of the series or L_t. The change in levels ($L_t - L_{t-1}$) is used to determine the trend during the period t. Then, the trend is smoothed using the previously forecast value, that is, $T_t = (1 - \beta)T_{t-1} + \beta{*}(L_t - L_{t-1})$. This method is also called Double Exponential Smoothing (DES). The formulae are:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$

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

$$F_{t+m} = L_t + mT_t, \qquad m = 1, 2, \ldots$$

In order to start the forecast, one may set $L_1 = Y_1$ and the slope can be obtained by regressing initial values of the series against time. Search methods can be used to select the "optimal" values of the two smoothing constants, α and β. (The word optimal is in quotes because the criterion for optimization could be minimizing different types of errors, including errors one step or two steps ahead.) An example is shown in sheet "Holt" of spreadsheet "FA-Excel Template.xlsx". Appendix 1 section "Holt method" lists the R command. The data is shown in Table 12.6 and the output in Table 12.10.

2.2.5 Holt–Winters' trend and seasonality method

Holt–Winters' method is a smoothing method that takes both trend and seasonality into account known as Error, Trend, and Seasonality (ETS) or triple exponential smoothing as three components (viz., level, trend, and seasonality) in the data are used and smoothened to arrive at forecast values. It is a variant of Holt method of exponential smoothing in which a component of seasonality index along with trend and level is also added to arrive at forecast:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1})$$

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

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

$$F_{t+m} = (L_t + T_t m)S_{t-s+m}.$$

where S_t denotes the seasonal component, s is the length of a season, and γ is the seasonal smoothing factor. Note that after each step we need to renormalize the seasonal factors to add up to k ("Periods in Season"). The initial values for L_s, b_s, and S_s can be initially calculated as:

$$L_s = \frac{1}{s}(Y_1 + Y_2 + \cdots + Y_s)$$

$$b_s = \frac{1}{s}\left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \cdots + \frac{Y_{s+s} - Y_s}{s}\right]$$

$$S_1 = \frac{Y_1}{L_s}, S_2 = \frac{Y_2}{L_s}, \ldots, S_s = \frac{Y_s}{L_s}.$$

In the additive form seasonality is added to the forecast, instead of being multiplied.

$$L_t = \alpha\,(Y_t - S_{t-s}) + (1-\alpha)\,(L_{t-1} + T_{t-1})$$

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

$$S_t = \gamma\,(Y_t - L_t) + (1-\gamma)\,S_{t-s}$$

$$F_{t+m} = L_t + T_t m + S_{t-s+m}.$$

The initial values for level and trend can be chosen like in the multiplicative method. The seasonality values can be estimated as below to start the forecast:

$$S_1 = Y_1 - L_1, S_2 = Y_2 - L_2, \ldots, \quad S_s = Y_s - L_s.$$

The data is shown in Table 12.6 and the forecast output (produced by R) is shown in Table 12.10 for both methods. The R command is listed in the Appendix in the section "Holt–Winters Method."

2.2.6 Damped Exponential Smoothing for Holt's Method

When the trend in the observation has a nonlinear pattern, the damped method of exponential smoothing can be used. It is a variant of Holt's method in which only a fraction of trend forecast values of current and earlier periods are added to L_t to arrive at F_{t+1}:

$$L_t = \alpha Y_t + (1-\alpha)\,(L_{t-1} + \phi T_{t-1})$$

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

$$F_{t+m} = L_t + \left(\phi + \phi^2 + \cdots + \phi^m\right) T_t.$$

where φ is the damped parameter for the trend coefficient T_t. The forecast can be started just as in the Holt's method for FIT. Usually, the damping parameter is set to be greater than 0.8 but less than 1.

The data is given in Table 12.7 and the output in Table 12.11. The same example is given in sheet "Damped Holt" of spreadsheet "Forecasting Analytics-Excel Template". The R command is listed in the Appendix in "Damped Holt Method" section.

2.2.7 The Theta model

This methodology provides a procedure to exploit the embedded useful data information components in the form of short-term behavior and long-term trend before applying a forecasting method. The idea is to modify the local curvature of the time series before forecasting.

In a simple version, the Theta model decomposes the seasonally adjusted series into two data series called Theta lines and the forecast is a combination of the values obtained from the two theta lines (Assimakopoulos and Nikolopoulos 2000; Thomakos and Nikolopoulos 2014). The forecast from the first Theta line provides the long-term trend of the data, and is obtained from a regression line $\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_t t$, where $\widehat{Y}_t$ is forecast at time t. The second Theta line is computed by first setting a new time series equal to $2Y_t - \widehat{Y}_t$. The forecast value for the second line is obtained using SES which is discussed earlier. The point forecasts of the two Theta lines are combined using equal weight of ½.[12]

The data is given in Table 12.8 and output in Table 12.7. The same example is given in sheet "Theta" of spreadsheet "Forecasting Analytics-Excel Template". The R command is listed in the Appendix in the "Theta method" section.

2.2.8 Advances in Time-Series Processes

The Autoregressive Integrated Moving Average (ARIMA) Framework: Extrapolation models are most frequently and widely used in forecasting with a large dataset, and among them, the exponential smoothing forecast approaches have been the most popular method (Petropoulos et al. 2014). The other advances in quantitative forecasting approaches and class of models are based on the ARMA framework (Box and Jenkins 1970), Bayesian method of forecasting (Harrison and Stevens 1976), state space models (Chatfield 2005), and application of neural networks (Andrawis et al. 2011; Tseng et al. 2002). Only the ARMA method is discussed below.

ARMA approach is warranted when there is evidence of autocorrelation. Usually, the first step in applying the ARMA framework would be to study the data for evidence of stationarity. This can be done by looking at the plots of ACF and PACF as described earlier. For example, the 95% critical values are approximately at $\pm 1.96/n^{0.5}$, where n is the number of data points (these are shown in the plots below as dotted lines). Other methods include *unit root* tests, such as the Dickey-Fuller test. One of the common techniques of removing nonstationarity is differencing and seasonal differencing. In seasonal differencing, values one season apart are differenced.

[12] https://link.springer.com/chapter/10.1007/978-3-642-25646-2_56 *accessed on Sep 11, 2017.*

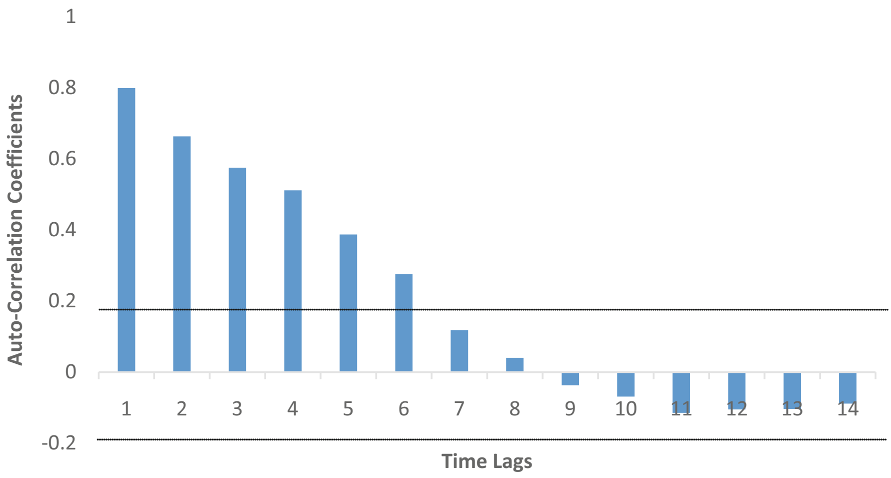

Fig. 12.5 Autocorrelation function

Box and Jenkins developed the ARMA model in 1970. The autoregressive part (or AR) of the ARMA model can be written as $y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t$, where e_t is white noise and p lagged values are used. This is a multiple regression with lagged values of y_t as predictors. The lagged explanatory variable becomes stochastic and contemporaneously correlated with the error term, making the forecast stochastic and creating bias leading to loss of confidence that comes with estimator bias and variance (please see Chap. 7 on Regression for details). The moving average (or MA) part of the model includes $y_t = c + e_t + \omega_1 e_{t-1} + \omega_2 e_{t-2} + \cdots + \omega_p e_{t-q}$, which is a multiple regression with q past errors as predictors. (A common confusion is with the MA methods discussed earlier. There the data itself was averaged. Here, the errors are averaged.)

The ACF and PACF can be used to identify the lag structure of an ARMA model. ACF is used to estimate the MA-part and PACF is used to estimate the AR-part, for example, in Fig. 12.5 we show the ACF and PACF plots of difference in data. Both ACF and PACF are decaying, there is a drop off after the time-lag 6 in ACF (Fig. 12.5), and there is spike at the time-lag 1 in PACF (Fig. 12.6). Therefore, the appropriate lag structure could be ARMA (1, 6).

The ARIMA model: ARIMA forecasting is used when the condition of no-autocorrelation and homoscedasticity are violated. Then, it requires transformation of the data series to stabilize both variance as well as mean. A data series is said to be stationary when the mean and variance are constant over time. ARIMA was developed to handle nonstationary data by differencing d-times. Later, Engle (1982) introduced autoregressive conditional heteroscedastic (ARCH) models which "describe the dynamic changes in conditional variance as a deterministic function of past values." When "additional dependencies are permitted on lags of

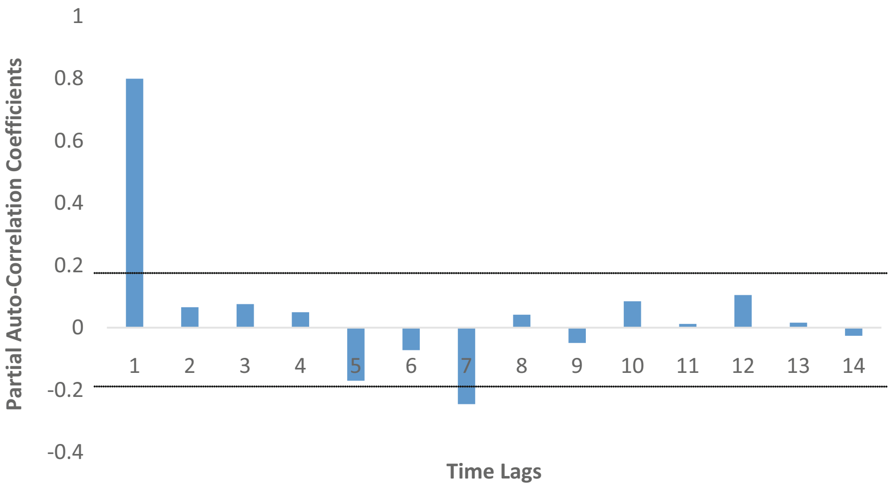

Fig. 12.6 Partial autocorrelation function

the conditional variance the model" is called generalized ARCH (GARCH) model and share many properties of ARMA (Bollerslev et al. 1994; Taylor 1997).

The data series is plotted against time to identify nonstationary, that is, changing means and variances over time. For a nonstationary series, the value of the autocorrelation coefficient, r_1, is often large and positive and the autocorrelation function (ACF) decreases slowly while it drops to zero relatively quickly for stationary data. To stabilize the varying mean due to seasonality and trend, data differencing is done, while AR and MA processes are used to incorporate autocorrelation in lagged values of the time series and the linear combination of error terms whose values change contemporaneously over time. Combining autoregressive and moving average models, the ARIMA ($p; d; q$) model can be written as: $y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t + \omega_1 e_{t-1} + \omega_2 e_{t-2} + \cdots + \omega_p e_{t-q}$, where, AR: $p =$ order of the autoregressive part, I: $d =$ degree of first differencing involved, and MA: $q =$ order of the moving average part. While it appears that one has to search for a number of values, practically just values of 0, 1, and 2 for p, d, q suffice to generate a large number of models.

Maximum Likelihood Estimation (MLE) of ARIMA Model: Having specified the model order, after checking for stationarity, the ARIMA parameters are estimated using the MLE method and use of a nonlinear numerical optimization technique. One can minimize AIC $= -2\ \log(L) + 2\ (P + q + k + 1)$ or BIC $=$ AIC $+ \log(T)(p + q + k\text{-}1)$, where L is likelihood of the data, k $=$ 1, if constant $\neq 0$ and k $=$ 0, if constant $= 0$ to get a good model. An approximate estimate of $-2\log(L)$ is given by $n(1 + \log(2\pi)) + n \log(\sigma^2)$, where n is the number of data points and σ^2 is the variance of the residuals.

The R command is given in the Appendix in "ARIMA method" section. The data and summary output of R on an example is given in Table 12.1. The complete

Table 12.1 Forecasting Sofa demand using ARIMA

Months	Demand	Months	Demand	Months	Demand	Months	Demand
1	98	14	99	27	81	40	93
2	82	15	93	28	93	41	90
3	84	16	82	29	91	42	84
4	85	17	84	30	81	43	82
5	99	18	88	31	86	44	82
6	90	19	93	32	81	45	98
7	92	20	83	33	97	46	91
8	83	21	95	34	88	47	85
9	86	22	93	35	96	48	86
10	90	23	92	36	96	49	88
11	95	24	92	37	97	50	90
12	91	25	97	38	90		
13	87	26	88	39	88		
Three months ahead forecast values (See Table 12.2 for calculations)						**51**	**92.84**
						52	**92.47**
						53	**91.39**

Table 12.2 Estimated ARIMA(2,1,2) Model

Variables	Coefficients	Standard error
AR1	0.9298	0.1429
AR2	−0.2561	0.1471
MA1	−1.9932	0.1048
MA2	0.9999	0.1048
Log likelihood = −151.47, aic = 312.94		

output is in Table 12.13. The same data can be found in sheet "Data - ARIMA" in csv format.

EXAMPLE: The monthly demand of sofa (in thousands) by a company for the last 50 months is given below in Table 12.1. The problem is to provide the forecast of sofa demand for the company for the next 3 months using the ARIMA model.

Solution:
Assume that we want to fit the ARIMA model (2, 1, 2). Assume that the data is named as ARCV. The R command is: *fitted* ← arima(ARCV, order = c(2, 1, 2)). Here, fitted is where the output will be placed.

The ARIMA parameters of order p = 2, d = 1, q = 2 are estimated using MLE (maximum likelihood estimation) methods and automated nonlinear numerical optimization techniques. The coefficients are obtained by calling *fitted*. The output is given in Table 12.2.

The forecast equation is (to be written by the user): $Y_t = 0.929{*}Y_{t\text{-}1} - 0.256{*}Y_{t\text{-}2} - 1.993{*}e_{t\text{-}1} + 0.999{*}e_{t\text{-}2}$ + Error. The example reveals that after the estimate the equation has to be written in the forecast equation form to predict values in the future. The forecast values can be obtained by using the command forecast (*fitted*, h = 3).

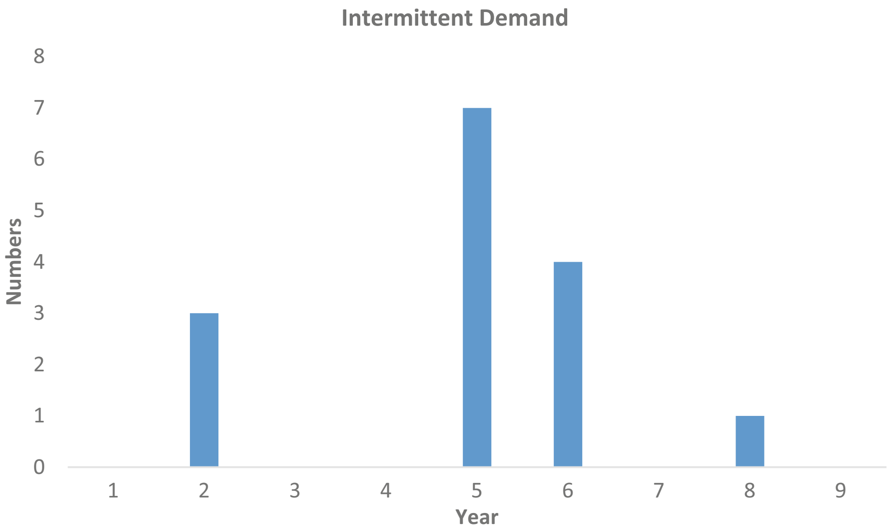

Fig. 12.7 Intermittent demand

2.3 Forecasting Intermittent Demand

The SES method assumes a constant probability for the occurrence of nonzero values which is often violated leading to count data or intermittent series (Lindsey and Pavur 2008). It is found that around 60% of the stock-keeping units in industrial settings can be characterized as intermittent (Johnston et al. 2003). Intermittent demand is characterized by infrequent demand arrivals and variable demand sizes when demand occurs. As Fig. 12.7 shows, there are "periods with demand followed by periods of no demand at all, and on top of this even the demand volume (when realized) comes with significant variation. There are two things to forecast: when the next demand period is going to be realized? And, whenever demand is realized, what will be the volume of this demand?" The basic technique is to combine different time block and different methods have been proposed to doing so. The SES method performs poorly in cases of stochastic intermittent demand.

Croston (1972) developed methodology for forecasting such cases and suggested decomposition of intermittent series into nonzero observations and the time intervals between successive nonzero values. The two series, namely, quantity and the intervals, are extrapolated separately. An updating is done for both quantity and interval series only after a nonzero value occurs in quantity series.

2.3.1 Croston's Approach (CR)

This approach applies SES independently to demand size y and inter-demand interval τ independently, where $\tau = 1$ for non-intermittent demand:

$$F_{t+1} = \frac{\widehat{y}_{t+1}}{\widehat{\tau}_{t+1}}.$$

where $\widehat{y}_{t+1}$ and $\widehat{\tau}_{t+1}$ are the forecast of the demand size and interval. Both are updated at each time t for which $y_t \neq 0$. An example is provided in sheet "Croston and SBA" in spreadsheet "Forecasting Analytics-Excel Template". The R command is given in the Appendix in the "Croston and SBA method" section. The data is in Table 12.9 and output is in Table 12.14 in the Appendix.

2.3.2 Syntetos and Boylan Approximation (SBA)

Syntetos and Boylan (2001) found that Croston's methodology provides upward biased forecast. Subsequently, they proposed an improved Croston's methodology in which the final forecasts are multiplied by a *debiasing factor* derived from the value of the smoothing parameter of intervals (Syntetos and Boylan 2005). Syntetos and Boylan (2005) found that Croston method is biased on stochastic intermittent demand and corrected the bias by modifying the forecasts to:

$$F_{t+1} = \left(1 - \frac{\beta}{2}\right) \frac{\widehat{y}_{t+1}}{\widehat{\tau}_{t+1}}.$$

SBA works well for intermittent demand but is biased for non-intermittent demand. Syntetos and Boylan (2001) avoided this problem by using a forecast:

$$F_t = \left(1 - \frac{\beta}{2}\right) \frac{\widehat{y}_{t+1}}{\widehat{\tau}_{t+1} - \frac{\beta}{2}}.$$

This removes the bias but it increases the variance of the forecast. Other variants include that of Leven and Segerstedt (2004).

None of these variants handle obsolescence well. When obsolescence occurs these methods continue to forecast a fixed nonzero demand forever. An example of SBA is provided in sheet "Croston and SBA" in spreadsheet "Forecasting Analytics – Excel Template". The R command is given in the Appendix in the "Croston and SBA Method" section. The data is shown in Table 12.9 and output is in Table 12.14 in the Appendix.

Recent development in the area of forecasting intermittent demand include the work of Babai et al. (2012), Kourentzes (2014), Kourentzes et al. (2014), Nikolopoulos et al. (2011a, b), Prestwich et al. (2014), Rostami-Tabar et al. (2013), Spithourakis et al. (2011), and Teunter et al. (2011).

2.4 Bootstrapping Method

Often, the forecasting task is to predict demand over a fixed leadtime. In this case bootstrapping might be used. Bootstrapping (Efron 1979) is a statistical method of inference that uses draws from sample to create an approximate distribution. Willemain et al. (2004) produce accurate forecasts of the demand of nine companies over a fixed lead time compared to exponential smoothing or Croston's method. We illustrate with an example.

EXAMPLE: Demand for an Automobile Part
Suppose, we would like to forecast the automobile part demand for the next 3 months. Historically, the 24 monthly demand for the part is given as follows (Table 12.3).

Solution:
Bootstrap scenarios of possible total demands for 3-month lead periods are created by taking random sample with replacement as follows:

1. Months: 3,17,21; demand: $7 + 0 + 0 = 7$.
2. Months: 1,20,8; demand: $0 + 13 + 0 = 13$.
3. Months: 6,14,19; demand: $2 + 9 + 5 = 16$.

Continuing this process, we can build the demand distribution for the given lead time.

3 AFA—Applied Forecasting Analytics Process

Forecasting in business is performed at an operational, tactical, and strategic level:

- At the *operational* level—where the focus of this chapter is, we are mostly interested in being "*roughly right within the limited available time*," given in order to prepare the forecasts. This involves short-term forecasting tasks; real-life applications are usually a few weeks/months ahead. For some category of products (e.g., dairy) we may even need more frequent forecasts (every day or every other day).
- *Tactical* forecasting involves short- to mid-term forecasting, usually 3–12 months ahead. Cumulative and individual point forecasts are needed for this period, as well as, we would incorporate the effect of forthcoming events like promotions and supply interruptions.

Table 12.3 Demand data for automobile part

Month	1	2	3	4	5	6	7	8	9	10	11	12
Demand	0	0	7	4	0	2	0	0	0	11	0	0
Month	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Demand	15	9	0	0	0	0	5	13	0	21	0	0

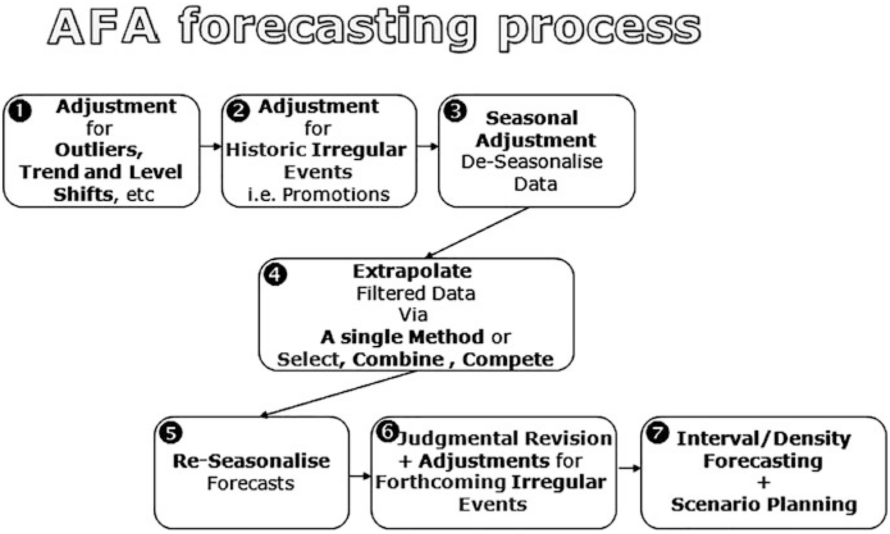

Fig. 12.8 AFA forecasting process

- At the *strategic* level we usually look into forecasting horizons that go beyond a year and involve the impact of rare events (like major international crises as the recent one regarding energy prices and the global credit system), new product development, product withdrawals, capacity amendments, and scenario planning.

The aforementioned *forecasting horizons* are only indicative, and often met in supply chain forecasting. There are many forecasting applications where a strategic forecast is just for a few months ahead! So, in order to avoid any confusion, we use the terms: "forecast for x steps ahead" or "forecast for x periods ahead", without specifying what steps/periods stand for. These steps could be anything from minutes to years depending on the application area. Typically *short-term* forecasting involves 1–3 steps ahead; *medium-term* or mid-term is for 4–12 steps ahead; and *long-term* anything over 12 steps ahead.[13]

This chapter proposes a simple *seven-step* forecasting process tailored for operational forecasting tasks. This process is illustrated in Fig. 12.8, and is abbreviated as "*AFA forecasting process*" or just *AFA* for short.

AFA provides detailed guidance on how to prepare operational forecasts for a *single* product. This process should be:

(a) *Repeated* for every product in your inventory, and
(b) *Rerun* each time new demand/sales data becomes available.

Thus, if we observe our inventory of 100 products on a monthly basis, we should run AFA every month for each of the hundred products we manage.

[13] For this and other types of forecast classifications, see Hibon and Makridakis (2000).

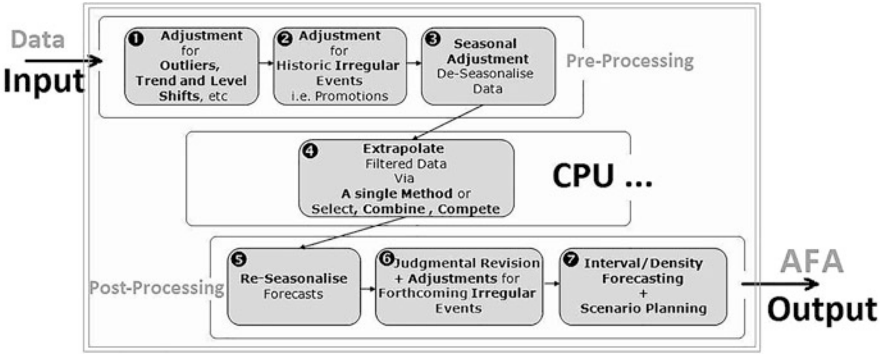

Fig. 12.9 I/O and CPU of the AFA forecasting process!

Let us start decoding what these boxes stand for; there is one box for each of the *seven steps* of the AFA process. The upper three form the *preprocessing* phase, the one in the middle the *main-forecasting-task*, and the latter three the *post-processing* phase (Fig. 12.9).

It looks like a typical Black-Box approach.[14] However, we believe it is more like a "Grey-Box" approach! A situation where you will be able to understand most of the things that are happening throughout AFA, however rely on automated tools to deliver for you!

Let us explore the AFA process as illustrated in Fig. 12.8:

- First Box: the BAD things . . .

  Each single time series comes with a number of problems. Some of these are dead-obvious but some are well hidden. To cut the long story short, we must deal with all these "bad things" and prepare a series with no missing values, no extremely low/high values (outliers), no level or trend shifts; this would involve automated detection algorithms for such problems and suggested solutions in order to adjust the original series into new series, filtered for all the aforementioned problems.
- Second Box: the GOOD things . . .

  In time-series forecasting it is often very difficult to tell good things from bad ones . . . just like in real life! A 'good thing' in a time series is a *special event* (SE), often termed in literature as *irregular*, *infrequent* or *rare* event; it could be a promotion, a production interruption, news, regulation, etc., in general anything that could make demand deviate substantially from regular levels! But why is something irregular good? Simply because it is an information-rich period, a period with special interest where an external/exogenous event has driven excess or limited demand respectively. So it would look exactly like an outlier, but we will know what exactly happened. From a mathematics perspective, the way you

[14] A standard engineering expression, for a situation or a solution where something seems to work fine, but we are not sure why and definitely do not know for how long it will keep on working!!

detect and subsequently adjust periods with special events, is identical to the one used to treat outliers.

- Third Box: the REGULAR things . . .

  In forecasting, finding regularities and patterns in a series is an essential task; usually termed *periodicity*, things that repeat themselves on a regular basis. If the regularity, the repetition happens within a year then we will call this phenomenon *seasonality*; for less than a year *mini-cycles*, while for more than a year, big *economic/financial-cycles*. In any case, removing the effect of these cycles at this stage of the AFA process (and reintroduce them later on) has been empirically proven to work very well, as argued in various empirical investigations.[15]

After successfully completing these three steps of the preprocessing phase of our time series, we should by now have a nice *filtered*[16]—*smoothed*—series, that will look almost[17] like a straight *line*, either entirely flat or with a constant-*ish* trend (upward or downward). Now, it is about time to extend this line into the future . . . Now we are ready to forecast!

- THE (fourth) BOX: FORECASTING . . .

  This is where all the fun is . . . . Let us try forecasting: extrapolate the series in the future. We will not just choose a method—and that is it! (where would the fun be after all . . . ?).

  We basically employ three fundamental strategies . . . the "three forecasting tricks" as I fancy calling them:

  – "**Select**": my mother always says that "Experience is all that matters . . . "; and she is probably right. Thus if a method worked well in the past, we should probably stick to it, and keep on selecting that same one for the task of extrapolation. Furthermore, some methods may have been proven to work better for some products while other methods better for other products; so there are "horses for courses" and once again we are better off sticking to them. In essence, we could build a nice table—a *selection protocol* (SP) as we will call it more formally, where in one column there is a list of our products, while in the other column the forecasting methods and models that have worked well in the past for the respective products. An illustrative example is shown in Table 12.4.

– "**Combine**":

  When in doubt . . . combine!

  a very good piece of advice I dare say. When a method has worked well in the past for a certain product, but the new Statistician on the block . . . insists that method *X* is the new panacea in time-series forecasting, then why not combine those two? So get a set of forecasts from your trusty chosen method, get another set of forecasts from the new highly promising method, and then average these

[15] For more information please visit http://www.forecastingprinciples.com/, or read "Principles of Forecasting: A Handbook for Researchers and Practitioners, J. Scott Armstrong (ed.): Norwell, MA: Kluwer Academic Publishers, 2001".

[16] A term often used in engineering applications.

[17] Of course there would still be some *noise* over this line.

Table 12.4 Forecasting selection protocols

Company X	
'Best' FORECASTING methods over the last 2 years	
Product 1	Method A
Product 2	Method B
Product 3	Method A
...	...

to get a final set of forecasts. If you believe more in the former (or the later) you could easily differentiate the weights respectively as to express your belief, for example, via a 30% weight to the experience-based method and 70% to the new one. Rule of thumb: "*Combining always works!*" (In other words: Combining most of the times outperforms the individual performance of the methods being combined.)

– "**Compete**," the true reason forecasters exist: (empirical) *Forecasting Competitions*! We do not trust anything, and from all the available methods and models, applied on all the available history, we will find the one that forecasts "best." These criteria typically include average or median error metrics like MAE, MdAE, MAPE, MDAPE, MASE, and MdASE.

Sometimes, we even apply these tricks simultaneously—for example, (a) we compete only among methods that have performed well in the past, or (b) we combine the winner of the competition and the top performing method in the past as described in a selection protocol, or ...

- Fifth Box: Superimpose regular patterns.

  By now, our forecasts should look like a straight line, either *flat* or with a certain *slope*. If we have identified regularities in step three, then we need to bring them back into the game, in other words we superimpose these patterns onto the extrapolation. Once this step completed, our forecasts will have *ups* and *downs,* and will look like a natural extension of the cycles and seasonality observed in the history of the series.
- Sixth Box: Human Judgment.

  This is where humans come into the game. No matter how sophisticated the process so far, people—usually referred to as *forecasters* or *experts*—want to intervene at this stage; primarily to introduce market intelligence? This is usually performed in two phases: (a) an initial phase, where experts roughly revise all provided forecasts by changing[18] them by a percentage x% (e.g., increase all monthly forecasts for the full next year by 10%), and (b) a more targeted one, where some specific forecasts in the future are adjusted for the potential impact of special events like promotions (e.g., increase by an extra 1000 units the sales forecast for next September due to an expected advertising campaign).
- Seventh (Last) Box: Density forecasting + SCENARIOS; living with Uncertainty!

[18]Usually increasing the forecasts, due to an optimism bias (more on this and other types of bias in Chap. 13).

Table 12.5 AFA output

– A list of *adjustments* made to the original data due to problems
– A list of *adjustments* made to the original data due to special events
– A set of *Seasonal Indices* if seasonality was identified
– Sets of *Cyclical indices* if min-cycles or major economic cycles were identified
– A set of statistical *point-forecasts* (each one for each respective forecasting horizon)
– A set of judgmentally *revised* forecasts plus *Notes* explaining the reasons for adjustment
– Two sets of *prediction intervals*, under and over the provided forecasts
– An estimation of the *Bias* of those forecasts: A tendency to consistently under-forecast or over-forecast
– An estimation of the *expected accuracy* of those forecasts: In the form of past errors
– An estimation of the *uncertainty* of those forecasts: In the form of the standards deviation of the forecasts
– An estimation of the *endogenous difficulty* of the forecasting task; in the form of the *noise*[19] existing in the original time series. In statistical and mathematical sciences, we believe that under any observed phenomena (time series in our case), there is an underlying signal where whatever is not explained and captured from it, is described as noise in the series
– *Statistical significance*[20] of the forecasts: In the form of comparisons with standard forecasting benchmarks
– *Economics significance*[21] of the forecasts: In the form of the financial implications of our forecasts as in stock holding costs [22], or trading financial results [23]

- In this final step, we try to cope with the *uncertainty* that comes with the produced forecasts. Firstly, we usually provide a set of *confidence* or *prediction* intervals, associated with the point forecasts for the full forecasting horizon, as shown in Fig. 12.2; this is also known as *density* forecasting. There are theoretical as well empirical ways so as to produce these intervals. The most popular way to deal with the uncertainty around the provided forecast is by building scenarios. These practically derivate from the produced forecasts, but we will treat them as an indispensable part of the AFA process.

We have seen the input; we have roughly seen the steps within the "grey-box"; let us stay a bit more on the output of AFA. When you started reading this chapter, you probably thought it would all be about a number or a few numbers—if forecasts for more periods ahead were required. By now, it should have become obvious that far more output—in numerical and narrative form—will be available. Practically every step of the AFA process is producing some output, which is consisted by-and-large of what is contained in Table 12.5. AFA Output (which is not exhaustive).

[19]Noise is a term met in many sciences. I prefer the electrical engineering definition of it where Noise can block, distort, or change/interfere with the meaning of a message in both human and electronic communication.

[20]Armstrong 2001.

[21]Timmerman and Granger 2004.

[22]Syntetos et al. 2010.

[23]Maris et al. 2007; Bozos et al. 2008.

4 A Few More Interesting FORECASTING Applications

So what are we really fascinated to forecast?

Whenever I say outside my inner academic circle that I am a forecaster or into forecasting/predictive analytics, I typically get three responses:

What's going to be the weather tomorrow?
Can you forecast the numbers for the lottery?
Can you forecast the stock market?

My answer to all these is: ... "Unfortunately NO". And this brings us back to the fundamental question: "what can realistically be forecasted, and what can not?" Maybe more interestingly the latter is what people are really interested in forecasting ...

The aforementioned questions are clearly beyond the scope of this chapter ... particularly the weather! The following list is not exhaustive, but we believe captures most of the things people are interested to forecast:

- Gambling / Individual and team performance in sports.
- Weather forecasting.
- Transportation forecasting.
- Economic forecasting/Major Economic shocks.
- Technology forecasting.
- Earthquake prediction/Major catastrophes.
- Land use/Real estate forecasting.
- Long term/Strategic forecasting/Foresight.

5 Evaluating Forecast Accuracy

The measures such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percent error (MAPE) are used to evaluate accuracy of a particular forecast (Hyndman 2014).

Let, y_t denote the tth observation and $\widehat{y}_{t|t-1}$ denotes its forecast based on all previous data, where $t = 1,2 \ldots T$. Then, the following measures, mean absolute error, means square error, root mean squared error, and mean absolute percentage error are useful.

$$MAE = T^{-1}\sum_{t=1}^{T} \left|\left(y_t - \widehat{y}_{t|t-1}\right)\right|$$

$$MSE = T^{-1}\sum_{t=1}^{T} \left(y_t - \widehat{y}_{t|t-1}\right)^2$$

$$RMSE = \sqrt{\left(T^{-1}\sum\nolimits_{t=1}^{T}\left(y_t - \widehat{y}_{t|t-1}\right)^2\right)}$$

$$MAPE = 100T^{-1}\sum\nolimits_{t=1}^{T}\frac{\left|y_t - \widehat{y}_{t|t-1}\right|}{|y_t|}$$

Some practical considerations: It is not appropriate to compare these error measures across models that do and do not incorporate trend and seasonality. The reason is that even those accommodations can be somewhat ad hoc. Moreover, the errors could be correlated, and therefore adding them across periods might create a wrong notion of overall accuracy. ACF diagnoses of the errors might reveal patterns that can help identify seasonality and trends as mentioned in several places in the chapter. In addition, out of sample testing is also recommended.

The error estimates are for the forecast for the current period. However, they can also be computed two steps or three steps ahead. This is done by comparing the actual versus the forecast obtained for that period, but two or three periods ago.

Often, prediction intervals are necessary to the user. For example, one might like to know how the uncertainty in the forecast values changes for different forecast horizons. The inventory planner would like to use the prediction interval to source sufficient number of parts to ensure s/he does not run out of stock more than 10% of the time. A farmer might like to know what extent price can deviate if the produce is harvested and sold next week instead of right now. As one may anticipate, making these types of predictions involves making assumptions about the structure of error distribution. Forecasting packages often produce 1.25 times the MAD as an estimate of standard deviation. One may also compute a heuristic interval by using past data, generate a forecast, and determine the interval by trial and error.

6 Conclusion

This chapter aims to provide managers/executives as well as managers-to-be (PG/MBA students), with the necessary background knowledge and software tools to run a successful business forecasting analytics function. Ok … but there at least 50 titles out there in business forecasting! ... Why do we need another one recapturing the analytics end of? This chapter is not about all the things you could possibly do when you are faced with a forecasting task. It is not or about guiding you through a methodology[24] tree, where all possible options are given, and it is up to you to decide where to go. If this is what you are looking for, then the best place to go is www.forecastingprinciples.com (accessed on Feb 22, 2018); led by

[24] www.forecastingprinciples.com [Accessed on Oct 1, 2017].

J. Scott Armstrong[25] and the International Institute of Forecasters,[26] where you get a gateway to the amazing world of forecasting free of cost.

Furthermore, this chapter is not about giving you all the underlying theory and mathematics of the discipline. In fact, mathematics and statistics, theorems, and axioms are kept to the absolute minimum. "Everything is kept as simple as possible, … but not simpler!"[27] So there will be a few formulae, but expressed in a way that does not require a mathematical background to follow. If you were looking for the mathematics of forecasting then the leading textbook of the field "Forecasting Methods and Applications"—by Makridakis et al. (1998)—is your reference point. For engineers like me, that prefer the "do it yourself" approach, the second edition of the latter book is particularly useful as most of the forecasting algorithms are presented in such a way that their implementation is very straightforward in a standard programming language.

Now, if you need subjective approach then and judgmental forecasting is your weapon of choice when approaching forecasting tasks, asked then Goodwin (2006) and along with Wright and Goodwin (1998), are probably the way to go.

A *forecasting process* we strongly believe will significantly enhance the forecasting performance in your company/private or public organization; and is a process consisting roughly of two basic elements:

(a) A fairly accurate set of *forecasts*.
(b) A good estimate of the *uncertainty* around them.

Of course, it would be up to you, once faced with real-life problems, how to use these forecasts, and more importantly how to take countermeasures and back-up policies as to cope with the predicted uncertainty. Living with scenarios built around this uncertainty is the key to your business success.

Electronic Supplementary Material

All the datasets, code, and other material referred in this section are available in www.allaboutanalytics.net.

- Data 12.1: Data - ARIMA.csv
- Data 12.2: Data - Croston and SBA.csv
- Data 12.3: Data - Damped Holt.csv
- Data 12.4: Data - SES, ARRSES, Holt, HoltWinter.csv
- Data 12.5: Data - Theta.csv
- Data 12.6: FA - Excel Template.xlsx
- Code 12.1: Forecasting Analytics.R
- Data 12.7: Forecasting chapter - Consolidated Output.xlsx

[25] Professor J. Scott Armstrong, http://www.jscottarmstrong.com/ [Accessed on Oct 1, 2017].
[26] International Institute of Forecasters, https://forecasters.org/ [Accessed on Oct 1, 2017].
[27] A famous quote attributed to Albert Einstein.

Exercises

Ex. 12.1 Using time-series data on annual production of tractors in India from 1991 to 2016, provide forecast of the production of tractors (in millions) for the year 2017 using Theta model (combining regression and SES methods, $\alpha = 0.4$).

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Production	14.5	14.8	15.1	15.4	15.7	16.0	16.3	16.7	17.0	17.3	17.7	18.0	18.4
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Production	18.8	19.3	19.8	20.3	20.8	21.3	21.9	22.4	23.0	23.5	24.1	24.7	25.4

Ex. 12.2 Given the monthly production (in millions) of mobiles by a company for the last 20 months, provide the forecast of mobile production for the company for the next 3 months using the ARIMA (2,1,2) model.

Months	1	2	3	4	5	6	7	8	9	10
Production	3.40	3.43	3.47	3.50	3.54	3.57	3.61	3.65	3.68	3.72
Months	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Production	3.77	3.82	3.77	3.82	3.77	3.82	3.77	3.82	3.77	3.82

Ex. 12.3 Monthly demand (in millions) for an automobile spare part is given as follows:

Months	1	2	3	4	5	6	7	8	9	10	11	12
Demand	3	0	1	0	0	8	0	0	0	2	0	5
Months	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Demand	0	0	0	1	4	0	0	0	3	?	?	?

Based on the above time-series data, provide 3 months ahead forecast for the spare part using Croston and SBA methods.

Ex. 12.4 Using the provided Excel templates, create:

(a) A version of Holt exponential smoothing where both the level smoothing parameter and the trend smoothing parameter are equal,
(b) A version of damped Holt exponential smoothing where alpha (α) = a, beta (β) = a2, and phi (φ) = a3.

Appendix 1

Example: The monthly sales (in million USD) of Vodka is given for the period 1968–1970. We want to forecast the sales for the year 2016 using various forecasting methods—SES, ARRSES, Holt, Holt–Winters (Additive/Multiplicative).

Data: The data can be downloaded from the book's website and the dataset name is "*Data - SES, ARRSES, Holt, HoltWinter.csv*". You can also refer to Table 12.6 for data.

R Code (to read data)

```
read.csv ("filename.ext", header = TRUE)
```

SES Method

Install forecast package

```
install.packages("forecast")
```

R function

```
ses (<Univariate vector of observations>, h = <number of periods
 to forecast)
```

Note: The *ses* function in R by default optimizes both the value of alpha and the initial value.

In case you prefer the output for a specified alpha value then use parameter

```
<initial = "simple">
```

and set the alpha value in the parameters.

```
ses(<univariate vector of observations>, h = <number of periods
 to forecast>, alpha = < >, initial = "simple")
```

The above code will set the first forecast value equal to first observation. If alpha is omitted it will optimize for alpha.

Table 12.6 Data for SES, Holt, ARRSES, and Holt–Winters method

Period (t)	Vodka (Y_t)	Period (t)	Vodka (Y_t)	Period (t)	Vodka (Y_t)
Jan-68	42	Jan-69	21	Jan-70	47
Feb-68	40	Feb-69	31	Feb-70	38
Mar-68	43	Mar-69	33	Mar-70	91
Apr-68	40	Apr-69	39	Apr-70	107
May-68	41	May-69	70	May-70	89
Jun-68	39	Jun-69	79	Jun-70	116
Jul-68	46	Jul-69	86	Jul-70	117
Aug-68	44	Aug-69	125	Aug-70	274
Sep-68	45	Sep-69	55	Sep-70	137
Oct-68	38	Oct-69	66	Oct-70	171
Nov-68	40	Nov-69	93	Nov-70	155
Dec-68	49	Dec-69	99	Dec-70	143

Holt Method

Install forecast package

```
install.packages("forecast")
```

R function.

```
holt (<univariate vector of observations>, h = <number of
 periods to forecast>)
```

Note: The *holt* function by default optimizes both the value of alpha and the initial value.

In case you prefer the output for a specified alpha and beta value then use

```
<initial = "simple">
```

parameter and set the alpha and beta values in the parameters.

```
holt (<univariate vector of observations>, h = <number of
 periods to forecast>, alpha = < >, beta = < >, initial =
 "simple")
```

The above code sets first level equal to first value and trend as difference of first two values. If alpha is omitted it will optimize for alpha.

Holt–Winters Method

Install stats package

```
install.packages("stats")
```

R function

```
HoltWinters (<name of dataset>, alpha = <>, beta = <>, gamma =
 <>, seasonal = c("additive", "multiplicative"), start.periods =
 2, l.start = NULL, b.start = NULL, s.start = NULL, optim.start
 = c(alpha = 0.3, beta = 0.1, gamma = 0.1),
 optim.control = list())
```

The value of *alpha*, *beta*, and *gamma* can be either initialized by specifying <alpha>, <beta>, <gamma> and if they are NULL it will optimize the values as specified in optim.start. You can also specify starting values of *alpha*, *beta*, and *gamma* to optimize using <*optim.start*> parameter. Seasonality can be considered *additive* or *multiplicative*. The <*start.periods*> is the initial data used to start the forecast (minimum 2 seasons of data). Starting values of level <*l.start*>, trend <*b.start*>, and seasonality <*s.start*> can be either be initialized or optimized by setting equal to NULL.

For the *HoltWinters* function, the dataset must be defined as a time-series (ts) type. A dataset can be converted to time-series type, using the below code:

```
ts (<name of dataset>, frequency = number of periods in a season)
```

Damped Holt Method

Data: The data can be downloaded from the book's website and the dataset name is "*Data - Damped Holt.csv*". You can also refer to Table 12.7 for data.

Install forecast package

```
install.packages("forecast")
```

Table 12.7 Data for damped exponential smoothing using Holt's method

Period (t)	Demand (Y_t)	Period (t)	Demand (Y_t)
1	818	8	805
2	833	9	808
3	817	10	817
4	818	11	836
5	805	12	855
6	801	13	853
7	803	14	851

Table 12.8 Data for Theta model

Time (t)	Cars (Y_t)	Time (t)	Cars (Y_t)
1	13.31	13	18.12
2	13.6	14	18.61
3	13.93	15	19.15
4	14.36	16	19.55
5	14.72	17	20.02
6	15.15	18	20.53
7	15.6	19	20.96
8	15.94	20	21.47
9	16.31	21	22.11
10	16.72	22	22.72
11	17.19	23	23.3
12	17.64	24	23.97

R function

```
holt (<univariate vector of observations>, h = <number of
 periods to forecast>, damped = TRUE)
```

Note: The *holt* function by default optimizes both the value of alpha and the initial value.

In case you prefer the output for a specified alpha, beta, and phi values then use <*initial* = ”*simple*”> parameter and set the alpha, beta, and phi values in the parameters.

```
holt (<univariate vector of observations>, h = <number of
 periods to forecast>, damped = TRUE, alpha = < >, beta = < >,
 phi =  < >)
```

If *alpha*, *beta*, and *phi* are omitted, it will optimize for these values.

Theta method

Data: The data can be downloaded from the book's website and the dataset name is "*Data - Theta.csv*". You can also refer to Table 12.8 for data.

Install forecTheta package

```
Install.packages("forectheta")
```

R function

```
stm (ts (<univariate vector of observations>), h = <number of
 periods to forecast>, par_ini = c (y[1]/2, 0.5,2))
```

Refer https://cran.r-project.org/web/packages/forecTheta/forecTheta.pdf for more details.

Note: You may try either “stm” or “stheta.” There is a slight difference in the implementation of the original method.

ARIMA method

Data: The data can be downloaded from the book’s website and the dataset name is “*Data - ARIMA.csv*”.

Install forecast package

```
install.packages("forecast")
```

R function

```
arima (ts (<univariate vector of observations>, freq = <period
 of data>), order = c(<p>,<d>,<q>))
```

To view the fitted coefficients, store the output and call that array.

To forecast, use the command:

```
forecast (<name of output>, h = <number of periods to forecast>)
```

Croston and SBA method

Data: The data can be downloaded from the book’s website and the dataset name is “*Data - Croston and SBA.csv*”. You can also refer to Table 12.9 for data.

Install tsintermittent package

```
install.packages("tsintermittent")
```

R function

```
crost (ts(<univariate vector of observations>), h = <number of
 periods to forecast>, w = c(<>,<>), init = c(<>,<>), type =
 "croston", init.opt = FALSE)
```

Refer https://cran.r-project.org/web/packages/tsintermittent/tsintermittent.pdf for more details.

<*crost*> function operates on the time-series vector. Initial values can be either differently chosen or provided as a vector of demand and interval value. <*type*> refers to the model used. Cost to the optimization criterion. If <*init.opt*> is *TRUE*, it will optimize the initial values. If <*w*> is *NULL*, it will optimize the smoothing parameters.

Table 12.9 Data for Croston and SBA model

Months (t)	Actual Demand Number (Y_t)
1	5
2	0
3	7
4	28
5	0
6	0
7	11
8	0
9	4
10	19
11	0

Consolidated Forecast Output for Vodka Example

See Tables 12.10, 12.11, 12.12, 12.13, and 12.14.

Table 12.10 Consolidated output of SES, ARRSES, Holt, Holt–Winters methods (*R Output, ˆExcel Output)

Period	Vodka	SESˆ	ARRSESˆ	Holtˆ	Holt–Winters Additive*	Holt–Winters Multiplicative*
t	Y_t	F_t	F_t	F_t	F_t	F_t
Jan-68	42	42.0000	42.0000	42.0000	–	–
Feb-68	40	42.0000	42.0000	42.0000	–	–
Mar-68	43	41.0000	41.4000	41.2474	–	–
Apr-68	40	42.0000	41.8800	43.7804	–	–
May-68	41	41.0000	41.3160	43.2701	–	–
Jun-68	39	41.0000	41.1590	43.5794	–	–
Jul-68	46	40.0000	39.9062	42.5508	–	–
Aug-68	44	43.0000	45.2167	46.1132	–	–
Sep-68	45	43.5000	44.4632	46.6752	–	–
Oct-68	38	44.2500	44.5858	47.4834	–	–
Nov-68	40	41.1250	42.1887	43.7545	–	–
Dec-68	49	40.5625	40.5625	43.0670	–	–
Jan-69	21	44.7813	47.5915	47.9867	13.3136	17.0694
Feb-69	31	32.8906	34.1680	34.0851	25.4230	24.3483
Mar-69	33	31.9453	31.8512	33.1256	30.4248	25.8316
Apr-69	39	32.4727	32.7419	33.9700	40.4559	31.7168
May-69	70	35.7363	36.4148	38.0008	71.4978	57.2978
Jun-69	79	52.8682	45.2818	58.0939	79.4149	64.2547
Jul-69	86	65.9341	75.3264	73.0838	73.0716	61.8609
Aug-69	125	75.9670	85.5710	84.1672	84.3076	65.0093
Sep-69	55	100.4835	123.8677	111.8418	117.0539	73.0378
Oct-69	66	77.7418	55.6361	84.0182	82.3369	65.9793
Nov-69	93	71.8709	59.8772	76.2076	75.8113	71.2471
Dec-69	99	82.4354	65.2778	87.9222	93.4888	86.3133
Jan-70	47	90.7177	80.4700	97.0133	69.5189	44.0442
Feb-70	38	90.7177	55.8669	97.0133	63.7863	63.7920
Mar-70	91	90.7177	53.2861	97.0133	46.3351	66.0826
Apr-70	107	90.7177	70.0641	97.0133	75.7723	77.0393
May-70	89	90.7177	85.5019	97.0133	126.1014	136.6830
Jun-70	116	90.7177	88.0616	97.0133	114.8522	150.6100
Jul-70	117	90.7177	109.1998	97.0133	109.6536	158.7803
Aug-70	274	90.7177	116.2344	97.0133	119.5669	219.8930
Sep-70	137	90.7177	262.3861	97.0133	208.5555	98.9493
Oct-70	171	90.7177	137.8308	97.0133	185.8300	114.7506
Nov-70	155	90.7177	143.9737	97.0133	195.4645	157.7524
Dec-70	143	90.7177	145.0455	97.0133	–	166.1216

Table 12.11 Forecast using Damped Holt method

Period	Demand	Forecast 1 period ahead^
t	Y_t	F_t
1	818	818.0000
2	833	824.3000
3	817	835.8020
4	818	829.1143
5	805	823.9495
6	801	811.7567
7	803	802.4828
8	805	799.7078
9	808	800.7737
10	817	804.2789
11	836	812.5885
12	855	829.5993
13	853	850.6078
14	851	858.8332

Table 12.12 Forecast using Theta method

	Time	Cars	Forecast
Year	t	Y_t	F_t
1993	1	13.31	
1994	2	13.6	
1995	3	13.93	13.848
1996	4	14.36	14.110
1997	5	14.72	14.440
1998	6	15.15	14.781
1999	7	15.6	15.158
2000	8	15.94	15.564
2001	9	16.31	15.944
2002	10	16.72	16.320
2003	11	17.19	16.709
2004	12	17.64	17.131
2005	13	18.12	17.564
2006	14	18.61	18.016
2007	15	19.15	18.483
2008	16	19.55	18.979
2009	17	20.02	19.437
2010	18	20.53	19.899
2011	19	20.96	20.381
2012	20	21.47	20.842
2013	21	22.11	21.322
2014	22	22.72	21.867
2015	23	23.3	22.437
2016	24	23.97	23.012
2017	**25**		**23.6244**

Table 12.13 Forecast using ARIMA method

Month	Production of Sofa (in Thousands)	Forecast*F_t
1	98	97.9020
2	82	92.7517
3	84	86.6899
4	85	87.2354
5	99	90.2576
6	90	92.3250
7	92	89.9272
8	83	89.0541
9	86	87.5629
10	90	89.2152
11	95	90.6584
12	91	90.7141
13	87	89.0086
14	99	89.5715
15	93	90.8333
16	82	87.9907
17	84	86.9563
18	88	89.0024
19	93	90.6564
20	83	90.5054
21	95	89.8928
22	93	91.6043
23	92	90.1840
24	92	89.4144
25	97	89.2052
26	88	88.5374
27	81	86.6061
28	93	87.6779
29	91	90.0032
30	81	88.9099
31	86	88.8512
32	81	90.9446
33	97	92.2831
34	88	93.9458
35	96	92.0745
36	96	92.2296
37	97	90.4160
38	90	88.5144
39	88	86.7300
40	93	86.9814
41	90	87.4583
42	84	86.7020
43	82	86.9679

(continued)

Table 12.13 (continued)

Month	Production of Sofa (in Thousands)	Forecast*F_t
44	82	88.7995
45	98	91.3190
46	91	92.8558
47	85	90.5459
48	86	90.3639
49	88	91.6516
50	90	92.5493
51		**92.84**
52		**92.47**
53		**91.39**

Table 12.14 Forecast using Croston and SBA methods

Months	Actual demand, Number	Croston ForecastˆF_t	SBA ForecastˆF_t
1	5	5.0000	4.0000
2	0	5.0000	4.0000
3	7	4.1429	3.3143
4	28	11.8387	9.4710
5	0	11.8387	9.4710
6	0	11.8387	9.4710
7	11	6.7942	5.4354
8	0	6.7942	5.4354
9	4	4.8438	3.8750
10	19	8.4280	6.7424
11	0	8.4280	6.7424

References

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. *International Journal of Forecasting, 27*, 672–688.

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Dordrecht: Kluwer Academic Publishers.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting, 16*, 521–530.

Babai, M. Z., Ali, M., & Nikolopoulos, K. (2012). Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *OMEGA: The International Journal of Management Science, 40*, 713–721.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2959–3038). Amsterdam: North-Holland.

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. San Francisco, Holden Day (revised ed. 1976).

Bozos, K., Nikolopoulos, K., & Bougioukos, N. (2008). Forecasting the value effect of seasoned equity offering announcements. In *28th international symposium on forecasting ISF 2008, June 22–25 2008*. France: Nice.

Brown, R. G. (1956). *Exponential smoothing for predicting demand*. Cambridge, MA: Arthur D. Little Inc.

Chatfield, C. (2005). Time-series forecasting. *Significance, 2*(3), 131–133.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly, 23*, 289–303.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*, 126.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica, 50*, 987–1008.

Goodwin, P. (2006). Decision Analysis for Management Judgement, 3rd Edition Chichester: Wiley.

Hanke, J. E., & Wichern, D. W. (2005). *Business forecasting* (8th ed.). Upper Saddle River: Pearson.

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society (B), 38*, 205–247.

Hibon, M., & Makridakis, S. (2000). The M3 competition: Results, conclusions and implications. *International Journal of Forecasting, 16*, 451–476.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages. O. N. R. Memorandum 52/1957. Pittsburgh: Carnegie Institute of Technology. Reprinted with discussion in 2004. *International Journal of Forecasting, 20*, 5–13.

Hyndman, R. J. (2014). *Forecasting – Principle and practices*. University of Western Australia. Retrieved July 24, 2017, from robjhyndman.com/uwa.

Johnston, F. R., Boylan, J. E., & Shale, E. A. (2003). An examination of the size of orders from customers, their characterization and the implications for inventory control of slow moving items. *Journal of the Operational Research Society, 54*(8), 833–837.

Jose, V. R. R., & Winkler, R. L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting, 24*(1), 163–169.

Keast, S., & Towler, M. (2009). *Rational decision-making for managers: An introduction*. Hoboken, NJ: John Wiley & Sons.

Kourentzes, N. (2014). Improving your forecast using multiple temporal aggregation. Retrieved August 7, 2017, from http://kourentzes.com/forecasting/2014/05/26/improving-forecasting-via-multiple-temporal-aggregation.

Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting, 30*, 291–302.

Leven and Segerstedt. (2004). Referred to in Syntetos and Boylan approximation section.

Lindsey, M., & Pavur, R. (2008). A comparison of methods for forecasting intermittent demand with increasing or decreasing probability of demand occurrences. In K. D. Lawrence & M. D. Geurts (Eds.), *Advances in business and management forecasting (advances in business and management forecasting)* (Vol. 5, pp. 115–132). Bingley, UK: Emerald Group Publishing Limited.

Makridakis, S., Hogarth, R., & Gaba, A. (2009). *Dance with chance: Making luck work for you*. London, UK: Oneworld Publications.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). New York: John Wiley and Sons.

Maris, K., Nikolopoulos, K., Giannelos, K., & Assimakopoulos, V. (2007). Options trading driven by volatility directional accuracy. *Applied Economics, 39*(2), 253–260.

Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., & Petropoulos, F. (2011a). The theta model: An essential forecasting tool for supply chain planning. *Advances in Automation and Robotics, 2*, 431–437.

Nikolopoulos, K., Syntetos, A., Boylan, J., Petropoulos, F., & Assimakopoulos, V. (2011b). ADIDA: An aggregate/disaggregate approach for intermittent demand forecasting. *Journal of the Operational Research Society, 62*, 544–554.

Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'Horses for Courses' in demand forecasting. *European Journal of Operational Research, 237*, 152–163.

Prestwich, S. D., Tarim, S. A., Rossi, R., & Hnich, B. (2014). Forecasting intermittent demand by hyperbolic-exponential smoothing. *International Journal of Forecasting, 30*(4), 928–933.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A. A., & Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics, 60*, 479–498.

Spithourakis, G. P., Petropoulos, F., Babai, M. Z., Nikolopoulos, K., & Assimakopoulos, V. (2011). Improving the performance of popular supply chain forecasting techniques: An empirical investigation. *Supply Chain Forum: An International Journal, 12*, 16–25.

Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics, 71*, 457–466.

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting, 21*, 303–314.

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting, 26*, 134–143.

Taylor, A. R. (1997). On the practical problems of computing seasonal unit root tests. *International Journal of Forecasting, 13*(3), 307–318.

Teunter, R. H., Syntetos, A., & Babai, Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research, 214*, 606–615.

Thomakos, D. D., & Nikolopoulos, K. (2014). Fathoming the theta method for a unit root process. *IMA Journal of Management Mathematics, 25*, 105–124.

Timmerman, A., & Granger, C. W. J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting, 20*, 15–27.

Tseng, F., Yu, H., & Tzeng, G. (2002). Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change, 69*, 71–87.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20*, 375–387.

Wright, G., & Goodwin, P. (1998). *Forecasting with judgement*. Chichester and New York: John Wiley and Sons.