

Time Series and Forecasting

A time series is a sequence of observations of a random variable. Hence, it is a stochastic process. Examples include the monthly demand for a product, the annual freshman enrollment in a department of a university, and the daily volume of flows in a river. Forecasting time series data is an important component of operations research because these data often provide the foundation for decision models. An inventory model requires estimates of future demands, a course scheduling and staffing model for a university requires estimates of future student inflow, and a model for providing warnings to the population in a river basin requires estimates of river flows for the immediate future.

Time series analysis provides tools for selecting a model that can be used to forecast of future events. Modeling the time series is a statistical problem. Forecasts are used in computational procedures to estimate the parameters of a model being used to allocate limited resources or to describe random processes such as those mentioned above. Time series models assume that observations vary according to some probability distribution about an underlying function of time.

Time series analysis is not the only way of obtaining forecasts. Expert judgment is often used to predict long-term changes in the structure of a system. For example, qualitative methods such as the Delphi technique may be used to forecast major technological innovations and their effects. Causal regression models try to predict dependent variables as a function of other correlated observable independent variables.

In this chapter, we only begin to scratch the surface of the field, restricting our attention to using historical time series data to develop time-dependent models. The methods are appropriate for automatic, short-term forecasting of frequently used information where the underlying causes of time variation are not changing markedly. In practice, forecasts derived by these methods are likely to be modified by the analyst upon considering information not available from the historical data.

Several methods are described in this chapter, along with their strengths and weaknesses. Although most are simple in concept, the computations required to estimate parameters and perform the analysis are tedious enough that computer implementation is essential. For a more detailed treatment of the field, the reader is referred to the bibliography.

22.1 Time Series Models

An example of a time series for 25 periods is plotted in Fig. 1 from the numerical data in the Table 1. The data might represent the weekly demand for some product. We use x to indicate an observation and the subscript t to represent the index of the time period. For the case of weekly demand the time period is measured in weeks. The observed demand for time t is specifically designated x_t . The lines connecting the observations on the figure are provided only to clarify the graph and otherwise have no meaning.

Table 1. Random Observations of Weekly Demand

Time	Observations									
1 - 10	4	16	12	25	13	12	4	8	9	14
11 - 20	3	14	14	20	7	9	6	11	3	11
20 - 25	8	7	2	8	8	10	7	16	9	4

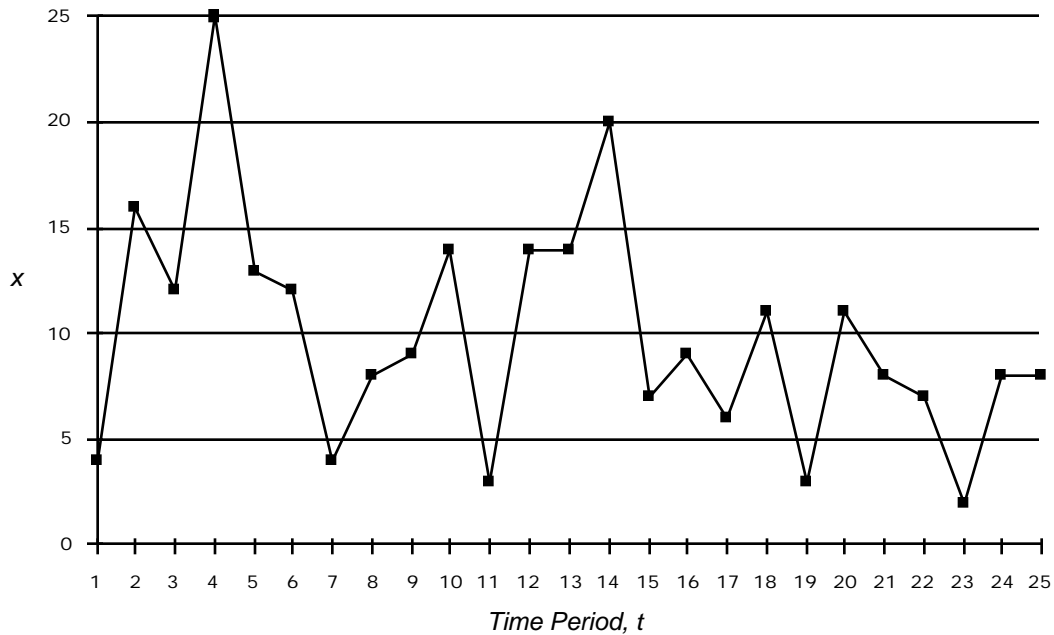


Figure 1. A time series of weekly demand

Mathematical Model

Our goal is to determine a model that explains the observed data and allows extrapolation into the future to provide a forecast. The simplest model suggests that the time series in Fig. 1 is a constant value b with variations about b determined by a random variable ε_t .

$$X_t = b + \varepsilon_t \quad (1)$$

The upper case symbol X_t represents the random variable that is the unknown demand at time t , while the lower case symbol x_t is a value that has actually been observed. The random variation ε_t about the mean value is called the *noise*, and is assumed to have a mean value of zero and a given variance. It is also common to assume that the noise variations in two different time periods are independent. Specifically

$$E[\varepsilon_t] = 0, \text{Var}[\varepsilon_t] = \sigma_\varepsilon^2, E[\varepsilon_t \varepsilon_w] = 0 \text{ for } t \neq w.$$

A more complex model includes a linear trend b_1 for the data.

$$X_t = b_0 + b_1 t + \varepsilon_t. \quad (2)$$

Of course, Eqs. (1) and (2) are special cases of a polynomial model.

$$X_t = b_0 + b_1 t + b_2 t^2 + \dots + b_n t^n + \varepsilon_t$$

A model for a seasonal variation might include transcendental functions. The cycle of the model below is 4. The model might be used to represent data for the four seasons of the year.

$$X_t = b_0 + b_1 \sin \frac{2\pi t}{4} + b_1 \cos \frac{2\pi t}{4} + \varepsilon_t$$

In every model considered here, the time series is a function only of time and the parameters of the models. We can write

$$X_t = f(b_0, b_1, b_2, \dots, b_n, t) + \varepsilon_t.$$

Because the value of f is a constant at any given time t and the expected value of ε_t is zero,

$$E[X_t] = f(b_0, b_1, b_2, \dots, b_n, t) \text{ and } \text{Var}[X_t] = V[\varepsilon_t] = \sigma_\varepsilon^2.$$

The model supposes that there are two components of variability for the time series; the mean value varies with time and the difference from the mean varies randomly. Time is the only factor affecting the mean value, while all other factors are subsumed in the noise component. Of course, these assumptions may not in fact be true, but this chapter is devoted to cases that can be abstracted to this simple form with reasonable accuracy.

One of the problems of time series analysis is to find the *best* form of the model for a particular situation. In this introductory discussion, we are primarily concerned about the simple constant or trend models. We leave the problem of choosing the best model to a more advanced text.

In the following subsections, we describe methods for fitting the model, forecasting from the model, measuring the accuracy of the forecast and forecasting ranges. We illustrate the discussion of this section with the *moving average* forecasting method. Several other methods are described later in the chapter.

Fitting Parameters of the Model

Once a model is selected and data are collected, it is the job of the statistician to estimate its parameters; i.e., to find parameter values that *best* fit the historical data. We can only hope that the resulting model will provide good predictions of future observations.

Statisticians usually assume that all values in a given sample are equally valid. For time series, however, most methods recognize that recent data are more accurate than aged data. Influences governing the data are likely to change with time so a method should have the ability of de-emphasizing old data while favoring new. A model estimate should be designed to reflect changing conditions.

In the following, the time series model includes one or more parameters. We identify the estimated values of these parameters with *hats* on the parameters. For instance,

$$\hat{b}_1, \hat{b}_2, \dots, \hat{b}_n.$$

The procedures also provide estimates of the standard deviation of the noise, call it σ_ε . Again the estimate is indicated with a hat, $\hat{\sigma}_\varepsilon$. We will see that there are several approaches available for estimating σ_ε .

To illustrate these concepts consider the data in Table 1. Say that the statistician has just observed the demand in period 20. She also has available the demands for periods 1 through 19. She does not know the future, so the data points shown as 21 through 25 are not available to her. The statistician thinks that the factors that influence demand are changing very slowly, if at all, and proposes the simple *constant* model for the demand given by Eq. (1)

With the assumed model, the values of demand are random variables drawn from a population with mean value b . The *best* estimator of b is the average of the observed data. Using all 20 points, the estimate is

$$\hat{b} = \frac{1}{20} \sum_{t=1}^{20} x_t = 11.3.$$

This is the best estimate for the 20 data points; however, we note that x_1 is given the same weight as x_{20} in the computation. If we think that the model is actually changing over time, perhaps it is better to use a method that gives less weight to old data and more weight to the new. One possibility is to include only recent data in the estimate. Using the last 10 observations and the last 5, we obtain

$$\hat{b} = \frac{1}{10} \sum_{t=10}^{20} x_t = 11.2 \text{ and } \hat{b} = \frac{1}{5} \sum_{t=15}^{20} x_t = 9.4.$$

which are called *moving averages*.

Which is the better estimate for the application? We really can't tell at this point. The estimator that uses all data points will certainly be the best if the time series follows the assumed model; however, if the model is only approximate and the situation is actually changing, perhaps the estimator with only 5 data points is better.

In general, the moving average estimator is the average of the last m observations.

$$\hat{b} = \frac{1}{m} \sum_{i=k}^t x_i,$$

where $k = t - m + 1$. The quantity m is the *time range* and is the parameter of the method.

Forecasting from the Model

The main purpose of modeling a time series is to make forecasts which are then used directly for making decisions, such as ordering replenishments for an inventory system or developing staff schedules for running a production facility. They might also be used as part of a mathematical model for a more complex decision analysis.

In the analysis, let the current time be T , and assume that the demand data for periods 1 through T are known. Say we are attempting to forecast the demand at time $T + \tau$ in the example presented above. The unknown demand is the random variable $X_{T+\tau}$, and its ultimate realization is $x_{T+\tau}$. Our forecast of the realization is $\hat{x}_{T+\tau}$. Of course, the best that we can hope to do is estimate the mean value of $X_{T+\tau}$. Even if the time series actually follows the assumed model, the future value of the noise is unknowable.

Assuming the model is correct

$$X_{T+\tau} = E[X_{T+\tau}] + \varepsilon_t \text{ where } E[X_{T+\tau}] = f(b_0, b_1, b_2, \dots, b_n, T+\tau).$$

When we estimate the parameters from the data for times 1 through T , we have an estimate of the expected value for the random variable as a function of τ . This is our forecast.

$$\hat{x}_{T+\tau} = f(\hat{b}_0, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_n, T+\tau).$$

Using a specific value of τ in this formula provides the forecast for period $T+\tau$. When we look at the last T observations as only one of the possible time series that could have been obtained from the model, the forecast is a random variable. We should be able to describe the probability distribution of the random variable, including its mean and variance.

For the moving average example, the statistician adopts the model

$$X_t = b + \varepsilon_t.$$

Assuming T is 20 and using the moving average with 10 periods, the estimated parameter is

$$\hat{b} = 11.2.$$

Because this model has a constant expected value over time, the forecast is the *same* for all future periods.

$$\hat{x}_{T+\tau} = \hat{b} = 11.2 \text{ for } \tau = 1, 2, \dots$$

Assuming the model is correct, the forecast is the average of m observations all with the same mean and standard deviation σ_ε . Because the noise is normally distributed, the forecast is also normally distributed with mean b and standard deviation

$$\frac{\sigma_\varepsilon}{\sqrt{m}}$$

Measuring the Accuracy of the Forecast

The error in a forecast is the difference between the realization and the forecast,

$$e_\tau = x_{T+\tau} - \hat{x}_{T+\tau}.$$

Assuming the model is correct,

$$e_\tau = E[X_{T+\tau}] + \varepsilon_t - \hat{x}_{T+\tau}.$$

We investigate the probability distribution of the error by computing its mean and variance. One desirable characteristic of the forecast $\hat{x}_{T+\tau}$ is that it be *unbiased*. For an unbiased estimate, the expected value of the forecast is the same as the expected value of the time series. Because ε_t is assumed to have a mean of zero, an unbiased forecast implies

$$E[e_\tau] = 0.$$

Moreover, the fact that the noise is independent from one period to the next means that the variance of the error is

$$\begin{aligned} \text{Var}[e_\tau] &= \text{Var}[E[X_{T+\tau}] - \hat{x}_{T+\tau}] + \text{Var}[\varepsilon_{T+\tau}] \\ \sigma_\varepsilon^2(\tau) &= \sigma_E^2(\tau) + \sigma^2. \end{aligned}$$

As we see, this term has two parts: (1) that due to the variance in the estimate of the mean $\sigma_E^2(\tau)$, and (2) that due to the variance of the noise σ^2 . Due to the inherent inaccuracy of the statistical methods used to estimate the model parameters and the possibility that the model is not exactly correct, the variance in the estimate of the means is an increasing function of τ .

For the moving average example,

$$\sigma_\varepsilon^2(\tau) = \frac{\sigma^2}{m} + \sigma^2 = \sigma^2[1 + (1/m)].$$

The variance of the error is a decreasing function of m . Obviously, the smallest error comes when m is as large as possible, *if the model is correct*. Unfortunately, we cannot be sure that the model is correct, and we set m to smaller values to reduce the error due to a poorly specified model.

Using the same forecasting method over a number of periods allows the analyst to compute measures of quality for the forecast for given values of τ . The forecast error, e_t , is the difference between the forecast and the observed value. For time t ,

$$e_t = x_t - \hat{x}_t.$$

Table 2 shows a series of forecasts for periods 11 through 20 using the data from Table 1. The forecasts are obtained with a moving average for $m = 10$ and $\tau = 1$. We make a forecast at time t with the calculation

$$\hat{x}_{t+1} = \frac{1}{10} \sum_{i=t-9}^t x_i.$$

Although in practice one might round the result to an integer, we keep fractions here to observe better statistical properties. The error of the forecast is the difference between the forecast and the observation.

Table 2. Forecast Error for a Moving Average

Data	Time, t									
	11	12	13	14	15	16	17	18	19	20
Observation	3	14	14	20	7	9	6	11	3	11
Forecast	11.7	11.6	11.4	11.6	11.1	10.5	10.2	10.4	10.7	10.1
Error	-8.7	2.4	2.6	8.4	-4.1	-1.5	-4.2	0.6	-7.7	0.9

One common measure of forecasting error is the mean absolute deviation, MAD.

$$\text{MAD} = \frac{\sum_{i=1}^n |e_i|}{n}$$

where n error observations are used to compute the mean.

The sample standard deviation of error is also a useful measure,

$$s_e = \frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n - p} = \frac{\sum_{i=1}^n e_i^2 - n(\bar{e})^2}{n - p}$$

where \bar{e} is the average error and p is the number of parameters estimated for the model. As n grows, the MAD provides a reasonable estimate of the sample standard deviation

$$s_e \approx 1.25 \text{ MAD}$$

From the example data we compute the MAD for the 10 observations.

$$\text{MAD} = (8.7 + 2.4 + \dots + 0.9)/10 = 4.11.$$

The sample error standard deviation is computed as follows.

$$\bar{e} = (-8.7 + 2.4 \dots 0.9)/10 = -1.13.$$

$$s_e^2 = \frac{(-8.7^2 + 2.4^2, \dots, 0.9^2) - 10(-1.13)^2}{9} = 27.02$$

$$s_e = 5.198.$$

We see that $1.25(\text{MAD}) = 5.138$ is approximately equal to the sample standard deviation. Because it is easier to compute the MAD, this measure is used in our examples.¹

The value of s_e^2 for a given value of τ is an estimate of the error variance, $\sigma_e^2(\tau)$. It includes the combined effects of errors in the model and the noise. If one assumes that the random noise comes from a normal distribution, an interval estimate of the forecast can be computed using the *Students t* distribution.

$$\hat{x}_{T+\tau} \pm t_{\alpha/2} s_e(\tau)$$

The parameter $t_{\alpha/2}$ is found in a *Students t* distribution table with $n - p$ degrees of freedom.

¹The time series used as an example is simulated with a constant mean. Deviations from the mean are normally distributed with mean 0 and standard deviation 5. One would expect an error standard deviation of $5\sqrt{1+1/10} = 5.244$. The observed statistics are not far from this value. Of course, a different realization of the simulation will yield different statistical values.

22.2 Analysis of the Constant Model

In this section, we investigate two procedures for estimating and forecasting based on a constant model. The next section considers a model involving a linear trend. For all cases we assume the data from previous periods: x_1, x_2, \dots, x_T , is available and will be used to provide the forecast.

To illustrate the methods, we propose a data set that incorporates changes in the underlying mean of the time series. Figure 2 shows the time series used for illustration together with the mean demand from which the series was generated. The mean begins as a constant at 10. Starting at time 21, it increases by one unit in each period until it reaches the value of 20 at time 30. Then it becomes constant again. The data is simulated by adding to the mean, a random noise from a normal distribution with 0 mean and standard deviation 3. Table 3 shows the simulated observations. When we use the data in the table, we must remember that at any given time, only the past data are known.

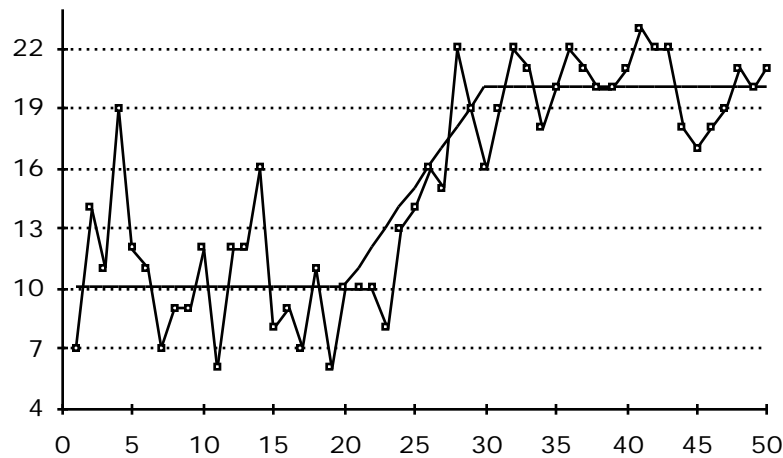


Figure 2. Simulated data for model with a linear trend

Table 3. Simulated Observations

Time	Observations									
1 - 10	7	14	11	19	12	11	7	9	9	12
11 - 20	6	12	12	16	8	9	7	11	6	10
21 - 30	10	10	8	13	14	16	15	22	19	16
31 - 40	19	22	21	18	20	22	21	20	20	21
41 - 50	23	22	22	18	17	18	19	21	20	21

Moving Average

This method assumes that the time series follows a constant model, i.e., Eq. (1) given by $X_t = b + \varepsilon_t$. We estimate the single parameter of the model as average of the last m observations.

$$\hat{b} = \frac{1}{m} \sum_{i=k}^t x_i,$$

where $k = t - m + 1$. The forecast is the same as the estimate.

$$\hat{x}_{T+\tau} = \hat{b} \text{ for } \tau > 0$$

The moving average forecasts should not begin until m periods of data are available. To illustrate the calculations, we find the estimate of the parameter at $t = 20$ using $m = 10$.

$$\hat{b} = \frac{1}{10} \sum_{i=11}^{20} x_i = 9.7$$

The estimates of the model parameter, \hat{b} , for three different values of m are shown together with the mean of the time series in Fig. 3. The figure shows the moving average estimate of the mean at each time and not the forecast. The forecasts would shift the moving average curves to the right by τ periods.

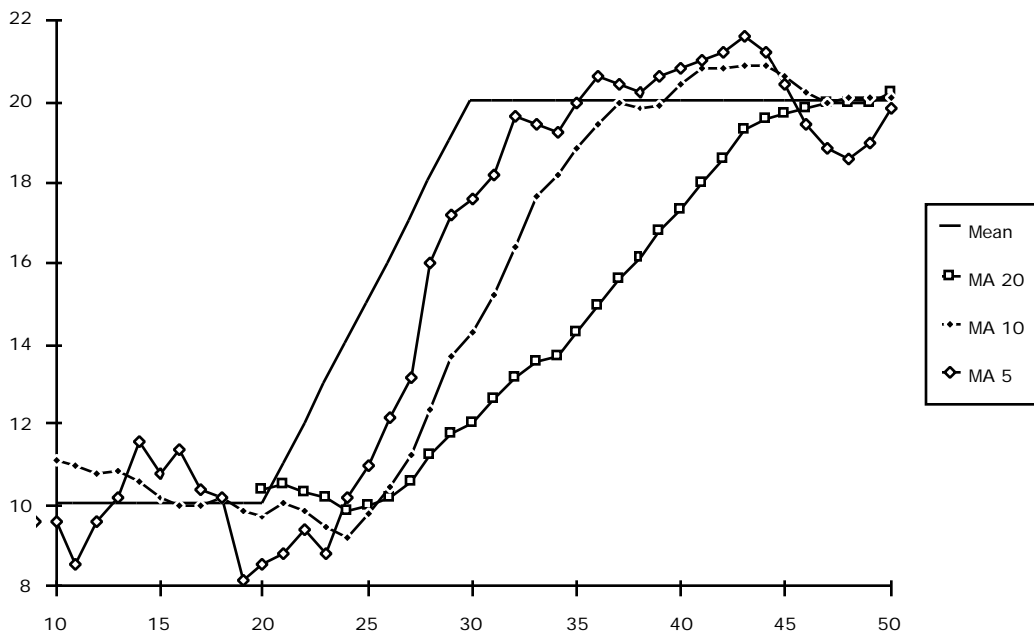


Figure 3. Moving average response to changes

One conclusion is immediately apparent from Fig. 3. For all three estimates the moving average lags behind the linear trend, with the lag increasing with m . Because of the lag, the moving average underestimates the observations as the mean is increasing. The lag in time and the bias introduced in the estimate are

$$\text{lag} = (m - 1)/2, \text{ bias} = -a(m - 1)/2.$$

The moving average forecast of τ periods into the future increases these effects.

$$\text{lag} = \tau + (m - 1)/2, \text{ bias} = -a[\tau + (m - 1)/2].$$

We should not be surprised at this result. The moving average estimator is based on the assumption of a constant mean, and the example has a linear trend in the mean. Because real time series will rarely obey the assumptions of any model exactly, we should be prepared for such results.

We can also conclude from Fig. 3 that the variability of the noise has the largest effect for smaller m . The estimate is much more volatile for the moving average of 5 than the moving average of 20. We have the conflicting directives to increase m to reduce the effect of variability due to the noise, and to decrease m to reduce the effect of the variability due to changes in mean.

Exponential Smoothing for the Constant Model

Again, this method assumes that the time series follows a *constant model*, $X_t = b + \varepsilon_t$. The parameter value b is estimated as the weighted average of the last observation and the last estimate.

$$\hat{b}_T = \alpha x_T + (1 - \alpha)\hat{b}_{T-1},$$

where α is a parameter in the interval $[0, 1]$. Rearranging, obtains an alternative form.

$$\hat{b}_T = \hat{b}_{T-1} + \alpha(x_T - \hat{b}_{T-1}).$$

The new estimate is the old estimate plus a proportion of the observed error.

Because we are supposing a constant model, the forecast is the same as the estimate.

$$\hat{x}_{T+\tau} = \hat{b} \text{ for } \tau > 0.$$

We illustrate the method using the parameter value $\alpha = 0.2$. The average of the first 10 periods was used as initialize the estimate at time 0. The first 10 observations were then used to *warm up* the procedure. Subsequent observations and estimates are shown in Table 4.

Table 4. Results of the Exponential Moving Forecast

Time	11	12	13	14	15	16	17	18	19	20
Observation	6	12	12	16	8	9	7	11	6	10
Estimate	10.7	9.763	10.21	10.57	11.65	10.92	10.54	9.831	10.06	9.252

At time 21 we observe the value 10, so the estimate of the mean at time 21 is

$$\hat{b}_{21} = \hat{b}_{20} + 0.2(x_{21} - \hat{b}_{20}) = 9.252 + 0.2(10 - 9.252) = 9.412.$$

Only two data elements are required to compute the new estimate, the observed data and the old estimate. This contrasts with the moving average which requires m old observations to be retained for the computation.

Replacing \hat{b}_{T-1} with its equivalent, we find that the estimate is

$$\hat{b}_T = \alpha x_T + (1 - \alpha)(\alpha x_{T-1} + (1 - \alpha)\hat{b}_{T-2}).$$

Continuing in this fashion, we find that the estimate is really a weighted sum of all past data.

$$\hat{b}_T = \alpha(x_T + (1 - \alpha)x_{T-1} + (1 - \alpha)^2 x_{T-2} + \cdots + (1 - \alpha)^{T-1} x_1).$$

Because α is a fraction, recent data has a greater weight than more distant data. The larger values of α provide relatively greater weight to more recent data than smaller values of α . Figure 4 shows the parameter estimates obtained for different values of α together with the mean of the time series.

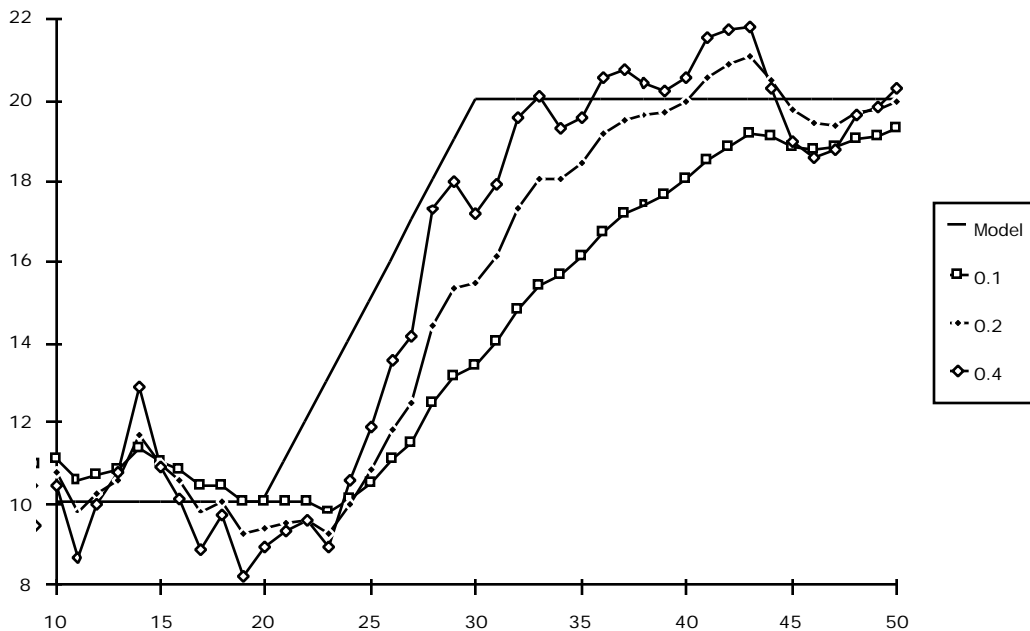


Figure 4. Exponential smoothing for the example time Series

A lag characteristic, similar to the one associated with the moving average estimate, can also be seen in Fig. 4. In fact, one can show comparable results

$$lag = \frac{1 - \alpha}{\alpha}, \quad bias = -\frac{\alpha(1 - \alpha)}{\alpha}$$

For larger value of α we obtain a greater lag in response to the trend.

To investigate the error associated with exponential smoothing we again note that the error is

$$e_{\tau} = x_{T+\tau} - \hat{x}_{T+\tau}.$$

Assuming the model is correct, we have the following.

$$e_{\tau} = E[X_{T+\tau}] + \varepsilon_{\tau} - \hat{x}_{T+\tau}.$$

$$E[e_{\tau}] = E[X_{T+\tau}] + E[\varepsilon_{\tau}] - E[\hat{x}_{T+\tau}]$$

$$E[\hat{x}_{T+\tau}] = \alpha b [1 + (1 - \alpha) + (1 - \alpha)_2 + \dots + (1 - \alpha)^{T-1}].$$

As the T goes to infinity, the series in the brackets goes to $1/\alpha$, and we find that

$$E[\hat{x}_{T+\tau}] = b \text{ and } E[e_{\tau}] = 0.$$

Because the estimate at any time is independent of the noise at a future time, the variance of the error is

$$\text{Var}[e_{\tau}] = \text{Var}[\hat{x}_{T+\tau}] + \text{Var}[\varepsilon_{T+t}]$$

$$\sigma_e^2(\tau) = \sigma_E^2(\tau) + \sigma^2.$$

The variance of the error has two parts, the first due to the variance in the estimate of the mean $\sigma_E^2(\tau)$, and the second due to the variance of the noise σ^2 . For exponential smoothing,

$$\sigma_E^2(\tau) = \frac{\alpha\sigma^2}{2 - \alpha}.$$

Thus assuming the model is correct, the error of the estimate increases as α increases.

This result shows an interesting correspondence to the moving average estimator. Setting the estimating error for the moving average and the exponential smoothing equal, we obtain.

$$\sigma_E^2(\tau) = \frac{\sigma^2}{m} = \frac{\alpha\sigma^2}{2 - \alpha}.$$

Solving for α in terms of m , we obtain the relative values of the parameters that give the same error

$$\alpha = \frac{2}{m + 1}.$$

Thus the parameters used in the moving average illustrations of Fig. 3 ($m = 5, 10, 20$) and roughly comparable to the parameters used for exponential smoothing in Fig. 4 ($\alpha = 0.4, 0.2, 0.1$). Using this same relation between the parameters of the two methods, we find that the biased introduced by the trend will also be the same.

22.3 Analysis of the Linear Trend Model

One way to overcome the problem of responding to trends in the time series is to use a model that explicitly includes a trend component,

$$X_t = a + bt + \varepsilon_t$$

which is Eq. (2). In the following, we use a linear model of the form

$$X_T = a_T + b_T(t - T) + \varepsilon_t.$$

Now, we must estimate two parameters \hat{a}_T and \hat{b}_T from the observations previous to time T . Forecasts will be made by projecting the estimated model into the future.

$$\hat{x}_{T+\tau} = \hat{a}_T + \tau \hat{b}_T \text{ for } \tau > 0$$

Regression Analysis

One obvious way to estimate the parameters is to *fit* the last m observations with a straight line. This is similar to the moving average idea, except that in addition to a constant term we also estimate the slope.

The formulas for determining the parameter estimates to minimize the least squares differences between the line and the observations are well known. The results are repeated here. The last m observations are used for the estimate: $x_{T-m+1}, x_{T-m+2}, \dots, x_T$.

Define the following sums

$$S_1(t) = \sum_{k=t-m+1}^t x_k$$

$$S_2(t) = \sum_{k=t-m+1}^t (k-t)x_k.$$

The estimates of the parameters are determined by operations on these sums.

$$\hat{a}_T = \frac{6}{m(m+1)}S_2(T) + \frac{2(2m-1)}{m(m+1)}S_1(T)$$

$$\hat{b}_T = \frac{12}{m(m^2-1)}S_2(T) + \frac{6}{m(m+1)}S_1(T)$$

The expressions for the sums are awkward for spreadsheet computation. Nevertheless, the computations can be carried out easily in a sequential fashion by noting

$$S_1(t) = S_1(t-1) - x_{t-m} + x_t,$$

$$S_2(t) = S_2(t-1) - S_1(t-1) + mx_{t-m}.$$

As with the moving average estimate, the regression method requires that the last m observations to be saved for computation.

To illustrate the computations, we derive estimates for $T = 20$. The relevant quantities are shown in Table 5.

Table 5. Illustration of Linear Model

Data	Time, t										Total
	11	12	13	14	15	16	17	18	19	20	
x_t	6	12	12	16	8	9	7	11	6	10	97
$(t - T)x_t$	-54	-96	-84	-96	-40	-36	-21	-22	-6	0	-455

$$\hat{a}_T = (0.05455)(-455) + (0.34545)(97) = 8.69$$

$$\hat{b}_T = (0.01212)(-455) + (0.05455)(97) = -0.22$$

Forecasts from time 20 are then computed from

$$\hat{x}_{20+\tau} = 8.69 - 0.22\tau \quad \text{for } \tau > 0.$$

Figure 5 shows the regression estimates of \hat{a} for three different values of m . Although there is a lag in response when the ramp begins, the estimate gradually grows to meet the ramp when $m = 5$ and $m = 10$. When $m = 20$, there is insufficient time for the computations to match the slope before the mean value becomes constant again. We observe considerably more variability for the same values of m when compared to the moving average estimates in Fig. 3. By allowing the flexibility of responding to a trend, we have increased the variability of the estimates when the time series is constant.

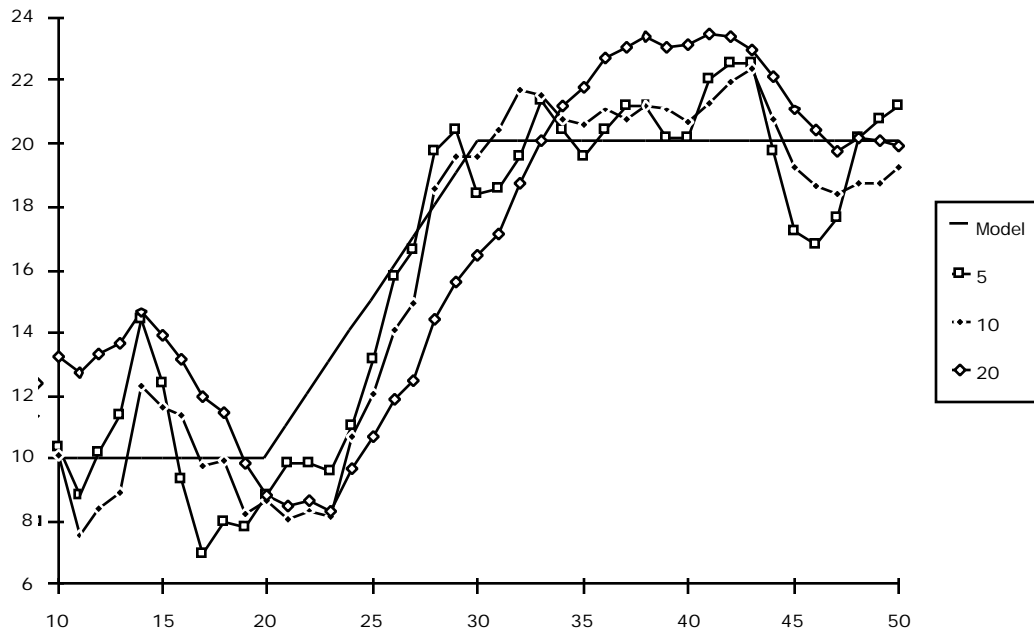


Figure 5. The linear regression estimate for the time series

The ability of the regression estimate to respond to a trend is more clearly illustrated when we remove the noise from the time series (the noise variance is set to 0). The result is shown in Fig. 6 where the estimate adjusts to the changing mean. Because of a lag effect, there are periods of over- and under-correction after the points in time when the slope changes.

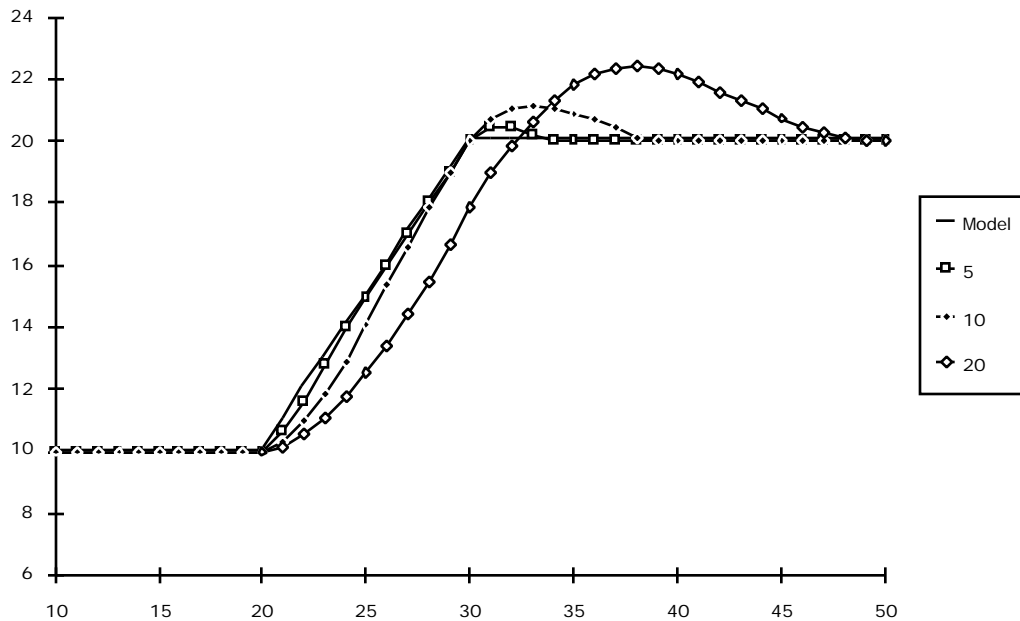


Figure 6. Linear regression with zero noise variance

Exponential Smoothing Adjusted for Trend

There is also a variation of the exponential smoothing method that explicitly accounts for a trend component. Again we assume the linear model

$$X_T = a_T + b_T(t - T) + \varepsilon_t.$$

The new method simultaneously estimates the constant and trend component using two parameters α and β .

$$\hat{a}_T = \alpha x_T + (1 - \alpha)(\hat{a}_{T-1} + \hat{b}_{T-1})$$

$$\hat{b}_T = \beta (\hat{a}_T - \hat{a}_{T-1}) + (1 - \beta)\hat{b}_{T-1}$$

Forecasts are made with the expression

$$\hat{x}_{T+\tau} = \hat{a}_T + \hat{b}_T\tau.$$

At any time T , only three pieces of information are necessary to compute the estimates \hat{a}_{T-1} , \hat{b}_{T-1} , and x_T . The exponential smoothing method is much simpler to implement than the regression method.

There is justification for expressing both smoothing factors in terms of a single parameter. Here

$$\alpha = 1 - (1 - \delta)^2, \beta = \frac{\delta^2}{1 - (1 - \delta)^2}.$$

We use these formulas in the following computations. The values of δ and the associated smoothing parameters are shown in Table 6.

Table 6. Data for Exponential Smoothing Example

Parameter	Case 1	Case 2	Case 3
δ	0.4	0.2	0.1
α	0.64	0.36	0.19
β	0.25	0.1111	0.0526

For purposes of illustration, we used a regression model with $m = 10$ to find initial estimates of the constant and trend components. Starting at time 10 we then applied the exponential smoothing equations. The first 10 estimates are meaningless and constitute a *warm up* period for the method.

Assuming $\delta = 0.2$, at time 20 we have

$$x_{20} = 10, \hat{a}_{19} = 8.23936, \hat{b}_{19} = -0.253$$

$$\hat{a}_{20} = 0.36x_{20} + (1 - 0.36)(\hat{a}_{19} + \hat{b}_{19}) = 8.71$$

$$\hat{b}_{20} = 0.1111(\hat{a}_{20} - \hat{a}_{19}) + (1 - 0.1111)\hat{b}_{19} = -0.172$$

$$\text{Forecasting: } \hat{x}_{20+\tau} = 8.71 + 0.172\tau.$$

Estimates of the time series using three values of the parameter are shown in Fig. 7. Again we see that for greater values of δ , the estimate begins to track the mean value after an initial overshoot when the trend begins. The lower values of δ have less variability during the constant portions of the time series, but react to the trend more slowly.

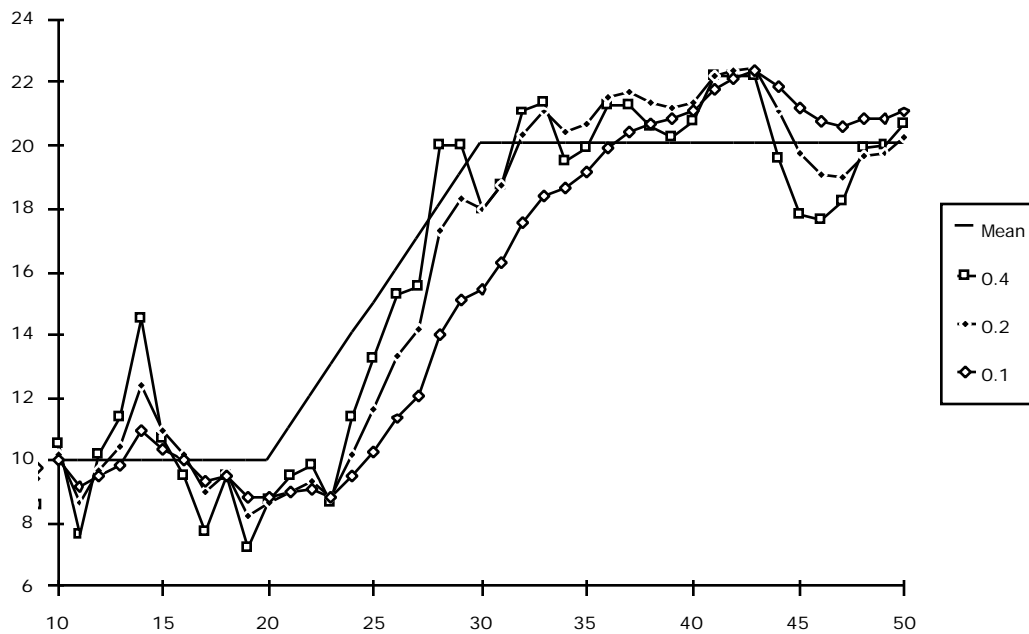


Figure 7. Example with estimates using exponential smoothing with a trend

Compared to the regression model, the exponential smoothing method never entirely forgets any part of its past. Thus it may take longer to recover in the event of a perturbation. This is illustrated in Fig. 8 where the variance of the noise is set to 0.

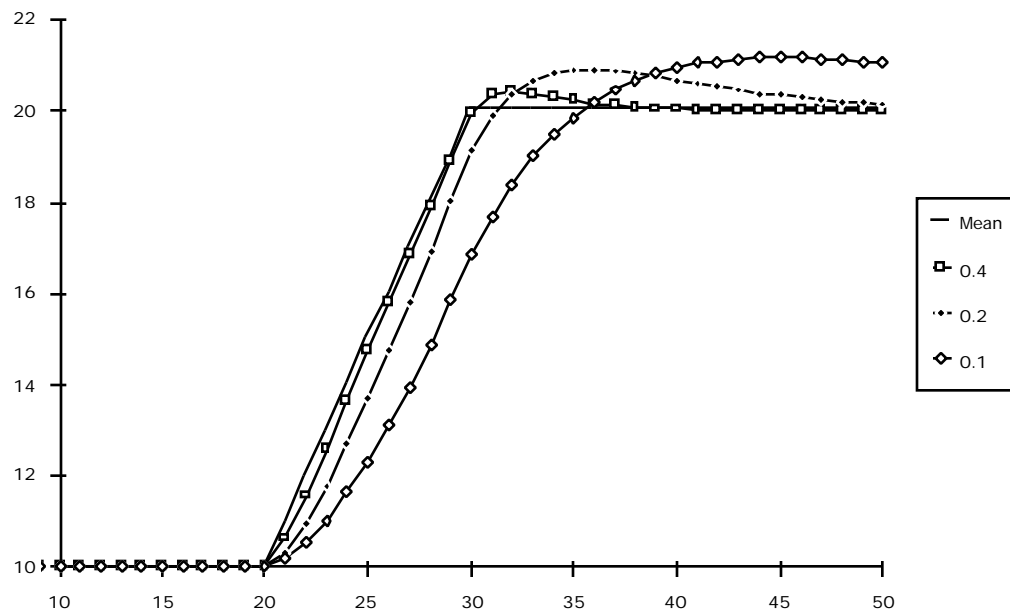


Figure 8. Exponential smoothing with 0 noise variance

22.4 Selecting a Forecasting Method

The selection of a forecasting method is a difficult task that must be based in part on knowledge concerning the quantity being forecast. We can, however, point out some simple characteristics of the methods that have been described. With forecasting procedures, we are generally trying to recognize a change in the underlying process of a time series while remaining insensitive to variations caused by purely random effects. The goal of planning is to respond to fundamental changes, not to spurious effects.

With a method based purely on historical data, it is impossible to filter out all the noise. The problem is to set parameters that find an acceptable tradeoff between the fundamental process and the noise. If the process is changing very slowly, both the moving average and the regression approach should be used with a long stream of data. For the exponential smoothing method, the value of α should be small to de-emphasize the most recent observations. Stochastic variations will be almost entirely filtered out.

If the process is changing rapidly with a linear trend, the moving average and the exponential smoothing methods are at a disadvantage because they are not designed to recognize trends. Because of the rapid changes, the time range of the moving average method must be set small and the α parameter of the exponential smoothing method must be set to a larger value so that the forecasts will respond to the change. Nevertheless, these two methods will always fall behind a linear trend. The forecasts will never converge to a trend line even if there is no random variation. Of course, with the adjustment of parameters to allow a response to a process change, the forecasts become more sensitive to random effects.

The exponential smoothing method with a trend adjustment and the regression method are both designed to respond to a linear trend and will eventually converge to a trend line. Thus in the absence of a change in trend, the time range of the regression data can be large, and the α and β values of the exponential smoothing method can be small, thus reducing the random effects.

If the process is changing rapidly with rapid changes in the linear trend, each of the methods described in the chapter will have trouble, because it is difficult to separate changes in the process from random changes. The time ranges must be set small for the moving average and regression methods, resulting in sensitivity to random effects. Similarly, the α and β parameters for exponential smoothing must be set to larger values with a corresponding increase in sensitivity to randomness.

Both the moving average and regression methods have the disadvantage that they are most accurate with respect to forecasts in the middle of the time range. Unfortunately, all interesting forecasts are in the future, outside the range of the data. With all methods, though, the accuracy of the results decreases with the distance into the future one wishes to forecast.

22.5 Exercises

- Use the data in Table 3 and the analytical results presented in Section 22.2. Assume that you have observed the time series for 21 periods and let the next data point in the series be $x_{22} = 10$. Update each of the forecasts described in Section 22.2 for $t = 23$.
- Use the data in Table 3 for times 35 through 44.

Time, t	35	36	37	38	39	40	41	42	43	44
Observation	53	55	44	41	48	42	33	38	26	23

Provide forecasts for times 45 through 50 using exponential smoothing with and without a trend adjustment. For purposes of exponential smoothing use

$$\alpha = 0.3 \text{ and } \hat{b}_{43} = 36.722.$$

For exponential smoothing with the trend adjustment, use

$$\alpha = 0.3, \beta = 0.2, \hat{a}_{43} = 32.22, \hat{b}_{43} = -2.632.$$

- The table below shows 60 observations from a time series.

Time	Observations									
1 - 10	99	104	112	85	95	98	91	87	112	103
11 - 20	106	96	109	69	78	98	74	79	65	93
21 - 30	58	76	70	61	75	58	63	32	43	41
31 - 40	25	44	25	54	44	49	39	35	49	56
41 - 50	49	45	38	45	44	50	62	51	55	58
51 - 60	58	49	61	65	50	71	77	87	70	68

Using the data provided for times 1 through 10, generate a forecast for time 11. Then sequentially observe x_{11} through x_{20} (as given in the table), and after each observation x_t , forecast the mean of the next observation $E[\hat{x}_{t+1}]$. Compare the forecasts with the actual data and compute the mean absolute deviation of the 10 observations. Do this for the following cases.

- Moving average with $n = 5$
- Moving average with $n = 10$
- Regression with $n = 5$
- Regression with $n = 10$
- Exponential smoothing with $\alpha = 0.2$ and $\hat{b}_9 = 100$.
- Exponential smoothing with $\alpha = 0.1$ and $\hat{b}_9 = 100$
- Exponential smoothing with $\alpha = 0.2$, $\beta = 0.2$, $\hat{a}_9 = 100$, $\hat{b}_9 = 0$

h. Exponential smoothing with $\alpha = 0.1$, $\beta = 0.1$, $\hat{a}_9 = 100$, $\hat{b}_9 = 0$

Note: The data given for the exponential smoothing cases are not computed from the data but are the model values for this point.

Computer Problems

For the following exercises, implement the appropriate forecasting methods using a spreadsheet program such as Excel. Try to use the most efficient approach possible when designing the calculations.

4. The table below shows the random variations added to the model of Section 22.2 to obtain the data of Table 3. Double the values of these deviations and add them to the model to obtain a new set of data. Repeat the computations performed in Sections 22.2 and 22.3, and compare the mean residuals for the four forecasting methods (moving average and exponential smoothing with and without a trend).

Time	Variations									
1 - 10	-3	4	1	9	2	1	-3	-1	-1	2
11 - 20	-4	2	2	6	-2	-1	-3	1	-4	0
21 - 30	-1	-2	-5	-1	-1	0	-2	4	0	-4
31 - 40	-1	2	1	-2	0	2	1	0	0	1
41 - 50	3	2	2	-2	-3	-2	-1	1	0	1

5. Use the data from Exercise 3 and experiment with the parameters of all methods discussed in the chapter. Try to find the best parameters for this data set. Compare the mean absolute deviations for the best parameters found.
6. The data below is simulated from a model that has a constant value of 20 for times 1 through 15. At times 16 through 30, the model jumps to 40. For times 31 through 50, the model goes back to 20. Experiment with the parameters of the forecasting methods described in the chapter to find the one that best forecasts this time series.

Time	Observations									
1 - 10	21	24	16	21	24	25	24	25	21	20
11 - 20	20	21	24	20	17	35	46	48	44	43
21 - 30	37	49	39	42	46	40	40	32	46	49
31 - 40	18	31	24	28	23	27	18	17	26	22
41 - 50	15	22	12	11	25	22	17	14	21	20

Bibliography

Box, G.E.P and G.M. Jenkins, *Time Series Analysis, Forecasting, and Control*, Holden-Day, Oakland, CA, 1970.

Box, G.E.P and G.M. Jenkins and G.D. Reinsel, *Time Series Analysis, Forecasting, and Control*, Third Edition, Prentice Hall, Englewood Cliffs, NJ, 1993.

Chatfield, C., *The Analysis of Time Series: An Introduction*, Fifth Edition, Chapman & Hall, Boca Raton, FL, 1996.

Fuller, W.A., *Introduction to Statistical Time Series*, Second Edition, John Wiley & Sons, New York, 1996.

Gaynor, P.E. and R.C. Kirkpatrick, *Introduction to Time-Series Modeling and Forecasting in Business and Economics*, McGraw-Hill, New York, 1994.

Shumway, R.H. and D.S. Stoffer, *Time Series Analysis and Its Applications*, Springer-Verlag, New York, 2000.

West, M. and J. Harrison, *Baysian Forecasting and Dynamic Models*, Second Edition, Springer-Verlag, New York, 1997.