4 – Exponential Smoothing Methods

If you have to forecast, forecast often.

EDGAR R. FIEDLER, *American economist*

Videos about Smoothing Methods

Galit Shmueli

1. Moving Average for visualization
2. Moving Average for forecasting
3. Differencing
4. Simple exponential smoothing (SES)
5. Holt's exponential smoothing
6. Winter's exponential smoothing
7. Automation and prediction intervals

Signal and Noise

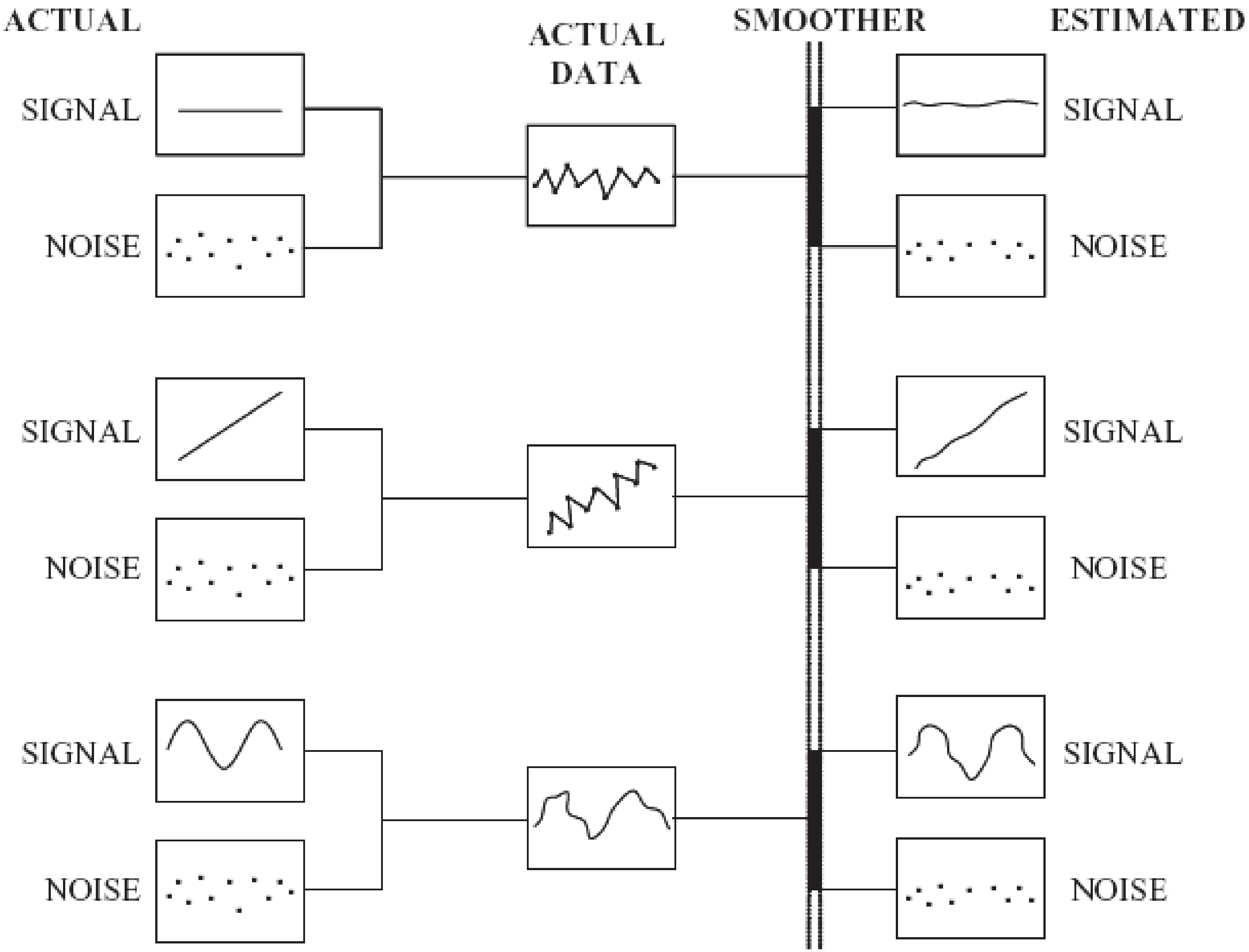
We can often think of a data set as consisting of two distinct components: **signal** and **noise**. Signal represents any pattern caused by the intrinsic dynamics of the process from which the data is collected. These patterns can take various forms from a simple constant process to a more complicated structure that cannot be extracted visually or with any basic statistical tools. The constant process, for example, is represented as

$$y_t = \mu + \varepsilon_t \tag{4.1}$$

where μ represents the underlying constant level of system response and ε_t is the noise at time t. The ε_t are often assumed to be uncorrelated with mean 0 and constant variance σ_ε^2.

Smoothing can be seen as a technique to separate the signal and the noise as much as possible and in that a smoother acts as a filter to obtain an "estimate" for the signal.

Smoothing a data set



Smoothing a Constant Process

A constant process can be smoothed by replacing the current observation with the best estimate for μ. Using the least squares criterion, we define the error sum of squares, SS, for the constant process as

$$SS_E = \sum_{t=1}^{T} (y_t - \mu)^2$$

The least squares estimate of μ can be found by setting the derivative of SS with respect to μ to 0. This gives

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t \tag{4.2}$$

where $\hat{\mu}$ is the least squares estimate of μ. Equation (4.2) shows that the least squares estimate of μ is indeed the average of observations up to time T.

Dow Jones Index (1999-2001) – Constant Process

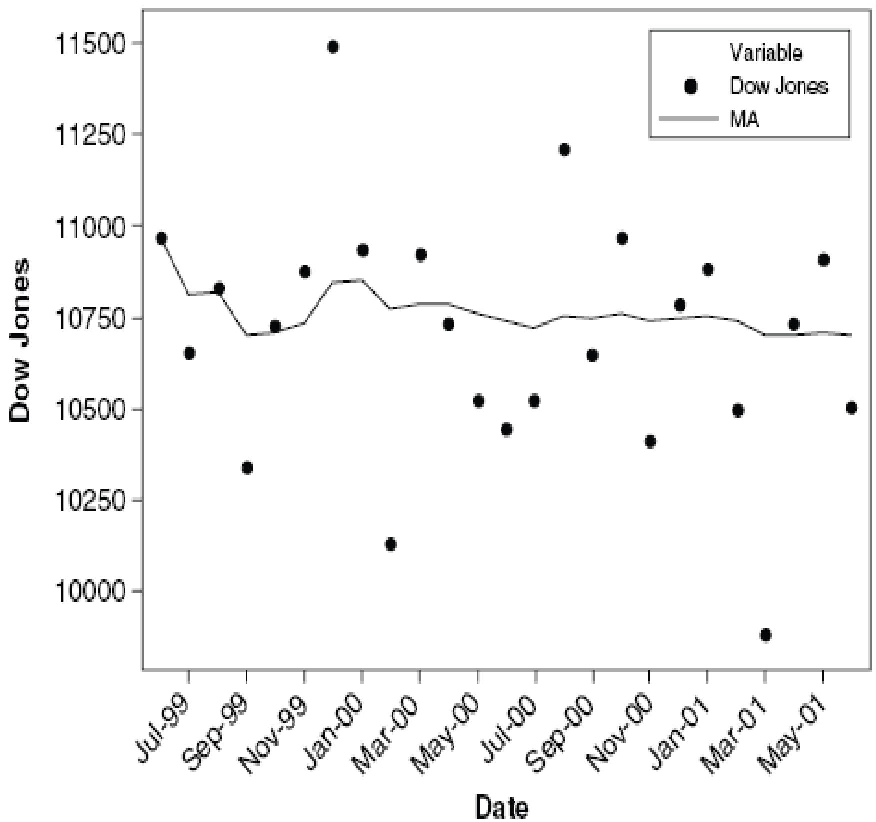


FIGURE 4.2 The Dow Jones Index from June 1999 to June 2001.

A constant model can be used to describe the general pattern of the data.

For the constant process, the smoother in (4.2) does a good job

Dow Jones Index (1999-2006)

How realistic is the assumption of “constant” process though?

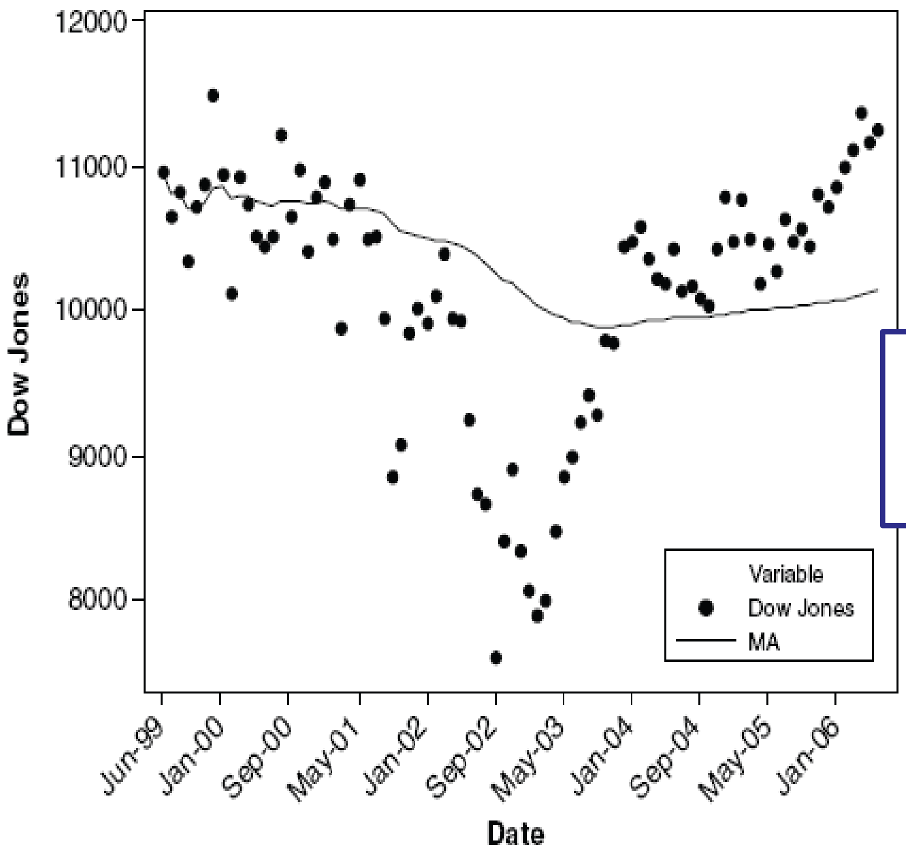


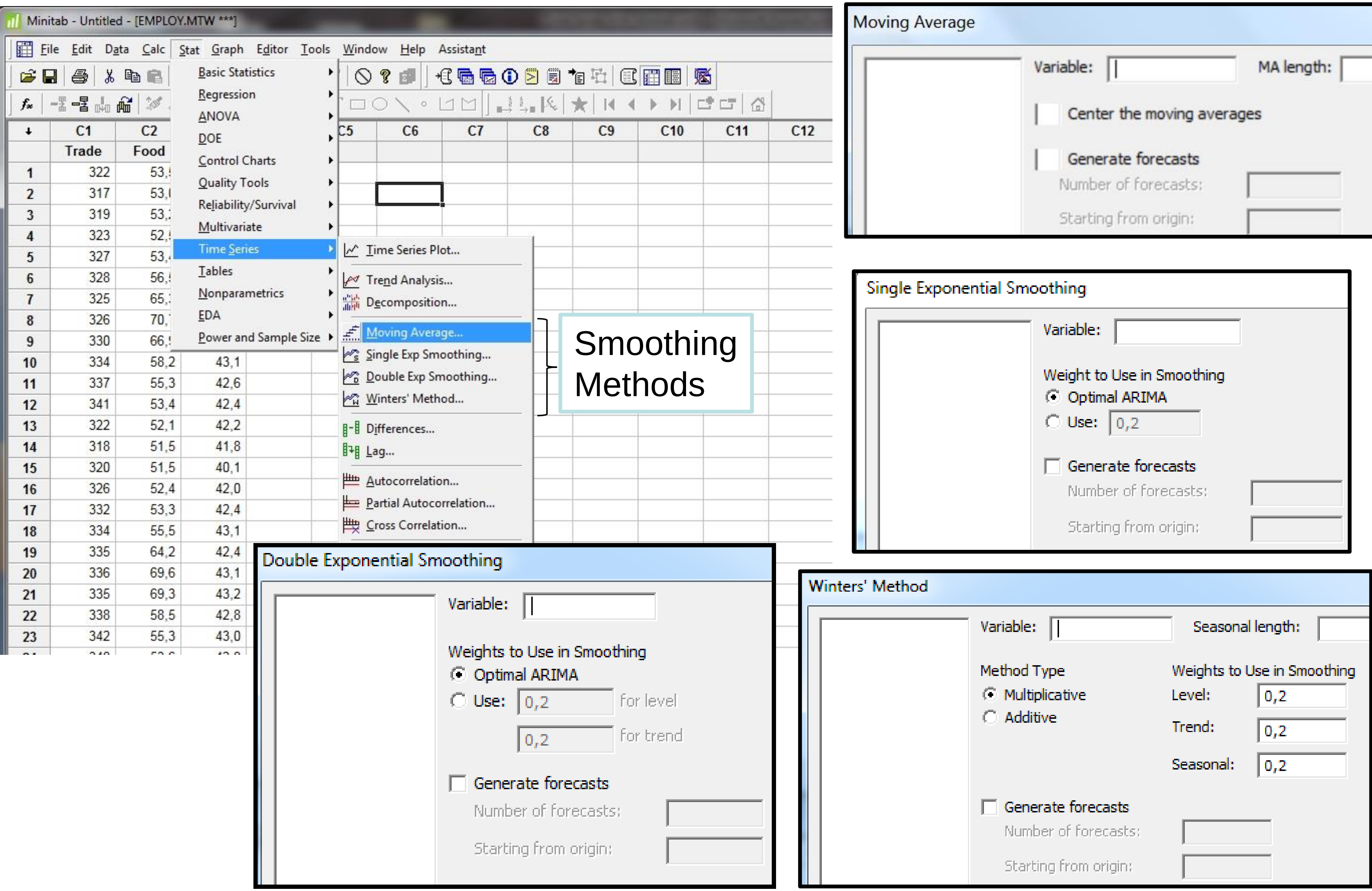
FIGURE 4.3 The Dow Jones Index from June 1999 to June 2006.

The constant process assumption is no longer valid

In real life …

- A constant process is **not** the norm but the exception
- Second Law of Thermodynamics says so: left to its own, any process will deteriorate
- Trying to smooth a non-constant process with the average of the data points up to the current time does not look too promising

Minitab Smoothing Methods

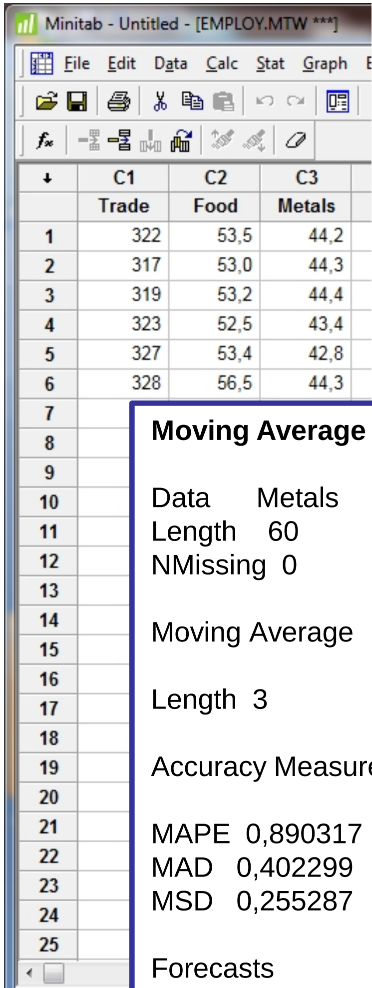


Simple Moving Average

- Average of the data points in a moving window of length *N*

$$M_T = \frac{y_T + y_{T-1} + \cdots + y_{T-N+1}}{N} = \frac{1}{N} \sum_{t=T-N+1}^{N} y_t$$

- Of course the question is: what should *N* be?



Moving Average for Metals

Data Metals
Length 60
NMissing 0

Moving Average

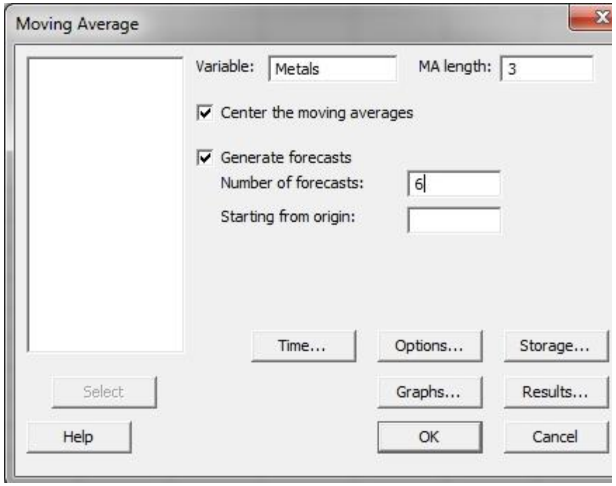
Length 3

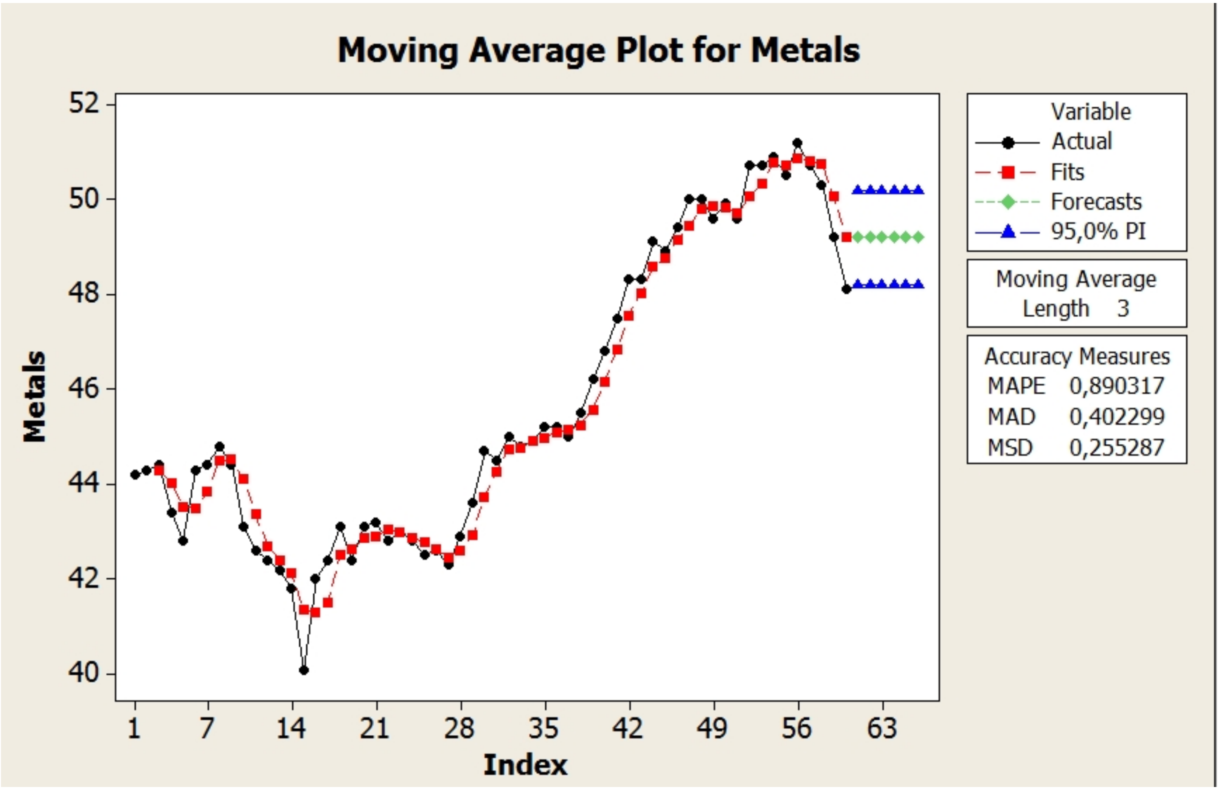
Accuracy Measures

MAPE 0,890317
MAD 0,402299
MSD 0,255287

Forecasts

Period	Forecast	Lower	Upper
61	49,2	48,2097	50,1903
62	49,2	48,2097	50,1903
63	49,2	48,2097	50,1903
64	49,2	48,2097	50,1903
65	49,2	48,2097	50,1903
66	49,2	48,2097	50,1903





To calculate a moving average, Minitab averages consecutive groups of observations in a series. For example, suppose a series begins with the numbers 4, 5, 8, 9, 10 and you use the moving average length of 3. The first two values of the moving average are missing. The third value of the moving average is the average of 4, 5, 8; the fourth value is the average of 5, 8, 9; the fifth value s the average of 8, 9, 10.

Centered moving average

By default, moving average values are placed at the period in which they are calculated. For example, for a moving average length of 3, the first numeric moving average value is placed at period 3, the next at period 4, and so on.

When you center the moving averages, they are placed at the center of the range rather than the end of it. This is done to position the moving average values at their central positions in time.

· **If the moving average length is odd:** Suppose the moving average length is 3. In that case, Minitab places the first numeric moving average value at period 2, the next at period 3, and so on. In this case, the moving average value for the first and last periods is missing (*).

· **If the moving average length is even:** Suppose the moving average length is 4. The center of that range is 2.5, but you cannot place a moving average value at period 2.5. This is how Minitab works around the problem. Calculate the average of the first four values, call it MA1. Calculate the average of the next four values, call it MA2. Average those two numbers (MA1 and MA2), and place that value at period 3. Repeat throughout the series. In this case, the moving average values for the first two and last two periods are missing (*).

The most crucial issue in simple moving averages is the choice of the **span**, N. A simple moving average will react faster to the changes if N is small. However, we know from Section 2.2.2 that the variance of the simple moving average with uncorrelated observations with variance σ^2 is given as

$$\mathrm{Var}(M_T) = \frac{\sigma^2}{N}$$

This means that as N gets small, the variance of the moving average gets bigger. This represents a dilemma in the choice of N. If the process is expected to be constant, a large N can be used whereas a small N is preferred if the process is changing. In Figure 4.4, we show the effect of going from a span of 10 observations to 5 observations. While the latter exhibits a more jittery behavior, it nevertheless follows the actual data more closely. A more thorough analysis on the choice of N can be performed based on the prediction error. We will explore this for exponential smoothers in Section 4.6.1, where we will discuss forecasting using exponential smoothing.

Moving average methods tem sempre a desvantagem de possuírem Autocorrelação!

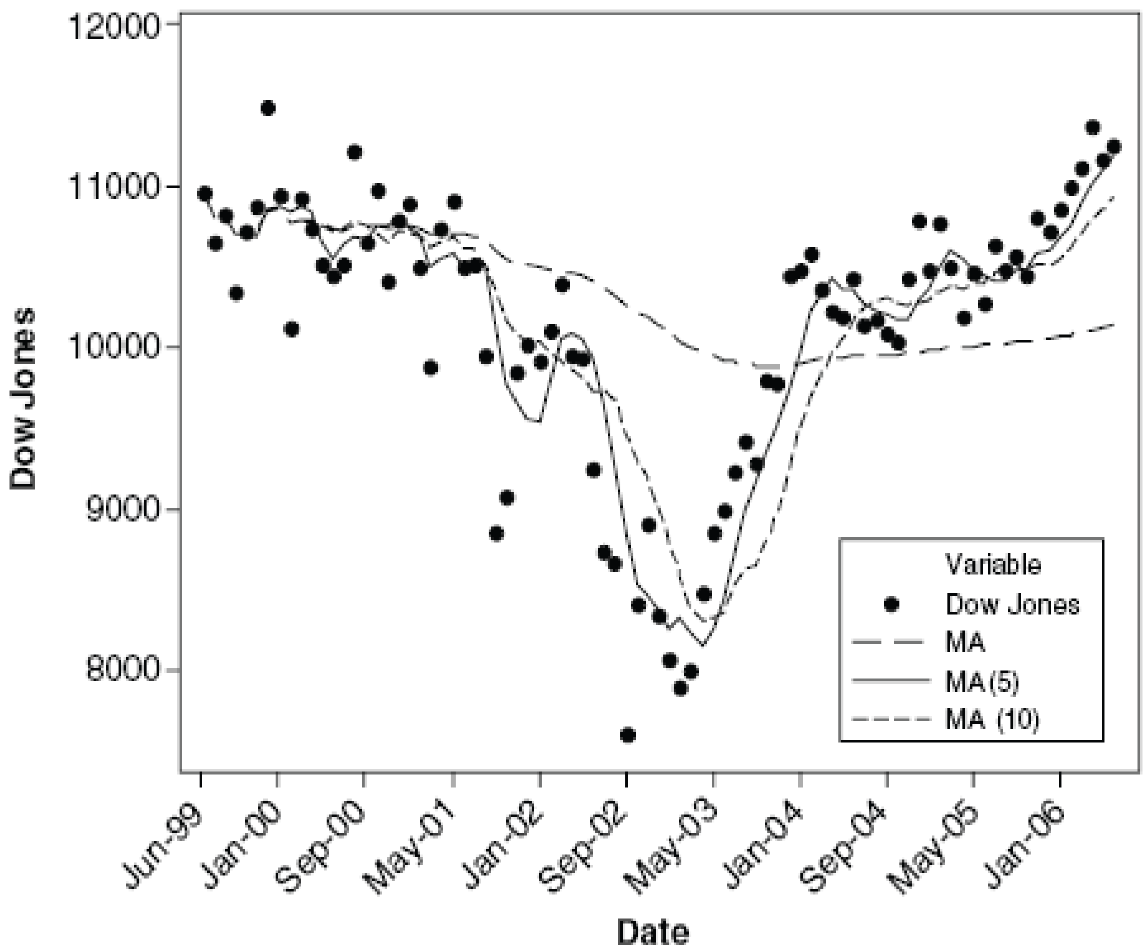


FIGURE 4.4 The Dow Jones Index from June 1999 to June 2006 with moving averages of span 5 and 10.

As *N* gets smaller, MA(*N*) reacts faster to the changes in the data!

If the process is expected to be constant, a large N can be used whereas a small N is preferred if the process is changing. (quanto maior o número de N maior a suavização)

First-Order Exponential Smoothing

- For a faster reacting smoother, have a weighted average with exponentially decreasing weights

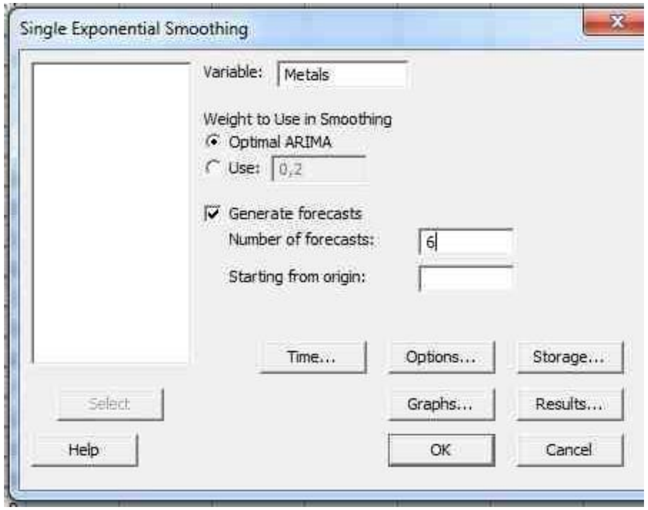
- The equation for the model is

$$S_t = \alpha Y_{t-1} + (1-\alpha) S_{t-1}$$

where α: the smoothing parameter, $0 \le \alpha \le 1$

Y_t: the value of the observation at time t

S_t: the value of the smoothed obs. at time t.



Use for:

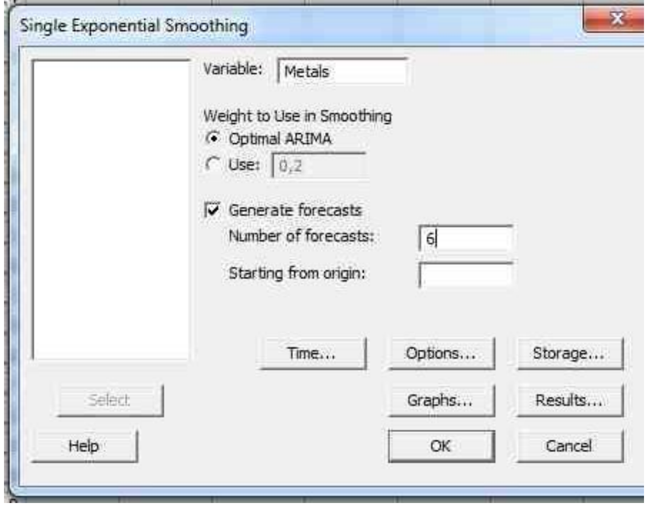
- Data with no trend, and
- Data with no seasonal pattern
- Short term forecasting

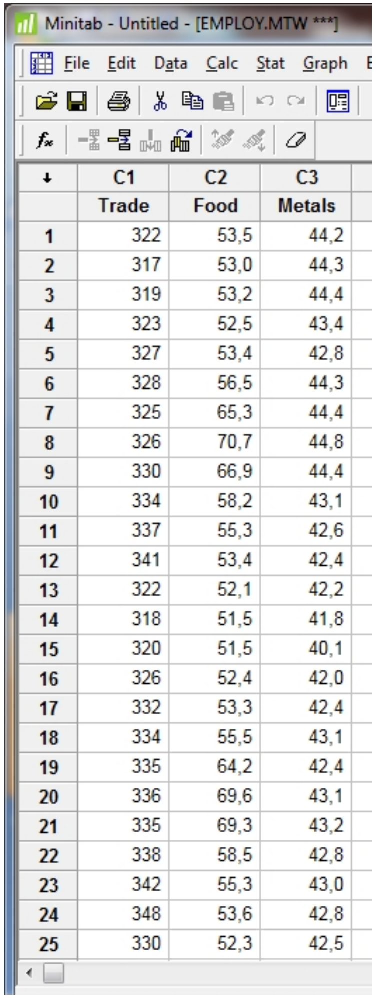
Forecast profile:

- Flat line

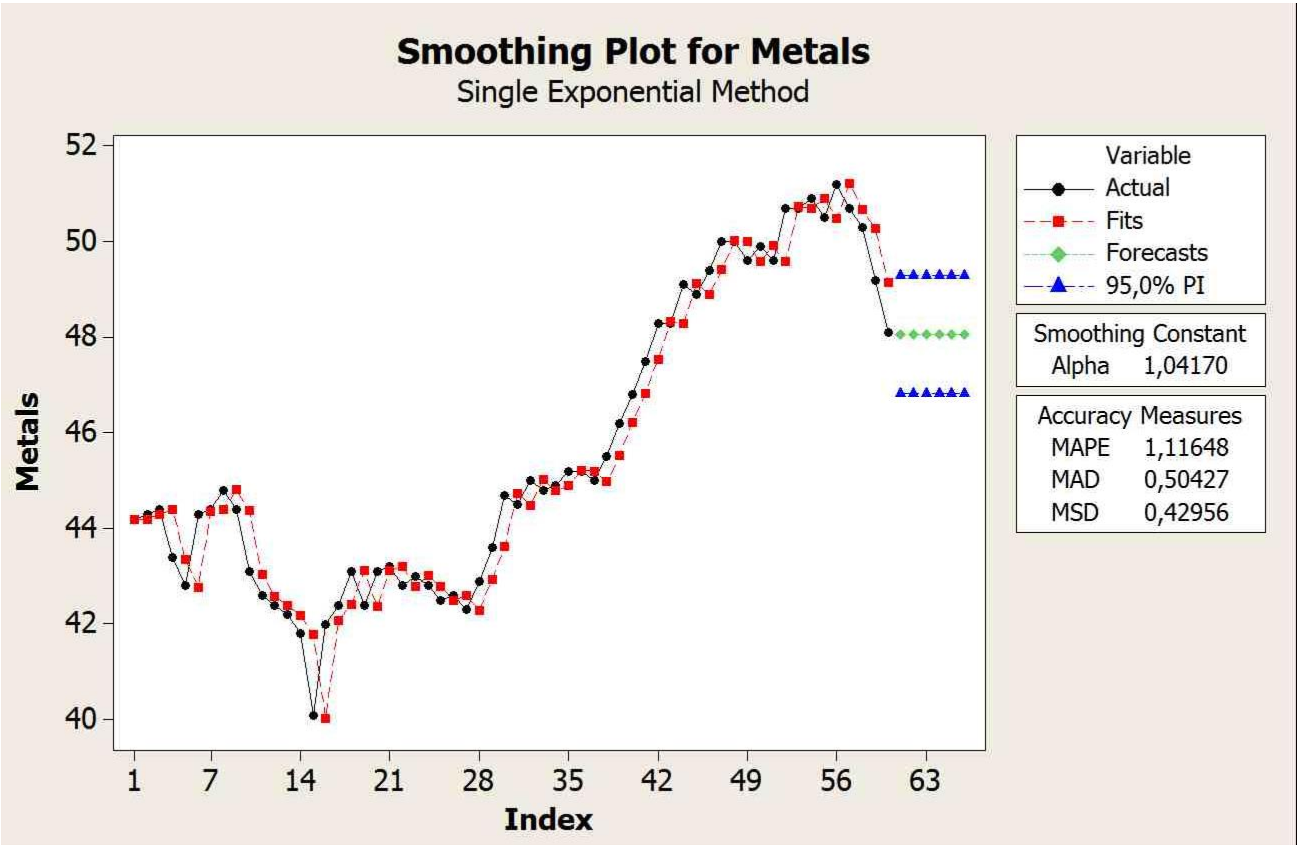
ARIMA equivalent: (0,1,1) model

Single Exponential Smoothing





↓	C1 Trade	C2 Food	C3 Metals
1	322	53,5	44,2
2	317	53,0	44,3
3	319	53,2	44,4
4	323	52,5	43,4
5	327	53,4	42,8
6	328	56,5	44,3
7	325	65,3	44,4
8	326	70,7	44,8
9	330	66,9	44,4
10	334	58,2	43,1
11	337	55,3	42,6
12	341	53,4	42,4
13	322	52,1	42,2
14	318	51,5	41,8
15	320	51,5	40,1
16	326	52,4	42,0
17	332	53,3	42,4
18	334	55,5	43,1
19	335	64,2	42,4
20	336	69,6	43,1
21	335	69,3	43,2
22	338	58,5	42,8
23	342	55,3	43,0
24	348	53,6	42,8
25	330	52,3	42,5



Optimal ARIMA weight

1 Minitab fits with an **ARIMA (0,1,1)** model and stores the fits.

2 The smoothed values are the ARIMA model fits, but lagged one time unit.

3 Initial smoothed value (at time one) by backcasting:

initial smoothed value = [smoothed in period two - a (data in period 1)] / (1 - a)

where 1- a estimates the MA parameter.

Specified weight

1 Minitab uses the average of the first six (or N, if N < 6) observations for the initial smoothed value (at time zero). Equivalently, Minitab uses the average of the first six (or N, if N < 6) observations for the initial fitted value (at time one).

Fit(i) = Smoothed(i-1).

2 Subsequent smoothed values are calculated from the formula:

smoothed value at time t = a (data at t) + (1 - a) (smoothed value at time t - 1)

where a is the weight.

A value between 0.1 and 0.4 is commonly recommended

Prediction limits

Based on the mean absolute deviation (MAD). The formulas for the upper and lower limits are:

Upper limit = Forecast + 1.96 * 1.25 * MAD

Lower limit = Forecast - 1.96 * 1.25 * MAD

The value of 1.25 is an approximate proportionality constant of the standard deviation to the mean absolute deviation. Hence, 1.25 * MAD is approximately the standard deviation.

Single Exponential Smoothing for Met

Data Metals
Length 60

Smoothing Constant

Alpha 1,04170

Accuracy Measures

MAPE	1,11648
MAD	0,50427
MSD	0,42956

Forecasts

Period	Forecast	Lower	Upper
61	48,0560	46,8206	49,2914
62	48,0560	46,8206	49,2914
63	48,0560	46,8206	49,2914
64	48,0560	46,8206	49,2914
65	48,0560	46,8206	49,2914
66	48,0560	46,8206	49,2914

The three accuracy measures, MAPE, MAD, and MSD, were 1.12, 0.50, and 0.43, respectively for the single exponential smoothing model, compared to 1.55, 0.70, and 0.76, respectively, for the moving average fit (see Example of moving average). Because these values are smaller for single exponential smoothing, you can judge that this method provides a better fit to these data.

Second-Order Exponential Smoothing

- If the data shows a linear trend, a more appropriate model will be

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Use for:

- Data with constant or non-constant **trend**, and
- Data with no seasonal pattern
- Short term forecasting

Forecast profile:

- Straight line with slope equal to last trend estimate

ARIMA equivalent: (0,2,2) model

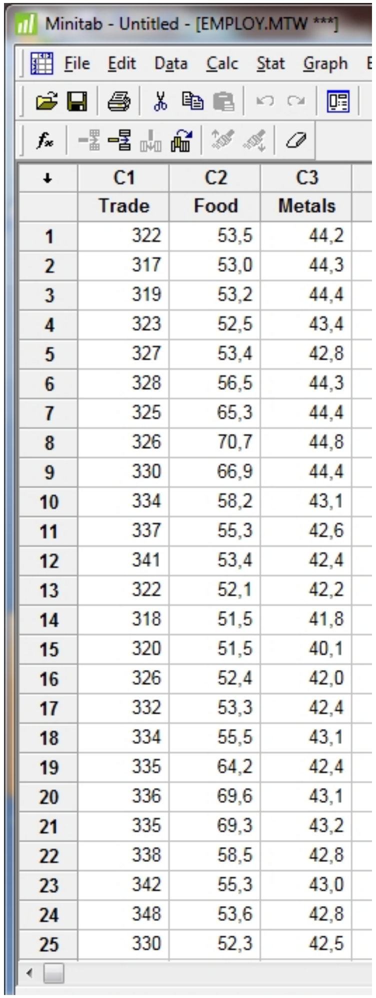
Second-Order Exponential Smoothing

Double exponential smoothing (also called **Holt's method**) smoothes the data when a trend is present. The double exponential smoothing equations are:

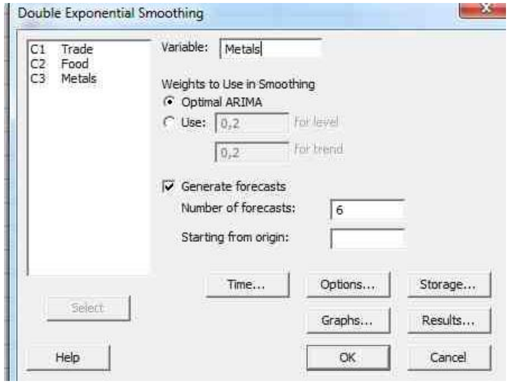
$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$
$$\hat{Y}_t = L_{t-1} + T_{t-1},$$

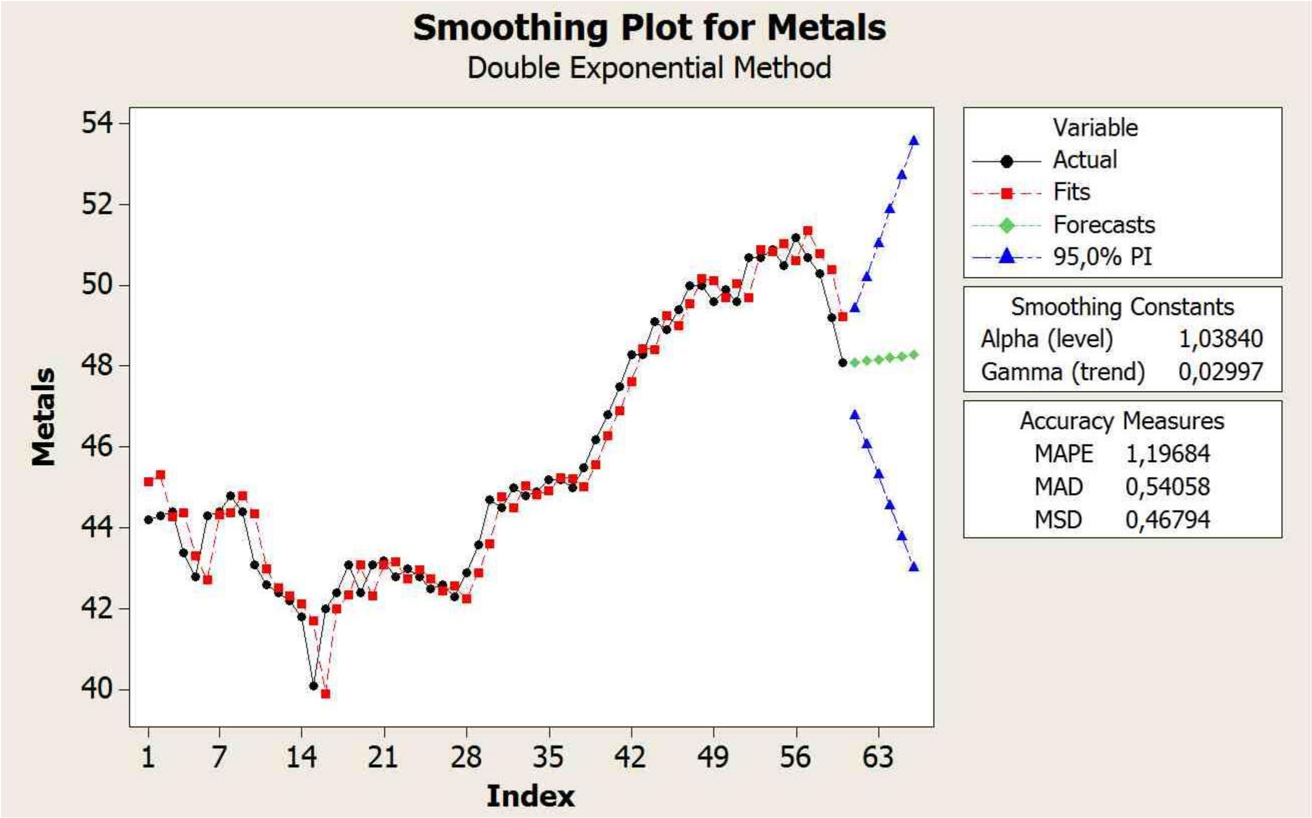
where L_t is the level at time *t*, α is the weight (or smoothing constant) for the level, T_t is the trend at time *t*, β is the weight (or smoothing constant) for the trend, and all other quantities are defined as earlier. An equivalent ARIMA(0,2,2) model can be constructed to represent the double exponential smoother.

Double Exponential Smoothing



	C1 Trade	C2 Food	C3 Metals
1	322	53,5	44,2
2	317	53,0	44,3
3	319	53,2	44,4
4	323	52,5	43,4
5	327	53,4	42,8
6	328	56,5	44,3
7	325	65,3	44,4
8	326	70,7	44,8
9	330	66,9	44,4
10	334	58,2	43,1
11	337	55,3	42,6
12	341	53,4	42,4
13	322	52,1	42,2
14	318	51,5	41,8
15	320	51,5	40,1
16	326	52,4	42,0
17	332	53,3	42,4
18	334	55,5	43,1
19	335	64,2	42,4
20	336	69,6	43,1
21	335	69,3	43,2
22	338	58,5	42,8
23	342	55,3	43,0
24	348	53,6	42,8
25	330	52,3	42,5





Weights

Optimal ARIMA weights

1 Minitab fits with an ARIMA (0,2,2) model to the data, in order to minimize the sum of squared errors.

2 The trend and level components are then initialized by backcasting.

Specified weights

1 Minitab fits a linear regression model to time series data (y variable) versus time (x variable).

2 The constant from this regression is the initial estimate of the level component, the slope coefficient is the initial estimate of the trend component. When you specify weights that correspond to an equal-root ARIMA (0, 2, 2) model, Holt's method specializes to Brown's method.[1]

[1] N.R. Farnum and L.W. Stanton (1989). *Quantitative Forecasting Methods*. PWS-Kent.

Double Exponential Smoothing for Metals

```
Data    Metals
Length  60


Smoothing Constants

Alpha (level)  1,03840
Gamma (trend)  0,02997


Accuracy Measures

MAPE  1,19684
MAD   0,54058
MSD   0,46794


Forecasts

Period  Forecast    Lower    Upper
61       48,0961  46,7718  49,4205
62       48,1357  46,0600  50,2113
63       48,1752  45,3135  51,0368
64       48,2147  44,5546  51,8747
65       48,2542  43,7899  52,7184
66       48,2937  43,0221  53,5652
```

Because the difference in accuracy measures for the two exponential smoothing methods are small, you might consider the type of forecast (horizontal line versus line with slope) in selecting between methods. Double exponential smoothing forecasts an employment pattern that is slightly increasing though the last four observations are decreasing. A higher weight on the trend component can result in a prediction in the same direction as the data, which may be more realistic, but the measured fit may not be as good as when Minitab generated weights are used.

Modeling Seasonal Data

- Additive Seasonal Model
  - If the amplitude of the seasonal pattern is independent of the average level within the season

- Multiplicative Seasonal Model
  - If the amplitude of the seasonal pattern is proportional to the average level within the season

Winters' Method, Additive Model

Use for:

- Data with or without trend
- Data with seasonal pattern
- Size of seasonal pattern not proportional to data
- Short to medium range forecasting

Forecast profile:

- Straight line with seasonal pattern added on

ARIMA equivalent: none

Winters' Method, Multiplicative Model

Use for:

- Data with or without trend
- Data with seasonal pattern
- Size of seasonal pattern proportional to data
- Short to medium range forecasting

Forecast profile:

- Straight line multiplied by seasonal pattern

ARIMA equivalent: none

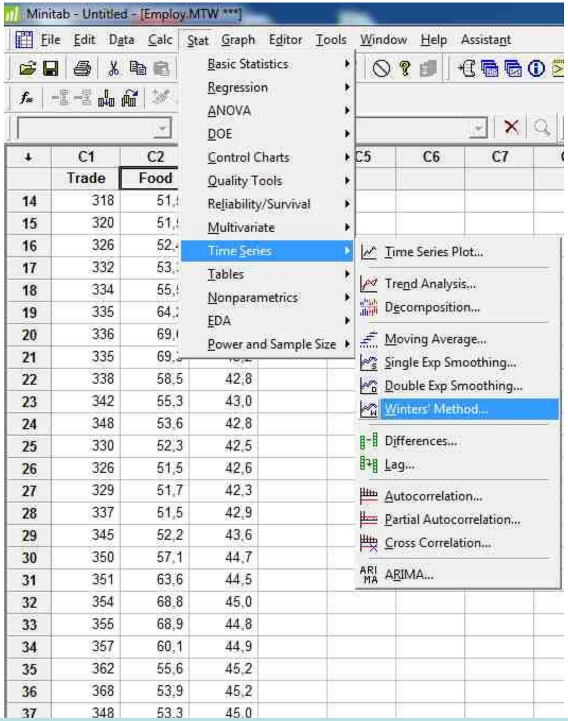
Finally, **Holt-Winters exponential smoothing** smoothes the data when trend and seasonality are present; however, these two components can be either additive or multiplicative. For the additive model, the equations are:

$$
\begin{aligned}
L_t &= \alpha(Y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \\
T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \\
S_t &= \delta(Y_t - L_t) + (1-\delta)S_{t-p} \\
\hat{Y}_t &= L_{t-1} + T_{t-1} + S_{t-p}.
\end{aligned}
$$

For the multiplicative model, the equations are:

$$
\begin{aligned}
L_t &= \alpha(Y_t / S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \\
T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \\
S_t &= \delta(Y_t / L_t) + (1-\delta)S_{t-p} \\
\hat{Y}_t &= (L_{t-1} + T_{t-1})S_{t-p}.
\end{aligned}
$$

For both sets of equations, all quantities are the same as they were defined in the previous models, except now we also have that S_t is the seasonal component at time t, δ is the weight (or smoothing constant) for the seasonal component, and p is the seasonal period.



Multiplicative model for decomposition

Use when the size of the seasonal pattern depends on the level of the data. This model assumes that as the data increase, so does the seasonal pattern. Most time series plots exhibit such a pattern. In this model, the trend and seasonal components are multiplied and then added to the error component.

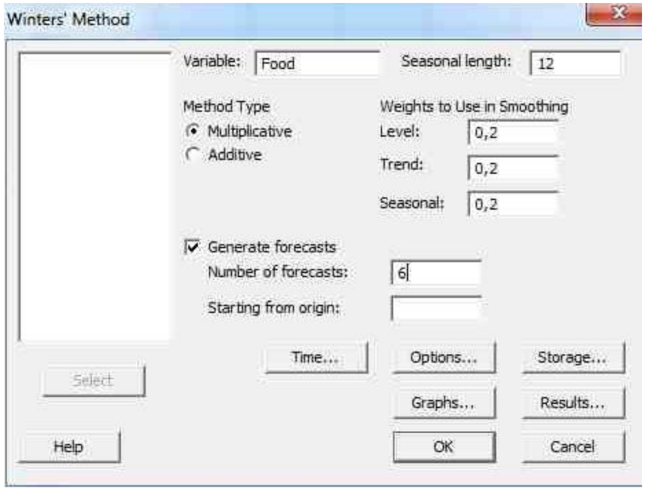
Additive model

A data model in which the effects of individual factors are differentiated and added together to model the data. They appear in several Minitab commands:

· An additive model is optional for Decomposition procedures and for Winter's method. Use an additive model when the magnitude of the data does not affect its seasonal pattern.

· An additive model is optional for two-way ANOVA procedures. Choose this option to omit the interaction term from the model.

· Median polish creates an additive model to describe data in a two-way table. This additive model describes raw data as the sum of row effects, column effects, residuals, and a common effect.

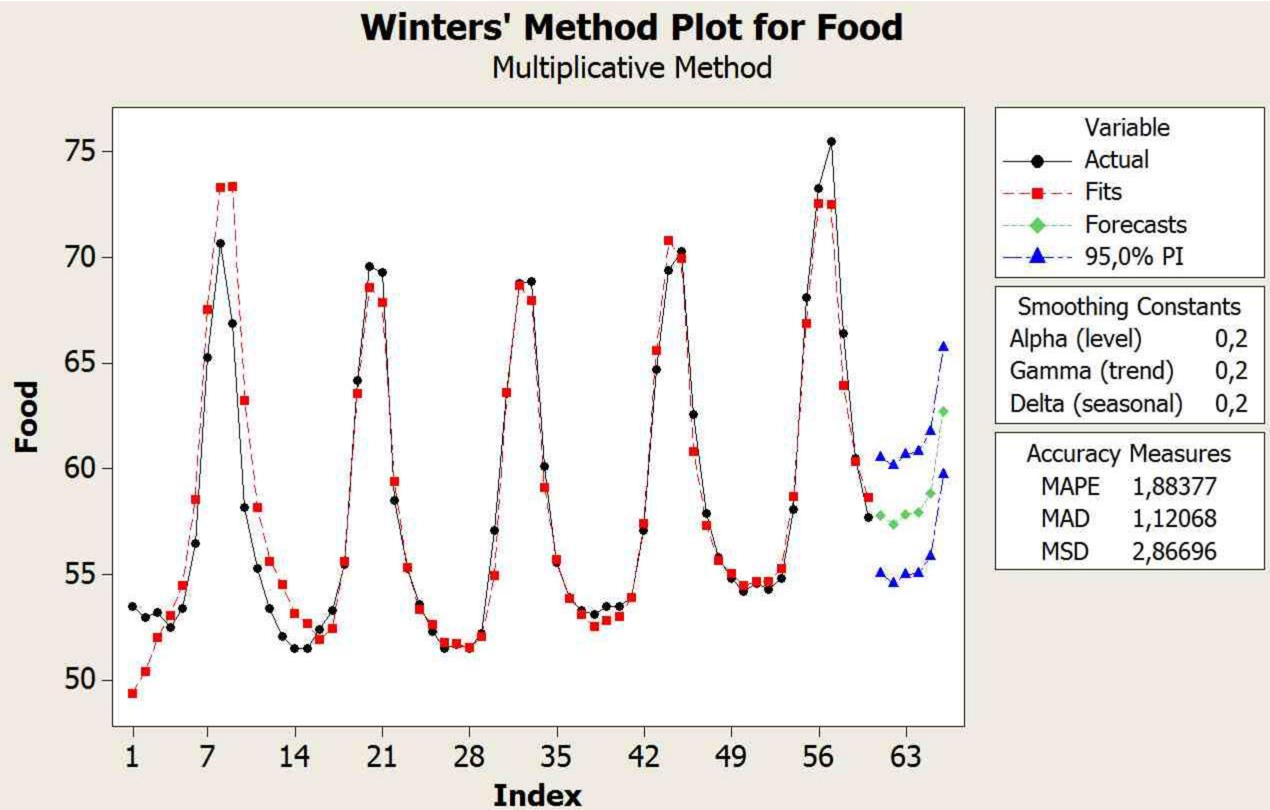


Model fitting

Winters' method employs a level component, a trend component, and a seasonal component at each period. It uses three weights, or smoothing parameters, to update the components at each period. Initial values for the level and trend components are obtained from a linear regression on time. Initial values for the seasonal component are obtained from a dummy-variable regression using detrended data.

You can enter weights, or smoothing parameters, for the level, trend, and seasonal components. The default weights are 0.2 and you can enter values between 0 and 1. Since an equivalent ARIMA model exists only for a very restricted form of the Holt-Winters model, Minitab does not compute optimal parameters for Winters' method as it does for single and double exponential smoothing.

Regardless of the component, large weights result in more rapid changes in that component; small weights result in less rapid changes. The components in turn affect the smoothed values and the predicted values. Thus, small weights are usually recommended for a series with a high noise level around the signal or pattern. Large weights are usually recommended for a series with a small noise level around the signal.



Winters' Method for Food

Multiplicative Method

Data Food
Length 60

Smoothing Constants

Alpha (level) 0,2
Gamma (trend) 0,2
Delta (seasonal) 0,2

Accuracy Measures

MAPE 1,88377
MAD 1,12068
MSD 2,86696

Forecasts

Period	Forecast	Lower	Upper
61	57,8102	55,0646	60,5558
62	57,3892	54,6006	60,1778
63	57,8332	54,9966	60,6698
64	57,9307	55,0414	60,8199
65	58,8311	55,8847	61,7775
66	62,7415	59,7339	65,7492

For these data, MAPE, MAD, and MSD were 1.88, 1.12, and 2.87, respectively, with the multiplicative model. MAPE, MAD, and MSD were 1.95, 1.15, and 2.67, respectively (output not shown) with the additive model , indicating that the multiplicative model provided a slightly better fit according to two of the three accuracy measures.

Additive Seasonal Model

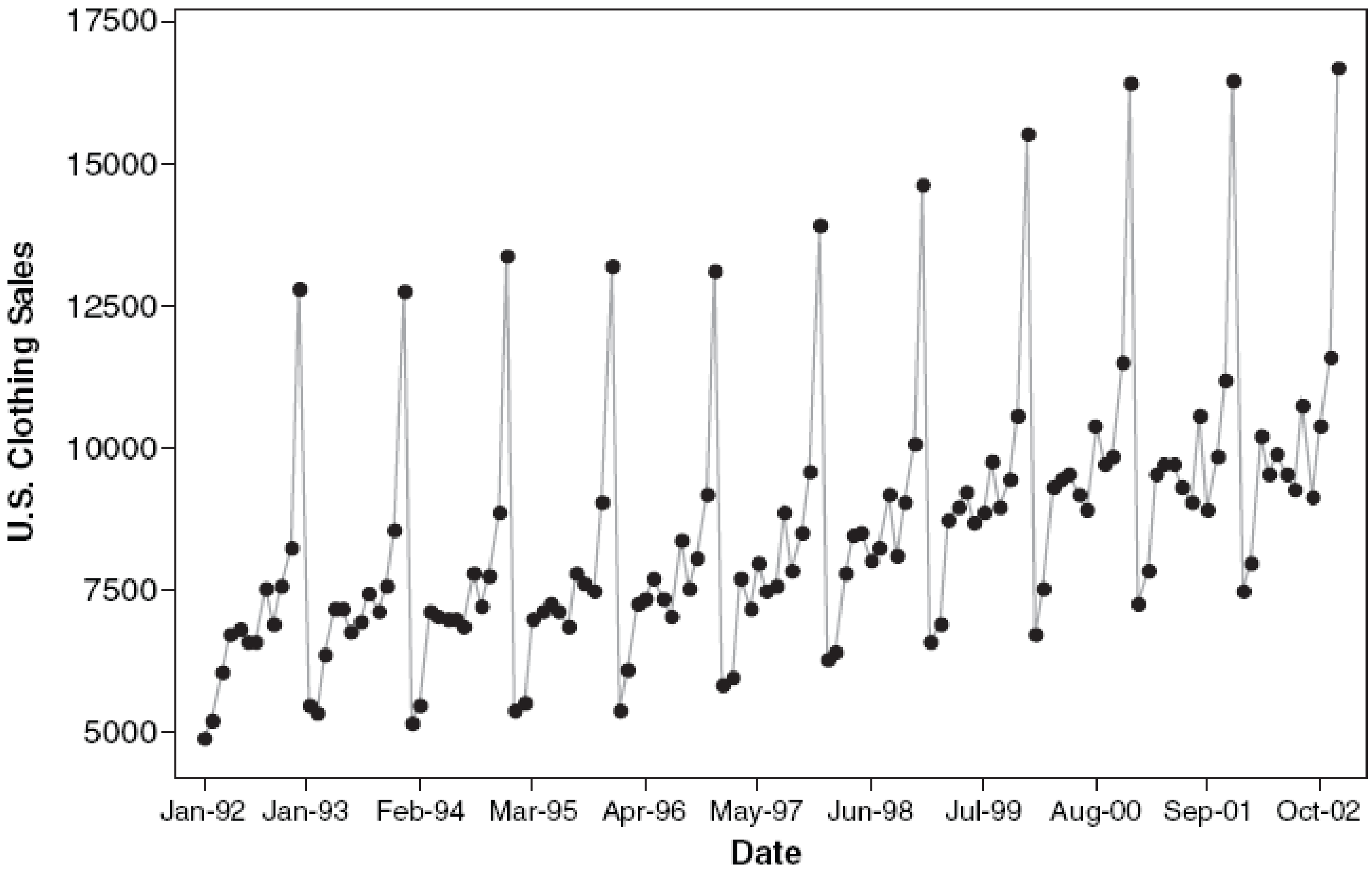


FIGURE 4.26 Time series plot of U.S. clothing sales from January 1992 to December 2003.

Additive Seasonal Model

We will for this case assume that the seasonal time series can be represented by the following model:

$$y_t = L_t + S_t + \varepsilon_t \tag{4.44}$$

where L_t represents the linear trend component and can in turn be represented by $\beta_0 + \beta_1 t$; S_t represents the seasonal adjustment with $S_t = S_{t+s} = S_{t+2s} = \cdots$ for $t = 1, \ldots, s-1$, where s is the length of the period of the cycles; and the ε_t are assumed to be uncorrelated with mean 0 and constant variance σ_ε^2. One usual restriction on this model is that the seasonal adjustments add to zero during one period,

$$\sum_{t=1}^{s} S_t = 0 \tag{4.45}$$

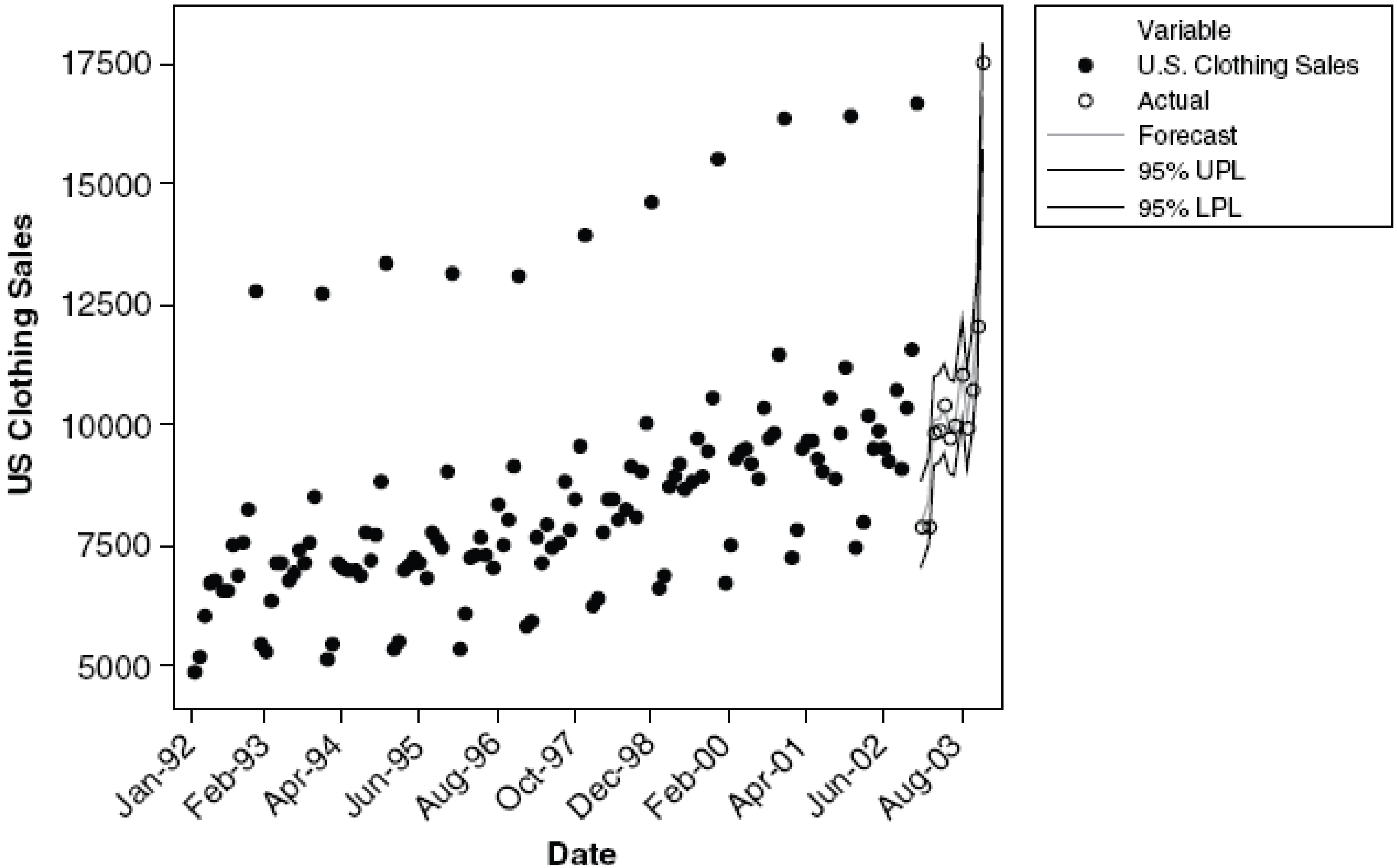


FIGURE 4.28 Forecasts for 2003 for the U.S. clothing sales.

Multiplicative Seasonal Model

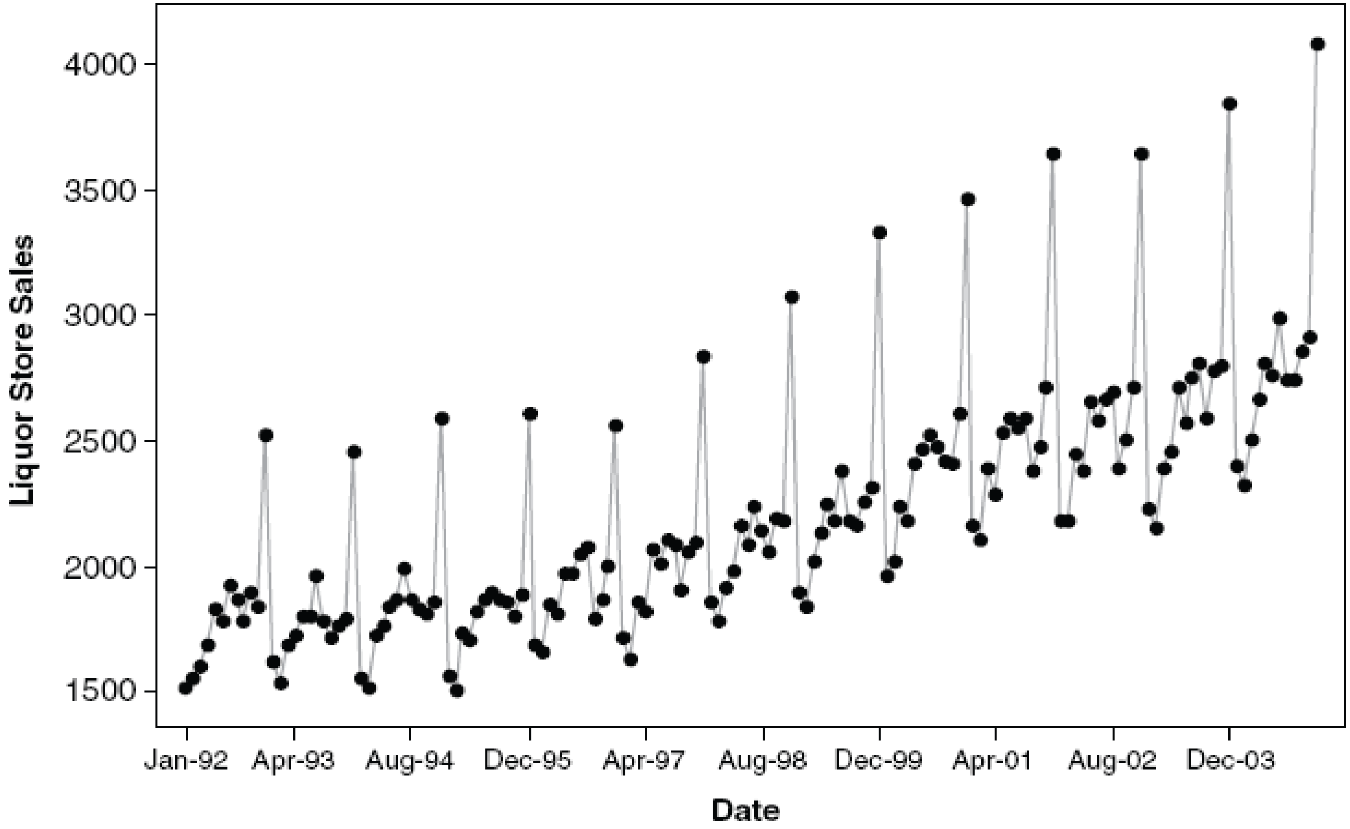


FIGURE 4.29 Time series plot of liquor store sales data from January 1992 to December 2004.

Note that the amplitude gets bigger as average level of the time series gets big.

Multiplicative Seasonal Model

If the amplitude of the seasonal pattern is proportional to the average level of the seasonal time series, as in the liquor store sales data given in Figure 4.29, the following multiplicative seasonal model will be more appropriate:

$$y_t = L_t S_t + \varepsilon_t \qquad (4.52)$$

where L_t once again represents the trend component (i.e., $\beta_0 + \beta_1 t$); S_t represents the seasonal adjustment with $S_t = S_{t+s} = S_{t+2s} = \cdots$ for $t = 1, \ldots, s-1$, where s is the length of the period of the cycles; and the ε_t are assumed to be uncorrelated with mean 0 and constant variance σ_ε^2. The restriction for the seasonal adjustments in this case becomes

$$\sum_t^s S_t = s \qquad (4.53)$$

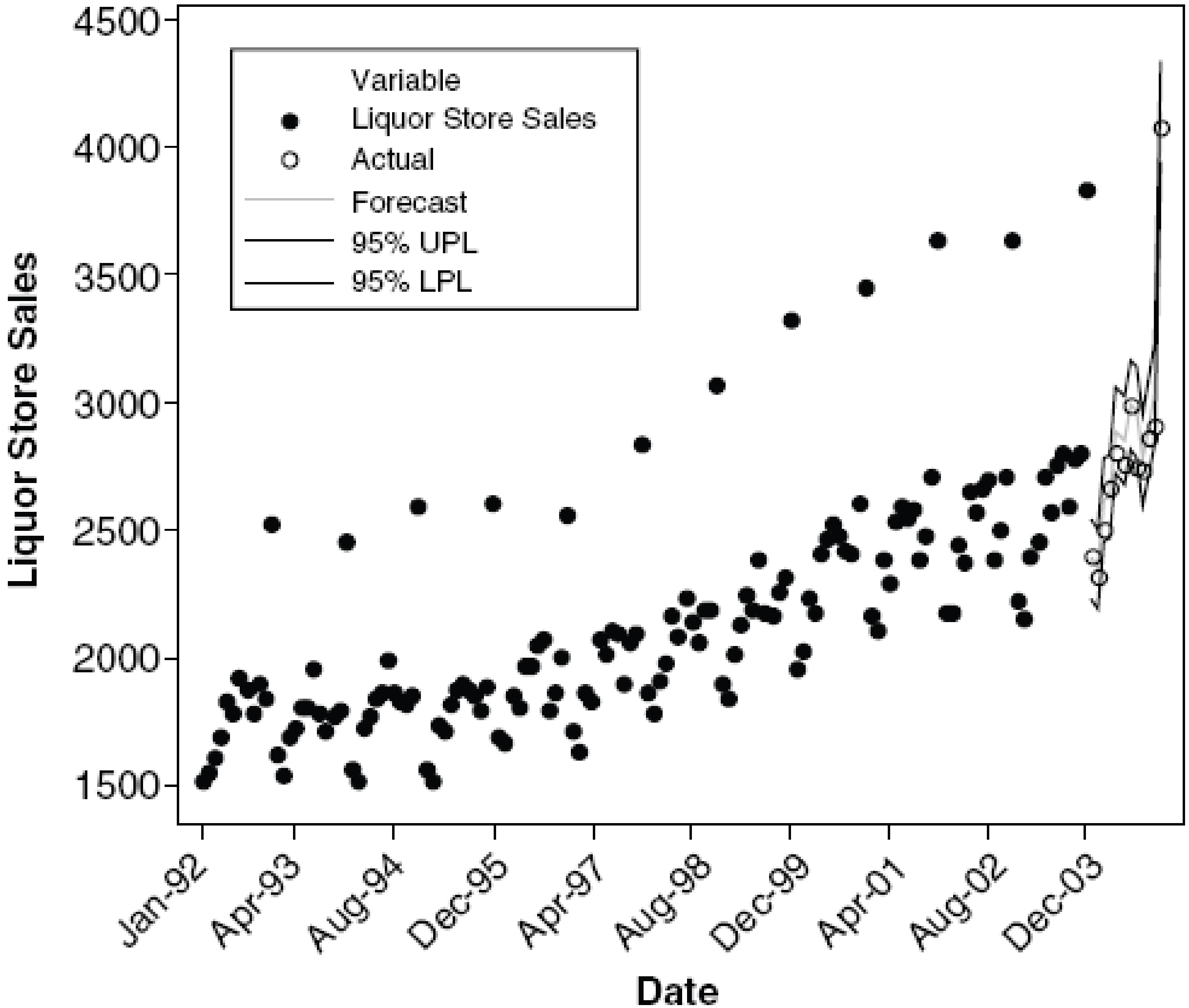


FIGURE 4.32 Forecasts for the liquor store sales for 2004 using the multiplicative model.