Exponential Smoothing Methods

1 Introduction

Signal and Noise in Time Series

- **Signal**: The underlying pattern in the data, caused by the process itself.
- **Noise**: Random fluctuations, unpredictable errors around the signal.

For example, a constant process can be expressed as:

$$y_t = \mu + \varepsilon_t$$

where:

- μ= constant mean level
- ε_t = noise at time t, assumed to be independent, mean zero, constant variance

Smoothing as Filtering

Smoothing separates signal from noise by replacing current values with functions of past values. This reveals the underlying pattern.

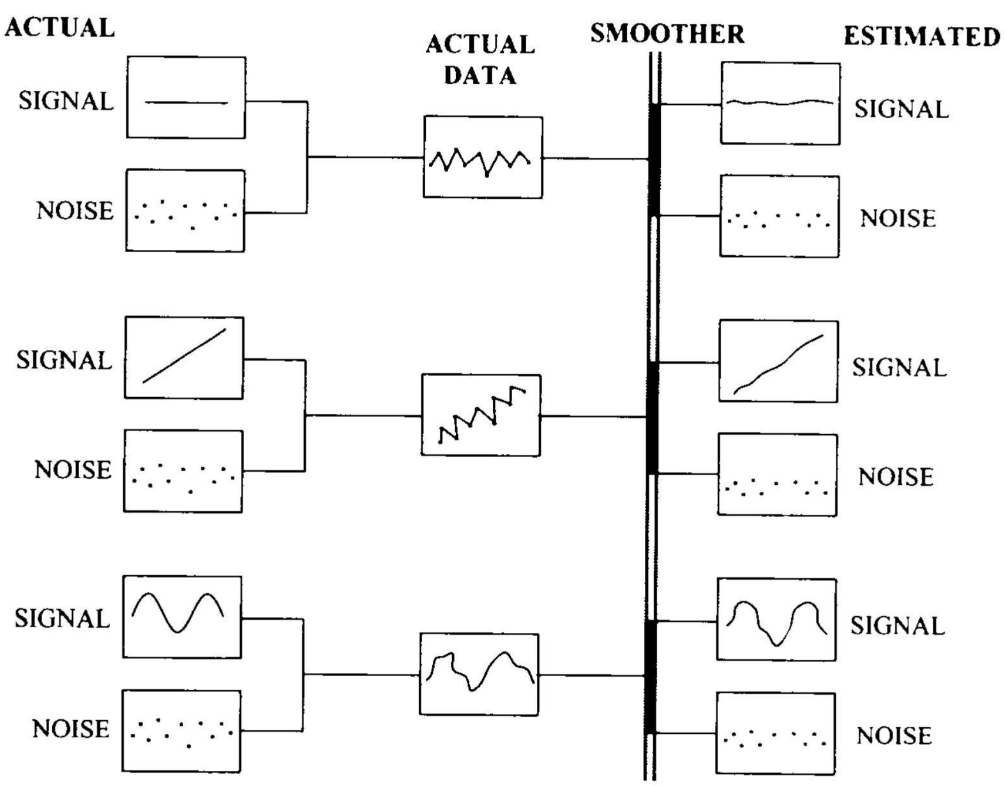


FIGURE 4.1 The process of smoothing a data set.

- A simple smoother: replace YTY_T with the average of all previous values. This is equivalent to the least squares estimate of μ\mu.
- Problem: If the process changes (e.g., trends or shifts), a simple average is too slow to react.

Key Idea: Instead of treating all past values equally, give more weight to recent data and less weight to older data.

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

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

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

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

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.

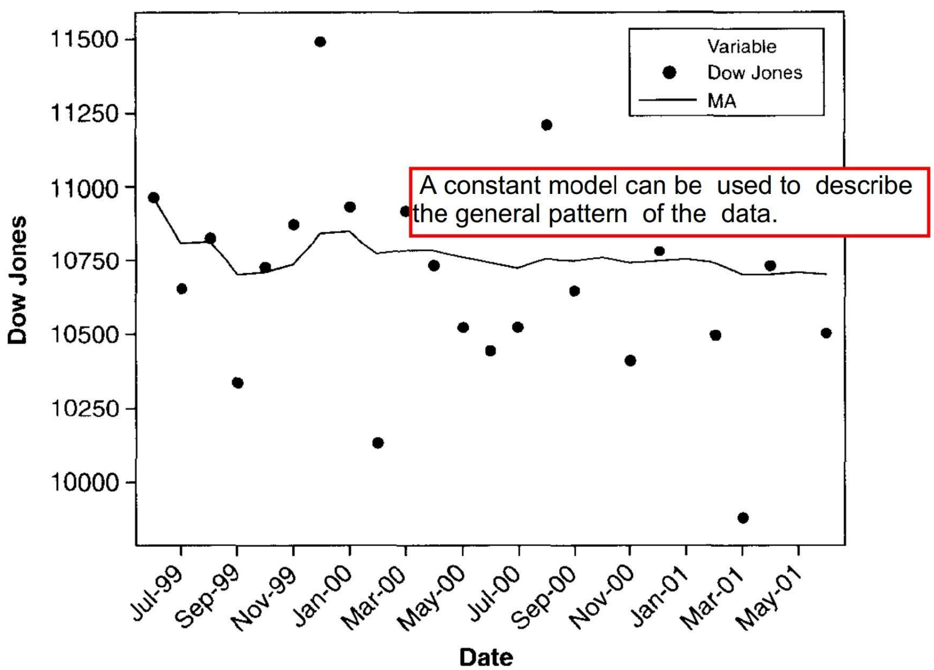


FIGURE 4.2 The Dow Jones Index from June 1999 to June 2001.

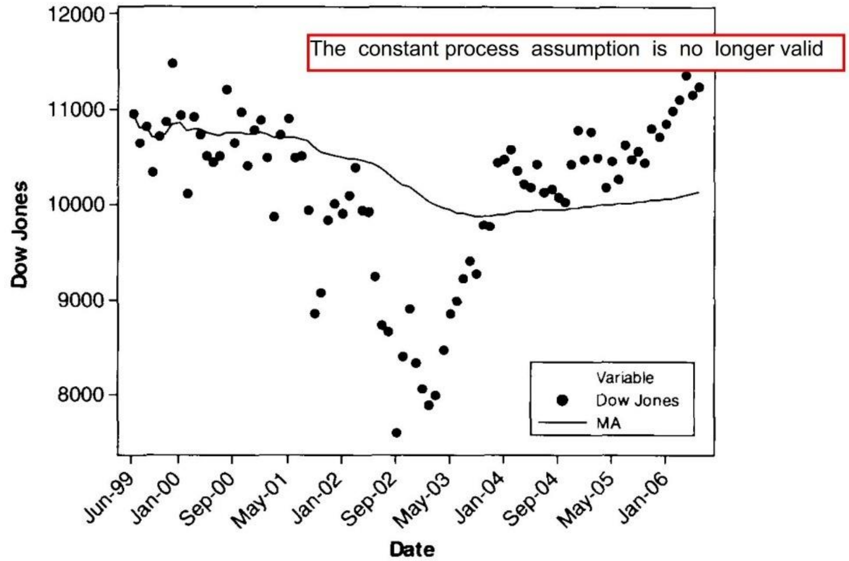


FIGURE 4.3 The Dow Jones Index from June 1999 to June 2006.

When there is a change in the process, earlier data no longer carry the information about the change in the process, yet they contribute to this inertia at an equal proportion compared to the more recent (and probably more useful) data. The most obvious choice is to somehow discount the older data. A common solution is to use the simple moving average. The most crucial issue in simple moving averages is the choice of the span, N. As N gets small, the variance of the moving average gets bigger. This represents a dilemma in the choice of N. A final note on the moving average is that even if the individual observations are independent, the moving averages will be autocorrelated as two successive moving averages contain the same N – 1 observations.

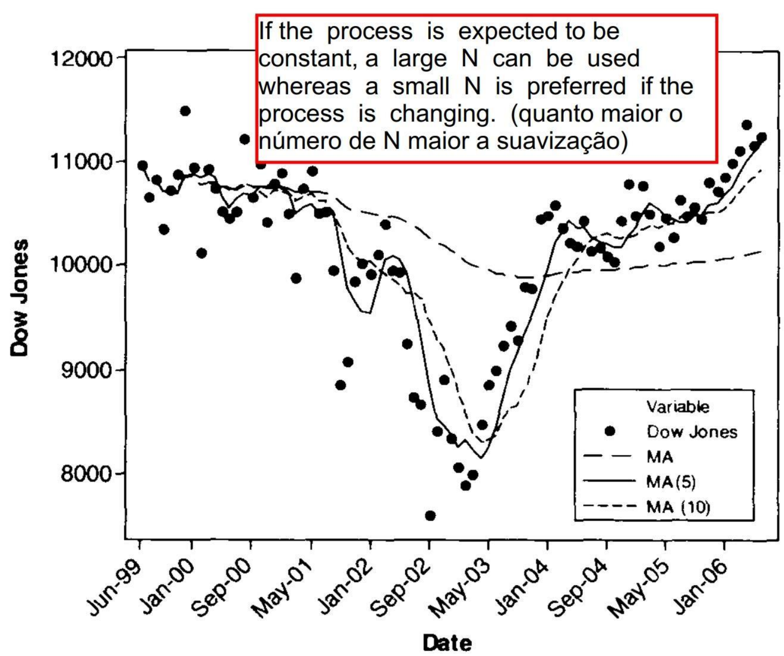


FIGURE 4.4 The Dow Jones Index from June 1999 to June 2006 with moving averages of span 5 and 10.

2 First-Order (Simple) Exponential Smoothing

Another approach to obtain a smoother that will react to process changes faster is to give geometrically decreasing weights to the previous observations. Hence an exponentially weighted smoother is obtained by introducing a discount factor θ as

$$\boxed{\sum_{t=0}^{T-1} \theta^t y_{T-t} = y_T + \theta\, y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1} \tag{4.3}$$

Please note that if the previous observations are to be discounted in a geometrically decreasing manner, then we should have $|\theta| < 1$. However, the smoother in Eq. (4.3) is not an *average* as the sum of the weights is

$$\boxed{\sum_{t=0}^{T-1} \theta^t = \frac{1-\theta^T}{1-\theta}} \tag{4.4}$$

Soma dos termos de uma PG finita diferente de 1

and hence does not necessarily add up to 1. For that we can adjust the smoother in Eq. (4.3) by multiplying it by $(1-\theta)/(1-\theta^T)$. However, for large T values, θ^T goes to zero and so the exponentially weighted average will have the following form:

$$\boxed{\begin{aligned} \tilde{y}_T &= (1-\theta) \sum_{t=0}^{T-1} \theta^t y_{T-t} \\ &= (1-\theta)\left(y_T + \theta\, y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1\right) \end{aligned}} \tag{4.5}$$

Agora a soma é igual a 1, caracterizando uma distribuição

An alternate expression in a recursive form for simple exponential smoothing is given by

$$\boxed{\begin{aligned} \tilde{y}_T &= (1-\theta)\, y_T + (1-\theta)\left(\theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1\right) \\ &= (1-\theta)\, y_T + \theta \underbrace{(1-\theta)\left(y_{T-1} + \theta^1 y_{T-2} + \cdots + \theta^{T-2} y_1\right)}_{\tilde{y}_{T-1}} \\ &= (1-\theta)\, y_T + \theta \tilde{y}_{T-1} \end{aligned}} \tag{4.6}$$

Definition

Exponential smoothing applies geometrically decreasing weights to past data:

$$S_T = \alpha Y_T + (1 - \alpha) S_{T-1}$$

where:

- S_T: smoothed value at time T
- Y_T: actual observed value at time T
- α (0 < α < 1): smoothing constant

This recursive form means each smoothed value is a weighted combination of the latest observation and the previous smoothed value.

Interpretation of α (alpha)

- Small α (close to 0): more smoothing, slower to adapt to changes
- Large α (close to 1): less smoothing, follows data closely

Extreme cases:

- α=0: smoothing produces a constant line (maximum smoothing)
- α=1: smoothed series equals the original data (no smoothing)

Initial Value (S0)

To start the recursion, S0 must be chosen. Options:

1. Set S0=Y1
2. Set S0 equal to the average of first few observations

In practice, the influence of S0 diminishes over time.

Example: Dow Jones Index (1999–2006)

Using α=0.2 gives a smoother curve that lags behind data. Increasing to α=0.4 makes it follow data more closely but still filters out some noise.

(alpha in file are the lambda in book pics)

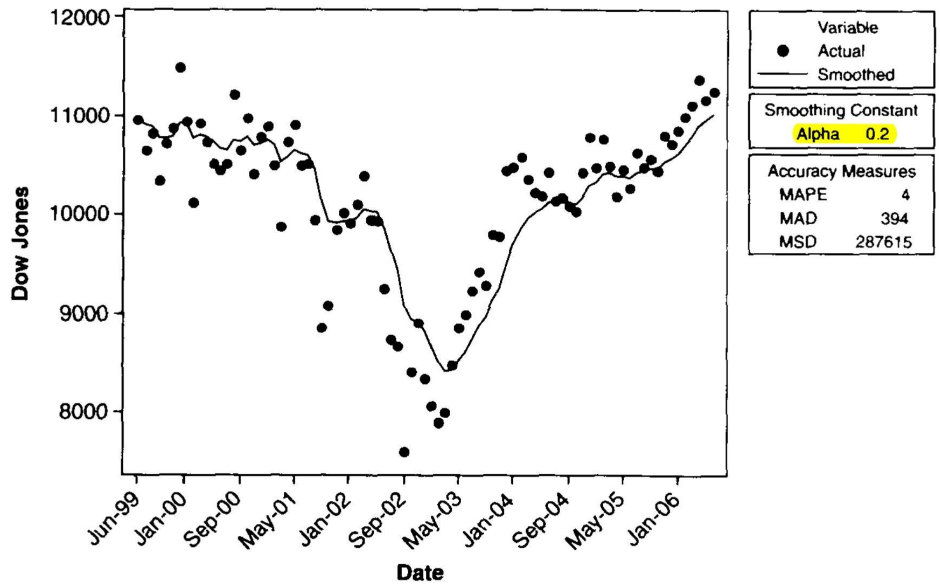
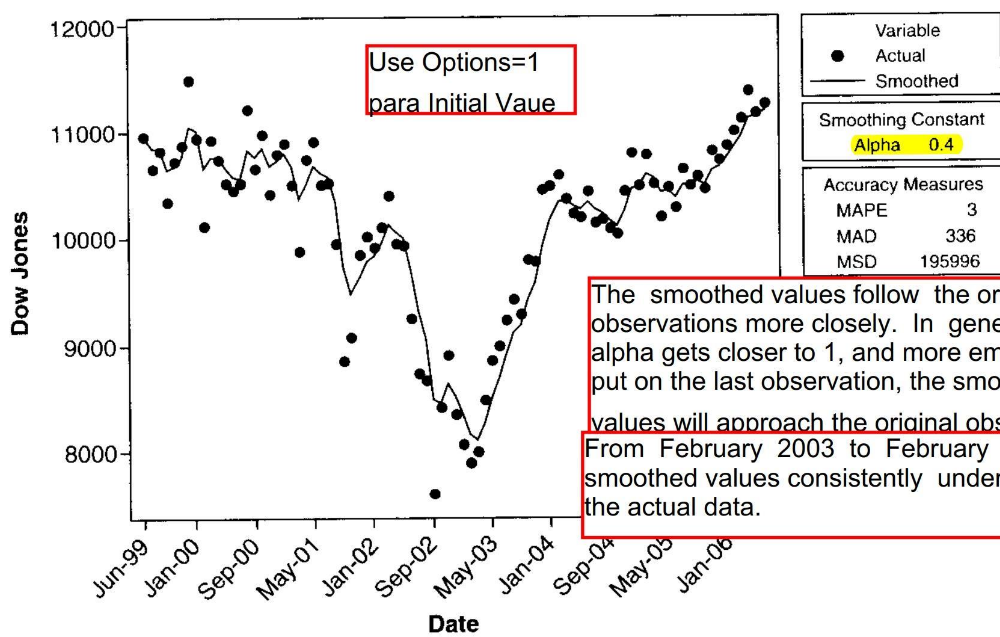


FIGURE 4.5 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.2$.



The smoothed values follow the original observations more closely. In general, as alpha gets closer to 1, and more emphasis is put on the last observation, the smoothed values will approach the original observations.

From February 2003 to February 2004) the smoothed values consistently underestimate the actual data.

FIGURE 4.6 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.4$.

Accuracy Measures

1. **MAPE** (Mean Absolute Percentage Error): $\mathrm{MAPE} = \frac{1}{T}\sum\left|\frac{Y_t - S_t}{Y_t}\right| \times 100$
2. **MAD** (Mean Absolute Deviation): $\mathrm{MAD} = \frac{1}{T}\sum|Y_t - S_t|$
3. **MSD** (Mean Squared Deviation): $\mathrm{MSD} = \frac{1}{T}\sum(Y_t - S_t)^2$

3 Exponential Smoothing as Model Estimation

Exponential smoothing can be viewed as a weighted least squares estimator:

$$SSE = \sum_{t=1}^{T} \beta^{T-t}(Y_t - \mu)^2$$

where past errors are weighted by a decreasing factor. This shows exponential smoothing estimates the level (μ) while giving more importance to recent data.

4 Second-Order Exponential Smoothing

When is First-Order Not Enough?

If the series has a trend, simple exponential smoothing produces biased estimates. It tends to underestimate upward trends and overestimate downward trends.

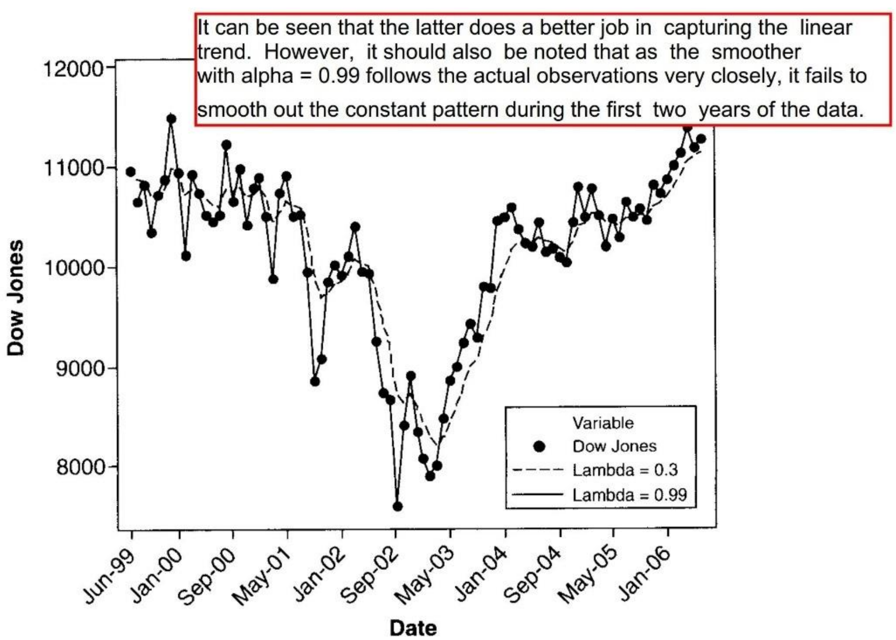


FIGURE 4.11 The Dow Jones Index from June 1999 to June 2006 using exponential smoothing with $\lambda = 0.3$ and 0.99.

Assume a local linear (trend) model for the mean of the series:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \qquad \mathbb{E}(\varepsilon_t) = 0,\ \mathrm{Var}(\varepsilon_t) = \sigma^2.$$

Define the first- and second-order exponential smoothers (with the same smoothing constant $\alpha \in (0,1)$):

$$S_t^{(1)} = \alpha Y_t + (1-\alpha) S_{t-1}^{(1)}, \qquad S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}.$$

Under $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$, the expected value of the simple (first-order) exponential smoother is

$$\mathbb{E}\left[S_T^{(1)}\right] = \beta_0 + \beta_1 T - \frac{1-\alpha}{\alpha}\,\beta_1.$$

Thus the first-order smoother lags a linear trend by a constant bias $-\big((1-\alpha)/\alpha\big)\beta_1$

Let $c = \dfrac{1-\alpha}{\alpha}$ for brevity

Smoothing $S_t^{(1)}$ again yields a second-order smoother that also has a constant bias under the same trend.

$$\begin{aligned}
\mathbb{E}\left[S_T^{(1)}\right] &= \beta_0 + \beta_1 T - c\,\beta_1, \\
\mathbb{E}\left[S_T^{(2)}\right] &= \beta_0 + \beta_1 T - 2c\,\beta_1.
\end{aligned}$$

Subtract the two equations:

$$\mathbb{E}\left[S_T^{(1)} - S_T^{(2)}\right] = (\beta_0 + \beta_1 T - c\beta_1) - (\beta_0 + \beta_1 T - 2c\beta_1) = c\,\beta_1.$$

Hence an unbiased estimator of β_1 is

$$\hat{\beta}_{1,T} = \frac{1}{c}\left(S_T^{(1)} - S_T^{(2)}\right) = \frac{\alpha}{1-\alpha}\left(S_T^{(1)} - S_T^{(2)}\right),$$

From the first bias equation,

$$\beta_0 + \beta_1 T = \mathbb{E}\left[S_T^{(1)}\right] + c\,\beta_1 \quad \Longrightarrow \quad \hat{\beta}_{0,T} = S_T^{(1)} + c\,\hat{\beta}_{1,T} - \hat{\beta}_{1,T} T.$$

Substitute $\hat{\beta}_{1,T} = \dfrac{\alpha}{1-\alpha}\left(S_T^{(1)} - S_T^{(2)}\right)$ and $c = \dfrac{1-\alpha}{\alpha}$ and simplify

Second-Order Smoothing Formula

We smooth the smoothed values: $S_t^{(1)} = \alpha Y_t + (1-\alpha) S_{t-1}^{(1)}$, $S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$

From these, we estimate:

- Level: $a_t = 2S_t^{(1)} - S_t^{(2)}$
- Trend: $b_t = \frac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right)$

Forecast: $F_{t+m} = a_t + b_t m$

Example: U.S. Consumer Price Index

- First-order smoothing underestimates CPI because of linear upward trend.
- Second-order smoothing captures both level and trend, removing bias.

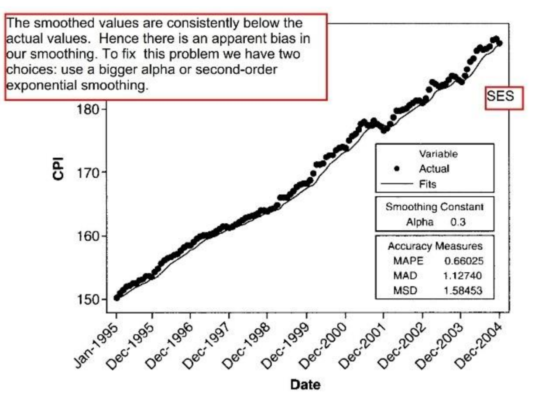


FIGURE 4.14 Single exponential smoothing of the U.S. Consumer Price Index (with $\tilde{y}_0 = y_1$).

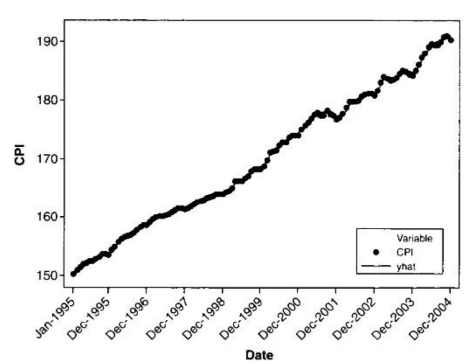


FIGURE 4.15 Second-order exponential smoothing of the U.S. Consumer Price Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$).

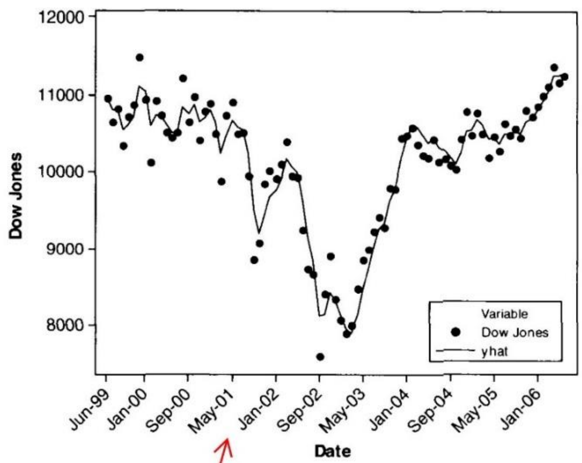


FIGURE 4.17 The second-order exponential smoothing of the Dow Jones Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$).

5 Higher-Order Exponential Smoothing

For a polynomial trend of degree n: $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2/2! + \ldots + \beta_n t^n/n! + \varepsilon_t$ we use $n+1$-order exponential smoothing.

- Rarely used in practice due to complexity.
- For higher-order trends, ARIMA models are more efficient.

6 Forecasting with Exponential Smoothing

Constant Process Forecast

Forecast for all future periods is simply the last smoothed value: $F_{t+m} = S_t$

Updated recursively: $F_{t+1}(t) = F_t(t-1) + \alpha e_t$ where $e_t = Y_t - F_t(t-1)$ is the forecast error.

Linear Trend Forecast

Using second-order smoothing: $F_{t+m} = a_t + b_t m$

Choosing α (Optimization)

Select α by minimizing: $SSE(\alpha) = \sum e_t^2$

Prediction Intervals

For confidence level $1-\alpha$: $F_{t+m} \pm z_{\alpha/2} \cdot \hat{\sigma}_e$ where $\hat{\sigma}_e$ is the estimated forecast error standard deviation.

7 Exponential Smoothing for Seasonal Data

When data has seasonality (repeated patterns over time), simple smoothing is inadequate.

Holt-Winters Method

Separates series into:

1. Level (L_t)
2. Trend (T_t)
3. Seasonal (S_t)

Additive Form (stable seasonal effect):

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

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

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

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

Multiplicative Form (proportional seasonal effect):

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

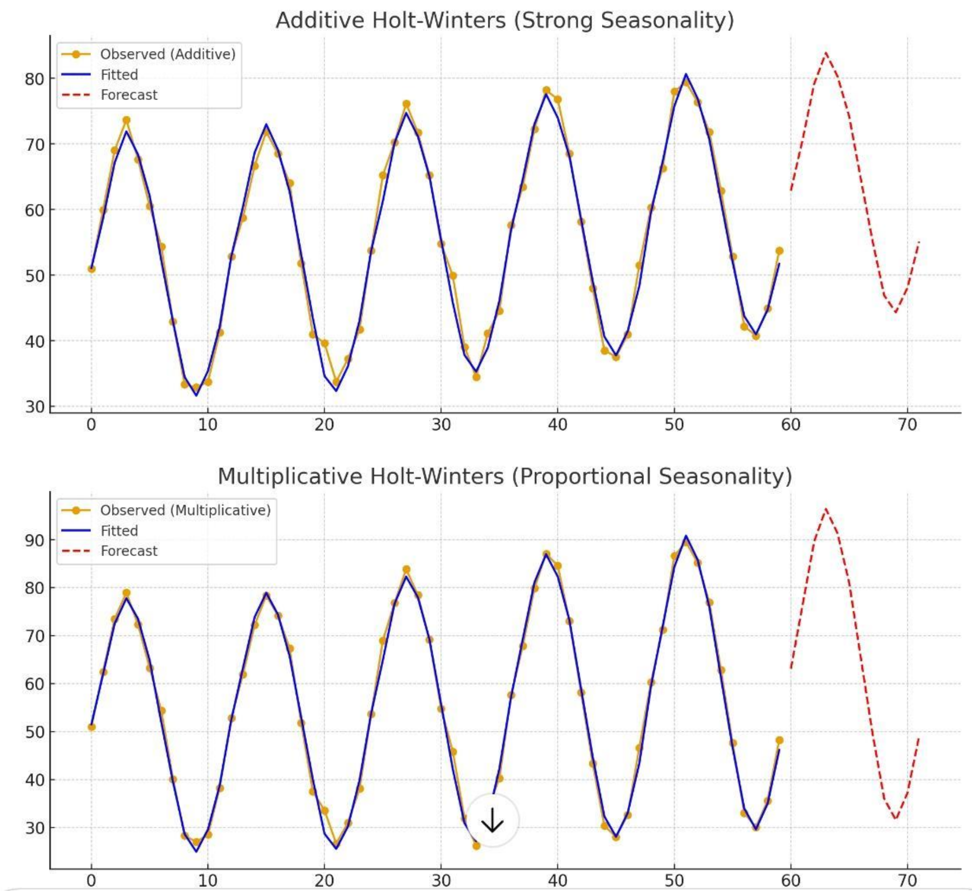
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

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

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

Choosing Between Additive and Multiplicative

- Use additive if seasonal variation is roughly constant in magnitude.
- Use multiplicative if seasonal variation grows with the series level.



- **Additive model**: the seasonal peaks and troughs remain about the same height over time, even as the trend grows.
- **Multiplicative model**: the seasonal swings get larger as the overall level increases (the amplitude of the seasonal effect scales with the series).