

10 Forecasting Techniques

- 10.1 The ARAR Algorithm
- 10.2 The Holt–Winters Algorithm
- 10.3 The Holt–Winters Seasonal Algorithm
- 10.4 Choosing a Forecasting Algorithm

We have focused until now on the construction of time series models for stationary and nonstationary series and the determination, assuming the appropriateness of these models, of minimum mean squared error predictors. If the observed series had in fact been generated by the fitted model, this procedure would give minimum mean squared error forecasts. In this chapter we discuss three forecasting techniques that have less emphasis on the explicit construction of a model for the data. Each of the three selects, from a limited class of algorithms, the one that is optimal according to specified criteria.

The three techniques have been found in practice to be effective on wide ranges of real data sets (for example, the economic time series used in the forecasting competition described by Makridakis et al. 1984).

The ARAR algorithm described in Section 10.1 is an adaptation of the ARARMA algorithm (Newton and Parzen 1984; Parzen 1982) in which the idea is to apply automatically selected “memory-shortening” transformations (if necessary) to the data and then to fit an ARMA model to the transformed series. The ARAR algorithm we describe is a version of this in which the ARMA fitting step is replaced by the fitting of a subset AR model to the transformed data.

The Holt–Winters (HW) algorithm described in Section 10.2 uses a set of simple recursions that generalize the exponential smoothing recursions of Section 1.5.1 to generate forecasts of series containing a locally linear trend.

The Holt–Winters seasonal (HWS) algorithm extends the HW algorithm to handle data in which there are both trend and seasonal variation of known period. It is described in Section 10.3.

Each of these three algorithms can be applied to specific data sets with the aid of the ITSM options `Forecasting>ARAR`, `Forecasting>Holt-Winters` and `Forecasting>Seasonal Holt-Winters`.

10.1 The ARAR Algorithm

10.1.1 Memory Shortening

Given a data set $\{Y_t, t = 1, 2, \dots, n\}$, the first step is to decide whether the underlying process is “long-memory,” and if so to apply a memory-shortening transformation before attempting to fit an autoregressive model. The differencing operations permitted under the option `Transform` of ITSM are examples of memory-shortening transformations; however, the ones used by the option `Forecasting>ARAR` selects are members of a more general class. There are two types allowed:

$$\tilde{Y}_t = Y_t - \hat{\phi}(\hat{\tau}) Y_{t-\hat{\tau}} \quad (10.1.1)$$

and

$$\tilde{Y}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2}. \quad (10.1.2)$$

With the aid of the five-step algorithm described below, we classify $\{Y_t\}$ and take one of the following three courses of action:

- **L.** Declare $\{Y_t\}$ to be long-memory and form $\{\tilde{Y}_t\}$ using (10.1.1).
- **M.** Declare $\{Y_t\}$ to be moderately long-memory and form $\{\tilde{Y}_t\}$ using (10.1.2).
- **S.** Declare $\{Y_t\}$ to be short-memory.

If the alternative L or M is chosen, then the transformed series $\{\tilde{Y}_t\}$ is again checked. If it is found to be long-memory or moderately long-memory, then a further transformation is performed. The process continues until the transformed series is classified as short-memory. At most three memory-shortening transformations are performed, but it is very rare to require more than two. The algorithm for deciding among L, M, and S can be described as follows:

1. For each $\tau = 1, 2, \dots, 15$, we find the value $\hat{\phi}(\tau)$ of ϕ that minimizes

$$\text{ERR}(\phi, \tau) = \frac{\sum_{t=\tau+1}^n [Y_t - \phi Y_{t-\tau}]^2}{\sum_{t=\tau+1}^n Y_t^2}.$$

We then define

$$\text{Err}(\tau) = \text{ERR}(\hat{\phi}(\tau), \tau)$$

and choose the lag $\hat{\tau}$ to be the value of τ that minimizes $\text{Err}(\tau)$.

2. If $\text{Err}(\hat{\tau}) \leq 8/n$, go to L.
3. If $\hat{\phi}(\hat{\tau}) \geq 0.93$ and $\hat{\tau} > 2$, go to L.
4. If $\hat{\phi}(\hat{\tau}) \geq 0.93$ and $\hat{\tau} = 1$ or 2 , determine the values $\hat{\phi}_1$ and $\hat{\phi}_2$ of ϕ_1 and ϕ_2 that minimize $\sum_{t=3}^n [Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}]^2$; then go to M.
5. If $\hat{\phi}(\hat{\tau}) < 0.93$, go to S.

10.1.2 Fitting a Subset Autoregression

Let $\{S_t, t = k + 1, \dots, n\}$ denote the memory-shortened series derived from $\{Y_t\}$ by the algorithm of the previous section and let \bar{S} denote the sample mean of S_{k+1}, \dots, S_n .

The next step in the modeling procedure is to fit an autoregressive process to the mean-corrected series

$$X_t = S_t - \bar{S}, \quad t = k + 1, \dots, n.$$

The fitted model has the form

$$X_t = \phi_1 X_{t-1} + \phi_{l_1} X_{t-l_1} + \phi_{l_2} X_{t-l_2} + \phi_{l_3} X_{t-l_3} + Z_t,$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and for given lags, l_1, l_2 , and l_3 , the coefficients ϕ_j and the white noise variance σ^2 are found from the Yule–Walker equations

$$\begin{bmatrix} 1 & \hat{\rho}(l_1 - 1) & \hat{\rho}(l_2 - 1) & \hat{\rho}(l_3 - 1) \\ \hat{\rho}(l_1 - 1) & 1 & \hat{\rho}(l_2 - l_1) & \hat{\rho}(l_3 - l_1) \\ \hat{\rho}(l_2 - 1) & \hat{\rho}(l_2 - l_1) & 1 & \hat{\rho}(l_3 - l_2) \\ \hat{\rho}(l_3 - 1) & \hat{\rho}(l_3 - l_1) & \hat{\rho}(l_3 - l_2) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_{l_1} \\ \phi_{l_2} \\ \phi_{l_3} \end{bmatrix} = \begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(l_1) \\ \hat{\rho}(l_2) \\ \hat{\rho}(l_3) \end{bmatrix}$$

and

$$\sigma^2 = \hat{\gamma}(0) [1 - \phi_1 \hat{\rho}(1) - \phi_{l_1} \hat{\rho}(l_1) - \phi_{l_2} \hat{\rho}(l_2) - \phi_{l_3} \hat{\rho}(l_3)],$$

where $\hat{\gamma}(j)$ and $\hat{\rho}(j), j = 0, 1, 2, \dots$, are the sample autocovariances and autocorrelations of the series $\{X_t\}$.

The program computes the coefficients ϕ_j for each set of lags such that

$$1 < l_1 < l_2 < l_3 \leq m,$$

where m can be chosen to be either 13 or 26. It then selects the model for which the Yule–Walker estimate σ^2 is minimal and prints out the lags, coefficients, and white noise variance for the fitted model.

A slower procedure chooses the lags and coefficients (computed from the Yule–Walker equations as above) that maximize the Gaussian likelihood of the observations. For this option the maximum lag m is 13.

The options are displayed in the ARAR Forecasting dialog box, which appears on the screen when the option Forecasting>ARAR is selected. It allows you also to bypass memory shortening and fit a subset AR to the original (mean-corrected) data.

10.1.3 Forecasting

If the memory-shortening filter found in the first step has coefficients $\psi_0 (= 1), \psi_1, \dots, \psi_k (k \geq 0)$, then the memory-shortened series can be expressed as

$$S_t = \psi(B)Y_t = Y_t + \psi_1 Y_{t-1} + \dots + \psi_k Y_{t-k}, \quad (10.1.3)$$

where $\psi(B)$ is the polynomial in the backward shift operator,

$$\psi(B) = 1 + \psi_1 B + \dots + \psi_k B^k.$$

Similarly, if the coefficients of the subset autoregression found in the second step are $\phi_1, \phi_{l_1}, \phi_{l_2}$, and ϕ_{l_3} , then the subset AR model for the mean-corrected series $\{X_t = S_t - \bar{S}\}$ is

$$\phi(B)X_t = Z_t, \quad (10.1.4)$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and

$$\phi(B) = 1 - \phi_1 B - \phi_{l_1} B^{l_1} - \phi_{l_2} B^{l_2} - \phi_{l_3} B^{l_3}.$$

From (10.1.3) and (10.1.4) we obtain the equations

$$\xi(B)Y_t = \phi(1)\bar{S} + Z_t, \quad (10.1.5)$$

where

$$\xi(B) = \psi(B)\phi(B) = 1 + \xi_1 B + \cdots + \xi_{k+l_3} B^{k+l_3}.$$

Assuming that the fitted model (10.1.5) is appropriate and that the white noise term Z_t is uncorrelated with $\{Y_j, j < t\}$ for each t , we can determine the minimum mean squared error linear predictors $P_n Y_{n+h}$ of Y_{n+h} in terms of $\{1, Y_1, \dots, Y_n\}$, for $n > k + l_3$, from the recursions

$$P_n Y_{n+h} = - \sum_{j=1}^{k+l_3} \xi_j P_n Y_{n+h-j} + \phi(1)\bar{S}, \quad h \geq 1, \quad (10.1.6)$$

with the initial conditions

$$P_n Y_{n+h} = Y_{n+h}, \quad \text{for } h \leq 0. \quad (10.1.7)$$

The mean squared error of the predictor $P_n Y_{n+h}$ is found to be (Problem 10.1)

$$E[(Y_{n+h} - P_n Y_{n+h})^2] = \sum_{j=0}^{h-1} \tau_j^2 \sigma^2, \quad (10.1.8)$$

where $\sum_{j=0}^{\infty} \tau_j z^j$ is the Taylor expansion of $1/\xi(z)$ in a neighborhood of $z = 0$. Equivalently the sequence $\{\tau_j\}$ can be found from the recursion

$$\tau_0 = 1, \quad \sum_{j=0}^n \tau_j \xi_{n-j} = 0, \quad n = 1, 2, \dots \quad (10.1.9)$$

10.1.4 Application of the ARAR Algorithm

To determine an ARAR model for a given data set $\{Y_t\}$ using ITSM, select `Forecasting>ARAR` and choose the appropriate options in the resulting dialog box. These include specification of the number of forecasts required, whether or not you wish to include the memory-shortening step, whether you require prediction bounds, and which of the optimality criteria is to be used. Once you have made these selections, click `OK`, and the forecasts will be plotted with the original data. Right-click on the graph and then `Info` to see the coefficients $1, \psi_1, \dots, \psi_k$ of the memory-shortening filter $\psi(B)$, the lags and coefficients of the subset autoregression

$$X_t - \phi_1 X_{t-1} - \phi_{l_1} X_{t-l_1} - \phi_{l_2} X_{t-l_2} - \phi_{l_3} X_{t-l_3} = Z_t,$$

and the coefficients ξ_j of B^j in the overall whitening filter

$$\xi(B) = (1 + \psi_1 B + \cdots + \psi_k B^k) (1 - \phi_1 B - \phi_{l_1} B^{l_1} - \phi_{l_2} B^{l_2} - \phi_{l_3} B^{l_3}).$$

The numerical values of the predictors, their root mean squared errors, and the prediction bounds are also printed.

Example 10.1.1 To use the ARAR algorithm to predict 24 values of the accidental deaths data, open the file `DEATHS.TSM` and proceed as described above. Selecting `Minimize WN variance [max lag=26]` gives the graph of the data and

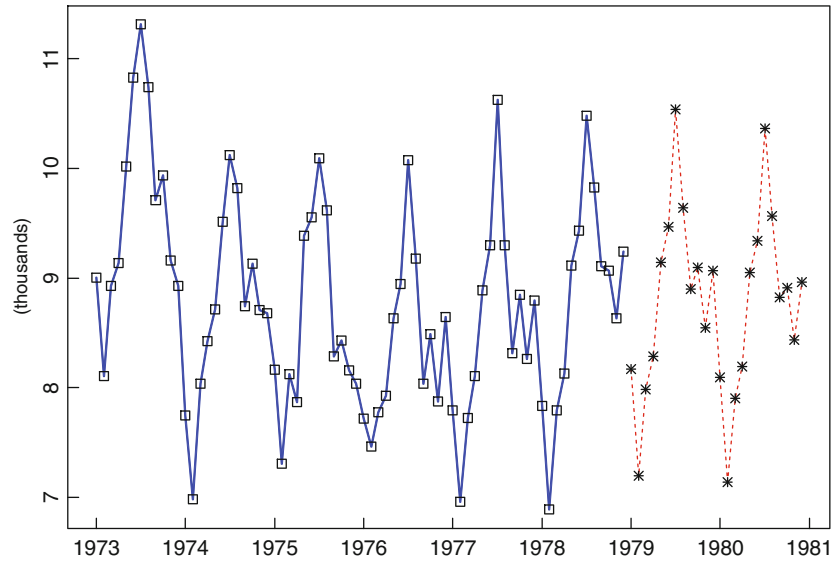


Figure 10-1

The data set DEATHS.TSM with 24 values predicted by the ARAR algorithm

predictors shown in Figure 10-1. Right-clicking on the graph and then Info, we find that the selected memory-shortening filter is $(1 - 0.9779B^{12})$. The fitted subset autoregression and the coefficients ξ_j of the overall whitening filter $\xi(B)$ are shown below: □

Optimal lags	1	3	12	13
Optimal coeffs	0.5915	-0.3822	-0.3022	0.2970
WN Variance:	0.12314E+06			
COEFFICIENTS OF OVERALL WHITENING FILTER:				
1.0000	-0.5915	0.0000	-0.2093	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	-0.6757	0.2814	0.0000
0.2047	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	-0.2955
0.2904				

In Table 10.1 we compare the predictors of the next six values of the accidental deaths series with the actual observed values. The predicted values obtained from ARAR as described in the example are shown together with the predictors obtained by fitting ARIMA models as described in Chapter 6 (see Table 10.1). The observed root mean squared errors (i.e., $\sqrt{\sum_{h=1}^6 (Y_{72+h} - P_{72}Y_{72+h})^2 / 6}$) for the three prediction methods are easily calculated to be 253 for ARAR, 583 for the ARIMA model (6.5.8), and 501 for the ARIMA model (6.5.9). The ARAR algorithm thus performs very well here. Notice that in this particular example the ARAR algorithm effectively fits a causal AR model to the data, but this is not always the case. □

10.2 The Holt–Winters Algorithm

10.2.1 The Algorithm

Given observations Y_1, Y_2, \dots, Y_n from the “trend plus noise” model (1.5.2), the exponential smoothing recursions (1.5.7) allowed us to compute estimates \hat{m}_t of the trend at times $t = 1, 2, \dots, n$. If the series is stationary, then m_t is constant and the exponential smoothing forecast of Y_{n+h} based on the observations Y_1, \dots, Y_n is

$$P_n Y_{n+h} = \hat{m}_n, \quad h = 1, 2, \dots \quad (10.2.1)$$

If the data have a (nonconstant) trend, then a natural generalization of the forecast function (10.2.1) that takes this into account is

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h, \quad h = 1, 2, \dots, \quad (10.2.2)$$

where \hat{a}_n and \hat{b}_n can be thought of as estimates of the “level” a_n and “slope” b_n of the trend function at time n . Holt (1957) suggested a recursive scheme for computing the quantities \hat{a}_n and \hat{b}_n in (10.2.2). Denoting by \hat{Y}_{n+1} the one-step forecast $P_n Y_{n+1}$, we have from (10.2.2)

$$\hat{Y}_{n+1} = \hat{a}_n + \hat{b}_n.$$

Now, as in exponential smoothing, we suppose that the estimated level at time $n + 1$ is a linear combination of the observed value at time $n + 1$ and the forecast value at time $n + 1$. Thus,

$$\hat{a}_{n+1} = \alpha Y_{n+1} + (1 - \alpha)(\hat{a}_n + \hat{b}_n). \quad (10.2.3)$$

We can then estimate the slope at time $n + 1$ as a linear combination of $\hat{a}_{n+1} - \hat{a}_n$ and the estimated slope \hat{b}_n at time n . Thus,

$$\hat{b}_{n+1} = \beta (\hat{a}_{n+1} - \hat{a}_n) + (1 - \beta)\hat{b}_n. \quad (10.2.4)$$

In order to solve the recursions (10.2.3) and (10.2.4) we need initial conditions. A natural choice is to set

$$\hat{a}_2 = Y_2 \quad (10.2.5)$$

and

$$\hat{b}_2 = Y_2 - Y_1. \quad (10.2.6)$$

Then (10.2.3) and (10.2.4) can be solved successively for \hat{a}_i and \hat{b}_i , $i = 3, \dots, n$, and the predictors $P_n Y_{n+h}$ found from (10.2.2).

The forecasts depend on the “smoothing parameters” α and β . These can either be prescribed arbitrarily (with values between 0 and 1) or chosen in a more systematic way to minimize the sum of squares of the one-step errors $\sum_{i=3}^n (Y_i - P_{i-1} Y_i)^2$, obtained

Table 10.1 Predicted and observed values of the accidental deaths series for $t = 73, \dots, 78$

t	73	74	75	76	77	78
Observed Y_t	7798	7406	8363	8460	9217	9316
Predicted by ARAR	8168	7196	7982	8284	9144	9465
Predicted by (6.5.8)	8441	7704	8549	8885	9843	10,279
Predicted by (6.5.9)	8345	7619	8356	8742	9795	10,179

when the algorithm is applied to the already observed data. Both choices are available in the ITSM option `Forecasting>Holt-Winters`.

Before illustrating the use of the Holt–Winters forecasting procedure, we discuss the connection between the recursions (10.2.3) and (10.2.4) and the steady-state solution of the Kalman filtering equations for a local linear trend model. Suppose $\{Y_t\}$ follows the local linear structural model with observation equation

$$Y_t = M_t + W_t$$

and state equation

$$\begin{bmatrix} M_{t+1} \\ B_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_t \\ B_t \end{bmatrix} + \begin{bmatrix} V_t \\ U_t \end{bmatrix}$$

[see (9.2.4)–(9.2.7)]. Now define \hat{a}_n and \hat{b}_n to be the filtered estimates of M_n and B_n , respectively, i.e.,

$$\hat{a}_n = M_{n|n} := P_n M_n,$$

$$\hat{b}_n = B_{n|n} := P_n B_n.$$

Using Problem 9.17 and the Kalman recursion (9.4.16), we find that

$$\begin{bmatrix} \hat{a}_{n+1} \\ \hat{b}_{n+1} \end{bmatrix} = \begin{bmatrix} \hat{a}_n + \hat{b}_n \\ \hat{b}_n \end{bmatrix} + \Delta_n^{-1} \Omega_n G' (Y_n - \hat{a}_n - \hat{b}_n), \quad (10.2.7)$$

where $G = [1 \ 0]$. Assuming that $\Omega_n = \Omega_1 = [\Omega_{ij}]_{i,j=1}^2$ is the steady-state solution of (9.4.2) for this model, then $\Delta_n = \Omega_{11} + \sigma_w^2$ for all n , so that (10.2.7) simplifies to the equations

$$\hat{a}_{n+1} = \hat{a}_n + \hat{b}_n + \frac{\Omega_{11}}{\Omega_{11} + \sigma_w^2} (Y_n - \hat{a}_n - \hat{b}_n) \quad (10.2.8)$$

and

$$\hat{b}_{n+1} = \hat{b}_n + \frac{\Omega_{12}}{\Omega_{11} + \sigma_w^2} (Y_n - \hat{a}_n - \hat{b}_n). \quad (10.2.9)$$

Solving (10.2.8) for $(Y_n - \hat{a}_n - \hat{b}_n)$ and substituting into (10.2.9), we find that

$$\hat{a}_{n+1} = \alpha Y_{n+1} + (1 - \alpha) (\hat{a}_n + \hat{b}_n), \quad (10.2.10)$$

$$\hat{b}_{n+1} = \beta (\hat{a}_{n+1} - \hat{a}_n) + (1 - \beta) \hat{b}_n \quad (10.2.11)$$

with $\alpha = \Omega_{11} / (\Omega_{11} + \sigma_w^2)$ and $\beta = \Omega_{21} / \Omega_{11}$. These equations coincide with the Holt–Winters recursions (10.2.3) and (10.2.4). Equations relating α and β to the variances σ_u^2 , σ_v^2 , and σ_w^2 can be found in Harvey (1990).

Example 10.2.1

To predict 24 values of the accidental deaths series using the Holt–Winters algorithm, open the file `DEATHS.TSM` and select `Forecasting>Holt-Winters`. In the resulting dialog box specify 24 for the number of predictors and check the box marked `Optimize coefficients` for automatic selection of the smoothing coefficients α and β . Click `OK`, and the forecasts will be plotted with the original data as shown in Figure 10-2. Right-click on the graph and then `Info` to see the numerical values of the predictors, their root mean squared errors, and the optimal values of α and β . The predicted and observed values are shown in Table 10.2.

□

The root mean squared error $(\sqrt{\sum_{h=1}^6 (Y_{72+h} - P_{72} Y_{72+h})^2 / 6})$ for the nonseasonal Holt–Winters forecasts is found to be 1143. Not surprisingly, since we have not taken

seasonality into account, this is a much larger value than for the three sets of forecasts shown in Table 10.1. In the next section we show how to modify the Holt–Winters algorithm to allow for seasonality.

10.2.2 Holt–Winters and ARIMA Forecasting

The one-step forecasts obtained by exponential smoothing with parameter α (defined by (1.5.7) and (10.2.1)) satisfy the relations

$$P_n Y_{n+1} = Y_n - (1 - \alpha)(Y_n - P_{n-1} Y_n), \quad n \geq 2. \quad (10.2.12)$$

But these are the same relations satisfied by the large-sample minimum mean squared error forecasts of the invertible ARIMA(0,1,1) process

$$Y_t = Y_{t-1} + Z_t - (1 - \alpha)Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2). \quad (10.2.13)$$

Forecasting by exponential smoothing with optimal α can therefore be viewed as fitting a member of the two-parameter family of ARIMA processes (10.2.13) to the data and using the corresponding large-sample forecast recursions initialized by $P_0 Y_1 = Y_1$. In ITSM, the optimal α is found by minimizing the average squared error of the one-step forecasts of the *observed* data Y_2, \dots, Y_n , and the parameter σ^2 is estimated by this average squared error. This algorithm could easily be modified to minimize other error measures such as average absolute one-step error and average 12-step squared error.

In the same way it can be shown that Holt–Winters forecasting can be viewed as fitting a member of the three-parameter family of ARIMA processes,

$$(1 - B)^2 Y_t = Z_t - (2 - \alpha - \alpha\beta)Z_{t-1} + (1 - \alpha)Z_{t-2}, \quad (10.2.14)$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. The coefficients α and β are selected as described after (10.2.6), and the estimate of σ^2 is the average squared error of the one-step forecasts of Y_3, \dots, Y_n obtained from the large-sample forecast recursions corresponding to (10.2.14).

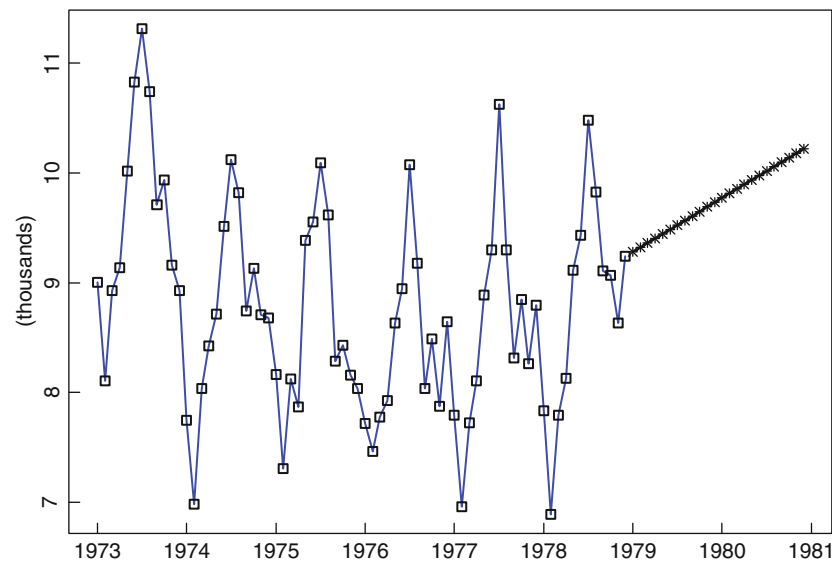


Figure 10-2

The data set DEATHS.TSM with 24 values predicted by the nonseasonal Holt–Winters algorithm

Table 10.2 Predicted and observed values of the accidental deaths series for $t = 73, \dots, 78$ from the (nonseasonal) Holt–Winters algorithm

t	73	74	75	76	77	78
Observed Y_t	7798	7406	8363	8460	9217	9316
Predicted by HW	9281	9322	9363	9404	9445	9486

10.3 The Holt–Winters Seasonal Algorithm

10.3.1 The Algorithm

If the series Y_1, Y_2, \dots, Y_n contains not only trend, but also seasonality with period d [as in the model (1.5.11)], then a further generalization of the forecast function (10.2.2) that takes this into account is

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h + \hat{c}_{n+h}, \quad h = 1, 2, \dots, \quad (10.3.1)$$

where \hat{a}_n , \hat{b}_n , and \hat{c}_n can be thought of as estimates of the “trend level” a_n , “trend slope” b_n , and “seasonal component” c_n at time n . If k is the smallest integer such that $n + h - kd \leq n$, then we set

$$\hat{c}_{n+h} = \hat{c}_{n+h-kd}, \quad h = 1, 2, \dots, \quad (10.3.2)$$

while the values of \hat{a}_i , \hat{b}_i , and \hat{c}_i , $i = d+2, \dots, n$, are found from recursions analogous to (10.2.3) and (10.2.4), namely,

$$\hat{a}_{n+1} = \alpha (Y_{n+1} - \hat{c}_{n+1-d}) + (1 - \alpha)(\hat{a}_n + \hat{b}_n), \quad (10.3.3)$$

$$\hat{b}_{n+1} = \beta (\hat{a}_{n+1} - \hat{a}_n) + (1 - \beta)\hat{b}_n, \quad (10.3.4)$$

and

$$\hat{c}_{n+1} = \gamma (Y_{n+1} - \hat{a}_{n+1}) + (1 - \gamma)\hat{c}_{n+1-d}, \quad (10.3.5)$$

with initial conditions

$$\hat{a}_{d+1} = Y_{d+1}, \quad (10.3.6)$$

$$\hat{b}_{d+1} = (Y_{d+1} - Y_1)/d, \quad (10.3.7)$$

and

$$\hat{c}_i = Y_i - (Y_1 + \hat{b}_{d+1}(i-1)), \quad i = 1, \dots, d+1. \quad (10.3.8)$$

Then (10.3.3)–(10.3.5) can be solved successively for \hat{a}_i , \hat{b}_i , and \hat{c}_i , $i = d+1, \dots, n$, and the predictors $P_n Y_{n+h}$ found from (10.3.1).

As in the nonseasonal case of Section 10.2, the forecasts depend on the parameters α , β , and γ . These can either be prescribed arbitrarily (with values between 0 and 1) or chosen in a more systematic way to minimize the sum of squares of the one-step errors $\sum_{i=d+2}^n (Y_i - P_{i-1} Y_i)^2$, obtained when the algorithm is applied to the already observed data. Seasonal Holt–Winters forecasts can be computed by selecting the ITSM option `Forecasting>Seasonal Holt-Winters`.

Example 10.3.1 As in Example 10.2.1, open the file DEATHS.TSM, but this time select `Forecasting>Seasonal Holt-Winters`. Specify 24 for the number of predicted

values required, 12 for the period of the seasonality, and check the box marked `Optimize Coefficients`. Click `OK`, and the graph of the data and predicted values shown in Figure 10-3 will appear. Right-click on the graph and then on `Info` and you will see the numerical values of the predictors and the optimal values of the coefficients α , β , and γ (minimizing the observed one-step average squared error $\sum_{i=14}^{72} (Y_i - P_{i-1}Y_i)^2/59$). Table 10.3 compares the predictors of Y_{73}, \dots, Y_{78} with the corresponding observed values. □

The root mean squared error ($\sqrt{\sum_{h=1}^6 (Y_{72+h} - P_{72}Y_{72+h})^2/6}$) for the seasonal Holt–Winters forecasts is found to be 401. This is not as good as the value 253 achieved by the ARAR model for this example but is substantially better than the values achieved by the nonseasonal Holt–Winters algorithm (1143) and the ARIMA models (6.5.8) and (6.5.9) (583 and 501, respectively).

10.3.2 Holt–Winters Seasonal and ARIMA Forecasting

As in Section 10.2.2, the Holt–Winters seasonal recursions with seasonal period d correspond to the large-sample forecast recursions of an ARIMA process, in this case defined by

$$\begin{aligned} (1 - B)(1 - B^d)Y_t &= Z_t + \dots + Z_{t-d+1} + \gamma(1 - \alpha)(Z_{t-d} - Z_{t-d-1}) \\ &\quad - (2 - \alpha - \alpha\beta)(Z_{t-1} + \dots + Z_{t-d}) \\ &\quad + (1 - \alpha)(Z_{t-2} + \dots + Z_{t-d-1}), \end{aligned}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. Holt–Winters seasonal forecasting with optimal α , β , and γ can therefore be viewed as fitting a member of this four-parameter family of ARIMA models and using the corresponding large-sample forecast recursions.

Table 10.3 Predicted and observed values of the accidental deaths series for $t = 73, \dots, 78$ from the seasonal Holt–Winters algorithm

t	73	74	75	76	77	78
Observed Y_t	7798	7406	8363	8460	9217	9316
Predicted by HWS	8039	7077	7750	7941	8824	9329

10.4 Choosing a Forecasting Algorithm

Real data are rarely if ever generated by a simple mathematical model such as an ARIMA process. Forecasting methods that are predicated on the assumption of such a model are therefore not necessarily the best, even in the mean squared error sense. Nor is the measurement of error in terms of mean squared error necessarily always the most appropriate one in spite of its mathematical convenience. Even within the framework of minimum mean squared-error forecasting, we may ask (for example) whether we wish to minimize the one-step, two-step, or twelve-step mean squared error.

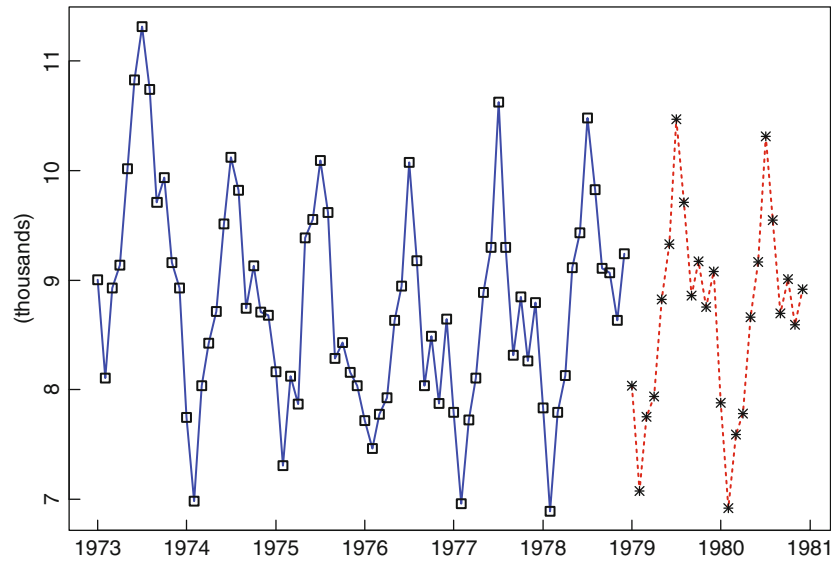


Figure 10-3

The data set DEATHS.TSM with 24 values predicted by the seasonal Holt–Winters algorithm

The use of more heuristic algorithms such as those discussed in this chapter is therefore well worth serious consideration in practical forecasting problems. But how do we decide which method to use? A relatively simple solution to this problem, given the availability of a substantial historical record, is to choose among competing algorithms by comparing the relevant errors when the algorithms are applied to the data already observed (e.g., by comparing the mean absolute percentage errors of the 12-step predictors of the historical data if 12-step prediction is of primary concern).

It is extremely difficult to make general theoretical statements about the relative merits of the various techniques we have discussed (ARIMA modeling, exponential smoothing, ARAR, and HW methods). For the series DEATHS.TSM we found on the basis of average mean squared error for predicting the series at times 73–78 that the ARAR method was best, followed by the seasonal Holt–Winters algorithm, and then the ARIMA models fitted in Chapter 6. This ordering is by no means universal. For example, if we consider the natural logarithms $\{Y_t\}$ of the first 130 observations in the series WINE.TSM (Figure 1-1) and compare the average mean squared errors of the forecasts of Y_{131}, \dots, Y_{142} , we find (Problem 10.2 that an MA(12) model fitted to the mean corrected differenced series $\{Y_t - Y_{t-12}\}$ does better than seasonal Holt–Winters (with period 12), which in turn does better than ARAR and (not surprisingly) dramatically better than nonseasonal Holt–Winters. An interesting empirical comparison of these and other methods applied to a variety of economic time series is contained in Makridakis et al. (1984).

The versions of the Holt–Winters algorithms we have discussed in Sections 10.2 and 10.3 are referred to as “additive,” since the seasonal and trend components enter the forecasting function in an additive manner. “Multiplicative” versions of the algorithms can also be constructed to deal directly with processes of the form

$$Y_t = m_t s_t Z_t, \quad (10.4.1)$$

where m_t , s_t , and Z_t are trend, seasonal, and noise factors, respectively (see, e.g., Makridakis et al. 1997). An alternative approach (provided that $Y_t > 0$ for all t) is to apply the linear Holt–Winters algorithms to $\{\ln Y_t\}$ (as in the case of WINE.TSM in the preceding paragraph). Because of the rather general memory shortening permitted by the ARAR algorithm, it gives reasonable results when applied directly to series of the form (10.4.1), even without preliminary transformations. In particular, if we

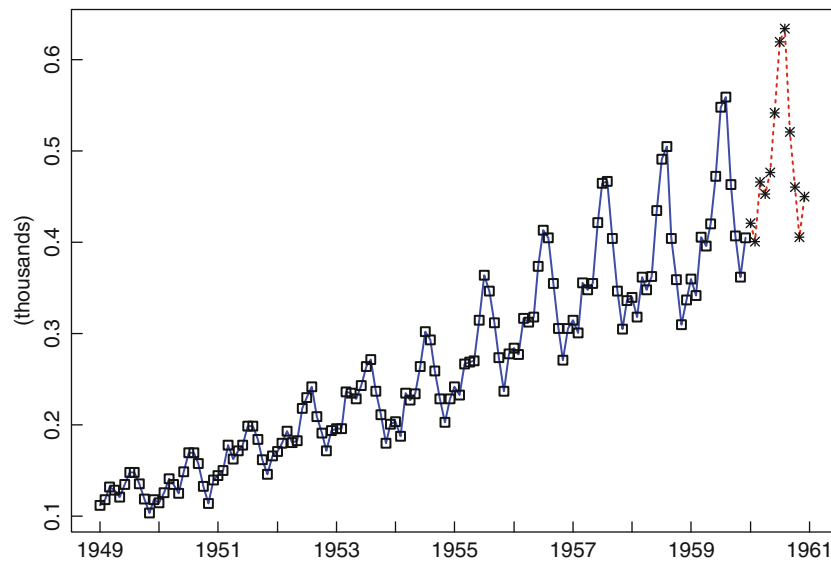


Figure 10-4
The first 132 values of the data set AIRPASS.TSM and predictors of the last 12 values obtained by direct application of the ARAR algorithm

consider the first 132 observations in the series AIRPASS.TSM and apply the ARAR algorithm to predict the last 12 values in the series, we obtain (Problem 10.4) an observed root mean squared error of 18.22. On the other hand if we use the same data take logarithms, difference at lag 12, subtract the mean and then fit an AR(13) model by maximum likelihood using ITSM and use it to predict the last 12 values, we obtain an observed root mean squared error of 21.17. The data and predicted values from the ARAR algorithm are shown in Figure 10-4.

Problems

- 10.1** Establish the formula (10.1.8) for the mean squared error of the h -step forecast based on the ARAR algorithm.
- 10.2** Let $\{X_1, \dots, X_{142}\}$ denote the data in the file WINE.TSM and let $\{Y_1, \dots, Y_{142}\}$ denote their natural logarithms. Denote by m the sample mean of the differenced series $\{Y_t - Y_{t-12}, t = 13, \dots, 130\}$.
- Use the program ITSM to find the maximum likelihood MA(12) model for the differenced and mean-corrected series $\{Y_t - Y_{t-12} - m, t = 13, \dots, 130\}$.
 - Use the model in (a) to compute forecasts of $\{X_{131}, \dots, X_{142}\}$.
 - Tabulate the forecast errors $\{X_t - P_{130} X_t, t = 131, \dots, 142\}$.
 - Compute the average squared error for the 12 forecasts.
 - Repeat steps (b), (c), and (d) for the corresponding forecasts obtained by applying the ARAR algorithm to the series $\{X_t, t = 1, \dots, 130\}$.
 - Repeat steps (b), (c), and (d) for the corresponding forecasts obtained by applying the seasonal Holt–Winters algorithm (with period 12) to the logged data $\{Y_t, t = 1, \dots, 130\}$. (Open the file WINE.TSM, select Transform>Box-Cox with parameter $\lambda = 0$, then select Forecasting>Seasonal Holt-Winters, and check Apply to original data in the dialog box.)

- (g) Repeat steps (b), (c), and (d) for the corresponding forecasts obtained by applying the nonseasonal Holt–Winters algorithm to the logged data $\{Y_t, t = 1, \dots, 130\}$. (The procedure is analogous to that described in part (f).)
- (h) Compare the average squared errors obtained by the four methods.
- 10.3** In equations (10.2.10) and (10.2.11), show that $\alpha = \Omega_{11} / (\Omega_{11} + \sigma_w^2)$ and $\beta = \Omega_{21} / \Omega_{11}$.
- 10.4** Verify the assertions made in the last paragraph of Section 10.4, comparing the forecasts of the last 12 values of the series AIRPASS.TSM obtained from the ARAR algorithm (with no log transformation) and the corresponding forecasts obtained by taking logarithms of the original series, then differencing at lag 12, mean-correcting, and fitting an AR(13) model to the transformed series.