B.Sc./Grad. Dip.: Probability Models and Time Series
MAS programmes: Stochastic Models and Time Series

10 Algorithmic Updating of Forecasts

10.1 Box-Jenkins Forecasting

10.1.1 Updating $\hat{y}_T(1)$ w.r.t. to T for the ARIMA(0,1,1) model

Suppose that, as successive values of the time series become available at unit intervals of time, we wish to update our forecasts of future values of the series. We may do this by using the forecasting equation derived in the previous chapter i.e.

$$\hat{y}_T(h) \approx (1+\theta)\sum_{k=0}^{T-1}(-\theta)^k y_{T-k}. \tag{1}$$

for successive values of T, yielding $\hat{y}_{T+1}(1), \hat{y}_{T+2}(1), \ldots, \ldots$. However, there is a simpler way which is computationally more efficient. Recalling the equation

$$e_T(1) = \epsilon_{T+1},$$

replace T by $T-1$ to obtain

$$y_T - \hat{y}_{T-1}(1) = e_{T-1}(1) = \epsilon_T,$$

the value of which becomes apparent at time T. Substituting this result into the expression

$$\hat{y}_T(h) = y_T + \theta\epsilon_T, \qquad h \geq 1$$

derived previously, in the case where $h = 1$, yields

$$\hat{y}_T(1) = y_T + \theta[y_T - \hat{y}_{T-1}(1)].$$

Rearranging the terms on the right hand side,

$$\hat{y}_T(1) = (1+\theta)y_T - \theta\hat{y}_{T-1}(1). \tag{2}$$

Equation (2) is an *updating equation*, which (if θ is negative) expresses the forecast at time T as a weighted average of the forecast at time $T-1$ and the observed value of the process at time T.

Assuming that the parameter θ is known from previous data, if a time series $y_1, y_2, \ldots$ starts to be observed at time 1 then the initial forecast $\hat{y}_1(1)$ may be taken to be y_1 or 0 or some other suitable value based on previous knowledge. The initial value suggested here differs from the approximation of Equation (1), according to which $\hat{y}_1(1) = (1+\theta)y_1$. Thereafter, Equation (2) may be used to update the forecast for successive values of T (assuming that the model for the data and corresponding parameter values are still correct). In any case, the effect of the value chosen for the initial forecast dies away exponentially with time.

10.1.2 Updating $\hat{y}_T(h),\ \ h \geq 1$ w.r.t. to T for the ARIMA(0,1,1) model

The expression of (1) can be rewritten as

$$\hat{y}_T(h) \approx (1+\theta)y_T + (1+\theta)\sum_{k=1}^{T-1}(-\theta)^k y_{T-k}$$

$$= (1+\theta)y_T + (1+\theta)\sum_{k=0}^{T-2}(-\theta)^{k+1} y_{T-1-k}$$

$$= (1+\theta)y_T + (-\theta)(1+\theta)\sum_{k=0}^{T-2}(-\theta)^k y_{T-1-k}$$

$$= (1+\theta)y_T - \theta\hat{y}_{T-1}(h).$$

10.1.3 Revision of forecast for Y_{T+h} in moving from $T-1$ to T for ARIMA(p,d,q)

For both the stationary $d=0$ and non-stationary $d \geq 1$ scenarios, it was previously discussed that

$$\begin{aligned} Y_{T+h} &= \mu + \sum_{i=0}^{\infty}\psi_i \epsilon_{T+h-i} \\ &= \hat{y}_T(h) + \sum_{i=0}^{h-1}\psi_i \epsilon_{T+h-i}. \end{aligned} \tag{3}$$

Putting $h=1$ into (3), one obtains

$$Y_{T+1} = \hat{y}_T(1) + \epsilon_{T+1}. \tag{4}$$

Taking expectations conditional upon $\mathcal{H}_{T+1}$ (instead of $\mathcal{H}_T$) in Equation (3) and noting that ϵ_{T+1} is determined by $\mathcal{H}_{T+1}$, we obtain, for $h \geq 2$,

$$\begin{aligned} \hat{y}_{T+1}(h-1) &= E[Y_{T+h}|\mathcal{H}_{T+1}] \\ &= \hat{y}_T(h) + \psi_{h-1}\epsilon_{T+1}. \end{aligned} \tag{5}$$

Replacing T by $T-1$ and h by $h+1$ in Equation (5) we obtain

$$\hat{y}_T(h) = \hat{y}_{T-1}(h+1) + \psi_h \epsilon_T, \qquad h \geq 1. \tag{6}$$

Equation (6) is yet another *updating equation*. It provides us with a method of updating our forecast $\hat{y}_T(h)$ of Y_{T+h} at time T from our forecast $\hat{y}_{T-1}(h+1)$ of Y_{T+h} at time $T-1$. The value of the second term on the right hand side of equation (6) becomes indirectly observable at time T, as is made explicit in the following.

For $j \leq 0$, replacing T by $T+j-1$ in Equation (4), and rearranging we obtain

$$\begin{aligned} \epsilon_{T+j} &= y_{T+j} - \hat{y}_{T+j-1}(1) && (7) \\ &= e_{T+j-1}(1). && (8) \end{aligned}$$

The two terms on the right hand side of Equation (7) are both known at time T, i.e., are determined by $\mathcal{H}_T$. So, in forecasting, all past error terms can be replaced by forecast errors. In particular, substituting from Equation (7), with $j = 0$, into Equation (6) we obtain

$$\hat{y}_T(h) = \hat{y}_{T-1}(h+1) + \psi_h(y_T - \hat{y}_{T-1}(1)), \qquad h \geq 1, \tag{9}$$

which is an alternative form of the updating equation.

10.2 Exponential smoothing

Forecasting methods for time series were in use well before the advent of ARIMA modelling, although they were of a somewhat *ad hoc* nature, without a solid theoretical basis. The simplest such *ad hoc* method of forecasting is what is often known as *exponential smoothing*, which uses an *exponentially weighted moving average*.

The method is based upon the recursion

$$\hat{y}_T(1) = \lambda y_T + (1-\lambda)\hat{y}_{T-1}(1), \tag{10}$$

which is identical with the updating equation (2) if λ is identified with $1 + \theta$. As we saw for the ARIMA(0,1,1) model, and as may be checked directly, use of this recursion is equivalent to using an exponentially weighted moving average of the observed values.

What the ARIMA modelling approach shows is that exponential smoothing is optimal, in the sense that it gives minimum mean square forecasts, if the data may be modelled as arising from an ARIMA(0,1,1) process.

The choice of a value for the *smoothing parameter* λ might be made on *a priori* grounds from previous experience, or by trial and error, to be one that gave the best forecasts.

10.3 Holt-Winters forecasting

Holt-Winters forecasting, an extension of the method of exponential smoothing, is an older ad hoc method that can nevertheless under some circumstances perform as well as Box-Jenkins forecasting and, furthermore, has the merit of simplicity.
The Holt-Winters method uses a forecast function of the form

$$\hat{y}_T(h) = m_T + b_T h, \qquad h \geq 1, \tag{11}$$

a linear function of h, where m_T represents an estimated underlying process mean at time T and b_T the estimated slope of a locally linear trend. The values of m_T and b_T are updated using the updating equations

$$m_T = \lambda_0 y_T + (1-\lambda_0)(m_{T-1} + b_{T-1}), \qquad T \geq 2 \tag{12}$$

and

$$b_T = \lambda_1(m_T - m_{T-1}) + (1-\lambda_1)b_{T-1}, \qquad T \geq 2, \tag{13}$$

where the *smoothing constants*, λ_0 and λ_1, satisfy $0 < \lambda_0 < 1$ and $0 < \lambda_1 < 1$. The above equations tells how to move from (m_{T-1}, b_{T-1}) at time $T-1$ to (m_T, b_T) at time T upon the appearance of the new datum y_T at time T.

Appropriate choices for the value of λ_0 and λ_1 need to be "learnt" through further investigation.

Suitable starting values, m_1 and b_1, may be suggested by previous experience. Another possibility is to take $m_1 = y_1$ and $b_1 = 0$ or else to start at $T = 2$ with $m_2 = y_2$ and $b_2 = y_2 - y_1$.

Noting, from (11), that $m_{T-1} + b_{T-1} = \hat{y}_{T-1}(1)$, we may rewrite Equation (12) as

$$m_T = \lambda_0 y_T + (1 - \lambda_0)\hat{y}_{T-1}(1).$$

If we are looking for an ARIMA(p, d, q) model that will give similar forecasts to the Holt-Winters method, comparing Equation (11) with the eventual forecast function for ARIMA(p, d, q), namely

$$\hat{y}_T(h) = \sum_{k=1}^{p} A_k \alpha_k^h + \sum_{k=0}^{d-1} B_k h^k$$

then we see that we should have $p = 0$ and $d = 2$. The appropriate value of q is suggested by the fact that there are two parameters, λ_0 and λ_1, that appear in the updating equations (12) and (13). It turns out that with an appropriate choice of moving average parameters, θ_1 and θ_2, the ARIMA(0,2,2) model gives exactly the same updating equations as the Holt-Winters approach.