Forecasting with Box-Jenkins Models

1. Minimum Mean Square Error Forecast

In this section we turn to the issue of determining point forecasts and their confidence intervals for some selected ARMA(p,q) and ARIMA(p,d,q) models. Suppose our objective is to minimize the mean square error of forecasting y_t h periods ahead. Let T represent the last time period $t = 1,2\cdots,T$ for which observations on the time series y_t is available and $\hat{y}_{T+h}$ denote the h-step-ahead forecast of y_{T+h}. The **mean square error** of the forecast $\hat{y}_{T+h}$ in predicting y_{T+h} is defined to be

$$\begin{aligned} MSE(\hat{y}_{T+h}) &\equiv E(y_{T+h} - \hat{y}_{T+h})^2 \\ &= E(\hat{y}_{T+h} - E(\hat{y}_{T+h}))^2 + [E(y_{T+h} - \hat{y}_{T+h})]^2 \\ &= Var(\hat{y}_{T+h}) + \{Bias(\hat{y}_{T+h})\}^2 \ . \qquad (1) \end{aligned}$$

That is, the mean square error of the forecast $\hat{y}_{T+h}$ is equal to the sum of the variance of the forecast and the squared bias of the forecast.

Now consider the conditional mean of $y_{T+h} = E(y_{T+h} \mid y_T, y_{T-1,} \cdots, a_T, a_{T-1}, \cdots)$. Denote this conditional mean by m_h. Then any predictor $\hat{y}_{T+h}$ can be represent by $\hat{y}_{T+h} = m_h + d$ where d is the difference between the proposed predictor $\hat{y}_{T+h}$ and m_h . The mean square error of forecast is then

$$\begin{aligned} E(y_{T+h} - \hat{y}_{T+h})^2 &= E(y_{T+h} - m_h - d)^2 \\ &= E(y_{T+h} - m_h)^2 - 2dE(y_{T+h} - m_h) + d^2 \\ &= E(y_{T+h} - m_h)^2 + d^2 \qquad (2) \end{aligned}$$

where $E(y_{T+h} - m_h) = 0$ has been used. Therefore, to minimize the mean square error of forecast we should choose $d = 0$ and the forecast $\hat{y}_{T+h} = m_h$ (the conditional mean of y_{T+h}) to minimize the mean square error of the forecast.

2. Forecasting with the ARMA(0,0) Model

Consider the simpliest Box-Jenkins model

$$y_t = \phi_0 + a_t , \qquad (3)$$

the white noise model ARMA(0,0). Now let us determine the minimum mean square h-step-ahead forecast $E(y_{T+h}\,|\,\bullet)$, where $\bullet = (y_T, y_{T-1}, \cdots, a_T, a_{T-1}, \cdots)$ is the conditioning set for the conditional expectation. Then

$$E(y_{T+h}\,|\,\bullet) = E[(\phi_0 + a_{T+h})\,|\,\bullet] = \phi_0 \quad . \tag{4}$$

Therefore, the h-step-ahead minimum mean square error forecast of y_{T+h} for the ARMA(0,0) model is

$$\hat{y}_{T+h} = \phi_0 , \quad h = 1,2,\cdots . \tag{5}$$

However, the forecast $\hat{y}_{T+h}$ is not feasible (operational) because it is dependent on the unknown intercept ϕ_0 . The intercept can, of course, be consistently estimated by the sample mean $\hat{\phi}_0 = \bar{y} = \sum_{t=1}^{T} y_t / T$. Therefore, an approximate h-step-ahead minimum mean square error forecast is

$$\hat{\hat{y}}_{T+h} = \hat{\phi}_0 = \bar{y} \quad . \tag{6}$$

Of course, as the sample size for the time series goes to infinity $(y_t, t \rightarrow \infty)$, the approximate h-step-ahead minimum mean square error forecast (6) approaches the theoretical h-step-ahead minimum mean square error forecast (5).

3. Forecasting with the AR(1) Model

Now let us turn to prediction in the AR(1) model. First, consider the problem of one-step-ahead forecasting in the AR(1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t \quad . \tag{7}$$

Writing equation (7) for time period $T+1$ we have

$$y_{T+1} = \phi_0 + \phi_1 y_T + a_t \quad . \tag{8}$$

Then

$$\begin{aligned} E(y_{T+1}\,|\,\bullet) &= E[(\phi_0 + \phi_1 y_T + a_{T+1})\,|\,\bullet)] \\ &= \phi_0 + \phi_1 y_T = \hat{y}_{T+1} \end{aligned} \tag{9}$$

is the minimum mean square error one-step-ahead forecast of y_{T+1} .

Similarly,

$$
\begin{aligned}
E(y_{T+2} \mid \bullet) &= E[(\phi_0 + \phi_1 y_{T+1} + a_{T+2}) \mid \bullet)] \\
&= E(\phi_0 \mid \bullet) + \phi_1 E(y_{T+1} \mid \bullet) + E(a_{T+2} \mid \bullet) \\
&= \phi_0 + \phi_1(\phi_0 + \phi_1 y_T) \\
&= \phi_0 + \phi_0\phi_1 + \phi_1^2 y_T = \hat{y}_{T+2} \qquad (10)
\end{aligned}
$$

is the minimum mean square error two-step-ahead forecast. Likewise, the minimum mean error h-step-ahead forecast is derived as

$$
\begin{aligned}
\hat{y}_{T+h} &= \phi_0(1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{h-1}) + \phi_1^h y_T \\
&= \frac{\phi_0}{1-\phi_1}(1 - \phi_1^h) + \phi_1^h y_T \\
&= \frac{\phi_0}{1-\phi_1} + \phi_1^h (y_T - \frac{\phi_0}{1-\phi_1}) \\
&= \mu + \phi_1^h (y_T - \mu) \qquad (11)
\end{aligned}
$$

where recall that the unconditional mean of y is $\mu = \phi_0 /(1 - \phi_1)$.

From (11) we can see that the optimal h-step-ahead forecast in the AR(1) model requires that the mean of y, μ, be "add-factored." The add factor, $\phi_1^h(y_T - \mu)$, is dependent on the position of the last available observation y_T relative to the mean, $y_T - \mu$, and the first order autocorrelation coefficient discounted h periods, ϕ_1^h. Suppose that the time series y_t is positively autocorrelated ($0 < \phi_1 < 1$) and the last available observation y_T is below the mean ($y_T - \mu < 0$). Then the one-step-ahead forecast $\hat{y}_{T+1}$ will be below the mean μ and ϕ_1 of the distance between the last available observation y_T and μ. (See the lead production example examined in exercise 1.) The two-step-ahead forecast will likewise be below the mean but it will only be ϕ_1^2 of the distance between the last available observation and the overall mean of the data. Obviously, as the forecast horizon, h, increases to infinity, the optimal forecast approaches the overall mean of the data because the add-factor, $\phi_1^h(y_T - \mu)$, approaches zero as $h \to \infty$. This is typical behavior for stationary Box-Jenkins models. As $h \to \infty$, the optimal forecast approaches the overall mean in the data. (As we will see in the MA(1) model, this approach of the overall mean is sooner than later.)

Obviously the optimal forecasts of (11) are going to forecast y_{T+h} with error. The mean square error of the one-step-ahead forecast is calculated as

$$\begin{aligned} MSE(\hat{y}_{T+1}) &= E(y_{T+1} - \hat{y}_{T+1})^2 \\ &= E[\phi_0 + \phi_1 y_T + a_{T+1} - (\phi_0 + \phi_1 y_T)]^2 \\ &= E(a_{T+1}^2) = \sigma_a^2 \quad . \end{aligned} \tag{12}$$

By definition then, the standard error of the one-step-ahead forecast is

$$se(\hat{y}_{T+1}) = \sqrt{MSE(\hat{y}_{T+1})} = \sqrt{\sigma_a^2} = \sigma_a \quad . \tag{13}$$

The mean square error of the two-step-ahead forecast is calculated as

$$\begin{aligned} MSE(\hat{y}_{T+2}) &= E(y_{T+2} - \hat{y}_{T+2})^2 \\ &= E[\phi_0 + \phi_1 y_{T+1} + a_{T+2} - (\phi_0 + \phi_1\phi_0 + \phi_1^2 y_T)]^2 \\ &= E\ [\phi_0 + \phi_1(\phi_0 + \phi_1 y_T + a_{T+1}) + a_{T+2} - (\phi_0 + \phi_1\phi_0 + \phi_1^2 y_T)]^2 \\ &= E[\phi_1 a_{T+1} + a_{T+2}]^2 \\ &= E\phi_1^2 a_{T+1}^2 + E(\phi_1 a_{T+1} a_{T+2}) + E a_{T+2}^2 \\ &= \phi_1^2 \sigma_a^2 + \sigma_a^2 = \sigma_a^2(1 + \phi_1^2) \quad . \end{aligned} \tag{14}$$

Similarly, the mean square error of the optimal h-step ahead forecast is

$$\begin{aligned} MSE(\hat{y}_{T+h}) &= E(y_{T+h} - \hat{y}_{T+h})^2 \\ &= E(\phi_1^{h-1} a_{T+1} + \phi_1^{h-2} a_{T+2} + \cdots + a_{T+h})^2 \\ &= \sigma_a^2 (1 + \phi_1^2 + \phi_1^4 + \cdots + \phi_1^{2(h-1)}) \quad . \end{aligned} \tag{15}$$

By definition, the standard error of the h-step-ahead forecasts for the AR(1) model is

$$\begin{aligned} se(\hat{y}_{T+h}) &= \sqrt{MSE(\hat{y}_{T+h})} = \sqrt{\sigma_a^2 (1 + \phi_1^2 + \phi_1^4 + \cdots + \phi_1^{2(h-1)})} \\ &= \sigma_a (1 + \phi_1^2 + \phi_1^4 + \cdots + \phi_1^{2(h-1)})^{1/2} . \end{aligned} \tag{16}$$

Of course, the theoretical minimum mean square error h-step-ahead forecasts (11) are not operational because the formula depends on the unknown parameter values μ and ϕ_1 . They can, of course, be estimated by $\hat{\mu} = \bar{y}$ and $\hat{\phi}_1 = r_1$, the sample mean and the sample first-order autocorrelation coefficient, respectively. Therefore, the **approximate** minimum mean square h-step-ahead forecast for the AR(1) model is

$$\hat{\hat{y}}_{T+h} = \hat{\mu} + \hat{\phi}_1^h (y_T - \hat{\mu}) = \bar{y} + r_1^h (y_T - \bar{y}) \qquad (17)$$

with an **approximate** standard error of forecast of

$$s\hat{e}(\hat{y}_{T+h}) = \hat{\sigma}_a (1 + \hat{\phi}_1^2 + \hat{\phi}_1^4 + \cdots + \hat{\phi}_1^{2(h-1)})^{1/2} \qquad (18)$$

where $\hat{\sigma}_a = \sqrt{\sum_{t=1}^{T} \hat{a}_t^2 \Big/ T}$ is the standard error of the residuals of the AR(1) model and the residuals are defined as $\hat{a}_t = y_t - \hat{y}_t = y_t - (\hat{\phi}_0 + \hat{\phi}_1 y_{t-1})$. See exercise 2 and the calulation of the 12 step ahead forecasts of lead production and their standard errors. You might note that letting $h \to \infty$ in equation (15) implies that the mean square error of the infinite horizon forecast is just

$$MSE(\hat{y}_{T+\infty}) = \frac{\sigma_a^2}{1 - \phi_1^2} \qquad (19)$$

which is the unconditional variance of the time series y_t. That is, the uncertainty in your forecasts can never be greater than the unconditional variance of the series itself and approach this limit as the time horizon of the forecast increases.

4. Forecasting with the MA(1) model

Now let us turn to the derivation of the minimum mean square forecasts for the MA(1) model

$$y_t = \phi_0 + a_t - \theta_1 a_{t-1} \quad . \qquad (20)$$

Considering the time period T + 1, the MA(1) model becomes

$$y_{T+1} = \phi_0 + a_{T+1} - \theta_1 a_T \, . \qquad (21)$$

Taking the conditional expectation of (21) assuming a_T is known we have

$$E(y_{T+1} \mid \bullet) = E[(\phi_0 + a_{T+1} - \theta_1 a_T) \mid \bullet]$$

$$= \phi_0 - \theta_1 a_T = \hat{y}_{T+1} \tag{22}$$

as the minimum mean square error one-step-ahead forecast for the MA(1) model.

The approximate minimum mean square error one-step-ahead forecast then becomes

$$\hat{\hat{y}}_{T+1} = \hat{\phi}_0 - \hat{\theta}_1 \hat{a}_T \tag{23}$$

where $\hat{a}_T$ is the residual at time T and, say, using the method of moments we can estimate ϕ_0 and θ_1 by, respectively, $\hat{\phi}_0 = \bar{y}$ and $\hat{\theta}_1$ so as to satisfy the moment condition

$$r_1 = \frac{-\hat{\theta}_1}{1+\hat{\theta}_1^2} \tag{24}$$

and, at the same time, the invertibility condition $\left|\hat{\theta}_1\right| < 1$. Again, r_1 is the first-order sample autocorrelation coefficient of the time series y_t. The two roots that will satisfy (24) are

$$\hat{\theta}_1 = \frac{-1 \pm \sqrt{1-4r_1^2}}{2r_1} \tag{25}$$

as long as $r_1 \le 1/2$. One then just chooses the root $\hat{\theta}_1$ that satisfies the invertibility condition.

The minimum mean square error two-step-ahead forecast for the MA(1) model is obtained by solving

$$E(y_{T+2} \mid \bullet) = E[(\phi_0 + a_{T+2} - \theta_1 a_{T+1}) \mid \bullet]$$

$$= \phi_0 = \hat{y}_{T+2} \tag{26}$$

with an approximate minimum mean square error forecast of

$$\hat{\hat{y}}_{T+2} = \hat{\phi}_0 = \bar{y}\,. \tag{27}$$

It can easily be shown that the minimum mean square error h-step-ahead forecast for $h \ge 2$ is likewise $\hat{y}_{T+h} = \phi_0$. In summary, the (approximate) minimum mean square error h-step ahead forecasts for the MA(1) model are

$$\hat{\hat{y}} = \begin{cases} \bar{y} - \hat{\theta}_1 \hat{a}_T, h = 1 \\ \bar{y}, h \geq 2 \end{cases} \tag{28}$$

As can be seen from (28), the optimal forecasts for the MA(1) model requires that we add-factor the mean for one period and then we adopt the mean for forecasts two or more periods ahead. Then the forecast profile of the MA(1) model directly reflects the type of memory that the MA(1) model has – a one period memory. That is, recall that the MA(1) model has a nonzero correlation at lag one, namely, $\rho_1 = -\theta_1 /(1+\theta_1^2)$ but $\rho_j = 0$ for $j \geq 2$. The MA(1) model has a one-period memory and, correspondingly, the optimal forecasts for the MA(1) model calls for add-factoring the mean for one period but not thereafter. Note that the one-period add-factor is quite intuitive. If the last available time series observation, y_T, is larger than expected ($\hat{a}_T > 0$) and if the data are positively autocorrelated at one lag ($\hat{\theta}_1 < 0$), then next period's forecast will be above the mean but two and further step-ahead forecasts will adopt the mean as the optimal forecast. Similarly the minimum mean square error forecasts of the MA(2) leads to an add-factoring of the same mean for two periods but, for three or more periods ahead, the sample mean is used. In general the minimum mean square error forecasts for the MA(q) model add-factors the sample mean for q periods-ahead and then thereafter, the sample mean is used.

Looking back over sections 2, 3, and 4 where we derived the minimum mean square forecasts for the ARMA(0,0), AR(1), and MA(1) models, we can see a pattern of add factoring the sample mean according to the type of memory that the data has. In the case of the white noise model, the mean is always used and is never add-factored because white noise data has no memory. When data follow the AR(1) model, the mean is always add-factored but in a diminishing way reflecting the infinite but diminishing memory of the AR(1) process. In the case of the MA(1) model, the data has a one-period memory, therefore, the forecasts add-factor the mean for one-period-ahead forecasting but adopts the mean thereafter. Even though we don't derive the minimum mean square error forecasts for the ARMA(1,1) model, they behave much like the forecasts of the AR(1) model. Given, that the autocorrelation function of the ARMA(1,1) model is diminishing, its memory is infinitely-lived but diminishing. Therefore, it logically follows that the minimum mean square error forecasts of the ARMA(1,1) model add-factors the sample mean but in a diminishing way as the forecast horizon increases and only in the infinite horizon is the sample mean used. In short, the Box-Jenkins methodology generates forecasts that carefully take into account the location of the last observation vis-à-vis the sample mean and the type of memory that characterizes the data. Moreover the forecasts of stationary data eventually (sometimes sooner than later) achieve the mean and the standard errors of the forecasts approach the unconditional variance of the data as the forecast horizon approaches infinity.

5. Forecasting with Integrated ARIMA(p,d,q) models

Recall that for the Dow-Jones data (see the SAS program DOW.sas and Exercise 4) we required that the data be differenced in order to make it stationary. Then how do we forecast with ARIMA(p,d,q) models that require data to be differenced? Let us focus on the case where the data y_t needs to be differenced only once (d=1) before it becomes stationary. That is, the transformed series $\Delta y_t = y_t - y_{t-1}$ is assumed to have a constant mean, constant variance, and constant covariance for each of the lags j = 1, 2, … . In general, let us assume that we have the optimal h-step-ahead forecasts of the differences in y_t, namely $\hat{\hat{\Delta}} y_{T+h}$, available to us. Given the last available observation, y_T, we can construct the minimum mean square error forecasts of the original (level) data as follows:

$$\begin{aligned} &\hat{\hat{y}}_{T+1} = y_T + \hat{\hat{\Delta}} y_{T+1}\,; \\ &\hat{\hat{y}}_{T+2} = \hat{\hat{y}}_{T+1} + \hat{\hat{\Delta}} y_{T+2} = y_T + \hat{\hat{\Delta}} y_{T+1} + \hat{\hat{\Delta}} y_{T+2}\,; \\ &\qquad \text{etc.} \end{aligned} \tag{29}$$

For example, we saw that the Dow-Jones data followed an ARIMA(0,1,1) Box-Jenkins model. Equivalently, the differences of the data follow an MA(1) model. Then in generating the forecasts of the Dow-Jones Index we first need to generate the forecasts of the differences Δy_{T+h} and then integrate them into the last available observation, as in (29) to obtain the forecasts of the original data. Before doing that, let us first generate the optimal forecasts for the ARIMA(0,1,0), i.e. random walk, model.

The random walk model with drift (ARIMA(0,1,0)) is written as

$$y_t = \phi_0 + y_{t-1} + a_t \ . \tag{30}$$

The drift parameter is ϕ_0. If $\phi_0 = 0$, the data is "flat" and is neither drifting up or down. If $\phi_0 > 0$, the data is drifting upward. If $\phi_0 < 0$, the data is drifting downward. In differenced form (30) is written as

$$\Delta y_t = \phi_0 + a_t \ . \tag{31}$$

That is, the differenced data, Δy_t, follows a white noise model (ARMA(0,0)). It follows that the approximate minimum mean square forecasts for Δy_{T+h} are

$$\hat{\hat{\Delta}} y_{T+h} = \overline{\Delta y} \quad \text{for } h \geq 1 \tag{32}$$

where $\overline{\Delta y} = \sum_{t=2}^{T} (\Delta y_t /(T-1))$ is the sample mean of the differenced data, Δy_t .

Therefore, the approximate minimum mean square forecasts for y_{T+h} are

$$\hat{\hat{y}}_{T+1} = y_T + \hat{\hat{\Delta}} y_{T+1} = y_T + \overline{\Delta y} \ ;$$

$$\hat{\hat{y}}_{T+2} = \hat{\hat{y}}_{T+1} + \hat{\hat{\Delta}} y_{T+2} = y_T + \overline{\Delta y} + \overline{\Delta y} = y_T + 2\overline{\Delta y} \, ;$$

and, in general,

$$\hat{\hat{y}}_{T+h} = y_T + h\overline{\Delta y} \ , \quad h \geq 1 \qquad . \tag{33}$$

Thus, in the ARIMA(0,1,0) model the optimal forecasts begin with the last available observation, y_T, and move in lock step in increments of $\overline{\Delta y}$ which is the “average drift” in the data.

Now consider forecasting with the ARIMA(1,1,0) model or, equivalently, an ARMA(1,0) model in the differences Δy_t. The optimal forecasts of the AR(1) model imply that

$$\hat{\hat{\Delta}} y_{T+h} = \overline{\Delta y} + \hat{\phi}_1^h (\Delta y_T - \overline{\Delta y}) \ , \quad h \geq 1 \tag{34}$$

where now $\hat{\phi}_1$ represents the estimated first-order autocorrelation coefficient for the AR(1) model of the $\Delta y's$. It follows that the level forecasts are

$$\hat{\hat{y}}_{T+1} = y_T + \hat{\hat{\Delta}} y_{T+1} = y_T + \overline{\Delta y} + \hat{\phi}_1 (\Delta y_T - \overline{\Delta y}) \ ;$$

$$\hat{\hat{y}}_{T+2} = \hat{\hat{y}}_{T+1} + \hat{\hat{\Delta}} y_{T+2} = y_T + 2\overline{\Delta y} + (\Delta y_T - \overline{\Delta y})(\hat{\phi}_1 + \hat{\phi}_1^2) \, ;$$

and, in general,

$$\begin{aligned} \hat{\hat{y}}_{T+h} &= y_T + h\overline{\Delta y} + (\Delta y_T - \overline{\Delta y})(\hat{\phi}_1 + \hat{\phi}_1^2 + \cdots + \hat{\phi}_1^h) \\ &= y_T + h\overline{\Delta y} + (\Delta y_T - \overline{\Delta y}) \frac{\hat{\phi}_1}{1-\hat{\phi}_1} (1 - \hat{\phi}_1^h) \ . \end{aligned} \tag{35}$$

Compare the h-step-ahead forecasts of the ARIMA(0,1,0) model, (33), with the h-step-ahead forecasts of the ARIMA(1,1,0) model, (35). The ARIMA(1,1,0) forecasts are not equal to the trend line forecasts $y_T + h\overline{\Delta y}$ of the ARIMA(0,1,0) model. The basic trend $y_T + h\overline{\Delta y}$ is add-factored by the amount $(\Delta y_T - \overline{\Delta y})(\hat{\phi}_1 + \hat{\phi}_1^2 + \cdots + \hat{\phi}_1^h)$. As the forecast horizon approaches infinity ($h \to \infty$), the forecasts of the ARIMA(1,1,0) model approaches the trend line $y_T + h\overline{\Delta y} + (\Delta y_T - \overline{\Delta y})\dfrac{\hat{\phi}_1}{1-\hat{\phi}_1}$.

Finally, consider forecasting with the ARIMA(0,1,1) model or, equivalently, an ARMA(0,1) model in the differences Δy_t. The optimal forecasts of the MA(1) model imply that

$$\hat{\hat{\Delta}} y_{T+h} = \begin{cases} \overline{\Delta y} - \hat{\theta}_1 \hat{a}_T, h = 1 \\ \overline{\Delta y}, h \geq 2 \end{cases} \tag{36}$$

where $\hat{a}_T$ now represents the T-th residual of the MA(1) model of the $\Delta y's$ and $\hat{\theta}_1$ is the estimated MA(1) coefficient of the same model. It follows that the level forecasts are

$$\hat{\hat{y}}_{T+1} = y_T + \hat{\hat{\Delta}} y_{T+1} = y_T + \overline{\Delta y} - \hat{\theta}_1 \hat{a}_T \;;$$

$$\hat{\hat{y}}_{T+2} = \hat{\hat{y}}_{T+1} + \hat{\hat{\Delta}} y_{T+2} = y_T + 2\overline{\Delta y} - \hat{\theta}_1 \hat{a}_T \,;$$

and, in general,

$$\hat{\hat{y}}_{T+h} = y_T + h\overline{\Delta y} - \hat{\theta}_1 \hat{a}_T \quad . \tag{37}$$

Compare the h-step-ahead forecasts of the ARIMA(0,1,0) model, (33), with the h-step-ahead forecasts of the ARIMA(0,1,1), (37). The ARIMA(0,1,1) forecasts are not equal to the trend line forecasts $y_T + h\overline{\Delta y}$ of the ARIMA(0,1,0) model. The basic trend $y_T + h\overline{\Delta y}$ is add-factored once by the amount $-\hat{\theta}_1 \hat{a}_T$ and then changes by the "average drift", $\overline{\Delta y}$, thereafter. After the first forecast, the forecasts follow the trend line $y_T + h\overline{\Delta y} - \hat{\theta}_1 \hat{a}_T$. (See the level forecasts of the Dow-Jones Index analyzed in Exercise 4.) For the ARIMA(0,1,1) model, the basic trend line $y_T + h\overline{\Delta y}$ is add-factored once (reflecting the one period memory of the $\Delta y's$ in the MA(1) model) while, for the ARIMA(1,1,0) model, the trend line $y_T + h\overline{\Delta y}$ is continually add-factored but in a diminishing way (reflecting the infinite but diminishing memory of the $\Delta y's$ in the AR(1) model).