Minimum Mean Square Error Forecasts

Consider the forecasting, or prediction, of z_{t+l} given $z_t, z_{t-1}, \ldots,$ assuming $\{z_t\}$ follows some *known* ARIMA(p,d,q) model.

Let $\tilde{z}_t(l)$ be any function of $Z_t = (z_t, z_{t-1}, \ldots)$. The mean square error of forecasting z_{t+l} by $\tilde{z}_t(l)$ is seen to satisfy

$$\begin{aligned} E[z_{t+l} - \tilde{z}_t(l)]^2 &= E[z_{t+l} - E[z_{t+l}|Z_t]]^2 + E[E[z_{t+l}|Z_t] - \tilde{z}_t(l)]^2 \\ &\geq E[z_{t+l} - E[z_{t+l}|Z_t]]^2. \end{aligned}$$

Hence, $E[z_{t+l}|Z_t]$ is the minimum mean square error forecast of z_{t+l} given Z_t, to be denoted by $\hat{z}_t(l)$.

Recall the truncated MA form $z_{t+l} = I_t(l) + C_t(l)$, where $I_t(l) = \sum_{j=0}^{l-1} \psi_j a_{t+l-j}$ and $C_t(l)$ is the complimentary function at origin t. It is easily seen that $\hat{z}_t(l) = C_t(l)$. $I_t(l)$ is the forecasting error, to be denoted by $e_t(l)$.

Other Forms of Forecasts, Updating

Consider the difference equation form of the model

$$z_{t+l} = \sum_{j=1}^{p+d} \varphi_j z_{t+l-j} + a_{t+l} - \sum_{j=1}^{q} \theta_j a_{t+l-j}.$$

Taking conditional expectations at time t, one has

$$\hat{z}_t(l) = \sum_{j=1}^{p+d} \varphi_j \hat{z}_t(l-j) - \sum_{j=l}^{q} \theta_j a_{t+l-j},$$

where $\hat{z}_t(k) = z_{t+k}$ for $k \leq 0$.

Based on the AR form of the model, $z_{t+l} = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} + a_{t+l}$, one has $\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j \hat{z}_t(l-j)$, where $\hat{z}_t(k) = z_{t+k}$ for $k \leq 0$.

From $z_{t+l} = \sum_{j=0}^{l-1} \psi_j a_{t+l-j} + \hat{z}_t(l)$, it is easy to show that

$$\hat{z}_{t+1}(l-1) = \hat{z}_t(l) + \psi_{l-1} a_{t+1}.$$

As soon as z_{t+1} becomes available, one may calculate $a_{t+1} = z_{t+1} - \hat{z}_t(1)$ and update the forecast of z_{t+l} by $\hat{z}_{t+1}(l-1)$.

Examples

Consider an ARI(1,1,0) model $(1-1.8B+.8B^2)z_t = a_t$. One has

$$\hat{z}_t(1) = 1.8z_t - .8z_{t-1},$$

$$\hat{z}_t(2) = 1.8\hat{z}_t(1) - .8z_t,$$

$$\hat{z}_t(l) = 1.8\hat{z}_t(l-1) - .8\hat{z}_t(l-2), \quad l > 2.$$

The ψ weights for updating are given by $\psi_j = 1.8\psi_{j-1} - .8\psi_{j-2}$, $j > 0$, with $\psi_0 = 1$, $\psi_{-1} = 0$.

Consider an IMA(0,2,2) model $\nabla^2 z_t = (1-.9B+.5B^2)a_t$. One has

$$\hat{z}_t(1) = 2z_t - 1z_{t-1} - .9a_t + .5a_{t-1},$$

$$\hat{z}_t(2) = 2\hat{z}_t(1) - z_t + .5a_t,$$

$$\hat{z}_t(l) = 2\hat{z}_t(l-1) - \hat{z}_t(l-2), \quad l > 2.$$

The ψ weights are given by $\psi_j = 2\psi_{j-1} - \psi_{j-2}$, $j > 2$, with $\psi_0 = 1$, $\psi_1 = 2 - .9 = 1.1$, $\psi_2 = 2(1.1) - 1 + .5 = 1.7$.

Probability Limits of Forecasts

The forecasting error $e_t(l) = \sum_{j=0}^{l-1} \psi_j a_{t+l-j}$ has variance

$$V(l) = (1 + \sum_{j=1}^{l-1} \psi_j^2)\sigma_a^2,$$

which naturally increases with l. The formula can be used to calculate "prediction intervals" for z_{t+l},

$$\hat{z}_t(l) \pm 1.96\sqrt{(1 + \sum_{j=1}^{l-1} \psi_j^2)}\sigma_a.$$

For the examples above, $(1 - 1.8B + .8B^2)z_t = a_t$ (model A) and $\nabla^2 z_t = (1 - .9B + .5B^2)a_t$ (model B), $\sqrt{V(l)/\sigma_a^2}$ at $l = 1, \ldots, 6$ are calculated and listed below.

l	1	2	3	4	5	6
A	1	2.06	3.19	4.35	5.50	6.62
B	1	1.49	2.26	3.22	4.34	5.57

Forecast Function and Weights

For $l > q$, $\hat{z}_t(l) = \sum_{j=1}^{p+d} \varphi_j \hat{z}_t(l-j)$, so the "eventual" forecasting function satisfies the equation $\varphi(B)\hat{z}_t(l) = 0$, hence are of the form

$$\hat{z}_t(l) = b_0^{(t)} f_0(l) + \cdots + b_{p+d-1}^{(t)} f_{p+d-1}(l),$$

where $f_j(l)$ are determined by the roots of $\varphi(B)$ and $b_j^{(t)}$ by the initial values. Recall the form of complimentary function $C_t(l)$.

From $\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j \hat{z}_t(l-j)$, one may express $\hat{z}_t(l)$ directly in terms of $z_t, z_{t-1}, \ldots$, $\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} z_{t-j+1}$.

$$\begin{aligned}
\hat{z}_t(1) &= \pi_1 z_t + \pi_2 z_{t-1} + \pi_3 z_{t-2} + \cdots \\
\hat{z}_t(2) &= \pi_1 \hat{z}_t(1) + \pi_2 z_t + \pi_3 z_{t-1} + \cdots \\
&= (\pi_1\pi_1 + \pi_2) z_t + (\pi_1\pi_2 + \pi_3) z_{t-1} + \cdots \\
&\cdots\cdots
\end{aligned}$$

General Form of Forecast Weights

From $z_{t+l} = \sum_{k=0}^{l-1} \psi_k a_{t+l-k} + \hat{z}_t(l)$, one has

$$\begin{aligned}\hat{z}_t(l) &= \hat{z}_{t+l-1}(1) - \textstyle\sum_{k=1}^{l-1} \psi_k a_{t-l-k} \\ &= \pi_1 z_{t+l-1} + \pi_2 z_{t+l-2} + \cdots + \pi_{l-1} z_{t+1} + \pi_l z_t + \pi_{l+1} z_{t-1} + \cdots \\ &\quad + \psi_1(-z_{t+l-1} + \pi_1 z_{t+l-2} + \cdots + \pi_{l-2} z_{t+1} + \pi_{l-1} z_t + \pi_l z_{t-1} + \cdots) \\ &\quad + \cdots + \psi_{l-1}(-z_{t+1} + \pi_1 z_t + \pi_2 z_{t-1} + \cdots).\end{aligned}$$

Adding up the coefficients of $z_t, z_{t-1}, \ldots$, one has

$$\pi_j^{(l)} = \pi_{l+j-1} + \psi_1 \pi_{l+j-2} + \cdots + \psi_{l-1}\pi_j = \pi_{j+1}^{(l-1)} + \psi_{l-1}\pi_j.$$

For example, $\pi_j^{(2)} = \pi_{j+1} + \psi_1 \pi_j$,

$$\pi_j^{(3)} = \pi_{j+2} + \psi_1 \pi_{j+1} + \psi_2 \pi_j = \pi_{j+1}^{(2)} + \psi_2 \pi_j.$$

The coefficients of $z_{t+l-1}, \ldots, z_{t+1}$ vanish as $\sum_j \pi_j \psi_{k-j} = 0$ for $k > 0$, where $\psi_0 = -\pi_0 = 1$, $\psi_j = \pi_j = 0$, $j < 0$.

Example: IMA(0,1,1)

Consider the model $\nabla z_t = a_t - \theta a_{t-1}$. One has

$$\hat{z}_t(1) = z_t - \theta a_t, \qquad \hat{z}_t(l) = \hat{z}_t(l-1) = \hat{z}_t(1), \quad l > 1,$$

which give a constant forecast function. Since $\psi_j = 1 - \theta$, $j > 0$, the forecast function can be updated through

$$\hat{z}_{t+1}(l) = \hat{z}_{t+1}(l-1) = \hat{z}_t(l) + (1-\theta)a_{t+1}, \quad l > 1.$$

The π weights are $\pi_j = (1-\theta)\theta^{j-1}$. Note that

$$\pi_j^{(2)} = \pi_{j+1} + \psi_1 \pi_j = (1-\theta)\theta^j + (1-\theta)^2\theta^{j-1} = \pi_j,$$

so there is no surprise here. The calculation applies recursively to $\pi_j^{(3)}, \pi_j^{(4)}, \ldots$. The variance of $e_t(l)$ is easily seen to be

$$V(l) = \sigma_a^2(1 + (l-1)(1-\theta)^2).$$

Example: IMA(0,2,2)

Consider the model $\nabla^2 z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$. One has

$$\hat{z}_t(1) = 2z_t - 1z_{t-1} - \theta_1 a_t - \theta_2 a_{t-1},$$
$$\hat{z}_t(2) = 2\hat{z}_t(1) - z_t - \theta_2 a_t,$$
$$\hat{z}_t(l) = 2\hat{z}_t(l-1) - \hat{z}_t(l-2), \quad l > 2.$$

Since $\psi_j = \lambda_0 + \lambda_1 j$, where $\lambda_0 = 1 + \theta_2$ and $\lambda_1 = 1 - \theta_1 - \theta_2$,

$$\hat{z}_{t+1}(l-1) = \hat{z}_t(l) + (\lambda_0 + \lambda_1(l-1))a_{t+1}, \quad l > 1.$$

In the form of the complimentary function, $\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l$, and

$$b_0^{(t+1)} = b_0^{(t)} + b_1^{(t)} + \lambda_0 a_{t+1}, \quad b_1^{(t+1)} = b_1^{(t)} + \lambda_1 a_{t+1}.$$

The variance of $e_t(l)$ is given by

$$V(l) = \sigma_a^2(1 + (l-1)\lambda_0^2 + \tfrac{1}{6}l(l-1)(2l-1)\lambda_1^2 + \lambda_0\lambda_1 l(l-1)).$$

Examples: AR(1) and ARI(1,1,0)

Consider the model $z_t = \phi z_{t-1} + a_t$. One has $\hat{z}_t(l) = z_t \phi^l$. Since $\psi_j = \phi^j$, the variance of $e_t(l)$ is given by

$$V(l) = \sigma_a^2 \sum_{j=0}^{l-1} \phi^{2j} = \sigma_a^2 (1 - \phi^{2l})/(1 - \phi^2).$$

Consider the model $(1 - \phi B)(1 - B) z_t = a_t$. One has

$$\hat{z}_t(l) - \hat{z}_t(l-1) = \phi^l (z_t - z_{t-1}).$$

It follows that $\hat{z}_t(l) = z_t + (\sum_{j=1}^{l} \phi^j)(z_t - z_{t-1})$, $l > 0$, or

$$\hat{z}_t(l) = z_t + (z_t - z_{t-1}) \phi (1 - \phi^l)/(1 - \phi),$$

which "converges" to $z_t + (z_t - z_{t-1})\phi/(1 - \phi)$. It can be shown that $\psi_j = (1 - \phi^{j+1})/(1 - \phi)$, so the variance of $e_t(l)$ is given by

$$V(l) = \frac{\sigma_a^2}{(1-\phi)^2} \{ l + \frac{\phi^2 (1 - \phi^{2l})}{1 - \phi^2} - 2 \frac{\phi(1 - \phi^l)}{1 - \phi} \}.$$

Example: ARMA(1,1)

Consider the model $z_t = \phi z_{t-1} + a_t - \theta a_{t-1}$. One has

$$\hat{z}_t(1) = \phi z_t - \theta a_t,$$

$$\hat{z}_t(l) = \phi \hat{z}_t(l-1) = \phi^{l-1} \hat{z}_t(1), \quad l > 1.$$

Since $\psi_j = (\phi - \theta)\phi^{j-1}$, $j > 0$, so for $l > 1$,

$$\hat{z}_{t+1}(l-1) = \hat{z}_t(l) + (\phi - \theta)\phi^{l-2} a_{t+1}.$$

In particular, one has the updating formula

$$\hat{z}_{t+1}(1) = \hat{z}_t(2) + (\phi - \theta) a_{t+1} = \phi \hat{z}_t(1) + (\phi - \theta) a_{t+1}.$$

The π weights are $\pi_j = (\phi - \theta)\theta^{j-1}$, so $\pi_j^{(l)} = \phi^{l-1}(\phi - \theta)\theta^{j-1}$. The variance of $e_t(l)$ is seen to be

$$V(l) = \sigma_a^2 \{1 + (\phi - \theta)^2 (1 - \phi^{2(l-1)})/(1 - \phi^2)\}.$$

Example: ARIMA(1,1,1)

Consider the model $(1-\phi B)(1-B)z_t = a_t - \theta a_{t-1}$. One has

$$\hat{z}_t(1) = (1+\phi)z_t - \phi z_{t-1} - \theta a_t,$$

$$\hat{z}_t(l) = (1+\phi)\hat{z}_t(l-1) - \phi\hat{z}_t(l-2), \quad l > 1.$$

Since $(1-\phi B)(1-B)\hat{z}_t(l) = 0$, $l > 1$, $\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)}\phi^l$, where

$$b_0^{(t)} = (z_t - \phi z_{t-1} - \theta a_t)/(1-\phi) = z_t - b_1^{(t)},$$

$$b_1^{(t)} = (\theta a_t - \phi(z_t - z_{t-1}))/(1-\phi),$$

yielding $\hat{z}_t(l) = z_t + \phi\frac{1-\phi^l}{1-\phi}(z_t - z_{t-1}) - \theta\frac{1-\phi^l}{1-\phi}a_t$. As $a_t = z_t - \sum_{j=1}^{\infty}\pi_j z_{t-j}$, where $\pi_1 = 1+\phi-\theta$, $\pi_j = (1-\theta)(\theta-\phi)\theta^{j-2}$, $j > 1$, some algebra yields

$$\hat{z}_t(l) = (1-\alpha_l)z_t + \alpha_l\{(1-\theta)\textstyle\sum_{j=1}^{\infty}\theta^{j-1}z_{t-j}\},$$

where $\alpha_l = (\theta-\phi)(1-\phi^l)/(1-\phi)$. Now $\psi_j = [(1-\theta) + \phi^l(\theta-\phi)]/(1-\phi)$, so

$$V(l) = \frac{\sigma_a^2}{(1-\phi)^2}\{l(1-\theta)^2 + (\theta-\phi)^2\frac{1-\phi^{2l}}{1-\phi^2} + 2(1-\theta)(\theta-\phi)\frac{1-\phi^l}{1-\phi}\}.$$

Forecasting with Finite Samples

With finite samples $(z_t, z_{t-1}, \ldots, z_1)$, the procedure developed above works without a problem for $q = 0$. For $q > 0$, however, $a_t, \ldots, a_{t-q+1}$ appearing in $\hat{z}_t(l)$ also depend on $z_{-1}, z_{-2}, \ldots$, so modifications are needed.

For invertible models, the π weights decay exponentially, so it is reasonable to set $z_{-1} = z_{-2} = \cdots = 0$ when calculating a_k from $z_k, z_{k-1}, \ldots$.

Using the innovations algorithm, one can calculate the exact one step forecast $\hat{z}_{t+1} = E[z_{t+1}|z_t, \ldots, z_1]$ with error variance v_t. The exact multiple step forecast with finite samples will be discussed along with the state space models.