김은영

제5장 ARMA모형의 예측

5.1 최소평균제곱오차예측값

■ 시점 n가지의 시계열 관측값이 있을 때 이후 시점의 시계열 값을 예측하는 문제

■ 예측값

k단계 이후 예측값(k-step ahead forecast): $f_{n,k}$ $(k=1,\ 2,\ \cdots)$

· $k=1$: $f_{n,1}$ [일단계 이후 예측값(one-step ahead forecast)]

■ $f_{n,k}$를 과거 시계열 관측값들의 선형결합으로 예측한다고 하자.

$$f_{n,k}=b_k Z_n+b_{k+1}Z_{n-1}+\cdots \quad \text{여기서, } b_k,\ b_{k+1},\ \cdots : \text{상수}$$

■ ARMA모형의 MA 표현방식

$$Z_t=a_t-\psi_1 a_{t-1}-\psi_2 a_{t-2}-\cdots=\sum_{j=0}^{\infty}\psi_j a_{t-j} \qquad \text{단, } \psi_0=1$$

$$f_{n,k}=c_k a_n+c_{k+1}a_{n-1}+\cdots=\sum_{j=k}^{\infty}c_j a_{n+k-j} \quad \text{여기서, } c_k,\ c_{k+1},\ \cdots : \text{추정해야 할 상수}$$

■ 예측 관련 평균제곱오차(mean squared error)

$$Q = E[(Z_{n+k} - f_{n,k})^2] = E\left[\left(\sum_{j=0}^{\infty}\psi_j a_{n+k-j} - \sum_{j=k}^{\infty}c_j a_{n+k-j}\right)^2\right]$$

$$Q = Var\left[\sum_{j=0}^{\infty}\psi_j a_{n+k-j} - \sum_{j=k}^{\infty}c_j a_{n+k-j}\right] + E^2\left[\sum_{j=0}^{\infty}\psi_j a_{n+k-j} - \sum_{j=k}^{\infty}c_j a_{n+k-j}\right]$$

$$Q = Var\left[\sum_{j=0}^{\infty}\psi_j a_{n+k-j}\right] + Var\left[\sum_{j=k}^{\infty}c_j a_{n+k-j}\right] - 2Cov\left[\sum_{j=0}^{\infty}\psi_j a_{n+k-j}, \sum_{j=k}^{\infty}c_j a_{n+k-j}\right]$$

여기서,

$$Var\left[\sum_{j=0}^{\infty}\psi_j a_{n+k-j}\right] = \sigma_a^2\sum_{j=0}^{\infty}\psi_j^2$$

$$Var\left[\sum_{j=k}^{\infty}c_j a_{n+k-j}\right] = \sigma_a^2\sum_{j=k}^{\infty}c_j^2$$

$$Cov\left[\sum_{j=0}^{\infty}\psi_j a_{n+k-j}, \sum_{j=k}^{\infty}c_j a_{n+k-j}\right] = \sigma_a^2\sum_{j=k}^{\infty}\psi_j c_j$$

$$\rightarrow \quad Q = \sigma_a^2\sum_{j=0}^{k-1}\psi_j^2 + \sigma_a^2\sum_{j=k}^{\infty}(\psi_j - c_j)^2 = \sigma_a^2\sum_{j=0}^{k-1}\psi_j^2 + \sigma_a^2\sum_{j=0}^{\infty}(\psi_{j+k} - c_{j+k})^2$$

· 위 식을 최소로 하는 상수 c_{j+k}의 추정치: $\hat{c}_{j+k} = \psi_{j+k}$

$$\Rightarrow f_{n,k} = \psi_k a_n + \psi_{k+1}a_{n-1} + \cdots = \sum_{j=k}^{\infty}\psi_j a_{n+k-j}$$

■ $Z_{n+k} = \sum_{j=0}^{\infty}\psi_j a_{n+k-j} = a_{n+k} + \psi_1 a_{n+k-1} + \cdots + \psi_{k-1}a_{n+1} + \sum_{j=k}^{\infty}\psi_j a_{n+k-j}$

다음 조건부 기댓값이 $f_{n,k}$와 동일하다.

$$E[Z_{n+k} | Z_n, Z_{n-1}, \cdots] = \psi_k a_n + \psi_{k+1}a_{n-1} + \cdots = \sum_{j=k}^{\infty}\psi_j a_{n+k-j}$$

※ 과거 시계열이 주어진 경우 과거 오차항은 실현된 것이기 때문에 상수로 간주한다.

$$f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \cdots], \ k = 1, 2, \cdots$$

■ k단계 이후 예측값에 대한 예측오차(forecast error)

$$e_{n,k} = Z_{n+k} - f_{n,k} = a_{n+k} + \psi_1 a_{n+k-1} + \cdots + \psi_{k-1} a_{n+1}$$

여기서, $Z_{n+k} = \sum_{j=0}^{\infty} \psi_j a_{n+k-j} = a_{n+k} + \psi_1 a_{n+k-1} + \cdots + \psi_{k-1} a_{n+1} + \sum_{j=k}^{\infty} \psi_j a_{n+k-j}$

$$f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \cdots] = \psi_k a_n + \psi_{k+1} a_{n-1} + \cdots = \sum_{j=k}^{\infty} \psi_j a_{n+k-j}$$

■ k단계 이후 예측값에 대한 예측오차(forecast error)의 분산

$$v_{n,k} = Var(e_{n,k}) = \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2 = Var[Z_{n+k} | Z_n, Z_{n-1}, \cdots]$$

· 이를 바탕으로 예측구간(prediction interval)을 구할 수 있다.

cf) 일단계 이후 예측오차: $e_{n,1} = Z_{n+1} - f_{n,1} = a_{n+1}$

일단계 이후 예측오차의 분산: $v_{n,1} = Var(e_{n,1}) = Var(a_{n+1}) = \sigma_a^2$

5.2 ARMA모형의 예측

5.2.1 AR모형의 예측

■ AR(p)모형

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t \ , \qquad a_t \sim N(0,\ \sigma_a^2)$$

· 시점 n까지의 관측값$(z_1,\ z_2,\ \cdots,\ z_n)$이 주어질 때, 일단계 이후 예측값은

$$\begin{aligned} f_{n,1} &= E[Z_{n+1} | Z_n, Z_{n-1}, \cdots] = E[\phi_1 Z_n + \cdots + \phi_p Z_{n+1-p} + a_{n+1} | Z_n, Z_{n-1}, \cdots] \\ &= \phi_1 Z_n + \cdots + \phi_p Z_{n+1-p} \end{aligned}$$

· 일단계 이후 예측오차의 분산

$$\begin{aligned} v_{n,1} = Var(e_{n,1}) &= Var[Z_{n+1} | Z_n, Z_{n-1}, \cdots] \\ &= Var[\phi_1 Z_n + \cdots + \phi_p Z_{n+1-p} + a_{n+1} | Z_n, Z_{n-1}, \cdots] \\ &= Var(a_{n+1}) = \sigma_a^2 \end{aligned}$$

· 시점 n까지의 관측값$(z_1,\ z_2,\ \cdots,\ z_n)$이 주어질 때, 이단계 이후 예측값은

$$f_{n,2} = E[Z_{n+2} | Z_n, Z_{n-1}, \cdots] = E[\phi_1 Z_{n+1} + \cdots + \phi_p Z_{n+2-p} + a_{n+2} | Z_n, Z_{n-1}, \cdots]$$

여기서, $E(Z_j | Z_n,\ \cdots) = \begin{cases} f_{n,j-n}, & j \ge n+1 \\ Z_j \quad , & j \le n \end{cases}$

· 시점 n까지의 관측값$(z_1,\ z_2,\ \cdots,\ z_n)$이 주어질 때, k단계 이후 예측값은

$$f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \cdots] = E[\phi_1 Z_{n+k-1} + \cdots + \phi_p Z_{n+k-p} + a_{n+k} | Z_n, Z_{n-1}, \cdots]$$
$$= \phi_1 f_{n,1} + \phi_2 Z_n + \cdots + \phi_p Z_{n+1-p}$$

· k단계 이후 예측오차의 분산

$$\upsilon_{n,k} = Var(e_{n,k}) = Var[Z_{n+k} | Z_n, Z_{n-1}, \cdots]$$
$$= Var[\phi_1 Z_{n+k-1} + \cdots + \phi_p Z_{n+k-p} + a_{n+k} | Z_n, Z_{n-1}, \cdots]$$
$$= \sigma_a^2 + \sum_{i=1}^{p} \phi_i^2 Var[Z_{n+k-i} | Z_n,\ \cdots] + 2 \sum_{1 \le i < j \le p} \phi_i \phi_j Cov[Z_{n+k-i},\ Z_{n+k-j} | Z_n,\ \cdots]$$

여기서, $Var[Z_{n+k-i} | Z_n,\ \cdots] = \begin{cases} \upsilon_{n,k-i}, & k-i \ge 1 \\ 0, & k-i \le 0 \end{cases}$

5.2.2 MA모형의 예측

■ MA(1)모형

$$Z_t = a_t - \theta_1 a_{t-1}$$

· k단계 이후 예측값

$$f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \cdots] = E[a_{n+k} - \theta_1 a_{n+k-1} | Z_n, Z_{n-1}, \cdots]$$
$$= \begin{cases} -\theta_1 a_n, & k = 1 \\ 0, & k \ge 2 \end{cases}$$

→ a_n: 실현된 것. 상수로 간주. 과거 시계열 값으로 추정할 수 있다

$$a_n = (1 - \theta_1 B)^{-1} Z_n = \sum_{j=0}^{\infty} (\theta_1)^j Z_{n-j}$$

⇒ 실제로는 다음 관계를 이용하여 예측오차[잔차]로 추정하는 것이 편리하다

$$\hat{a}_n = e_{n-1,1} = Z_n - f_{n-1,1}$$

· 일단계 이후 예측오차의 분산

$$\upsilon_{n,1} = Var(e_{n,1}) = Var[Z_{n+1} | Z_n, Z_{n-1}, \cdots]$$
$$= Var[a_{n+1} - \theta_1 a_n | Z_n, Z_{n-1}, \cdots]$$
$$= Var(a_{n+1}) = \sigma_a^2$$

· k단계 이후 예측오차의 분산

$$v_{n,k} = Var(e_{n,k}) = Var[Z_{n+k} | Z_n, Z_{n-1}, \cdots]$$

$$= Var[a_{n+k} - \theta_1 a_{n+k-1} | Z_n, Z_{n-1}, \cdots] = (1+\theta_1^2)\sigma_a^2,\ k \ge 2$$

■ MA(q)모형

$$Z_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

· MA(q)모형의 일단계 이후 예측값

$$f_{n,1} = E[Z_{n+1} | Z_n, Z_{n-1}, \cdots] = E[a_{n+1} - \theta_1 a_n - \cdots - \theta_q a_{n+1-q} | Z_n, Z_{n-1}, \cdots]$$

$$= -\theta_1 a_n - \cdots - \theta_q a_{n+1-q}$$

· MA(q)모형의 k단계 이후 예측값

$$\begin{aligned} f_{n,k} &= E[Z_{n+k} | Z_n, Z_{n-1}, \cdots] \\ &= E[a_{n+k} - \theta_1 a_{n+k-1} - \cdots - \theta_q a_{n+k-q} | Z_n, Z_{n-1}, \cdots] \\ &= E[-\theta_1 a_{n+k-1} - \cdots - \theta_q a_{n+k-q} | Z_n, Z_{n-1}, \cdots] \end{aligned}$$

여기서, 오차항에 대한 조건부 기댓값은

$$E[a_j | Z_n,\ Z_{n-1},\ \cdots] = \begin{cases} 0\ , & j \ge n+1 \\ a_j, & j \le n \end{cases}$$

· MA(q)모형에서 k단계 이후 예측에 대한 예측오차분산

$$v_{n,k} = Var(e_{n,k}) = Var[a_{n+k} - \theta_1 a_{n+k-1} - \cdots - \theta_q a_{n+k-q} | Z_n, Z_{n-1}, \cdots]$$

$$= \sigma_a^2 + \sum_{i=1}^{q} \theta_i^2 Var[a_{n+k-i} | Z_n,\ \cdots]$$

여기서, $Var[a_{n+k-i} | Z_n,\ \cdots] = \begin{cases} \sigma_a^2, & k-i \ge 1 \\ 0\ , & k-i \le 0 \end{cases}$

5.2.3 ARMA모형의 예측

■ ARMA(1,1)모형

$$Z_t = \phi Z_{t-1} + a_t - \theta a_{t-1}\ , \qquad a_t \sim N(0,\ \sigma_a^2)$$

· 일단계 이후 예측값

$$f_{n,1} = E[Z_{n+1} | Z_n, Z_{n-1}, \cdots] = E[\phi Z_n + a_{n+1} - \theta a_n | Z_n, Z_{n-1}, \cdots] = \phi Z_n - \theta a_n$$

· k단계 이후 예측값

$$f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \cdots] = E[\phi Z_{n+k-1} + a_{n+k} - \theta a_{n+k-1} | Z_n, Z_{n-1}, \cdots]$$

$$\rightarrow f_{n,k} = \phi f_{n,k-1} = \phi^{k-1} f_{n,1}, \quad k = 2,\ 3,\ \cdots$$

· k단계 이후 예측오차의 분산

$$\upsilon_{n,k} = Var(e_{n,k}) = Var[Z_{n+k} | Z_n, Z_{n-1}, \cdots]$$

$$= Var[\phi Z_{n+k-1} + a_{n+k} - \theta a_{n+k-1} | Z_n, Z_{n-1}, \cdots]$$

→ ARMA(1,1)모형의 예측오차분산

➤ $\upsilon_{n,1} = \sigma_a^2$

➤ $\upsilon_{n,2} = \left[1 + (\phi - \theta)^2\right] \sigma_a^2$

➤ $\upsilon_{n,k} = \phi^2 \upsilon_{n,k-1} + \left[1 + \theta^2 - 2\phi\theta\right] \sigma_a^2,\ k = 2,\ 3,\ \cdots$

5.3 예측식의 갱신

■ 갱신식이 필요한 이유

· 매시점마다 일단계 이후 또는 k단계 이후 값을 예측하는 경우 과거 시계열 값을 반복적으로 예측값에 대입하는 것은 번거로울 수 있다.

· 어떤 시점에서 미래를 예측할 때 이전 시점에서 산출된 예측값과 새로운 관측값만을 사용하는 갱신식(updating formula)이 있다면 편리할 수 있을 것이다.

■ 일단계 이후 예측식에 대한 갱신식

$$f_{n,1} = \psi_1 a_n + \psi_2 a_{n-1} + \cdots$$

$$f_{n,2} = \psi_2 a_n + \psi_3 a_{n-1} + \cdots$$

$$f_{n+1,1} = \psi_1 a_{n+1} + \psi_2 a_n + \cdots \rightarrow f_{n+1,1} = \psi_1 a_{n+1} + f_{n,2} \leftarrow e_{n,1} = Z_{n+1} - f_{n,1} = a_{n+1}$$

$$\Rightarrow f_{n+1,1} = f_{n,2} + \psi_1 (Z_{n+1} - f_{n,1})$$

· 시점 n에서 일단계 이후 및 이단계 이후 예측값이 있다면 시점 $n+1$에서의 일단계 이후 예측값을 상기 식을 이용하여 구할 수 있다.

■ k단계 이후 예측식에 대한 갱신식을 유도하기 위해 상기 과정을 반복하면

$$f_{n+1,k} = f_{n,k+1} + \psi_k (Z_{n+1} - f_{n,1}),\ k = 1,\ 2,\ \cdots$$