第八讲：ARMA 和 ARIMA 模型预测

概要： 本讲介绍均值预测。预测的结果至少要有两个：均值和方差。本讲介绍均值的预测，主要内容包括：预测的原理，以及 AR 模型、MA 模型、ARMA 模型和 ARIMA 模型的预测，最后介绍修正预测的原理及优点。

所谓预测，就是利用序列已经观测到的样本值对序列在未来某个时刻的取值进行估计，未来某个时刻的值是个随机变量，预测的结果至少要有两个：均值和方差，这是描述随机变量分布特征最重要的两个统计量。目前，对平稳序列最常用的预测方法是线性最小方差预测，线性是指预测值为观测值序列的线性函数，最小方差是指预测方差达到最小。

假定，已知观测值序列$x_1, x_2, \dots, x_t$，参数估计得到的结果$\widehat{\phi_0}, \widehat{\phi_1}, \widehat{\phi_2}, \dots, \widehat{\phi_p}$，$\widehat{\theta_1}, \widehat{\theta_2}, \dots, \widehat{\theta_q}$，且$\{X_t\}$满足正态分布，$\{\varepsilon_t\}$为白噪声序列。

一 预测均值（线性预测函数）

第四讲中讲到，根据平稳 ARMA(p,q)模型的可逆性，可以用 AR 结构（逆转形式）表达任意一个平稳的 ARMA 模型，即：

$$\varepsilon_t = \frac{\Phi(B)}{\Theta(B)} X_t = \sum_{j=0}^{\infty} I_j X_{t-j} = I(B) X_t \qquad 6.13$$

$$E(\varepsilon_t) = E\left(\sum_{j=0}^{\infty} I_j X_{t-j}\right) = 0$$

$$I_0 X_t + I_1 X_{t-1} + I_2 X_{t-2} + \cdots = 0$$

$$X_t = -I_1 X_{t-1} - I_2 X_{t-2} - \cdots$$

其中，$I_0, I_1, I_2, \dots$.就是逆函数。这意味着适用递推法，基于现有的序列观测值，可以预测未来任意时刻的序列值（k-steps-ahead forecasting，前向 k 步预测）：

$$\hat{x}_t(1) = -I_1\, x_t - I_2\, x_{t-1} - I_3\, x_{t-2} - \cdots \qquad 6.14$$

$$\hat{x}_t(2) = -I_1\, \hat{x}_t(1) - I_2\, x_t - I_3\, x_{t-1} - \cdots \qquad 6.15$$

$$\hat{x}_t(3) = -I_1\, \hat{x}_t(2) - I_2\, \hat{x}_t(1) - I_3\, x_t - \cdots \qquad 6.16$$

将 6.14 代入 6.15，经整理得：

$$\hat{x}_t(2) = (I_1^2 - I_2) x_t + (I_1 I_2 - I_3) x_{t-1} + (I_1 I_3 - I_4)\, x_{t-2} \quad 6.17$$

将 6.14，6.17 代入 6.16，经整理得：

$$\hat{x}_t(3) = (-I_1^3 + 2I_1I_2 - I_3)x_t + (-I_1^2I_2 + I_1I_3 + I_2^2 - I_4)x_{t-1} + (-I_1^2I_3 + I_1I_4 + I_2I_3 - I_4)x_{t-2} + \cdots \qquad 6.18$$

由以上推导可以知道：

$$\hat{x}_t(k) = D_0x_t + D_1x_{t-1} + D_2x_{t-2} + \cdots = \sum_{k=0}^{\infty} D_k\, x_{t-k} \qquad 6.19$$

式 6.19 就是线性预测函数，说明前向 k 步预测都是历史值的线性函数。该线性函数的系数D_i是 ARMA 模型逆函数I_k的函数，即$D_i = f_i(I_1, I_2, I_3, \dots.)$，而逆函数$I_k$与 ARMA 模型系数的关系为（逆函数的递推公式，具体推导过程不要求，但要记住公式）：

$$\begin{cases} I_0 = 1 \\ I_k = \sum_{j=1}^{k} \hat{\theta}_j' I_{k-j} - \hat{\phi}_k' \qquad k \geq 1 \end{cases} \qquad 6.20$$

式中，

$$\hat{\theta}_j' = \begin{cases} \hat{\theta}_j & 1 \leq j \leq q \\ 0 & j > q \end{cases} \qquad \hat{\phi}_k' = \begin{cases} \hat{\phi}_k & 1 \leq k \leq p \\ 0 & k > p \end{cases}$$

二、预测方差

前向 k 步预测的误差为：

$$e_t(k) = x_{t+k} - \hat{x}_t(k) \qquad 6.21$$

为了便于误差方差的计算，有必要将 ARMA(p,q)模型转换为 MA 结构（传递形式），即：

$$X_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t = \sum_{j=0}^{\infty} G_j\, \varepsilon_{t-j} = G(B)\varepsilon_t \qquad 6.22$$

其中，$G_0, G_1, G_2, \dots.$就是 Green 函数。Green 函数与 ARMA 模型系数的关系为（Green 函数的递推公式，具体推导过程不要求，但要记住公式）：

$$\begin{cases} G_0 = 1 \\ G_k = \sum_{j=1}^{k} \hat{\phi}_k' G_{k-j} - \hat{\theta}_j' \qquad k \geq 1 \end{cases} \qquad 6.23$$

式中，

$$\hat{\theta}_j' = \begin{cases} \hat{\theta}_j & 1 \leq j \leq q \\ 0 & j > q \end{cases} \qquad \hat{\phi}_k' = \begin{cases} \hat{\phi}_k & 1 \leq k \leq p \\ 0 & k > p \end{cases}$$

式 6.21 的右边x_{t+k}采用 6.22 的形式，$\hat{x}_t(k)$结合式 6.19 和 6.22 形式，展开得到：

$$e_t(k) = x_{t+k} - \hat{x}_t(k) = \sum_{j=0}^{\infty} G_j\, \varepsilon_{t+k-j} - \sum_{k=0}^{\infty} D_k \sum_{j=0}^{\infty} G_j\, \varepsilon_{t+k-j} = \sum_{j=0}^{\infty} G_j\, \varepsilon_{t+k-j} - \sum_{j=0}^{\infty} W_j\, \varepsilon_{t-j}$$

$$= \sum_{j=0}^{k-1} G_j\, \varepsilon_{t+k-j} + \sum_{j=0}^{\infty} (G_{k+j} - W_j)\varepsilon_{t-j} \qquad 6.24$$

预测方差为：

$$Var[e_t(k)] = \left[\sum_{j=0}^{k-1} {G_j}^2 + \sum_{j=0}^{\infty} \left(G_{k+j} - W_j\right)^2\right] \widehat{\sigma_\varepsilon^2} \geq \sum_{j=0}^{k-1} {G_j}^2\, \widehat{\sigma_\varepsilon^2} \qquad 6.25$$

显然，要使预测方差达到最小，必须

$$G_{k+j} = W_j \quad j = 0,1,2,\dots\dots. \qquad 6.26$$

这时，x_{t+k}的预测值为：

$$\hat{x}_t(k) = \sum_{k=0}^{\infty} D_k\, x_{t-k} = \sum_{k=0}^{\infty} D_k \sum_{j=0}^{\infty} G_j\, \varepsilon_{t+k-j} = \sum_{j=0}^{\infty} W_j\, \varepsilon_{t-j} = \sum_{j=0}^{\infty} G_{k+j}\, \varepsilon_{t-j} \qquad 6.27$$

预测误差为：

$$e_t(k) = \sum_{j=0}^{k-1} G_j\, \varepsilon_{t+k-j} \qquad 6.28$$

由于$\{\varepsilon_t\}$为白噪声序列，因此，

$$E[e_t(k)] = 0$$

$$Var[e_t(k)] = \sum_{j=0}^{k-1} {G_j}^2\, \widehat{\sigma_\varepsilon^2}$$

在正态性假设下，向前 k 步预测值（均值）是$\hat{x}_t(k)$，预测误差为$e_t(k)$，预测误差的方差为$var[e_t(k)]$，则向前 k 步预测的 95%概率保证的置信区间为：

$$(\hat{x}_t(k) - 1.96var[e_t(k)],\ \hat{x}_t(k) + 1.96var[e_t(k)])$$

三、 线性最小方差预测性质

序列真实值采用 Green 函数可以进行如下分解：

$$x_{t+k} = \sum_{j=0}^{\infty} G_j\, \varepsilon_{t+k-j} = (G_0\varepsilon_{t+k} + G_1\varepsilon_{t+k-1} + \cdots + G_{k-1}\varepsilon_{t+1}) + (G_k\varepsilon_t + G_{k+1}\varepsilon_{t-1} + \cdots)$$

$$= e_t(k) + \hat{x}_t(k) \qquad 6.29$$

而依据 6.19 得：

$$\hat{x}_t(k) = D_0 x_t + D_1 x_{t-1} + D_2 x_{t-2} + \cdots = \sum_{k=0}^{\infty} D_k x_{t-k}$$

说明，未来任意 k 期的预测值都可以表达成历史信息的线性函数，在$x_t, x_{t-1}, x_{t-2}, \dots x_1, x_0$ 已知的条件下，$\hat{x}_t(k)$值为常数，其方差为 0，即：

$$E(\hat{x}_t(k)|x_1, x_2, \dots, x_t) = \hat{x}_t(k)$$

$$Var(\hat{x}_t(k)|x_1, x_2, \dots, x_t) = 0$$

则式 6.29 两边 取期望和方差得：

$$E(x_{t+k}|x_1, x_2, \dots, x_t) = E[e_t(k)|x_1, x_2, \dots, x_t] + E[\hat{x}_t(k)|x_1, x_2, \dots, x_t] = \hat{x}_t(k) \qquad 6.30$$

$$Var(x_{t+k}|x_1, x_2, \dots, x_t) = Var[e_t(k)|x_1, x_2, \dots, x_t] + Var[\hat{x}_t(k)|x_1, x_2, \dots, x_t]$$

$$= Var[e_t(k)] \qquad 6.31$$

式 6.30 和 6.31 说明，估计值$\hat{x}_t(k)$是序列值x_{t+k}在$x_1, x_2, \dots, x_t$已知情况下的无偏中最小方差估计值，且预测方差只与预测步长 *k* 有关，而与预测起点 *t* 无关。预测步长 *k* 越大，预测方差也越大，因而为了保证预测精度，时序数据只适合 短期预测（通常 k=3~5）

四、 AR(p)模型预测

掌握 AR(1)和 AR(2)模型的前向 3 期预测的计算。

先看最为简单的 AR(1)模型：$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t$

已知观测值序列，$x_1, x_2, \dots, x_t$，则前向 1 步预测均值如下：

$$\hat{x}_t(1) = E(x_{t+1}|x_1, x_2, \dots, x_t) = E\left(\widehat{\phi_0} + \widehat{\phi_1} x_t + \varepsilon_{t+1}|x_1, x_2, \dots, x_t\right) = \widehat{\phi_0} + \widehat{\phi_1} x_t \qquad 6.32$$

这里表达为条件期望的形式，说明在 *t* 时刻，我们可以获得的信息是$x_1, x_2, \dots, x_t$，所做的预测是基于观测值信息获得的。

前向 2 步预测均值如下：

$$\hat{x}_t(2) = E(x_{t+2}|x_1, x_2, \dots, x_t) = E\left(\widehat{\phi_0} + \widehat{\phi_1}\hat{x}_t(1) + \varepsilon_{t+2}|x_1, x_2, \dots, x_t\right)$$

$$= \widehat{\phi_0}\left(1 + \widehat{\phi_1}\right) + \widehat{\phi_1}^2 x_t \qquad 6.33$$

即，*t+1* 时刻预测值$\hat{x}_t(1)$可以用来预测 *t+2* 时刻预测值$\hat{x}_t(2)$。

前向 *3* 步预测均值如下：

$$\hat{x}_t(3) = E(x_{t+3}|x_1, x_2, \dots, x_t) = E\left(\widehat{\phi_0} + \widehat{\phi_1}\hat{x}_t(2) + \varepsilon_{t+3}|x_1, x_2, \dots, x_t\right)$$

$$= \widehat{\phi_0} + \widehat{\phi_1}\left[\widehat{\phi_0}\left(1 + \widehat{\phi_1}\right) + \widehat{\phi_1}^2 x_t\right]$$

$$= \widehat{\phi_0}\left(1 + \widehat{\phi_1} + \widehat{\phi_1}^2\right) + \widehat{\phi_1}^3 x_t \qquad 6.34$$

依此类推，前向 k 步预测值$\hat{x_t}(k)$的预测均值为：

$$\hat{x_t}(k) = E(x_{t+k}|x_1, x_2, \dots, x_t) = \widehat{\phi_0}\left(1 + \widehat{\phi_1} + \widehat{\phi_1}^2 + \cdots + \widehat{\phi_1}^{k-1}\right) + \widehat{\phi_1}^k x_t \qquad 6.35$$

表示所有向前 k 步的预测值都可以由 t 时刻的已知信息推导出（只要知道x_t的值就可以了）。AR(1)模型是平稳的，则式 6.35 的$|\phi_1| < 1$，当$k \to \infty$时，有

$$\hat{x_t}(k) = E(x_{t+k}|x_1, x_2, \dots, x_t) \to \frac{\widehat{\phi_0}}{1 - \widehat{\phi_1}} \qquad 6.36$$

说明，x_{t+k}的条件期望将收敛于 AR(1)模型的均值（无条件期望）。

定义向前 k 步的预测误差为$e_t(k)$，其定义式为：

$$e_t(k) = x_{t+k} - \hat{x_t}(k) \qquad 6.37$$

前向 1 步预测误差为：

$$e_t(1) = x_{t+1} - \hat{x_t}(1) = \widehat{\phi_0} + \widehat{\phi_1}x_t + \varepsilon_{t+1} - \widehat{\phi_0} - \widehat{\phi_1}x_t = \varepsilon_{t+1} \qquad 6.38$$

正好等于x_{t+1}中不可预测的部分（在已知 t 时刻信息的条件下）。预测方差为：

$$Var[e_t(1)] = var(\varepsilon_{t+1}) = \sigma_\varepsilon^2 \qquad 6.39$$

前向 2 步预测误差为：

$$e_t(2) = x_{t+2} - \hat{x_t}(2) = \widehat{\phi_0} + \widehat{\phi_1}x_{t+1} + \varepsilon_{t+2} - \widehat{\phi_0} - \widehat{\phi_1}\hat{x_t}(1) = \varepsilon_{t+2} + \widehat{\phi_1}\left(x_{t+1} - \hat{x_t}(1)\right)$$
$$= \varepsilon_{t+2} + \widehat{\phi_1}\varepsilon_{t+1} \qquad 6.40$$

预测方差为：

$$Var[e_t(2)] = var\left(\varepsilon_{t+2} + \widehat{\phi_1}\varepsilon_{t+1}\right) = \left(1 + \widehat{\phi_1}^2\right)\sigma_\varepsilon^2 \qquad 6.41$$

依次类推，AR(1)模型向前 k 步的预测误差为：

$$e_t(k) = \varepsilon_{t+k} + \widehat{\phi_1}\varepsilon_{t+k-1} + \widehat{\phi_1}^2\varepsilon_{t+k-2} + \cdots + \widehat{\phi_1}^{k-1}\varepsilon_{t+1} \qquad 6.42$$

方差为：

$$var[e_t(k)] = var\left[\varepsilon_{t+k} + \widehat{\phi_1}\varepsilon_{t+k-1} + \widehat{\phi_1}^2\varepsilon_{t+k-2} + \cdots + \widehat{\phi_1}^{k-1}\varepsilon_{t+1}\right]$$
$$= \sigma_\varepsilon^2\left[1 + \widehat{\phi_1}^2 + \widehat{\phi_1}^4 + \cdots + \widehat{\phi_1}^{2(k-1)}\right] \qquad 6.43$$

说明预测误差的方差随着预测步长 k 的增大而增加，或者说，平稳序列的预测只适用于短期（通常步长为 3~5）。当$k \to \infty$时，式 6.20 得：

$$var[e_t(k)] \to \frac{\sigma_\varepsilon^2}{1 - \widehat{\phi_1}^2} \qquad 6.44$$

说明，预测误差的方差将收敛于 AR(1)模型的方差（无条件方差）。

因此，如果$\{\varepsilon_t\}$序列是正态分布，就可以构造出预测的置信区间。前向 1 步预测均值为

$\widehat{\phi_0}+\widehat{\phi_1}x_t$，预测误差方差为$\sigma_\varepsilon^2$，则前向 1 步预测的 95%概率保证的置信区间为：

$$\widehat{\phi_0}+\widehat{\phi_1}x_t \pm 1.96\sigma_\varepsilon$$

前向 2 步预测均值为$\widehat{\phi_0}\left(1+\widehat{\phi_1}\right)+\widehat{\phi_1}^2 x_t$，预测误差方差为$\sigma_\varepsilon^2\left[1+\widehat{\phi_1}^2\right]$，那么前向 2 步预测的 95%概率保证的置信区间为：

$$\left(\widehat{\phi_0}\left(1+\widehat{\phi_1}\right)+\widehat{\phi_1}^2 x_t-1.96\sigma_\varepsilon\sqrt{1+\widehat{\phi_1}^2},\quad \widehat{\phi_0}\left(1+\widehat{\phi_1}\right)+\widehat{\phi_1}^2 x_t+1.96\sigma_\varepsilon\sqrt{1+\widehat{\phi_1}^2}\right)$$

AR(2)模型前向 3 期预测值和预测方差的推导，大家作为课后作业。

五、 MA(q)模型预测

掌握 MA(1)和 MA(2)模型的前向 3 期预测的计算。

先看 MA(2)模型：$X_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}$。

已知观测值序列，$x_1,x_2,\ldots,x_t$，和估计值序列，$\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}$。

则前向 1 步预测均值为：

$$\widehat{x_t}(1)=E\left(x_{t+1}|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right)=E\left(\varepsilon_{t+1}-\widehat{\theta_1}\widehat{\varepsilon_t}-\widehat{\theta_2}\widehat{\varepsilon_{t-1}}|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right)$$
$$=E\left[\varepsilon_{t+1}-\widehat{\theta_1}(x_t-\widehat{x_t})-\widehat{\theta_2}(x_{t-1}-\widehat{x_{t-1}})|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right]$$
$$=-\widehat{\theta_1}(x_t-\widehat{x_t})-\widehat{\theta_2}(x_{t-1}-\widehat{x_{t-1}}) \qquad 6.45$$

预测方差为：

$$Var[\widehat{x_t}(1)]=Var\left[\varepsilon_{t+1}-\widehat{\theta_1}(x_t-\widehat{x_t})-\widehat{\theta_2}(x_{t-1}-\widehat{x_{t-1}})|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right]$$
$$=Var(\varepsilon_{t+1})=\sigma_\varepsilon^2 \qquad 6.46$$

前向 2 步预测均值为：

$$\widehat{x_t}(2)=E\left(x_{t+2}|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right)=E\left(\varepsilon_{t+2}-\widehat{\theta_1}\varepsilon_{t+1}-\widehat{\theta_2}\widehat{\varepsilon_t}|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right)$$
$$=E\left[\varepsilon_{t+2}-\widehat{\theta_1}\varepsilon_{t+1}-\widehat{\theta_2}(x_t-\widehat{x_t})|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right]$$
$$=-\widehat{\theta_2}(x_t-\widehat{x_t}) \qquad 6.47$$

预测方差为：

$$Var[\widehat{x_t}(2)]=Var\left[\varepsilon_{t+2}-\widehat{\theta_1}\varepsilon_{t+1}-\widehat{\theta_2}(x_t-\widehat{x_t})|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right]$$
$$=Var\left(\varepsilon_{t+2}-\widehat{\theta_1}\varepsilon_{t+1}\right)=(1+\widehat{\theta_1}^2)\sigma_\varepsilon^2 \qquad 6.48$$

前向 3 步预测均值为：

$$\widehat{x_t}(3)=E\left(x_{t+3}|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right)$$
$$=E\left(\varepsilon_{t+3}-\widehat{\theta_1}\varepsilon_{t+2}-\widehat{\theta_2}\varepsilon_{t+1}|x_1,x_2,\ldots,x_t,\widehat{x_1},\widehat{x_2},\ldots,\widehat{x_t}\right)=0 \qquad 6.49$$

预测方差为：

$$Var[\hat{x}_t(3)] = Var\left[\varepsilon_{t+3} - \widehat{\theta_1}\varepsilon_{t+2} - \widehat{\theta_2}\varepsilon_{t+1} \middle| x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}\right]$$

$$= (1 + \widehat{\theta_1}^2 + \widehat{\theta_2}^2)\sigma_\varepsilon^2 \qquad 6.50$$

以此类推，当$k > 3$时，前向 k 步预测均值为：

$$\hat{x}_t(k) = E(x_{t+k}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t})$$

$$= E\left(\varepsilon_{t+k} - \widehat{\theta_1}\varepsilon_{t+k-1} - \widehat{\theta_2}\varepsilon_{t+k-2}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}\right) = 0 \qquad 6.51$$

预测方差为：

$$Var[\hat{x}_t(k)] = Var[\varepsilon_{t+k} - \widehat{\theta_1}\varepsilon_{t+k-1} - \widehat{\theta_2}\varepsilon_{t+k-2}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}]$$

$$= (1 + \widehat{\theta_1}^2 + \widehat{\theta_2}^2)\sigma_\varepsilon^2 \qquad 6.51$$

如果$\{\varepsilon_t\}$序列是正态分布，就可以构造出预测的置信区间。那么前向 k 步预测的 95%概率保证的置信区间为：

$$\left(-1.96\sigma_\varepsilon\sqrt{(1 + \widehat{\theta_1}^2 + \widehat{\theta_2}^2)}, \quad +1.96\sigma_\varepsilon\sqrt{(1 + \widehat{\theta_1}^2 + \widehat{\theta_2}^2)}\right)$$

这说明 MA(q)序列理论上只能预测 q 步之内的序列走势，超过 q 步预测值恒等于序列均值，同时，预测方差也固定不变。这是由 MA(q)序列自相关 q 步截尾的性质决定的。

Qianxiang

MA(1)模型前向 3 期预测值和预测方差的推导，大家作为课后作业。

六、ARMA(p,q)模型预测

主要掌握 ARMA(1,1)模型前向 3 期的预测。

ARMA(1,1)模型：$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1}$

已知观测值序列，$x_1, x_2, \dots, x_t$，和估计值序列，$\widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}$。

则前向 1 步预测均值为：

$$\hat{x}_t(1) = E(x_{t+1}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t})$$

$$= E\left(\widehat{\phi_0} + \widehat{\phi_1}x_t + \varepsilon_{t+1} - \widehat{\theta_1}\hat{\varepsilon}_t|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}\right)$$

$$= E\left[\widehat{\phi_0} + \widehat{\phi_1}x_t + \varepsilon_{t+1} - \widehat{\theta_1}(x_t - \hat{x}_t)\middle|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}\right]$$

$$= \widehat{\phi_0} + \widehat{\phi_1}x_t - \widehat{\theta_1}(x_t - \hat{x}_t) \qquad 6.52$$

预测方差为：

$$Var[\widehat{x_t}(1)] = Var[e_t(1)] = G_0{}^2\sigma_\varepsilon^2 = \widehat{\sigma_\varepsilon^2}$$

前向 2 步预测均值为：

$$\widehat{x_t}(2) = E(x_{t+2}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t})$$

$$= E\left(\widehat{\phi_0} + \widehat{\phi_1}\widehat{x_t}(1) + \varepsilon_{t+2} - \widehat{\theta_1}\varepsilon_{t+1}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}\right) = \widehat{\phi_0} + \widehat{\phi_1}\widehat{x_t}(1)$$

$$= \widehat{\phi_0} + \widehat{\phi_1}\left[\widehat{\phi_0} + \widehat{\phi_1}x_t - \widehat{\theta_1}(x_t - \widehat{x_t})\right]$$

$$= \widehat{\phi_0}\left(1 + \widehat{\phi_1}\right) + \widehat{\phi_1}^2 x_t - \widehat{\phi_1}\,\widehat{\theta_1}(x_t - \widehat{x_t}) \qquad 6.54$$

预测方差为：

$$Var[\widehat{x_t}(2)] = Var[e_t(2)] = ({G_0}^2 + {G_1}^2)\widehat{\sigma_\varepsilon^2} = \left(1 + {G_1}^2\right)\widehat{\sigma_\varepsilon^2} = [1 + (\widehat{\phi_1} - \widehat{\theta_1})^2]\widehat{\sigma_\varepsilon^2}$$

前向 3 步预测均值为：

$$\widehat{x_t}(3) = E(x_{t+3}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t})$$

$$= E\left(\widehat{\phi_0} + \widehat{\phi_1}\widehat{x_t}(2) + \varepsilon_{t+3} - \widehat{\theta_1}\varepsilon_{t+2}|x_1, x_2, \dots, x_t, \widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_t}\right) = \widehat{\phi_0} + \widehat{\phi_1}\widehat{x_t}(2)$$

$$= \widehat{\phi_0} + \widehat{\phi_1}\left[\widehat{\phi_0}\left(1 + \widehat{\phi_1}\right) + \widehat{\phi_1}^2 x_t - \widehat{\phi_1}\,\widehat{\theta_1}(x_t - \widehat{x_t})\right]$$

$$= \widehat{\phi_0}\left(1 + \widehat{\phi_1} + \widehat{\phi_1}^2\right) + \widehat{\phi_1}^3 x_t - \widehat{\phi_1}^2\,\widehat{\theta_1}(x_t - \widehat{x_t}) \qquad 6.55$$

预测方差为：

$$Var[\widehat{x_t}(3)] = Var[e_t(3)] = ({G_0}^2 + {G_1}^2 + {G_2}^2)\widehat{\sigma_\varepsilon^2} = \left(1 + {G_1}^2 + {G_2}^2\right)\widehat{\sigma_\varepsilon^2}$$

$$= [1 + \left(\widehat{\phi_1} - \widehat{\theta_1}\right)^2 + \widehat{\phi_1}^2\left(\widehat{\phi_1} - \widehat{\theta_1}\right)^2]\widehat{\sigma_\varepsilon^2}$$

从上述计算可知，方差计算的关键就是要记住 Green 函数的递推公式（式 6.23）。

前向 1、2、3 步预测的 95%概率保证的置信区间，大家自己写写看。

七、 ARIMA(p,d,q)模型预测

在最小均方误差预测原理下，ARIMA 模型和 ARMA 模型的预测方法非常类似。掌握 ARIMA(1,1,1)模型的预测。

ARIMA 模型，即：

$$\Phi(B)(1-B)^d X_t = \Theta(B)\,\varepsilon_t \qquad 6.56$$

可以通过变换，将式 6.56 转换成 MA 结构（传递形式），即：

$$X_t = \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \cdots = \Psi(B)\varepsilon_t \qquad 6.57$$

式 6.57 中的$\Psi_1, \Psi_2, \dots$类似 ARMA 模型中的 Green 函数，其值由下式决定：

$$\Phi(B)(1-B)^d\Psi(B) = \Theta(B) \qquad 6.58$$

关键是要找到$\Psi_1, \Psi_2, \dots$和$\phi_1, \phi_2, \dots \phi_p, \theta_1, \theta_2, \dots \theta_q$之间的关系。现以 ARIMA(1,1,1)为例说明。

$$(1-B)\Psi(B) = \frac{1-\theta_1 B}{1-\phi_1 B} \qquad 6.59$$

$$(1-B)(1+\Psi_1 B+\Psi_2 B^2+\cdots)=(1-\theta_1 B)(1+\phi_1 B+\phi_1^2 B^2+\phi_1^3 B^3+\cdots) \qquad 6.60$$

$$1+(\Psi_1-1)B+(\Psi_2-\Psi_1)B^2+\cdots=1+(\phi_1-\theta_1)B+(\phi_1^2-\phi_1\theta_1)B^2+\cdots \qquad 6.61$$

式 6.61 两边相等，得到$\Psi_1,\Psi_2,\Psi_3\ldots$和$\phi_1,\theta_1$的关系式，即：

$$\Psi_1-1=\phi_1-\theta_1$$

$$\Psi_2-\Psi_1=\phi_1^2-\phi_1\theta_1$$

$$\Psi_3-\Psi_2=\phi_1^3-\phi_1^2\theta_1$$

$$\ldots\ldots$$

那么，以 t 时刻为起点，未来 k 期的真实值x_{t+k}为：

$$x_{t+k}=(\varepsilon_{t+k}+\Psi_1\varepsilon_{t+k-1}+\cdots+\Psi_{k-1}\varepsilon_{t+1})+(\Psi_k\varepsilon_t+\Psi_{k+1}\varepsilon_{t-1}+\cdots) \qquad 6.62$$

由于$\varepsilon_{t+1},\varepsilon_{t+2},\ldots\varepsilon_{t+k-1},\varepsilon_{t+k}$是未来的值，$t$ 时刻无法获取，所以x_{t+k}的估计值只能是：

$$\widehat{x_t}(k)=\Psi_0^*\varepsilon_t+\Psi_1^*\varepsilon_{t-1}+\Psi_2^*\varepsilon_{t-2}+\cdots \qquad 6.63$$

真实值x_{t+k}和估计值$\widehat{x_t}(k)$的误差为：

$$\begin{aligned}e_t(k)&=x_{t+k}-\widehat{x_t}(k)\\&=(\varepsilon_{t+k}+\Psi_1\varepsilon_{t+k-1}+\cdots+\Psi_{k-1}\varepsilon_{t+1})+[(\Psi_k-\Psi_0^*)\varepsilon_t+(\Psi_{k+1}-\Psi_1^*)\varepsilon_{t-1}\\&\quad+\cdots] \qquad 6.64\end{aligned}$$

方差为：

$$Var[e_t(k)]=(1+\Psi_1^2+\Psi_2^2\ldots+\Psi_{k-1}^2)\sigma_\varepsilon^2+[(\Psi_k-\Psi_0^*)^2+(\Psi_{k+1}-\Psi_1^*)^2+\cdots]\sigma_\varepsilon^2 \qquad 6.65$$

要使方差最小，当且仅当

$$\Psi_k-\Psi_0^*=0$$

$$\Psi_{k+1}-\Psi_1^*=0$$

$$\Psi_{k+2}-\Psi_2^*=0$$

$$\ldots\ldots$$

因此，最小预测值为：

$$\widehat{x_t}(k)=\Psi_k\varepsilon_t+\Psi_{k+1}\varepsilon_{t-1}+\Psi_{k+2}\varepsilon_{t-2}\ldots \qquad 6.66$$

预测误差为：

$$e_t(k)=\varepsilon_{t+k}+\Psi_1\varepsilon_{t+k-1}+\cdots+\Psi_{k-1}\varepsilon_{t+1} \qquad 6.67$$

以 t 时刻为起点，未来 k 期的真实值x_{t+k}就是预测值$\widehat{x_t}(k)$加上预测误差$e_t(k)$，即：

$$x_{t+k} = \hat{x}_t(k) + e_t(k)$$
$$= (\Psi_k \varepsilon_t + \Psi_{k+1}\varepsilon_{t-1} + \Psi_{k+2}\varepsilon_{t-2} + \cdots)$$
$$+ (\varepsilon_{t+k} + \Psi_1 \varepsilon_{t+k-1} + \cdots + \Psi_{k-1}\varepsilon_{t+1}) \qquad 6.68$$

预测误差方差为：

$$Var[e_t(k)] = (1 + \Psi_1^2 + \Psi_2^2 \ldots + \Psi_{k-1}^2)\sigma_\varepsilon^2 \qquad 6.69$$

八、 修正预测

从上述的预测过程可知，预测的实质就是根据已有的观测值序列$x_1, x_2, \ldots, x_t$，对未来的某个时刻的发展水平x_{t+k}，$k = 1,2, \ldots.$，做出估计$\hat{x}_t(k)$。需要估计的时期越长（即 k 越大），未知信息就越多，估计精度就越低。随着时间的推移，会获得新的观测值$x_{t+1}, x_{t+2}, \ldots$，这就意味着未知信息减少，如果把新的观测值加进来，x_{t+k}的估计精度就会提高。例如：已知观测序列$x_1, x_2, \ldots, x_t$，要预测（t+k）时刻的值，表示为$\hat{x}_t(k)$；如果得到新的观测值x_{t+1}，观测值序列更新为$x_1, x_2, \ldots, x_t, x_{t+1}$则预测（$t$+$k$）时刻的值，表示为$\widehat{x_{t+1}}(k-1)$，所对应的预测方差，$\widehat{x_{t+1}}(k-1)$通常小于$\hat{x}_t(k)$。

所谓修正预测，就是研究如何利用新的信息去获得精度更高的预测值。最为直接的方法是，将信息加入到旧信息中，重新拟合模型，再利用拟合后的模型重新预测x_{t+k}的值。在新的信息量比较大且使用统计软件很便利的时候，这不失为一种可行的方法。但在新的数据量不大时，可以根据平稳时序预测的性质，寻求更为简便的修正方法。

在已知观测值序列$x_1, x_2, \ldots, x_t$时，x_{t+k}可以分解成如下 MA 结构：

$$x_{t+k} = G_0\varepsilon_{t+k} + G_1\varepsilon_{t+k-1} + \cdots + G_{k-1}\varepsilon_{t+1} + G_k\varepsilon_t + G_{k+1}\varepsilon_{t-1} + \cdots \qquad 6.70$$

其中：$e_t(k) = G_0\, \varepsilon_{t+k} + G_1\, \varepsilon_{t+k-1} + G_2\, \varepsilon_{t+k-2} + \cdots + G_{k-1}\, \varepsilon_{t+1}$

$$\hat{x}_t(k) = G_k\varepsilon_t + G_{k+1}\varepsilon_{t-1} + G_{k+2}\varepsilon_{t-2} \ldots$$

即，x_{t+k}的预测值$\hat{x}_t(k)$的表达式如上。

现获得新的信息x_{t+1}，观测值序列更新为$x_1, x_2, \ldots, x_t, x_{t+1}$，重新预测$x_{t+k}$为：

$$\widehat{x_{t+1}}(k-1) = G_{k-1}\varepsilon_{t+1} + G_k\varepsilon_t + G_{k+1}\varepsilon_{t-1} \ldots = G_{k-1}\varepsilon_{t+1} + \hat{x}_t(k) \qquad 6.71$$

预测误差为：

$$e_{t+1}(k-1) = G_0\, \varepsilon_{t+k} + G_1\, \varepsilon_{t+k-1} + G_2\, \varepsilon_{t+k-2} + \cdots + G_{k-2}\, \varepsilon_{t+2} \qquad 6.72$$

而采用旧信息获得的预测值$\hat{x}_t(k)$所对应的预测误差为：

$$e_t(k) = G_0\, \varepsilon_{t+k} + G_1\, \varepsilon_{t+k-1} + G_2\, \varepsilon_{t+k-2} + \cdots + G_{k-2}\, \varepsilon_{t+2} + G_{k-1}\, \varepsilon_{t+1} \qquad 6.73$$

式 6.58 和式 6.59 的区别是，后者比前者多了一项$G_{k-1}\, \varepsilon_{t+1}$。

预测误差的方差之差为：

$$var[e_t(k)] - var[e_{t+1}(k-1)] = G_{k-1}^2\sigma_\varepsilon^2 \qquad 6.74$$

说明修正前的误差方差要大于修正后的误差方差，添加新的信息x_{t+1}后，预测精度得到了提高。