ARMA forecast

Assume x_t causal and invertible. Let x^n_{n+m} be the mean square predictor of x_{n+m} based on $x_{1:n}$, that is

$$x^n_{n+m} = E(x_{n+m}|x_n, \ldots, x_1).$$

Also define predictor defined on the complete past

$$\tilde{x}_{n+m} = E(x_{n+m}|x_n, \ldots, x_1, x_0, x_{-1}, \ldots),$$

which equals

$$\begin{aligned} & E(x_{n+m}|x_n, \ldots, x_1, x_0, x_{-1}, \ldots) \\ = \quad & E(\Sigma_{j=0}^{\infty}\psi_j w_{m+n-j}|x_n, \ldots, x_1, x_0, x_{-1}, \ldots) \\ = \quad & \Sigma_{j=m}^{\infty}\psi_j w_{m+n-j}, \end{aligned}$$

since by causality and invertibility

$$\tilde{w}_t = E(w_t|x_n, \ldots, x_1, x_0, x_{-1}, \ldots) = \begin{cases} 0 & t > n \\ w_t & t \leq n. \end{cases}$$

Also by causality and invertibility

$$\begin{aligned} 0 = \quad & E(w_{n+m}|x_n, \ldots, x_1, x_0, x_{-1}, \ldots) \\ = \quad & E(\Sigma_{j=0}^{\infty}\pi_j x_{n+m-j}|x_n, \ldots, x_1, x_0, x_{-1}, \ldots) \\ = \quad & \Sigma_{j=0}^{m-1}\pi_j \tilde{x}_{n+m-j} + \Sigma_{j=m}^{\infty}\pi_j x_{n+m-j}. \end{aligned}$$

Because $\pi_0 = 1$

$$\tilde{x}_{n+m} = -\Sigma_{j=1}^{m-1}\pi_j \tilde{x}_{n+m-j} - \Sigma_{j=m}^{\infty}\pi_j x_{n+m-j}.$$

Now prediction using $\tilde{x}_{n+m}$ can be done recrursivly starting with

m=1 $\tilde{x}_{n+1} = -\Sigma_{j=1}^{\infty}\pi_j x_{1+n-j}$

m=2 $\tilde{x}_{n+2} = -\pi_1 \tilde{x}_{n+1} - \Sigma_{j=2}^{\infty}\pi_j x_{2+n-j}$

m=3 $\tilde{x}_{n+3} = -\pi_1 \tilde{x}_{n+1} - \pi_2 \tilde{x}_{n+2} - \Sigma_{j=3}^{\infty}\pi_j x_{3+n-j}$

etc. Also

$$x_{n+m} - \tilde{x}_{n+m} = \Sigma_{j=0}^{\infty}\psi_j w_{n+m-j} - \Sigma_{j=m}^{\infty}\psi_j w_{n+m-j} = \Sigma_{j=0}^{m-1}\psi_j w_{n+m-j}$$

so the *mean square prediction error* is

$$P_{n+m}^n = E(x_{n+m} - \tilde{x}_{n+m})^2 = \sigma^2 \Sigma_{j=0}^{m-1} \psi_j^2$$

The prediction errors are correlated so

$$E(x_{n+m} - \tilde{x}_{n+m})(x_{n+m+k} - \tilde{x}_{n+m+})^2 = \sigma^2 \Sigma_{j=0}^{m-1} \psi_j \psi_{j+k}.$$

Assume x_t ARMA with mean μ_x. Then

$$\tilde{x}_{n+m} = \mu_x + \Sigma_{j=m}^{\infty} \psi_j w_{m+n-j}$$

so $\tilde{x}_{n+m} \to \mu_x$ as $m \to \infty$ and $P_{n+m}^n \to \sigma^2 \Sigma_{j=0}^{\infty} \psi_j^2 = \gamma_x(0) = var(x_t)$.

Truncated prediction

Remember

$$\tilde{x}_{n+m} = -\Sigma_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \Sigma_{j=m}^{\infty} \pi_j x_{n+m-j}.$$

The variables $x_0, x_{-1}, \ldots$ are not observable so a truncated version is obtained by setting them equal to 0, their expected values, giving the *truncated predictor*

$$\tilde{x}_{n+m}^n = -\Sigma_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j}^n - \Sigma_{j=m}^{n+m-1} \pi_j x_{n+m-j}.$$

which also can be computed recursively for $m = 1, 2, \ldots$.

For AR(p) models $\tilde{x}_{n+m}^n = \tilde{x}_{n+m} = x_{n+m}^n$ when $n > p$ so the approximations are exact.

For ARMA(p,q) where $q > 1$ the situation is a bit more complicated. For example if x_t is ARMA(1,1), $x_{n+1} = \phi x_n + w_{n+1} + \theta w_t$. The truncated one step ahead forecast is

$$\tilde{x}_{n+1}^n = \phi x_n + 0 + \tilde{w}_n^n$$

setting $\tilde{w}_{n+1}^n = 0$ and for $m \geq 2$

$$\tilde{x}_{n+m}^n = \phi \tilde{x}_{n+m-1}$$

setting $\tilde{w}_t^n = 0$ when $t > n$. Then $\tilde{w}_n^n$ is initiated using $w_t = x_t - \phi x_{t-1} - \theta w_{t-1},\ t = 1, \ldots, n$. Set $\tilde{w}_0^n = 0, x_0 = 0$ and iterate

$$\tilde{w}_t^n = x_t - \phi x_{t-1} - \theta \tilde{w}_{t-1}^n,\ t = 1, \ldots, n$$

The general ARMA(p,q) is treated similarly using the model equation setting $\tilde{x}_t^n = x_t$ when $1 \leq t \leq n$ and $\tilde{x}_t^n = 0$ when $t \leq 0$ so

$$\tilde{x}_{n+m}^n = \phi_1 \tilde{x}_{n+m-1}^n + \cdots + \phi_p \tilde{x}_{n+m-p}^n + \theta_1 \tilde{w}_{n+m-1}^n + \cdots + \theta_q \tilde{w}_{n+m-q}^n$$

For $t = 1, \cdots, n$ the model equation with $\tilde{x}_t^n$ and $\tilde{w}_t^n = 0$ when $t > n$ or $t \leq 0$ are used to obtain $\tilde{w}_t^n$.

Thus,

$$\tilde{w}_n^t = \phi_1 \tilde{x}_t^n + \cdots + \phi_p \tilde{x}_{t-p}^n - \theta_1 \tilde{w}_{t-1} n - \cdots - \theta_q \tilde{w}_{t-q}^n$$

For ARMA(p,q) the weights ψ_j are known so the P_{n+m}^n derived earlier can be used as an approximate forecast variance. These and the predictions x_{n+m}^n or their approximations can be used to construct *prediction intervals*.

Backcasting

In backcasting x_{1-m} is predicted using data $x_{1:n}$. If $x_{1-m}^n = \alpha_n x_n + \alpha_{n-1} x_{n-1} + \cdots + \alpha_1 x_1$ the prediction equations take the form

$$\Sigma_{j=1}^n \alpha_j E(x_j x_k) = E(x_{1-m}^n x_k),\ k = 1, \cdots, n$$

or

$$\Sigma_{j=1}^n \alpha_j \gamma(k-j) = \gamma(m+k-1),\ k = 1, \cdots, n.$$

which are the same as for the forward case so the solutions $\alpha_j = \phi_{nj}^m$ and the backcasts are given by

$$x_{1-m}^n = \phi_{nn}^n x_n + \cdots + \phi_{n1}^m x_1.$$

Now consider the ARMA(1,1) process $x_t = \phi x_{t-1} + w_t + \theta w_{t-1}$, where the weights ψ_j are found by expanding $\frac{1+\theta z}{1-\phi z}$. The weights for the model $x_t = \phi x_{t+1} + v_t + \theta v_{t+1}$ is determined by expanding $\frac{1+\theta(1/z)}{1-\phi(1/z)}$ so x_t may be expressed as $x_t = \Sigma_{j=0}^\infty \psi_j v_{t+j}$. The two models have therefore the same autocovariance function if the $\sigma_w = \sigma_v$. If they in addition are iid $N(0, \sigma_w^2)$, the two models are equivalent.

Given data $x_{1:n}$, set $v_n^n = 0$ and generate the errors backwards

$$\tilde{v}_t^n = x_t - \phi x_{t+1} - \theta \tilde{v}_{t+1}^n,\ t = (n-1), (n-2), \ldots, 1.$$

and

$$\tilde{x}_0^n = \phi x_1 + \tilde{v}_1^n + \tilde{v}_0^n = \phi x_1 + \tilde{v}_1^n + 0$$

since $\tilde{v}_t^n,\ t \leq 0$. Continuing

$$\tilde{x}_{1-m}^n = \phi \tilde{x}_{2-m}^n,\ m = 2, 3, \cdots$$

To backcast: reverse the data, fit the model and predict.