Introduction to ARIMA model and how to do point/interval estimates

"ARIMA" stands for Auto-Regressive Integrated Moving Average. An ARIMA model is classified as an "ARIMA(p, d, q)" model, where p is the number of autoregressive terms, d is the number of nonseasonal differences needed for stationarity, and q is the number of lagged forcast error in the prediction equation.

The parameters (p, d, q) are determined usually through Box-Jenkins model.

Suppose $y_1, \cdots, y_n$ are the observed values. Let Y denote the d-th difference of y; i.e.,

$$Y_t = \sum_{i=0}^{d} (-1)^i \binom{d}{i} y_{t-i}.$$

The general forecasting equation of ARIMA(p, q, d) is given by

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q},$$

where $\hat{Y}_t$ is the prediction value (of the d-th difference series) at time t and e_t is the error term at time t given by $\hat{Y}_t - Y_t$. For $t > n$, the "observed value" Y_t is replaced by the prediction value $\hat{Y}_t$ and the error term is replaced by 0 (since there is no way to find out the true value y_t). The coefficients $\mu, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q$ are determined through fitting historical data. Point estimates (of y_t) are made by using the above equation to determine Y_t and by reversing the difference process.

For the interval estimate, rewrite the ARIMA model into an infinite MA series:

$$\hat{Y}_t = \mu + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots$$

This is always achievable when the ARIMA model is stable (or in mathematical terms, all roots of the polynomial $-1 + \phi_1 t + \cdots + \phi_p t^p$ are outside the closed unit disk). The α prediction interval for $\hat{Y}_t$ when $t > n$ is determined by

$$\hat{Y}_t \pm z_\alpha \sigma \sqrt{\left(1 + \sum_{i=1}^{t-n-1} \psi_i^2\right)},$$

where z_α is the $(1 - \alpha)$ critical value for standard normal distribution.

In particular, the 90% prediction interval for $\hat{Y}_{n+1}$ is determined by

$$\hat{Y}_{n+1} \pm 1.28\sigma.$$

The prediction interval for y_t for $t > n$ is again determined by reversing the difference process. In particular, the 90% prediction interval for y_{n+1} is determined by

$$\hat{y}_{n+1} \pm 1.28\sigma.$$

1. How to write ARIMA model into infinite MA series

Suppose the ARIMA forecast equation is

$$Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}.$$

Suppose L is the lag operator. For example $LY_t = Y_{t-1}$ and $L^2 Y_t = Y_{t-2}$. Using L, the ARIMA forecast equation can be rewritten as:

$$(I - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = \mu + (-\theta_1 L - \cdots - \theta_q L^q) e_t,$$

where I is the identity map; i.e., $IY_t = Y_t$. By fundamental theorem of algebra, there exists p roots of the polynomial $1 - \phi_1 t - \phi_2 t^2 - \cdots - \phi_p t^p$. Let those p roots be $\lambda_1, \cdots, \lambda_p$. Then

$$1 - \phi_1 t - \phi_2 t^2 - \cdots - \phi_p t^p = (-1)^p \phi_p (\lambda_1 - t)(\lambda_2 - t) \cdots (\lambda_p - t).$$

Hence, we can further rewrite the forecast equation as

$$(-1)^p \phi_p (\lambda_1 - L)(\lambda_2 - L) \cdots (\lambda_p - L) Y_t = \mu + (-\theta_1 L - \cdots - \theta_q L^q) e_t,$$

and thus

$$Y_t = (-1)^p \phi_p^{-1} (\lambda_p - L)^{-1} (\lambda_{p-1} - L)^{-1} \cdots (\lambda_1 - L)^{-1} (\mu + (-\theta_1 L - \cdots - \theta_q L^q) e_t). \tag{1}$$

When the ARIMA model is stable, $|\lambda_i|$ will be greater than 1. In this case

$$(\lambda_i - L)^{-1} = \frac{1}{\lambda_i} \sum_{j=0}^{\infty} \left(\frac{L}{\lambda_i} \right)^j .$$

This is easy to check:

$$\begin{aligned}
&\frac{1}{\lambda_i} \sum_{j=0}^{\infty} \left(\frac{L}{\lambda_i} \right)^j (\lambda_i - L) \\
=&\sum_{j=0}^{\infty} \left(\frac{L}{\lambda_i} \right)^j (I - \frac{L}{\lambda_i}) \\
=&\sum_{j=0}^{\infty} \left(\frac{L}{\lambda_i} \right)^j - \sum_{j=0}^{\infty} \left(\frac{L}{\lambda_i} \right)^{j+1} \\
=&\sum_{j=0}^{\infty} \left(\frac{L}{\lambda_i} \right)^j - \sum_{j=1}^{\infty} \left(\frac{L}{\lambda_i} \right)^j \\
=&I.
\end{aligned}$$

(1) is the corresponding infinite MA series.