## ARIMA method

ARIMA stands for Auto-Regressive Integrated Moving Average while ARMA is integrated with differencing. A nonseasonal ARIMA model is classified as ARIMA(p,d,q) where, p is the order of AR (Auto-Regressive) terms, d is the number of nonseasonal differences needed for stationarity (a series without trend and constant variability i.e., constant mean and variance), and q is the order of MA terms.

In an AR model of a time series the value of X at time t is a linear function of the values of X at time t - 1, t - 2...t - p and is written as:

$$AR(p): X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t$$

where,  $\epsilon_t$  is assumed to be iid (identical and independent of  $X_t$  distributed) and is normally distributed i.e.,  $\epsilon_t N(0, \sigma^2)$ .

The model is also written using the backshift operator B  $(B^k X_t = X_{t-k})$  as follows:

$$AR(p): X_t = \sum_{i=1}^p \phi_i B^i X_t$$

An alternative notation using  $\phi(B)$  as a polynomial of B is:

$$\phi(B)X_t = \epsilon_t$$

where,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2, ..., -\phi_p B^p = 1 - \sum_{i=1}^p \phi_i B^i$ .

By a MA(q) model of a time series the value of X at time t depends linearly on its own current and previous stochastic terms as follows:

$$MA(q): X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots, + \theta_q \epsilon_{t-q} = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Using the backshift operator B, the MA(q) model is simplified as:

$$MA(q): X_t = (1 + \theta_1 B + \theta_2 B^2 +, ..., +\theta_q B^q) \epsilon_t = (1 + \sum_{i=1}^q \theta_i B^i) \epsilon_t = \theta(B) \epsilon_t$$

Thus, an ARMA(p,q) model consists of both autoregressive (AR) part and moving average (MA) part:

$$ARMA(p,q): X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

or using the backshift operator B, we obtain the equation:

$$ARMA(p,q): \phi(B)X_t = \theta(B)\epsilon_t$$

In ARIMA models the original series is differenced to get a stationary series i.e.,  $\nabla X_t = X_t - X_{t-1} = (1-B)X_t$  where B is the backshift operator. Stationary series requires that all

roots of  $\phi(z) = 0$  are outside the unit circle thus, the final stationary model of ARIMA(p,d,q) is:

$$ARIMA(p, d, q) : \phi(B)(1 - B)^{d}X_{t} = \theta(B)\epsilon_{t}$$

Finally, the ARIMA model can be generalized as follow:

$$ARIMA(p, d, q) : \phi(B)\alpha(B)X_t = \theta(B)\epsilon_t$$

where,  $\phi(B)$  is an autoregressive polynomial with all roots are outside unit circle,  $\alpha(B)$  is the differencing filter renders the data stationary in which all roots are on the unit circle, and  $\theta(B)$  is a moving average polynomial with all roots are outside unit circle (to assure that  $\theta(B)$  is invertible). Alternatively,  $X_t$  can be expressed as:

$$X_t = \frac{\theta(B)}{\phi(B)\alpha(B)} \epsilon_t$$

Let us define an ARIMA model for the series  $X_t^*$  subject to m outliers defined as  $L_j(B)$  with weights w:

$$X_t^* = \sum_{j=1}^m w_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)} \epsilon_t$$

where  $I_t(t_j)$  is an indicator variable containing the value 1 at observation  $t_j$  where the j-th outlier arises and 0 otherwise,  $w_j$  denotes the magnitude of the  $j^{th}$  outlier effect, and  $L_j(B)$  depends on the type of the  $j^{th}$  outlier. Three types of outliers are considered in MODS:

- 1. Additive Outlier (AO): L(B) = 1. This outlier is a one time jump that gets back to the original values after one period.
- 2. Level Shift (LS):  $L(B) = \frac{1}{1-B}$ . This outlier is a jump to new levels.
- 3. Temporary Changes (TC):  $L(B) = \frac{1}{1-\delta B}$ . This outlier decays back to the original values after the jump in a pace of  $\delta$ .

Comments:

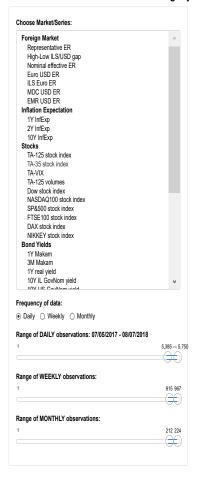
- 1. If level shifts are found at consecutive time points, only then point with higher t-statistic in absolute value is kept.
- 2. If more than one type of outlier exceed a predetermined threshold 'cval' at a given time point, the type of outlier with higher t-statistic in absolute value is kept and the others are removed. Screen example is depicted below:

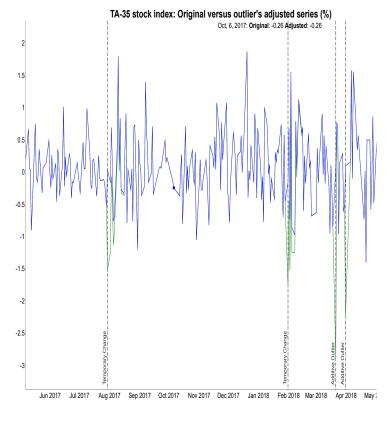
## References

Chen, C., and Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421), 284-297.

17/07/2018

## MODS - Multilier Outliers Detecting System





http://127.0.0.1:5016/