Nov 15 Notes

Hongdou Li

11/15/2018

Addressing Seasonality

Ordinary differencing $\nabla Y_t = (1-B)Y_t = Y_t - Y_{t-1}$ doesn't work to remove seasonal effects. For this, we need to perform **Seasonal differencing**.

Notation: $\nabla_k = (1 - B^k)$

$$\nabla_k Y_t = (1-B^k)Y_t = Y_t - Y_{t-k} <= \text{“lag k” differencing}$$

* This is different from ∇^k $((1-B)^k)$ which signifies k iterations of ordinary (lag 1) differencing.

Idea: A seasonal effect of period m manisfests itself as $s_t = s_{t\pm m}$. This sort of seasonal effect can be eliminated/ mitigated by finitely many applications of lag lag-m differencing.

Ex: $Y_t = s_t + \epsilon_t$ where s_t is a seasonal effect with period m.

$$\begin{aligned}
D_m Y_t &= (1-B^m)Y_t = (1-B^m)(s_t + \epsilon_t) \\
&= (s_t - s_{t-m}) + (\epsilon_t - \epsilon_{t-m}) \\
&= 0 + (\epsilon_t - \epsilon_{t-m}) \\
&= \nabla_m \epsilon_t
\end{aligned}$$

The goal is ingeneral to use ordinary differencing to eliminate trend and seasonal differencing to account for seasonality. Thus, both types of differencing may be necessary.

* the order of differencing does not matter. Mathematically the resultant series will be identical.

Illustration: suppose we need to ordinarily difference d times, and lag-m difference D times.

$$\nabla^d \nabla_m^D Y_t = (1-B)^d (1-B^m)^D Y_t$$
$$\nabla_m^D \nabla^d Y_t = (1-B^m)^D (1-B)^d Y_t$$

How do we choose m? The period m is the number of lags required for one iteration of the seasonal effect on an ACF plot.

SARIMA “Seasonal ARIMA”

$\{Y_t\} \sim SARIMA(p,d,q)_\times (P,D,Q)_m$ if $X_t = (1-B)^d(1-B^m)^D Y_t$ can be modeled by a stationary ARMA model:

$$\phi^*(B)X_t = \theta^*(B)\epsilon_t$$
$$\phi^*(B) = \phi(B)\Phi(B^n)$$
$$\theta^*(B) = \theta(B)\Theta(B^m)$$

$\phi(Z) = 1 - \phi_1 Z - \phi_2 Z^2 - ... - \phi_p Z^p <= p^{th}$ degree polynomial

$\Phi(Z) = 1 - \Phi_1 Z - \Phi_2 Z^2 - ... - \Phi_P Z^P <= P^{th}$ degree polynomial

$\theta(Z) = 1 + \theta_1 Z + \theta_2 Z^2 + ... + \theta_q Z^q <= q^{th}$ degree polynomial

$\Theta(Z) = 1 + \Theta_1 Z + \Theta_2 Z^2 + ... + \Theta_Q Z^Q$ <= Q^{th} degree polynomial

$$\phi(B)\Phi(B^m)(1-B)^d(1-B^m)^D Y_t = \theta(B)\Theta(B^m)\epsilon_t$$

- Idea: The data within a season can be viewed as a **within-season** time series. The data between seasons can be viewed as a **between-season** time series. These two time series may have different ARMA representations.

EX: Suppose $\{Y_t\}$ is observed quarterly and so it has a seasonal effect of period m=4

$$\begin{matrix} y_1 & y_2 & y_3 & y_4 \\ y_5 & y_6 & y_7 & y_8 \\ y_9 & y_{10} & y_{11} & y_{12} \\ y_{13} & y_{14} & y_{15} & y_{16} \\ & ... & ... & \end{matrix}$$

- Rows represent with-season time series which may be modeled by ARMA(p,q)
- Columns represent between-season time series which may be modeled by ARMA(P,Q)
- p,q = AR, MA orders of the within season model.
- P,Q = AR, MA orders of the between season model.

Order Selection

- STEP 1: Choose d,m,D such that $X_t = (1-B)^d(1-B^m)^D Y_t$ is stationary.
- STEP 2: Examine ACF and PACF plots of $\{X_t\}$ to determine p,P,q,Q

  => p and q are choosen such that $\rho(1), \rho(2), ..., \rho(m-1)$ and $\alpha(1), \alpha(2), ...\alpha(m-1)$ are consistent with ARMA(p,q)

  => P and Q are choosen such that $\rho(km)$ and $\alpha(km)$ for k=1,2,3... are consistent with ARMA(P,Q)

* This procedure will provide sensible first guesses for p,q, P,Q, but optimal orders should be determined via comparison of goodness-of-fit metrics and likelihood ratio tests.