Data Analytics (UE18CS312)
Unit 3

Aronya Baksy

October 2020

1 Time Series Analysis

- Time series data is data on some response variable Y_t observed at various points in time t.
- Time Series data can be univariate (one variable at different points in time) or multivariate (multiple variables at different points in time).

1.1 Components of Time Series Data

- **Trend Component** (T_t): Consistent long-term upward/downward movement of the data.
- **Seasonal Component** (S_t): Fluctuations within a calendar year caused by events that occur at approximately the same time every year (eg: festivals, business practices like end-of-season sale, school holidays).
- **Cyclical Component** (C_t): Cyclical component is fluctuation around the trend line that happens due to macro-economic changes such as recession, unemployment, etc. Periodicity of cyclical fluctuations is not constant, while the periodicity of seasonal fluctuations is constant.
- **Irregular Component** (I_t): Irregular component is the white noise or random uncorrelated changes that follow a normal distribution with 0 mean and constant Standard Deviation.

1.2 Time Series Models

- An additive model is of the form:
$$Y_t = T_t + S_t + C_t + I_t$$
- This assumes that S_t and C_t are independent of T_t, which is often not a valid observation in the real world.
- A multiplicative model is of the form:
$$Y_t = T_t S_t C_t I_t$$
- These are more often used, and closer to the real world behaviour. The simpler form
$$Y_t = T_t S_t$$
- Additive models are appropriate if S_t is fixed while the mean of the data changes, while multiplicative models are more appropriate if S_t is correlated with the mean of the data.

1.3 Error Metrics in TSA

- **Mean Absolute Error**: Given by
$$MAE = \sum_{t=1}^{n} \frac{|Y_t - F_t|}{n}$$
- **Mean Absolute Percentage Error**: Given by
$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y_t - F_t|}{Y_t}$$
- **Mean Squared Error**: Given by
$$MSE = \frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2$$
- **Root Mean Squared Error**: Given by
$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2}$$

2 Forecasting Techniques

2.1 Moving Average Method

- The future value is an average (weighted or not) of the past N data points in the series.
- The **Simple Moving Average** (SMA) is represented as:
$$F_{t+1} = \frac{1}{n} \sum_{k=t+1-N}^{t} Y_k \tag{1}$$
- The **Weighted Moving Average** is given as:
$$F_{t+1} = \sum_{k=t+1-N}^{t} W_k Y_k \tag{2}$$
where W_k is the weight given to the kth point in the series.
- In a weighted moving average, past observations are given differential weights (usually the weights decrease as the data becomes older). The last N weights must follow the condition
$$\sum_{k=t+1-N}^{t} W_k = 1$$

2.2 Single Exponential Smoothing

- Single Exponential smoothing allows the older samples to be assigned smaller decaying weights. The formula is given by
$$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t \tag{3}$$
- Recursively expanding this, we get
$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + ... + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t F_1$$

- The initial forecasted value F_1 can be assumed equal to Y_1.
- SES has the advantages that it uses the entire historical data (unlike MA that uses only the last N data points), and that it assigns decreasing weights to older weights.
- The disadvantages of SES are that it lags behind the trend as it uses past observations, and that forecast bias and systematic errors occur when the observations exhibit strong trend or seasonal patterns.

2.3 Double Exponential Smoothing (Holt's Method)

- Double exponential smoothing uses two equations to forecast the future values of the time series, one for forecasting the level (short term average value) and another for capturing the trend.
- The **Level Equation** is given as:

$$L_t = \alpha Y_t + (1-\alpha)F_t \tag{4}$$

- The **Trend Equation** is given as:

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{5}$$

- The forecast at time $t+1$ and time $t+n$ is given as:

$$F_{t+1} = L_t + T_t$$
$$F_{t+n} = L_t + nT_t$$

2.4 Triple Exponential Smoothing (Holt-Winter's Model)

- This method is used to forecast for time series where seasonal components are present. It uses threee equations for the level, the trend and the seasonal components.
- Let c is the number of seasons (if it is monthly seasonality, then $c = 12$; in case of quarterly seasonality $c = 4$; and in case of daily data $c = 7$).
- The **Level Equation** is given as:

$$L_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)(L_{t-1} + T_{t+1}) \tag{6}$$

- The **Trend Equation** is given as:

$$T_t = \beta(L_t + L_{t-1}) + (1-\beta)T_{t-1} \tag{7}$$

- The **Seasonal Equation** is given as:

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-c} \tag{8}$$

- The forecast at time $t+1$ is given as:

$$F_{t+1} = (L_t + T_t) \times S_{t+1-c} \tag{9}$$

- The inital value of L_t can be calculated as either

$$L_t = Y_t$$

or

$$L_t = \frac{1}{c}(Y_1 + Y_2 + Y_3 + ... + Y_c)$$

- The initial value of T_t is calculated as the average of seasonal indices, given by:

$$T_t = \frac{1}{c}\left(\frac{Y_t - Y_{t-c}}{12} + \frac{Y_{t-1} - Y_{t-1-c}}{12} + .. + \frac{Y_{t-c+1} - Y_{t-2c+1}}{12}\right)$$

- The following algorithm is used to calculate the initial value for S_t.
  1. Calculate season-wise averages $\overline{Y}_1, \overline{Y}_2, \overline{Y}_3, .., \overline{Y}_c$.
  2. Calculate the average of season averages $\overline{\overline{Y}}$
  3. The seasonality index of the season k is given as $\overline{Y}_k / \overline{\overline{Y}}$

3 Regression Models for Forecasting

- The forecasted value at time t, F_t can be written as:

$$F_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + ... + \beta_k X_{kt} + \varepsilon_t \tag{10}$$

  where $X_{1t}, X_{2t}, ...$ are the values of the variables X_1, X_2 at time t.

- All regression models must be passed through the Durbin-Watson test to prove that there is no auto-correlation.

3.1 Regression Model for Forecasting with Seasonality

- Divide each data point by the seasonality index of its season ($Y_{d,t} = Y_t / S_t$). This produces a de-seasonalized data $Y_{d,t}$ from the original Y_t
- Build a regression model on the de-seasonalized data.
- Forecast for time $t+1$ is $F_{t+1} = F_{d,t+1} * S_{t+1}$

3.2 Conditions for a stationary time series

- The mean values of Y_t at different values of t are constant.
- The variances of Y_t at different time periods are constant (Homoscedasticity).
- The covariances of Y_t and Y_{t-k} for different lags depend only on k and not on time t.

4 Auto-Regressive Models

- Auto regression is regression of a variable on itself measured at different time points.
- An AR(1), meaning Auto regression with lag 1, process can be written as

$$Y_{t+1} = \beta Y_t + \varepsilon_{t+1} \tag{11}$$

- A general $AR(p)$ process with lag p is written as

$$Y_{t+1} = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + ... + \beta_p Y_{t-p+1} + \varepsilon_{t+1} \tag{12}$$

- This equation can be rewritten as

$$Y_{t+1} - \mu = \beta^t (Y_0 - \mu) + \sum_{k=1}^{t-1} (\beta^{t-k} \varepsilon_k) + \varepsilon_{t+1} \tag{13}$$

- The OLS estimate of β is

$$\beta = \frac{\sum_{t=2}^{n}(Y_t - \mu)(Y_{t-1} - \mu)}{\sum_{t=2}^{n}(Y_{t-1} - \mu)^2}$$

4.1 Model Parameter Estimation for AR Model

- The *auto correlation* between k lags of Y is given as

$$\rho_k = \frac{\sum_{t=k+1}^{n}(Y_{t-k} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

- It is the correlation between Y_t and Y_{t-k}, ie. different values of Y that are k time apart.
- A plot of this auto correlation value ρ_k for different integer values of k is called the **autocorrelation function (ACF)**
- The *partial autocorrelation* ρ_{pk} is the correlation between Y_t and Y_{t-k}, ie. different values of Y that are k time apart, with the intermediate values $Y_{t-1}, Y_{t-2}, .., Y_{t-k+1}$ removed (partial out).
- A plot of partial auto correlation with differnet values of k is called the **Partial Autocorrelation Function (PACF)**
- The hypothesis test for ACF is:
  - **Null Hypothesis** H_0: $\rho_k = 0$
  - **Alternate Hypothesis** H_1: $\rho_k \neq 0$
- The hypothesis test for PACF is:
  - **Null Hypothesis** H_0: $\rho_{pk} = 0$
  - **Alternate Hypothesis** H_1: $\rho_{pk} \neq 0$
- The null hypotheses are rejected when $\rho_k > 1.96/\sqrt{n}$ and $\rho_{pk} > 1.96/\sqrt{n}$. The lines in PACF and ACF plots represent these confidence limits
- Rule to determine the value of p in $AR(p)$ from ACF and PACF plots are:
  - In PACF, $\rho_{pk} > 1.96/\sqrt{n}$ for the first p lags, then cuts off to 0.
  - The ACF decreases exponentially.

5 Moving Average Process

- A moving average process $MA(q)$ of lag q is written as

$$Y_{t+1} = \mu + \alpha_1\varepsilon_t + \alpha_2\varepsilon_{t-1} + ... + \alpha_q\varepsilon_{t-q+1} + \varepsilon_{t+1}$$

-
- Rule to determine the value of q in $MA(q)$ from ACF and PACF plots are:
  - In ACF, $\rho_k > 1.96/\sqrt{n}$ for the first q lags, then cuts off to 0.
  - The PACF decreases exponentially.

6 ARMA Model

- $ARMA(p,q)$ is a linear combination of $AR(p)$ and $MA(q)$. The general form of $ARMA(p,q)$ is:

$$Y_{t+1} = (\beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + ... + \beta_p Y_{t-p+1} + \varepsilon_{t+1}) + (\mu + \alpha_1 \varepsilon_t + \alpha_2 \varepsilon_{t-1} + ... + \alpha_q \varepsilon_{t-q+1} + \varepsilon_{t+1}) + \varepsilon_{t+1}$$

- Rule to determine the value of p, q in $ARMA(p,q)$ from ACF and PACF plots are:
  - In ACF, $\rho_k > 1.96/\sqrt{n}$ for the first q lags, then cuts off to 0.
  - In PACF, $\rho_{pk} > 1.96/\sqrt{n}$ for the first p lags, then cuts off to 0.

7 ARIMA Model

- Autoregressive *Integrated* Moving Average (ARIMA) is a model that is used for non-stationary data.
- Model building of $ARIMA(p,d,q)$:
  1. Plot ACF and PACF
  2. If series is stationary, then model is $ARIMA(p,0,q)$, the same as $ARMA(p,q)$.
  3. If series is not stationary, find the order of differencing d required to make the series stationary. The model is $ARIMA(p,d,q)$

7.1 Differencing Method

- Differencing is a method used to convert the non-stationary time series into stationary.
- For $d = 1$, the first difference is given as

$$\nabla Y_t = Y_t - Y_{t-1}$$

- The second order difference, corresponding to $d = 2$ is given as:

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = Y_t - 2Y_{t-1} + Y_{t-2}$$

- In most cases, $d \leq 2$ is sufficient.

7.2 Dickey-Fuller Test for Stationarity

- Considering the $AR(1)$ process below,

$$Y_{t+1} = \beta Y_t + \varepsilon_{t+1}$$

- The Dickey Fuller test is a test for stationarity of the time series. $AR(1)$ can now be written as

$$Y_{t+1} - Y_t = (\beta - 1)Y_t + \varepsilon_{t+1} = \psi Y_t + \varepsilon_{t+1}$$

  - **Null Hypothesis** H_0: $\psi = 0$ (Non-stationary)
  - **Alternate Hypothesis** H_1: $\psi < 0$ (Stationary)

- The Dickey-Fuller test statistic is given as

$$D = \frac{\psi}{S_e(\psi)}$$

7.3 Augmented Dickey-Fuller Test

- Instead of only one lag, p lags of the $AR(1)$ process are considered as when ε_{t+1} is not entirely random (white noise), then there may be more significant lags.
- Now the $AR(1)$ is written as

$$Y_{t+1} - Y_t = \psi Y_t + \sum_{i=0}^{p} \alpha_i Y_{t-i} + \varepsilon_{t+1}$$

  - **Null Hypothesis** H_0: $\psi = 0$ (Non-stationary)
  - **Alternate Hypothesis** H_1: $\psi < 0$ (Stationary)

7.4 Ljung-Box Test for Autocorrelation

- The hypothesis are:
  - **Null Hypothesis** H_0: Model does not show lack of fit
  - **Alternate Hypothesis** H_1: Model shows lack of fit
- The test statistic Q is given as:

$$Q = n(n+2) \sum_{k=1}^{m} \frac{\rho_k^2}{n-k} \tag{14}$$

  where n is the lengtjh of the time series, ρ_k is the autocorrelation value for k lags, and m is the total number of lags.
- Q-statistic is an approximate chi-square distribution with $m - p - q$ degrees of freedom where p and q are the AR and MA lags.

7.5 Theil's Coefficient: Power of a forecasting Model

- The power of forecasting model is a comparison between Naïve forecasting model and the model developed.
- In the Naïve forecasting model, the forecasted value for the next period is same as the last period's actual value ($F_{t+1} = Y_t$).
- Theil's coefficient U is the ratio of MSE of forecasting model to the naïve model.

$$U = \frac{\sum_{i=1}^{n} (Y_{t+1} - F_{t+1})^2}{\sum_{i=1}^{n} (Y_{t+1} - Y_t)^2} \tag{15}$$

- The value of $U < 1$ indicates that forecasting model is better than the Naïve forecasting model, while $U > 1$ indicates that the forecasting model is not better than Naïve model