Time Series Primer

Forecasting Problems

- Short-term: predicting only a few periods ahead. Typically based on modelling and extrapolating patterns in the data
- Medium-term: one to two years into the future
- Long Term: several years into the future

Stochastic Process and Time Series: sample realization of a stochastic process is a time series

First Order Properties

- ***Data Inspection***
- ***Trend***
  - Long-run behavior of the time series
  - Deterministic: evolves in a perfectly predictable way
  - Stochastic: evolves in an unpredictable way. (Stochastic process)
  - Time is treated as a dummy variable
  - A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we will refer to a trend as "changing direction," when it might go from an increasing trend to a decreasing trend.
- ***Seasonality***
  - Additive: works well when seasonality behavior is constant. Example: retail sales in December
    - Y_adjusted = y_t – S_t
    - In an additive timeseries, the components add together to make the time series. If you have an increasing trend, you still see roughly the same size peaks and troughs throughout the time series.
  - Multiplicative: seasonal fluctuations vary with time
    - Y_adjusted = y_t/S_t
    - In a multiplicative timeseries, the components multiply together to make the time series. If you have an increasing trend, the amplitude of seasonal activity increases. Everything becomes more exaggerated. This is common when you're looking at web traffic.
  - A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known frequency. The monthly sales of antidiabetic drugs above shows seasonality which is induced partly by the change in the cost of the drugs at the end of the calendar year.
- ***Cycles***
  - After taking away trend and seasonality, we are left with cycles
  - To observe cycles, we need long term data
  - Deterministic: common in engineering, physical sciences etc
  - Stochastic: economics, business

- A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency. These fluctuations are usually due to economic conditions and are often related to the "business cycle." The duration of these fluctuations is usually at least 2 years.

<u>Note on Seasonality vs Trend:</u> Many people confuse cyclic behavior with seasonal behavior, but they are really quite different. If the fluctuations are not of a fixed frequency, then they are cyclic; if the frequency is unchanging and associated with some aspect of the calendar, then the pattern is seasonal. In general, the average length of cycles is longer than the length of a seasonal pattern, and the magnitudes of cycles tend to be more variable than the magnitudes of seasonal patterns.

<u>Second Order Properties</u>

- ***<u>Series Decomposition</u>***
  - Data is made of a trend, seasonal and irregular component. When they are added together, we can forecast.
- **<u>Autocorrelation Functions</u>**
  - **<u>Note</u>**
    - **<u>Autocovariance Function</u>**
      -
  - Running y_t with multiple lags
  - ***<u>Autocorrelation Functions (ACF)</u>***
    - autocorrelation function is a measure of the correlation between observations of a time series that are separated by k time units (yt and yt–k).
    - Bars that extend across the red line are statistically significant.
  - ***<u>Partial Autocorrelation Function (PACF)</u>***
    - The partial autocorrelation function is similar to the ACF except that it displays only the correlation between two observations that the shorter lags between those observations do not explain
    - partial correlation for each lag is the unique correlation between those two observations after partialling out the intervening correlations.
    - partial autocorrelation function (PACF) is more useful during the specification process for an autoregressive model
  - ***<u>Sample Partial Autocorrelation</u>***
    -
- **<u>Stationarity</u>**
  - Stationary series: showcase variation around the mean
    - First Order Strongly Stationary – Unlikely
      - Same means and same variances
      - Example: radio waves
    - First Order Weakly Stationary
      - Only requirement is that it is mean reverting
      - Take first difference data – to make stationary

        - Second Order Weakly Stationary (Covariance Stationarity)
            - Mean reverting and volatility is constant
            - Time independent covariances
            - Take log of the data, an then take first difference
    - Non-stationary series: does not move around the mean
        - Transformation of nonstationary processes
            - First Difference of data – for first order weakly stationary
            - Take log(y_t) – for second order weakly stationary
                - Then take first difference for transformed series
        - The Augmented Dickey Fuller Test (ADF) is unit root test for stationarity
- **Lag Operator**

Forecasting Environments

- *Estimation Sample:* This sample is used for estimating the model and respective
- parameters.
- *Prediction Sample*: This sample is used to assess the accuracy of the forecast
- Forecasting Methods
    - Recursive
    - Rolling
    - Fixed
    - Note: *We can use CV for more a robust model assessment*

Forecasting Challenges

- Lack of understanding of the phenomenon
- Lack of statistical methods
- High uncertainty
- Lack of integration of skills

Recursive Scheme

- Divide data into estimation and prediction
- Use estimation data to predict data
- Then train on a period y(0) to y(n) then predict y_hat(n+1). Then we train on y(0) to y(n+1) and predict y_hat(n+2) and so forth. ***The window we train on gets bigger, and we do one-step ahead predictions.***

Rolling Scheme

- we train on a period y(0) to y(n) then predict y^(n+1). Then we train on y(1) to y(n+1) and predict y^(n+2) and so forth. The ***size of the window we train on stays the same size, and we do one-step ahead predictions.***

Fixed Scheme

- train on y(0) to y(n), then predict y^(n+1) to y^(n+m)
- For subsequent forecasts we continue to use that estimate.
- In this scheme we estimate the model parameter once only, for the first forecast period. To be precise we use observations 1 to R to obtain βˆ1,R where the subscript reflects the observations on the basis of which the estimate is obtained. For subsequent forecasts we continue to use that estimate.

Forecasting Performance

- **Mean Absolute Error/Deviation (MAE):** MAE is defined as the average of absolute difference between forecasted values and true values.

$$MAE = \frac{1}{n}\sum_{t=1}^{t=n} |y' - y|$$

  -
  - *Interpretation and limitation*: The lower the MAE value, the better the model is; a value of zero means there is no error in the forecast. In other words, when comparing multiple models, the model with the lowest MAE is considered better.
  - However, MAE does not indicate the relative size of the error and it becomes difficult to differentiate big errors from small error

- **Mean Absolute Percentage Error (MAPE):** MAPE is defined as the percentage of the average of absolute difference between forecasted values and true values, divided by true value.

$$MAPE = \frac{1}{n}\sum_{t=1}^{t=n} \frac{|y' - y|}{y} * 100\%$$

  -
  - Because of the in denominator, it works best with data without zeros and extreme values. If this value is extremely small or large, the MAPE value also takes an extreme value.

- *Interpretation and limitation*: The lower the MAPE, the better the model is. Keep in mind that MAPE works best with data without zeros and extreme values. Like MAE, MAPE also understates the impact of large but infrequent errors due to extreme values.

- **Mean Squared Error (MSE):** MSE is defined as the average of squares of the error3. It is also defined as the metric that assesses the quality of forecasting model or predictor. MSE also incorporates both the variance (the spread of predicted values from each other) and bias (the distance of predicted value from its true value).

$$MSE = \frac{1}{n}\sum_{t=1}^{t=n}(y' - y)^2$$

  -
  - MSE is almost always positive and values closer to zero are better. Because of the square term (as seen in the formula above), this metric penalizes large errors or outliers more compared to the small errors
  - *Interpretation and limitation*: The closer to zero MSE is, the better. While it solves the extreme value and zero problem of MAE and MAPE, it might be disadvantageous in some cases.

- **Mean Forecast Error (MFE or Bias):** Measures average deviation of forecast from actuals.

$$MFE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)$$

Model Selection

- There are many metrics used for model selection such as e.g., MSE, AIC, SIC, Mallows CP, etc.
- As the number of parameters increases, the MSE performance deteriorates (overfitting)!
- MSE = Var + Bias^2
    - Done at training and testing level
    - Need to include penalty for adding more predictors

Forecasting Trend

- Point Forecast
- Interval Forecast
- Density Forecast
- Quadratic Trend
- Keep trend simple – linear and quadratic

Seasonality

- Seasonal Pattern: Is a pattern that repeats itself every year.
- Deterministic Seasonality: When the annual repetition is exact.
- Stochastic Seasonality: When the annual repetition is approximate.
- Sources of Seasonality: links to the calendar, technologies, preferences, institutions, etc.

Modelling and Forecasting Seasonality

- *Seasonal Dummy Variables*: Indicate which season we are in
- *Seasonal Factors*: coefficients of the seasonal dummy variables
- *Pure Seasonal Dummy Model*:

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- *Seasonal Dummy Model Including Linear Trend*

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

White noise is not predictable. A time series is white noise if the variables are independent and identically distributed with a mean of zero.This means that all variables have the same variance (sigma^2) and each value has a zero correlation with all other values in the series

Moving Average Models

- A moving average model uses past forecast errors in a regression-like model.
- Moving-average (MA) models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant.
- Moving average processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time.
- **The MA model is a simple extension of the white noise series**
- **Invertibility:** An MA(1) process is invertible if $|\theta| < 1$. Otherwise, if $|\theta| \geq 1$, the process is noninvertible.
- **Use ACF for understanding the degree of the model**

Using ACF and PACF to choose model order

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Autoregressive Models (Use PACF to understand model)

- Autoregressive processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock.
- AR processes are stochastic difference equations. They are used for modeling discrete-time stochastic dynamic processes (among others).
- **ACF:**
  - We would expect to see that ρ1 = p1 = φ, and all others decay to zero according to ρk= φk.
  - ACF decaying to zero, ***if persistence is small (phi is small), decay to zero is faster. (for e.g phi = 0.4)***
    - ***If persistence is large (for e.g phi = 0.95), slower decay to zero***
- **PACF:**
  - We would expect to see only 1 spike different from zero, i.e., p1 = φ, and all others equal to zero (pk=0, k>1). **For AR(1)**
  - ***A necessary and sufficient condition for an AR(1) process*** Yt = c +φYt-1 + εt to be covariance stationary is that |φ|<1.
  - ***A useful Property for an AR(2) Process: The necessary conditions for an AR(2) process to be covariance stationary are that*** -1<φ2<1 and -2<φ1<2,and the sufficient conditions are that φ1+φ2<1 and φ2-φ1<1.
- For AR(2) unconditional mean is given by:

$$\boxed{\mu = \frac{c}{1 - \phi_1 - \phi_2}}$$

Seasonal Autoregressive Models S-AR(p)

- For Seasonal AR
  - **Use PACF to understand AR(q)**
  - **Use ACF to understand at what duration spikes occur (monthly, quarterly, daily etc.)**

Seasonal Moving Average Models S-MA(q)

- For Seasonal MA
  - **Use ACF to understand MA(p)**
  - **Use PACF to understand at what duration spikes occur**

ARMA Models

- The ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization.
- For an ARMA(p,q) process, lower-order models are better. For example, ARMA(1,1) is better than AR(3).

Auto ARIMA – very little flexibility to incorporate subject matter expertise

CUSUM

On a tabular CUSUM chart, look for the following:

- Upward or downward trends in the upper and lower CUSUMs. The plotted points should fluctuate randomly around zero. If an upward or downward trend develops, the process mean has shifted and the process may be affected by special causes.
- Plotted points that are located beyond the control limits, which indicates that the process is out of control

Recursive estimates provide information about parameter stability

---

Holt-Winters Method

- decomposes the series into level, trend, seasonality. Future values are predicted by combining these systematic factors based on recent history. The intuitive idea here is that the future will behave very similar to recent past, we just have to find how much of the past is relevant. The three systematic components are:
    - Level, (alpha): Average value around which the series varies. For a seasonal time series, level is obtained by first de-seasonalizing the series and then averaging. Alpha value determines how much of the past to consider and is between [0,1]. alpha=1 means give importance only to the last data point (naive forecast)
    - Trend, (beta): Trend is how the level changes over time. Similar to alpha, a beta value closer to 1 indicates the model is considering only the recent trend. Trend also has a damping factor (phi) which determines how much of the recent trend to 'forget'. Consider it as a de-rating factor on trend.
    - Seasonality (gamma): This factor models how the series behaves in each time period for full season.

- method is called "Exponential" because each of the above factors give exponential weightage to the past values.
- Additive model = (Level + Trend) + Seasonality
- Multiplicative Model = (Level Trend) Seasonality

ETS

- ETS standards for **Error, Trend, Seasonality model**. It is similar to Holt-Winter's model above but with a State Space statistical framework. In Holt Winter's, the time series is decomposed into trend and seasonality and exponential weights are used to make the forecast. **In a State Space approach, the underlying statistical process is identified, and errors are factored in to make the forecast**
- ETS models follow a taxonomy of ETS(XYZ) where:
  - X:Error (aka innovations). It can be Additive (A) or Multiplicative (M)
  - Y:Trend. Trend component can be No trend (N), additive (A), Multiplicative (M) or damped (Ad)
  - Z:Seasonality, Null (N), additive (A) or multiplicative (M)

Facebook Prophet

- Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well
- Prophet was built for high frequency data like daily, hourly, minute etc.. It may not work very well on monthly, quarterly data, but you won't know until you try.
- In addition to forecasting, it also provide changepoints, anomalies which are great for detecting sudden changes in the time series
- ETS/HW & SARIMA cannot work with multiple seasonalities & high frequency data. Prophet can also include effect of holidays.
- Prophet requires the data to be in specific format. Dataframe must have time column ds and time series observations in column y
- Though Prophet is designed mainly for high frequency data, it can be used for monthly/quarterly/yearly data with some tweaks.