Assignment-2 (Experiential Learning)

Course No: MATH F432 (Applied Statistical Methods)
Weightage: 10%
Last Date of Submission: 27 November, 2020

Problem 1: Data Analysis for Investments and Portfolio Management

Preparation files:
1. Visit www1.nseindia.com
2. Locate "products" tab, and then under "capital markets" click on "indices"
3. On this page click on "+Historical Data"
4. Then, click on "search" belonging to "Historical Index Data"
5. Select Index :: Nifty 50 // Select a time period 02.04.2018 to 31.03.2019 -- Click "Get Data"; A data set will be displayed. Scroll down and click on "download file in csv format"
6. Repeat the above process to also get data for 01.04.2019 to 30.03.2020. Now you merge the dataset to get two years of data.
7. A file "data.csv" will be obtained containing two years of data.
9. Go back to Step 2 above :: Now click on "+About indices"
10. Click on "Broad Market Indices"
11. Click on "Nifty 50 Index"
12. Download these three files and save them on your computer
(a) Methodology (.pdf)
(b) List of NIFTY 50 stocks (.csv)
(c) Fact Sheet of NIFTY 50 (.pdf)
*(a) and (c) are for our general information.
13. Open .csv file containing the list of NIFTY 50 stocks and find and note down the "symbol" of "Wipro Ltd."
14. Go back to Step 2 above. This time click on 'equities' and then on "+Historical Data" and then on "search" under "Security-wise Price/Volume Archives"
15. Now get historical data for "all" for period 02.04.2018 to 30.03.2020 (as previous) and save this as .csv file on your computer.
16. Now locate "live market" tab, go to "live analysis", and then click on "top ten gainers and losers"
17. Browse through other tabs and details to enrich your knowledge

Guidelines for report writing: Assume that you have to explain the importance, rationale, and methodology of stock exchange market to a layman who just have some theoretical exposure of Statistics. Therefore, prepare a detailed group report (limited to about 5000 words, or about 8 pages) in a lucid yet technical manner including tables/graphs that must include (but not limited to)

1. What are NSE and Nifty 50 Index? What is the index methodology? (explain basics of share market)
2. Why do we need to invest? Does it increase my stress level? Why shouldn't I deposit my money in a reputed bank, and get the nominal interest without any worry?
3. How can one be an effective investor (based on your analysis from the historical data and associated information)? Shall I invest some amount daily, or in a particular day of a month, or once in a month? Is the stock market operational daily, 24 hours? If not, at what time it becomes operational and when (e.g., early or late hours) should I invest? In which part of industry (IT, Telecom, …), should I invest? Shall I target to become a domestic or an international investor?
4. Is it possible to determine the underlying best-fit probability distribution for the opening/closing prices of WIPRO? Using graphical tools, when you plot the open/close prices of WIPRO over time, do you see any trend/pattern? You are free to use any statistical software should there be any requirement.
5. What is a portfolio? What is the concept of portfolio management? (No need to demonstrate in details)

Problem 2: Statistical Analysis and Forecasting of Solar Energy (Intra-State)

Preparation files and guidelines:

1. Visit www.mnre.gov.in and understand about various renewable energy resources.
2. Note that our aim is to analyse solar energy for five study regions of the same state. If possible, we may like to forecast solar energy for next week or even for next month.
3. Obtain 2000-2014 hourly data of five solar parks for the state of Rajasthan (visit MATHF432 google drive).
4. Look at various terms such as DHI, DNI, GHI, dew point, temperature, pressure, relative humididty, wind speed, etc. Understand which terms/parameters are relevant to solar energy. In fact, we shall use GHI data for time-series analysis.
5. Perform several graphical plots or compute various descriptive statistics to understand the DHI, DNI, and GHI data, their correlation, etc.
6. Now let us concentrate only on the GHI data. Does the dataset (for a solar park) exhibit a normal distribution? If not, identify the underlying probability distribution of the GHI data for a park.
7. Let us now use a time–series analysis to GHI data. Do you see any trend or seasonality in the GHI data? Could you decompose this time series into various components? Is the data stationary?
8. Use AR, MA, ARMA, ARIMA, and SARIMA to the GHI data for daily/weekly forecasting.
9. Discuss briefly how you have validated your results of time-series forecasting.
10. **Optional**: Can we use some other technique (say, machine learning) to forecast the solar energy?

Problem 3: Statistical Analysis and Forecasting of Solar Energy (Inter-states)

Preparation files and guidelines:

1. Visit www.mnre.gov.in and understand about various renewable energy resources.
2. Note that our aim is to analyse solar energy for five study regions of the same state. If possible, we may like to forecast solar energy for next week or even for next month.
3. Obtain 2000-2014 hourly data of four solar parks situated in Rajasthan, MP, Andhra Pradesh, and TN (visit MATHF432 google drive).
4. Look at various terms such as DHI, DNI, GHI, dew point, temperature, pressure, relative humididty, wind speed, etc. Understand which terms/parameters are relevant to solar energy. In fact, we shall use GHI data for time-series analysis.
5. Perform several graphical plots or compute various descriptive statistics to understand the DHI, DNI, and GHI data, their correlation, etc.
6. Now let us concentrate only on the GHI data. Does the dataset (for a solar park) exhibit a normal distribution? If not, identify the underlying probability distribution of the GHI data for a park.
7. Let us now use a time–series analysis to GHI data. Do you see any trend or seasonality in the GHI data? Could you decompose this time series into various components? Is the data stationary?
8. Use AR, MA, ARMA, ARIMA, and SARIMA to the GHI data for daily/weekly forecasting.
9. Discuss briefly how you have validated your results of time-series forecasting.
10. **Optional**: Can we use some other technique (say, machine learning) to forecast the solar energy?

Problem 4: Statistical Analysis and Forecasting of Wind Energy (Intra-State)

Preparation files and guidelines:

1. Visit www.mnre.gov.in and understand about various renewable energy resources.
2. Note that our aim is to analyse wind energy for five study regions of the same state. If possible, we may like to forecast wind energy for next week or even for next month.
3. Obtain 2000-2014 hourly wind speed data of five locations in the state of Rajasthan (visit MATHF432 google drive).
4. Look at various terms such as DHI, DNI, GHI, dew point, temperature, pressure, relative humididty, wind speed, etc. Understand which terms/parameters are relevant to wind energy. In fact, we shall use wind speed data for time-series analysis.
5. Perform several graphical plots or compute various descriptive statistics to understand the wind speed data and their correlation to other parameters.
6. Now let us concentrate only on the wind speed data. Does the dataset (for a location) exhibit a normal distribution? If not, identify the underlying probability distribution of the wind speed data for a location.
7. Let us now use a time–series analysis to the wind speed data. Do you see any trend or seasonality in the data? Could you decompose this time series into various components? Is the data stationary?
8. Use AR, MA, ARMA, ARIMA, and SARIMA to the wind data for daily/weekly forecasting.
9. Discuss briefly how you have validated your results of time-series forecasting.
10. **Optional**: Can we use some other technique (say, machine learning) to forecast the wind energy?

Problem 5: Statistical Analysis and Forecasting of Wind Energy (Inter-states)

Preparation files and guidelines:

1. Visit www.mnre.gov.in and understand about various renewable energy resources.
2. Note that our aim is to analyse wind energy for five study regions of the same state. If possible, we may like to forecast wind energy for next week or even for next month.
3. Obtain 2000-2014 hourly wind speed data of four locations from the state of Rajasthan, MP, Andhra Pradesh, and TN (visit MATHF432 google drive).
4. Look at various terms such as DHI, DNI, GHI, dew point, temperature, pressure, relative humididty, wind speed, etc. Understand which terms/parameters are relevant to wind energy. In fact, we shall use wind speed data for time-series analysis.
5. Perform several graphical plots or compute various descriptive statistics to understand the wind speed data and their correlation to other parameters.
6. Now let us concentrate only on the wind speed data. Does the dataset (for a location) exhibit a normal distribution? If not, identify the underlying probability distribution of the wind speed data for a location.
7. Let us now use a time–series analysis to the wind speed data. Do you see any trend or seasonality in the data? Could you decompose this time series into various components? Is the data stationary?
8. Use AR, MA, ARMA, ARIMA, and SARIMA to the wind data for daily/weekly forecasting.
9. Discuss briefly how you have validated your results of time-series forecasting.
10. **Optional**: Can we use some other technique (say, machine learning) to forecast the wind energy?

Problem 6: Earthquake Forecasting in Indonesia: Analysis of Interevent Time Distribution

Preparation files and guidelines:

1. Earthquake interevent time analysis helps up to assess earthquake hazards of a seismic region. Several probability distributions are used for this purpose. Before we proceed, let us first understand a few basic concepts about earthquakes. Visit https://earthquake.usgs.gov/education/ for details.
2. What is an earthquake? Is Indonesia prone to earthquakes? What do you mean by earthquake prediction? Is earthquake prediction same as earthquake forecasting? Find at https://www.usgs.gov/faqs/can-you-predict-earthquakes?qt-news_science_products=0#qt-news_science_products
3. Now let us collect previous earthquakes for the Sumatra and adjacent regions. For this, visit http://www.isc.ac.uk/iscbulletin/search/catalogue/, in ISC Bulletin → CSV formatted catalogue → rectangular search region, latitude -10 to 10 deg N, longitude 90-110 deg E → time period 1900.01.01 to 2020.10.20 → additional parameters, depth 0-200 km, magnitude 6-10, magnitude type 'any', magnitude author ISC →output event catalogue. Thus you get a list of earthquakes that occurred in the region after 1900.
4. In order to remove the dependent events, such as foreshocks, aftershocks and seismic clusters, let us apply a dynamic window-based spatio-temporal filtering algorithm as:

   Search radius $r = \exp(-1.024 + 0.804M) \pm 15,$ and

   time window $t = \exp(-2.870 + 1.235M) \pm 60.$ If there is any event falling within the search radius and/or within the time window, we remove that event. In this way, we find a catalog of i.i.d events.
5. The interevent times of the declustered catalog can be obtained by subtracting the occurrence time from the next occurrence time. For example, if two events occurred on April 04, 1905 and August 20, 1908, then the difference of these dates will be the interevent time. You can express the interevent times in day or year, as per your choice.
6. Having the list of interevent times, now we are ready for modeling using various probability distributions, such as exponential, gamma, Weibull, and lognormal.
7. Use MLE and MoM parameter estimations to obtain the estimated parameters.
8. Now in order to prioritize the candidate probability distributions, let us apply two tests: AIC and K-S.
9. You may use inbuilt tool of R, MATLAB, or any other software to fit interevent time data to different probability models.
10. Now, having done the above steps, how can you forecast earthquakes? See https://www.tandfonline.com/doi/full/10.1080/19475705.2018.1466730 for reference.

Problem 7: Earthquake Nowcasting in Indonesia: Analysis of Interevent Count Distribution

Preparation files and guidelines:

1. Earthquake nowcasting uses interevent count (natural time) analysis to assess earthquake hazard of a city at current time (see, https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/2016EA000185). Several probability distributions are used to fit the natural time data. Before we proceed, let us first understand a few basic concepts about earthquakes. Visit https://earthquake.usgs.gov/education/ for details.
2. What is an earthquake? Is India prone to earthquakes? What do you mean by earthquake prediction? Is earthquake prediction same as earthquake forecasting? Find at https://www.usgs.gov/faqs/can-you-predict-earthquakes?qt-news_science_products=0#qt-news_science_products
3. Now let us collect previous earthquakes for the Sumatra and adjacent regions. For this, visit http://www.isc.ac.uk/iscbulletin/search/catalogue/, in ISC Bulletin → CSV formatted catalogue → rectangular search region, latitude -10 to 10 deg N, longitude 90-110 deg E → time period 1970.01.01 to 2020.10.20 → additional parameters, depth 0-200 km, magnitude 4-10, magnitude type 'any', magnitude author ISC →output event catalogue. Thus you get a list of earthquakes that occurred in the region after 1970.
4. As the earthquake nowcasting approach can consider dependent events, we no need to remove foreshocks, aftershocks, or any seismic swarms.
5. The interevent counts (natural times) of the catalog can be obtained by finding the cumulative counts of "small" events (say, $4 \le M < 6.5$) between two successive "large" earthquakes (say, $M \ge 6.5$). For example, if two subsequent large events occurred on Dec 15, 2017 and Aug 02, 2019, then the smaller number of earthquakes (say, $4 \le M < 6.5$) is one interevent count. Discard those cases where the interevent count is 0.
6. Having the list of non-zero interevent times, now we are ready to model natural times using various probability distributions, such as exponential, gamma, and Weibull.
7. Use MLE and MoM parameter estimations to obtain the estimated parameters.
8. Now in order to prioritize the candidate probability distributions, let us apply two tests: AIC and K-S.
9. You may use inbuilt tool of R, MATLAB, or any other software to fit natural time data to different probability models.
10. Now, having done the above steps, how can you nowcast earthquake for two cities, Aceh and Medan, for which the current small event counts are 275 and 400, respectively. See https://link.springer.com/article/10.1007/s00024-018-2037-0 for reference.