**IMF MCMCO Financial Sector Expert Test:**

This document is the write-up for the test

**Scripts submitted for each question**:

Question 1: Q1.ipynb and ets\_mod.py and Q1.R

Question 2: Q2\_Q3.ipynb

(DEXUSEU.csv as an input data file)

Question 3: Q2\_Q3.ipynb

(DEXUSEU.csv as an input data file)

Question 4: Q4.ipynb

(Communication.zip as the input data file)

Question 5: Q5.ipynb

(Labeled sentences.csv and New sentences.csv file as input data files)

Q1. General and Statistical Programming in Python and R

Please complete the following tasks. Please do not forget to include your debugged ets\_mod.py in your submission!

**Please check script *Q1.ipynb* , script *ets\_mod.py* , and *Q1.R* for details.**

(a)Retrieve currency in circulation times series from the Bank of Zambia website.

(b)There are bugs and errors in ets\_mod.py. Please correct them as many as possible.

(c)Please complete method plot() in ets\_mod.py. When this method is invoked, the program should produce a line chart with both original data and the ETS model fitted data.

(d)Please write a user case script to instantiate ets\_mod.py. Build an ETS(A,A,A) model (or any other configurations you deem appropriate with explanation) on the dataset retrieved in item (a). In your user case script, please invoke all methods inside the class and comment on the purpose of each method.

All results can be generated by running the code in ***Q1.ipynb*** and ***ets\_mod.py***

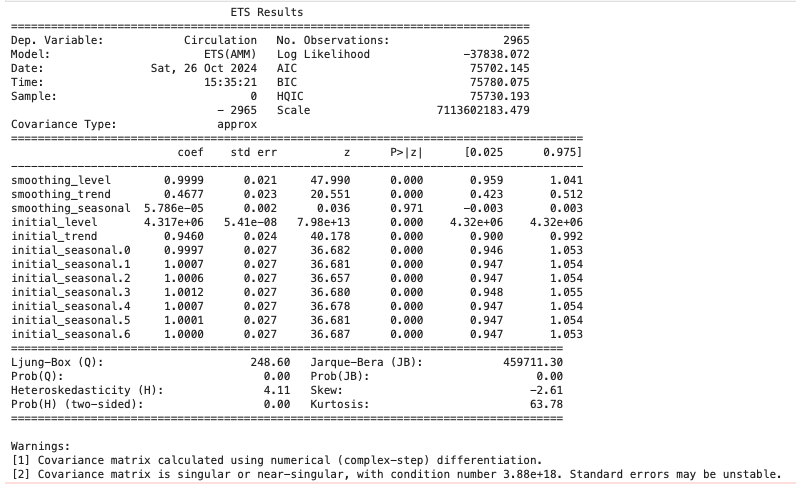
I use the following parameters for ETS(A,A,A) model as an instance:

**trend="mul", seasonal="mul", seasonal\_periods=7**

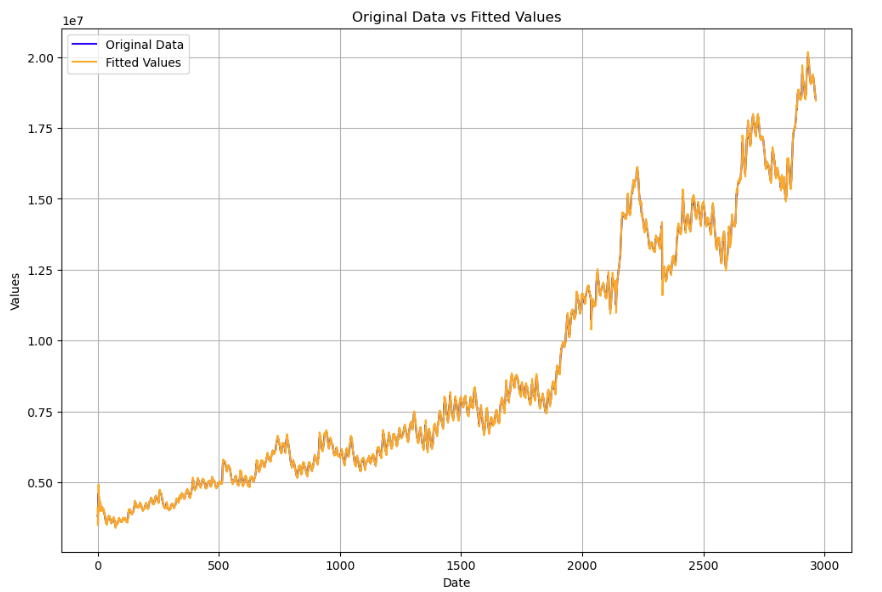
There are 3 methods in ets\_mod.py: **plot(), summary(), forecast(days=7)**

Here is an example output by run the ETS model

**Model summary**: summary() method shows the summary of the model fit



**Model plot**: plot() method shows the plot of original data and the model fitted values



**Model forecast**: please refer to question (e) below for details

(e)Please forecast the currency in circulation for the next week from the end of dataset based on the model above.

The forecast currency in circulation for next week is generated by the model.forecast(days=7) method.

Here are the values:

**7-day forecast: [18422009, 18361521, 18280121, 18223492, 18164051, 18112086, 18060944]**

This can be generated by running the code in ***Q1.ipynb***

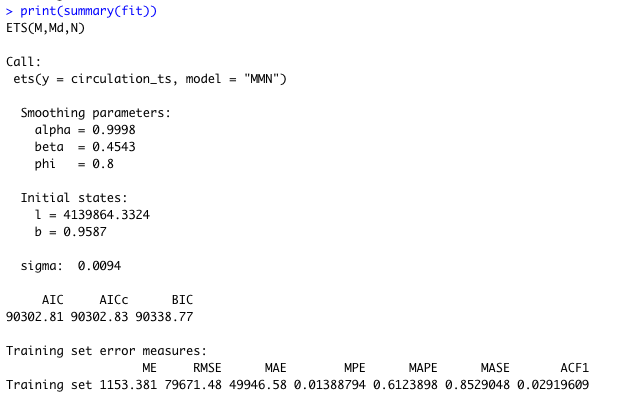
(f)Replicate a) through d) in R. You do not need to write it as a class. However, class realization in R will be awarded more credits.

Please check the script ***Q1.R*** for details.

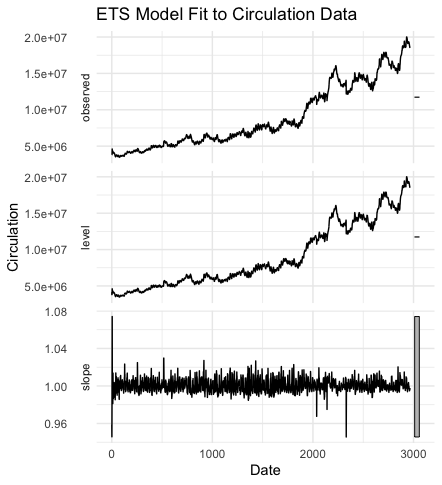
Here are the results: ~ All can be generated by running Q1.R script

Note: I picked up the “MMN” setting for the model this time, it can be changed.

**Model Summary**:



**Model Plot:**



**Model forecast (next 7 days):**

[18463619, 18430906, 18404777, 18383900, 18367215, 18353878, 18343215]

(g)Please forecast the currency in circulation for the next week from the end of dataset using the Long Short-Term Memory (LSTM) supervised machine learning model. Alternatively, you may utilize any other suitable machine-learning model, provided that you justify the choice of that model in comparison to LSTM. You can either invoke this method inside the class in previous user case python script or create a new python script for this question g). You do not need to replicate the codes in R for question g).

I use the LSTM model, please run the code in Q1.ipynb to generate the result.

**7-day forecast: [18452314, 18410356, 18372172, 18336848, 18302896, 18268996, 18234210]**

Q2. Quantitative Modelling in Python or R

Please collect the EUR/USD daily exchange rate from FRED and complete the following tasks related to Value-at-Risk forecasting from a GARCH-type model with a Python or R script:

**Please check the script *Q2\_Q3.ipynb* for details.**

**Note:**

* **Question 2 and Question 3 are all in this script**
* ***DEXUSEU.csv* is an input dataset used by the script**

1. Collect a daily series of a macroeconomic/financial indicator, which will influence the log return of EUR/USD from publicly available source. Please describe your choice and explain why.

**Choice of Indicator**: Effective Federal Funds Rate

**Reason:** The Effective Federal Funds Rate (EFFR) is a determinant factor of USD's value against other currencies, including the EUR. EFFR is set by the Federal Reserve and reflects the US central bank’s monetary policy stance. EFFR is also a primary tool to manage inflation. A higher EFFR makes a stronger dollar, while a lower EFFR weakens the dollar. Interest rate differentials between Japan and US affect the exchange rate of EUR/USD.

1. Run a GARCH-type model on the log return of EUR/USD with the series in a) as an exogenous variable in the mean process.

Here is the model summary:

AR - GARCH Model Results

==============================================================================

Dep. Variable: Log\_Return R-squared: 0.003

Mean Model: AR Adj. R-squared: 0.001

Vol Model: GARCH Log-Likelihood: 4798.46

Distribution: Normal AIC: -9584.92

Method: Maximum Likelihood BIC: -9554.42

No. Observations: 1193

Date: Fri, Oct 25 2024 Df Residuals: 1190

Time: 13:42:13 Df Model: 3

Mean Model

==================================================================================

coef std err t P>|t| 95.0% Conf. Int.

----------------------------------------------------------------------------------

Const -3.4178e-05 1.171e-04 -0.292 0.770 [-2.637e-04,1.953e-04]

Log\_Return[1] 0.0363 2.746e-02 1.323 0.186 [-1.748e-02,9.014e-02]

Log\_Return[2] -0.0380 1.379e-02 -2.756 5.849e-03 [-6.503e-02,-1.098e-02]

Volatility Model

============================================================================

coef std err t P>|t| 95.0% Conf. Int.

----------------------------------------------------------------------------

omega 4.2005e-07 2.089e-11 2.011e+04 0.000 [4.200e-07,4.201e-07]

alpha[1] 0.0500 1.171e-02 4.269 1.964e-05 [2.704e-02,7.296e-02]

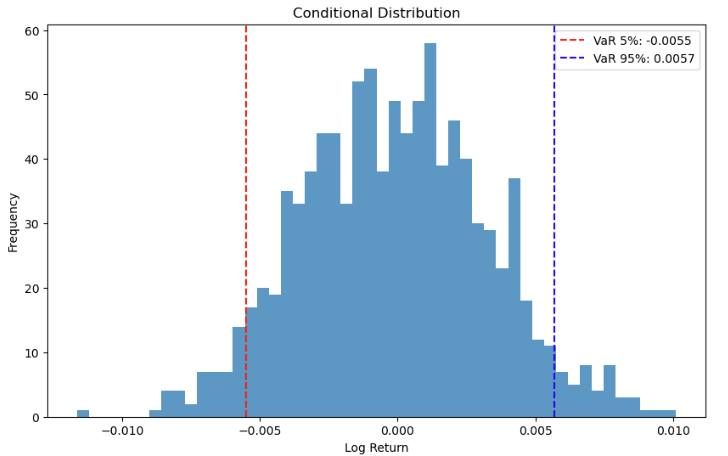
beta[1] 0.9300 1.119e-02 83.074 0.000 [ 0.908, 0.952]

============================================================================

Covariance estimator: robust

1. Based on b), please prepare a plot of conditional distribution curve for a specific date, on which you should indicate the value at risk of 10% in total, both tails. Prepare a summary to interpret this plot.

I fit GARCH model to the log returns EUR/USD exchange rate with the Federal Funds Rate as an exogenous variable. The plot is the conditional distribution for a specific date (choice of date 2023-12-29), indicating the Value at Risk (VaR) at the 10% level (both tails).



The primary goal is to visualize the risk associated with the exchange rate fluctuations and to indicate the Value at Risk (VaR) at the 10% level, considering both tails (5% on each tail).

* The Bell Curve (Histogram) is the simulated log returns on this specific date. The returns are generated based on the conditional variance from the GARCH model.
* The Red line (VaR 5%) and Blue line (VaR 95%): the 5th & 95th percentile of the simulated returns. They represent the lower & upper tail risk, which means there is a 5% probability that the log return will be below or above this value.

For risk management, the plot reflects potential extreme losses and gains. Therefore, we can set risk limits based on this information.

Q3. Machine-learning-based Modelling in Python or R

Instead of using a GARCH-type model for value-at-risk forecasting in Q2, now please use a machine-learning based model to predict the log return of EUR/USD, such as Long Short-Term Memory (LSTM). You can use instead other suitable recurrent neural network, provided that a suitable justification is given. You can choose to include your selected exogenous indicator from Q2 or not in your answer.

**Please check the script *Q2\_Q3.ipynb* for details.**

**Note:**

* **Question 2 and Question 3 are all in this script**
* ***DEXUSEU.csv* is an input dataset used by the script**

I use the LSTM model. Please run the script to generate the result

Predicted log return for next day (2024-10-19) is -0.0013728945050388575

Q4. Natural Language Processing in Python: Unsupervised Machine Learning

Using central bank communications from the folder (Communication), group the 100 documents into 15 topics using the Latent Dirichlet Allocation (LDA).

**Please check the script *Q4.ipynb* for details.**

**(*Communication.zip* file is the input data file to run the script)**

1. Please prepare a well-commented Python script/notebook that shows the data preprocessing steps and the LDA. Please provide a label for each topic.

Group the 100 documents into 15 topics using LDA:

Topic 1: ['cent', 'rate', 'monetary', 'inflation', 'policy']

Topic 2: ['bey', 'legislative', 'thoug', 'expiration', 'siness']

Topic 3: ['deputy', 'entered', 'path', 'rate', 'repo']

Topic 4: ['2020', 'governing', 'policy', 'inflation', 'percent']

Topic 5: ['bey', 'legislative', 'thoug', 'expiration', 'siness']

Topic 6: ['economic', 'japan', 'policy', 'mr', 'bank']

Topic 7: ['fourth', 'end', '10', 'february30', 'cutting']

Topic 8: ['support', 'economy', 'development', 'policy', 'financial']

Topic 9: ['bank', 'figure', 'percent', 'zealand', 'new']

Topic 10: ['trillion', 'month', 'operation', 'yen', 'approximately']

Topic 11: ['forecast', 'rate', '2023', 'growth', 'inflation']

Topic 12: ['policy', 'committee', 'bank', 'inflation', 'growth']

Topic 13: ['2018', 'inflation', 'committee', 'bank', 'growth']

Topic 14: ['march', 'inflation', 'preliminary', 'weather', 'q1']

Topic 15: ['bey', 'legislative', 'thoug', 'expiration', 'siness']

1. Compute the sentiment of each document using FINBERT.

**Please run the code to generate the result of sentiment computation.**

**Here are some example outputs:**

document label score

The Swedish economy has developed strongly so ... Positive 0.999511

21 January 2021 The Governing Council decided ... Neutral 0.999915

Deputy Governor Lars E.O. Svensson entered a r... Neutral 0.999654

Deputy Governor Karolina Ekholm entered a rese... Neutral 0.999662

10 June 2021 At today’s meeting, the Governing... Neutral 0.999946

.. ... ... ...

\n \n \n \n \n \n \n \nMonetary Policy S umma... Neutral 0.999990

MPC Statement 23 November 2023 Page 1 \n \n \... Negative 0.995403

\nMPC Statement 26 January 2023 Page 1 \n \... Positive 0.951902

1June 15, 2010 \nBank of Japan \n \nStatement... Neutral 0.946743

1August 10, 2010 \nBank of Japan \n \nStateme... Neutral 0.958936

[100 rows x 3 columns]

1. Please present a workflow chart illustrating the steps of your analysis along with the techniques/functions used in each step.

Load Data:

Description: Extract files in the zip file and load the documents. Then read each document (PDF and txt files)

Vectorization:

Function/Technique: CountVectorizer()

Description: Convert the text data into a numerical format for machine learning model to use.

Fit LDA Model:

Function/Technique: LatentDirichletAllocation(n\_components)

Description: Use LDA model to group 100 documents into 15 topics

Summary:

Description: Display the 15 topics generated from previous step

FINBERT for Sentiment Computation:

Function/Technique: transformers package TFBertForSequenceClassification, BertTokenizer and pipeline

Description: Compute sentiment for each document by using FINBERT

Q5. Natural Language Processing in Python: Supervised Machine Learning

Create a Large Language Model (LLM) that classifies sentences into two categories (inflation and exchange rate) using RoBERTa.

To facilitate the training, you will select 12 samples from each of the two categories, resulting in a total of 24 training samples. For this purpose, use the first 24 sentences as the training sample in the csv file named Labeled sentences.csv. The remaining samples will be reserved for testing purposes.

**Please check the script *Q5.ipynb* for details.**

**(*Labeled sentences.csv* file and *New sentences.csv* file are the input data files to run the script)**

• Provide the Recall, Precision and F1 statistics to assess the performance of the model.

{'inflation': {'precision': 0.9857142857142858, 'recall': 0.92, 'f1-score': 0.9517241379310346, 'support': 75.0},

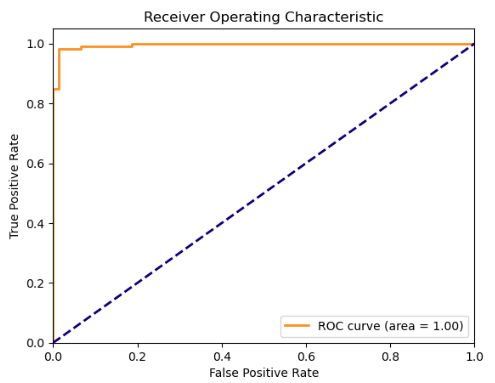
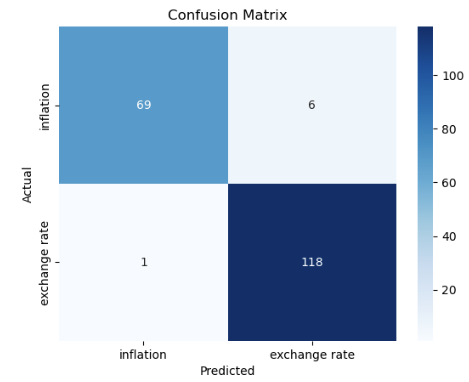
'exchange rate': {'precision': 0.9516129032258065, 'recall': 0.9915966386554622, 'f1-score': 0.9711934156378601, 'support': 119.0},

'accuracy': 0.9639175257731959,

'macro avg': {'precision': 0.9686635944700461, 'recall': 0.9557983193277311, 'f1-score': 0.9614587767844474, 'support': 194.0},

'weighted avg': {'precision': 0.9647964273837238, 'recall': 0.9639175257731959, 'f1-score': 0.963666633019242, 'support': 194.0}}

• Provide the confusion matrix and the ROC Curve.



• Make predictions on new data file (New sentences.csv).

New Sentence Prediction

0 The finances of central banks is a topic of r... inflation

1 Inflation dynamics are closely monitored, and ... inflation

2 The opposition party accuses the ruling admini... inflation

3 Our Proprietary Moisture Management technology. inflation

4 The cat lazily stretched out on the sun-warmed... exchange rate

5 Central banks intervened in the FX market to c... inflation

6 Grassroots activism is reshaping the political... inflation

7 We document the rise of China in offshore capi... inflation

8 Communication strategies aim to clarify the ce... exchange rate

9 Political gamesmanship, financial crises, and ... inflation

10 The incumbent president faces mounting scrutin... inflation

11 The aroma of freshly brewed coffee filled the ... inflation

12 Stars shimmered like diamonds against the velv... inflation

13 A gentle breeze rustled the leaves of the anci... inflation

14 Political polarization continues to deepen, fu... inflation

15 Voter turnout surged in the recent election, s... inflation

16 Transparent communication on foreign exchange ... exchange rate

17 Corruption scandals have rocked the foundation... inflation

18 Negotiations between rival factions have stall... inflation

19 International diplomacy plays a crucial role i... inflation

20 Clear and consistent communication on inflatio... inflation

1. Please prepare a well-commented Python script/notebook for the learning model.

**Please check the script *Bodi\_Yang\_Q5.ipynb* for details.**

* ***Labeled sentences.csv* file and *New sentences.csv* file are the input data files to run the script**

1. Please present a workflow chart illustrating steps of your analysis along with techniques/functions used in each step.

Load Data:

Function/Technique: pandas.read\_csv()

Description: Load the labeled sentences and new sentences from Labeled sentences.csv files.

Preprocess Data:

Function/Technique: RoBERTa

Description: Tokenization and text normalization using RoBERTa.

Vectorization:

Function/Technique: CountVectorizer()

Description: Convert the processed text data into a numerical format for machine learning model to use.

Train Classifier:

Function/Technique: LogisticRegression()

Description: Train logistic regression model on the vectorized training data.

Model Summary:

Description: Generate model summary, confusion matrix

Prediction:

Function/Technique: model.predict()

Description: Use the trained model to make predictions on New\_sentences.csv file (data processing, vectorization are also required for the new file)