## QUANTITATIVE test for MCMCO recruitment

**General instructions**

**Python,** **R, and RoBERTa** should be used for this test. Please submit the following:

* Your codes commented and easy to read. Please submit your **codes in .py or .ipynb or .R format**. Do not put your codes in Word or PDF. Please make sure you also include any **underlying datasets/documents** in your submission and make sure your codes are directly executable.
* If required, your **answers to the questions** below with the **charts/tables** when necessary.

The charts should be readable and intuitive. Software default style is fine. Use different line styles/colors, etc. to distinguish lines in lineplots. More generally, don’t hesitate to take initiative to produce intuitive and reader-friendly charts.

**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!

1. Retrieve currency in circulation times series from the [Bank of Zambia website](https://www.boz.zm/monetary-and-financial-statistics-Dec2022.htm).
2. There are bugs and errors in ets\_mod.py. Please correct them as many as possible.
3. 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.
4. 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.
5. Please forecast the currency in circulation **for the next week from the end of dataset** based on the model above.
6. 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.
7. 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).

**Q2. Quantitative Modelling in Python or R**

Please collect the EUR/USD daily exchange rate from [FRED](https://fred.stlouisfed.org/series/DEXUSEU) and complete the following tasks related to Value-at-Risk forecasting from a GARCH-type model with a Python or R script:

a) 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.

b) 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.

c) 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.

**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.

**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).

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.
2. Compute the sentiment of each document using FINBERT.

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

**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.

* Provide the Recall, Precision and F1 statistics to assess the performance of the model.
* Provide the confusion matrix and the ROC Curve.
* Make predictions on new data file (New sentences.csv).

1. Please prepare a well-commented Python script/notebook for the learning model.
2. Please present a workflow chart illustrating steps of your analysis along with techniques/functions used in each step.