Project Title: Quantifying the Impact of Geopolitical Risk, Trade War, and Environmental Shock on Global Financial and Commodity Markets

1. Introduction

This document explains how the source code for our capstone project was put together. The project is about analyzing how geopolitical, trade, and environmental shocks affect global financial and commodity markets. Python was used to write the codes and econometrics models were used in combination with machine learning models.

2. Objectives of the Code

- Load and prepare financial and risk indicator data from several sources that have been cleansed in Excel
- Perform exploratory data analysis (EDA)
- Use statistical modeling in this case, MS-GARCH and DCC-GARCH.
- Apply structural break detection (Chow & Bai-Perron)
- Create and assess an LSTM model for forecasting
- Display the results of the models for interpretation

3. Technologies Used and their Purposes

- The primary programming language is python
- Pandas - Data wrangling and manipulation
- NumPy - Numerical calculations
- Datetime, timedata - Forhandling time and date information
- Matplotlib, Seaborn - Data visualization
- Arch - GARCH modeling
- Statsmodels - Time series tests and regressions
- Scikit-learn - Preprocessing and evaluation metrics
- Keras/TensorFlow - LSTM model implementation
- Ruptures - Used for identifying structural break
- Warnings - For suppressing warnings for cleaner output

4. Code Modules and Workflow

4.1 Data Preparation

- Dataset loaded: dataset.xlsx containing SP500 returns, Oil returns, GPR, TPU, ENV
- Preprocessing: Alignment by date, handling of missing values, calculating log returns
- Output: Cleaned and merged dataframe

4.2 Exploratory Data Analysis

- Descriptive statistics: mean, std, kurtosis, skewness
- Correlation matrix: Heatmap to check for linear dependencies
- Time-series plots: To visually check for volatility and trends

4.3 Structural Break Analysis

- Chow Test: Tested for breaks at some specific dates
- Bai-Perron Test: Detected unknown structural breakpoints
- Purpose: To justify the selection of the regime-switching model

4.4 MS-GARCH Modeling

- The tool used here is the `arch` library and the logic behind it is to fit Markov-Switching GARCH to returns series with the goal to identify volatility patterns which are regime-dependent.

4.5 DCC-GARCH Modeling

- The logic was Dynamic Conditional Correlation analysis between financial and commodity market, the output was a time-varying correlation matrix to check the co-movement and this is all relevant because it measures how connected markets are under different shocks

4.6 LSTM Model

The libraries used here are Keras, TensorFlow and steps taken include reshaping data into 3D format, defining the model (LSTM layers → Dense output), training with return sequences, assessment using RMSE and visual prediction vs actual plot. The purpose of this model is Out-of-sample forecasting to improve prediction accuracy.

5. Input/Output Summary

Module	Inputs	Outputs
Data Collection & Cleaning	Raw data from S&P 500, Oil prices, GPR, TWAR, ENV indices	Cleaned, merged, and time-aligned DataFrame
Exploratory Data Analysis	Preprocessed DataFrame	Summary statistics, correlation matrix, trend plots

Stationarity & Structural Tests	Time series data (returns)	ADF/PP test results, Bai-Perron & Chow structural breakpoints
MS-GARCH Modeling	Return series (SP500 and Oil)	Conditional volatility estimates across regimes
DCC-GARCH Modeling	Paired return series (SP500 & Oil)	Time-varying correlations; dynamic co-movement patterns
Shock Impact Estimation	Risk indices and asset returns	Coefficients of GPR, TWAR, ENV under calm and crises regimes
LSTM Forecasting	Scaled and reshaped time series data	Forecasted returns; RMSE, MAE, visual comparison with actual values
Visualization & Reporting	All model outputs and diagnostics	Charts, plots, tables for interpretation and presentation

6. Model Evaluation

- GARCH models were interpreted through regime classification and volatility shifts while structural breaks helped us to validate crisis dates. For LSTM, RMSE was used to assess the accuracy of the forecast. Then visualizations were used to support interpretation and storytelling.

7. Limitations

- Granularity of LSTM performance is reduced by monthly data and multicollinearity was not deeply tested between risk indicators. DCC-GARCH assumes that there is linear correlation dynamics.

8. Future Improvements

- Models can be used for other asset types and regional indexes.
- Transformer-based time series models can be used to enhance LSTM.
- For improved interpretability, explainable AI technologies (like SHAP) can be incorporated.
- Real-time risk alert functionality can be enabled using API feeds