Homework / Project 1:
Financial Stylized Facts from OHLC Time Series

Quantitative Risk Management (QRM)

Instructor: Pasin Marupanthorn, Ph.D., CQF
Contact: quantfilab@gmail.com | Line ID: `oporkab`
Course repository: https://github.com/QuantFILab/QRM-NIDA

Overview

In this assignment, you will use real-world OHLC (Open–High–Low–Close) price data to empirically verify *financial stylized facts* commonly observed in asset returns and volatility. You must select **two** time series from the dataset provided and present evidence for **at least two** stylized facts (you may show more if you find additional evidence).

Learning Objectives

By completing this assignment, you will be able to:

1. Transform OHLC data into return series suitable for statistical analysis.
2. Identify and empirically evaluate stylized facts using plots and summary statistics.
3. Communicate findings clearly with reproducible code and well-labeled figures.

Data Source

The OHLC dataset will be provided via Google Drive:

- **Data folder:** Google Drive: OHLC Dataset (May time frame)

Notes:

- The dataset includes multiple OHLC time series within a May time frame (as provided by the instructor).
- You must choose **two** series (two assets, indices, FX pairs, crypto, etc.) from the provided files.
- Use the **Close** price for return construction unless you justify an alternative.

Requirements (What You Must Do)

You must complete the following for **each** of your two selected time series.

1) Data Preparation and Quality Checks

1. Load the OHLC data and verify:
   - time ordering of observations,
   - missing timestamps or missing OHLC values,
   - duplicated timestamps,
   - obvious data errors (e.g., negative prices).
2. Construct (log) returns from close prices:

$$r_t = \log\left(\frac{C_t}{C_{t-1}}\right), \tag{1}$$

   where C_t denotes the close price at time t.
3. Briefly describe how you handled missing values and irregular sampling (if any).

2) Verify Stylized Facts (Minimum: Two Facts Total)

You must empirically verify **at least two** stylized facts across your analysis (per asset or jointly). You may include as many stylized facts as you can support with clear evidence.

Possible stylized facts (choose at least two, more is encouraged):

- **Heavy tails / non-normality:** return distribution deviates from Gaussian (fat tails, excess kurtosis).
- **Volatility clustering:** periods of high volatility tend to cluster together.
- **Weak autocorrelation of returns, strong autocorrelation of magnitudes:** ACF(r_t) near zero while ACF($|r_t|$) or ACF(r_t^2) is positive and persistent.
- **Time-varying volatility (heteroskedasticity):** volatility changes over time (e.g., rolling volatility).
- **Leverage effect (asymmetry):** negative returns associated with higher future volatility (market-dependent).
- **Co-movement (optional):** dependence between your two chosen series; correlation changes in stress.
- **Jumps / extreme moves:** occasional large returns/outliers consistent with jump-like behavior.

3) Evidence Standard (How to Support Each Claim)

For **each** stylized fact you claim, include:

- at least one **figure** (plot) that directly supports the claim, and
- at least one **quantitative summary** (statistic, diagnostic, or model output), and
- a short explanation of **methodology** and **interpretation** (what you did and what it means).

Suggested Analyses (Optional but Recommended)

You may use any combination of the following:

- **Distribution diagnostics:** histogram of r_t, QQ-plot vs Normal (and/or Student-t), skewness and kurtosis.
- **Normality checks:** Jarque–Bera or Shapiro–Wilk (interpret cautiously; large samples reject easily).
- **Volatility visualization:** rolling standard deviation (e.g., 10–30 period window), EWMA volatility.
- **Dependence in time:** ACF/PACF plots for r_t, $|r_t|$, and r_t^2.
- **Simple volatility model:** fit GARCH(1,1) and comment on persistence (optional extension).
- **Cross-asset comparison:** correlations between the two series (overall vs high-volatility subperiods).

Deliverables

Submit a **single PDF** exported from your:

- Jupyter Notebook (`.ipynb`) **or**
- Google Colab notebook,

including **code, outputs, and all figures**.

PDF Content Requirements

Your PDF must contain:

1. **Title section:** name, student ID, chosen time series (two assets), and sample period (May time frame).
2. **Data description:** file names, frequency (daily/intraday), any cleaning decisions.
3. **Methods and results:** stylized facts tested, plots, quantitative summaries, and interpretation.
4. **Conclusion:** key findings, limitations, and what you would test with a longer sample.

Submission Instructions

- Submit the **PDF** via the course submission channel (as announced in class / repository).
- Ensure reproducibility:
  - If you use randomness (bootstrapping, simulations), set and report a random seed.
  - List external libraries used (e.g., `pandas`, `numpy`, `matplotlib`, `statsmodels`).
  - Your notebook should run from top to bottom without manual intervention (except downloading/loading the provided data).

Grading Rubric (20 points)

Criterion	Points
Data loading, cleaning, return construction, and clear description of the two selected series	4
Evidence for stylized facts (plots + quantitative summaries; correct mapping to claims)	8
Methodological correctness and quality of interpretation (explain *what* and *why*)	4
Presentation quality (structure, labeling, readability, captions, professionalism)	2
Depth/initiative (additional stylized facts, cross-asset comparison, thoughtful extensions)	2

Academic Integrity

You may discuss general ideas with classmates; however, the submitted notebook (code, analysis, and writing) must be your own work. If you use external references or code snippets, cite them clearly in the notebook.

Helpful Tips

- Use clear figure titles and axis labels (e.g., "ACF of squared returns").
- Do not claim a stylized fact without evidence. If the evidence is weak for your May sample window, explain why.
- Keep the narrative concise and professional: state the stylized fact, show the evidence, interpret, and move on.