In the name of GOD.

Sharif University of Technology

Stochastic Processes

Fall 2024

Hamid R. Rabiee

Project | Deadline : 1403/11/09

You are tasked with building and analyzing models to simulate and predict financial data using the S&P500 dataset. This dataset contains the following columns: **Timestamp, Open, High, Low, Close, Adj Close, Volume.**

Literature Review Requirement: For each part, you must perform a brief (1-2 paragraphs) literature review **inside your Jupyter Notebook**.

1. **Data Preparation (15 points)**

   (a) Perform a literature review on stationarity testing methods.

   (b) Download S&P500 data and visualize the Close prices. Discuss stationarity.

   (c) Transform the data to achieve stationarity, if necessary.

   (d) Test and justify stationarity of the transformed data. Justification must be presented as **hypothesis testing**, with clear arguments supported by minimizing the p-value. Report your p-value.

   Question: Discuss the limitations of your approach.

2. **Random Walk Model (10 points)**

   (a) Review literature on the Random Walk hypothesis in financial markets and discuss well-known conclusions regarding the predictability (or lack thereof) of stock prices under this model.

   (b) Fit a random walk model to the data and estimate parameters.

   (c) Simulate Close price paths and compare with actual prices (plot results).

   (d) Compute log-likelihood, MAE, and RMSE.

   Question: What are the assumptions of the random walk model, and why might it be too simplistic for financial data?

3. **Geometric Brownian Motion (GBM) (15 points)**

   (a) Review literature on GBM models used in mathematical finance.

   (b) Fit a GBM model to the data (use raw, non-stationary prices) and estimate parameters.

   (c) Simulate Close price paths and compare with actual prices (plot results).

   (d) Compute log-likelihood, MAE, and RMSE.

Question: How does the GBM model handle price trends and volatility, and what are its limitations in capturing real-world financial data?

4. **Gaussian Process Regression (GPR) (20 points)**

   (a) Review literature on Gaussian Process Regression applied to financial modeling. Identify standard kernels.

   (b) Define a custom kernel using features derived from available columns (e.g., price ranges, Volume).

   (c) Fit a GPR model and estimate parameters.

   (d) Simulate Close price paths with confidence intervals and compare with actual prices.

   (e) Calculate MAE, RMSE and the percentage of actual prices within the confidence intervals.

   Questions:

   - Could the GPR model be overfitted? Justify your conclusion based on its performance on training vs. test data.

5. **HMM with Gaussian Mixture Model (GMM) Emissions (25 points)**

   (a) Review literature on Hidden Markov Models (HMMs) applied to financial time series. Discuss how hidden states have been interpreted as market regimes.

   (b) Define hidden states for the market.

   (c) Fit GMMs for each state and estimate the HMM parameters.

   (d) Simulate Close price paths and compare with actual prices.

   (e) Compute MAE, RMSE.

   Questions:

   - What do hidden states in your HMM reveal about market dynamics? Interpret the meaning of the states.
   - Could your HMM model be over-fitted? Discuss risks and justify your conclusion based on the complexity and performance of the model.

6. **HMM-Based Automated Trading Strategy (15 points)**

   (a) Review literature on HMM-based trading strategies and regime-switching models in portfolio management. Summarize common approaches and their reported performance.

   (b) Design a trading strategy using the HMM states.

   (c) Decode the most likely sequence of states.

   (d) Backtest the strategy on historical data and report cumulative returns.

   (e) Explore whether features from additional columns (e.g., Volume) improve your strategy.

   (f) Visualize your trading actions (buy/sell/hold signals) alongside the price history plot.

Disclaimer: This exercise is intended solely for educational purposes to demonstrate the application of Hidden Markov Models in financial data analysis. It is not an endorsement or encouragement to engage in automated trading. Automated trading involves significant ethical, legal, and financial considerations, including market risks, regulatory compliance, and the potential for market manipulation. Always seek professional advice and adhere to regulatory requirements before implementing any trading strategies in real-world scenarios.

Deliverables

- A **Jupyter Notebook** containing all your code and report. The notebook should be well-documented with comments and markdown cells explaining each step.
- Your notebook must include:
  - Literature reviews for each part.
  - Explanation of data preparation steps and stationarity tests.
  - Parameter estimation procedures for each model.
  - Resulting plots and visualizations for each model and simulation.
  - Computed metrics (MAE, RMSE, log-likelihood, cumulative returns, etc.).
  - Answers to the posed questions.
- Ensure all output plots and figures are clearly labeled and embedded in the Jupyter Notebook.