Term Project Proposal

Group Members (Group 3):

Thu Trang Bach, Rhichard Koh, Mohsin Mohammed, Thu Tra Truong

1. Business Questions:

Volatility Forecasting

"How can we forecast the volatility of a portfolio of financial stocks over the next 30 days?"

Value-at-Risk Analysis

"What is the probability that this portfolio loses more than 5% in a single day?"

Expected Shortfall

"What is the average loss the portfolio experiences on the worst 5% of trading days?"

Diversification Impact

"How much does diversification across sectors reduce portfolio risk?"

Crisis Risk Forecasting

"How did portfolio risk behave during market crises (e.g., COVID crash in March 2020), and how can we forecast such spikes?"

2. Datasets to be Used:

We will use a dataset from Kaggle.com to answer our business question. Here is a link to the dataset:

[https://www.kaggle.com/datasets/nikitamanaenkov/stock-portfolio-data-with-prices-and-indices?resource=download].

(We use the portfolio_prices.csv file as this is our basket of stocks/assets)

We can also do a comparative analysis of the S&P 500 benchmark index (also priced in the dataset) to see how our portfolio performs in comparison.

We may also use yfinance for more data if needed.

3. High-Level Description of the Dataset:

Our dataset includes 35,397 observations and 9 variables. The key variables that we believe will be most useful for achieving our project's goals are Date, Ticker, Adjusted (adjusted closing price), Returns, and Volume, as these provide the necessary information to track stock performance, compute returns, and analyze portfolio trends over time.

Depending on our project's goal (forecasting, portfolio optimization, risk analysis, etc.), here are the variables we expect to be the most useful:

- Date - to index time, align series, compute returns
- Ticker / Asset identifier - so you can manage multiple stocks
- Close (or Adj_Close) - to compute returns, log returns, and compare across stocks
- Open, High, Low - for intraday or volatility features (e.g. high-low spread)
- Volume - trading activity, liquidity signal
- Market indices (e.g. S&P 500, Nasdaq, etc.) - to assess market movements / beta / relative performance
- Portfolio weight / holdings (if available) - to relate returns to allocation
- Derived variables (not originally in data but computed): e.g. daily return, cumulative return, rolling volatility, momentum indicators, correlation with market, sharpe ratio, drawdowns, etc.

Our dataset(s) consist of multiple CSVs providing stock-level and index-level daily price histories. After merging, we'll obtain a panel with (time × assets) observations. The most important variables for our goals will be the adjusted close / close prices, volume, market index series, and portfolio weights / holdings (if present). From these, we can engineer returns, risk / volatility metrics, and relative performance indicators.

4. Similar Projects:

https://www.kaggle.com/code/yousefsaeedian/var-cvar-analysis-and-sharp-ratio-calculation#%F0%9F%94%B88.-Sharpe-Ratio

While the Kaggle project *"VaR, CVaR Analysis and Sharpe Ratio Calculation"* focuses primarily on calculating portfolio risk and performance metrics using historical simulation

on a static dataset, our project expands this framework toward forecasting and forward-looking risk evaluation. Specifically, we build upon this work by

- Applying it to a different dataset containing a portfolio of multiple financial stocks with market index benchmarks
- Extending the analysis to forecast 30-day portfolio volatility using time-series modeling techniques, and
- Incorporating Value-at-Risk estimation under simulated future return scenarios rather than historical data alone. Furthermore, our study adds a diversification and crisis-risk component, examining how sectoral diversification influences overall portfolio volatility and how portfolio risk evolves during stress periods such as the COVID-19 market crash.

In essence, while the referenced project provides static risk metrics, our work aims to model and forecast dynamic portfolio risk behavior. enhancing practical financial insights for portfolio management and stress testing.

Summary

This project applies a purely statistical and econometric framework, using time series analysis, probability distributions and compensation based risk metrics to forecast, quantify and interpret portfolio risk dynamics. No machine learning and deep learning will be employed for this project.