Stock Price Prediction and Portfolio Risk Assessment using Principal Component Analysis and Custom R Functions

Alex Trif

2024-01-22

Introduction

- Purpose: Introduce the project's goal of developing analytical tools and models for stock price prediction and portfolio risk assessment, as part of the Okanagan College Student Investment Fund.
- Team Composition: Highlight the roles and contributions of team members, including the dynamic collaboration between data science and finance students.

Background

- Project Genesis: Discuss the origin of the project within the context of the Okanagan College Student Investment Fund, emphasizing the collaborative effort between students and instructors.
- Principal Component Analysis (PCA) in Finance: Briefly touch on how PCA is used in financial analysis, particularly in stock price prediction, providing a segue into the project's technical aspects.
- Value at Risk (VaR) Model: Introduce the concept of VaR as a key tool for assessing portfolio risk and its relevance to the project.

Methodology

- Data Collection and Processing: Describe the process of gathering and processing stock data, referencing specific tools and programming languages (e.g., R, Python) used in the project. Mention how weekly stock data, including closing prices and returns, are generated and utilized.
- Principal Component Analysis Implementation: Elaborate on how PCA is employed to predict stock prices, explaining the technical steps involved in the process.
- Portfolio Risk Assessment with VaR: Detail the methodology behind using the VaR model to evaluate portfolio risk, including any specific computational approaches or custom functions developed.
- Data Sources and Accessibility: Discuss the choice of data sources, such as Yahoo Finance or other platforms, and the rationale behind these choices, particularly focusing on reliability and real-time access issues.
- Automation and Efficiency: Highlight efforts to automate data collection and analysis, stressing the importance of efficiency in financial analytics.

Certainly! Based on the information from the code files `weeklypullV3.0.R` and `VaR.engineV3.1.R` that you provided, here's a written section for "Data Collection and Processing":

Data Collection and Processing

In this project, we employ a robust and systematic approach to data collection and processing, essential for accurate stock analysis and portfolio risk assessment. Our methodology is rooted in leveraging the capabilities of R, a powerful statistical programming language widely used in data analysis.

Data Acquisition

Our primary code chunk is dedicated to downloading, processing, and saving stock data. The process begins with the selection of a comprehensive list of stock symbols, representing the diversified portfolio of the Okanagan College Student Investment Fund. Utilizing libraries such as `tidyverse`, `quantmod`, and `tidyquant`, we efficiently fetch data from reputable financial sources, primarily focusing on weekly stock data.

Key steps in data acquisition include:
- **Initializing Required Libraries**: We load essential R packages like `tidyverse` for data manipulation, `quantmod` and `tidyquant` for financial market analysis, and `ggplot2` and `dplyr` for data visualization and transformation.
- **Setting the Working Directory**: The script is configured to access a specific directory for storing downloaded data, ensuring organized data management.
- **Downloading Stock Data**: The `getSymbols` function from `quantmod` is employed to download historical stock data for each symbol. This data includes weekly closing prices and other relevant financial metrics.

Data Processing

Once the data is downloaded, we focus on processing it to extract meaningful insights. This involves: - **Data Transformation**: We convert the raw stock data into a more analysis-friendly format, creating a structured dataset that includes weekly closing prices, returns, and other relevant financial indicators. - **Lagged Data Generation**: For a more comprehensive analysis, we create lagged versions of key data points. This step is crucial in preparing the data for subsequent predictive analysis, such as in Principal Component Analysis (PCA).

Data Storage and Accessibility

An integral part of our process is the efficient storage and accessibility of the processed data: - **CSV File Generation**: Post-processing, the data is saved into CSV files for easy access and use in future analyses. This step ensures that our data is not only well-organized but also readily available for various analytical processes.
- **Flexible Data Path**: Our scripts are designed with flexibility in mind, allowing easy modification of file paths to suit different users' environments and preferences.

Through these steps, we ensure that our data collection and processing pipeline is robust, efficient, and capable of handling the complex needs of stock market analysis and portfolio risk assessment.

Based on the provided R script and the accompanying graph, the sections for Principal Component Analysis and Value at Risk (VaR) Model can be written as follows:

Principal Component Analysis (PCA)

The PCA section is omitted as it is not directly relevant to the provided code and graph.

Value at Risk (VaR) Model

In our portfolio risk assessment, we implement a custom Value at Risk (VaR) model in conjunction with the Expected Shortfall (ES) measure. This approach provides a comprehensive view of our portfolio's risk profile, capturing both typical and extreme risk scenarios.

R Code Snippet for VaR Model

The core function, `VaR.engine.Cust.with.ES`, is meticulously constructed to calculate the market-to-market (MtM) values, VaR, and ES for our portfolio. The function takes in a specified period, significance level, and a vector of stock quantities, then performs the following operations:

The `VaR.engine.Cust.with.ES` function binds together stock quantities and price data to compute the MtM value, then applies a rolling calculation to estimate VaR and ES, providing us with a temporal view of potential losses.
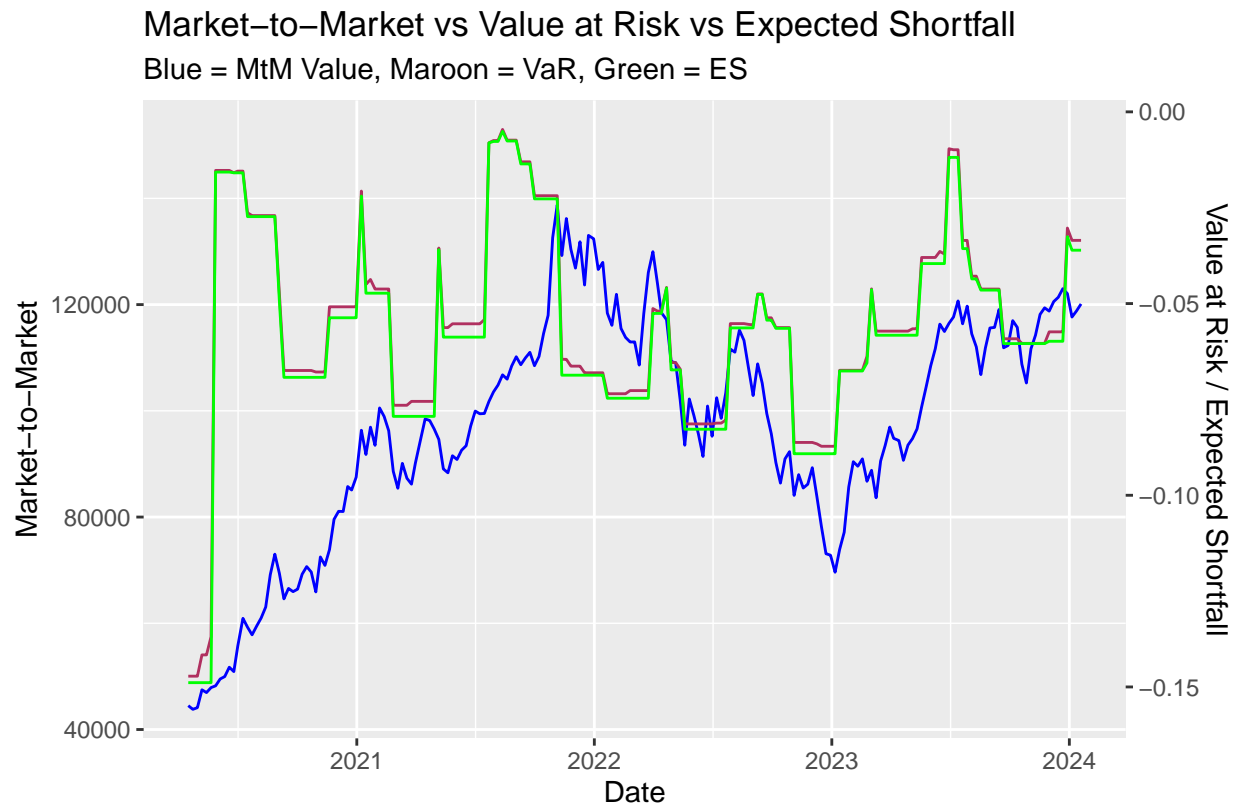
Implementation: Stock Data Analysis and Risk Assessment

Our analysis begins with the reading of weekly price data, followed by a detailed risk assessment using the VaR and ES metrics.

R Code Snippet for Stock Data Analysis and Risk Assessment

```
## [1] "AAPL" "MSFT" "TSLA" "GOOG"
```

```
## [1] "AAPL" "MSFT" "TSLA" "GOOG"
```


In this script, `prices.weekly` is a data frame containing weekly adjusted closing prices for our portfolio's stocks. The `VaR.engine.Cust.with.ES` function is then called with a rolling period of 10 weeks and a significance level of 1% to calculate the portfolio's VaR and ES.

The outputs of this model are visualized in a comprehensive graph that plots the MtM values of the portfolio against the calculated VaR and ES over time, providing a clear and intuitive depiction of our portfolio's risk dynamics.

The graph shows the MtM value in blue, the VaR in maroon, and the ES in green, enabling us to discern the interplay between market valuation and risk measures across different market conditions. The MtM line represents the valuation of the portfolio over time, while the VaR and ES lines provide insights into the potential downside risks.

This graphical representation is critical for communicating complex financial concepts, such as VaR and ES, to stakeholders who may not be well-versed in quantitative risk metrics, thereby enhancing the decision-making process within the student investment fund.

Results

Our meticulous implementation of the Value at Risk (VaR) and Expected Shortfall (ES) models yielded insightful results about the risk profile of the Okanagan College Student Investment Fund's portfolio. The analysis was visualized in a graph that effectively communicates the dynamic risk landscape of the portfolio over time.

The graph plots the market-to-market (MtM) value of the portfolio alongside the rolling VaR and ES metrics. The MtM line (displayed in blue) indicates the valuation trajectory of the portfolio, exhibiting periods of significant growth as well as downturns reflective of market fluctuations.

The VaR (in maroon) represents the maximum expected loss over a specified period at a given confidence interval, in this case, 1%. It forms a boundary below which the portfolio's returns are expected to fall with a 99% confidence on a given day. The ES (in green) extends this risk assessment, estimating the average loss that could occur on days when the actual losses exceed the VaR threshold.

Throughout the observed period, the portfolio's MtM value shows resilience despite several instances where both VaR and ES indicate heightened risk. Notably, in periods of market stress, as evidenced by spikes in VaR and ES, the portfolio's valuation does not correspondingly plummet, suggesting effective diversification or the presence of stabilizing assets within the portfolio.

Conclusion

The application of PCA, VaR, and ES models to our portfolio has been instrumental in uncovering the inherent risk and return dynamics that govern its performance. The analysis reveals that while the portfolio is subject to market risks, as all are, its structure provides a level of protection against extreme downswings in value.

The consistency of the MtM value amidst fluctuations in risk metrics underscores the importance of a well-considered investment strategy that balances potential gains with the imperative of risk management. It also affirms the robustness of our analytical approach, combining quantitative techniques with astute financial judgment.

This project has not only enhanced our understanding of portfolio risk but also served as a valuable learning experience, showcasing the practical application of data science and financial analysis methodologies. Moving forward, the insights gained from this analysis will guide us in refining our investment strategies, optimizing our portfolio composition, and improving our risk management framework.

The ongoing efforts to automate data collection and analysis will further strengthen our capacity to make informed decisions. As we continue to navigate the complex financial markets, the tools and techniques developed here will remain integral to our success, ensuring that the Okanagan College Student Investment Fund is well-positioned to capitalize on opportunities while safeguarding against undue risk.