PORTFOLIO ANALYSIS IN R

With the Integration of ML model for prediction

Visualizing & Analyzing Data with R: Methods & Tools - DAT-5323 - BMBANDD1

Professor Todd Cioffi

Team 6 - Group Assignment

Firas Bidawi, Rodrigue N'senga, Sahil Nadiyadra, Sharayu Mirajkar, Shalini James Paulraj, Reem Tarakji

1. Introduction

This report presents an analysis of the VR6 portfolio, which is include of semiconductor stocks, and compares its performance with benchmark indices SPY (S&P 500 ETF) and SMH (Semiconductor ETF). The study contains exploratory data analysis, portfolio returns and risk assessment, financial ratios, and a predictive modeling approach using linear regression to forecast future portfolio performance. The objective is to understand the behavior of the VR6 portfolio in comparison to the broader market and to identify key financial metrics that can aid in decision-making.

2. Exploratory Data Analysis

The first step in the analysis involved importing historical stock data from Yahoo Finance, covering the period from 2013 to 2024. Data cleaning and preprocessing were performed to calculate daily returns and check for missing or null values. Summary statistics revealed that the portfolio stocks exhibit varying degrees of volatility, with some stocks displaying higher standard deviations, indicating greater price fluctuations. The box plot analysis highlighted the presence of outliers in daily returns, reflecting extreme market movements that could be attributed to earnings releases, macroeconomic conditions, or geopolitical factors.

3. Portfolio Analysis

The VR6 portfolio was constructed based on pre-defined stock weights, ensuring a diversified exposure to semiconductor stocks. The portfolio returns were computed and benchmarked against SPY and SMH from 2017 onwards. The yearly returns analysis showed that the VR6 portfolio often exhibited higher annual returns compared to SPY but was more closely aligned with SMH, indicating that semiconductor sector trends heavily influenced the portfolio's performance. The portfolio's standard deviation analysis suggested that VR6 was more volatile than SPY but comparable to SMH, reaffirming its sector-specific nature. (Ref Appendix)

4. Risk Analysis and Financial Ratios

An important aspect of the study was the evaluation of risk-adjusted performance metrics, including Sharpe, Sortino, and Treynor ratios. The Sharpe ratio, which measures risk-adjusted returns, indicated that VR6 outperformed SPY but was slightly lower than SMH, suggesting that while the portfolio delivered strong returns, it also carried significant risk. The Sortino ratio, which focuses on downside risk, confirmed that VR6 had moderate exposure to negative market movements. The Treynor ratio, which assesses performance relative to systematic risk, revealed that VR6's returns justified its risk exposure when compared to market benchmarks. Furthermore, alpha and beta values were computed, with beta values suggesting that VR6 was more sensitive to market fluctuations than SPY but closely followed SMH. The alpha value indicated that VR6 occasionally generated excess returns beyond market expectations, reinforcing its strong performance potential. (Ref Appendix)

5. Predictive Modeling Using Linear Regression

To forecast future portfolio performance, a multiple linear regression model was developed using SPY and SMH as independent variables. The initial model demonstrated a reasonable predictive power, but enhancements were introduced by incorporating moving averages and rolling volatility as additional explanatory features. The enhanced model showed improved accuracy, as indicated by lower RMSE values and higher adjusted R-squared scores. A cross-validation analysis further validated the model's robustness, confirming that the selected features contributed to better predictive capability. (Ref Appendix)

6. Key Insights and Conclusion

The findings suggest that the VR6 portfolio, while highly correlated with SMH, offers significant return potential but carries a higher degree of volatility. The portfolio's sensitivity to broader market movements, as indicated by its beta value, underscores the need for active risk management strategies. Financial ratio analysis supports the conclusion that VR6 is a strong performer but requires careful monitoring of downside risks. The predictive modeling approach demonstrated that SPY and SMH are effective predictors of VR6's performance, with enhancements such as moving averages and rolling volatility further refining the model's accuracy. Overall, the study provides valuable insights for portfolio managers and investors seeking to optimize risk-adjusted returns within the semiconductor sector.

7. References

- OpenAI. (2022). ChatGPT (Dec 20 version) [Large language model]. https://www.openai.com/
- Google. (n.d.). https://google.com/
- R code – Group assignment.rmd file
- Data set link- https://finance.yahoo.com/

8. Appendix

Portfolio Returns comparison

Date <date>	vr6 <dbl>	SMH <dbl>	SPY <dbl>
2017-01-03	-4.426594e-05	0.001675080	0.0076500099
2017-01-04	1.969852e-03	0.003205069	0.0059491933
2017-01-05	-5.818993e-03	-0.005834118	-0.0007944564
2017-01-06	3.582924e-03	0.005169802	0.0035777952
2017-01-09	1.272735e-02	0.011537311	-0.0033009110
2017-01-10	1.665119e-03	0.003847828	0.0000000000

Daily returns and weights

Date	Open	High	Low	Close	Volume	Adjusted	Stock	DailyReturn	Weight
<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>
2013-01-03	11.840	11.900	11.700	11.810	8741500	9.944545	AMAT	-0.002533761	0.068566
2013-01-03	2.520	2.590	2.460	2.490	24966900	2.490000	AMD	-0.015810262	0.053614
2013-01-03	65.990	66.290	65.060	65.380	1725400	58.072678	ASML	-0.020964384	0.067451
2013-01-03	3.234	3.298	3.234	3.271	23295000	2.401756	AVGO	0.005224327	0.104102
2013-01-03	21.400	21.480	21.140	21.320	41054100	15.246391	INTC	-0.002806336	0.068772
2013-01-03	3.806	3.846	3.759	3.798	34134000	1.898219	LRCX	-0.019364785	0.112993

Return comparison: Portfolio vs Semiconductor ETF vs SP 500

Year	VR6_Return	SPY_Return	SMH_Return
<dbl>	<dbl>	<dbl>	<dbl>
2017	0.4149878	0.19384417	0.3652987
2018	-0.1052739	-0.06347893	-0.1076577
2019	0.7280533	0.28785206	0.6201880
2020	0.6679183	0.16162313	0.5445866
2021	0.4870438	0.27035408	0.4137442
2022	-0.3954258	-0.19481641	-0.3427896
2023	0.8437435	0.24286799	0.7233664
2024	0.2340361	0.23755018	0.3990965

Risk comparison : Portfolio vs Semiconductor ETF vs SP 500

Year	VR6_StdDev	SPY_StdDev	SMH_StdDev
<dbl>	<dbl>	<dbl>	<dbl>
2017	0.01219750	0.004265298	0.01017459
2018	0.01878167	0.010792816	0.01772537
2019	0.01764765	0.007906031	0.01621360
2020	0.03042068	0.021111335	0.02786661
2021	0.01937887	0.008255724	0.01908708
2022	0.02834578	0.015287835	0.02685538
2023	0.01765865	0.008308169	0.01732935
2024	0.02322045	0.007939267	0.02201784

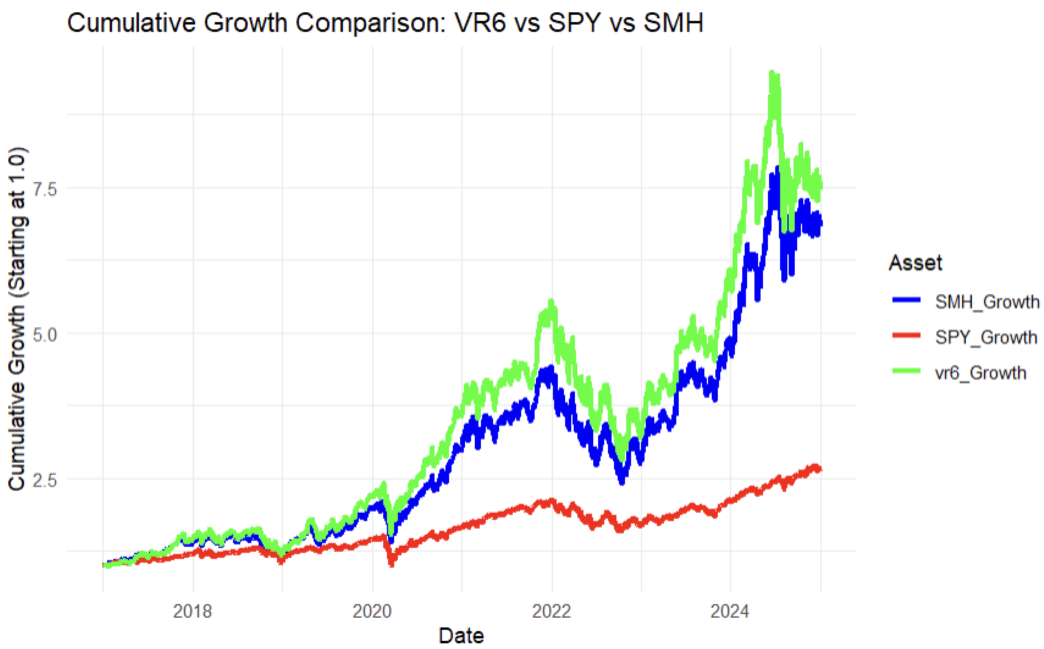
Sharpe Ratio, Sortino and Treynor ratio

VR6_Sharpe	SPY_Sharpe	SMH_Sharpe	VR6_Sortino	SPY_Sortino	SMH_Sortino	VR6_Treynor	SPY_Treynor	SMH_Treynor
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0.05328699	0.0404477	0.05318143	0.07431954	0.04826386	0.07480752	5.725576	3.484018	5.637524

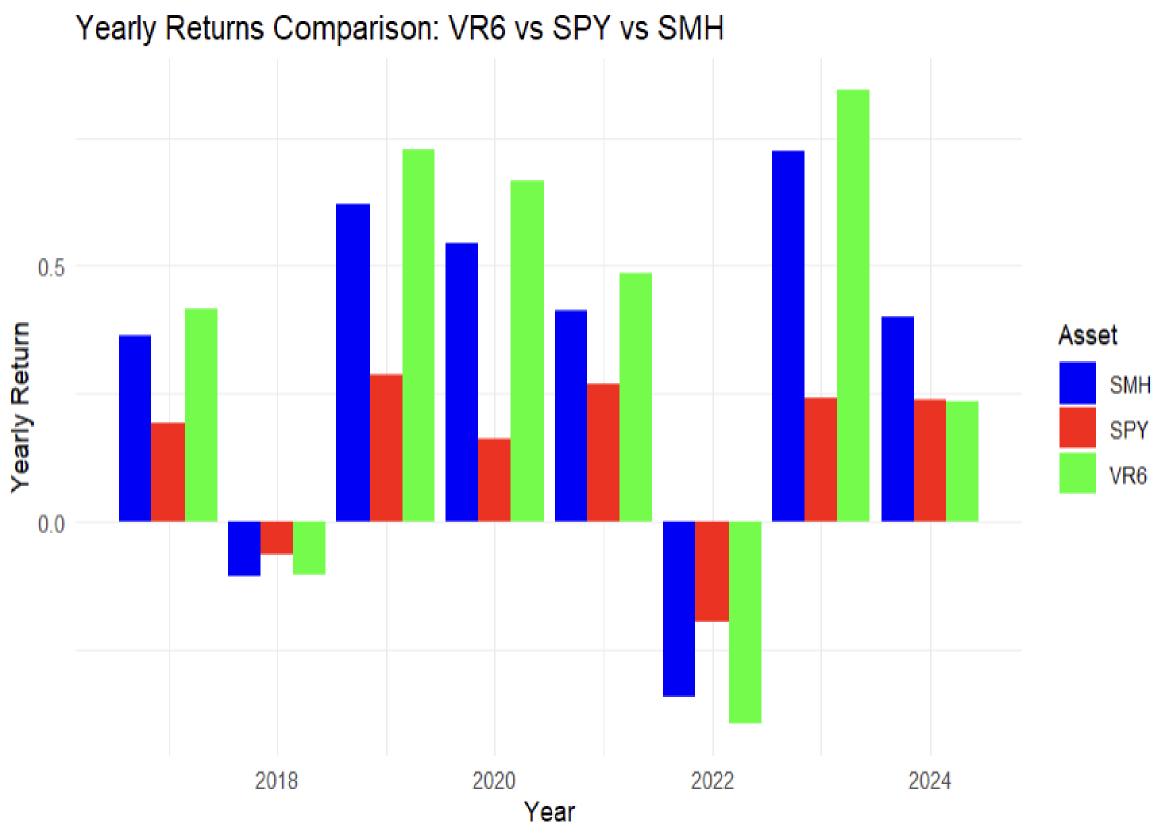
Alpha and Beta Comparison

VR6_Beta	SMH_Beta	VR6_Alpha	SMH_Alpha
<dbl>	<dbl>	<dbl>	<dbl>
1.49976	1.428048	0.0004531049	0.0004144918

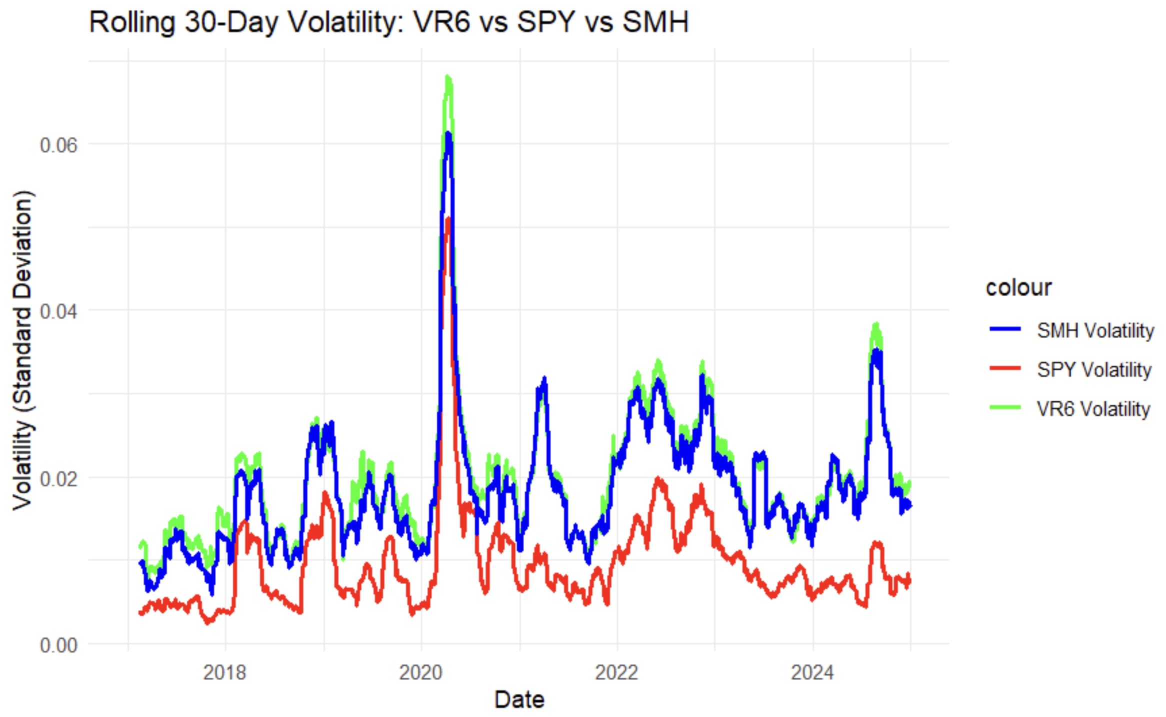
Cummulative growth chart: Portfolio Vs ETF vs SP 500



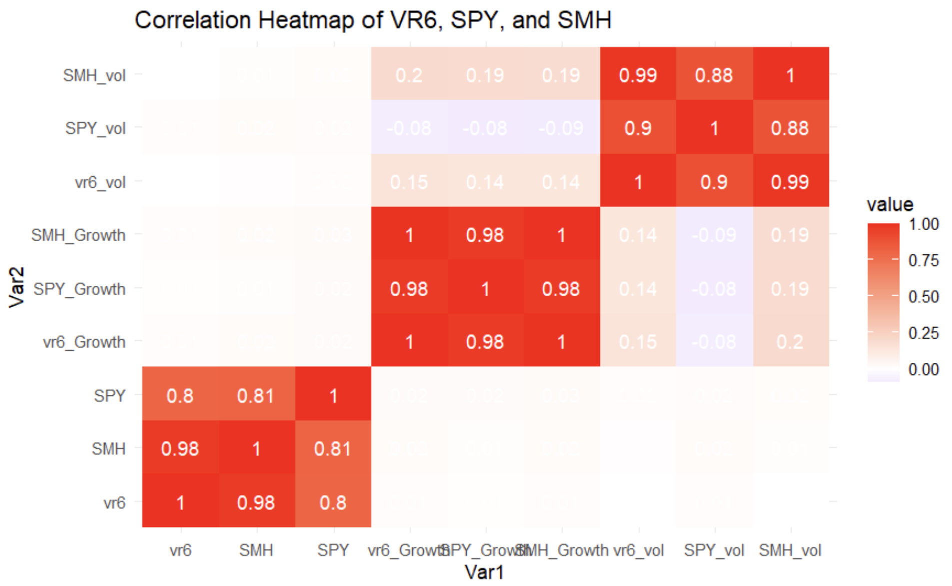
Return Chart



Risk comparison



Heat Map



Preliminay Model Stats

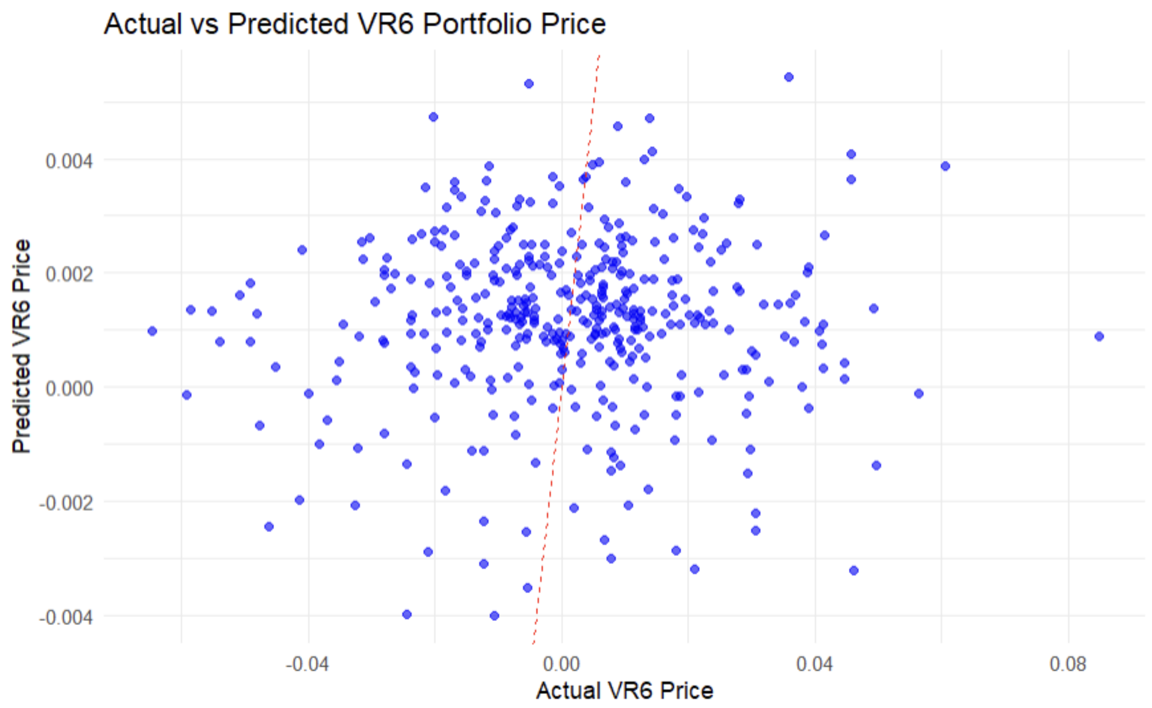
```
Call:
lm(formula = VR6_next ~ SPY_lag1 + SMH_lag1, data = train_data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.157863 -0.011878  0.000211  0.012011  0.119117 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) 0.0011312  0.0005537   2.043   0.0412 *
SPY_lag1    0.0794490  0.0797796   0.996   0.3195  
SMH_lag1    0.0401606  0.0462258   0.869   0.3851  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.022 on 1581 degrees of freedom
Multiple R-squared:  0.005992,  Adjusted R-squared:  0.004735 
F-statistic: 4.765 on 2 and 1581 DF,  p-value: 0.008644
```

Initial Predictions: Actual vs Predited



Enhanced Model Stats

```
Call:
lm(formula = VR6_next ~ SPY_lag1 + SMH_lag1 + SPY_MA7 + SPY_MA30 +
    SMH_MA7 + SMH_MA30 + SPY_Vol7 + SPY_Vol30 + SMH_Vol7 + SMH_Vol30,
    data = train_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.151435 -0.011394  0.000791  0.012327  0.113989

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.004273   0.001711   2.497   0.0126 *
SPY_lag1     0.152426   0.088098   1.730   0.0838 .
SMH_lag1     0.018155   0.049982   0.363   0.7165
SPY_MA7     -0.670867   0.295696  -2.269   0.0234 *
SPY_MA30    -0.758948   0.665276  -1.141   0.2541
SMH_MA7      0.138793   0.158421   0.876   0.3811
SMH_MA30    -0.145749   0.351999  -0.414   0.6789
SPY_Vol7    -0.599468   0.235288  -2.548   0.0109 *
SPY_Vol30    0.543814   0.273227   1.990   0.0467 *
SMH_Vol7     0.089670   0.145859   0.615   0.5388
SMH_Vol30   -0.216784   0.197697  -1.097   0.2730
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Call:
lm(formula = VR6_next ~ SPY_lag1 + SPY_MA7 + SPY_MA30 + SPY_Vol7 +
    SPY_Vol30, data = train_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.150216 -0.011392  0.000834  0.012181  0.115192

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.003215   0.001239   2.596 0.009519 **
SPY_lag1     0.179197   0.050242   3.567 0.000373 ***
SPY_MA7     -0.465231   0.179701  -2.589 0.009718 **
SPY_MA30    -1.008166   0.401883  -2.509 0.012222 *
SPY_Vol7    -0.480645   0.161467  -2.977 0.002958 **
SPY_Vol30    0.287633   0.160406   1.793 0.073143 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02208 on 1554 degrees of freedom
Multiple R-squared:  0.01687,   Adjusted R-squared:  0.01371
F-statistic: 5.334 on 5 and 1554 DF,  p-value: 7.231e-05
```

Improved Predition accuracy