Style analysis of Warren Buffet using Fama-French five-factor benchmark model

Astarag Mohapatra[a]
[a]Indiana University at Bloomington, Luddy School of Informatics, 47408

Abstract. This paper presents an in-depth analysis of Mr. Warren Buffet's investment style, showcasing his strategy of value investing with conservative stocks. Our study explores the correlation between the returns generated by Berkshire Hathaway, Mr. Buffet's holding company and the multi-factor Fama-French model. We employ the five-factor Fama-French model to assess the HML (High minus Low) and CMA (Conservative minus Aggressive) factors' factor loadings, which suggest a positive correlation with the returns. Conversely, the SMB (Small minus Big) factor demonstrates a negative correlation, indicating Mr. Buffet's tendency to avoid small-cap companies and invest in well-established firms. Furthermore, we conducted a grid search with multiple models and determined that the ARIMA(1,0) x SARMA(0) model had the best performance, as evidenced by the BIC and AIC scores (-15858.954 and -15899.773, respectively). It is crucial to note that these results establish a correlation, rather than causation, between the returns of value and conservative stocks and Mr. Buffet's portfolio. Our findings provide valuable information to investors seeking to understand Mr. Buffet's approach to investing and the factors that drive his success.

Keywords: style analysis, regression, time series, factor loadings, fama-french. SARIMAX, Warren Buffet.

1 Introduction

Fama-French in 1992 introduced factor models to access the risk of a portfolio. The CAPM (Capital Asset Pricing Model) had limitations as it didn't consider liquidity and financial distress premium. The CAPM only talks about the market premium factor, where we have a broad market index portfolio like SP500 and we compute its returns and subtract the risk-free rate from it to give the market return. Fama and French introduced two factors, namely SMB and HML, where the former creates a market-neutral portfolio by going long in small-cap stocks and going short in large-cap stocks, and the latter goes long on value stocks and short on growth stocks. Value stocks are companies that have a lower market value (number of shares outstanding multiplied by the share price) as compared to their book value. The growth companies trade at a premium compared to the value stocks, as the latter has a possibility of financial distress and bankruptcy risk, so this additional risk demands higher expected returns from the investors. Later on, factors like

CMW (Conservative minus Weak) and RMW (Robust minus Weak) were added to characterize the risk of a portfolio from robust and conservative stocks. The CMW and HML factors signal the propensity of investment toward conservative and value stocks respectively. This paper explores the style analysis of Mr. Warren Buffet, who is considered the greatest investor of all time. His focus on buying wonderful companies at wonderful prices has helped him to be the richest investor. He primarily focuses on value investing in conservative stocks, which we will explore from the fama-french model and show empirically in this paper. The paper is formatted as follows: sections 2 and 3 discuss the dataset and hypothesis, section 4 discusses the testing methodology, sections 4 and 5 shows the results and discussion, and Section 6 has conclusions, limitations, and future research directions.

2 Dataset description

First, we took the data from the 13F filling of Warren Buffet and computed his returns for the period 2000-2018. Then, we downloaded the Fama-French five-factor model data from here. It has daily data of the five factors with the description as follows

- SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios
- HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios,
- RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios

- CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios
- Mkt (Market Return): Return for a broad market index like $S\&P500$
- Rf is the risk-free rate like the treasury bill rate.

3 Hypothesis

The purpose of our study is to examine the factor loadings of a multi-factor Fama-French model and to determine whether the returns of Mr. Warren Buffet are more highly correlated with the value factor HML and conservative factor CMW than with other factors, with the exception of the market premium factor. Our null hypothesis posits zero factor loadings for HML and CMW, while the alternative hypothesis suggests non-zero and positive factor loadings for these two factors. Furthermore, we aim to demonstrate negative factor loadings for the SMB factor, indicating Mr. Buffet's preference for investing in large-cap stocks. This investigation seeks to contribute to the literature by shedding light on the investment style of one of the greatest investors of all time, and by providing empirical evidence regarding the relationship between Mr. Buffet's portfolio returns and the Fama-French factors.

4 Methodology

To begin our analysis, we examined the autocorrelations Appendix (Figure 1-6) and cross-correlations of the returns (Figure 7). As stock market prices are widely considered to be governed by random walks, the autocorrelation plots exhibit a rapid drop below the significance level after lag 1 for all input and output variables. Similarly, the cross-correlation plots reveal minimal correlations between variables, with no discernible trends. Additionally, we conducted a spectral analysis of the data, as illustrated in Figure 8. Here we can see that variables cover all the frequencies with no visible spikes. Based on the results of the autocorrelation, cross-correlation, and spectral analysis plots, we conclude that our independent variables are largely independent, with minimal cross-correlations. Hence, no preprocessing or detrending is required for our independent variables.

After conducting exploratory data analysis, we conducted a grid search SARMIAX model with parameters $ARMA(p, d, q)SARMA(sp, sd, sq, m)$. The hyperparameters p and sp were used for the AR component along with seasonality, while d and sd were used for the difference operator with and without seasonality. q and sq were used for MA components, with the latter for seasonality, and finally m was used for seasonal multiplicity. The range of hyperparameters tested were $p \in [0, 1, 2, 3, 4, 5]$, $d \in [0, 1, 2, 3, 4, 5]$, $q \in [0, 1, 2, 3, 4, 5]$, $sp \in [0, 1, 2, 3, 4, 5]$, $sd \in [0, 1, 2, 3, 4, 5]$, $sq \in [0, 1, 2, 3, 4, 5]$, and $m \in [0, 63, 252]$. Seasonal multiplicity refers to quarterly and yearly data seasonality. The weights and biases project report for all the grid search sweeps are here [1], [2], [3].

5 Results and Discussion

From our extensive grid search, we found that higher difference order hyperparameters lead to higher AIC and BIC values, hence the preferred value for it is 0. For our model selection, we choose the following configurations that gave the best results

- ARMA(1,0,0) SARMA(0,0,0,0): $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$, where y_t represents the time series variable at time t, α_0 is the intercept, α_1 is the coefficient of the AR term, y_{t-1} is the lag-1 value of the time series variable, and ϵ_t is the error term. Since the model has no MA terms, there are no additional error terms to consider. **BIC:-15858.954 and AIC: -15899.773**

- ARMA(5,3,4) SARMA(0,0,0,0): $(1-\phi_1 L-\phi_2 L^2-\phi_3 L^3-\phi_4 L^4-\phi_5 L^5)^3 y_t = c+(1+\theta_1 L+\theta_2 L^2+\theta_3 L^3+\theta_4 L^4)\epsilon_t$, where $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5$ are the AR coefficients, $\theta_1, \theta_2, \theta_3, \theta_4$ are the MA coefficients, L is the lag operator, c is a constant term, y_t is the time series data, and ϵ_t is the error term. The seasonal component of the model is not present as all SARMA coefficients are zero. **BIC:-15294.448 and AIC: -15382.678**

- ARMA(3,0,4) SARMA(0,0,0,63): $y_t = c+\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\theta_3\varepsilon_{t-3}+\phi_1 y_{t-4}+\phi_2 y_{t-5}+\phi_3 y_{t-6}+\varepsilon_{t-63}$. Here, y_t denotes the time series variable at time t, ε_t denotes the error term at time t, c is a constant term, θ_1, θ_2, and θ_3 are the autoregressive coefficients, ϕ_1, ϕ_2, and ϕ_3 are the moving average coefficients, and ε_{t-1}, ε_{t-2}, ε_{t-3}, and ε_{t-63} are the error terms at the previous time steps. The notation SARMA(0,0,0,63) indicates that there is no seasonal autoregressive or moving average component, but there is a seasonal component with a multiplicity of 63. **BIC:-15850.93 and AIC: -15930.229**

From the BIC and AIC scores for model selection, we choose the model ARMA(1,0,0) SARMA(0,0,0,0). Also, the residuals are white noise with p-value from box testing is 4.88×10^{-4}. So the residuals are white noise as the p-value is below the 95% confidence interval significance value. The plots for the periodogram and auto-correlation is in Figure 9 and 10.

Now, we will examine the coefficients for all the factors along with the AR component. It is presented in Table 1. From the table, you can see that the weighting for Market premium, HML and CMA is high. It is understandable that Market premium has the highest factor loading because the systematic risk component in a portfolio is always high. The HML and CMA factor is also pronounced with values of 0.1447 and 0.2266 respectively. Also, the SMB factor has a negative loading, which reinforces the hypothesis that Mr. Buffet stays away from small-cap stocks.

Also, note that all the coefficient values are statistically significant as the p-values from the model summary are below the statistical significance level of 0.05.

Factor	Coefficient
Market Premium (Mkt-RF)	0.6342
SMB (Small Minus Big)	-0.188
HML (High Minus Low)	0.1447
RMW (Robust Minus Weak)	-0.828
CMA (Conservative Minus Aggressive)	0.2266
Intercept	0.000105

Table 1: Coefficients of different factors

The main caveats are the relation between correlation and causation. The findings demonstrate that Mr. Buffet's investment strategy involves investing in value and conservative stocks, resulting in remarkable returns over the years. However, it does not imply that investing exclusively in value and conservative stocks will necessarily lead to similar returns as Mr. Buffet's. Achieving such outcomes demands a wealth of expertise in corporate finance and a deep understanding of business operations. Therefore, this analysis highlights the importance of investors performing their own due diligence in stock selection and gaining insights from Mr. Buffet's approach to investing in value-based companies with conservative business models.

6 Conclusion

In recent years, there has been a growing interest in understanding the investment strategies and stock selection criteria of successful investors. One of the most well-known investors is Warren

Buffett, who is renowned for his long-term approach to investing and his ability to consistently outperform the market. As such, there is considerable interest in analyzing the factors that drive Mr. Buffett's investment decisions. In this study, we seek to investigate the factor loadings of a multi-factor Fama-French model and assess whether the correlation between Mr. Buffett's returns and the value factor (HML) and conservative factor (CMA) is higher compared to other factors, except for the market premium factor. Understanding the factors that Mr. Buffett considers when making his investment decisions may provide valuable insights for other investors seeking to improve their investment strategies.

In addition to the analysis presented here, an interesting direction for future research could be to investigate whether there is any style drift in Mr. Buffet's investment strategy over time. This could be done by performing a rolling window analysis on the data and conducting style analysis on a yearly basis. The aim of this analysis would be to investigate whether there were any periods where Mr. Buffet deviated from his typical investment style, which could potentially provide insights into his decision-making process and the factors that influence his investment decisions.

References

1 Barberis, Nicholas, and Andrei Shleifer. "Style investing." Journal of financial Economics 68, no. 2 (2003): 161-199.

2 Froot, Kenneth, and Melvyn Teo. "Style investing and institutional investors." Journal of Financial and Quantitative Analysis 43, no. 4 (2008): 883-906.

3 Fama, Eugene F., and Kenneth R. French. "The cross-section of expected stock returns." the Journal of Finance 47, no. 2 (1992): 427-465.

4 Seabold, S. Perktold, J., 2010. statsmodels: Econometric and statistical modeling with python.

In 9th Python in Science Conference.

7 Appendix

```
import pandas as pd
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
plt.style.use('seaborn-dark-palette')
%matplotlib inline

#Reading the dataset
brka_rets = pd.read_csv('brka_d_ret.csv',parse_dates=True, index_col=0)
fff_data = pd.read_csv('F-F_Research_Data_5_Factors_2x3_daily.CSV')

#Calculating the compounding for returns
def compound(r):
    """
    returns the result of compounding the set of returns in r
    """
    return np.expm1(np.log1p(r).sum())

brka_m = brka_rets.resample('D').apply(compound).to_period('D')

brka_m['DATE'] = brka_m.index
brka_m['DATE'] = brka_m['DATE'].apply(lambda x:pd.to_datetime(str(x),format=
brka_m.set_index('DATE',inplace=True)

fff_data['DATE'] = fff_data['DATE'].apply(lambda x:pd.to_datetime(str(x),for

fff_data.set_index('DATE',inplace=True)
fff_data = fff_data/100

for col in fff_data.columns[1:5]:
    fff_data[col] = fff_data[col]-fff_data['RF']

df = pd.merge(left=brka_m, right=fff_data , left_index=True,how='left',right_in
df['BRKA-RF'] = df['BRKA']-df['RF']

from statsmodels.tsa.statespace.sarimax import SARIMAX
```

```
# from statsmodels.tsa.statespace.sarimax import

Y_train = df.loc["2008":"2017"]['BRKA-RF']
X_train = df.loc["2008":"2017"][['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA']]

Y_test = df.loc["2018":]['BRKA-RF']
X_test = df.loc["2018":][['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA']]

# order_ar = [0,1,2]
order_ar = [3,4,5]

order_diff = [0,1,2]
# order_diff = [3,4,5]
# order_mr = [0,1,2]
order_mr = [3,4,5]

# seasonal_ar = [0,1,2]
# seasonal_diff = [0,1,2]
# seasonal_mr = [0,1,2]
seasonal_m = [0,63,252]

sweep_config = {
    "method":"grid",
    "parameters":{
    "ar":{"values":order_ar},
    "diff":{"values":order_diff},
    "mr":{"values":order_mr},

    # "s_ar":{"values":seasonal_ar},
    # "s_diff":{"values":seasonal_diff},
    # "s_mr":{"values":seasonal_mr},
    "s_m":{"values":seasonal_m},
    }
}

def main():
    run = wandb.init("Time-Series-Project")

    start = time.time()
    model = SARIMAX(
        Y_train.values,
        X_train.values,
        order=(wandb.config.ar,wandb.config.diff,wandb.config.mr),
        seasonal_order=(
        # wandb.config.s_ar,wandb.config.s_diff,wandb.config.s_mr,
```

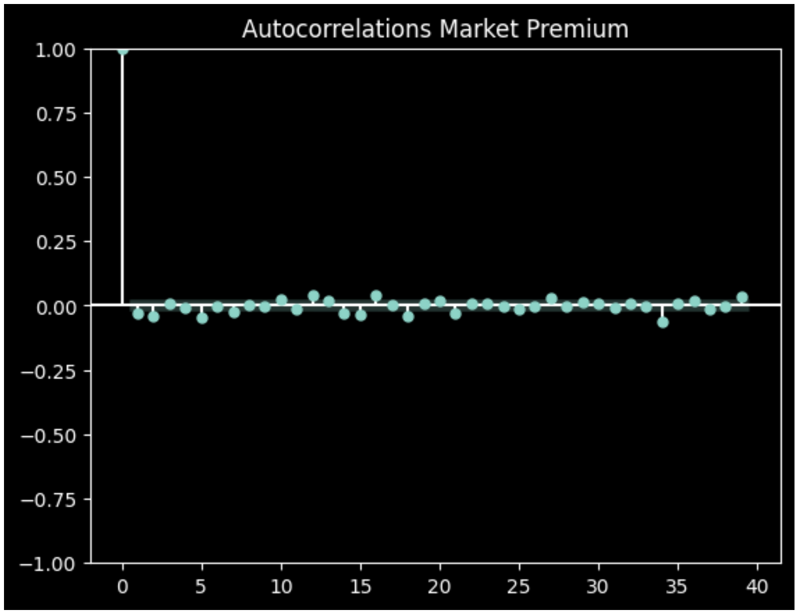

Fig 1: Autocorrelations for Market Returns

```
        0,0,0,
        wandb.config.s_m
        )
    )
    end = time.time()

    results = model.fit()
    coefs = results.params
    print(coefs)
    wandb.log({
        "AIC":results.aic,
        "BIC":results.bic,
        "Coefficients": coefs,
        "time":round(end-start,2)
    })

sweep_id = wandb.sweep(sweep=sweep_config, project="Time-Series-Project")
wandb.agent(sweep_id, function=main)

model = SARIMAX(
    Y_train, X_train,(1,0,0),(0,0,0,0)
)
res = model.fit()
test_res = model.predict(X_test)
res.summary()
```

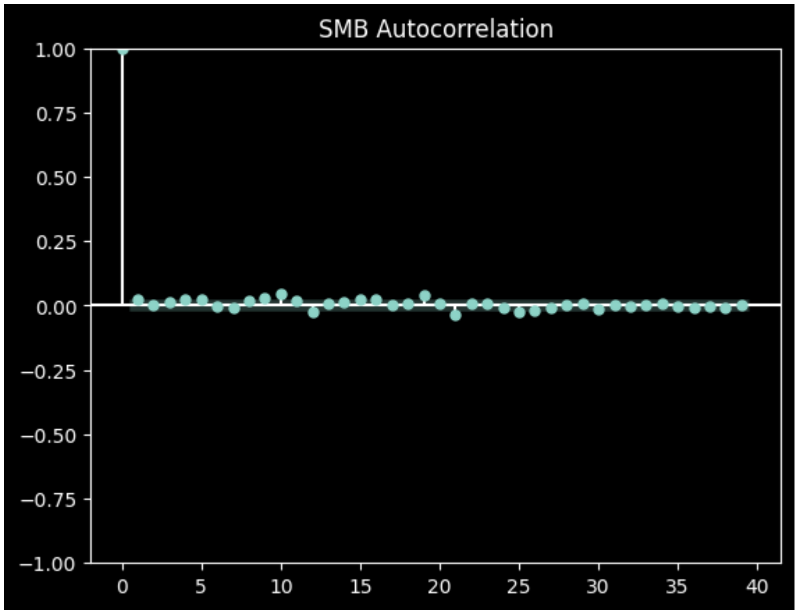

Fig 2: Autocorrelations for SMB

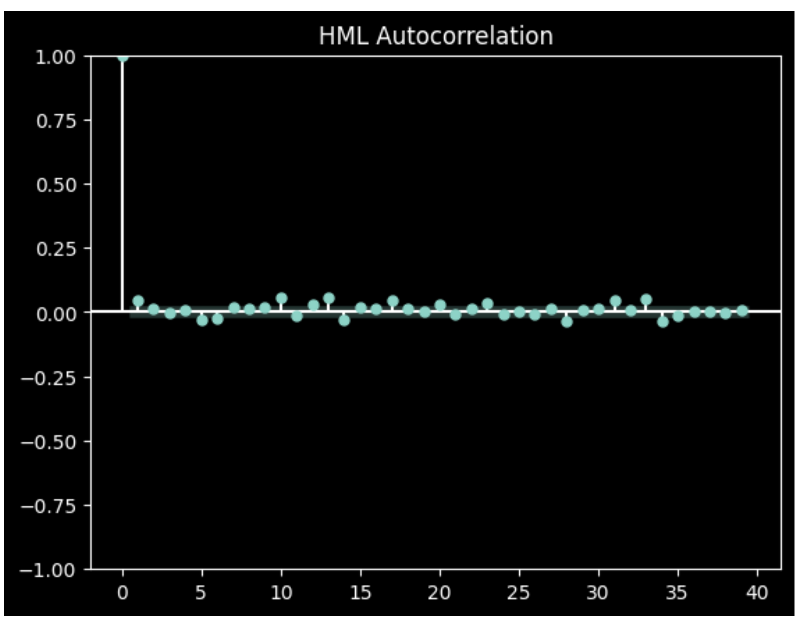

Fig 3: Autocorrelations for HML

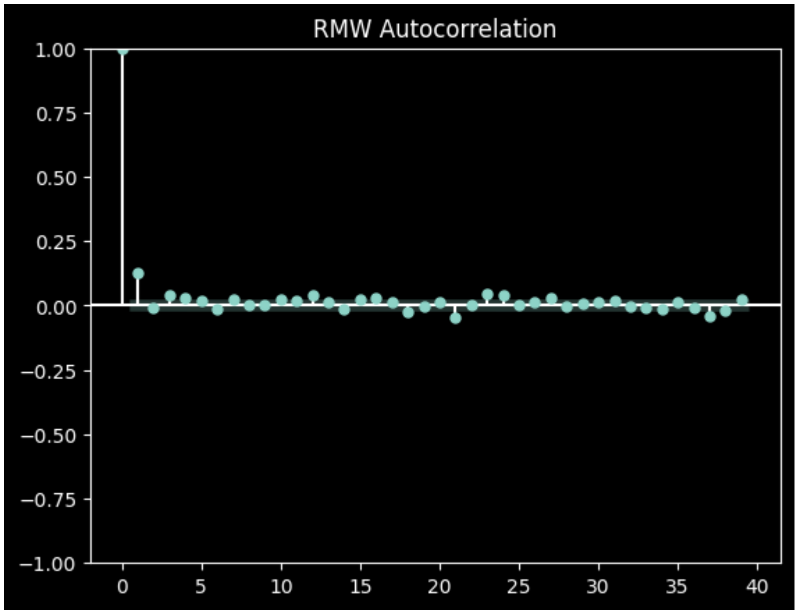

Fig 4: Autocorrelations for RMW

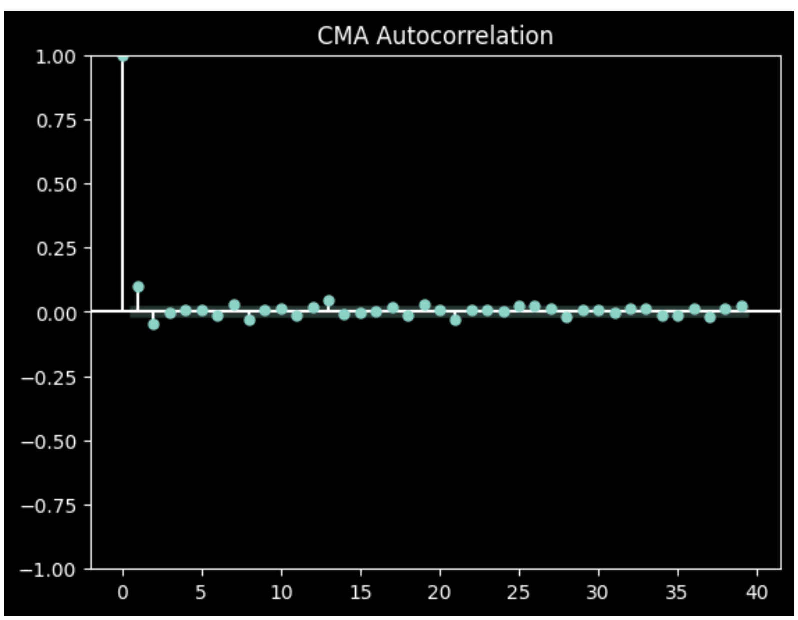

Fig 5: Autocorrelations for CMA

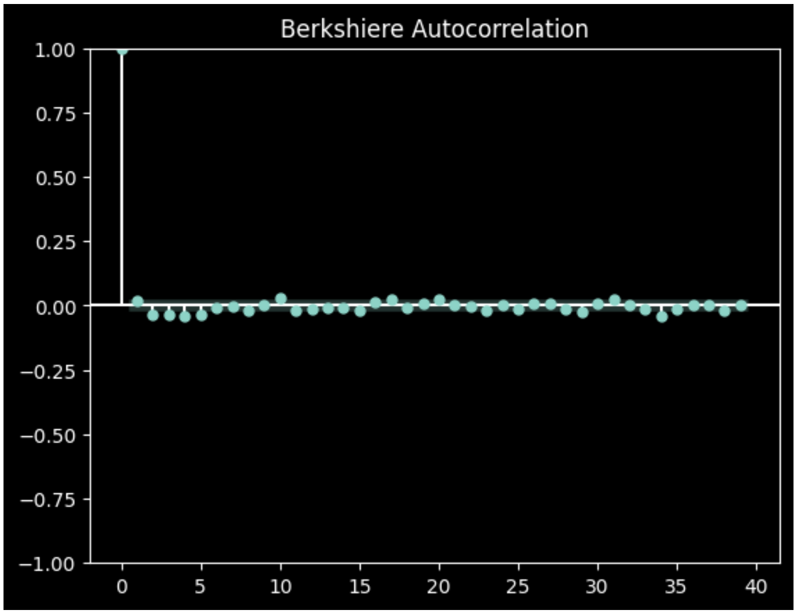

Fig 6: Autocorrelations for Berkshiere Returns

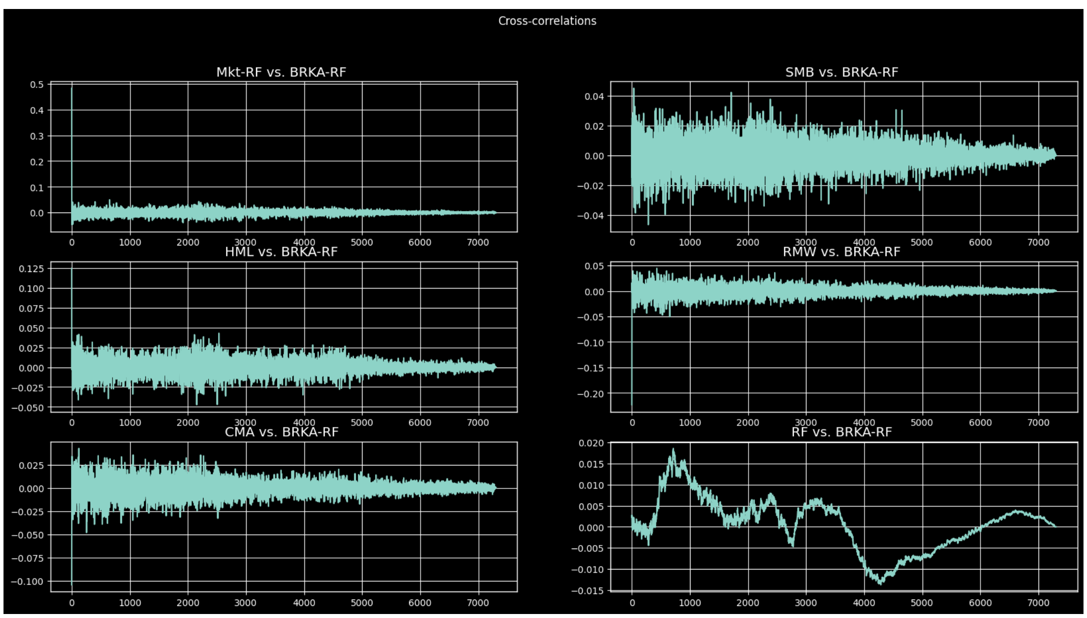

Fig 7: Cross-correlations of all the input variables

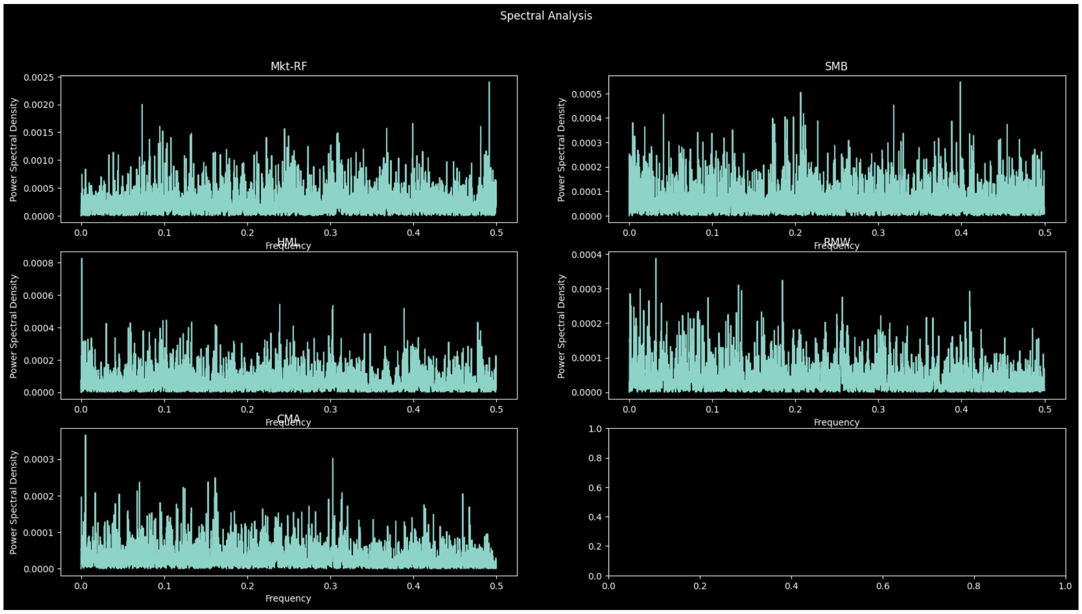

Fig 8: Spectral Analysis of all the input variables

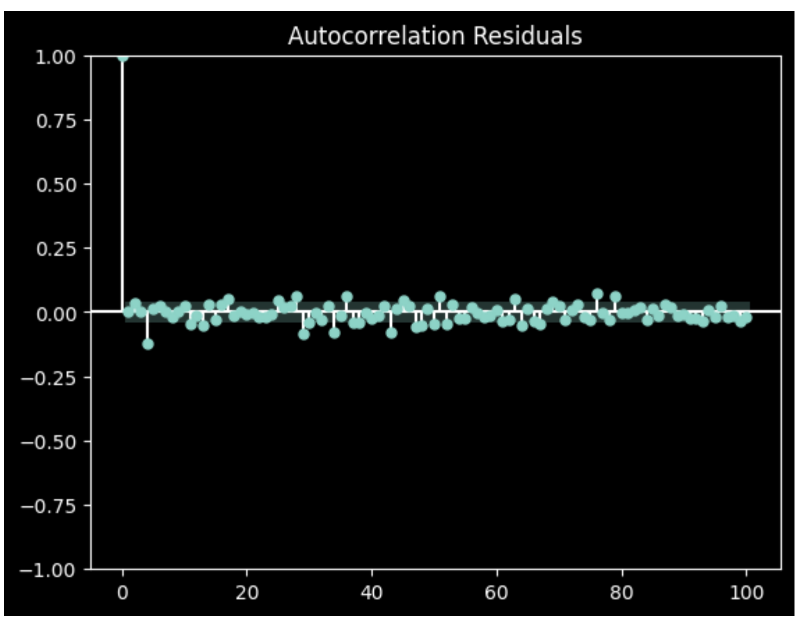

Fig 9: Auto-correlations of Residuals

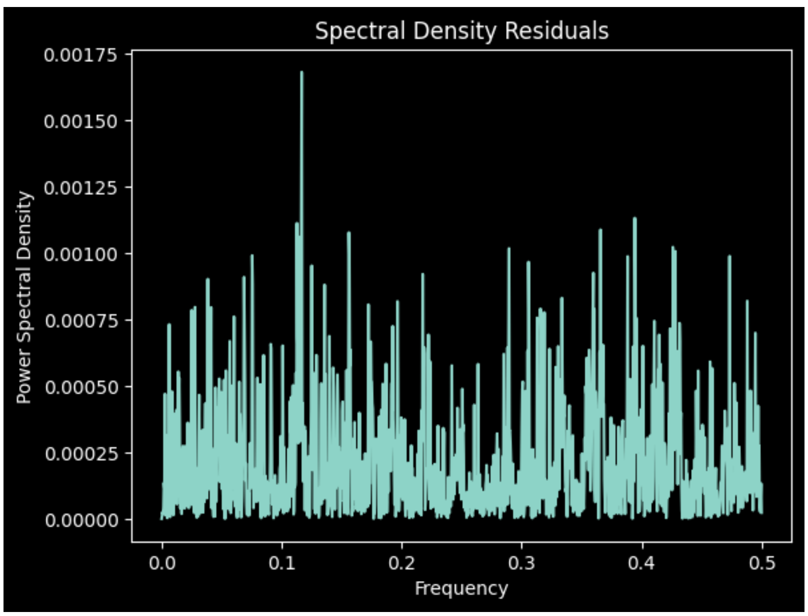

Fig 10: Residual Periodogram