Inter-Temporal Volatility Model: A Multifactor Linear Regression Approach

Akil Huang
University of Toronto, Department of Statistical Sciences
June 2024

Abstract

Financial market volatility is a critical indicator of market stability and investor sentiment. Understanding and predicting volatility is a key task in portfolio management, risk assessment, and asset hedging. This paper studies the relationship between macroeconomic and technical indicators, such as S&P500 Trading Volume, CBOE Volatility Index, and GDP with market volatility. Our results show that GLMs exhibit strong predictive power over rolling moving average models to robustly explain and predict volatility in the S&P500 financial market index.

Keywords: Volatility, Multiple Linear Regression, Autoregressive Volatility Model, Fixed Income Markets, CBOE Volatility Index, FRED.

1 Introduction

The S&P500 Index is a large-cap equities index tracking the top 500 leading publicly traded companies. As a stock index, it is subject to periods of low and high volatility, often contingent on major political or economic events. Since the 2008 global financial crisis, financial economists and large institutional investors – which comprise large mutual funds, pension plans and banks, or other financial institutions – have been particularly interested in forecasting and hedging against financial and macroeconomic risk and volatility.

In the context of portfolio management and trading, volatility is defined as the dispersion of returns on an asset or portfolio. As such, the identification of robust predictors of volatility allows institutional investors to make timely forecasts and hedge their portfolios against extreme market movements.

Academic literature on market volatility has predominantly employed GARCH and stochastic models for prediction, which can be difficult to interpret and lack consideration for economic variables (Lim & Sek, 2013). While linear models have a computational and interpretational advantage above other more complex autoregressive and nonlinear models, linear regression models of volatility are rarely used in studies of volatility and risk forecasting. Moreover, existing literature rarely employs models with both lagged autoregressive terms and macroeconomic variables. In fact, recent literature has shown that macroeconomic variables can be effective at predicting stock market movements (Ma et al., 2022). This paper seeks to understand the predictive power of blended multivariate linear regression models to forecast conditional future market volatility.

2 Methods

2.1 Data Collection

Data on key macroeconomic and technical indicators was gathered from Yahoo! Finance as well as the Federal Reserve Economic Data (FRED) online database. The time series datasets were pulled through the FRED API and YFinance Python API, an open source tool to scrape and manipulate publicly available data on the Yahoo! Finance platform.

A preliminary selection of time series data included the S&P500 index (adj. close) and daily Trading Volume, US Gross Domestic Product (GDP), Consumer Price Index (CPI), Federal Funds Rate (Interest Rate), US Effective Corporate Bond Yields, National Unemployment Rate, Privately-Owned Housing Unit Starts, Nominal Home Prices, the US Dollar Index, and the Chicago Board Options Exchange's (CBOE) Volatility Index (VIX). These independent variables were picked because of their macroeconomic significance and potential explanatory power.

Variables such as GDP, CPI, Unemployment Rate, Home Prices, and Housing Starts measure national economic activity through a variety of metrics: US GDP, or Gross Domestic Product, is a measure of aggregate income generated by all US residents in a given year; CPI, or Consumer Price Index, is a national index which measures the price of a selected basket of goods; Housing Starts measures the number of new housing projects across the US, in thousands of residence units; Home Prices is measured by the S&P Case-Shiller U.S. National Home Price Index, which measures the change in the aggregate value of all existing single-family housing.

Other variables, such as the US Dollar Index, the VIX, Corporate Bond Yields, Interest Rates, S&P500 Trading Volume, are technical indicators which may indirectly explain volatility. The VIX is an indexed measure of implied future volatility stemming from the options market's expectations of volatility 30 days into the future. Given its popular usage in measuring volatility and market sentiment, it was selected as a predictor of realized index volatility of the S&P500. Corporate Bond Yields, similar to the Federal Funds Rate (or Interest Rate), was gathered from ICE BofA BBB US Corporate Index. While Corporate Bond Yields are the posted rate of return an investor expects to receive, the Federal Funds Rate is the set rate at which US banks lend and loan reserves overnight. When interest rates are hiked, equity markets tend to falter (Kim, 2023).

In addition, the US Dollar Index measures the relative strength of the US Dollar compared to a basket of foreign currencies. The strength of the US Dollar has implications on domestic investment in capital (e.g. investment in large-cap stocks in the S&P500). Lastly, the level of daily S&P500 Trading Volume can provide important signals of volatility: when trading volume increases, this could be an indication of imbalances in order-flow or liquidity, which may cause increased price movement.

The response variable, Realized Volatility, was gathered from the S&P500 ticker on Yahoo! Finance. Throughout this paper, Realized Volatility (RV_t), will be used as a measure of empirical volatility. Realized Volatility is calculated as the sum of squared returns over a 21-day trading period. This measure is commonly used in existing literature to represent actual market volatility (Haugom et al., 2014). Realized volatility is measured in terms of percentage returns.

$$RV_t \equiv \sqrt{\sum_{i}^{n} r_{t,i}^2} \tag{1}$$

2.2 Data Cleaning

Individual time series were collected and concatenated into a DataFrame, containing data ranging from January 1, 2010 to January 1, 2019. This time period was selected to exclude particularly rare and significant market events (e.g. The 2008 GFC, the 2020 Covid-19 Pandemic), which affect the assumptions underlying linear regression models. Missing values were handled through forward-fill imputation. This procedure was appropriate given the nature of the missing data, which largely arose due to inconsistencies in reporting periods (e.g. daily vs. quarterly data). After scraping Yahoo! Finance for S&P500 index data, the adjusted close prices of the index were collected. Daily percentage returns and sum of squared returns were calculated.

Four additional time series were generated from existing time series data: GDP Growth was calculated as the quarter-over-quarter percentage change in nominal GDP; One-Month Lagged VIX was calculated by lagging the VIX by one month, of which there are 21 trading days; One-Month Lagged Volatility was calculated by lagging Realized Volatility by 21 trading days; One-Week Lagged Volatility was calculated by lagging Logged Volatility by 5 trading days.

$$Y_t = C + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t \tag{2}$$

In order to satisfy the assumptions of multiple linear regression, several preliminary data analyses were conducted. Data for the response variable, Realized Volatility, was plotted in time series plots and histograms to identify the distribution of the data. The plot showed a systematic imbalance in the response, which was corrected for with a variable logarithmic transformation.

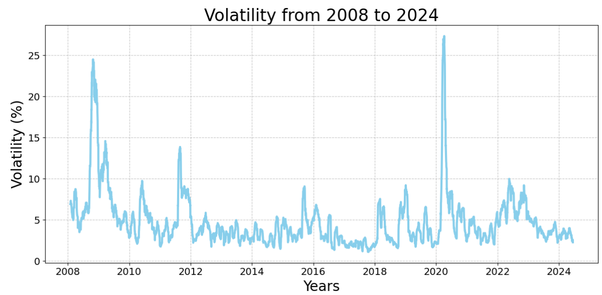

Figure 1: Plot of S&P500 Index volatility between 2008 to 2024.

Descriptive statistics (shown in appendix) were collected on all variables to identify discrepancies, which were addressed later on. Additionally, correlation plots of the regressors and VIF tests were conducted to identify and address strong multicollinearity between predictors. After having conducted an initial exploratory analysis, a preliminary multiple linear regression was run on a random subset of the initial dataset. Training, validation, and test sets were generated from the cleaned dataset and randomly sampled in an 80-10-10 split respectively. This was done to avoid data leakage when training the model. The initial linear regression was shown to have high AIC, high parameter standard errors, non-normally distributed residuals, and relatively poor fit to the data. To validate the model for prediction, the MSE of the model's prediction on a test set was benchmarked against other models in consideration to determine the model with the lowest prediction error. MSE is a widely used metric for prediction models on continuous data (Lim & Sek, 2013).

Several steps were undertaken to correct for problematic data. Regressor variables with large standard errors were standardized to reduce the range of their values. Moreover, various regressors such as the Lagged Volatility terms, were additionally log transformed to normalize their distributions. After independent and dependent variables were scaled and transformed, steps were taken to reduce the effect that outliers and highly influential points had on the regression model.

High influence/high leverage points and outliers were removed using a combined method of removing standardized residuals beyond the threshold ($t > 2$), high leverage points ($h_{ii} > 2\frac{p}{n}$), and points that exceeded the threshold Cook's Distance ($D_i >$ 50'th percentile of $F(2409, 2401)$). This allowed us to filter out anomalies and high-variance data, which would reduce the ability for the predictive model to generalize to future data. High leverage/influence points do not generally provide useful information for financial models, largely because incidents of inextricably high volatility tend to be very unpredictable (e.g. 2010 Dow Jones Flash Crash). Given that anomalies increase the standard deviation of parameters, as well as increase predictive bias, it is important to remove them. Additionally, high-variance variables were removed if they were determined to systematically skew the Q-Q plot of residuals, or if they exhibited high VIF scores.

$$X_i^* = \frac{X_i - \overline{X}}{S_X} \tag{2}$$

Finally, after correcting for issues in the data and model parameters, a secondary set of regression models were run. Through backwards-elimination involving analysis of model AIC scores, partial F-tests and F-statistics, adjusted R-squared, Durbin-Watson autocorrelation scores, and individual parameter t-tests, a final regression model was determined and applied to a validation dataset to test its predictive power. The model which balanced the lowest AIC, highest F-statistic and adjusted R-squared, fewest parameters, and ensured normality of the errors and response was chosen.

3 Results

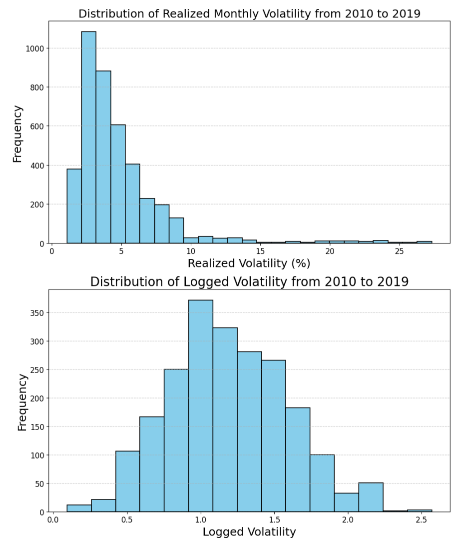

Figure 2: Plot of Realized Volatility before and after applying a transformation on the response.

In our exploratory analysis, plots of the response variable were generated to identify general trends and anomalies. A histogram of Volatility showed that the response variable followed a right-skewed distribution with heavy tail. A logarithmic transformation was applied to normalize the distribution of data. Log transformations also help to ensure stationarity, a key property required for time series data. The resulting distribution looked approximately normal, and this was later verified in a Q-Q plot (shown in appendix).

A preliminary VIF test identified heavy multicollinearity involving the GDP, CPI, Housing Starts, Unemployment Rate, and Home Prices (> 5). These predictors were removed to reduce the effect that intercorrelations had on widening standard error.

The resulting variables which were retained demonstrated an acceptable level of multicollinearity.

Feature	**VIF**
One-Month Lagged Volatility	2.635452
Volume	1.250249
One-Month Lagged VIX	3.561780
Interest Rates	1.724840
GDP Growth	1.156082
US Dollar Index	1.871245
Corporate Bond Yields	5.028675
Ten-Year Treasury Yield	3.673057

Table 1: Variance Inflation Factor (VIF) scores for selected predictive variables.

After ensuring normality of the response variable, a preliminary multiple linear regression was run on all non-multicollinear variables in an attempt to eliminate non-significant variables. The model's residuals were analyzed to ensure adherence to linear regression assumptions. A histogram plot of the residuals and a residuals vs. fitted values plot shows that the residuals were systematically unbalanced. Seeing this, GDP, S&P500 Trading Volume, and Housing Starts were standardized because they were found to have arbitrarily large values. After standardization, the model's condition number and coefficient SE's decreased, demonstrating a better model fit.

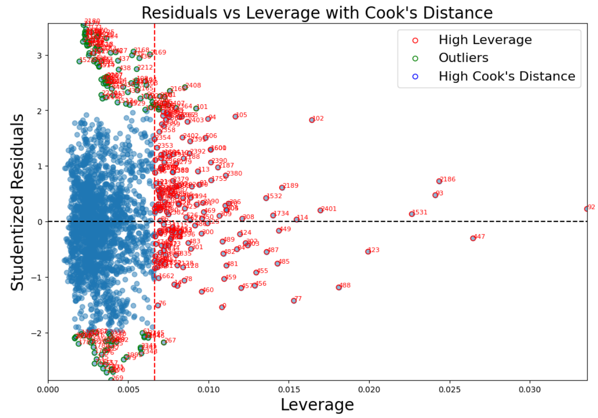

Figure 3: Plot of Residuals vs. Leverage. High leverage points, outliers, and influential points were removed to ensure high training precision and model explanatory power.

3.1 Model Selection

In addition to removing collinear variables, One-Week Lagged Volatility was also removed. Though the variable increased the coefficient of determination ($R^2 > 0.85$) of the model and reduced AIC, neither model metric were trustworthy because of the variable's tendency to skew the residuals and thus violate the assumptions of multiple linear regression. All other variables exhibited significant predictive power, other than Interest Rates. Interest Rates had a t-test p-value of $p \approx 0.01$, which was weaker than desired. As such, there were two models to test between: one with Interest Rates, and one without. A partial F-test/ANOVA test on a nested model revealed that Interest Rates contained predictive power ($p = 0.009643$) and thus was nto removed. Additionally, Interest Rates exhibited low multicollinearity with other variables, suggesting that there was not enough evidence to remove it.

The diagnostic summary of the final predictive model showed an F-statistic of 353.7, an adjusted R^2 of 0.619, and an AIC value of 99.28 – far below similar models, including the nested model without Interest

Rates. Additionally, the model's Durbin-Watson test score of serial autocorrelation was 1.941, which is within an acceptable range for $k = 8$ regressors. This demonstrates that the data transformations earlier, as well as the inclusion of a lagged response variable, corrected for dependence in the model's residual terms.

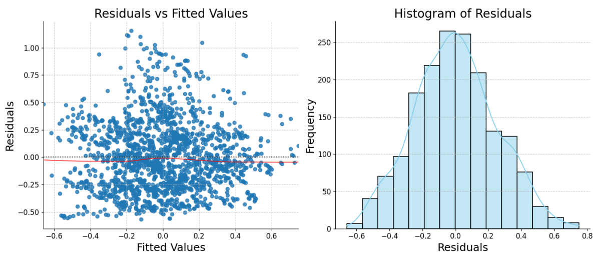

Figure 4: The distribution of residuals of the final model exhibits normality and eschews systematic patterns.

3.2 Model Prediction

After our preliminary data exploration and model tuning, we were able to generate a predictive model using our multiple linear regression (shown in appendix). The model was fit onto the training data and evaluated on the validation set. Finally, the tuned model was applied to predict logged volatility on the test set and MSE was calculated for each set.

The predictive model had a validation Mean-Squared Error (MSE) of 0.0560, a test MSE of 0.0582, and a MSE over the entire dataset of 0.605. The baseline MSE, or MSE if simply the mean of the response variable had been used for prediction, was 0.157. This disparity between a baseline model and the predictive general linear model shows that the predictive model successfully minimizes predictive errors. Moreover, the closeness in MSE between the validation set and Test set show that the model did not overfit to the training data. The standard deviation of the logged volatility was 0.401604 and the volatility response was 1.185172, suggesting that the predictions were well within a reasonable range of the response variable. Interestingly, a nested model which excluded Interest Rates achieved a similar test MSE (0.578), suggesting that the inclusion of Interest Rates may not necessarily have been better for the predictive model.

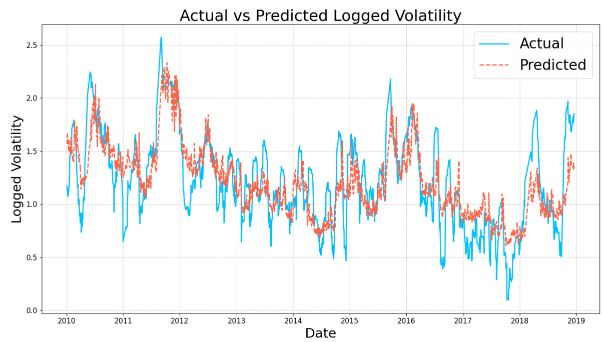

Figure 5: A plot of actual vs. predicted logged volatility shows that while general linear models do not perfectly predict volatility, they are robust methods of predicting volatility trends and movements.

4 Discussion

Looking at the model parameters, we notice interesting relationships between the dependent and independent variables. Since the primary goal of this paper was prediction, we can justifiably keep our response variable in log-transformed units in our analysis of the relationships between dependent and independent variables. While most macroeconomic indicators have negative variable coefficients, Lagged VIX, S&P500 Trading Volume, and Corporate Bond Yields exhibited positive linear relationships with Logged Volatility – with Corporate Bond Yields having the greatest linear effect on Logged Volatility ($\hat{\beta} = 0.2321$). These results correspond with existing literature given that increases in trading indicate an increase in liquidity, resulting in wider intervals of price movement (Johnson, 2008). Increases in one-month VIX forecasts also understandably result in increases in volatility.

Variables with negative signs, such as GDP Growth, signal that an increase in the independent variable results in a decrease in Logged Volatility. For instance, an inverse relationship with Treasury Yields suggests that increases in government bond yields reduce volatility. While empirical evidence shows that increased rates reduce stock prices (Arnott, 2024), we would also posit that investors have greater financial incentives in high-rate environments to substitute capital away from equity markets and towards fixed income markets, reducing the amount of overall price movement. However, it is not apparent why an increase in One-Month Lagged Volatility would signal a decrease in present-day volatility. One possible explanation is the idea of mean-reversion: as volatility increases, even though it may persist over time, eventually reverts back to its mean position – its mean level of volatility. This is perhaps why a one-unit increase in volatility from one month ago may suggest a fractional decrease in volatility in the present. In summary, the model seems to show that as the US economy tightens up (i.e. signalled by interest rate hikes and lower GDP Growth, for instance), Volatility tends to increase. This finding corroborates existing research into volatility and reinforces the use of both macroeconomic and technical indicators for forecasting.

It is important to note the limitations of this predictive model. Even with variable transformations, outlier detection, standardization, and careful feature selection, the model still had a tendency to slightly overfit on the training dataset. The same variables with significant predictors were not necessarily as significant when applied to a test or validation set during different time periods. This is likely due to the highly speculative and quickly-changing nature of equity markets.

These limitations, while important to consider, were not fully addressed due to capacity limitations. Gathering more data was impossible, and unlikely to increase the strength of the model given that Volatility data before 2008 likely would have had low signal for today's trading conditions. Additionally, issues of autocorrelation and systematic multicollinearity – which were mitigated through the addition of lagged AR terms – are modeled with far more precision with ARIMA and other higher-order AR models. Lastly, issues of non-constant variance, which were mostly remedied through transformations and normalization, may require higher-level procedures not taught in this course. Considering that the construction of most modern-day volatility forecasting models is often done in large financial institutions, with large teams that gather and clean private data sets, macro-factor models such as the one in this paper are still quite impressive.

Though GLMs have been shown to have limitations with regards to their strict assumptions and predictive limitations, their interpretability, scalability, and ability to robustly predict time-series volatility make them a strong candidate for market volatility prediction tasks. We developed a model using key macroeconomic indicators, technical indicators, as well as lagged autoregressive predictors which was capable of explaining variance in the response variable. Future research in this area should explore the use of GLMs with a wider range of alternative predictors, or include interaction terms that provide greater explanatory power.

5 Appendix

	Logged Volatility	One-Week Lagged Volatility	One-Month Lagged Volatility	Volume	One-Month Lagged VIX	Interest Rates	GDP	GDP Growth
mean	1.233321	1.232976	1.231350	3.712963e+09	16.995888	0.441985	17642.939507	0.994494
std	0.454221	0.452111	0.449655	8.458039e+08	5.5776856	0.577014	1780.806880	0.433561
min	0.093089	0.093089	0.093089	1.025000e+09	9.140000	0.070000	14764.610000	0.033265
25%	0.920107	0.921865	0.925154	3.224592e+09	13.020000	0.100000	16207.115000	0.666540
50%	1.181258	1.178750	1.183134	3.591965e+09	15.590000	0.160000	17804.228000	1.022055
75%	1.534230	1.528121	1.518129	4.100412e+09	19.202500	0.400000	18941.824000	1.342486
max	2.628552	2.628552	2.628552	1.061781e+10	48.000000	2.270000	20917.867000	1.865187

	CPI	Unemployment Rate	Housing Starts	Home Prices	US Dollar Index	Corporate Bond Yields	Ten-Year Treasury Yield
mean	235.067673	6.500332	960.753738	165.707000	87.301051	3.989319	2.432824
std	9.601243	1.969033	246.581846	21.817453	8.222655	0.505363	0.560458
min	217.199000	3.700000	517.000000	133.999000	72.930000	3.160000	1.370000
25%	228.590000	4.800000	723.000000	144.310000	80.169998	3.580000	2.000000
50%	236.222000	6.100000	1007.000000	166.240000	84.150002	3.910000	2.360000
75%	241.317250	8.200000	1177.000000	183.960000	95.250000	4.360000	2.830000
max	252.772000	9.900000	1357.000000	205.377000	103.290001	5.580000	4.010000

Table 2: Descriptive statistics for macroeconomic and technical variables before scaling and data transformations were applied.

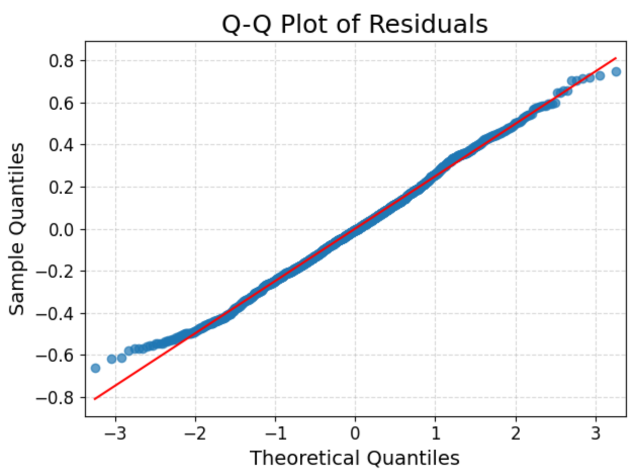

Figure 6: Q-Q plot of the residuals of the final predictive model. The residuals of the multiple linear regression model generally follow the theoretical quantiles when adjusted for the different ranges.

	coef	std err	t-values	$P > \|t\|$
Intercept	0.8042	0.117	6.886	0.000
	(0.575, 1.033)			
One-Month Lagged VIX	0.0441	0.002	19.121	0.000
	(0.040, 0.049)			
One-Month Lagged Volatility	-0.1064	0.022	-4.773	0.000
	(-0.150, -0.063)			
Volume	0.0279	0.008	3.378	0.001
	(0.012, 0.044)			
GDP Growth	-0.1924	0.015	-12.732	0.000
	(-0.222, -0.163)			
Interest Rates	-0.0386	0.015	-2.591	0.010
	(-0.068, -0.009)			
US Dollar Index	-0.0052	0.001	-5.235	0.000
	(-0.007, -0.003)			
Corporate Bond Yields	0.2321	0.028	8.243	0.000
	(0.177, 0.287)			
Ten Year Treasury Yield	-0.1956	0.021	-9.398	0.000
	(-0.236, -0.155)			
R-squared:		0.621		
Adj. R-squared:		0.619		
F-statistic:		353.7		
Prob (F-statistic):		0.000		
AIC:		99.28		
Validation MSE:		0.05600381257736679		
Test MSE:		0.05826594447454576		
Baseline MSE:		0.15743207315220623		

Table 3: Multiple Linear Regression summary and prediction metrics. Values in parentheses are the 95% confidence intervals.

References

[1] Lim, C. M., & Sek, S. K. (2013). Comparing the performances of GARCH-type models in capturing the stock market volatility in Malaysia. Procedia Economics and Finance, 5, 478–487. https://doi.org/10.1016/s2212-5671(13)00056-7

[2] Ma, F., Lu, X., Liu, J., & Huang, D. (2022). Macroeconomic attention and stock market return predictability. Journal of International Financial Markets, Institutions and Money, 79, 101603. https://doi.org/10.1016/j.intfin.2022.101603

[3] Kim, J. (2023). Stock market reaction to US interest rate hike: Evidence from an emerging market. Heliyon, 9(5). https://doi.org/10.1016/j.heliyon.2023.e15758

[4] Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the U.S. Oil Market. Journal of Banking & Finance, 47, 1–14. https://doi.org/10.1016/j.jbankfin.2014.05.026

[5] Johnson, T. C. (2008). Volume, liquidity, and liquidity risk. Journal of Financial Economics, 87(2), 388–417. https://doi.org/10.1016/j.jfineco.2007.03.006

[6] Arnott, A. C. (2024, April 23). How rising interest rates change the relationship between stocks and Bonds. Morningstar, Inc. https://www.morningstar.com/portfolios/what-rising-interest-rates-mean-stockbond-correlations