Time Series Analysis of Optimized Natural Gas Portfolios

Lewis Boslego

March 14, 2023

Executive Summary

In this case study, I set out to evaluate the U.S. natural gas industry from the perspective of a potential investor, and attempted to answer common questions that this investor may have. I began this study by collecting a grouping of five natural gas stocks that in order to represent this market. I then used this representative basket to address the first portion of questions that my theoretical investor was concerned with, which were about the past trends and behavior of this market. I found that this industry has high variability and responsiveness, and is especially sensitive to large economic events. I also found that this market has high covariance, as the stock returns follow similar structures.

The second portion of these questions were concerned with appropriate investment strategies. In order to address these concerns, I derived three optimized portfolios composed of the natural gas stocks, all of which having different attributes and strategies. These portfolios included an equal weighted portfolio, Global Minimum Variance portfolio, and maximum Sharpe ratio portfolio. The equal weight portfolio would be suitable for an investor interested in a simple and unbiased portfolio that would be easy to interpret and adjust. The Global Minimum Variance portfolio would be suitable for an investor whose primary concern is limiting their exposure to risk. Lastly, the maximum Sharpe ratio portfolio would be ideal for an investor interested in maximizing their risk adjusted return.

In order to further analyze these portfolios and answer the question of understanding their future returns, I undertook the Box-Jenkins modeling procedure and fit ARMA models to the time series of each portfolio. As a result of this, my findings were that these three portfolios follow the time series models, ARMA(1,1), MA(3), and MA(3), respectively. I found that these models were adequate in predicting the general trend of future returns, which was the goal, to determine generally the future trend of the returns, but their accuracy of prediction is not suited to predict exact returns for a given future date.

Introduction

Natural gas is a natural energy source that is considered to be one of the most efficient fossil fuels. This is because it has a high energy content per unit and burns cleanly relative to other fossil fuels, such as coal or oil. Natural gas is the largest source of electricity in the U.S., and was responsible for 40% of the electricity generated in 2022, according to the Energy Information Administration. It is also widely used for heating and cooking, and in chemical and industrial processes. Because of its relative cleanness and efficiency, natural gas is not under the same environmentalist pressure to be phased out and replaced by renewable energy sources as coal and crude oil are, and its demand is projected to grow in the coming decades, according to McKinsey's 2021 industry projection, "Global Gas Outlook to 2050."

There have been several developments over the past decades that have stimulated the United States natural gas industry. These developments include more efficient mining methods that have lowered the cost to produce natural gas, the lifting of government regulation that restricted exportation of natural gas, and the development of liquefied natural gas infrastructure, a technology used to transport natural gas more efficiently and without the use of pipelines. These combined factors have opened up natural gas U.S. companies to the global market by making the exportation of natural gas commercially viable and lucrative.

U.S. companies with the right infrastructure have since been able to serve the increasing worldwide demand for natural gas.

For these reasons and more, one might be interested in investing in the United States natural gas industry. An investor interested in the U.S. natural gas industry would likely be concerned with the following core questions regarding its returns:

What have the returns and volatility of this industry been?

How is this industry impacted by macroeconomic trends?

How do companies correlate with each other?

What could be an appropriate long-term investment strategy for this industry?

What can be expected from this investment strategy? What structure do the returns have and how do they change over time?

In this case study, I will address these questions by analyzing five U.S. natural gas companies whose stocks will be used to form a market basket to create a simple representation of this industry. First, I will study the stock returns of these companies individually, to consider their expected values, volatility, trends, and make other inferences. Then, I will aggregate this information to contrast these individual companies. To address the interest of investing in this industry, I will form potential portfolios comprised of these companies, state when they are useful, study and model their time series, and finally use the time series models to form forecasts, in order to understand what can be expected from these portfolios.

Literature review

Stock returns and portfolio composition are popular subjects for case studies, especially in the field of time series. This is largely because of the clear application of price prediction for financial gain, among other motivations. As a result of its usefulness, there are a number of helpful works that have been published with goals similar to my own.

One useful reference for the time series process is section 3.1 of the textbook, “An Introduction to Analysis of Financial Data with R,” by Ruey S. Tsay. In this section, Tsay presents a case study on gasoline prices in the United States. Tsay analyzes trends in gasoline price by modeling its growth rate using different iterations of ARMA models and time series regression. Tsay then uses these time series models to forecast gasoline prices. He concludes that his time series regression model that takes lagged values for crude oil price as an independent variable is the best model based on prediction accuracy. This work is relevant to my own study because Tsay’s methodology aligns with the Box-Jenkins framework I am following for time series analysis, and provides useful tutorials of the entire process.

The work, “Portfolio Optimization Using ARIMA-GMV Approach,” authored by Kamram Raiysat, studies stock price prediction and portfolio optimization using the Pakistan Stock Exchange. Raiysat uses ARIMA modeling to predict portfolio performance, and contrasts the portfolio optimization techniques of global minimum variance (GMV) and an ARIMA - GMV hybrid. Raiysat finds that the ARIMA GMV hybrid portfolio they construct performs better than the pure GMV portfolio, but is highly dependent on the accuracy of the predictions generated by the ARIMA model. This paper is a useful reference for the time series forecasting process, and in constructing global minimum variance portfolios, which I utilize later on.

The main reference that I will build from when deriving optimized portfolios is, “Portfolio Theory with Matrix Algebra,” by Eric Zivot, an economics professor at the University of Washington. In this work, Zivot, derives multiple portfolios using three risky assets, including a Global Minimum Variance portfolio, a maximized Sharpe ratio portfolio, and other efficient portfolios. He does this through the combination of the Lagrange multiplier, which is used to find maximums and minimums of a function that is subject to a constraint, and matrix algebra.

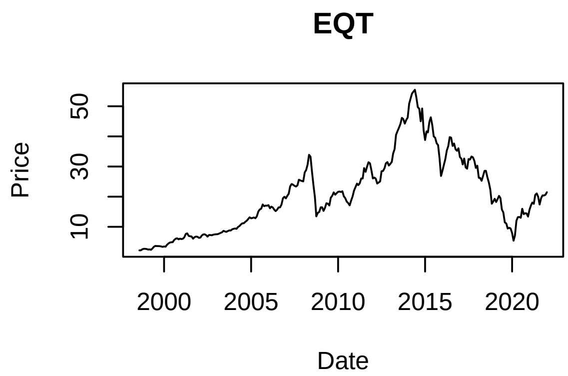
Data selection

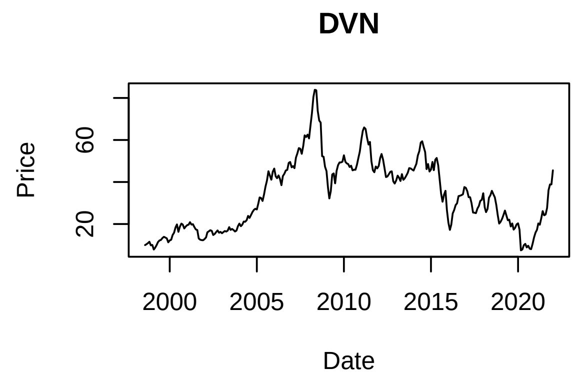
My data set is based on the daily adjusted closing price for the five natural gas companies, EQT Corporation (EQT), Devon Energy Corporation (DVN), Cheniere Energy, Inc (LNG), Enterprise Products Partners L.P. (EPD), and Sempra Energy (SRE). I sourced the individual price data for each of these companies from finance.yahoo.com, particularly the daily adjusted closing price, and converted it into monthly data by taking its arithmetic average for every calendar month. I then created a return time series for each stock, using logarithmic return, which is defined at time t as: $R_t = ln(\frac{P_t}{Pt-1})$. I chose this method for calculating returns because its decimal value can easily be interpreted as being equal to a percentage change in stock price. This formula also has the benefit of differencing the data, because it applies a logarithmic transformation that stabilizes the mean and variance. This collective data set spans from August, 1998, to February, 2023, and includes 294 observations. I have divided this data set into two, and the first 282 observations is now the training set and the last 12 observations is the testing set. All subsequent analysis will use the training set, and the testing set will be used to measure forecasting accuracy. I selected this relatively long time window because I wanted the results of this study to represent how this market responds to as many economic conditions as possible and to show a history of trends for this industry. These factors are especially important when planning long-term investments, which will also undoubtedly be subject to a variety of economic conditions. I chose the monthly sampling window for a similar reason, because a smaller time resolution is not as valuable in analyzing long-term investments and can lead to unwanted noise that is not indicative of a greater trend.

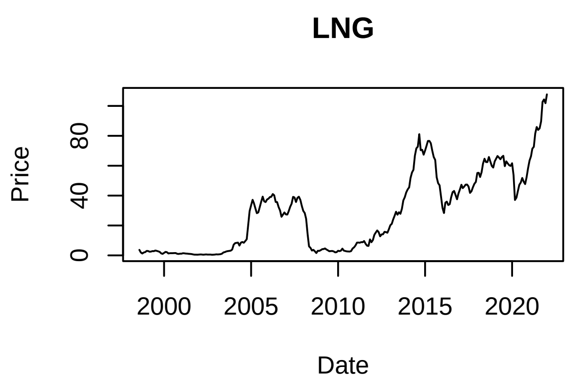
I chose to include each of these stocks based on the criteria that they have a notable dependency on natural gas, operate mainly in the United States, have been publicly traded for some time, and together represent many of the main components of the natural gas supply chain. Both EQT and DVN produce and store natural gas. LNG processes natural gas into liquefied natural gas, and transports and exports this natural gas product. EPD operates pipelines which transport natural gas domestically, and also stores and processes it. SRE processes natural gas into liquefied natural gas, operates natural gas pipelines, and supplies natural gas utilities. Most of these companies are diversified, and operate in different energy sectors as well, and therefore not perfect proxies for the U.S. natural gas industry.

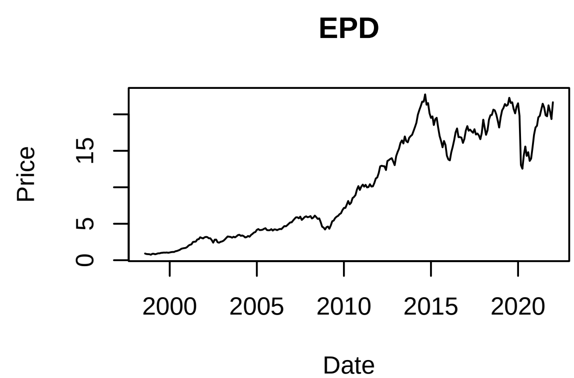
Data Analysis

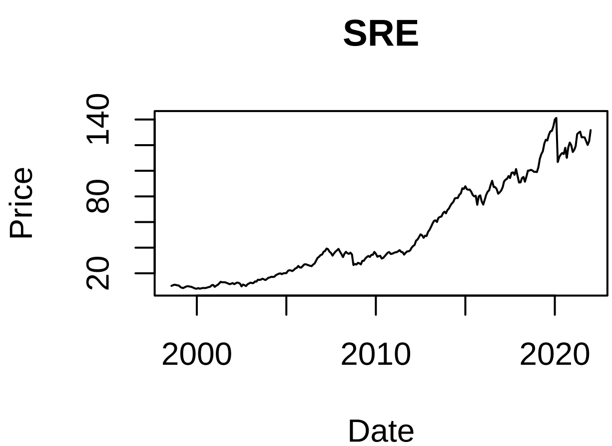
Let's now take a look at the data. The following plots are of the monthly average price for each of the selected stocks. These stock prices are not our main concern in this study, but are informative about the nature of the data and show how the companies relate to each other. These prices appear to make up two different shapes, where EQT and DVN appear to have peaked years ago, and their historic price forms a type of crude triangle. The other three stock prices, for LNG, EPD, and SRE, appear to have a more linear structure, with a somewhat positive slope. It is suggested by these plots that these companies generally have a volatile relationship with time. This is not much of a surprise, because these companies were selected due to their dependence on natural gas, which is inherently volatile due to its fluid demand.



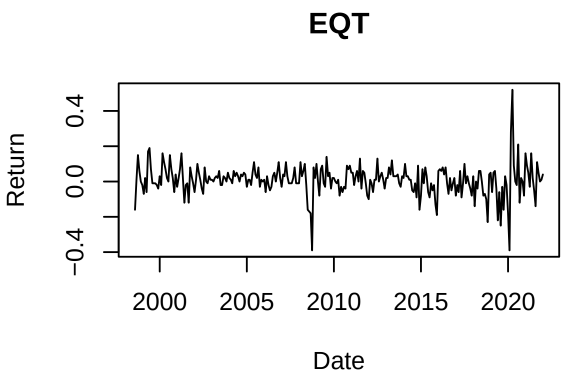


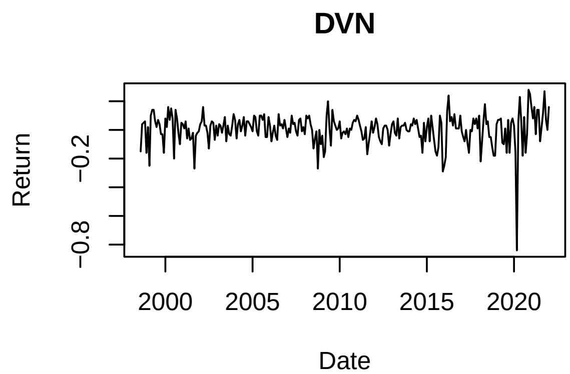


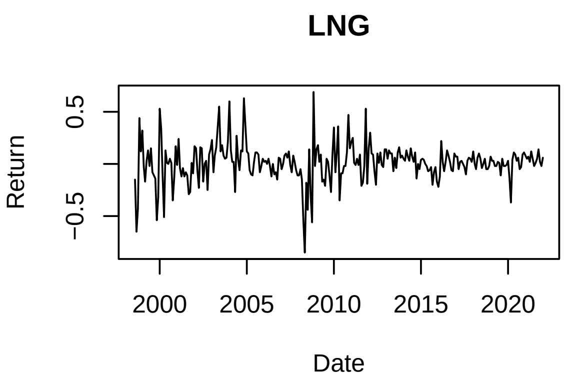


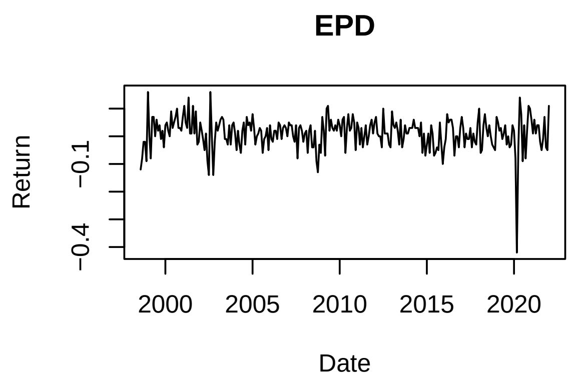


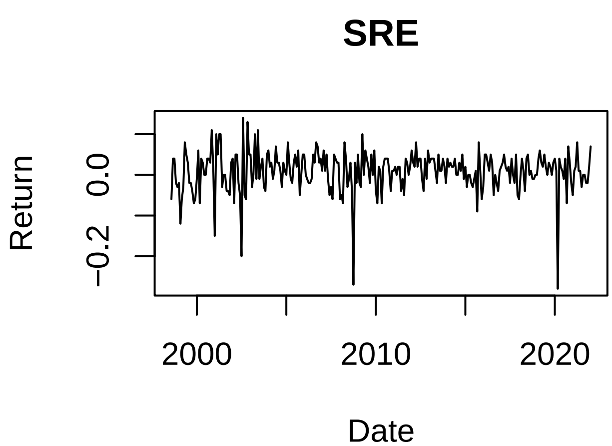
Because returns are our primary concern, let's now reflect on their structure. The returns for these stocks appear to be mostly constant, with stable means centered around 0. Their volatility appears to be mostly constant as well, but have large, often negative, spikes. These outliers correlate with greater economic events, particularly the Great Recession in 2008, and the Covid-19 pandemic in 2020, and tells us that this industry is sensitive to such factors.











Let's now uncover the summary statistics for these stock returns. The average values for each of these stocks are in the following table, "Table 1: Average Stock Returns". They are all positive, meaning that on average, these stocks produce positive returns. They range from .0049, for DVN, to .0117 for LNG. These averages are only part of the concern, and in order to fully understand the behavior of these returns, we need to see how the individual returns vary, and how they covary with each other. This information is provided in 'Table 2: Covariance of Stock Returns,' which is a covariance matrix for these stocks. As well as having the highest mean, LNG has the greatest variance. DVN and LNG have the greatest covariance, meaning that their returns move together most frequently out of this group.

Table 1: Average Stock Returns

Stock	mean return
EQT	0.0071
DVN	0.0049
LNG	0.0117
EPD	0.0110
SRE	0.0087

Table 2: Covariance of Stock Returns

	EQT	DVN	LNG	EPD	SRE
EQT	0.0073	0.0032	0.0026	0.0011	0.0009
DVN	0.0032	0.0113	0.0061	0.0037	0.0020
LNG	0.0026	0.0061	0.0335	0.0036	0.0022
EPD	0.0011	0.0037	0.0036	0.0032	0.0014
SRE	0.0009	0.0020	0.0022	0.0014	0.0025

Building Portfolios

In this case study, I am interested in investigating different investment strategies an investor may undertake, and will limit this analysis to portfolio weighting techniques. The portfolios these techniques create will be based on the key principle of Modern Portfolio Theory (MPT), which states that one can maximize their returns and minimize their risk by developing diversified portfolios. This simplified market basket I am analyzing is not a perfect application of MPT because the companies have relatively high covariance with each other. With this in consideration, I will carry out this analysis by creating three potential portfolios, including an equally weighted portfolio, a Global Minimum Variance Portfolio, and a maximum Sharpe ratio portfolio, all of which have varying levels of complexity and offer different approaches to this investment. In the construction of these portfolios, I will make simplifying assumptions in order to make them easier to model and interpret.

The equally weighted (EW) portfolio holds all assets in the same proportion, and is often seen as a naive approach because it does not consider any individual characteristics of its assets. This is, however, one of its strengths because this portfolio makes no assumptions on factors that could easily change and backfire on the investor. This portfolio also benefits from its simplicity and ease of use, which all together make this method a situationally viable strategy. Its construction follows the simple formula:

$$P_{EW} = \frac{1}{n}\sum_{i=1}^{n} A_i,$$

where P_{EW} is the equally weighted portfolio, and A_i is Asset i, $i = 1, .., n$.

The Global Minimum Variance (GMV) portfolio holds assets in proportions such that they minimize the overall volatility of the portfolio. The conservative nature of this strategy is its primary benefit, but the minimization also usually results in high risk-adjusted return, making these portfolios competitive with those that focus purely on risk-adjusted return. The formula for this portfolio is the following:

$$min_{\underline{w}}(\sigma^2_{p,\underline{w}}), s.t. \sum_{i=1}^{n} w_i = 1,$$

where $\underline{w}$ is the vector of asset weights, 1 to n, and $\sigma^2_{p,\underline{w}}$ is the portfolio variance with respect to the weights. The constraint, that weights must sum to 1, allows for short-selling of stocks, but not buying on margin. This constraint is made for the sake of simplicity in modeling, allowing the process to focus on long-term performance, without implementing another dimension of riskiness to handle. Back to the problem, this minimization is performed using the Lagrange multiplier, which is a constraint based optimization technique. In this context, it has the following form:

$$L(w, \lambda) = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_i w_j \sigma_{ij} - \lambda(\sum_{i=1}^{n} w_i - 1)$$

where L is the Lagrange operator, λ is the Lagrange multiplier, σ_i^2 is the variance of asset i, σ_{ij} is the covariance of assets i and j, and -1 is the constraint imposed. The minimizing weights can be found by taking the first order condition (FOC) for each asset and λ,

$$\frac{\partial L}{\partial w_i} = 0, i = 1, .., n, \ \frac{\partial L}{\partial \lambda} = 0$$

then, solving for the w, λ vector in:

$$\begin{bmatrix} 2\sigma_i^2 & 2\sigma_{ij} & \dots & 2\sigma_{in} & 1 \\ 2\sigma_{ij} & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ 2\sigma_{in} & \dots & \dots & 2\sigma_n^2 & 1 \\ 1 & \dots & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_i \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The last portfolio we will consider is the maximized Sharpe ratio portfolio. The Sharpe ratio measures risk adjusted return and equals:

$$\text{Sharpe} = \frac{R_p - R_f}{\sigma_p},$$

where R_p is portfolio returns, R_f is the returns of a risk free asset, and σ_p is the standard deviation of the portfolio returns. This ratio considers firstly the risk premium of the portfolio, found in the numerator of this ratio, which is the expected returns that the risky investment provides in excess of the returns provided by the risk free asset. If the risk premium is negative, the investor should invest fully in the risk free asset, as it has higher return than the risk adjusted return of the asset. For the purpose of this study, the risk free return will be set equal to 0. My goals is not to give real-world financial advice, but to experiment with our closed market, the natural gas basket. Now, going forward,

$$\text{Sharpe} = \frac{R_p}{\sigma_p}.$$

In order to maximize the risk-adjusted return of our portfolio, by maximizing the Sharpe ratio, we follow a similar process as in minimizing the GMV portfolio. We will again consider a Lagrange constraint based optimization technique. The process is as follows:

$$max_m \frac{\mu_{p,m}}{\sigma_{p,m}}, \frac{\mu_{p,m}}{\sigma_{p,m}} = \frac{m^T \mu}{(m^T \sum m)^{\frac{1}{2}}}, s.t \ \sum_{i=1}^{n} m_i = 1.$$

$\mu_{p,m}$ is the average return of the portfolio with weights, m, σ is its standard deviation, and $\sum$ is the covariance matrix. Here, we assume the same constraint that the weights must sum to 1, but can be negative. Here, the Lagrangian is:

$$L(m, \lambda) = (m^T \mu)(m^T \sum m)^{-1/2} + \lambda m^T 1 - 1, s.t \ \ m^T 1 = 1,$$

and the FOCs are:

$$\frac{\partial L}{\partial m} = 0, \frac{\partial L}{\partial \lambda} = 0.$$

m, the vector of weights, then equals:

$$m = \frac{\sum^{-1} \mu}{1^T \sum^{-1} \mu}.$$

Now that each of these portfolios has been properly defined, they can be derived using the stock returns sample statistics. Of course, the nature of the equal weight portfolio is that it does not consider these values, and has weights $\frac{1}{n} = \frac{1}{5}$, and is expressed as:

$$P_{EW} = \frac{1}{5}R_{EQT} + \frac{1}{5}R_{DVN} + \frac{1}{5}R_{LNG} + \frac{1}{5}R_{EPD} + \frac{1}{5}R_{SRE},$$

where R is the return for a given stock. The GMV portfolio has weights,

$$w = (0.1703922, -0.1111337, -0.0194195, 0.4235468, 0.5366141),$$

for EQT, DVN, LNG, EPD, and SRE, respectively, and is expressed as:

$$P_{GMV} = (0.1704) * R_{EQT} - (0.1111) * R_{DVN} - (0.0194) * R_{LNG} + (0.4235) * R_{EPD} + (0.5366) * R_{SRE}.$$

Lastly, the weights for the maximum Sharpe ratio portfolio are,

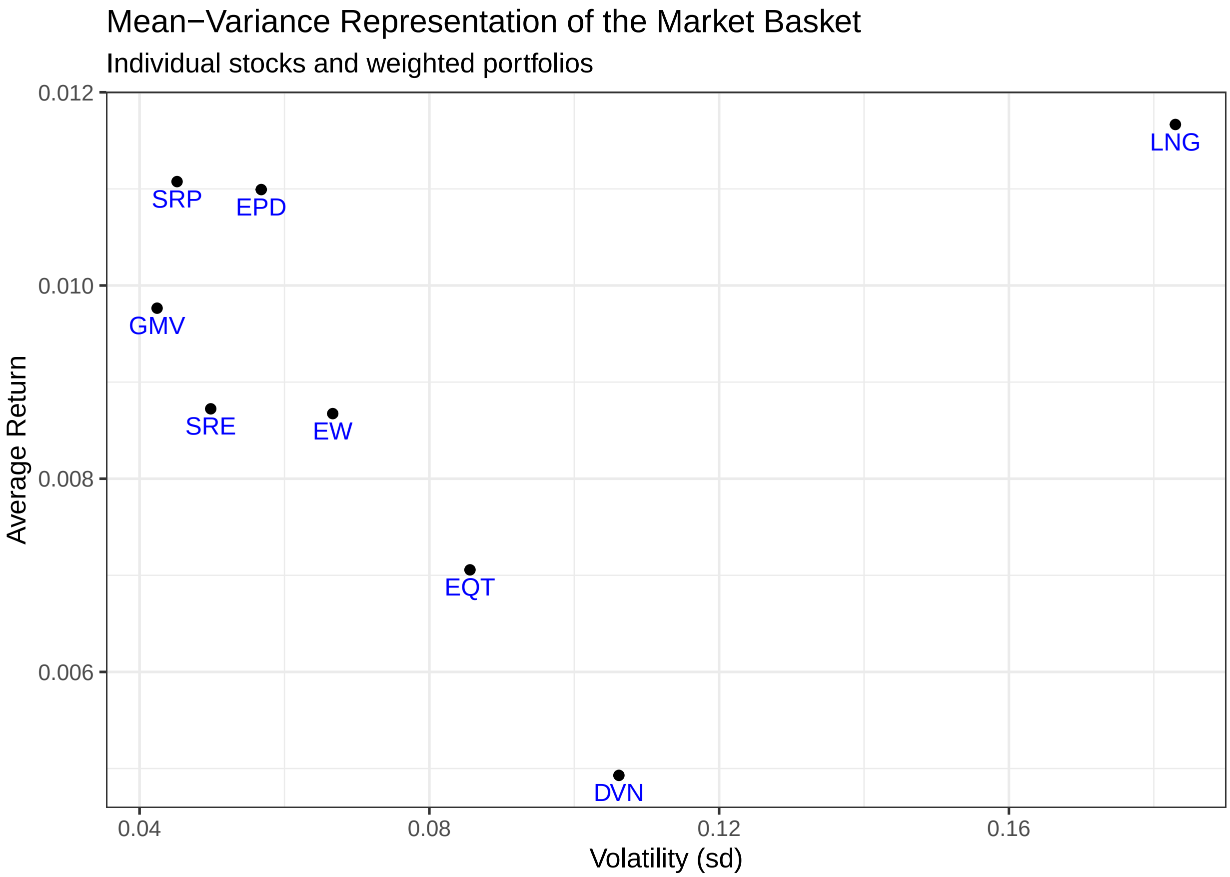
$$m = (0.134725966, -0.260390861, -0.001451409, 0.701833301, 0.425283002),$$

for EQT, DVN, LNG, EPD, and SRE, respectively, and this portfolio is expressed as:

$$P_{SRP} = (0.1307) * R_{EQT} - (0.2604) * R_{DVN} - (0.0015) * R_{LNG} + (0.7018) * R_{EPD} + (0.4253) * R_{SRE}.$$

Reflecting on the weights of the GMV portfolio and maximum Sharpe portfolio, these two seem to have somewhat similar compositions. The Sharpe portfolio has a larger short position in DVN, and a much larger long position in EPD, whereas the GMV portfolio has a larger position in both EQT and SRE. Neither of these portfolios are interested in a position LNG, which is interesting because it has by far the greatest average return, but makes logical sense because it also has the largest volatility by far. Because the portfolios have a close resemblance to each other, the risk of the assets seem to be dominant with respect to the optimal risk-adjusted portfolio.

Now that these three portfolios have been defined, we can add them to our consideration as parts of our market basket. We can compare this new basket with the use of Modern Portfolio Theory's mean-variance plane, which is plotted below. This graph, "Mean-Variance Representation of the Market Basket," plots the returns of each stock or portfolio against it's standard deviation, and helps to visual the risk-adjusted return of each. This also allows us to verify that the weighted portfolios were properly calculated. The GMV portfolio has the least volatility, and the Sharpe portfolio has the highest return, considering risk. The equally weighted portfolio (EW), appears to be in the center mass of the stocks, as expected. Like previously stated, the GMV and Sharpe portfolios have a close relationship, although the Sharpe portfolio appears to have substantially more return with a marginal increase in risk. The individual assets EPD and GMV are close by these portfolios, which is no surprise considering how much the portfolios rely on them. This visual really defines LNG's high volatility, as it's standard deviation is about double that of the next most volatile stock. It also explains why it has effectively no utilization by the risk concerned portfolios, as it has marginally better average return compared to EPD, but has way more risk. The low risk-reward of DVN is interesting, because the optimized portfolios are able to utilize it with a short position, which perhaps serves as a hedge against market volatility. The logic here is DVN is likely to have a large negative response to economic downturn, which would bring the portfolio positive returns, and would limit the portfolio's losses from the other stocks with long positions.



Now that we developed possible investment strategies in the form of the three portfolios, and have verified their respective properties, we can analyze them further and form models to forecast their future performance.

Experimental Design

In the following section, I will apply the Box-Jenkins methodology in order to model and make predictions for the returns for the three portfolios. The training data set will be used for all steps, and the testing data set will be used to measure prediction accuracy.

My workflow will follow the following steps:

1. Review and analyze plots of portfolio price and return, identify unusual values, make comments on stationarity of returns.
2. Verify weak stationarity of returns with the implementation of the Dickey-Fuller Hypothesis test.
3. Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the portfolio returns.
4. Use the results of the autocorrelation plots in order to identify candidate ARMA models.
5. Use R to generate and refine models.
6. Rank candidate models based on AIC.
7. Verify that model residuals are white noise with the Ljung-Box hypothesis test.

8. Generate a 12 month forecast and compare performance to the testing data set. Use this information to determine and select the best model.

Step 1 includes a visual inspection of original portfolio price data and the transformed data, the portfolio return. This step is focused on looking for major trends or sudden events in the original price data and considering their potential causes. When looking at the transformed data, the returns, the goal is to assess whether the data is weakly stationary. If this conclusion seems reasonable, it brings with it the assumptions that the data has a constant mean and variance, and that the autocovariance of the time series depends only on the time lag between the observations, not on the given time at which they are observed. If this conclusion is carried on, it is important to note here whether there are any points in the time series that deviate from its assumptions, because they could impact further results and conclusions.

Step 2 checks this assumption of stationary more formally than the previous visual analysis, with the use of the augmented Dickey-Fuller Hypothesis test. This test begins with the null hypothesis that there is a unit root, and assumes that the time series follows an AR(p) model, defined as:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t.$$

It then differences this model, by subtracting y_t, resulting in:

$$\Delta y_t = \mu + \delta y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \epsilon_t.$$

It then tests the hypotheses:
Ho: $\delta = 0$.
Ha: $\delta < 0$.
The test statistic is:

$$t_{\beta_i} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)},$$

which is then compared against the t-distribution, for each coefficient, i, and is subject to the rejection criterion of: $t_{\beta_i} < t_{\alpha=0.05}$. If rejected at the set t-distribution critical value, this would suggest that there is no unit root, and that the data are stationary.

In step 3, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are plotted in order to identify patterns present in a time series, which are characteristic of different autoregressive (AR) and moving average (MA) structures. The ACF measures the correlation between a stationary time series and its values at different lags. The PACF is the partial correlation between a time series and its values at different lags, which is then regressed against all shorter lags. This negates the effects of shorter lags, and removes the noise they would otherwise create when comparing a certain value of the time series with a given lagged value. These functions assume that the time series is stationary.
The ACF follows the formula:

$$ACF = \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{E[(Y_t - \mu)(Y_{t+k} - \mu)]}{E[(Y_t - \mu)^2]},$$

which can be estimated with:

$$A\hat{C}F = \rho(\hat{k}) = \frac{c(k)}{c(0)} = \frac{\hat{\gamma_k}}{\hat{\gamma_0}} = \frac{\sum_{t=k+1}^{T} (y_{k-1} - \hat{\mu})(y_t - \hat{\mu})}{\sum_{t=1}^{T} (y_t - \hat{\mu})^2}.$$

The PACF has the formula:

$$\alpha_Y(k) = \phi_{k,k} = \frac{\rho(k) - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(k-j)}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(j)},$$

and can be estimated with:

$$\hat{\phi}_k = R_k^{-1}\hat{\rho}_k, \text{ where } R(i,j) = Corr[Y_{t-i}, Y_{t-j}].$$

In step 4, our goal is to identify reasonable models for our time series, based on the correlation structure and ACF and PACF plots. These models include the autoregressive model of order p, AR(p) the moving average model of order q, MA(q), and the combined autoregressive moving average model of orders p and q, ARMA(p,q). There is no need to consider more complex models, such as ARIMA(p,d,q) or SARIMA(p,d,q)X(P,D,Q,s), because the portfolio returns are already differenced and do not exhibit seasonality.

An AR(p) model is appropriately applied to a time series if it is correlated with its own past values, because this model uses autocorrelation to explain the current time series value by using past time series values. the order p refers to the number of previous values of the time series that contribute to its current value, and are included in the model. This model has the general form:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t,$$

for the time series y_t, some constant μ, parameters ϕ_i, past values of the time series for i,..,p, and error term ϵ_i. This model makes the assumptions that the time series is linear and stationary, the error term is uncorrelated with itself at any lag, the error term is normally distributed, and that p is fixed over time. If any of these assumptions are violated, the model will likely be inaccurate and biased. This model has the unique property that it is always invertible. For an AR process to be stationary, it must have no unit roots, and determining this is the motivation for step 2.

An MA(q) model is appropriate when the time series has a moving average trend, where the series moves randomly around its mean. This model uses this structure to determine the causal relationship between past errors and the current value of a time series. The parameter q is the number of past errors that are included in the model. It has the following general form:

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

where y_t is the time series, μ is the constant mean, ϵ_t is current the error, ϵ_{t-j} are the past errors, and θ_j, are the model parameters, for $j = 1,..,n$. This model assumes that the time series is stationary, that the residuals have a constant zero mean, the residuals have constant variance, the residuals are independent, the residuals are normally distributed, and that q is constant. If any of these assumptions are violated, the model will likely be inaccurate and biased. This model has the unique characteristic that it is always stationary, and this is why in the augmented Dickey-Fuller test, we only test the AR model for stationarity. For an MA model to be invertible, it must be true that $|\theta_i| < 1$, for $i = 1,..,n$.

An ARMA(p,q) model is appropriate when the time series exhibits both the moving average and autoregressive patterns, that is, the time series depends on both its past values, and its past lags. It has the general form:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

with the same parameter definitions as in the AR(p) and MA(q) models, and combines their components.

In the ACF plots, AR(p) models can be identified if there is exponential decay or damped sinusoidal patterns in the autocorrelation structure. MA(q) models can be identified by spikes in autocorrelation from lags 1 to q, because for an MA(q) model, all autocorrelations beyond lag q are 0. ARMA(p,q) models exhibit exponential decay or a damped sinusoidal structure for autocorrelations at lags greater than q. In PACF plots, AR(p) models have significant autocorrelation spikes at lags 1 to p, MA(q) models have exponential decay or damped sinusoidal structure, and ARMA(p,q) models exhibit exponential decay or damped sinusoidal structure for lags greater than p. ACF plots are better suited to identify MA series and PACF are better for AR series. ACF and PACF plots are also useful in determining the sign of the coefficients, as the lag coefficient will have the same sign, positive or negative, as the autocorrelation at that lag.

In step 5, the candidate ARMA models from step 4 are generated using the `Arima()` function of the `forecast` package in R. This function uses maximum likelihood estimation (MLE) to estimate the parameters of an ARMA model by using the log-likelihood of the observed data. This function also returns many statistics that measure the goodness of fit of the given model, including the standard errors for the estimated parameters, the variance of errors of the model, the log-likelihood, information criteria statistics, and measures of training set errors.

In step 6, the candidate models are compared based on their AIC. This is done based on the statistics generated in step 5, particularly the Akaike Information Criteria (AIC). AIC is a model selection criterion that balances the goodness of fit of a given model with its complexity, and is based on the statistical principle of parsimony, which states that a model should be as simple as possible, while still being able to explain the data accurately. A lower AIC value correlates to greater parsimony, meaning this criterion is meant to be minimized. The AIC in the ARIMA setting defined as:

$$AIC = -2lnL[\phi_p, \theta_q, \frac{S(\phi_p, \theta_q)}{T}] + 2M,$$

where L is the log-likelihood, given the model coefficients ϕ_p, θ_q, the squared residuals of the model coefficients, $S(.)$, the sample size, T, and the penalty term, equal to $M = p + q + 1$.

Step 7 is dedicated to verifying that the selected ARMA model explains all of the autocorrelation by analyzing its residuals. The residuals are the difference between the predicted values of the time series and its observed values. If the ARMA model properly explains all time series structure, the residuals will follow a white noise distribution. This means that the residuals will have a constant zero mean, have constant variance, and have no autocorrelation. If these conditions do not hold, there is still autocorrelation structure and the model is insufficient. This will be verified with the Ljung-Box-Pierce statistic. This statistic is used to determine whether a time series is autocorrelated, and is applied through a hypothesis test. This test has the following form:

Ho: The data are independently distributed.

Ha: The data are not independently distributed, and instead exhibit serial correlation.

This hypothesis test is determined with the test statistic:

$$Q_K^* = T(T+2)\sum_{k=1}^{K}\frac{\hat{\rho}_\epsilon(k)}{T-l} \sim \chi^2_{K-p-q},$$

where T is the sample size, $\hat{\rho}$ is the sample autocorrelation of the error term at lag k, and χ^2_{K-p-q} is the Chi-squared distribution with $K - p - q$ degrees of freedom. This test has the rejection criterion:

$$Q_K^* > \chi^2_{K-p-q}.$$

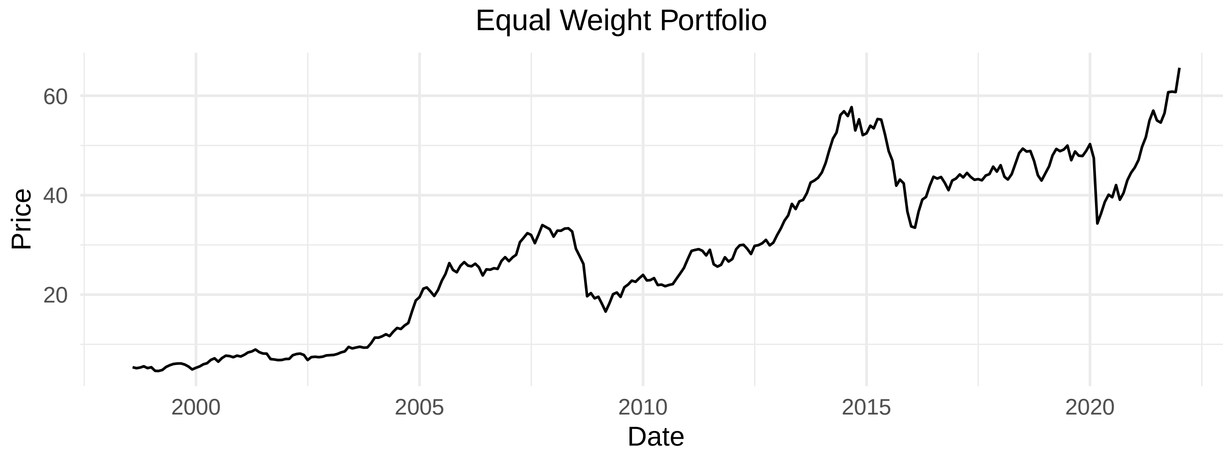
If this test is failed, the ACF and PACF plots for the residuals can be used to identify the remaining structure left in the residuals. It is important to note here, that the residuals

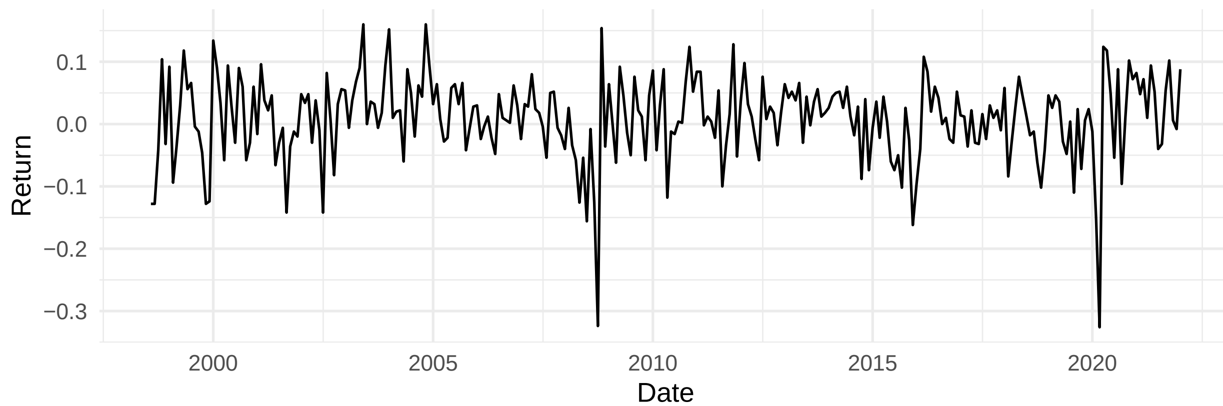
Finally, in step 8, the vetted ARMA models are used to generate forecasts. This is performed with the `forecast()` function of the forecast package in r, which takes in an ARIMA model and the number of periods to forecast, then returns a best estimate for each period, as well as 80% and 95% confidence intervals for the time series at each forecasting period. The accuracy of this prediction is then compared to the testing data set, that was partitioned in the beginning of the project, and includes the most recent 12 observations of data. Lastly, the ARMA model that performed the best in forecasting, with the lowest root-mean-squared error (RMSE), and mean absolute error (MAE) is selected.

Time Series Modeling and Forecasting

Equally Weighted Portfolio

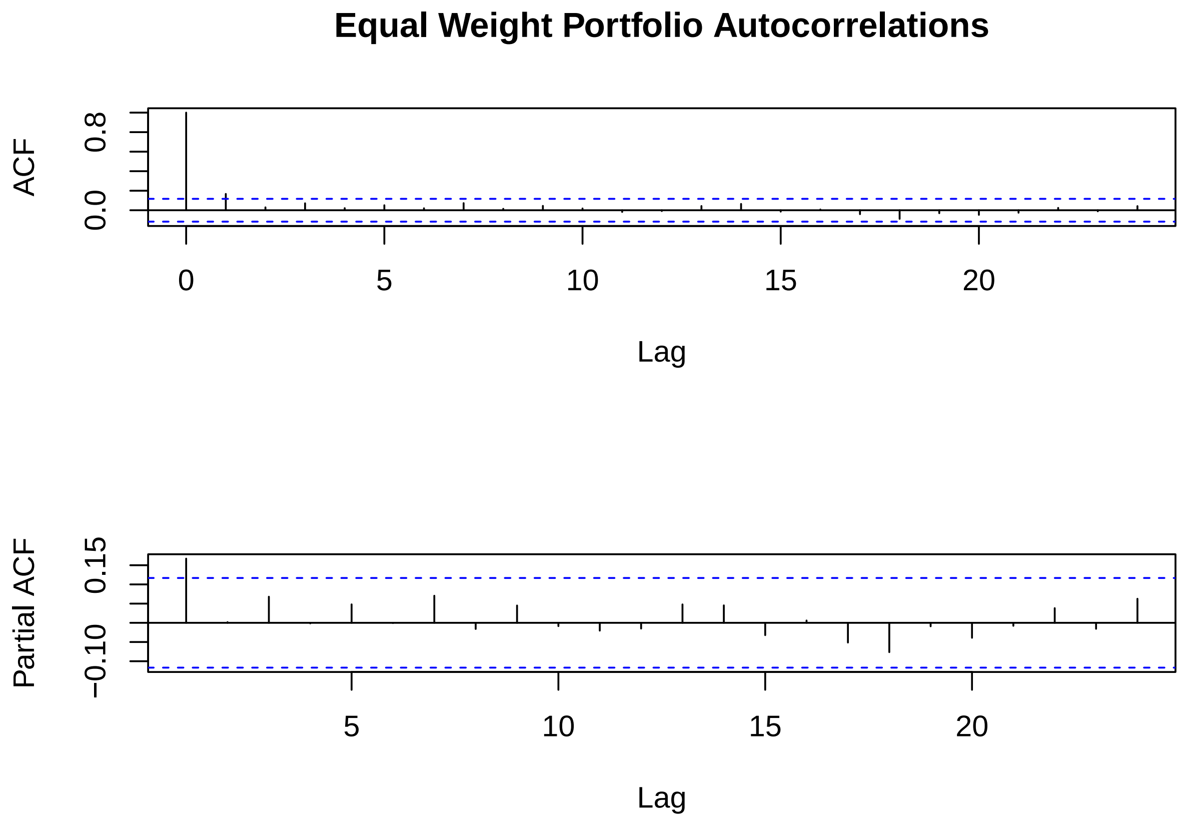
We will begin this framework with the portfolio of equally weighted stocks. Referring to the plot below, there is a positive trend in the price of this portfolio over time, and returns are centered around 0. There are two notable outliers in returns, which correlate with the market crashes associated with the Great Recession of 2008 and the Covid-19 pandemic, which negatively impacted every stock under consideration. The time series of the returns appears to be weakly stationary, as it has a constant mean over time, and constant variance, despite the outliers which may stretch this assumption.





The augmented Dickey-Fuller test of this portfolio's returns yields a test statistic of -5.3 and a p-value less than 0.01. This suggests that the null hypothesis of non-stationary is implausible, and that the alternative hypothesis of the time series being stationary is more likely.

Because this time series appears to be stationary, we can continue to the ACF and PACF plots, which are seen below. There is one significant autocorrelation spike at lag 1, after which the ACF and PACF exhibit constant insignificant structure. This could be indicative of an MA(1) model, as the ACF has a sharp drop off after lag 1, but the autocorrelations do not seem to slowly dissipate, like this model would suggest. Maybe this could be an AR(1) model, because the PACF drops off after the first lag, and afterwards the correlations weaken as time lag increases. These plots are not perfect examples of either of these models, and perhaps a hybrid ARMA(1,1) model is more appropriate.



Now that candidate models have been formed, it is time to fit these models to the training data. Fitting the MA(1) model using the `Arima()` function returns:

$$x_t = 0.009 + 0.17 * z_{t-1} + z_t,$$
$$\sigma^2 = 0.00433,$$

where z_t is white noise.the standard error for the intercept is 0.005 and the standard error for the coefficient is 0.059. This model has an AIC of -730.1. Fitting the AR(1) model returns:

$$y_t = 0.009 + 0.17 * y_{t-1} + e_t,$$
$$\sigma^2 = 0.00433,$$

where the standard error for the intercept is 0.005 and the standard error for the coefficient is 0.059. This model has an AIC of -730.2. The next model to fit is the ARMA(1,1):

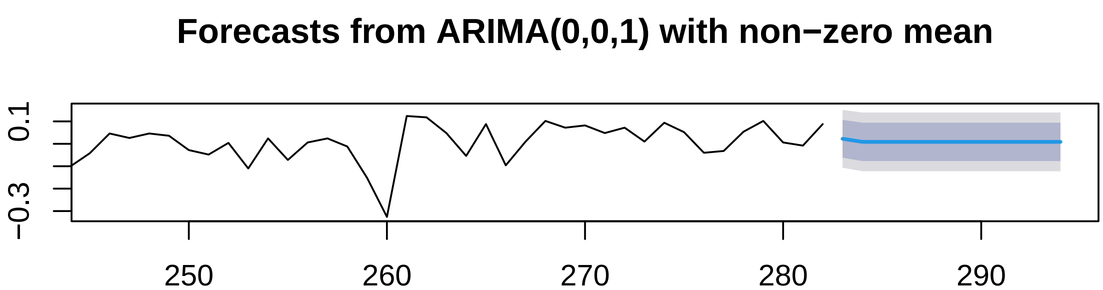
$$y_t = 0.009 + 0.289 * y_{t-1} - 0.123 * z_{t-1} + z_t,$$
$$\sigma^2 = 0.00435,$$

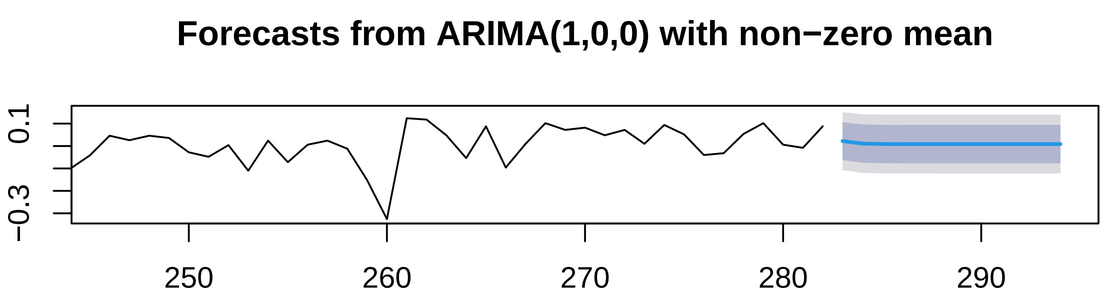
where the standard errors are 0.005, 0.767, and 0.803, for the intercept, autoregressive coefficient, and moving average coefficient, respectively. This model has an AIC of -728.2.

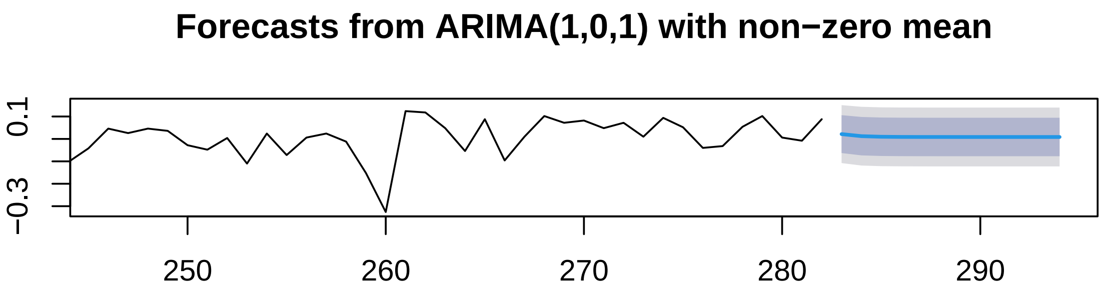
These three potential models, the MA(1), AR(1), and ARMA(1, 1) all have are very similar to each other, in that their estimated parameters are nearly identical, as well as their AIC statistics. This suggest that all of these models provide a similar quality of fit to the training data. The ARMA(1, 1) model may be overfit, because the standard errors for its estimated coefficients outweigh them considerably.

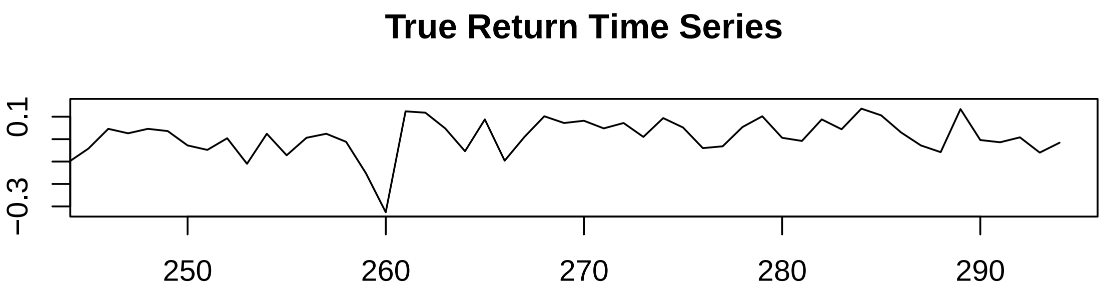
To continue our analysis, we will now consider the residuals for each of these models. Applying the Ljung-Box test to the MA(1) model provides a test statistic of $Q* = 4.4$, which corresponds to a p-value of 0.9. This p-value is statistically insignificant at the $\alpha = .05$ level, and we fail to reject the null hypothesis, that the residuals are uncorrelated. This conclusion is confirmed by the following plots, which show no remaining autocorrelation structure in the residuals. Now, to apply the same analysis to the AR(1) model. This model produces a Ljung-Box statistic of $Q^* = 4$, which corresponds to a p-value of 0.9. This p-value is again statistically insignificant at the $\alpha = .05$ level, and we fail to reject the null hypothesis, that the residuals are uncorrelated. The following plots confirm this, which also show no remaining autocorrelation structure. Lastly, lets look at the residuals for the ARMA(1,1) model. Performing the Ljung-Box test results in a test statistic of $Q^* = 3.8$, and p-value of 0.9 yet again. This is also statistically insignificant at the $\alpha = .05$ level. This conclusion is confirmed by the ACF plot, which shows no remaining significant structure.

Now that we have confirmed that the candidate models explain all of the structure of the data, we can use them to make predictions and compare the results to the testing data. For each of these models, we will generate predictions for the returns of this equally weighted portfolio over the next 12 months, using the `forecast()` command of the `forecast` R package.









As predicted from their other similarities, these models produce near identical forecasts, for the next 12 months. Now, we can compare these forecasts to the true return series for this 12 month period, in order to determine which model was most accurate. Visually, these forecasts appear to have captured the trend of the time series within their 80% confidence intervals, and are fairly, accurate. For a more formal comparison, lets look at each models testing RMSE and MAE, as described in "Table 3: Model Testing Accuracy." All of these models had very similar forecasting accuracy, but the ARMA(1,1) model had slightly less testing error, with the lowest RMSE of 0.06647 and MAE of 0.05178. For this reason, The ARMA(1,1) model is selected as the best model for the equally weighted portfolio time series.

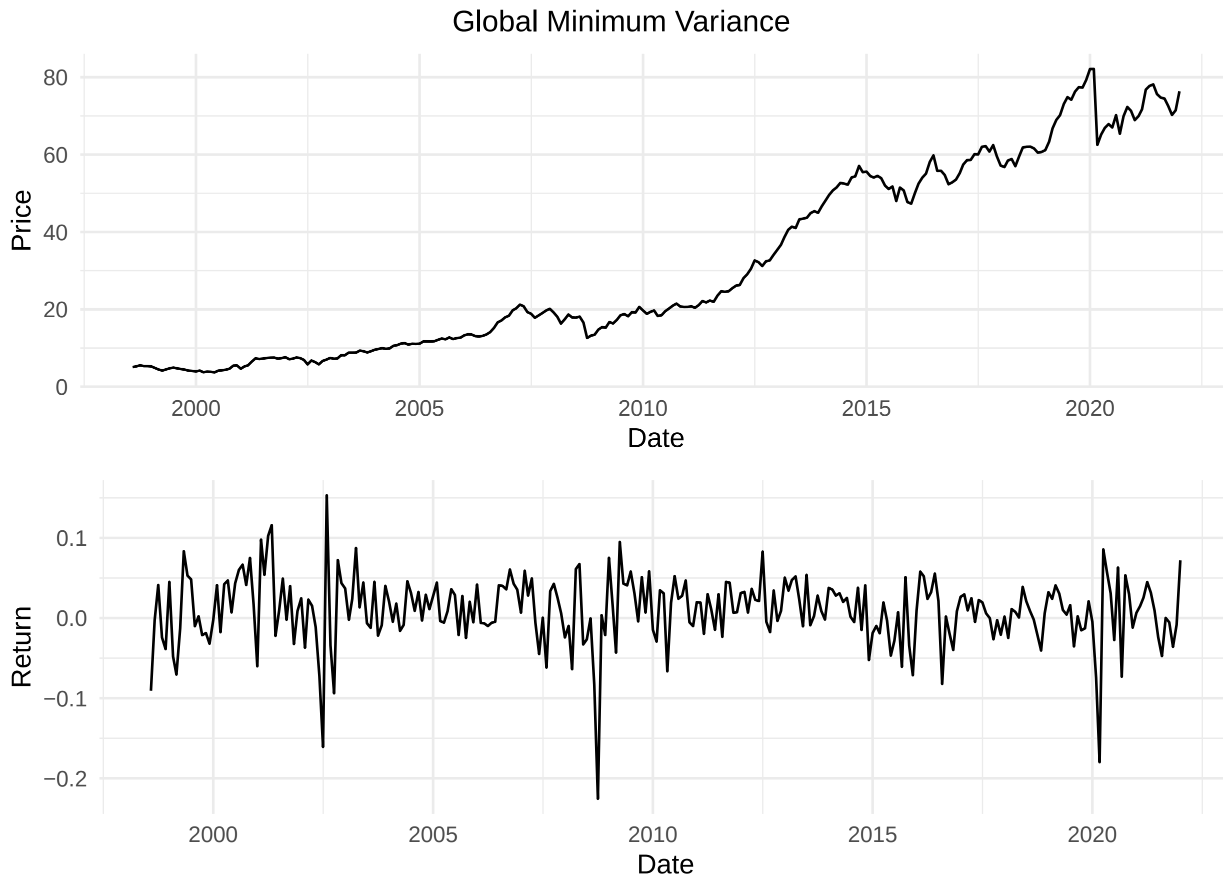
Table 3: Model Testing Accuracy

Model	RMSE	MAE
MA(1)	0.06715	0.05212
AR(1)	0.06675	0.05192
ARMA(1,1)	0.06647	0.05178

Now that we have selected an appropriate time series model for the returns of the equally weighted portfolio, we can move on the other two portfolios and model their returns.

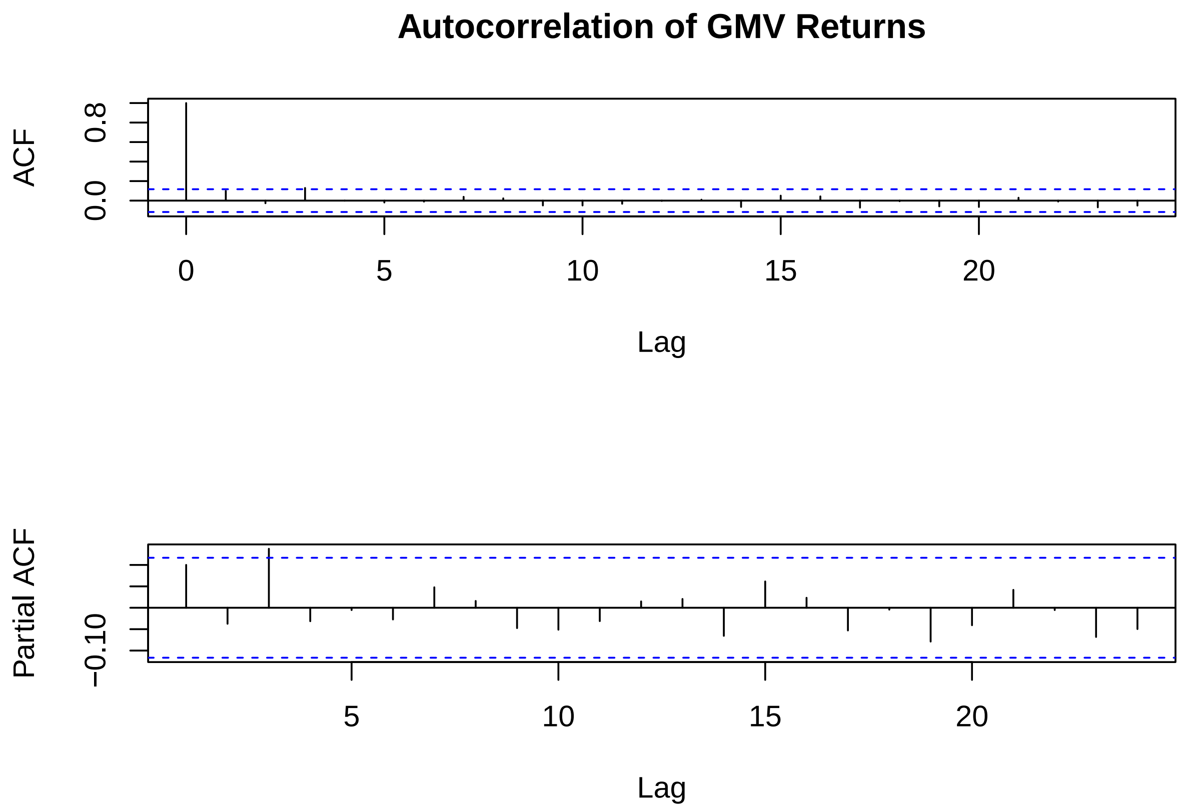
Global Minimum Variance Portfolio

The price and returns for the GMV portfolio are plotted below. In the price series, the trend appears more constant than for the equal weight portfolio, and the volatility of the returns appears to be more constant and smaller. These time series also appear to be correlated with market trends, with large downward spikes in returns around 2008 and 2020. The returns time series appears to be weakly stationary, with a constant mean centered around zero, and somewhat constant, finite variance. These assumptions are likely not perfectly accurate, as there are some outliers in the variance. This is important to note because later steps in modeling will assume that the process is stationary, and if this is not accurate, the modeling results will be negatively impacted.



Because of the importance of stationarity of the data, this will be tested again with the augmented Dickey-Fuller test. This test produces a test statistic equal to -6.1, which is statistically significant with a p-value < 0.01. This this is significant evidence against the null hypothesis, non-stationary, which is rejected. We can now more confidently assume stationarity.

We can now analyze the autocorrelation plots for this time series, in order to select candidate models. The autocorrelation structure in the ACF plot appears to be insignificant for the first 2 lags, then significant for the 3rd, and then the autocorrelation quickly dissipates. Based on the previously defined characteristics for the different time series models, this structure most closely resembles the description of an MA(q) model, where the autocorrelation goes to 0 after q. This definition suggests that this process is an MA(3) model, as the autocorrelation goes to 0 after that significant lag 3. It could also be the case that this is an MA(1) model, and the autocorrelation spike at lag 3 is non-deterministic. This could possibly also be best represented by an MA(0), because the autocorrelation at lag 1 is not significant, and at lag 2 it is less so, which could indicate the spike at lag 3 is non-deterministic, again.



Now to fit these 3 potential models. The MA(3) model has the form:

$$x_t = 0.01 + z_t + (0.117) * z_{t-1} - (0.039) * z_{t-2} + (0.138) * z_{t-3},$$
$$\sigma^2 = 0.00176,$$

where x_t is the time series, and z_{t-i} is the error for lags 0,..,3 of the time series. The standard errors are 0.03, 0.059, 0.061, and 0.06, for the intercept and coefficients respectively. This model has a log likelihood of 496.1 and an AIC of -982.3. The MA(1) model has the fit:

$$x_t = 0.1 + z_t + (0.116) * z_{t-1},$$
$$\sigma^2 = 0.00179,$$

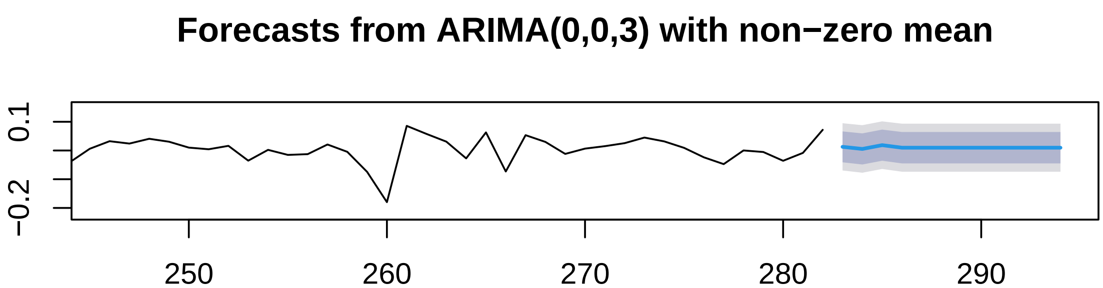
and this model has the same parameter definitions as the MA(3). This model has standard errors 0.003 and 0.065 for the mean and coefficient, and has a log-likelihood of 493.1 and AIC of -980.2. The next model, the MA(0) model has the fit:

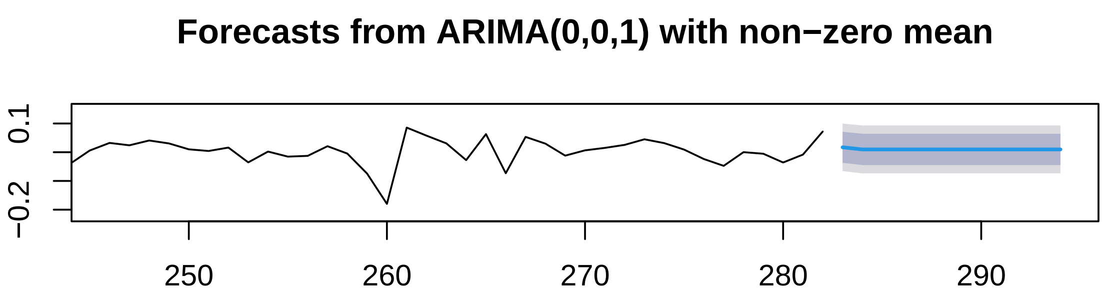
$$x_t = 0.1 + z_t,$$
$$\sigma^2 = 0.0019,$$

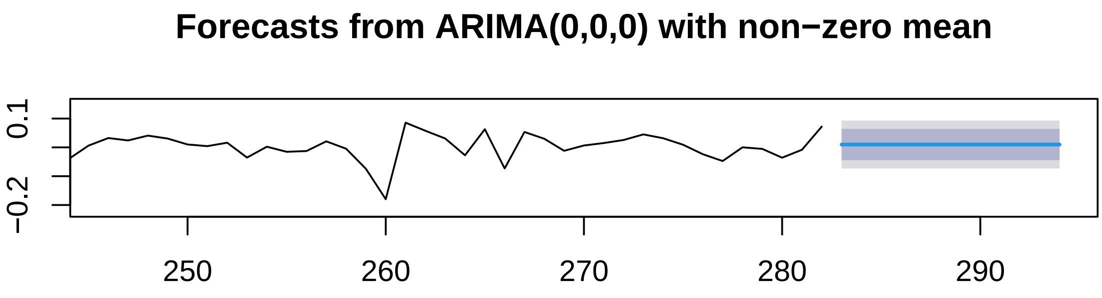
and the intercept has a standard error of 0.003, a log likelihood of 491.5, and an AIC of -979. These three models all have similar log likelihoods and AIC values, which suggests that they should also have comparable forecast performance.

Before we can evaluate their forecasting accuracy, it is important to reflect on their residuals, with the Ljung-Box test, to ensure there is no serial correlation. Applying this hypothesis test to the MA(3), MA(1), and MA(0) models, the respective test statistics are $Q^* = 2.3$, $Q^* = 7.7$, and $Q^* = 10$, which correspond to p-values of 0.9, 0.6, and 0.4. Because these p-values are statistically insignificant, we fail to reject the null hypotheses that the residuals for each of these models are independently distributed.

Now that the models have been vetted, we can use them to form forecasts and compare their results. The forecasts generated by each of the three models are plotted below, along with the true returns for the GMV portfolio from the testing data. Visually, it appears that the MA(3) model does the best in forecasting, because it predicts the first up and down pattern seen in the testing data. Comparing the RMSE and MAE for each of these models, which can be found in "Table 4: Model Testing Accuracy," this initial impression is confirmed, as the MA(3) model has the lowest MSRE and MAE, equaling 0.06301 and 0.04854. For these reasons, we can conclude this procedure by selecting the MA(3) model to represent the returns of the GMV portfolio.







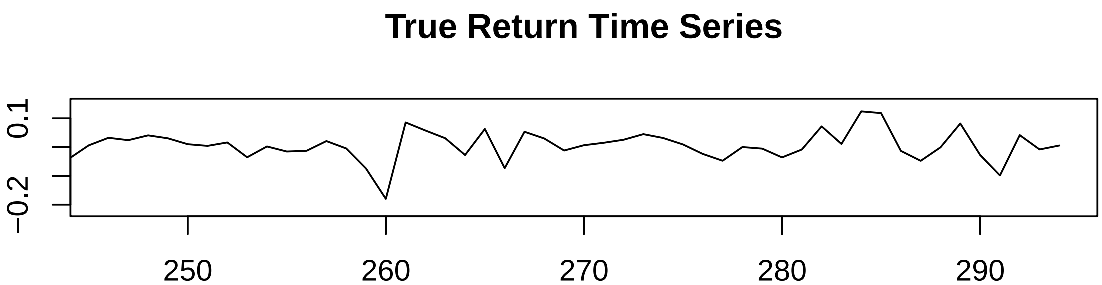
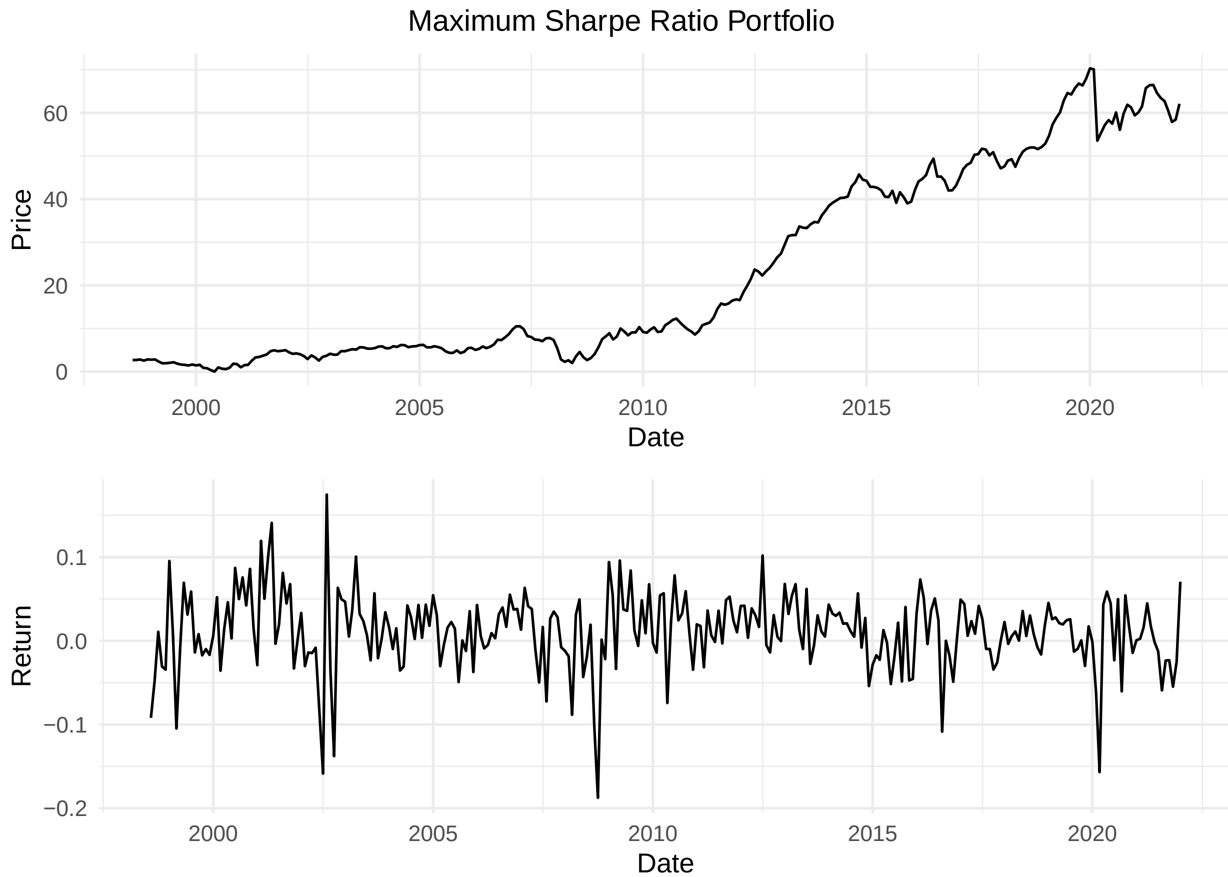


Table 4: Model Testing Accuracy

Model	RMSE	MAE
MA(3)	0.06301	0.04854
MA(1)	0.06357	0.04929
MA(0)	0.06354	0.04889

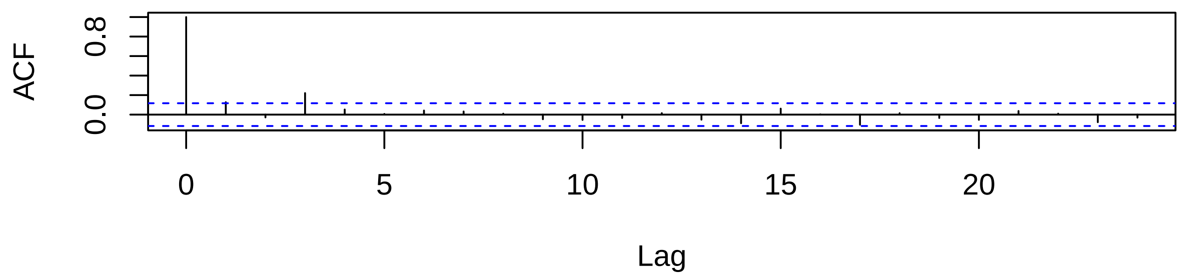
Maximum Sharpe Ratio Portfolio

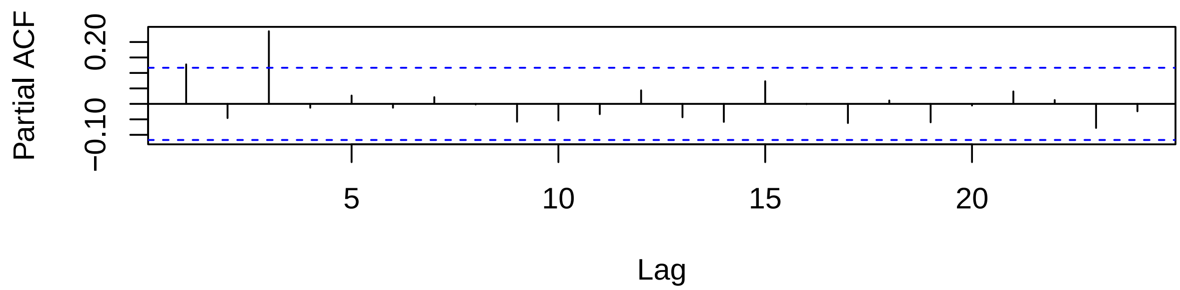
We can now follow model selection and estimation procedure one last time for the returns time series of the maximum Sharpe ratio portfolio. In the plots below, this time series closely resembles that of the GMV portfolio, which is no surprise considering their relationship that was uncovered earlier. This return time series exhibits the same key characteristics as the GMV portfolio series, in that it has a constant mean centered around 0, a mostly constant variance, despite a few outliers, and appears to be approximately weakly stationary. The deviations from this definition in this series, the outliers, may affect model accuracy and forecasting performance.



Let's again perform the augmented Dickey-Fuller test on the return series. This test results in a statistic equal to -5.4, which corresponds to a p-value less than 0.01, which is statistically significant at the threshold value, $\alpha = 0.05$, and we reject the null hypothesis that the series has a unit root and is non-stationary.

Plotting the ACF and PACF, The autocorrelation structure is quite similar to that found for the GMV portfolio, with a few significant autocorrelations then drop off, with a notable difference that the values for lag 1 and lag 3 are greater, and lag 1 is statistically significant. This autocorrelation structure suggests the MA(3) model. The insignificant autocorrelation at lag 2 may indicate a MA(1) model, however. Because this autocorrelation structure is not definitively moving average, I will also fit an AR(3) model in order to determine if it is a better fit of the data.





Now, fitting these possible models, the MA(3) model has the form:

$$x_t = 0.011 + z_t + (0.14) * z_{t-1} - (0.054)z_{t-2} + (0.205) * z_{t-3},$$
$$\sigma^2 = 0.00192,$$

where x_t is the time series, and z_{t-i} for $i = 0, .., 3$ are the time series errors. This model has standard errors 0.003, 0.059, 0.057, and 0.056 for the constant and the respective lags. This model has a log likelihood equal to 483.9, and an AIC of -957.9. The MA(1) model has the form:

$$x_t = 0.011 + z_t + (0.159) * z_{t-1},$$
$$\sigma^2 = 0.00201,$$

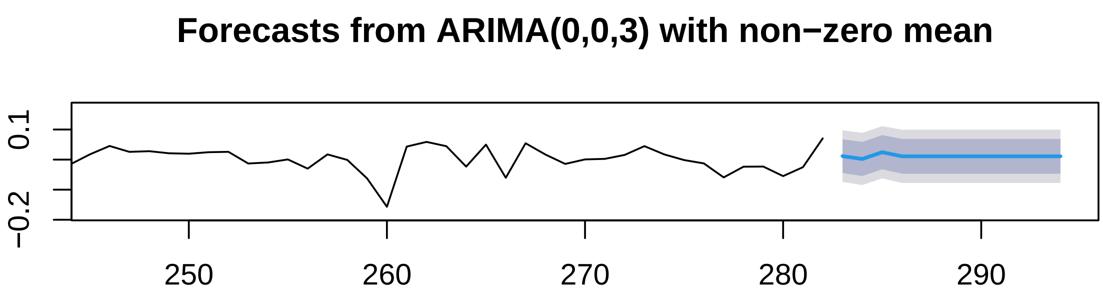
and has the standard errors 0.003 and 0.069 for the intercept and coefficient, and has a log likelihood of 476.5, and AIC of -947.1. The AR(3) model has the form:

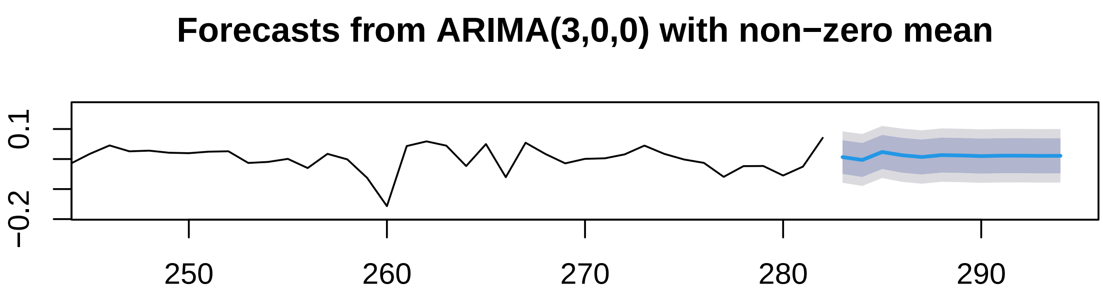
$$y_t = 0.011 + z_t + (0.145) * y_{t-1} - (0.077) * y_{t-2} + (0.24) * y_{t-3},$$
$$\sigma^2 = 0.00191,$$

with standard errors 0.004, 0.058, 0.059, and 0.059 for the intercept and coefficients respectively. This model has a log likelihood of 484.6, and AIC of -959.2. Reviewing the AIC statistics for each of these models, each appears to fit the data similarly well.

Continuing our analysis, it is now time to perform the Ljung-Box test on each model's residuals. This test produces statistics $Q^* = 3.2, 17, 2.3$, which correspond to the p-values 0.9, 0.04, and 0.9. Because the p-value of the MA(1) model is statistically significant at the $\alpha = 0.05$ level, we reject the null hypothesis that the residuals are independently distributed, in favor of the alternative hypothesis that there is still serial correlation present in the time series. Because of this, this model is inadequate and will be removed from further consideration. The p-values of the MA(3) and AR(3) models are statistically insignificant at $\alpha = 0.05$, and we fail to reject the null hypotheses that their residuals are independent.

Now, to decide between the two remaining models, the MA(3) and AR(3), we will generate forecasts using each and compare them to the testing data. The 12 month forecasts for these models are plotted below, along with the true time series for the Sharpe portfolio returns. Visually, it appears that the AR(3) captures more of the structure of the time series. The RMSE and MAE, shown in 'Table 5: Model Testing Accuracy,' for both of these forecasts result in the opposite conclusion, as the MA(3) model score lower in both of these error measures, with a RMSE of 0.0583, and MAE of 0.04677. Because the MA(3) model has lower forecasting error, we will select this model as the best fit of the maximum Sharpe ratio portfolio.





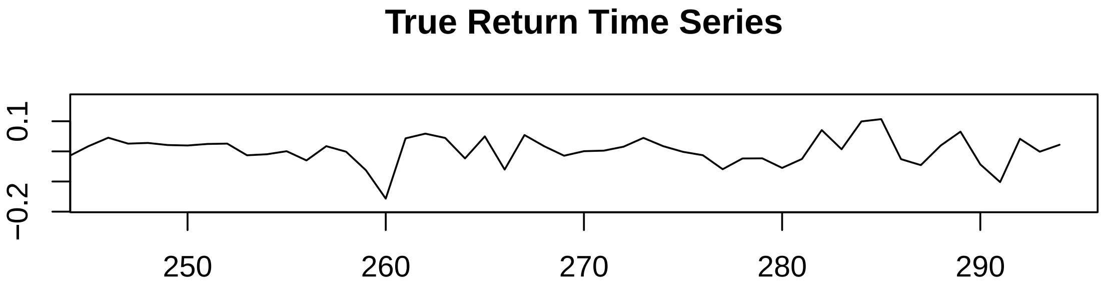


Table 5: Model Testing Accuracy

Model	RMSE	MAE
MA(3)	0.0583	0.04677
AR(3)	0.05895	0.04773

Discussion of Modeling Results

In the previous section, I followed the outlined modeling procedure and determined the most appropriate model for each of the portfolios. The equal weight portfolio was best represented by an ARMA(1,1) model, and both the Global Minimum Variance and maximum Sharpe ratio portfolios we best represented by MA(3) models. These results make sense, because the equal weight portfolio is an equal average of many time series, if any of those series has strong autocorrelation structure, it would be weakened by other other time series processes. Therefore, it makes sense that the equal weight portfolio would have a relatively simple model, with both moving average and autoregressive aspects. The GMV and Sharpe portfolios being best approximated by the same general model, the MA(3), is not much of a surprise, because these portfolios have similar stock compositions and time series. All of these models had similar accuracy in forecasting these portfolio returns, with RMSE's around 0.06. They were largely able to forecast 12 months of returns within their 80% confidence intervals. Because of this, these forecasts could be useful in forecasting returns in a broad sense, as they capture the general trends of the different time series, but should not be relied on to predict with high accuracy. This result makes sense in the context of stock returns, because of their high volatility and complexity of determinants.

Conclusion

Now that all of our analysis is compete, we can refer back to the problem statements in the introduction of this study, and address them with the answers that have been found. Restating these questions:

What have the returns and volatility of this industry been?

How is this industry impacted by macroeconomic trends?

How do companies correlate with each other?

What could be an appropriate long-term investment strategy for this industry?

What can be expected from this investment strategy? What structure do the returns have and how do they change over time?

The past returns in the industry have been volatile, but on average positive, and this industry has a high correlation with macroeconomic trends and events. It also has a high covariance, as the changes in stock returns tend to covary among different stocks. Regarding a long term strategy, I proposed three options in the forms of the equal weight, Global Minimum Variance, and maximum Sharpe ratio portfolios, which each have pros and cons to be considered. In order to better understand these portfolio strategies, I analyzed each under the Box-Jenkins framework and fit them with descriptive ARMA models. These models did not have much lag structure, as the longest lags considered were 3 periods back, meaning that they were somewhat ineffective in accurately forecasting long time periods, in that they would revert back to the time series average after a few periods. These forecasts did provide fairly accurate 80% and 95% confidence intervals, which are useful in higher level predictions, which is more palatable to our interest of long term investments where we are not reliant on the returns in the exact next period to profit. This modeling process was helpful in better understanding the portfolios, and demonstrated that their returns through time were mainly explained by past errors and random variation – at least at this level of analysis. These time series models did rely on certain assumptions that may have not been accurate to the data and that could explain some of their shortcomings.

Bibliography

Global gas outlook to 2050. (2021). Available at: https://www.mckinsey.com/~/media/mckinsey/industries/oil%20and%20gas/our%20insights/global%20gas%20outlook%20to%202050/global%20gas%20outlook%202050_final.pdf.

U.S. Energy Information Administration (2023). What is U.S. electricity generation by energy source? [online] Eia.gov. Available at: https://www.eia.gov/tools/faqs/faq.php?id=427&t=3.

Tsay, R.S.,2014. *An Introduction to Analysis of Financial Data with R*. John Wiley & Sons.

Portfolio Optimization Using ARIMA - Global Minimum Variance Approach. (n.d.). Available at: https://norma.ncirl.ie/4564/1/kamranraiysat.pdf

Zivot, E.. *Chapter 1 Portfolio Theory with Matrix Algebra.* Available at: https://faculty.washington.edu/ezivot/econ424/portfolioTheoryMatrix.pdf.