Project Report - Robert Murphy

GitHub URL

https://github.com/rob-p-murphy/14-02-23-CIDA_rmurphy

Abstract

This project analyzes the historical share price performance of four of the largest US life insurance companies relative to each other and the overall US stock market. The four companies in our sample are Reinsurance Group of America (RGA), Prudential Financial (PRU), MetLife (MET), and American International Group (RGA). Time-series analysis is used to consider the relative performance of the stock prices over time, while annual sharpe ratios are used to consider risk-adjusted returns over time. The results show that RGA outperformed its peer companies (and indeed the overall market index) for much of the time period considered on an absolute basis. The results were less conclusive on a risk-adjusted basis where there were no obvious differences in performance between the companies and the market.

After considering the performance of the stocks, the project analysed potential factors that could drive this performance. External economic factors (GDP growth, interest rates, and inflation rates) were considered, along with internal company-specific factors (Revenue growth, EBITDA growth, and shareholder equity growth). When running regression analyses for each of these factors against stock price performance, revenue growth showed the highest correlation amongst the factors considered.

Introduction

Analysis of asset prices and the factors driving them is a key element of my job. This project has allowed me to develop a basic structure to import data on asset prices, economic variables and company financial information from APIs. Going forward I can build on this project to import data on more assets over a longer time frame. The analysis can also be developed to include more complex statistical models.

Dataset

The project used three data sources for this analysis: Economic Data (interest rates, GDP, and inflation rates) from the World Bank API, Stock price data from Yahoo Finance API, and Company financials from Alpha Advantage API. The sources were selected because they are from reputable organisations that are widely used as data sources by companies in the financial services industry, which ensures data quality and accuracy. These sources are also free.

Four scripts were used to import the data into Python, create a dataframe, and export the data into different tabs of the “Data_index.xls” spreadsheet. These scripts have to be run in the following order:

1.GDP_IR_INFL_data_final

2. Stock_prices_export_final

3. US_All_Financials_Extraction_AIG_RGA

4.US_All_Financials_Extraction_PRU_MET

The first script imports economic data from the World Bank API as a pandas dataframe, and the second imports data stock prices for four large US life insurers along with two market indices from Yahoo Finance. The third and fourth scripts import the five most recent years of company financial data for AIG, RGA, PRU, and MET from Alpha

Advantage API. This data is then pasted into the Economic_data" tab, "Stockmarket_data" and "Company_data" tabs of the "data_source.xls" spreadsheet..

Implementation Process

Data importation, manipulation and cleaning

All data analysis is carried out by the "Analysis" script which takes the "data_source.xls" spreadsheet as an input.

comp_df data frame (imported from the Company_data" tab)

The company data shows annual balance sheet and income statement line items for the most recent 5 years of available data. However, in order to carry out our analysis we are interested in the annual change in three key metrics which we consider alongside movements in the stock price. These are as follows:

- "ebitda" - which is a broad measure of core earning capacity
- "totalRevenue" - which broadly captures sales volumes
- "totalShareholderEquity" - which is a broad measure of the value of shareholder's interest in the company based on accounting data.

We create three new variables showing the growth in these metrics over 2019, 2020, 2021 and 2022 ("EBITDA Growth", "Revenue Growth" and "SH Equity Growth"). It is these metrics that we will analyse alongside movements in stock prices. We also change the "fiscalDateEnding" column values to a YYYY string format which will assist with merging this data with other dataframes later in the process.

eco_df data frame (imported from the "Economic_data"" tab)

The economic data frame contains GDP values in monetary amounts for each year. These are converted into GDP growth figures ("GDP Growth"). The "Year" value is converted into a YYYY data type and also converted to a string type object to allow this dataframe to be merged with other dataframes later on in the process.

stock_df data frame (imported from the "Stockmarket_data"" tab)

As with the other dataframes we convert the "Date" column to a YYYY format. We also cut the dataframe off at the earliest date where there is data available for all 4 stocks and both indices (RGA was first listed on 2008-09-12 so our dataframe now starts here). In order to analyse the relative performance of the 4 stocks relative to our two indices, we rebase all of our stocks and indices at 2008-09-12 and consider their growth in percentage terms from that point on. To do this we create a "_rebased" column for each of the stocks and indices all of which has a value of 1 at 2008-09-12 and then increases over time based on the performance of the stock or index. This will make it easier to consider the relative performance of the stocks and indices over time.

Analysis

Stock price performance over time - (Fig 1 below)

This is a simple time series analysis which considers the relative performance of the 4 stocks along with the S&P500 over the time period under investigation. This uses the rebased returns data for each stock from stock_df

Risk reward analysis - (Fig 2 below)

In order to consider the risk/reward profile of the different stocks relative to each other and the index we need to calculate both the mean and standard deviations of daily returns over the period under investigation. In order to do this we have calculated the daily returns of each stock (which is set equal to (price(t+1)-price(t))/price(t)). We then group the daily returns by year using the pandas groupby function and calculate mean and standard deviation summary statistics. This new dataframe is called grouped_stock_df

I initially produced a scatter plot of the mean and standard deviation of annual returns for the 4 stocks and the S&P500. This graph was difficult to read or understand so I decided instead to calculate the Sharpe ratio for each stock and the S&P500 for each year of the analysis. The Sharpe ratio is a measure of risk adjusted returns. It was calculated as mean

daily price change/st. deviation of daily price change * sqrt(252) where 252 is the number of trading days in a year. This produced an annual _Sharpe_ratio value for each stock and the index. I then plotted a timeseries of this using seaboard

Economic driver analysis (Fig 3 below)
In order to identify potential external economic factors that could drive stock performance we merge the daily return values for each stock in the grouped_stock_df data with the economic variables in the eco_df dataframe using an inner join in year to create a new table called Final_df.

We then calculate a table of correlation coefficients between all variables in the Final_df table and consider their magnitude and direction in order to assess their potential relationship with stock price movements.

We also consider the correlation of the stock price growth of each individual company to the S&P500. This is due to the fact that the insurance companies themselves invest in the S&P500 and stocks are likely positively correlated (to different degrees) with the overall index. The S&P500 can therefore be considered a potential candidate of external driver of price alongside the economic variables.

Company driver analysis (Fig 4 below)
In order to identify internal company specific factors that could drive stock performance we have to merge the grouped_stock_df and the comp_df_ALL dataframe using an inner join in year to create a new table called comp_merge_df. The comp_df_ALL dataframe has been created by reshaping data in the comp_df dataframe so that is in the correct format to be joined with grouped_stock_df. The comp_merge_df has information on annual stock price growth and growth in revenue, ebidta and shareholder equity for each of the 4 companies. We estimate correlation coefficients between stock price growth and each of these factors to understand what internal metrics may drive stock prices.

Results

Fig1. Stock Price Growth Over Time - Selected Insurers

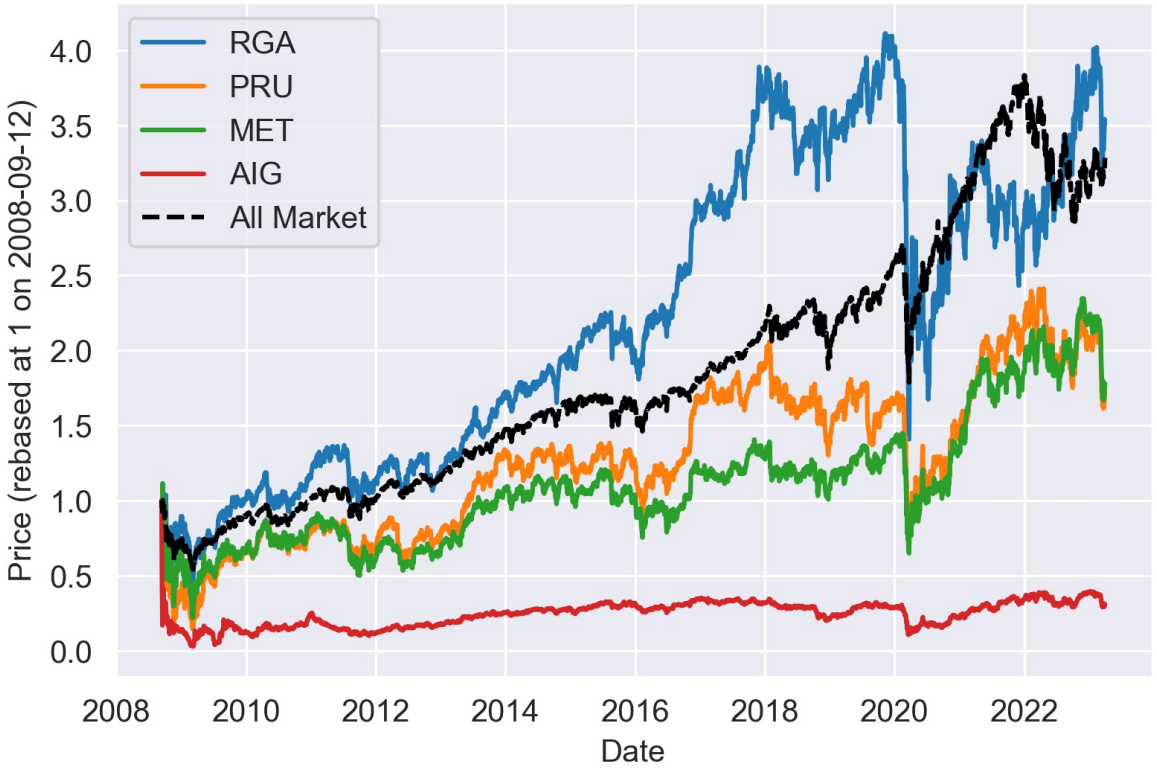


Fig2. Sharpe Ratio Over Time - Selected Insurers

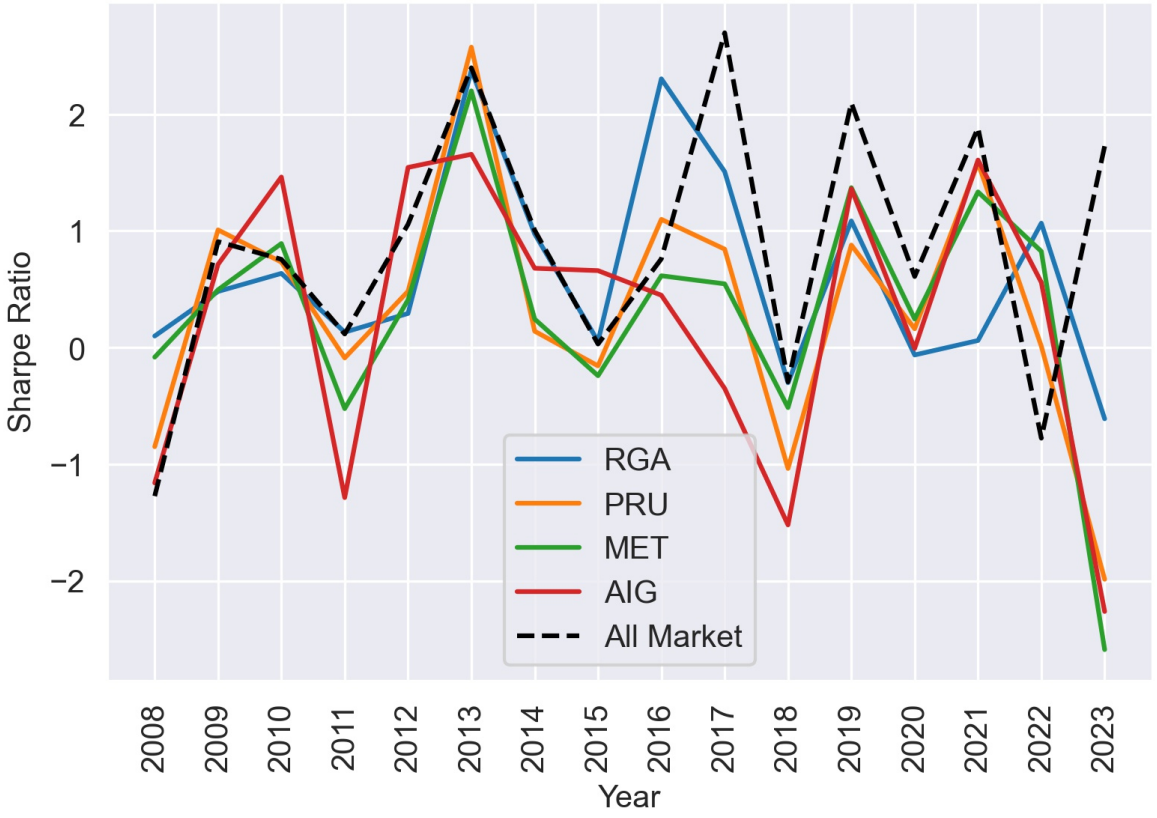
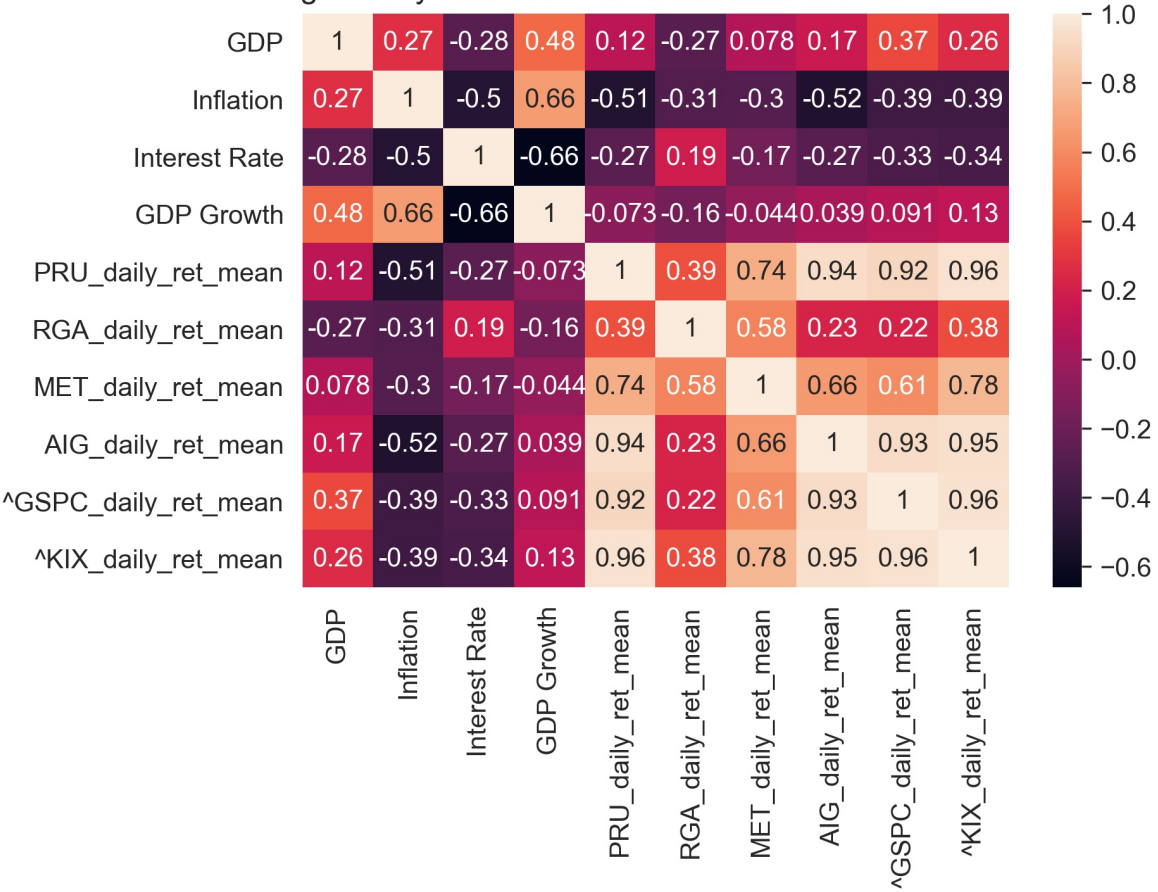
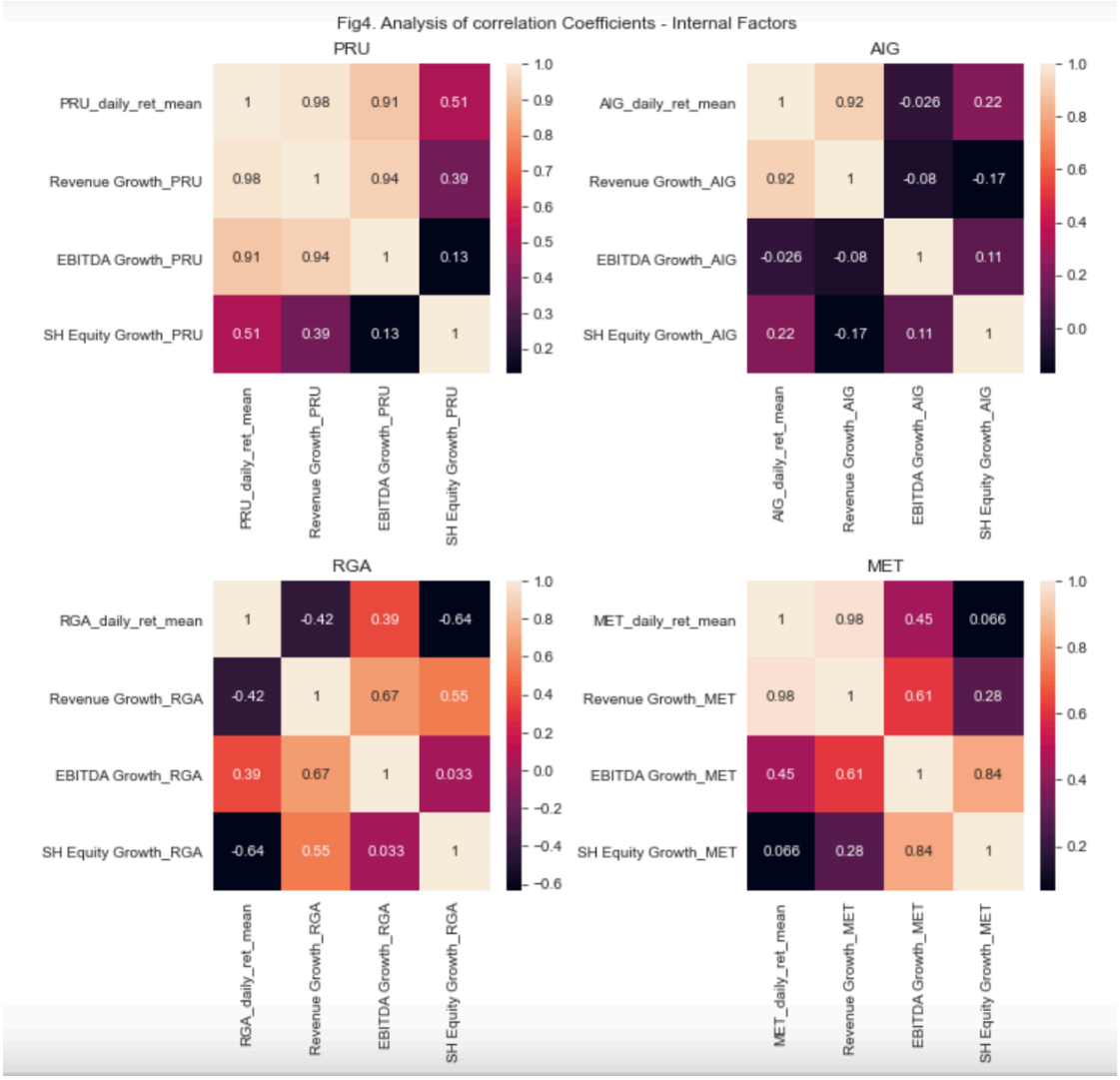


Fig3. Analysis of external factor correlation coefficients





Insights

Fig 1

1. Three of our 4 insurers significantly underperformed the overall market for the period under consideration. RGA outperformed the S&P 500 in general but did suffer a significant period of underperformance post covid. The underperformance of insurers in general is likely due to the significant impact that tech companies have had on S&P500 performance over the period. These companies generally grow profits and revenue at higher rates than insurers and therefore experience higher share price growth
2. While RGA (in blue) outperformed the other insurers and the market up until 2020, it suffered a larger fall than the other insurers and the market when covid struck (denoted by the large fall in all share prices at the beginning of 2020). It did however recover these losses over the next 3 years. This pattern of results leads us to consider that while RGA may be the best performing stock, it may also be more volatile than the other stocks. The lead me to consider Sharpe Ratios to measure risk adjusted returns

Fig 2

3. The analysis of the Sharpe Ratios over time do not highlight a clear pattern where any stock or indeed the S&P 500 delivers consistently higher risk adjusted returns. RGA's performance no longer looks like an outlier on a risk adjusted basis. We do however note that that from c.2016 onwards the the S&P500 outperforms our stocks on a risk-adjusted basis in most years.

Fig 3

4. The correlation matrix of external economic factors shows that neither inflation rates, interest rates nor GDP growth rates are particularly closely correlated to stock price growth for our 4 stocks of interest. We do note that three of our 4 stocks are highly correlated to the daily return of the overall S&P500 stock market as a whole (^GSPC). This is unsurprising given that asset prices across almost all asset classes globally

have increased since 2008 as a result of low interest rates and quantitative easing by central banks.

Fig 4

5. The matrices of internal company metrics to stock returns shows that revenue growth is highly positively correlated to daily stock price growth for three of our four stocks (PRU, MET, AIG). Only RGA again appears to be an outlier. EBITA also looks to have potential as a driver of stock market performance but shareholder equity growth doesn't really show a consistent pattern across the companies. These results are unsurprising given that insurance companies are difficult to analyse so investors often default to relying on revenue growth as a key predictor of company growth (which often translates into share price growth)

Machine Learning

Describe what kind of prediction you could perform in future using machine learning and/or deep learning

It may be possible to broaden out the company financial metrics (with an appropriate time lag) in order to identify drivers of future stock price movements. This would include getting more data and using some form of a logistic regression model to perform the prediction.

Would you use classification or regression methods?

A regression model would be more appropriate to predicting price given that stock prices are a continuous variable.

References

All data for the project was sourced from publicly available API's as described in the Dataset section above.