Machine Learning with FEARS index: Does the inclusion of investor sentiment improve a machine learning model's ability to predict volatility?

By

Student: Andrew James

JMSAND008

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In partial fulfilment of the requirements for the degree

M Com in Finance specialising in Financial Management

Faculty of Commerce

UNIVERSITY OF CAPE TOWN

Date of submission

Supervisor [s]: Chun-Sung Huang

Declaration

I, *Andrew James*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 05/02/2023

Plagiarism Declaration

COMPULSORY DECLARATION:

1. This dissertation has been submitted to Turnitin (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.
2. I certify that I have received Ethics approval (if applicable) from the Commerce Ethics Committee.
3. This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed, and has been cited and referenced.

Student number	JMSAND008
Student name	Andrew Michael James
Signature of Student	Signed by candidate
Date:	05/02/2023

Acknowledgements

Firstly, I would like to thank my supervisor Chun-Sung Huang for his invaluable advice during this research. Without his consistent guidance, this study would not have been able to be complete. Thank you for all of your assistance throughout the entire process.

I would also like to thank my wife, who was a constant ear for my frustration when I could not get something in my code to work. I will be forever grateful for your constant support and your willingness to listen even when you didn't know what I was talking about.

Abstract

The aim of this study is to determine whether the inclusion of investor sentiment allows machine learning methods to produce improved predictions of volatility in equity markets. Specifically, the investor sentiment measure is constructed as an index by using search volume data of different search terms obtained from Google Trends. The resulting Financial and Economic Attitudes Revealed by Search (FEARS) index is then utilised as a feature to forecast volatility via three different machine learning (ML) techniques, namely the Random Forest, Artificial Neural Network (ANN), and Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) methods. A consolidated dataset, where all G7 countries were combined into a single series, as well as an individualised dataset, where each individual country is analysed independently, were used to test the different ML methods' volatility forecasting ability. Our results show that, for the consolidated dataset, the inclusion of the FEARS index does not provide significant additional predictive power. However, through the individualised dataset, the FEARS index was shown in certain cases to provide greater predictive accuracy. Furthermore, it was observed that the LSTM-RNN outperformed the ANN and Random Forest methods, which indicates that our volatility prediction indeed benefits from elements of prior periods' volatilities as feature variables.

Contents

List of Figures

1. Introduction

The purpose of this study is to investigate whether the inclusion of an investor sentiment factor would allow machine learning (ML) models to produce improved predictions of Realised Volatility in equity markets.

The structure of this Chapter is as follows, first, we will introduce the reader to the research problem. We will then explore the objectives of the study and finally, we will show how this study contributes to the academic understanding of ML and investor sentiment.

1.1 Background

Rational investors have been shown to require a fixed rate of return, which shows us that stock prices should follow a random walk pattern (Samuelson, 1965). However, it was shown that stock price volatility is five to thirteen times too high to be justified by future dividends (Shiller, 1981). This tells us that market volatility is greater than the market expectation of volatility, researchers have hypothesized that this is because of noise traders, or investor sentiment in the market.

Investor sentiment has been studied extensively since 1965 and has been shown to provide mixed results in different countries. In some countries, investor sentiment has been shown to be a significant factor in explaining market returns and volatility, while in other countries it has been shown to have no significant explanatory power. Countries such as France were shown to have a positive correlation between returns and investor sentiment by Foucault, Sraer and Thesmar (2011). However, for countries such as Germany, there tends to be no significant relationship between market returns and investor sentiment (Finter, Niessen-Ruenzi, and Ruenzi, 2012).

Since the advent of the term noise traders, different proxies have been discussed in the academic literature. Da, Engelberg and Gao (2015) proposed the Financial and Economic Attitudes Revealed by Search (FEARS) Index, which makes use of Google Search Volumes as a proxy to determine the economic household sentiment of an economy. With the increased use of the internet and social media, the FEARS Index allows for the capturing of household sentiment towards the general economy. Subsequent studies, such as Solanki and Seetheram (2018), have shown that the FEARS Index holds explanatory power in several markets.

While the application of ML in the field of finance has existed for several decades, it has only become popular in more recent decades. This is mainly due to the huge growth in the computing power accessible by individuals. Which has increased to a point where a personal computer can now perform complex ML algorithms. Many different ML models have been used in finance, such as the Artificial Neural Network (ANN), the Random Forest, and the Long-Short Term Memory (LSTM) model.

In this study, we will utilise the LSTM, ANN, and Random Forest ML models. We contribute to the existing literature by providing further empirical evidence on the predictive accuracy of different ML models in forecasting Realised Volatility in the equity market. The robustness and accuracy of the models are evaluated through a set of well-accepted error measures.

1.2 Research Problem

Thaler (1993) was the first to define noise traders as an individual who do not trade on rationality but rather on sentiment. The study of behavioural finance has grown significantly since then as more research has shown that the market is not efficient, as described by Fama and French, due to the cognitive biases and emotions of traders. Since the definition of noise traders by Thaler (1993), Da, Engelberg and Gao (2015) used Search Volume data from Google Trends to create the FEARS Index as a proxy for household sentiment. Further research by Solanki and Seetharam (2018) found that the FEARS Index had explanatory power for stock market returns, across four of the G7 countries that were investigated.

In this research, we aim to investigate whether the inclusion of an investor sentiment factor, as proxied by the FEARS Index, improves ML models' ability to predict realised volatility in equity.

We answer this research question by training the ML models with the factors identified by Chen, Roll and Ross (1986), along with the FEARS Index. We will then attempt to predict Realised Volatility and compare that with the actual volatility for the period.

1.3 Research Contribution

The aim of this study is to contribute to the existing literature around ML and investor sentiment. We will do this by investigating whether the significant explanatory power that has been demonstrated by other academic research in the past also has predictive power. In particular, we make use of Search Volume data from Google Trends to construct the FEARS Index. The FEARS Index is subsequently used as a proxy for investor sentiment to improve the predictive accuracy for realised volatility.

1.4 Outline of Research

The remainder of this thesis is set out as follows. In Chapter 2, we explore the existing literature, including the theory of machine learning within finance as well as investor sentiment and the effect of that sentiment on the efficient market hypothesis. In Chapter 3, we will describe the data and methodology that was used in this study to create the FEARS Index, the ANN, the LSTM, and the Random Forest model as well as other components important for the study. The empirical results of our analyses for both the individual countries as well as the combined dataset will be presented in Chapters 4 and 5. Finally, Chapter 6 concludes the thesis, with the limitations of the study as well as areas for future research being explored.

2. Literature review

This chapter provides a review of the academic literature around asset pricing, the role that sentiment plays in markets, as well as the machine learning application in finance. This chapter is structured as follows. First, we will look at different asset pricing models. This is followed by a review of the literature on investor sentiment, which includes the theory of the efficient market hypothesis and how it relates to investor sentiment. This is followed by a review of the research regarding sentiments ability to explain movements in financial markets. Lastly, the chapter looks at the application of ML to financial market predictions.

2.1 Asset Pricing Models

2.1.1 Capital Asset Pricing Model

Asset pricing has been an area of interest for almost as long as exchanges have existed. In that time, many different models have been proposed and tested, including the Capital Asset Pricing Model (CAPM) that was introduced by Sharpe (1964) and Litner (1965). This is followed by the Arbitrage Pricing Theory (APT) introduced by Ross (1976). The APT model was proposed as an alternative to the CAPM due to some of the underlying issues of the mean-variance model that is employed by the CAPM, such as the assumption of normality in returns (Ross, 1976). Ross (1976) looked to formulate a model that would better explain expected returns. This was done by using significant factors that would still satisfy the initial assumption of the CAPM, whilst still acknowledging that a single factor cannot capture all of the risk in the market. Ross (1976) states that the APT relies on the arbitrage relationship rather than the equilibrium condition that CAPM relies on.

Formally, the CAPM can be represented as follows:

$$E(R_P) = r_f + \beta_m[E(R_m) - r_f]$$

Where $E(R_P)$ is the expected returns of the portfolio, r_f is the risk-free rate, β_m is the market beta and $E(R_m)$ is the expected returns on the market portfolio. Intuitively, the model suggests that the expected rate of return has a linear relationship with the factors described in the CAPM.

The Fama and French (1993) three-factor model attempts to further explain the risk and returns relationship by adding two additional factors to the CAPM. Specifically, the two factors are size and value. While the former is measured by market capitalisation and reflects the additional risk of smaller entities. The latter depicts the additional returns from owning out-of-favour stocks that have positive valuations. Fama and French (2014) then proposed two additional factors to formulate a five-factor model. The two additional factors were a profitability factor as well as an investment factor. The profitability factor is the difference between the returns on a diversified portfolio of shares with strong and weak profitability. The investment share is the difference in returns of a diversified portfolio with high and low investment shares. Similar to the Fama and French three-factor model, the five-factor model is designed to capture patterns in average returns. However, the theoretical underpinning of this is considered flimsy by many researchers (Fama & French, 2017). The empirical outcome of the five-factor model was not as favourable as the three-factor model as it failed to capture and explain the variation in returns (Solanki, 2017).

Formally, the Fama French Five-Factor model can be presented as follows:

$$E(R_P) = r_f + \beta_m\left[E(R_m) - r_f\right] + \beta_{SMB}(SMB) + \beta_{HML}(HML) + \beta_{RMW}(RMW) + \beta_{CMA}(CMA)$$

Where $E(R_P)$ is the expected returns of the portfolio, r_f is the risk-free rate, β_m is the market beta, $E(R_m)$ is the expected returns on the market portfolio, SMB represents the size factor, HML represents the value factor, RMW represents the profitability factor and CMA represents the investment factor.

2.1.2 Arbitrage Pricing Theory

The APT model in its simplest form attempts to explain asset returns using a list of factors. The model does not specify how many factors should be included nor does it specify the nature of those factors (Solanki & Seetharam, 2018). Empirical testing has found that the optimal number of factors for the APT model is between four and five factors (Roll and Ross, 1980). This lack of uniformity has resulted in the use of many different factors by different market participants. The first test of the APT model was performed by Gehr (1978). Where he found only one factor was significant for the forty-one companies that he analysed over the 30-year period being investigated (Connor and Korajczyk, 1992).

Formally, the APT model can be demonstrated as follows:

$$E_i = \rho + \beta_1 RP_1 + \beta_2 RP_2 + \beta_3 RP_3 + \dots + \beta_n RP_n$$

Where ρ is the risk-free rate of return, β_n is the sensitivity asset returns to a specific factor, and RP_n is the risk premium of factor n.

In Chen, Roll, and Ross (1986), the authors determined that certain macroeconomic factors, and the risk associated with these changing factors, were rewarded by returns in the stock market. These factors included the spread between long- and short-term interest rates, unexpected inflation, and industrial production, to name a few. Since the work of Chen, Roll, and Ross (1986) many more studies have been done to determine which factors are relevant in different countries. Kraft and Kraft (1977), Kia (2003) and Flannery and Protopapadakis (2002) found that factors such as money supply, commodity prices, balance of trade, and unemployment were factors that could be used to determine the returns of stock markets in the United States. Masuduzzaman (2012), and Sarwar, Mateus, and Todorovic (2015) found that significant factors in determining returns of the stock market in the United Kingdom, were money supply, exchange rates, credit spread, and GDP growth. Whereas factors such as money supply and exchange rates were found to have predictive power in Germany by Masuduzzaman (2012).

2.2 Investor Sentiment

The work of Selden (1912) was the first to provide evidence of a connection between finance and psychology. The author's premise throughout his book was that price movements in the market are driven by the mental attitudes of those investing (Solanki, 2017).

The efficient market hypothesis states that the current market price reflects all available information (Fama, 1970). Hence, an investor cannot make a return greater than the risk that they are taking on. In other words, investors cannot earn excess risk-adjusted average returns as the prices are set by market participants who are sensible and understand Bayes' law (Barberis and Thaler, 2003). Thus, as new information becomes available to investors they will change or modify their beliefs to align with new information received. The limitations of traditional financial theory led to the development of the study of Behavioural Finance. In this area of finance, investors are not assumed to be rational as was assumed by Fama (1970) in his efficient market hypothesis, instead, investors will let their emotions affect their trading decisions (Ricciardi and Simon, 2000).

2.2.1 Noise Trading Theory

Samuelson (1965) demonstrated that when a rational investor requires a fixed rate of return, stock prices should follow a random walk. This was later confirmed in Fama (1965) that stock prices are in fact close to a random walk. Shiller (1981) found that stock price volatility was five to thirteen times too high to be justified by future real dividends, this result was further confirmed in LeRoy and Porter (1981). Shleifer and Summers (1990) and De Long et al. (1990) found that volatility in the market was greater than the market expectation for volatility, leading to the conclusion that so-called noise traders risk needs to be considered in addition to the rational market expectation of volatility. Noise is the opposite of news, and rational traders make decisions based off of news, facts, and forecasts, while noise traders make decisions based off of anything else (Thaler, 1993).

According to Thaler (1993), noise traders do not trade on rationality alone but allow for sentiments, cognitive biases, and emotions to affect their trading decisions. These noise-trading investors tend to trade on what Black (1986) termed pseudo-signals rather than actual market signals. The presence of these noise traders in the market makes the risk of arbitrage higher as noise traders can make the stock price deviate even further from fundamentals (Solanki and Seetharam, 2018). Noise traders tend to overreact to good and bad news that comes into the market. This creates additional price pressure and lowers the expected returns, moreover, noise traders tend to have poor market timing. Noise traders tend to sell low and buy high. They do this because they tend to follow the crowd and trade when other noise traders are trading. Thus, noise traders tend to have greater capital losses due to poor market timing rather than from their change in perception (Lee, Jiang and Indro, 2002). The Friedman effect tells us that the above changes result in higher market risk and lower expected returns, the magnitude of the Friedman effect on returns depends on the climate that noise traders create in the market. A rise in the number of noise traders in the market can increase market uncertainty which will result in rational and informed risk-averse investors being crowded out of the market, therefore the larger the proportion of noise traders in the market the higher expected returns will be (Lee, Jiang and Indro, 2002).

Yu and Yuan (2011) and Karlson, Loewenstein and Seppi (2009), found that sentiment traders tend to trade more and more aggressively in periods of high sentiment. Barber and Odean (2008) confirmed that sentiment investors tend to have a moving effect on markets in periods of high attention days. That being days on which news puts additional focus on a specific equity. The authors were also able to find that the reason for this relationship of higher and more aggressive trading during high attention periods is because sentiment investors only sell stocks that they already own. Hence, they do not short sell in periods of low attention days decreasing their market activity in periods of low attention days. These sentiment traders also tend to lack the knowledge to accurately estimate the risk that is present in the market, these misestimates of the variance weaken the mean-variance relationship in the market (Yu and Yuan, 2011).

Fundamentally noise trading suggests that investors will trade on either the bullish or the bearish sentiment that they are experiencing (Solanki, 2017). In other words, investors will trade on the sentiment that they believe is in the market. The study of noise trading has allowed us to see the effect that irrational traders, who trade on sentiment, will have on the market, along with how they behave in the market (Solanki, 2017).

2.2.2 Sentiment Proxies

Proxies for investor sentiment have been investigated and used for as long as investor sentiment has been studied. Baker and Wurgler (2007) looked at several imperfect proxies for investor sentiment and settled on six proxies to make up their sentiment index. This includes factors such as trading volumes and the number of new IPOs, along with another 4 factors that were decided upon due to the ability to predict sentiment and the availability of data. Social media has been ranked as one of the most important forms of alternative data being used by many hedge funds. With many hiring experts in Natural Language Processing (NLP) seeking employees to scrap data from financial news sites, social media, and regulatory filings and releases to be used as news sentiment by the hedge funds in their trading strategies (Dixon and Halperin, 2019). The seminal work in sentiment proxies has been done by Baker, Bloom and Davis (2016), in which they developed the Economic Policy Uncertainty Index (EPU). The authors did this by using newspapers headlines as a proxy for economic outcomes, and found that economic uncertainty leads to negative economic outcomes. They also showed that economic uncertainty has increased worldwide since 2007. In addition to the above, Xu et al (2021) found that the Chinese EPU can be used to predict next month's returns of the Chinese A-share market. However, the reliability of predictions greatly decreases when an event that increases the uncertainty in the market occurs.

Da, Engelberg and Gao (2015) proposed an innovative way of measuring investor sentiment by using the Search Volume data provided by Google Trends. The authors then quantified the effect that the terms had on asset prices and fund flows, allowing them to develop an index they termed the Financial and Economic Attitudes Revealed by Search (FEARS) Index. To create the FEARS Index, the authors built a list of search terms that revealed sentiment toward economic conditions. This list included words such as 'bankruptcy', 'unemployment', 'crisis', 'inflation', 'recession', and 'security'. The FEARS index created in their study represented the sentiment of American households. In this process, they were able to determine that terms that had the most significant relationships with returns demonstrated exclusively negative relationships. Hence, the most significant search terms demonstrated a negative relationship with returns demonstrating that negative sentiment tends to be most significant than positive sentiment, despite the fact that their list of search terms included words with both positive and negative sentiment. This supported the results found by Tetlock (2007) that negative terms in the English language were more useful in determining sentiment than positive words.

2.2.3 Sentiment in explaining Returns and Volatility

It is clear from what is above that sentiment affects the returns in a market as well as the volatility. The theory of noise traders tells us that they will move prices in the market away from the intrinsic value making it possible to generate excess returns above what should be possible. Due to the fact that these noise traders also tend to move in groups. They will have an effect on the volatility of the market as they will create large sell-offs at certain times generating large variations in market prices and driving up the volatility of the market.

Foucault, Sraer and Thesmar (2011) looked at a subset of the French stock market and found that there was a positive correlation between retail trading and volatility. Mao, Counts and Bollen (2011) compared six different sentiment proxies. Including but not limited to Google Search volumes, Twitter Investor Sentiment (TIS), and investor sentiment survey data using the Investor Intelligence and Daily Sentiment Index surveys. The authors found that all six of the sentiment factors demonstrated a significant correlation with log returns, as well as Volatility Index (VIX) values. Using a Granger causality test they found that Google Search volumes improved weekly forecasting accuracy, but this was not the case with investor sentiment survey data. They further found that the inclusion of Google Search volumes improved the weekly forecasting accuracy of returns for the Dow Jones Industrial Average when returns are trending downwards, as well as for the predictions of VIX during periods of high volatility.

Dimpfl and Jank (2015) found that a rise in investor attention, as measured by an increase in Google Search volume, happened in times of heightened stock volatility. The authors found that including search queries in three different volatility prediction models provided more accurate predictions of the realised volatility. Finally, the authors found the greatest accuracy in predictions occurred in both the long run as well as periods of high volatility. This confirms the results found by Mao, Counts and Bollen (2011), that sentiment provides greater predictive power in times of higher volatility. Hamid and Heiden (2015) found that the inclusion of search volume data from Google Trends that daily predictability was not improved by the inclusion of Google Search volume data. The authors concluded that this was a result of Google Trends limiting access to daily Search Volume data, as the data is standardised into 90-day windows. However, Hamid and Heiden (2015) did find that Search Volume data provide additional predictive accuracy to their weekly volatility predictions.

Audrino, Sigrist and Ballinari (2020) determined the predictive power of economic and sentiment variables using a predictive regression model, they found that Google searches on key financial terms such as “financial markets” and “stock markets”. As well as daily volumes of company-specific messages posted on StockTwits were the most relevant factors in predicting volatility. The authors found that including a sentiment factor in a traditional heterogeneous autoregressive (HAR) model reduced the mean square prediction error (MSPE), especially in times of heightened volatility. Further, the authors found that although the improvements are significant, they tend to be small in magnitude. This confirmed the findings of Facault, Sraer and Thesmar (2011), Mao, Counts and Bollen (2011), Dimpfl and Jank (2015) and Hamid and Heiden (2015), that Search Volume data from Google Trends are significant in determining the volatility of the market, especially during high volatility periods. Audrino, Sigrist and Ballinari (2020) found that the predictive power of sentiment was generally lower for companies that have small market capitalisations and that the predictive power was greater for large market capitalisation.

Solanki and Seetharam (2018) followed a similar process to develop their FEARS index across multiple countries including the United States, Germany, the United Kingdom, and Japan. The authors found that the FEARS index was statistically significant in explaining returns at a 5% significance level in three of the four countries, with the only exception being Germany. The authors also found that the FEARS Index has significant explanatory power in both developed and developing markets.

Sentiment factors were shown to not influence excess returns on the FTSE 100 by Johnman, Vanstone and Gepp (2018). They used news articles from the Guardian Media Group from the beginning of 2000 until June 2016 to determine positive and negative sentiment. The authors then regressed the sentiment measured they calculated to daily excess returns and found that they did not influence those returns. This contradicts the findings of Solanki and Seetharam (2018) who they found that the FEARS index had significant explanatory power over the UK market. Johnman, Vanstone, and Gepp (2018) did however find that when these same sentiment factors were regressed against a proxy for volatility on the FTSE 100 that the negative sentiment increased the volatility and positive sentiment reduced the volatility in the market.

Finter, Niessen-Ruenzi, and Ruenzi (2012) looked at whether sentiment was a predictor of return spreads of 955 German stocks that were listed on the Frankfurt Stock Exchange in the period of 1993 and 2006. As a sentiment proxy, they followed the work set out by Baker and Wrugler (2006), as such Finter, Niessen-Ruenzi and Ruenzi (2012) applied several macroeconomic factors into an index. These factors included things such as employment level and retail sales, amongst others. Their study showed that sentiment only had weak predictive power in Germany due to the fact that the German market was largely populated by sophisticated institutional investors unlike what was found in US markets. These results are supported by the later study done by Solanki and Seetharam (2018) discussed above where the FEARS index appeared to have no statistical significance in explaining returns in Germany.

Sentiment was shown to have a significant predictive ability in the Japanese market by the work done by Vuong and Suzuki (2020), using two proxies for investor sentiment namely, the Consumer Confidence Index (CCI) and the VIX. Using these two proxies they found that there was a significant link between investor sentiment and returns in Japan, which is in line with the work done by Solanki and Seetharam (2018) that was discussed above. Further Vuong and Suzuki (2020) found that the effect of sentiment on stock returns seems to fall away fairly quickly as they found that neither the CCI nor the VIX were able to predict returns in the near or medium term.

You, Guo and Peng (2017) showed that there was no causal relationship between their happiness index and stock returns on Canada's Toronto Stock Exchange (S&P/TSX) along with several other international stock exchanges including the FTSE 100 and the S&P 500. The happiness index was created by extracting data from Twitter using Natural Language Processing ML techniques. These results support the results of Johnmann, Vanstone and Gepp (2018), where they found that sentiment did not influence excess returns on the FTSE 100.

Ho and Hung (2012) found that sentiment, as proxied by CCI, has predictive power in Italy, France and the US, where periods of high consumer confidence were followed by periods of low excess returns on their respective stock exchanges. They also found that shifts in consumer confidence affect conditional volatility in all three of the previously stated countries. In contrast to the work of You, Gou and Peng (2017) and Johnmann, Vanstone and Gepp (2018), Ho and

Huang (2012) showed that CCI, as a proxy for investor sentiment, had predictive power in the UK market, similar to the finding of Solanki and Seetharam (2018).

The inclusion of sentiment factors have shown mixed results in different countries as well as with different sentiment proxies. With some countries consistently showing that sentiment affects returns and volatility while other countries consistently display the opposite. The different proxies used in the academic literature also show varying relationships, with some proxies only showing significant relationships in certain countries. Further, some proxies did not show significant relationships in countries where other proxies have demonstrated that returns are affected by sentiment.

2.3 Machine Learning and its application in Finance

Machine learning has been being used in the financial services industry for over forty years (Dixon and Halperin, 2019). However, the last decade has seen a renaissance in supervised machine learning techniques (Chlebus, Dyczko and Woźniak, 2021), as the computing power of the average person has grown enough so that any person can partake in the use of machine learning. With that in mind, we know that the application of machine learning in finance is still in its infancy and is incomplete in many ways. Given that finance is a perfect area to deploy machine learning given the vast amounts of available data and the direct profit and loss implications, all of which exist within a highly competitive environment where every advantage is seized by industry experts (Dixon and Halperin, 2019).

Broadly speaking there are three different areas of machine learning, namely supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves the use of either parametric or non-parametric algorithms that learn the relationship between the feature variables (input data) and the target variables (output data) that the factors are being regressed to (Dixon and Halperin, 2019). For supervised learning the data being used must be labelled, meaning that input data must be paired with output data. Thereafter, the features and the factors are provided to the algorithm so that the relationship can be "learnt" by the model. Unsupervised learning is a data mining technique that separates and reduces the data's dimensionality. These techniques will change and generalise approaches to reduce data to their principal components (Dixon and Halperin, 2019). Unlike supervised learning, where labelled data is used, in unsupervised learning input data is not coupled with any output data. Finally, reinforcement learning is the most complicated of the three. However, it holds the most promise for use in finance, but due to the complexity is the most underutilised in the field.

Reinforcement learning will use stochastic control and use the feedback generated to alter the state of the input data (Dixon and Halperin, 2019).

In this work, we focus on the use of supervised learning for Realised Volatility forecasting. Specifically, we utilise deep learning, where linear regression and logistic linear regression models are generalised by adding multiple layers into the composition of a neural network. This allows the network to learn any linear as well as non-linear relationships that exist between the input and output variables (Dixon and Halperin, 2019). It is commonly known that many financial and economic variables are non-linear. ANN allows for both the linear and non-linear nature of variables to be applied to predictions (Enke and Thawornwong, 2005). Qi and Maddala (1999) found that ANN outperformed linear models in predicting excess returns. The early work of Kryzanowski, Galler and Wright (1993) found that ANN models demonstrated the ability to correctly classify 72% of the positive and negative returns.

Supervised learning techniques use multiple factors (or features) to explain stock returns. Abe and Nakayama (2018) used 25 features including but not limited to Book-to-market ratio, Earnings-to-price ratio, and dividend yield. Using these 25 factors they attempted to predict the stock returns for the MSCI Japan Universe by using the previous 5 data points. The authors then compared 3 different machine learning algorithms, a deep neural network, Support Vector Regression, and Random Forest. They found that deep neural networks with a greater number of layers performed better than shallower neural networks and found that the deep neural network also outperformed both the Support Vector Regression, as well as the Random Forest algorithm.

Most return predictions focus on predicting stock market returns rather than returns on an individual stock level and tend to largely rely on neural network algorithms to make these predictions (Abe and Nakayama, 2018). Cao, Leggio and Schniederjans (2005) were able to show that ANN better predicts stock market returns on the Shanghai Stock Exchange than linear models, these results were shown to be statistically significant across all firms that made up their sample.

Krauss, Do and Huck (2017), used three different machine learning algorithms namely, a deep neural network, gradient-boosted trees, and random forests, to develop a statistical arbitrage method on the S&P 500. They were able to find two significant observations through their research, firstly that the random forests were able to outperform both the deep neural network as well as the gradient-boosted trees in their application to develop a statistical arbitrage strategy. Secondly, they found that combining the predictions of the three models as an equal-weighted ensemble prediction outperformed all three of the models in isolation. Further, they noted that hyper-parameter optimisation of the deep neural network could still yield advantageous results.

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that was designed with the intention of overcoming the limitations of many other RNNs, such as the issue of the vanishing and exploding gradients (Sak, Senior, & Beaufays, 2014). Fisher and Krauss (2018) were amongst the first to apply an LSTM model to financial time series data. They applied an LSTM model to predict returns on the S&P500 using data between December 1992 up to October 2015 and compared the results to the results from a Random Forest model, a standard deep neural network, and a simple logistic regression model. They found that the LSTM model outperformed the other 3 models in all cases, except for during the 2008 Global Financial Crisis where the Random Forest performed best.

Although the application of machine learning in finance is relatively new, the testing that has been done has been fairly extensive, with the vast majority of supervised learning techniques being tested, as evidenced from the above. It is still an area that is in its infancy, as stated by Dixon and Halperin (2019), which leaves room for additional application of machine learning models to many areas of finance and huge potential for future research.

3. Data and Methodology

This chapter is laid out as follows, firstly the data used in the study is discussed, including the countries chosen and which factors were chosen as feature variables. The data section concludes with a discussion on the construction of the FEARS Index, and its application as a proxy for market sentiments.

Thereafter, we describe the methodology applied in the study. This discussion includes the relevant transformations made to the feature variables. The three ML models used in the study, namely the ANN, LSTM, and Random Forest model. Then a discussion on the metrics used to assess the accuracy of the model predictions is presented. Finally, the methodology section is concluded with a discussion on the feature importance functions that were used in analysing the models used in the study.

3.1 Data

The choice of sample period was limited by the data available on Google Trends that was used to create the FEARS Index. Google has only made publicly available the Search Volume data from 1 January 2004 on Google Trends. The observation period in this still will be between January 2004 and December 2021, which results in a total number of 216 monthly observations for each country.

We excluded data from 2022 from our sample period due to the large amount of economic uncertainty in 2022. This economic uncertainty is due to accelerating inflation and the unstable global political environment with the invasion of Ukraine. It was deemed that these observed global issues would introduce confounding variables into the study that would affect the results of the models being developed. ML techniques are notorious for their dependency on a large amount of observations to provide accurate predictions. To overcome the shortfall of our limited data, we carry out an additional set of analyses over and above the individual country volatility predictions. Specifically, this involves combining data from across the different regions to boost the number of observations in an attempt to improve the ANN and Random Forest predictions.

The analysis of this research will be twofold. Firstly, we will investigate the predictive ability of an overall dataset that combines the dataset from all of the countries in this study. Secondly, we will be analysing the forecasting of volatility across each individual country. The ANN and Random Forest model will be applied to both the combined dataset as well as the individual country dataset, whereas the LSTM model will only be applied to the individual country dataset. This is done due to the nature of the models and how they are structured, this will be discussed further below.

3.1.1 Choice of Countries

The G7 countries were chosen as the observation countries in this study. This choice was twofold, firstly the G7 countries include the most pertinent markets around the world. These are the markets that most of the investors are focussed on. Secondly, the availability of data, the G7 countries have more widely available data for longer periods of time, this allows for comparison across countries. An additional point of interest from the G7 countries is that in some of the country's investor sentiment has been shown to not have significant explanatory power in those markets, this includes such research as Johnman, Vanstone and Gepp (2018), and You, Gou and Peng (2017).

3.1.2 Volatility Data for each country

For all the G7 countries the daily closing price data for each respective index was collected for the stated sample period and was then used to calculate the 14-, 21- and 28-day realised volatility for each country in the chosen sample period.

3.1.3 Choice of APT factors

Following, the initial empirical work by Chen, Roll and Ross (1986), the following variables were used to derive the factors which are utilised as feature inputs into the ML models:

- Risk-free rate – the 90-day treasury bill rate of each country was selected to represent the respective risk-free rate.
- Long-term government bond yields – for all of the seven countries selected a 10-year government bond was used.
- Inflation – inflation was derived from the consumer price index (CPI) of each respective country being considered.
- Industrial Production – the monthly industrial production index was used.

The final variable that was used by Chen, Roll and Ross (1986) was the return on high-yield bonds. To obtain this Chen, Roll and Ross (1986) used the yield on corporate bonds at a rating of Baa or lower. In later studies such as those done by Sonlaki and Seetharam (2018), the yields on investment-grade corporate bonds were used due to the lack of available data that tracked yields on corporate bonds below investment grade. This factor was omitted from our study due to the lack of availability of components used to calculate returns on high-yield bonds. Although the data was available for some countries it was not available across all of the G7 countries. By excluding the return on high yield bonds from all of the countries it allows for comparison across all of the regions. Therefore, the return on high-yield bonds was not included as a factor in this study.

3.1.4 Investor Sentiment Proxy

In this study, we use the Search Volume data from Google Trends to construct the FEARS Index, which will be used as a proxy for investor sentiment. Da, Engelberg and Gao (2015) proposed the FEARS Index as an investor sentiment proxy. The FEARS index was chosen in this study as Google is used as the dominant search engine in all the countries in question, with Google holding over 80% of global search engine market share (Bianchi, 2024). The popularity of Google also translates to increased data availability, which is advantageous to this study given the heavy reliance of machine learning models on data. Further to this the choice was made not to use EPU due to the changing global and technological landscape, with Google being used more widely than newspapers the Google Trends data would provide a more accurate reflection of market sentiment. Finally, given the performance of the EPU in times of heightened uncertainty, as discussed above, the FEARS Index will likely be more reliable given the uncertainty created by the COVID 19 pandemic as well as the Ukraine war. The methodology in this research is built off of the methodology described by Da, Engelberg and Gao (2015).

In particular, we will adapt the methodology of Solanki and Seetheram (2018), which includes the use of the primitive list that they used in their 2018 study. The methodology described below, around the generation of the primitive list, will be the methodology applied by Solanki and Seetheram (2018). Following that, the methodology will return to the methodology used in this study.

The Harvard IV-4 and Lasswell Value Dictionary were used as a starting point as they classify words into categories such as "positive", "negative, "weak", "strong" and so on. The aim of the FEARS Index is to capture household sentiment towards the economy as such words that are "economic" and have either "positive" or "negative" sentiment were chosen. These search filters resulted in a list of 163 words being obtained from the dictionaries, of which 92 were related to positive sentiment and 71 were related to negative sentiment. This initial list of words included words such as ‘bankruptcy’, ‘unemployment’, ‘crisis’, ‘inflation’, ‘recession’, and ‘security’, this initial list was termed the primitive word list (Solanki and Seetharam, 2018).

Once, the primitive word list was generated each word was inputted into Google Trends. This then returns the top searches related to each word in the primitive word list. This produces a list of search terms that are related to each word found on the primitive word list. This generates a list of 640 search terms. This list of 640 search terms included terms such as “Depression”, “Cryptocurrency”, “International Financial Reporting Standard”, and “Currency War”.

The next step was to eliminate any duplicate search terms and eliminate any of the search terms that are not related to economics and finance. The reason for the existence of some of these terms was due to the fact that some of the words on the original primitive word list created search terms that related to non-economic terms. Google Trends allows for the data to be filtered by the specific category to which the search is related. For example, you can restrict the data to being finance-specific terms or finance-related searches. However, the related search terms included both finance and economic terms and non-finance and non-economic terms regardless of whether the search term filter on Google Trends was set to Finance or if it was set to All Categories. For example, the primitive list included the word “Depression”, and related search terms for the word depression included terms such as “Depression meme”, “Postpartum Depression” and “Anti-Depressants”. These terms do not relate to finance or economics and as such were removed. This was done for all the search terms generated from the primitive word list. Finally, any terms with insufficient data were removed, this was deemed to be any search term that had an SVI of less than 1000 for the sample period. This limit of sufficient data was in line with the work done by Solanki and Seetharam (2018). This resulted in a final list of 551 search terms that would form part of the FEARS Index.

The SVI for each of the 551 search terms was then downloaded from Google Trends for the sample period. Google Trends allows the user the ability to select the region for which the SVI should be related through the use of a country filter that uses the IP Address of each respective country to determine where the search is situated. In this case, the region was set to worldwide to create a FEARS Index that could be applied to all the countries included in the study rather than developing a FEARS Index for each country. The SVI for each search term was downloaded on a monthly basis so as to align with the macroeconomic data that was being used as factors in the APT Model.

The change in monthly SVI (ΔSVI) is then calculated for each of the downloaded search terms. This was calculated as follows:

$$\Delta SVI_t = \ln(SVI_t) - \ln(SVI_{t-1})$$

In previous work done by Da. Engelberg and Gao (2015) and Solanki and Seetheram (2018) a rolling window regression was then applied to the ΔSVI to produce the FEARS Index. In this study, the decision was made to not use the rolling regressions approach with only the top 30 words forming part of the FEARS Index. In this study the full 551 search was used to create the FEARS Index, the reason for this is threefold. Firstly, there does not appear to be any clear consensus as to why the top 30 words were chosen as opposed to the top 50 or more. Secondly, this was done to prevent the FEARS Index in this study from becoming over-polarised. Thirdly, the use of all of the search terms will create a more holistic sentiment proxy that does not only focus on extreme sentiment identified. Rather it will provide a full view of the household sentiment in a worldwide.

To create the FEARS Index we let the data identify which of the search terms is most significant in explaining returns. This was done by regressing the ΔSVI and the market returns, to determine the historic relationships between search values and market returns. For example, the sentiment value for January 2005 was calculated by regressing any given search term on the market returns for the beginning of the period, January 2004, until January 2005. This will then indicate if the search term has a strong relationship with the market returns, whether that relationship is positive or negative.

Following that the t-statistic of each word and each regression is calculated and ranked from the most negative to the most positive. Tetlock (2007) and Da, Engelberg and Gao (2015) found that in the English language negative terms appear to be the most useful in identifying sentiment, as a sanity check on the regression results in this study were compared with this finding of previous studies by Tetlock (2007) and Da, Engelberg and Gao (2015). The regression done in this study aligned with that found in previous studies providing confidence regarding the regression and the subsequent output. The t-statistic rankings are then multiplied by the ΔSVI for each word and summed to determine the for each month to determine the FEARS Index for each time period.

This can formally be represented as follows:

$$FEARS_t = \sum_{i=1}^{m} R^i(\Delta ASVI_t)$$

where $R^i(\Delta ASVI_t)$ is the change in average SVI for search terms that had a t-statistic rank i for the period January 2004 until time period t, m represents the number of search terms in the sample.

Using the historic regression approach is most robust as it lets the search terms to express their own sentiment relationship to the returns, even if this relationship is not the same as what the Harvard IV-4 and the Lasswell Dictionary suggested the relationship should be.

3.2 Methodology

3.2.1 Realised Volatility

This study will attempt to predict realised volatility. The target variable in all of the models will be the 14-, 21- or 28-day Realised Volatility. To calculate the Realised Volatility first the log returns of each day was calculated. Those returns were then used to calculate the volatility of each of the Realised Volatility time horizons in this study.

In all cases, the returns were calculated using the following formula:

$$R = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Where, R is the daily returns, and P is the index price.

The realised volatility for the 14-, 21- and 28-day was calculated as follows:

$$RV_t = \frac{1}{m}\sqrt{\sum_{t=1}^{m}(R_t - E[R])}$$

Where, RV_t is the Realised Volatility, m is the respective forward-looking time horizon, 14-,21- or 228 days R_t is the daily return on any given day, $E[R]$ is the expected return for the period being looked at, either 14-, 21-, or 28 days.

3.2.2 APT Factors

Our choice of APT factors follows that of Chen, Roll, and Ross (1986). In particular, the chosen variables and their definitions are listed in Table 1. Similarly, to the data regarding the APT factors above the methodology here will closely follow that done by Chen, Roll, and Ross (1986).

The following variables were obtained from Bloomberg and will be used to derive the feature variables:

Symbol	Variable	Definition
I	Inflation	The change in the specific countries' CPI
GB	Treasury Bills	Return on 1-year government bonds of the specific country
LTB	Long-term government bonds	Return on 10-year government bonds of the specific country
IP	Industrial Production	Percentage change in a country's Producer Price Index (PPI)

Table 1: Basic Series

The above variables have then been used to derive the below variables, the variable below in Table 2 will used as the feature variables in this study. The Feature variables were calculated as follows:

Symbol	Variable	Source
MP_t	Monthly growth in Industrial Production	$\left(\frac{IP_t}{IP_{t-1}}\right) - 1$
I_t	Inflation	$\left(\frac{CPI_t}{CPI_{t-1}}\right) - 1$
Rho_t	Real Interest Rates	$GB_{t-1} - I_t$
TS_t	Term Structure of Interest Rates	$LTB_t - GB_{t-1}$

Table 2: Derived series

A correlation analysis was then performed on each of the economic variables, as well as the FEARS Index for the respective volatility data. This was done to determine if there was any relationship between the feature variables as well as the target variables. A strong relationship between variables may be an indicator that a variable should be removed as a feature variable.

A high correlation between variables is an indicator that the information in one variable is already captured by another of the feature variables. This check allows us to reduce the computational power that is required by the ML model. This is especially important when working with ML models as they can be very inefficient if too much information is being fed into the model. Also providing additional information can result in ML models overfitting the data, which will produce a model that is poor at making predictions when exposed to non-training data.

The following correlations between variables were calculated:

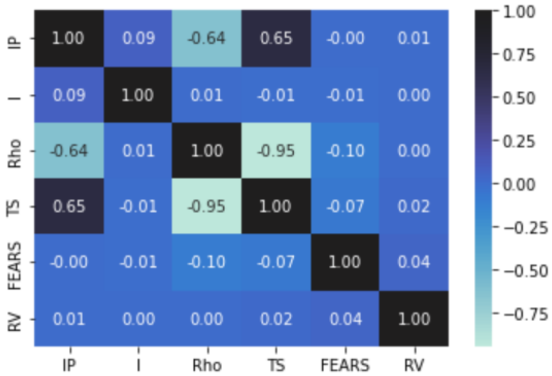


Table 3: Correlation plot for 14-Day Realised Volatility in the US

The correlation plots for both the 21- and 28-day realised volatility were similar to that of the 14-day volatility as such it has not been presented here.

After running the correlation analysis, it was determined that none of the variables should be removed as feature variables. The correlation between Rho and TS was noted however it was determined that both should be left in, for two reasons. Firstly, to align with the previous work done by Chen, Roll and Ross (1986) where both variables were included. Secondly, they explained different elements of the market. Both of which were important to investors and that the correlation was likely due to the fact that they were derived from a similar source, government bonds, rather than because they explained the same element of the market.

The work done by Chen, Roll and Ross (1986), and Solanki and Seetheram (2018) both looked at how the factors above related to market returns. However, in this study, we will be using the same factors to make Realised Volatility predictions. The same factors are being used to make Realised Volatility predictions. This can be done for two reasons 1. Realised Volatility is derived from returns therefore a relationship between returns and Realised Volatility exists and the factors as each of the variables have an effect on the Realised Volatility in the market. 2. each of the factors used have an effect on the volatility in the market.

Investors in industrial stocks will watch the Monthly Industrial Production Index as a determinant for their trading decisions. If the Index outperforms the expectations, then they will buy more industrial shares over other sectors in the market and vice versa. As such changes in Monthly Industrial production can cause market participants to buy or sell their positions which will increase the volatility in the market. A change in inflation will lead to investors having to look to markets where they can generate a real return. Thus, if inflation increases there will be a movement to equity markets that offer higher returns. This move to equity markets will cause an increase in the Realised Volatility in the market as more investors are buying shares. As bond yields change investors tend to move towards the market that is offering returns that are commensurate with the risk of the market. Therefore, as bond yields increase investors tend to move away from equities into lower risk bonds that are offering greater returns. Rho and TS derivation from bond yields and the behaviour of investors when bond yields change is the reason that a relationship exists between Rho and TS and Realised Volatility.

3.2.3 Train Test Split

When using supervised ML models, a split of the data into a training set as well as the test set is done. The training data is then split further into a true training set and a validation set that is used to validate the model before performing the final testing of the model on the test set. Each of the different sets serves a specific purpose when developing the model. The training data is inputted into the model to allow the model to determine the relationships that exist between the feature variables and target variables. The validation set will then be used to assess the accuracy of the relationships identified from the training set. The validation set is used to assess the performance of the model prior to introducing the test set to allow for improvement of the hyperparameters of the model before performing the final test of the model. This allows for quicker turnaround times when building and iterating on the model being developed.

The test set is used to compare the model's ability to make predictions regarding the target variables when new data is introduced to the model. For example, the feature variables for a particular month will be inputted into the trained and validated model. The model will then make predictions of what it believes the target variable should be based on the relationships identified. These predictions will then be compared with the actual target variables in the test set of data making it possible to assess the accuracy of the model predictions.

The aim of the test set of data is to compare Realised Volatility predicted by the model with the actual Realised Volatility for the time period in question. As such the model should not be exposed to the actual Realised Volatility for the period. Therefore, the test set is split into a test set with the feature variables and a second test set that consists of only the target variables. The feature variable test set is then introduced to the model so that the model can make predictions. These predictions are then compared to the actual Realised Volatility in the period, stored in the target variable test set. We can then determine the accuracy of the model using the error metrics discussed later.

The conventional approach within data science is to have 75% of your data form part of the training set, and 10% of the training data as a validation set, and the other 25% be used to create the test set (Chen, 2021). Given the limited amount of data available for use in this study and the large amount of data required for these models to learn appropriately, we have deviated slightly from these norms. In an attempt to provide the models with enough data to appropriately recognise relationships between the target variable and feature variables, a training and test split of 94%/6% was used. This allowed for additional data to be used to train and validate the model, as well as sufficient test data to assess the accuracy of the model predictions.

The training and test sets are normally created by randomly splitting the data. However, when working with time series data, as has been used in this study, a random split will introduce hindsight bias into the model. This will happen because the model will be exposed to future volatility which will provide the model with additional guidance for their predictions. This tends to make the model very good at making predictions within the test data, as the test data may be from a time period that falls in between a time the model has already been exposed to during the training and validation phase. For example, the test data may come from January 2010 but the 12 months on either side of January 2010 formed part of the training set which will cause the model to be biased due to its exposure to future volatility. This may make the model poor at making predictions when new data that did not form part of the original time period is entered into the model. As such for this study the data has not been randomly split and assigned to the various data sets but rather the data has been split on a time axis to remove the effects of hindsight bias. Practically, this translated to a test set being formed with data between December 2020 and December 2021. Similarly, the training data is formed with data between January 2004 and November 2020.

To avoid the random split, we manually processed the splitting of data in our study. The data are first being sorted from earliest to latest time period, i.e. January 2004 until December 2021, and the first 96%, January 2004 until November 2020, of data being taken as the training set, which is then split further into the true training set and the validation set, and the last 4%, December 2020 until December 2021, of dates being used as the test set.

3.2.4 Data scaling

Data scaling plays an important role in machine learning as it provides the model with standard datasets, which is crucial in ensuring the accuracy of the model's predictions (Ahsan et al., 2021). The scaling of data in ML allows for the data to be brought closer together, generally within a certain range such as between 0 and 1. By having the variables closer together the ML model is able to more quickly understand the data that is being presented to the model. It also allows for greater accuracy from the model, as the model will be better able to understand the relationships when the variance amongst the data is smaller.

There are several different scaling algorithms to choose from in ML, including the Standard Scaler, Min Max scaler, and Normalisation, but there is no conclusive evidence as to which scaler is best (Ahsan et al., 2021). Similar to many other aspects in ML it is down to the practitioner and researcher to decide which is best for their specific model. This is done through an iterative process of testing a model's accuracy with each of the different scalers to determine which is best for that specific model and dataset combination.

In the case of this study that was shown to be the Min Max scaler. The Scikit Learn Min Max scaler was used in this study to scale the data. The Min-Max scaler transforms each feature and target variable individually between the given range (Pedregosa et al., 2011), in this case between 0 and 1. To scale the data, the Python package will first calculate a standardised X value. This is done by calculating the difference between the current variable and the smallest variable in the data set and dividing it by the range of the data. The scaled X value is then determined by multiplying the standardised value by the range and adding the minimum value, which in the case of this study is 0.

The transformation can be formally described as follows:

$$X_{std} = (X - X_{min})/(X_{max} - X_{min})$$

$$X_{scaled} = X_{std} \times (X_{max} - X_{min}) + X_{min}$$

Where, X_{min}, X_{max} is the feature range, in the case of this study that would mean that min is 0 and max is 1.

After the data has been fed into the model and the model has made its predictions, the inverse Min Max scaler is applied to the predictions so that the output of the model can now be viewed by users of the model in a way that aligns with the scale of the original data for appropriate interpretation.

3.2.5 Hyperparameter Tuning

Hyperparameter tuning is an important part of optimising ML models to generate more accurate predictions the best possible predictions. This does, however, bring in some complexities when attempting to make comparisons between different models as the model will be optimised to correctly reflect the data that is being fed into the model. Hyperparameter tuning may result in models that we are comparing being very different in their structure. This may mean that we are comparing one model with 1 hidden layer and 2 neurons in that hidden layer with a second model that has 3 hidden layers and 6 neurons per layer. This results in two different possible approaches, either hyperparameters of different models are kept relatively similar or each model is optimised for best performance.

In this study the choice was made to optimise each model for best performance rather than keeping model hyperparameters similar, this was done as it allows for the best version of each model to be compared allowing for the best model to be determined based on the best-optimized version of each model.

The hyperparameters of each model and how the tuning of said hyperparameter are discussed after the details of each respective model are below.

3.2.6 Random Forest

3.2.6.1 The model

Random Forest models are widely utilised due to its ease of implementation. We shall follow the methodology as outlined in Nevasalmi (2020), and Tan, Yan and Zhu (2019).

The random forest model was first introduced by Breiman (2001) and has built on previous machine learning models such as bagging and various other boosting models. Random forest is one of the most successful ensemble models (Dietterich, 2000), and has been extensively applied in finance research for both returns and volatility predictions.

The random forest model in its simplest terms is just an amalgamation of a large number of gradient-boosted decision trees. The original idea of the random forest was to improve upon the classification ability of bagging. This was done by adding additional randomness into each component of the model, while it is being built. This reduces the correlation between each component of the model in the final ensemble. Another key difference of the random forest models, when compared with other models such as the gradient boosting models, is that each decision tree in a random forest model is trained independently of each other. The output of each tree is then aggregated to produce the prediction.

Gradient-boosted decision trees began with the classification algorithm, AdaBoost, which was introduced by Freund and Schapire (1996). However, this model remained controversial with ML practitioners until it was found that the AdaBoost algorithm fit an additive logistic regression model (Friedman, Hastie and Tibshirani, 2000). The goal of gradient boosting models is similar to that of many other ML algorithms, that being the aim of the model is to minimise the expected loss of a predetermined loss function.

Gradient boosting trees work by taking individual decision trees that generally produce poor predictions. Normally this means that they have a high error term, such as Mean Squared Error. These individual trees are then combined in such a way that the new connected trees generate lower error terms than the two trees separately were able to generate. Due to the connected nature of the trees, the algorithms tend to learn slowly but provide highly accurate predictions.

Although the Random Forest model is an amalgamation of gradient-boosted decision trees it differs significantly in how the decision trees are built. The gradient-boosted decision trees will develop an understanding of the relationship between the feature variables and the target variable. It will then use this understanding of the relationship to make predictions. Whereas the Random Forest model randomly selects feature and target variables to create several independently trained decision trees that have the predictions of each tree aggregated.

The other benefit of Random Forest models when compared with the gradient-boosted trees is that the Random Forest models are able to avoid some of the overfitting issues of gradient-boosted trees. By creating random subsets of features and building trees using those subsets and then combining these independent subsets. Although this approach does prevent some of the overfitting it does come with the downside of being more computationally expensive, which results in Random Forest models requiring longer training time than the gradient-boosted trees.

The Random Forest model can be graphically represented as follows in Figure 1:

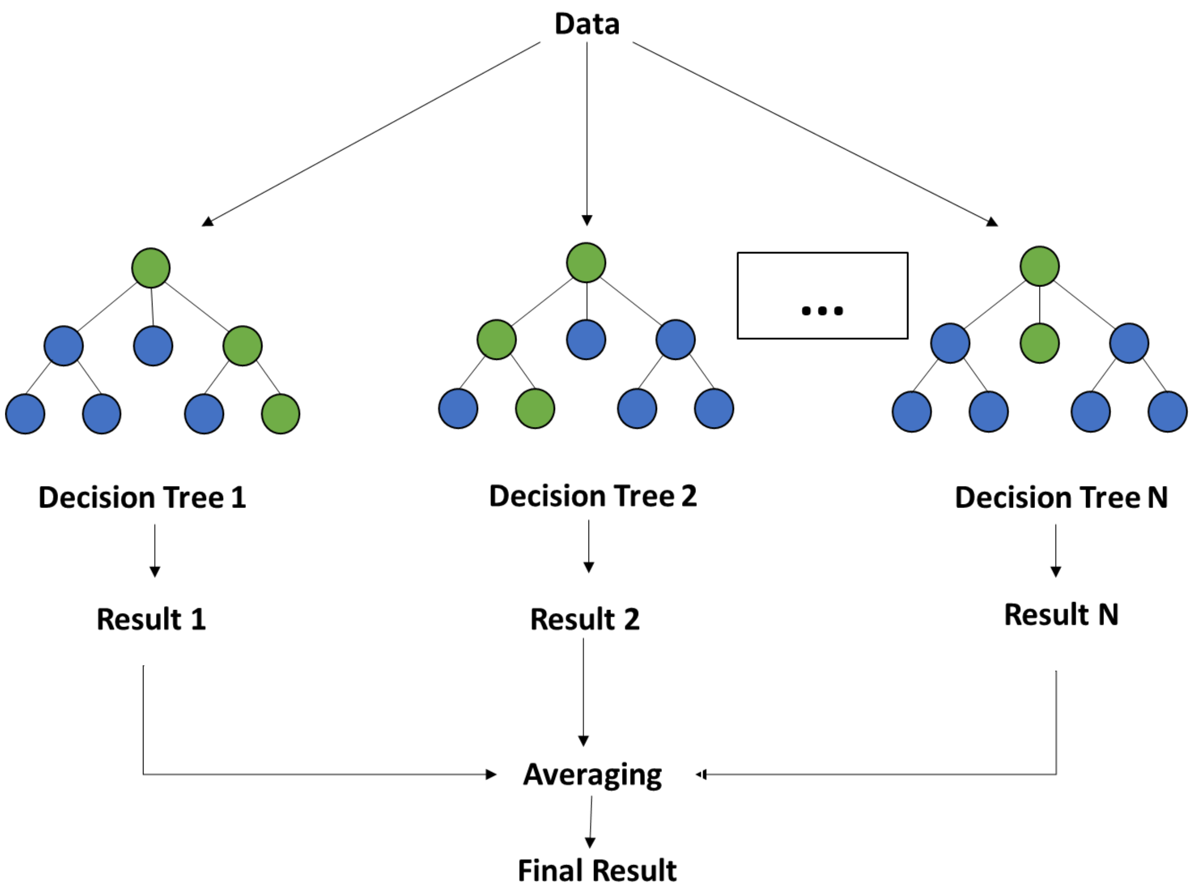


Figure 1: Random Forest model prediction process

3.2.6.2 Application of the Random Forest model in this study

As discussed above, the Random Forest model allowed for the use of a combined dataset to provide the model with additional learning data in the attempt to provide improved predictions. However, it is also possible to apply each country's data to the Random Forest model, which makes it possible to make comparisons between all the models employed. This resulted in 6 models being built using the combined dataset approach, one for each of the 14-day, 21-day, and 28-day Realised Volatility. All models built have a version with and without the FEARS Index to allow for an analysis of the difference that investor sentiment makes on the Random Forest model's ability to make predictions.

Another 42 models were then built for 14-day, 21-day, and 28-day Realised Volatility for each of the respective seven countries under observation. As with the above, all the models built have a model with and without the FEARS Index to allow for a comparison to evaluate the predictive ability of investor sentiment.

As with many ML models, the hyperparameters of random forest models are very important. The hyperparameters presented in Table 4 were considered with regards to Random Forest models:

Hyperparameter	Definition
n_estimators	This is the number of trees that the algorithm builds before taking the averages of all the predictions; in general, the higher the number the better the prediction performance of the model but the slower the model produces the predictions.
Max depth	This is the number of splits that each decision tree is allowed to make before making the prediction. If the split is too low the model may underfit the data and if it is too high, it may overfit the data.

Table 4: Random Forest Hyperparameters

3.2.7 Artificial Neural Network

3.2.7.1 The model

ANN models have become increasingly popular in the academic literature since the 1990s. Our methodology described below will follow that of Khandelwal, Adhikari, and Verma (2015), Jafar, Shahrour, and Juran (2010) and Nur Ozkan-Gunay and Ozkan (2007).

The ANN model was inspired by the biological nervous system process. It does this in practice with the use of artificial neurons that then combine input data on a number of these artificial neurons based on the weighted sums and biases of the neurons. The neurons will then each process the information that has been passed to it with non-linear decision functions to compute an output from that neuron. That output can either be passed on to another neuron or outputted as the final result of the model. The ability to explain non-linear relationships between variables in the dataset is one of the benefits of an ANN model over traditional regression models.

The weightings and biases form an integral part of an ANN model's learning process. The ANN model uses an iterative process to learn where it will adjust the weightings and biases as it collects and more data and information. The model does this by comparing the prediction and the actual target variable, which produces an error term. The weights are then adjusted iteratively in accordance with the error value by using a backward propagation technique to minimise the error.

The ANN model has been graphically represented in Figure 2 below:

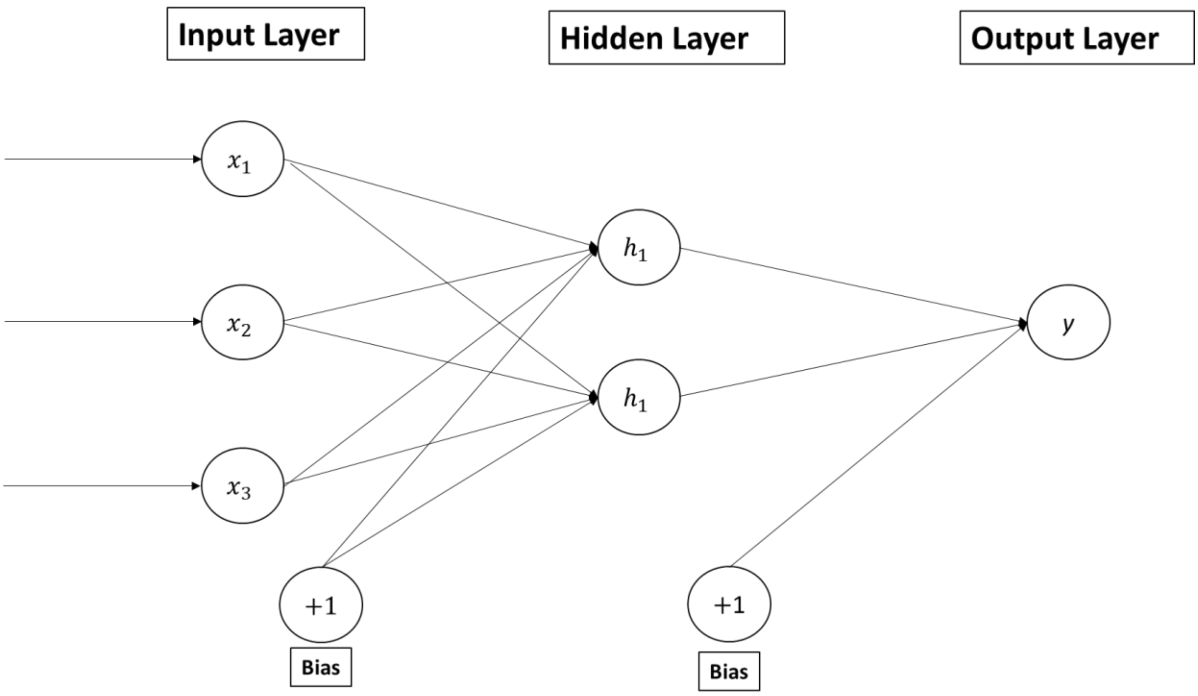


Figure 2: ANN Model Architecture

ANN models are sorted into layers of neurons. As a minimum ANN models will have an input and an output layer. The input layer is the first layer where the data is introduced, and the output layer is the layer that will output the model's prediction. Most ANN models will have several hidden layers between the input and output layers. Determining the number of hidden layers is an iterative process in ML and is discussed further below in Table 5.

Similar to the number of layers in a model the number of neurons per a layer is dependent on the type of data that is being passed into the model. This is discussed more below in the Table 5. The only exception to this is the output layer which must have the number of neurons that correlates with the number of predictions that the model should be making. In the case of this study that would mean that every output layer must have one neuron as we are only attempting to make 1 prediction, that being the respective volatility.

3.2.7.2 *Application of the ANN model in this study*

As discussed above, the ANN model allowed for the use of a combined dataset to provide the model with additional learning data in the hope that it would provide improved predictions. It was also possible to apply each separate country's data to the ANN model, which made it possible to make comparisons between the three ML models used in this study. This resulted in 6 models being built using the combined dataset approach, one for each of the following sets of data, 14-day, 21-day, and 28-day Volatility. All the models built have a version with and without the FEARS Index to allow for an analysis of the difference that investor sentiment makes on the ANN model's ability to make predictions.

As previously discussed hyperparameter tuning is a very important part of optimising ML models. The hyperparameters considered for the ANN model are presented below in Table 5.

Hyperparameter	Definition
Number of layers	The layers are the basic building blocks of a neural network; every neural network will have at least 3 layers. Namely the Input Layer, the Hidden Layer, and the Output Layer.
Number of neurons	A layer will have several neurons in it, the neurons operate similarly to a biological neuron; in that they receive information make small adjustments to the information, and then send the transformed information on.
Number of epochs	Epochs are the number of iterations that the training data will have through the model. i.e., a model with 50 epochs will have the training data pass through the model 50 times.
Model Optimiser	The model optimiser is a function that adjusts the models' attributes, to reduce the overall

	loss and improve the accuracy of the model. This is done by adjusting the weights and learning rates of the model. ML practitioners and researchers largely favour the use of the Adam optimiser as such all the models built in this study will use an Adam optimiser.
Activation function	The choice of an activation function defines how the inputted data into a specific layer will be transformed into an output. Namely, there are 3 types of commonly used activation functions that were considered when building the models: Tan-hyperbolic (Tanh) – this transforms the outputs to a number between -1 and 1, with more positive numbers being closer to 1 and more negative numbers being closer to -1. Sigmoid – also called the logistic function as it is the same function used in traditional logistic regression, it transforms the output data to a number between 0 and 1 with any negative number becoming 0. Rectified Linear Unit (ReLu) - similar to the sigmoid function the ReLu function will output data to a number between 0 and 1, with negative numbers becoming 0. However, it differs in the sense that it assumes a linear relationship for the data above 0 whereas the sigmoid function does not assume a linear relationship between the data.

Batch size	The batch size is the number of data points that are fed into the network. For example, if a batch size of 8 is chosen the model will take the first 8 data points, in the case of this study Jan 2004 – Aug 2004, and use that to train the model, it will then take the next 8 data points, Sep 2004 – Apr 2005, and use that to train the model. Smaller batch sizes are generally considered better to train a model but there is a trade-off between batch size and computational expense, with smaller batches requiring higher computational power.

Table 5: ANN Hyperparameters

3.3.8 Long Short-Term Memory (LSTM)

3.3.8.1 The model

The LSTM model has been used extensively in academic research in general as well as the finance research as such the methodology here will follow align closely with the methodology done by Jing, Wu and Wang (2021), Yang and Wang (2022), Moghar and Hamiche (2020), and Kumar, Goomer and Singh (2018).

An LSTM model is a type of Recurrent Neural Network, that uses a unique structure of forget and memory gates (cell) that are added on to the traditional recurrent neural network. The application of the memory cells allows the model to identify relationships that exist within time series data while ignoring relationships that are not important.

LSTM models employ a similar structure to that of the ANN models discussed above. Both ML models employ a multilayer multi-neuron structure. This allows the models to learn linear and non-linear relationships between the feature and target variables. However, the LSTM model employs a unique gate structure as was discussed previously. The input gate is the first layer of the LSTM model where the features for the current period are inputted into the model, and the output gate is the final layer where the model outputs the prediction, which in this study would be Realised Volatility for a respective time period for each respective country.

LSTM models have become particularly popular when making time series predictions, due to the use of the forget and memory gates. The memory gates allow for information from previous periods to be to be included in current period feature variables. The forget gate then limits the number of periods that are being carried into the current prediction period so as to avoid information overload within the model that will cause the model to create less accurate predictions due to overfitting. This function also makes LSTM models good at remembering relevant information from prior periods for a long period of time.

The LSTM model can be defined formally and visually as follows, where i_t^j is the input gate f_t^j is the forget gate, o_t^j is the output gate, c_t^j is the mememoryate and $c_t^{\sim j}$ is the input gate that is used to filter new information. All of which represent their respective gate for the j-th feature variable at time t.

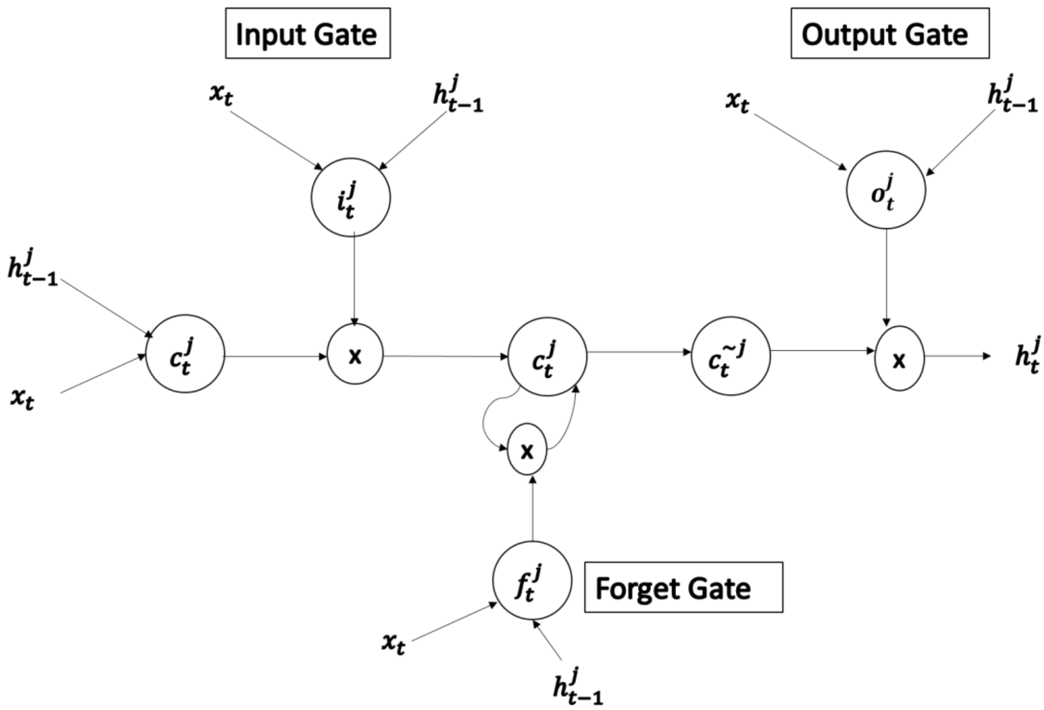


Figure 3: LSTM Architecture

$$i_t^j = \sigma(U_i x_t + w_i h_{t-1} + b_i)^j$$

$$f_t^j = \sigma(U_f x_t + w_f h_{t-1} + b_f)^j$$

$$o_t^j = \sigma(U_o x_t + w_o h_{t-1} + b_o)^j$$

$$c_t^j = f_t^j c_{t-1}^j + f_t^j c_t^{\sim j}$$

$$c_t^{\sim j} = \tanh\,(U_c x_t + w_c h_{t-1} + b_c)^j$$

$$h_t^j = o_t \tanh(c_t^j)$$

Where, U_i, U_f, U_o, U_c, w_i, w_f, w_o, and w_c are weight matrices in each neuron. b_i, b_f, b_o and b_c are bias terms ,a nd σ is a sigmoid function.

In the initial implementation of an LSTM model, the features for the first period, being January 2004 for this study, will be inputted into the input gate, the formula above will then be applied to the features. The transformed information will then be moved to the output gate where additional transformations are made, and the final prediction will then be outputted. The relevant information from the first input will be stored in the memory cell. In subsequent implementations the forget gate will control the amount of information from the previous periods, the new information for the current period is brought into the model through the input gate. The model will then use both the new information as well as the information stored in the memory cell to make predictions. The model will then eliminate the information from the previous periods and the current period that it considers unimportant for future predictions. The pertinent information will then be stored in the memory cell for use in the prediction for future time periods.

One of the weaknesses of the LSTM models ignore information from future periods which could have predictive power; thus, the Bidirectional LSTM was developed. The Bidirectional LSTM overcomes this shortcoming of the traditional LSTM model by using two separate hidden layers in conjunction that run the series data in both the forward time direction, i.e., in chronological order, as well as in reverse, i.e., reverse chronological order, the memory gate will then store information from the prior period as well as future periods to then make predictions.

The bidirectional LSTM can be represented visually as well as formally as follows:

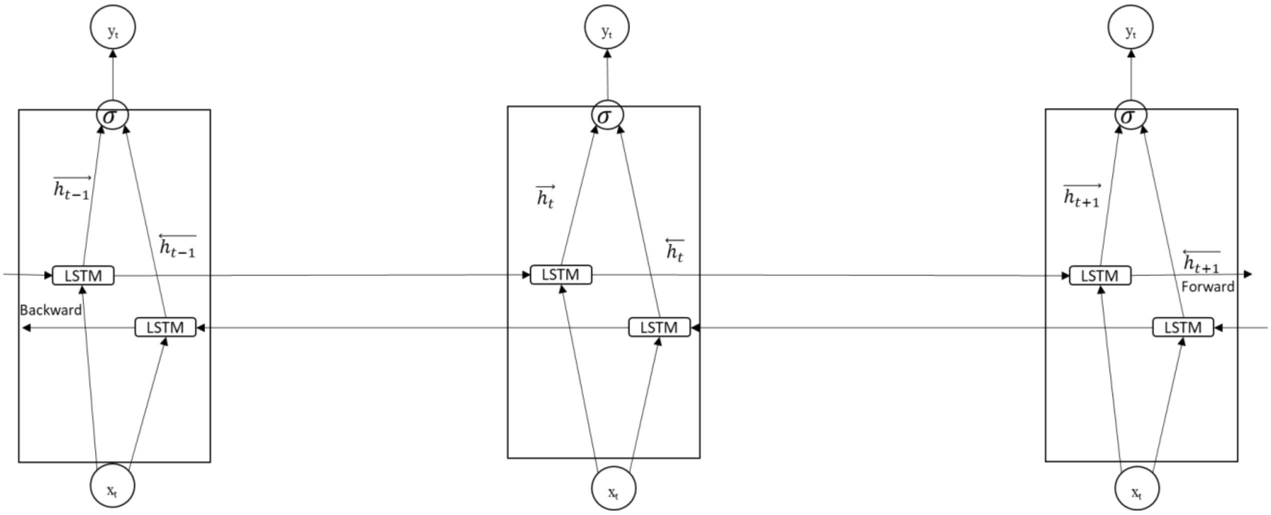


Figure 4: Bidirectional LSTM Architecture

$$\overrightarrow{h_t} = \sigma(W_{x\overrightarrow{h}}x_t + W_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{h}_{t-1} + b_{\overrightarrow{h}})$$

$$\overleftarrow{h_t} = \sigma(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}})$$

$$H_t = W_{x\overrightarrow{h}}\overrightarrow{h} + W_{\overleftarrow{h}y}\overrightarrow{h} + b_y$$

Where, H_t is the hidden layer output, $\overrightarrow{h_t}$ is the output generated from the forward structure, $\overleftarrow{h_t}$ is the output generated from the backward structure and σ is a sigmoid function.

3.3.8.2 Application of the LSTM model in this study

Due to the memory gate in the LSTM model, it was not possible to apply the combined dataset used in the Random Forest as well as the ANN to the LSTM model. As such, the LSTM model was applied on an individual country basis to all G7 countries used in this study. For each of the countries, the 14-day, 21-day, and 28-day realised volatility were predicted with and without the FEARS Index. This made it possible to assess the effectiveness of the FEARS Index in making predictions.

The following hyperparameters need to be considered when building an LSTM model. The hyperparameters in the LSTM model overlap extensively with those discussed above with the ANN model. As such only the new hyperparameters have been given detailed discussion below as the others have already been discussed above.

The following hyperparameters were considered in tuning the LSTM model:

Hyperparameter	Definition
Number of layers	Discussed above in the Application of the ANN model in this study section.
Number of neurons	Discussed above in the Application of the ANN model in this study section.
Number of epochs	Discussed above in the Application of the ANN model in this study section.
Model Optimiser	Discussed above in the Application of the ANN model in this study section.
Activation Function	Discussed above in the Application of the ANN model in this study section.
Dropout rate	Dropout rates are used during the training of the model, it will randomly drop a neuron from the next layer of the network based on the probability assigned to the dropout function.
Batch size	Discussed above in the Application of the ANN model in this study section.
Bidirectional or unidirectional LSTM	As discussed above, the bidirectional LSTM will model both forwards and backwards through the data whereas the unidirectional will only move forwards.

Table 6: LSTM Hyperparameters

3.3.9 Error Metrics

In order to evaluate the respective model's performance in making predictions four common metrics in prediction problems have been employed. The first two are the Mean Squared Error (MSE) and the Mean Absolute Error (MAE). Due to the heteroskedasticity found in financial data two additional error measures are being used to assess the predictions considering the heteroskedastic namely the Heteroskedastic Mean Squared Error (HMSE) and the Heteroskedastic Mean Absolute Error (HMAE).

Finance data inherently has conditional heteroskedasticity, meaning that the standard deviation of the returns is different in each period, as evidenced by the graph below in Figures 5 to 11 showing the daily returns of each country in this study over time.

The error metrics can be formally defined as follows:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i|$$

$$the\ HMSE = \frac{1}{m}\sum_{i=1}^{m}[1 - \frac{y_i}{\hat{y}_i}]^2$$

$$HMAE = \frac{1}{m}\sum_{i=1}^{m}\left|1 - \frac{y_i}{\hat{y}_i}\right|$$

Where, m is the respective Realised Volatility time horizon, y_i is the prediction made by the ML model and $\hat{y}_i$ is the actual Realised Volatility for that time period.

3.3.10 Feature Importance

Another tool that will be used to analyse the models that were used in this study is feature importance. Currently, there is no way to assess the feature importance of an LSTM model. However, there are tools available to assess the feature importance of both ANN and Random Forest models.

Feature Importance has provided ML practitioners and researchers with the ability to gain additional insight into what relationships the ML model is identifying. This is an important advancement in ML as for a long time ML models were seen as a "black box" meaning it was not possible to analyse the model.

For the Random Forest model, the Feature Importance tool within Sci-Kit Learn was used to assess the feature that the model believes has the largest predictive power in predicting volatility. The output of this model is presented as a percentage of importance, for example, if the output has a feature with 20% the feature has 20% importance in producing the predictions.

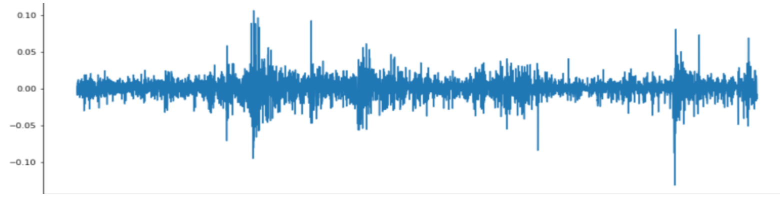


Figure 5: France Daily Returns over time

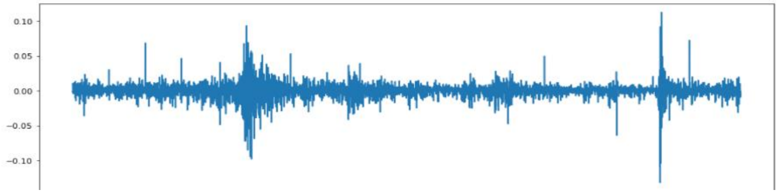


Figure 6:Canada Daily Returns over time

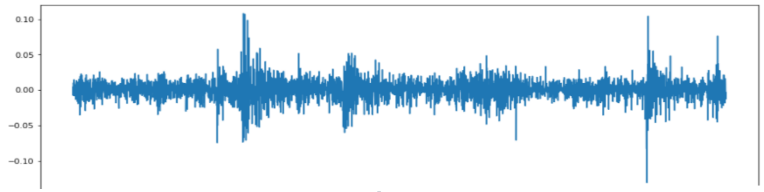


Figure 7: German Daily Returns over time

Figure 9: Italy Daily Returns over time

Figure 10: UK Daily Returns over time

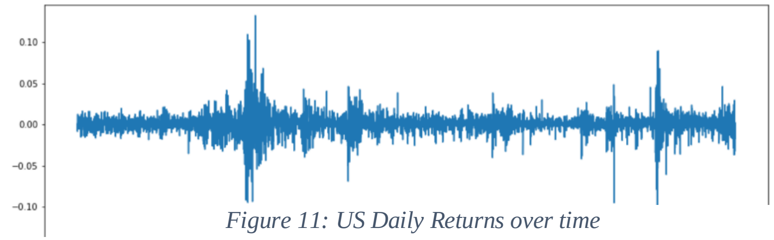


Figure 11: US Daily Returns over time

Figure 8: Japan Daily Returns over time

The other tool to be used will be applied to the ANN Model, the tool is the Shapley Additive Explanations (SHAP) this model uses a game theory approach to the output of ML models (Lipovetsky & Conklin, 2001). The output generated by the SHAP model shows the magnitude that the feature has on the prediction that is being made. Using this we can see relative to the other features in the model, which feature has the largest magnitude on the predictions that are being made.

These tools allow for additional insight into how the model is making predictions and where the model is placing the most significance when making predictions. This allows users to further analyse the features that are in the model. It allows the user to determine if there are any features that are not providing any predictive power and can thus be removed from the feature variables to save on computing power. The feature importance tools also make it possible to compare the features that have predictive power to previous studies that have used simple regression to determine which variables have explanatory power.

4. Individual Country Results

In this section we will discuss the hyperparameter tuning of the models that were fed the individual country datasets, this being the models that have been built using the data from only one of the countries rather than a combined dataset. In this section, the results of the LSTM, ANN, and Random Forest model will be presented for the three-target variable, that being the 14-day, 21-day, and 28-day Realised Volatility.

Each country's error metrics have been presented for the respective forecasting horizon for all of the different models used. Each model has a variant that includes the FEARS Index and one that excludes the FEARS Index to determine whether the FEARS Index provides the ML model with greater predictive ability. To determine whether the FEARS Index provides additional predictive ability to the models the error metrics discussed in Chapter 3 will be used. We will also be able to determine which of the three models provides the most accurate prediction by comparing these error metrics.

This presentation also allows for comparison between the LSTM, Random Forest, and ANN to determine which model provides better volatility predictions. This comparison will be done by looking at the error metrics of each of the models and comparing the model with and without the FEARS Index.

4.1 Hyperparameter Tuning

It is important to highlight upfront that there is currently no established method to determine how to best tune the hyperparameters in a ML model, with most practitioners and researchers using a trial-and-error approach to tune hyperparameters. Generally, this involves adjusting the hyperparameters of the model and assessing whether those changes improved the model examining the error metrics and loss functions as a guide as to whether the model had improved performance, i.e., if the error metrics decreased. This trial-and-error approach was adopted in this study. We examined across a range of parameters to determine where the minimum error terms may be.

Tables 7 and 8 show some of the tools that were used to tune the hyperparameters for the ANN and LSTM models:

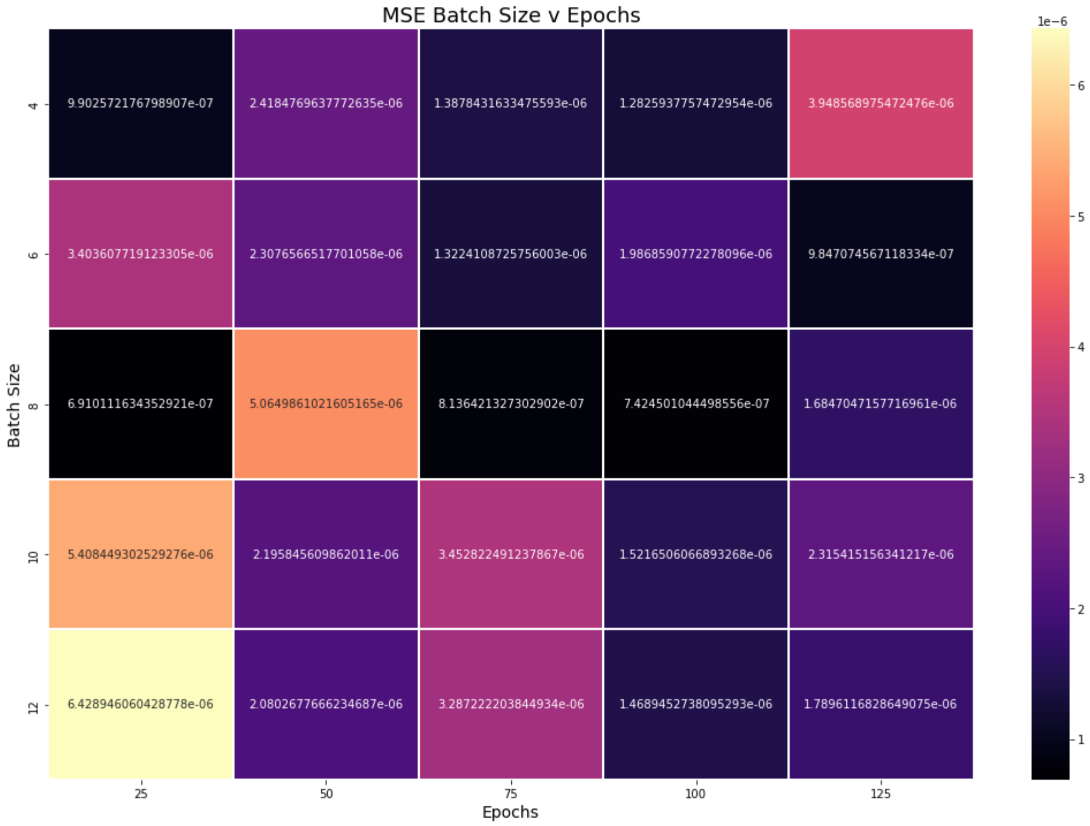


Table 7: Germany 14-day Realised Volatility Matrix Batch Size and Epochs

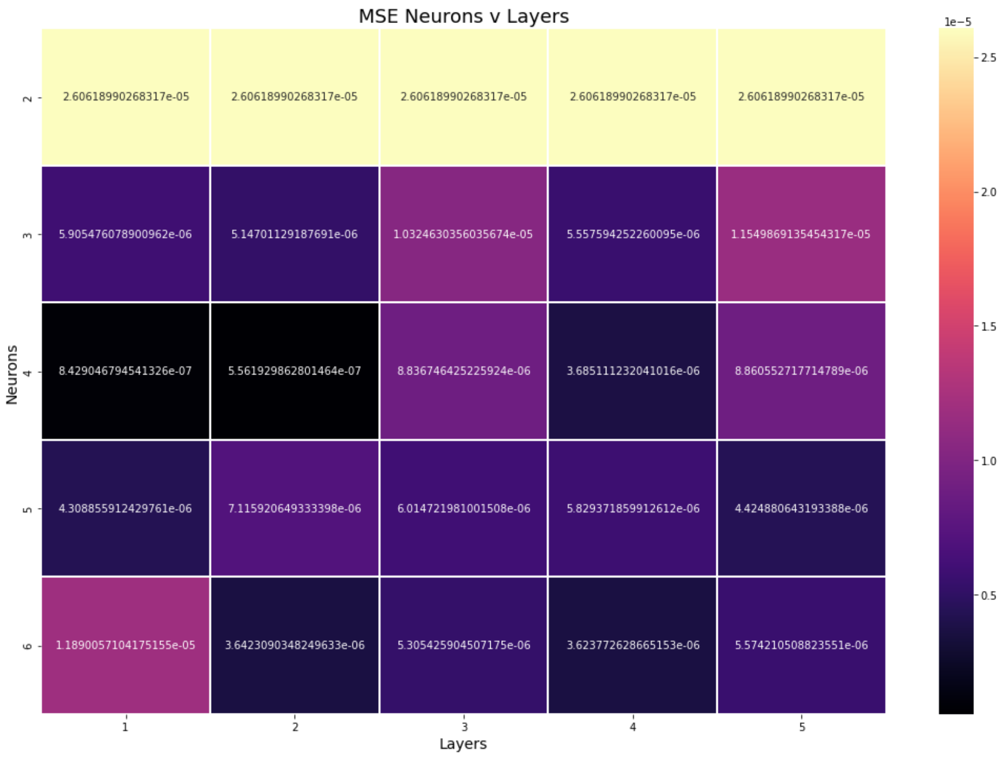


Table 8: Germany 14-day Realised Volatility Matrix Neurons and Layers

As part of the hyperparameter tuning process, and in an effort to reduce the time required for tuning, a for loop was created where the key variables were changed, and a matrix was created that displayed the MSE for each iteration of the model. This process was done, with 5 iterations per hyperparameter, with both the number of neurons and the number of layers, as well as the number of epochs and the batch size. These matrices were then used to optimise each model as they provided a range in which each respective hyperparameter would be optimised, small adjustments may then have been made to find the truly optimal position.

In Tables 7 and 8 above the German 14-Day realised volatility model with the FEARS Index matrices have been presented. The output of Table 7 was produced using the hyperparameter results from Table 8 to ensure that the Batch Size and Epochs were optimised for the optimal number of Neurons and Layers in the model. From Table 7 we can see the approximate optimal point for our Batch Size is 8 and the approximate optimal number of epochs is 25. From Table 8 we can see that the approximate optimal number of neurons is 4 with the approximate optimal number of layers being 2.

This implies that the model should consist of 4 total layers, an input layer, two hidden layers, and finally an output layer. The input layer and two hidden layers should then consist of 4 neurons each and finally, the output layer should have 1 neuron as that is the number of predictions being made. The Matrices for the Canadian 21-Day Realised Volatility model excluding the FEARS Index and the French 28-Day Realised Volatility model including the FEARS Index have been presented in Appendix 1.

Another tool used to determine whether the hyperparameters had been optimised was training and validation loss plots. When training the model, the model will calculate the loss function for each epoch for both the training and the validation set that is fed into the model. This can then be used to assess how well the model has been trained. If the training loss is decreasing, we know that the model is improving as the number of epochs increases. This allows us to optimise the number of epochs as it provides a visual indication of what number of epochs no longer provides a decrease in the error term, as such it provides an indication that the number of epochs should be at the point where the training error term is no longer decreasing.

In Figure 12 below, the loss function for the ANN 14-Day Realised volatility model, excluding the FEARS Index, for the US has been presented. Due to the limited amount of data available in the individual country datasets the validation loss function generally did not converge as closely with the training loss functions as those of the combined dataset, however, this does not indicate that the model will make poor predictions.

This is a tool that is used as an indicator however perfect convergence does not always result in the lowest error term. As such for the individual country datasets the plots were used as a determinant of whether the models were under or overfitting the data. To do this the validation loss plot was deemed to be acceptable when it flattened out and was not decreasing when the number of epochs ended and when the validation plot did not start rising, as this would be an indication of overfitting in the model.

From Figure 12 we can see that the validation plot has flattened out prior to the number of epochs ending and does not rise in the plot presented. As such it appears that the models were neither under nor overfitting the data provided.

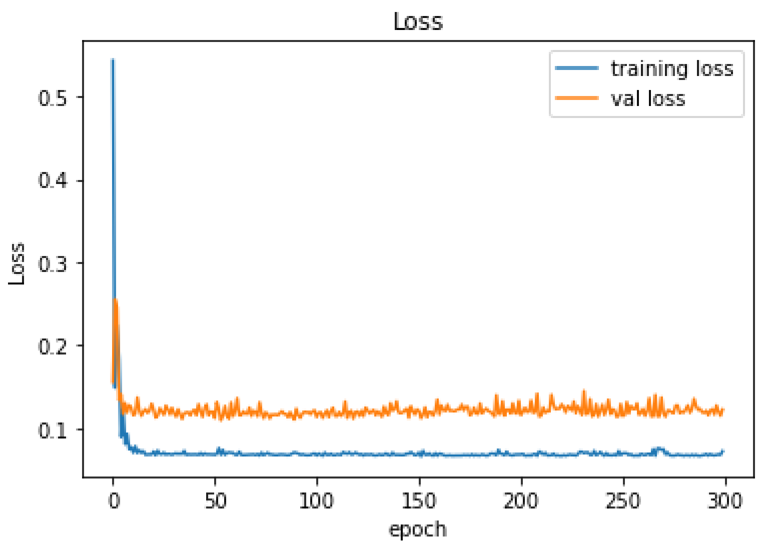


Figure 12: ANN US 14-Day Realised Volatility model excluding FEARS Loss function plot

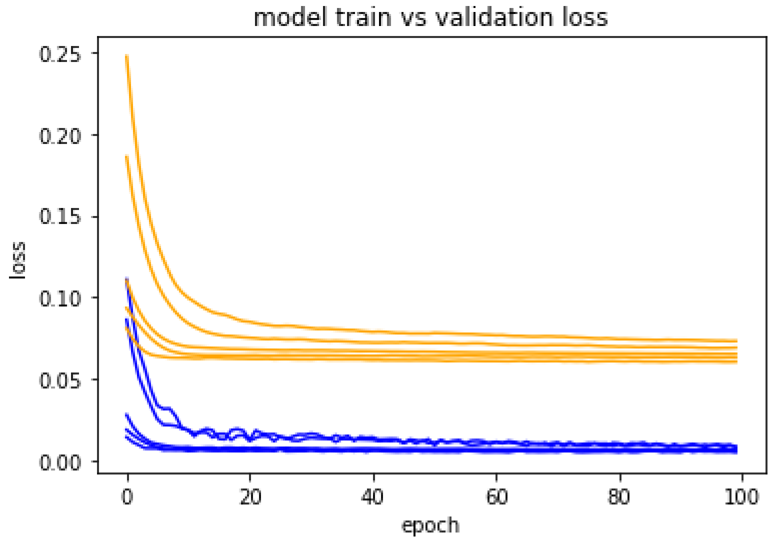


Figure 13: LSTM US 14-Day Realised Volatility model excluding FEARS Loss function plot

The training loss and validation loss plots for the LSTM models consist of multiple training and validation loss functions on the same set of axes, as seen in Figure 13 above. The reason for this is that LSTM models are stochastic, meaning they will make slightly different predictions each time that they are run. Therefore, to optimise across a variety of predictions that are made multiple training and validation loss plots have been plotted to ensure that there were no cases of under or overfitting. To create the plot presented in Figure 13, a for loop was created for 5 iterations of the model. Each of which is represented by a loss and validation plot, this allows for each of the iterations to be assessed as well as comparison across the 5 total iterations.

From Figure 13, it can again be seen that the validation plot did not fully converge on the training plot presented. This is again likely due to the limited number of observations presented in the individual country datasets. The plot was used to ensure that there was no over or underfitting. From Figure 13, we can see that the model does not appear to be over or underfitting the data that has been presented.

4.2 Results

In Table 9, the descriptive statistics for the US dataset have been presented, this shows us that the mean of the three realised volatilities for the US fall between 0.0113 and 0.0117 with the largest standard deviation being the standard deviation of the 14-day volatility, which aligns what was seen in the combined dataset.

	Growth IP	Inflation	Rho	TS	FEARS	14	21	28
count	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000
mean	0.043996	0.023721	-0.976441	1.593160	132394.211628	0.009617	0.009782	0.009736
std	1.273564	0.485534	1.756119	1.109016	69185.990970	0.007613	0.007244	0.006404
min	-13.245900	-2.600000	-6.954400	-1.601900	55796.000000	0.001562	0.003051	0.002690
25%	-0.288300	-0.200000	-1.765050	0.833200	93198.000000	0.005411	0.005613	0.005965
50%	0.133900	0.000000	-0.989500	1.605200	107587.000000	0.007572	0.007904	0.007880
75%	0.541050	0.300000	0.000000	2.431250	125130.500000	0.011194	0.011371	0.011258
max	5.318800	2.000000	3.873000	3.790200	343802.000000	0.062230	0.058659	0.048492

Table 9: Descriptive Statistics for US dataset

The descriptive statistics for the other 6 countries have been presented in Appendix 2, the mean of each of the countries differs slightly based on the volatility of that specific country however the standard deviations of all 6 countries are similar in that the highest standard deviation is found in the 14 day Realised Volatility.

The error metrics for all seven of the countries for the ANN, Random Forest, and LSTM models have been presented in Tables 10 to 16 below. It is important to highlight that the LSTM and ANN error metrics are an average of 5 training iterations of the model given the stochastic nature of the model whereas the Random Forest model error metrics are not averages but the results of a single training iteration as the Random Forest model is not stochastic and thus will produce the same predictions for every training iteration. It is also important to note that the models were created with MSE as the loss functions, therefore the models have been created to optimise the MSE.

The US results stand out most with FEARS providing additional predictive power across all 3 forecasting horizons for the ANN model, and across 14- and 21-day Realised Volatility for the LSTM model. This shows that the effects of investor sentiment, specifically in this case Search Volumes from Google Trends as a proxy for sentiment, are the strongest in the US. This is likely due to the amount of internet usage and use of Google specifically in the US providing a large portion of the population and allowing for Google SVI to provide a greater representation of household sentiment.

The German Error Metrics, Table 14, show that the addition of the FEARS Index as a feature variable does not provide any additional predictive ability across all three forecasting horizons when looking at the ANN and LSTM models. This aligns with the work done by Tetlock et al.

(2012) and Solanki and Seetharam (2018) who found that sentiment did not have statistically significant explanatory power in the German market.

The other five markets present in the study show differing results across the different forecasting horizons. The forecasting horizon where the FEARS Index provides the most predictive power is the 21-day Realised volatility. Three of the remaining five countries, namely the UK, Canada, and Italy, show improved predictions for the 21-Day Realised Volatility when the FEARS Index is added. This shows that Google Trends provide forward-looking explanatory power for volatility predictions in these markets of around 21 Days. The remaining two markets, Japan, Table 15, and France, Table 16, show that 14-Day and 28-Day realised volatility predictions are improved by the inclusion of the FEARS Index respectively. This shows us that the Japanese market is more reactive to sentiment with it pricing sentiment out of the market within a shorter time period, whereas the French markets are slower to react to market sentiment in the economy, specifically household sentiment.

The Canadian, Table 12, and French, Table 16, results show that the Random Forest model to be the best predictor of volatility in these markets. Whereas the UK and Italian, Table 13, markets have a mixture of models, with one of the forecasting horizons in each being best predicted by the ANN model. In the UK market that forecasting horizon is 21 Day and in the Italian market it was the 28-day forecasting horizon.

From these results, we can see that the LSTM model did not provide the best error metrics in any of the regions in this study. This is surprising given the LSTM model was specifically designed to make predictions for time series data, such as stock market predictions. However, it does provide an insight into the markets being studied, all of the markets in the study are from developed countries that have fairly sophisticated market participants which should improve the efficiency of the markets.

The LSTM model structure using that makes use of prior period results, through the use of the memory gate, assumes that the current volatility includes all of the information from prior periods, therefore assuming an efficient market. However, the results in this study seem to indicate that the markets under observation are not wholly efficient which aligns with the work done by Erdem and Ulucak (2016) who found that G7 country's markets are only show weak and semi-strong market efficiency. This is not to say that the LSTM model would not

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN US					
MSE	3.71E-04 *	4.45E-04	1.85E-04 *	1.86E-03	1.91E-04 *	7.29E-03
MAE	1.93E-02 *	2.11E-02	1.36E-02 *	4.31E-02	1.38E-02 *	8.54E-02
HMSE	1.25E-02 *	1.84E-01	1.62E-01 *	4.24E+00	9.49E-02 *	5.65E-01
HMAE	1.12E-01 *	4.29E-01	4.02E-01 *	2.06E+00	3.08E-01 *	7.51E-01
	Random Forest US					
MSE	1.94E-04	7.63E-05 *	2.07E-04	8.32E-05 *	1.12E-04	1.04E-04 *
MAE	1.39E-02	8.73E-03 *	1.44E-02	9.12E-03 *	1.06E-02	1.02E-02 *
HMSE	3.49E-01	2.49E-01 *	3.55E-01	2.66E-01 *	3.16E-01	2.95E-01 *
HMAE	5.91E-01	4.99E-01 *	5.96E-01	5.15E-01 *	5.62E-01	5.43E-01 *
	LSTM US					
MSE	4.17E-04 *	1.49E-03	1.16E-03	6.89E-04 *	1.67E-03 *	3.16E-03
MAE	2.92E-02 *	3.86E-02	3.41E-02	2.62E-02 *	4.08E-02 *	5.62E-02
HMSE	2.04E-02 *	7.84E-02	7.23E-02 *	4.54E-01	6.74E-02 *	1.04E-01
HMAE	1.71E-01 *	2.80E-01	2.69E-01 *	6.74E-01	2.60E-01 *	3.23E-01

Table 10: US Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN UK					
MSE	3.96E-05	3.66E-05 *	4.11E-06 *	2.37E-05	1.25E-04	3.47E-05 *
MAE	6.29E-03	6.05E-03 *	2.03E-03 *	4.87E-03	1.12E-02	5.89E-03 *
HMSE	2.38E-03 *	9.41E-03	8.41E-04 *	2.48E-03	1.81E-02	6.74E-03 *
HMAE	4.87E-02 *	9.70E-02	2.90E-02 *	4.98E-02	1.35E-01	8.21E-02 *
	Random Forest UK					
MSE	3.65E-04	6.22E-05 *	2.84E-04	5.65E-05 *	4.88E-05 *	5.70E-05
MAE	1.91E-02	7.88E-03 *	1.68E-02	7.52E-03 *	6.99E-03 *	7.55E-03
HMSE	5.31E-01	1.50E-01 *	4.79E-01	1.82E-01 *	1.24E-01 *	1.67E-01
HMAE	7.29E-01	3.87E-01 *	6.92E-01	4.27E-01 *	3.52E-01 *	4.09E-01
	LSTM UK					
MSE	1.59E-03	1.02E-03 *	1.99E-03	1.53E-03 *	3.01E-03	1.10E-03 *
MAE	3.98E-02	3.19E-02 *	4.47E-02	3.91E-02 *	5.49E-02	3.32E-02 *
HMSE	8.20E-02	6.37E-02 *	4.98E-01	4.03E-01 *	5.55E-01	2.81E-01 *
HMAE	2.86E-01	2.52E-01 *	3.49E-01	2.93E-01 *	3.61E-01	2.10E-01 *

Table 11: UK Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN Canada					
MSE	2.92E-05	2.87E-05 *	1.11E-05 *	2.63E-05	8.94E-05	1.93E-06 *
MAE	5.40E-03	5.36E-03 *	3.34E-03 *	5.12E-03	9.46E-03	1.39E-03 *
HMSE	7.48E-03	6.96E-03 *	1.20E-03 *	4.40E-03	1.94E-02	1.14E-03 *
HMAE	8.65E-02	8.35E-02 *	3.46E-02 *	6.64E-02	1.39E-01	3.38E-02 *
	Random Forest Canada					
MSE	7.58E-06	1.94E-06 *	7.41E-06	1.96E-06 *	8.63E-06	2.36E-06 *
MAE	2.75E-03	1.39E-03 *	2.72E-03	1.40E-03 *	2.94E-03	1.53E-03 *
HMSE	7.52E-02	3.53E-02 *	5.83E-02	3.50E-02 *	6.40E-02	3.96E-02 *
HMAE	2.74E-01	1.88E-01 *	2.41E-01	1.87E-01 *	2.53E-01	1.99E-01 *
	LSTM Canada					
MSE	3.10E-03	2.50E-03 *	9.45E-04 *	1.94E-03	2.39E-03	2.18E-03 *
MAE	5.57E-02	5.00E-02 *	3.07E-02 *	4.41E-02	4.89E-02	4.67E-02 *
HMSE	2.21E-01	1.98E-01 *	1.10E-01 *	1.72E-01	5.96E-01	5.87E-01 *
HMAE	4.70E-01	4.45E-01 *	3.32E-01 *	4.15E-01	4.16E-01	4.07E-01 *

Table 12: Canada Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN Italy					
MSE	8.88E-04	5.41E-04 *	1.81E-05 *	1.76E-04	2.14E-06 *	4.58E-05
MAE	2.98E-02	2.33E-02 *	4.26E-03 *	1.33E-02	1.46E-03 *	6.77E-03
HMSE	4.83E-02	3.36E-02 *	2.46E-03 *	1.10E-02	1.18E-04 *	2.64E-03
HMAE	2.20E-01	1.83E-01 *	4.96E-02 *	1.05E-01	1.09E-02 *	5.14E-02
	Random Forest Italy					
MSE	2.41E-05	1.05E-05 *	2.32E-05	1.22E-05 *	2.09E-05	1.41E-05 *
MAE	4.91E-03	3.24E-03 *	4.82E-03	3.49E-03 *	4.57E-03	3.76E-03 *
HMSE	4.94E-02 *	5.84E-02	4.92E-02 *	6.63E-02	4.48E-02 *	6.86E-02
HMAE	2.22E-01 *	2.42E-01	2.22E-01 *	2.57E-01	2.12E-01 *	2.62E-01
	LSTM Italy					
MSE	4.55E-03	1.68E-03 *	1.55E-03 *	3.94E-03	2.00E-03 *	7.88E-03
MAE	6.74E-02	4.10E-02 *	3.93E-02 *	6.28E-02	4.48E-02 *	8.88E-02
HMSE	1.58E-01	7.94E-02 *	7.28E-02 *	1.32E-01	6.78E-02 *	1.66E-01
HMAE	3.97E-01	2.82E-01 *	2.70E-01 *	3.63E-01	2.60E-01 *	4.07E-01

Table 13: Italy Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN Germany					
MSE	3.57E-03	6.01E-04 *	9.56E-04	7.21E-04 *	2.14E-04	1.63E-05 *
MAE	5.98E-02	2.45E-02 *	3.09E-02	2.68E-02 *	1.46E-02	4.04E-03 *
HMSE	1.10E-01	2.65E-02 *	4.68E-02	3.76E-02 *	8.53E-03	3.38E-04 *
HMAE	3.32E-01	1.63E-01 *	2.16E-01	1.94E-01 *	9.24E-02	1.84E-02 *
	Random Forest Germany					
MSE	1.82E-05 *	2.08E-05	2.32E-05	2.29E-05 *	2.07E-05 *	2.24E-05
MAE	4.27E-03 *	4.56E-03	4.81E-03	4.78E-03 *	4.55E-03 *	4.74E-03
HMSE	7.33E-02 *	7.98E-02	7.73E-02	7.31E-02 *	6.68E-02 *	8.02E-02
HMAE	2.71E-01 *	2.83E-01	2.78E-01	2.70E-01 *	2.58E-01 *	2.83E-01
	LSTM Germany					
MSE	3.45E-03	2.94E-03 *	3.16E-03	2.79E-03 *	3.43E-03	3.00E-03 *
MAE	5.88E-02	5.42E-02 *	5.62E-02	5.28E-02 *	5.85E-02	5.48E-02 *
HMSE	1.12E-01	1.00E-01 *	1.11E-01	1.02E-01 *	9.92E-02	9.03E-02 *
HMAE	3.35E-01	3.17E-01 *	3.33E-01	3.19E-01 *	3.15E-01	3.01E-01 *

Table 14: Germany Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN Japan					
MSE	1.89E-05 *	6.82E-04	5.33E-04	3.04E-06 *	5.28E-04	6.06E-05 *
MAE	4.35E-03 *	2.61E-02	2.31E-02	1.74E-03 *	2.30E-02	7.78E-03 *
HMSE	3.42E-03 *	9.87E-02	9.37E-02	1.99E-03 *	5.35E-02	1.58E-03 *
HMAE	5.85E-02 *	3.14E-01	3.06E-01	4.46E-02 *	2.31E-01	3.98E-02 *
	Random Forest Japan					
MSE	3.96E-06	2.94E-06 *	7.18E-06	2.43E-06 *	2.51E-05	3.44E-07 *
MAE	1.99E-03	1.71E-03 *	2.68E-03	1.56E-03 *	5.01E-03	5.87E-04 *
HMSE	2.37E-03	2.05E-04 *	8.73E-03	9.84E-04 *	7.53E-02	1.81E-04 *
HMAE	4.87E-02	1.43E-02 *	9.34E-02	3.14E-02 *	2.74E-01	1.35E-02 *
	LSTM Japan					
MSE	1.68E-05 *	9.10E-05	4.25E-04	3.46E-07 *	7.33E-04	1.94E-06 *
MAE	4.09E-03 *	9.54E-03	2.06E-02	5.88E-04 *	2.71E-02	1.39E-03 *
HMSE	8.27E-04 *	5.39E-03	4.80E-02	6.68E-05 *	2.70E-02	6.89E-05 *
HMAE	2.88E-02 *	7.34E-02	2.19E-01	8.18E-03 *	1.64E-01	8.30E-03 *

Table 15: Japan Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN France					
MSE	1.77E-03	3.97E-04 *	9.91E-04	4.87E-04 *	1.12E-04 *	1.27E-04
MAE	4.21E-02	1.99E-02 *	3.15E-02	2.21E-02 *	1.06E-02 *	1.13E-02
HMSE	6.31E-02	1.82E-02 *	4.80E-02	2.72E-02 *	7.12E-03 *	7.42E-03
HMAE	2.51E-01	1.35E-01 *	2.19E-01	1.65E-01 *	8.44E-02 *	8.61E-02
	Random Forest France					
MSE	5.03E-06	4.67E-06 *	1.20E-05	8.71E-06 *	9.79E-06	7.80E-06 *
MAE	2.24E-03	2.16E-03 *	3.47E-03	2.95E-03 *	3.13E-03	2.79E-03 *
HMSE	1.03E-02 *	2.93E-02	3.34E-02 *	5.33E-02	1.91E-02 *	3.91E-02
HMAE	1.02E-01 *	1.71E-01	1.83E-01 *	2.31E-01	1.38E-01 *	1.98E-01
	LSTM France					
MSE	1.33E-03 *	1.69E-03	3.58E-03	2.30E-03 *	2.25E-03 *	2.35E-03
MAE	3.65E-02 *	4.12E-02	5.98E-02	4.80E-02 *	4.74E-02 *	4.85E-02
HMSE	5.32E-02 *	6.59E-02	1.20E-01	8.83E-02 *	7.50E-02 *	7.70E-02
HMAE	2.31E-01 *	2.57E-01	3.46E-01	2.97E-01 *	2.74E-01 *	2.77E-01

Table 16: France Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

outperform the other models in different time periods, as the work done by Ito, Noda and Wada (2014) showed that market efficiency is time-dependent, as such the time period used in this study may be inefficient but other time periods may have greater market efficiency present.

In both the Italian and the French markets, it can be seen that the Random Forest Model was able to provide improved predictions with the FEARS Index when the heteroskedastic error metrics are considered. This shows that where the data is scaled to remove the heteroscedasticity present in Italy and France the FEARS Index improves the predictions. This may mean that these markets experienced higher levels of heteroscedasticity in the time period under observation than the other markets. It is only interesting to note that this relationship was only identified by the Random Forest model and not the other models in this study.

In Figures 14, 15, and 16 below the Feature Importance of the ANN models for the US, Japan, and UK have been presented for the 14-day Volatility models. Both the US and Japan showed that the inclusion of the FEARS Index improved predictions, as seen by the error terms, however in the UK the inclusion of the FEARS Index did not always result in better predictions.

It seems from the feature importance that in the models where the FEARS Index does not provide additional predictive ability the FEARS Index has been used as the most important feature in predictions, whereas the US and Japan have the FEARS Index as the second most important feature in making predictions. This shows that the FEARS Index as a proxy for investor sentiment is an important feature in making volatility predictions. However, it is not the most important feature when making predictions, and models that rank it as the most important feature tend to perform poorly when compared to the model that does not include the FEARS Index as a feature variable.

The feature importance for the UK, seen in Figure 16, is particularly interesting. The model with the FEARS Index did not perform very well, the model has placed the least importance on Growth in IP (IP). This weighting contradicts the findings of Solanki and Seetharam (2018), who they found that of the factors used in this study, only the FEARS Index and IP had statistically significant explanatory power. This shows that the model may be incorrectly identifying the relationships between the feature and target variables and further tuning of the hyperparameters of the models is required or a different model should be used to correctly identify the relationship.

This issue of overweighting the FEARS Index does not seem to be the case for the Random Forest model, the feature importance of the Random Forest model for the same dataset and

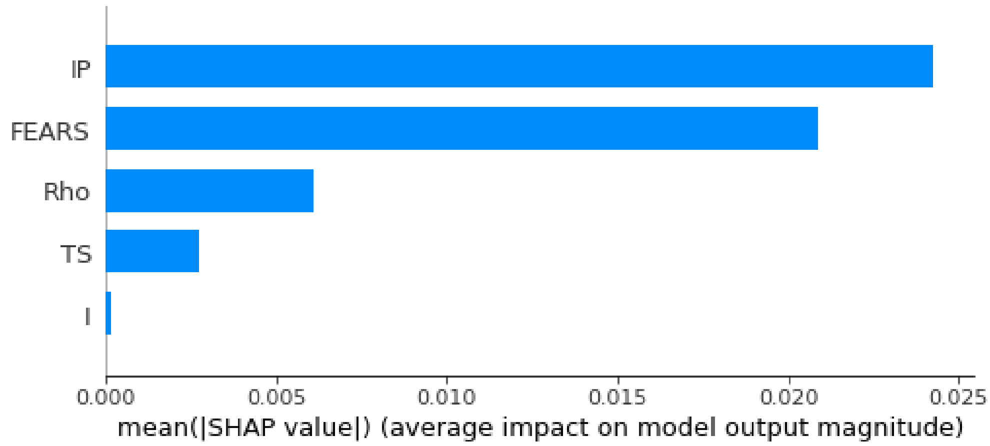


Figure 14: US ANN model 14-day Realised Volatility Feature Importance

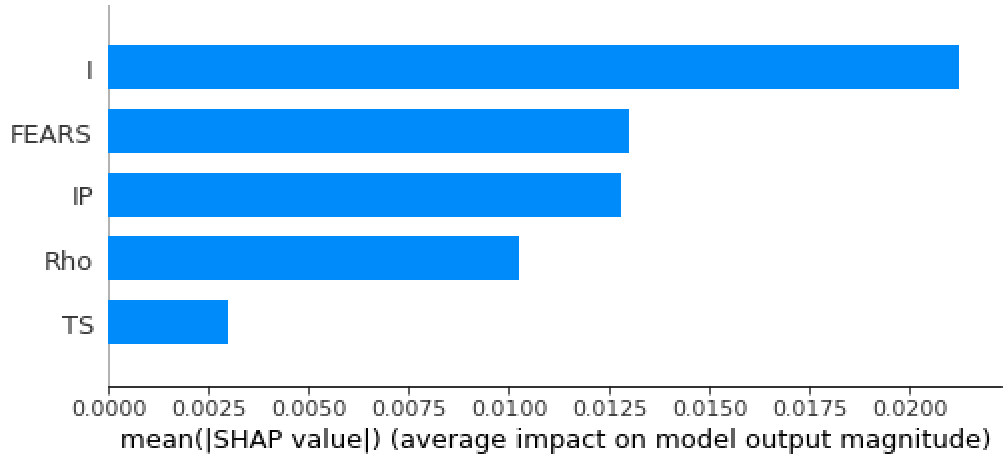


Figure 15: Japan ANN model 14-day Realised Volatility Feature Importance

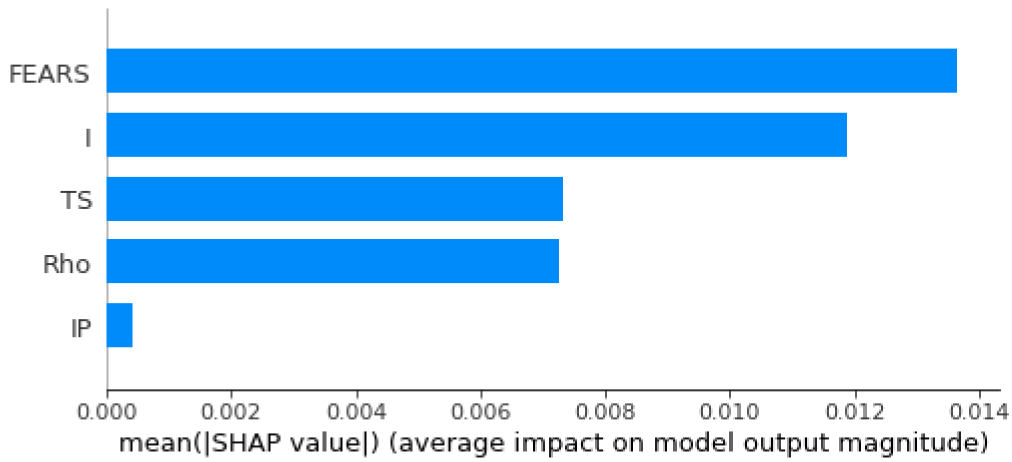


Figure 16: UK ANN model 14-day Realised Volatility Feature Importance

volatility period as presented for the ANN model has been presented below in Figures 17 to 20.

The Random Forest model does not follow a pattern in metrics of feature importance as was seen in the ANN models, it would appear that the Random Forest models do apply similar ranking to the features as the ANN model for example the feature importance for the US ANN model. Figure 17, matches that of the US Random Forest model in that IP is considered the most important, the FEARS Index being ranked as the second most important, and Rho ranked as the third most important. As discussed above the importance placed on IP differs from the Solanki and Seetharam (2018) findings, however, the reason for this difference could be two-fold, 1. This study was looking at predictive ability rather than explanatory power. 2. The time period under observation, the greater focus on America building their own IP in America and not in foreign countries in the 2016 – 2020 time period may have resulted in the change in the importance of the feature between this study and the Solanki and Seetharam (2018) study.

The Random Forest and ANN models then diverge and rank Inflation (I) and Term Structure of Interest Rates (TS) differently but they both consider them to be of low importance. Given the economic importance of these factors in practice as a discount rate in valuation techniques, this low weighting may be due to the fact that both the American and the Japanese markets during the period under observation had experienced persistent low and consistent short-term interest rates. The low amount of central bank intervention into the short end of their yield may mean that I and TS have not been fluctuating enough to assist in market predictions. This however may not be the case in the current high inflation period with the American market entering a tightening phase these factors may be relatively more important after 2020.

The extreme weightings that can be seen in the UK Random Forest model feature importance is further evidence to the error metrics that show that the Random Forest model did not perform well in predicting the 14-Day Realised Volatility for the region. This shows that it is likely that the Random Forest model was not able to correctly identify the relationships between the feature and target variables in the UK showing that it was not the correct model for the region.

The Random Forest model both placed high importance on Rho, ranking it as 3rd most important feature in the German market. Figure 20. This is likely due to the relative economic importance of short-term interest rates in the market as a monetary policy choice, this higher level of importance than that of the US or Japanese predictions is likely due to the single currency in the Eurozone. As the currency in Germany is centrally controlled Germany does

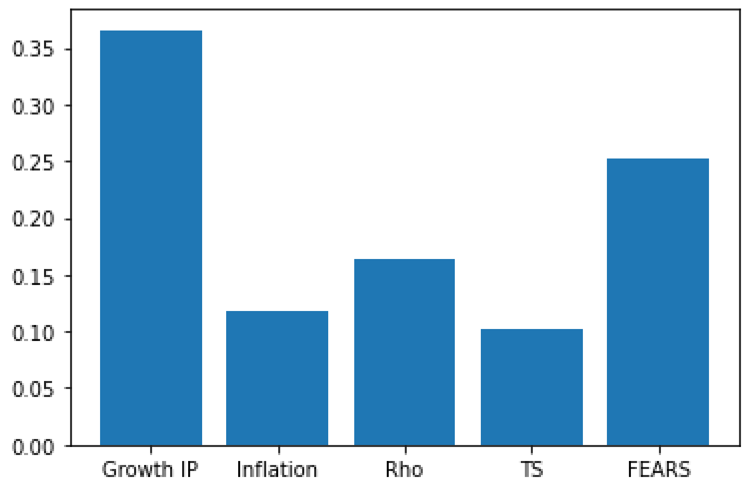


Figure 17: US Random Forest model 14-day Realised Volatility Feature Importance

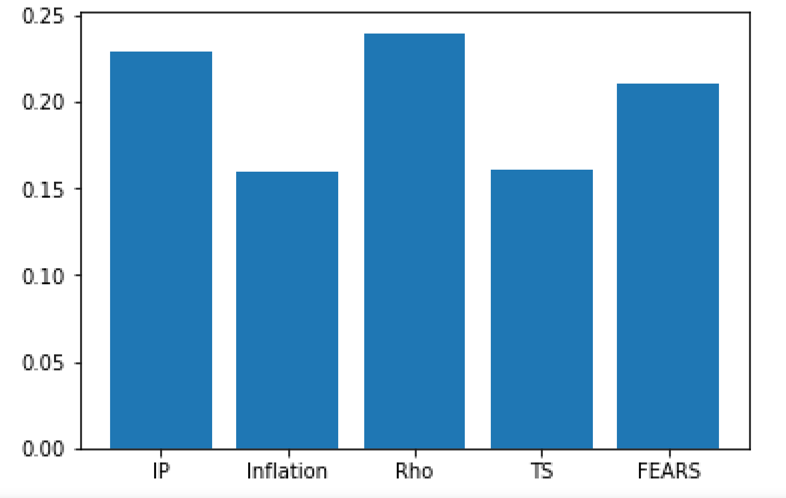


Figure 18: Japan Random Forest model 14-day Realised Volatility Feature Importance

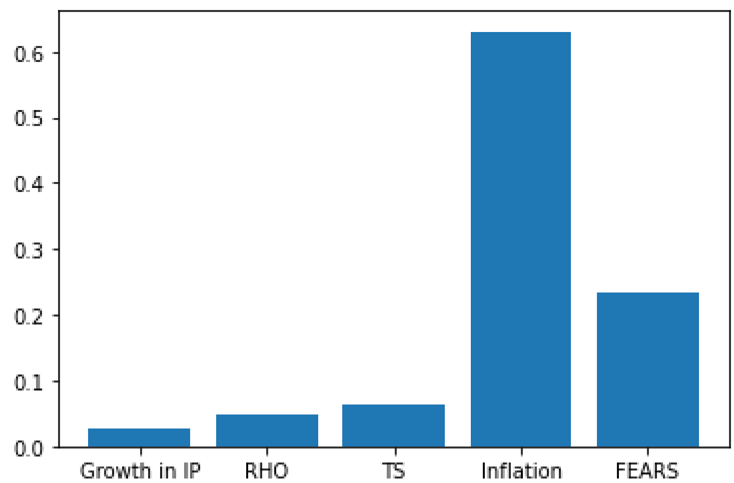


Figure 19: UK Random Forest model 14-day Realised Volatility Feature Importance

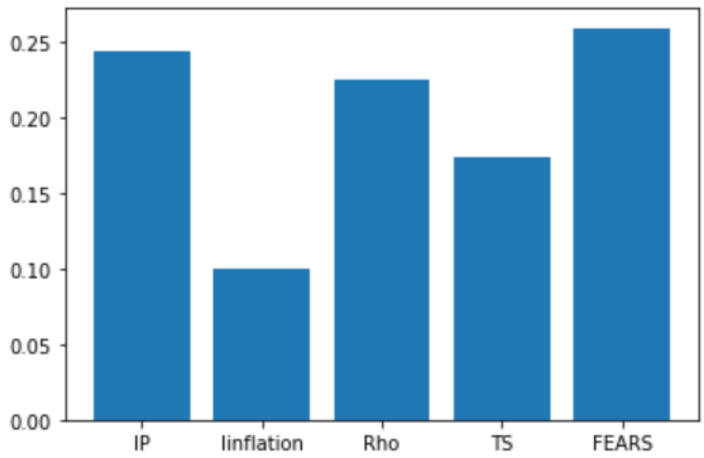


Figure 20: Germany Random Forest model 14-day Realised Volatility Feature Importance

not possess a fiscal policy tool to control inflation unless the entire Eurozone is experiencing high inflation, therefore increased importance is placed on the central banks' ability to control monetary policy through the short end of the yield curve.

We can therefore see from the feature importance graphs that the factors that the models have placed importance on when making predictions align with the economic understanding of the variables as well as the quality of the predictions made by the models. Models that were able to provide improved predictions placed more importance on factors that have economic importance within the region.

In this section the results from the individual country dataset have been presented, this has shown that the LSTM model was the worst at predicting volatility across all of the forecasting horizons as well as across all of the regions in this study. The other models provided mixed results with some predicting certain regions or forecasting horizons better than others. It was also seen that the inclusion of the FEARS Index did provide better predictive ability however it depended on the region as to what time horizon was best predicted by the inclusion of the FEARS Index. The inclusion of the FEARS Index only provided better predictive ability across all three forecasting horizons in the US in other markets it was shown that sentiment in the market affected different time horizons in different ways.

5. Combined Country Results

This section will follow the same structure as used above in Chapter 4. First, we will discuss the hyperparameter tuning of the combined dataset models. Secondly, the results of the ANN and Random Forest models with and without the FEARS Index as a feature variable will be discussed. This is done to assess whether the inclusion of the investor sentiment factor, the FEARS Index, provides the model with additional predictive power.

This presentation also allows for a comparison between the Random Forest and ANN to determine which model provides better volatility predictions. This comparison will be done by looking at the error metrics of each of the models and comparing the model with and without the FEARS Index.

5.1 Hyperparameter Tuning

As discussed above there currently is no established method to determine the optimal hyperparameters of a ML model. As such the trial-and-error method followed above will be followed here as well. The tools discussed above in Chapter 4 will again be used here to determine the approximate optimal point for the hyperparameter of the ANN model.

We examined across a range of parameters to determine where the minimum error terms may be. The following tools were used to tune the hyperparameters for the ANN models. As part of the hyperparameter tuning process, and in an effort to reduce the time required for tuning. A for loop was created where the key variables were changed, and a matrix was created that displayed the MSE of each iteration of the model. This process was done, with 5 iterations per a hyperparameter, with both the number of neurons and the number of layers, as well as the number of epochs and the batch size. These matrices, Tables 17 and 18, were then used to optimise each model as they provided an approximate range in which each respective hyperparameter would be optimised. Small adjustments may then have been made to find the truly optimal position.

Table 17 shows the table for batch size and epochs, from this we can see a batch size of 10 and 125 epochs was the approximate optimal point for this model.

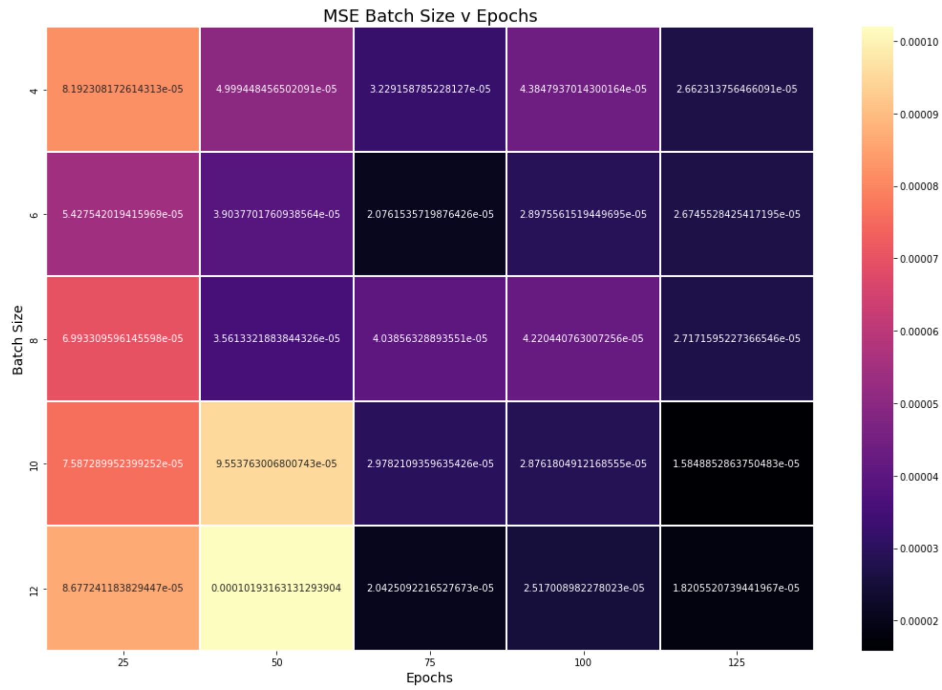


Table 17: Error Matrix Batch Size and Epochs

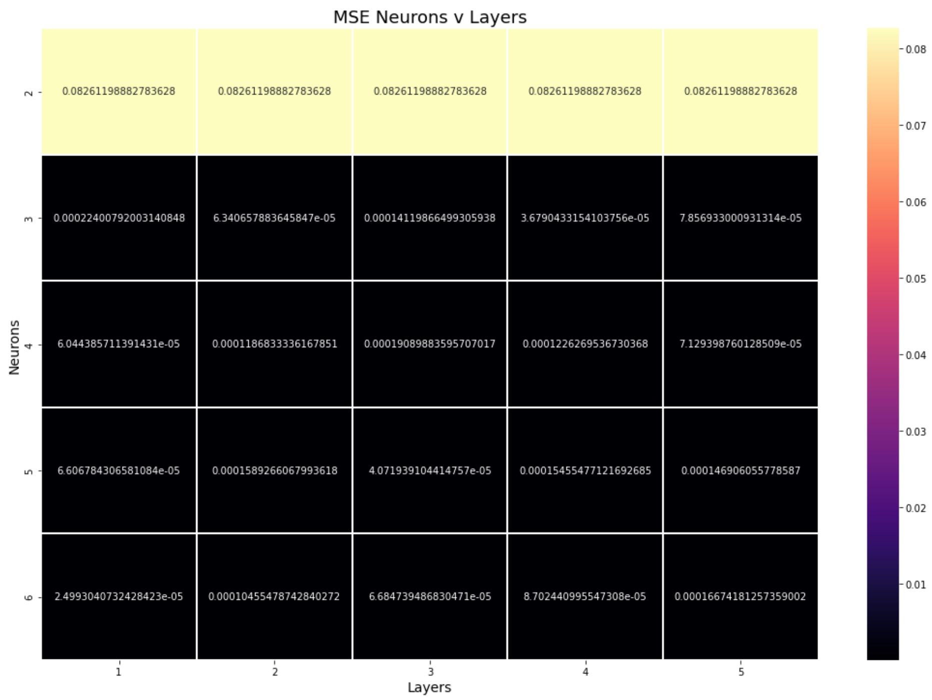


Table 18: Error Matrix Neurons and Layers

Table 18 then shows the matrix for a number of neurons and layers. From this, we can see the approximate optimal point is to have 6 neurons and 1 hidden layer in the model. This implies that the model should consist of an input layer of 6 neurons, a single hidden layer of 6 neurons, and finally, an output layer that should have only 1 layer. The hidden layer has only one neuron as this is the number of predictions that the model should be making, as discussed in Chapter 3 above.

Another tool that was used to hyperparameter tune the models was the training and validation loss functions, as discussed above in Chapter 4. The validation loss error metrics are then plotted on the same set of axes as the training loss error metrics, as seen in Figure 21. This allows us to assess the same information for the validation set. As the validation set is new data that the model is being exposed to it will show whether the model is good at predicting information from new data not just from the training data set. These plots allow us to see the optimal number of epochs, when analysing the validation plot the aim is to get the validation loss as close to the training loss. This will show us that the model is as good at making predictions with new data as it is with the training data.

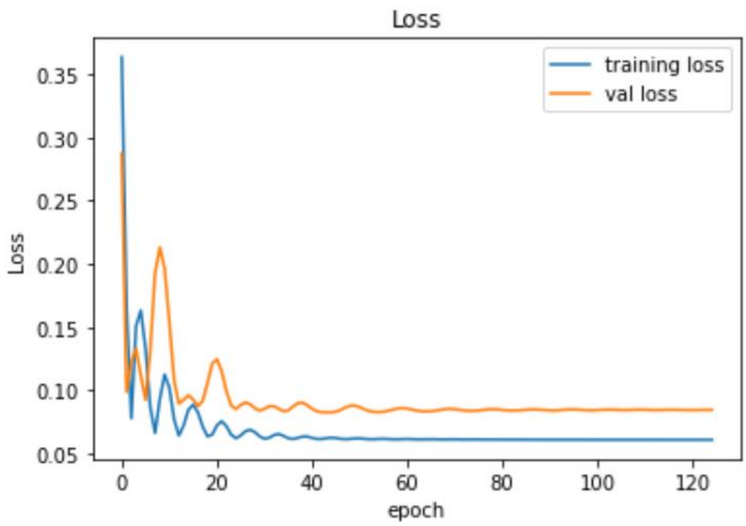


Figure 21: Loss function plot for 14 Day Realised Volatility with FEARS

Figure 21 above, shows us that the validation loss (orange plot) quickly converged with the training loss (blue plot) showing that the model is good at making predictions with new data as well as the training data and is not underfitting the data. The other piece of information that can be derived from Figure 21 is that the model is not overfitting the data. If the model were overfitting, we would see that the validation plot would converge with the training loss plot and then diverge away from the training loss plot again. Overfitting occurs when the model places too much importance on relatively unimportant variables which makes it very good at making predictions on the training data but not good at making predictions on the validation data. This creates the convergence and then the divergence of the validation and training loss plots discussed above.

5.2 Results

The descriptive statistics for the 14-, 21- and 28-day Realised Volatility have been presented in Table 19. This shows us that the mean of the three Realised Volatility is between 0.0107 and 0.0111. With the largest standard deviation existing amongst the 14-day Realised Volatility which makes sense given that the 21 and 28-day volatility will result in the volatility being more smoothed than the 14-day Realised Volatility.

	IP	I	Rho	TS	FEARS	14RV	21RV	28RV
count	1511.000000	1511.000000	1511.000000	1511.000000	1511.000000	1511.000000	1511.000000	1511.000000
mean	-0.246875	0.335076	0.563857	1.266905	132819.871277	0.010707	0.010914	0.011069
std	10.974670	2.432087	5.639699	5.502795	69239.720943	0.007098	0.006860	0.006577
min	-111.000000	-21.000000	-29.070000	-41.997000	55796.000000	0.001562	0.002363	0.002383
25%	-1.600000	-0.300000	-0.632000	0.401450	93198.000000	0.006418	0.006716	0.007016
50%	-0.200000	0.200000	0.261000	1.168000	107587.000000	0.009050	0.009329	0.009372
75%	1.000000	0.800000	1.309450	2.206500	125270.000000	0.012678	0.012781	0.012973
max	119.000000	35.800000	45.490000	32.848000	343802.000000	0.072784	0.069490	0.061633

Table 19: Descriptive Statistics for the combined dataset

As discussed above the MSE, MAE, HMSE, and HMAE error metrics will be used to assess the predictive ability of the models created. The error metrics compare the predictions made by the model on the test data set with the actual volatility in the market at the time. The error metrics for the ANN and Random Forest model are presented below in Table 20.

The rows in Table 20 present each of the four error metrics, while the columns are for each of the respective predictions made by the model. The first two columns show the four-error metrics for the 14-day Realised Volatility models, the first column is the model that includes the FEARS Index. With the second column showing the model that does not include the FEARS Index. This layout is repeated for the 21 and 28-day Realised Volatility error metrics.

The results presented in Table 20, show that the inclusion of the FEARS Index does not improve the model's ability to predict volatility. This is the case for both the ANN and Random

	14 Day Realised Volatility	14 Day Realised Volatility No FEARS	21-Day Realised Volatility	21-Day Realised Volatility No Fears	28-Day Realised Volatility	28-Day Realised Volatility No Fears
	ANN					
MSE	1.08E-03	9.47E-05*	2.44E-04	7.88E-06*	2.64E-03	1.27E-04*
MAE	3.29E-02	9.73E-03*	1.56E-02	2.81E-03*	5.13E-02	1.13E-02*
HMSE	6.46E-02	8.59E-03*	2.06E-02	7.42E-04*	2.33E-01	8.76E-03*
HMAE	2.54E-01	9.27E-02*	1.44E-01	2.72E-02*	4.82E-01	9.36E-02*
	Random Forest					
MSE	4.28E-06	3.05E-08*	5.72E-06	1.34E-06*	6.07E-06	2.50E-06*
MAE	2.07E-03	1.75E-04*	2.39E-03	1.16E-03*	2.46E-03	1.58E-03*
HMSE	9.36E-02	4.86E-03*	8.80E-02	3.36E-02*	8.15E-02	4.96E-02*
HMAE	3.06E-01	6.97E-02*	2.97E-01	1.83E-01*	2.86E-01	2.23E-01*

Table 20: Combined Dataset Error Metrics

* shows the lowest error measure for the Realised Volatility time horizon.

Forest model. Neither of the model's predictions are improved by the inclusion of the FEAR Index when using the combined dataset. The results also show us that when looking at the error metrics the Random Forest Model provides better predictions than both the ANN model across all of the forecasting horizons. This includes those that have the FEARS Index and those that do not. This shows that the Random Forest model is better able to identify the relationships that exist between the feature variables and Realised Volatility across all of the forecasting horizons.

The error metrics presented in Table 20 above do not allow for comparison between the 14, 21, and 28-day volatility as the error metrics are calculated based on the relative volatility of each volatility period. As such it is not possible to determine which of the volatilities is best predicted by each of the machine learning models from the error term presented above.

As discussed above in Chapter 3, ANN models are extremely data-dependent, i.e., they require a large amount of data to learn the patterns that are present between the feature and target variables. The data that is being presented to the model needs to be relevant to the predictions that are being made. As was seen above in Chapter 4 the ANN model is able to make more accurate predictions at an individual country level. This indicates that across the regions there is an inhomogeneity of sentiments even in regions that are close to each other, such as those in the Eurozone.

In Figure 22 and 23 below it can be seen that the ANN and Random Forest model have placed importance with regards to the feature variables in the dataset. The two different models have had their feature importance presented differently based on the functions available for each model, as discussed in Chapter 3 above. The ANN model has its feature importance presented as a portion of the magnitude that the feature provides to the prediction. Whereas the Random Forest feature importance is presented as what percentage effect does a feature have on the prediction made by the model.

From this, we can see that both the ANN and Random Forest model have placed the most importance on the FEARS Index in making their predictions. It can also be seen that apart from the FEARS Index both models placed the second most importance on Real Interest Rates. The relative importance placed on TS and Rho in Figures 22, 23, and 24 aligns with the findings of

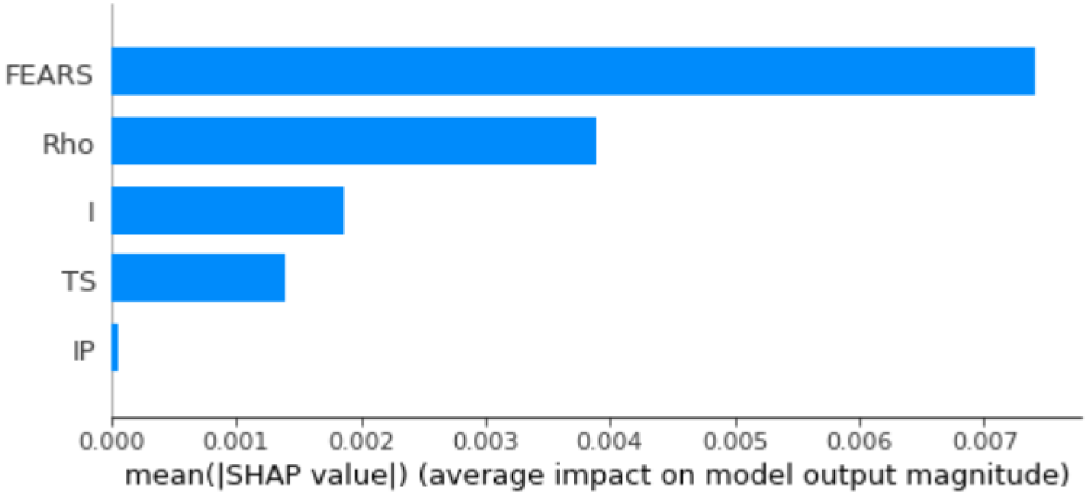


Figure 22: ANN 14 Day Realised Volatility Feature Importance with FEARS Index

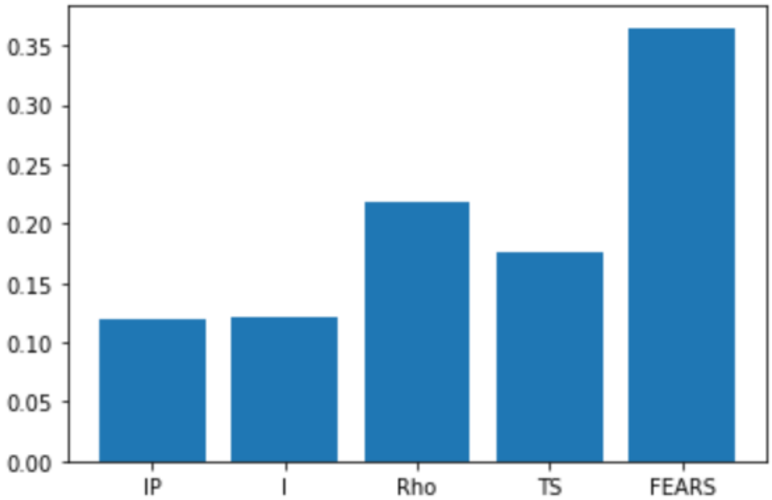


Figure 23: Random Forest 14 Day Realised Volatility Feature Importance with FEARS Index

Chen, Roll and Ross (1986) and Solanki and Seetharam (2018) who found that both of these factors had statistically significant explanatory power in stock market returns.

The importance placed on these factors aligns with the economic understanding of the factors. As Central Banks use short-term interest rates as a key monetary policy tool and the high degree of importance these short-term rates play in company valuations as they are used as discount rates in valuation techniques, such as the Discount Cash Flow methodology. Therefore, the inclusion of Treasury Bills in calculating both TS and Rho mean that they will capture movements in the index prices that will then translate into the volatility in the market. As these factors will capture changes in company valuations that will create movements, as investors buy or sell stocks based on the changes in intrinsic values due to changes in the discount rate, in the market creating additional market volatility.

Similar to the discussion above I will likely be a precursor to increases in short-term interest rates that are captured by TS and Rho. Therefore, the relative lack of importance placed on I likely means that the short-term interest rates are more directly affecting market volatility than changes in inflation. It can therefore be said that markets do not move for changes in inflation, rather markets only move when Central Banks make moves with the short-term interest rates.

In Figure 24 we can see that the model that excludes the FEARS Index. The noticeable difference between the two is that the without FEARS Index model appears to place relatively more weight on the IP. The difference in the weighting of the Growth in Industrial Production could be one of the reasons that the model with the FEARS Index is providing less accurate predictions. As it is relatively underweighting the predictive ability of Growth in Industrial Production. It can also be seen that the ANN model places very little importance on the Growth in Industrial Production which may be why it also does not provide as accurate a prediction as the Random Forest.

This relative importance placed on the Growth IP differs from the findings of Solanki and Seetharam (2018) who found that Growth IP only have statistically significant explanatory power in the UK. Given the importance that Growth IP had in contributing to the UK economies development given the industrial revolution (Solanki and Seetharam, 2018).

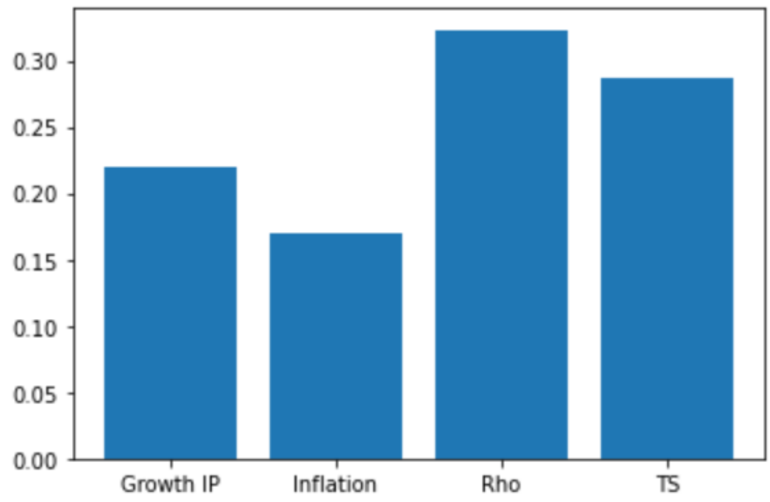


Figure 24: Random Forest 14 Day Realised Volatility Feature Importance without FEARS Index

A potential reason why the inclusion of the FEARS Index does not provide additional predictive power for the combined dataset may be the different countries present in the dataset. As some of the countries have shown that sentiment does not have explanatory or predictive power. This is seen in other research, such as the work done by Finter et al (2012), and above in Chapter 4.

In this chapter, we have seen that when using the combined dataset, the FEARS Index does not provide additional predictive power in both the ANN and the Random Forest model. Despite the fact that both models that include the FEARS Index have weighted it as the most important feature in making predictions the models without the FEARS Index are providing better predictions. It may be that the inclusion of countries where investor sentiment has been shown to not have explanatory power is creating confusion for the models preventing them from appropriately identifying relationships between the feature and target variables.

6. Conclusion

This chapter will provide an overview of the research findings and how they relate to the objectives of the research. Then some of the limitations of the research will also be presented, along with some areas of future research. Finally, the chapter concludes with some final remarks on the research.

The aim of this study was to determine whether the inclusion of an investor sentiment factor as a feature variable in a machine learning model would create improved volatility predictions. The following research question was investigated. Does the inclusion of an investor sentiment factor, as proxied by the FEARS Index, allow for ML models to make improved predictions of realised volatility?

This study attempted to answer this question by first reviewing the academic literature around both investor sentiment and machine learning applications in finance. Following this several LSTM, ANN, and Random Forest models were created for a combined dataset, which combined the volatility of all seven of the G7 countries. As well as for each of the seven individual countries. The results of this study showed that the inclusion of the FEARS Index did not provide any improved predictive ability for the combined dataset. However, with the individual countries, some of the predictions were improved and others were not.

6.1 Limitations of research

Although the results of the study show that in some cases the inclusion of the FEAR Index does provide improved volatility predictions it is not without its limitations.

Firstly, the limited amount of Search Volume data available from Google Trends meant that the ML models that are very data-hungry lacked an appropriate amount of data to make accurate predictions at all times.

The design of ML models requires that each model be built for the specific set of data that is being presented to the model. As such the results of this study are not generalisable as they will only speak to the specific countries that were under observation in this study. Due to the countries chosen, it is also not possible to determine whether investor sentiment has greater predictive power in developing markets. These markets tend to be less sophisticated and should thus be more susceptible to investor sentiment as seen in the academic literature.

6.2 Areas of future research

Future research into whether the inclusion of the FEARS Index as a feature variable would provide improved predictions of returns. This would show whether the academic literature regarding the explanatory power of the FEARS Index in relation to returns holds for the predictive ability of returns. Another area around investor sentiment that is ripe for further exploration would be to use a different investor sentiment proxy. Such as Twitter's sentiment towards the market as well as a more traditional investor sentiment factor such as the sentiment proxy used by Baker and Wurgler in their 2007 study. This would provide additional evidence as to whether investor sentiment has a predictive ability for volatility.

Additionally, it would be interesting to investigate whether using time-lagged feature variables would create better predictions, it was seen with the combined dataset that they seemed to make predictions that match the volatility of the previous period. An investigation into whether the t-1 period feature variables is able to make predictions that more closely match the time period t target variable. This would show that finance practitioners would be able to make use of the current periods variables to predict the volatility of the next time period.

Some potential future research would be to look at volatility spillover between regions. This could further investigate the finding of the inhomogeneity that was identified in Chapter 5. Where it was found that the individual country predictions were better than that of the combined dataset. This investigation could look at how volatility spills over specifically the spillover of the FEARS Index into other regions, the use of wavelet analysis to identify which countries have spillover as well as in which direction the spillover occurs.

Further, a look into how sentiment spillover occurs for online sentiment factors, such as the FEARS Index or the Twitter sentiment factor to see if spillover happens amongst online communities. Also, looking at whether the sentiment spillover has increased as the world has become more globalised would form part of the study.

Finally, work around different ML models' ability to make improved predictions with highly heteroskedastic data such as stock markets. This research could provide further insight into the results found in Chapter 4 regarding the Italian and French market predictions where the Random Forest model had lower heteroskedastic error measures than the non-heteroscedastic error measures.

6.3 Concluding remarks

This study attempted to determine whether the inclusion of the FEARS Index would improve a ML model's ability to predict volatility. This research has added to the body of knowledge around the use of investor sentiment in making volatility predictions as well as the use of machine learning in making volatility predictions. The results showed that the FEARS Index does provide improved volatility predictions but not in all cases. The results do not show a clear pattern as to when the FEARS Index does provide improved predictions. As such the study has failed to provide strong evidence in favour of the use of the FEARS Index as a feature variable in ML models when making volatility predictions.

Reference List

Abe, M. and Nakayama, H., 2018. Deep Learning for Forecasting Stock Returns in the Cross-Section. *Advances in Knowledge Discovery and Data Mining* , pp.273-284.

Ahsan, M., Mahmud, M., Saha, P., Gupta, K. and Siddique, Z., 2021. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies* , 9(3), p.52.

Alaloul, W.S. *et al.* (2018) "An artificial neural networks (ANN) model for Evaluating Construction Project Performance Based on Coordination Factors," *Cogent Engineering*, 5(1), p. 1507657. Available at: https://doi.org/10.1080/23311916.2018.1507657.

Audrino, F., Sigrist, F. and Ballinari, D., 2020. The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting* , 36(2), pp.334-357.

Baker, M. and Wurgler, J., 2007. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21(2), pp.129-151.

Baker, S.R., Bloom, N. and Davis, S.J. (2016). Measuring Economic Policy Uncertainty. *SSRN Electronic Journal*, [online] 131(4). doi:https://doi.org/10.2139/ssrn.2198490.

Barber, B. and Odean, T., 2008. All that glitters, is the effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2), pp785-818.

Barberis, N. and Thaler, R., 2003. A Survey of Behavioral Finance. In: *Handbook of the Economics and Finance*, 1B, pp. 1053-1128.

Bianchi, T (2024). Search engine market share worldwide | Statista. [online] Statista. Available at: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/.

Breiman, L., 2001. Random Forests. *Machine Learning* , 45, pp.5-32.

Cao, Q., Leggio, K. and Schniederjans, M., 2005. A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. *Computers & Operations Research* , 32(10), pp.2499-2512.

Chen, J., 2021. An Introduction to Machine Learning for Panel Data. *International Advances in Economic Research* , 27(1), pp.1-16.

Chen, N., Roll, R. and Ross, S., 1986. Economic Forces and the Stock Market. *The Journal of Business* , 59(3), p.383.

Chlebus, M., Dyczko, M. and Woźniak, M., 2021. Nvidia's Stock Returns Prediction Using Machine Learning Techniques for Time Series Forecasting Problem. *Central European Economic Journal* , 8(55), pp.44-62.

Connor, G. and Korajczyk, R., 1992 The Arbitrage Pricing Theory and Multifactor Models of Asset Returns, http://www.kellogg.northwestern.edu/faculty/korajczy/ftp/wp139.pdf p28

De Long, J., Shleifer, A., Summers, L. and Waldmann, R., 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy* , 98(4), pp.703-738.

Dietterich, T., 2000. *Ensemble Methods in Machine Learning* .

Dimpfl, T. and Jank, S., 2015. Can Internet Search Queries Help to Predict Stock Market Volatility?. *European Financial Management* , 22(2), pp.171-192.

Dixon, M. and Halperin, I., 2019. The Four Horsemen of Machine Learning in Finance. *SSRN Electronic Journal* ,.

Enke, D. and Thawornwong, S., 2005. The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications* , 29(4), pp.927-940.

Erdem, E. and Ulucak, R. (2016) “Efficiency of stock exchange markets in G7 countries: Bootstrap causality approach,” *Economics World*, 4(1). Available at: https://doi.org/10.17265/2328-7144/2016.01.003.

Fama, E. and French, K., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* , 33(1), pp.3-56.

Fama, E. and French, K., 2015. A five-factor asset pricing model. *Journal of Financial Economics* , 116(1), pp.1-22.

Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* , 25(2), p.383.

Fama, E.F.: The behavior of stock market prices. J. Bus. 38(1), 34–105 (1965)

Fama, E.F. and French, K.R. (2017) “International tests of a five-factor asset pricing model,” *Journal of Financial Economics*, 123(3), pp. 441–463. Available at: https://doi.org/10.1016/j.jfineco.2016.11.004.

Finter, P., Niessen-Ruenzi, A. and Ruenzi, S., 2012. The impact of investor sentiment on the German stock market. *Zeitschrift für Betriebswirtschaft* , 82(2), pp.133-163.

Flannery, M. and Protopapadakis, A., 2002. Macroeconomic Factors DO Influence Aggregate Stock Returns. *SSRN Electronic Journal*.

Foucault, T., Sraer, D. and Thesmar, D., 2011. Individual Investors and Volatility. *The Journal of Finance* , 66(4), pp.1369-1406.

Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, ICML'96. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 1996:148e156.

Friedman, J., Hastie, T. and Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics* , 28(2).

Hamid, A. and Heiden, M., 2015. Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization* , 117, pp.62-81.

Ho, J. and Hung, C., 2012. Predicting Stock Market Returns and Volatility with Investor Sentiment: Evidence from Eight Developed Countries. *SSRN Electronic Journal* ,.

Ito, M., Noda, A. and Wada, T. (2014) "International Stock Market Efficiency: A non-bayesian time-varying model approach," *Applied Economics*, 46(23), pp. 2744–2754. Available at: https://doi.org/10.1080/00036846.2014.909579.

Jafar, R., Shahrour, I. and Juran, I. (2010) "Application of artificial neural networks (ANN) to model the failure of urban water mains," *Mathematical and Computer Modelling*, 51(9-10), pp. 1170–1180. Available at: https://doi.org/10.1016/j.mcm.2009.12.033.

Jing, N., Wu, Z. and Wang, H., 2021. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications* , 178, p.115019.

Karlson, N., Loewenstein, G. and Seppi, D., 2009. The ostrich effect: selective attention to information. *Journal of Risk and Uncertainty,* 38, pp.95-115.

Khandelwal, I., Adhikari, R. and Verma, G. (2015) "Time series forecasting using Hybrid Arima and Ann models based on DWT decomposition," *Procedia Computer Science*, 48, pp. 173–179. Available at: https://doi.org/10.1016/j.procs.2015.04.167.

Kia, A., 2003. Forward-looking agents and macroeconomic determinants of the equity price in a small open economy. *Applied Financial Economics* , 13(1), pp.37-54.

Kraft, J. and Kraft, A., 1977. Determinants of Common Stock Prices: A Time Series Analysis. *The Journal of Finance* , 32(2), p.417.

Krauss, C., Do, X. and Huck, N., 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* , 259(2), pp.689-702.

Kryzanowski, L., Galler, M. and Wright, D., 1993. Using Artificial Neural Networks to Pick Stocks. *Financial Analysts Journal* , 49(4), pp.21-27.

Kumar, J., Goomer, R. and Singh, A., 2018. Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters. *Procedia Computer Science* , 125, pp.676-682.

Lee, W., Jiang, C. and Indro, D., 2002. Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking & Finance*, 26(12), pp. 2277-2299.

LeRoy, S. and Porter, R., 1981. The Present-Value Relation: Tests Based on Implied Variance Bounds. *Econometrica* , 49(3), p.555.

Lintner, J., 1965. Security Prices, Risk, and Maximal Gains From Diversification. *The Journal of Finance* , 20(4), p.587.

Lipovetsky, S. and Conklin, M. (2001) “Analysis of regression in game theory approach,” *Applied Stochastic Models in Business and Industry*, 17(4), pp. 319–330. Available at: https://doi.org/10.1002/asmb.446.

Masuduzzaman, M., 2012. Impact of the Macroeconomic Variables on the Stock Market Returns: The Case of Germany and the United Kingdom. *Global Journal of Management and Business Research* , 12(16), pp.23 - 34.

Moghar, A. and Hamiche, M., 2020. Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science* , 170, pp.1168-1173.

Nevasalmi, L., 2020. Forecasting Multinomial Stock Returns Using Machine Learning Methods. *SSRN Electronic Journal* ,.

Nur Ozkan-Gunay, E. and Ozkan, M. (2007) “Prediction of bank failures in emerging financial markets: An ann approach,” *The Journal of Risk Finance*, 8(5), pp. 465–480. Available at: https://doi.org/10.1108/15265940710834753.

Pedregosa, F., Varoquax, G., Gromfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, pp. 2825 – 2830

Qi, M. and Maddala, G., 1999. Economic factors and the stock market: a new perspective. *Journal of Forecasting* , 18(3), pp.151-166.

Ricciardi, V. and Simon, H., 2000. What is Behavioral Finance?. *Business, Education & Technology Journal* , 2(2), pp.1-9.

ROLL, R. and ROSS, S., 1980. An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of Finance* , 35(5), pp.1073-1103.

Ross, S., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* , 13(3), pp.341-360.

Sak, H., Senior, A.W., & Beaufays, F. (2014). Long short-term memory based recur- rent neural network architectures for large vocabulary speech recognition. CoRR , arXiv preprint arXiv: 1402.1128 .

Samuelson, P.A.: Proof that properly anticipated prices fluctuate randomly. Ind. Manag. Rev. 6(2), 41–49 (1965)

Sarwar, G., Mateus, C. and Todorovic, N., 2015. Macroeconomic Determinants of Cyclical Variations in Value, Size and Momentum Premiums in the UK. *SSRN Electronic Journal*.

Sharpe, W., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance* , 19(3), p.425.

Shiller, R.J.: The use of volatility measures in assessing market efficiency. NBER Working Paper No. 565 (1981)

Shleifer, A. and Summers, L., 1990. The Noise Trader Approach to Finance. *Journal of Economic Perspectives* , 4(2), pp.19-33.

Solanki, K. and Seetharam, Y., 2018. Is investor sentiment a relevant factor in determining asset prices?. *Investment Analysts Journal*, 47(3), pp.243-257.

Solanki, K., 2017. *Investor Sentiment as a Factor in an APT Model: An International Perspective Using the FEARS Index* . Masters. University of the Witwatersrand.

Tan, Z., Yan, Z. and Zhu, G., 2019. Stock selection with random forest: An exploitation of excess return in the Chinese stock market. *Heliyon* , 5(8), p.e02310.

Tetlock, P., 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3), pp.1139-1168.

Tetlock, P., Saar-Tsechansky, M. and Macskassy, S., 2007. More than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, 63(3), pp.1437 - 1466.

Thaler, R. (1993). Advances in behavioural finance. New York: Princeton University Press.

Uygur, U. and Taş, O., 2013. The impacts of investor sentiment on returns and conditional volatility of international stock markets. *Quality & Quantity* , 48(3), pp.1165-1179.

Vuong, N. and Suzuki, Y., 2020. Does Fear has Stronger Impact than Confidence on Stock Returns? The Case of Asia-Pacific Developed Markets. *Scientific Annals of Economics and Business* , 67(2), pp.157-175.

Xu, Y., Wang, J., Chen, Z. and Liang, C. (2021). Economic policy uncertainty and stock market returns: New evidence. *The North American Journal of Economics and Finance*, 58, p.101525. doi:https://doi.org/10.1016/j.najef.2021.101525.

Yang, M. and Wang, J., 2022. Adaptability of Financial Time Series Prediction Based on BiLSTM. *Procedia Computer Science* , 199, pp.18-25.

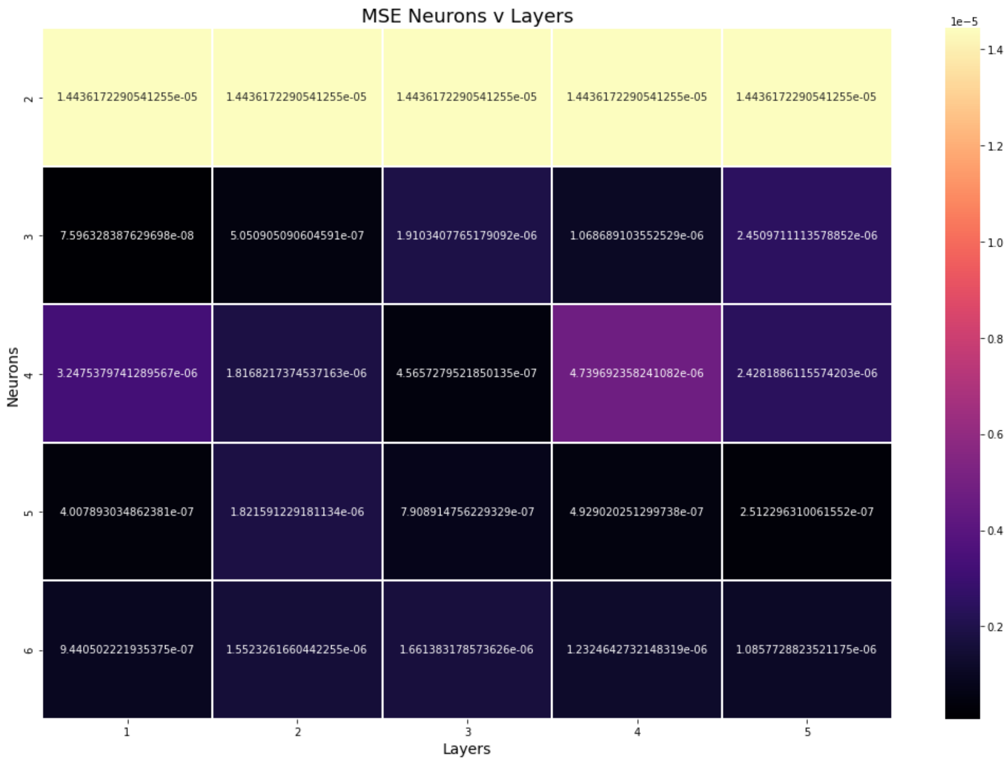
You, W., Guo, Y. and Peng, C., 2017. Twitter's daily happiness sentiment and the predictability of stock returns. *Finance Research Letters* , 23, pp.58-64.

Yu, J. and Yuan, Y., 2011. Investor sentiment and the mean–variance relation☆. *Journal of Financial Economics*, 100(2), pp.367-381.

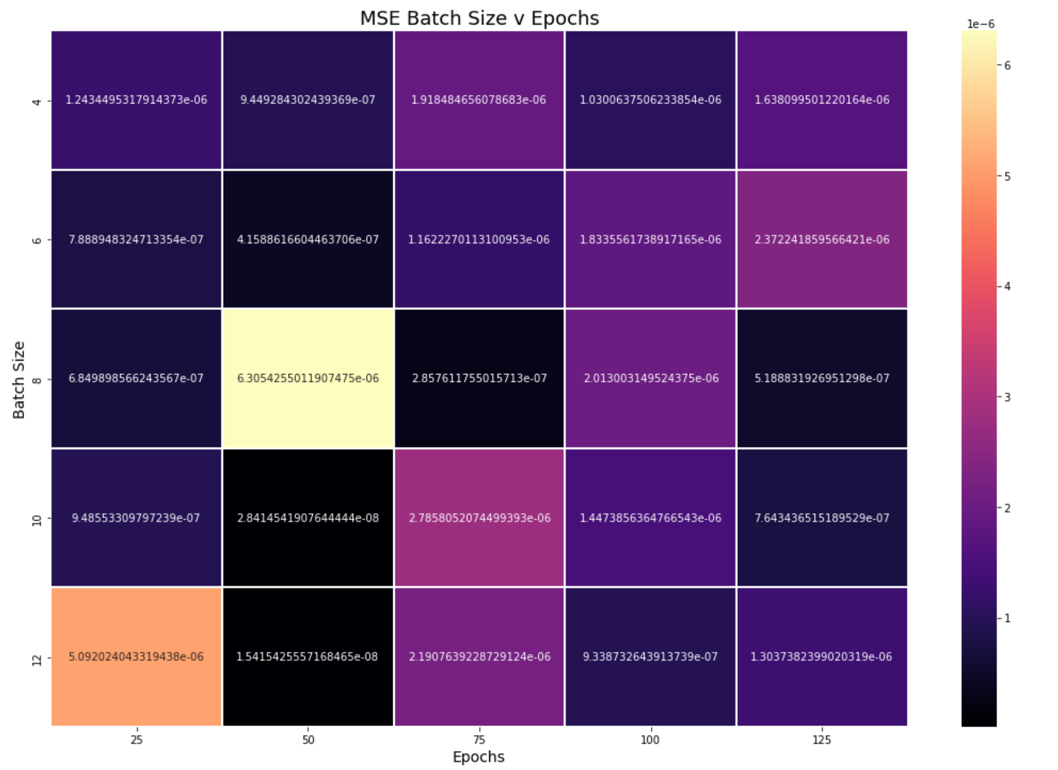
Appendix

Appendix 1 – Matrices

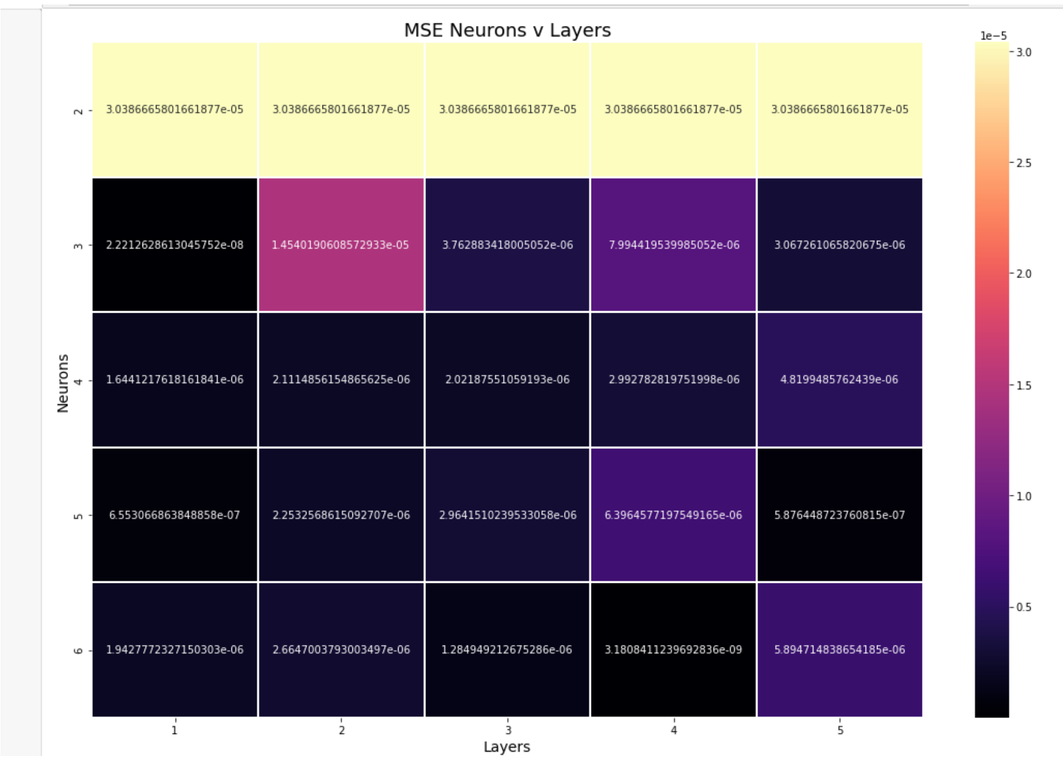
Canada 21-day Realised Volatility batch size and epochs matrix



Canada 21-day Realised Volatility neurons and layers matrix



France 28-day Realised Volatility batch size and epochs matrix



France 28-day Realised Volatility neurons and layers matrix

Appendix 2 – Descriptive statistics of for individual country dataset

	Growth IP	Inflation	Rho	TS	FEARS	14	21	28
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000
mean	-0.454019	0.156019	1.259736	1.156167	132890.486111	0.008472	0.008734	0.008788
std	3.342927	0.369111	1.260504	0.871458	69409.191795	0.006704	0.006729	0.006391
min	-15.250000	-1.000000	-0.637000	-0.782000	55796.000000	0.001868	0.002363	0.002383
25%	-1.426901	-0.100000	0.371750	0.565000	93227.500000	0.004911	0.005080	0.005346
50%	-0.500000	0.200000	0.886500	1.094000	107601.500000	0.006755	0.007003	0.007249
75%	0.256964	0.400000	1.845000	1.578500	125381.000000	0.009495	0.010056	0.009899
max	25.636364	1.200000	4.747000	3.362000	343802.000000	0.062927	0.062201	0.056136

Canada Descriptive Statistics

	IP	Iinflation	Rho	TS	FEARS	14	21	28
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000
mean	-0.000926	-0.000926	0.620394	1.047338	132890.486111	0.011199	0.011419	0.011697
std	3.000961	0.571981	1.637842	0.723533	69409.191795	0.006316	0.006184	0.006013
min	-9.700000	-1.800000	-1.566000	-0.324000	55796.000000	0.003119	0.003764	0.003819
25%	-1.525000	-0.300000	-0.681500	0.439000	93227.500000	0.007279	0.007615	0.008150
50%	-0.200000	0.000000	0.102500	1.010000	107601.500000	0.009915	0.010020	0.010131
75%	1.400000	0.300000	1.803000	1.560000	125381.000000	0.013098	0.012951	0.013154
max	27.100000	1.700000	4.621000	2.692000	343802.000000	0.048097	0.046828	0.045524

Germany Descriptive Statistics

	IP	Inflation	Rho	TS	FEARS	14	21	28
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000
mean	-0.001852	0.002315	0.999444	2.137213	132890.486111	0.012811	0.012988	0.013258
std	6.069703	1.278132	1.771502	1.079046	69409.191795	0.007710	0.007330	0.007171
min	-37.200000	-2.500000	-2.719000	0.197000	55796.000000	0.003672	0.003901	0.003923
25%	-1.600000	-0.900000	-0.261000	1.255000	93227.500000	0.008091	0.008247	0.008853
50%	0.000000	-0.100000	0.661000	2.185500	107601.500000	0.011204	0.011266	0.011630
75%	1.600000	0.925000	2.276000	2.994000	125381.000000	0.015212	0.015074	0.014916
max	65.800000	2.800000	5.863000	4.564000	343802.000000	0.059810	0.056624	0.051131

Italy Descriptive Statistics

	IP	Inflation	Rho	TS	FEARS	14	21	28
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000
mean	-0.004167	-0.037963	0.052046	0.684972	132890.486111	0.012336	0.012510	0.012610
std	3.445500	5.785462	0.389225	0.475045	69409.191795	0.007257	0.006811	0.006685
min	-17.100000	-21.000000	-2.025000	-0.087000	55796.000000	0.003668	0.003668	0.003668
25%	-1.500000	-3.125000	-0.184750	0.222000	93227.500000	0.008357	0.008752	0.008686
50%	-0.200000	0.200000	-0.001000	0.656500	107601.500000	0.010878	0.011072	0.011077
75%	1.525000	2.500000	0.300000	1.079750	125381.000000	0.014421	0.014430	0.014272
max	18.700000	35.800000	1.231000	1.812000	343802.000000	0.072784	0.069490	0.061633

Japan Descriptive Statistics

	Growth in IP	RHO	TS	Inflation	FEARS	14	21	28
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000
mean	-1.301387	1.174178	1.052202	2.199537	132890.486111	0.009253	0.009478	0.009663
std	14.797397	2.407726	1.054005	1.187865	69409.191795	0.006089	0.005945	0.005662
min	-111.000000	-7.083200	-0.758000	-0.100000	55796.000000	0.002599	0.003169	0.003334
25%	-1.754808	0.268500	0.149950	1.500000	93227.500000	0.005778	0.006017	0.006354
50%	-0.777358	0.503650	0.869500	2.200000	107601.500000	0.007633	0.008136	0.008346
75%	0.131159	4.079500	1.759250	2.900000	125381.000000	0.010409	0.011351	0.011220
max	119.000000	5.794000	3.366300	5.400000	343802.000000	0.052869	0.049217	0.045695

UK Descriptive Statistics

	Growth IP	Inflation	Rho	TS	FEARS	14	21	28
count	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000
mean	0.043996	0.023721	-0.976441	1.593160	132394.211628	0.009617	0.009782	0.009736
std	1.273564	0.485534	1.756119	1.109016	69185.990970	0.007613	0.007244	0.006404
min	-13.245900	-2.600000	-6.954400	-1.601900	55796.000000	0.001562	0.003051	0.002690
25%	-0.288300	-0.200000	-1.765050	0.833200	93198.000000	0.005411	0.005613	0.005965
50%	0.133900	0.000000	-0.989500	1.605200	107587.000000	0.007572	0.007904	0.007880
75%	0.541050	0.300000	0.000000	2.431250	125130.500000	0.011194	0.011371	0.011258
max	5.318800	2.000000	3.873000	3.790200	343802.000000	0.062230	0.058659	0.048492

France Descriptive Statistics