A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

# Revisiting Equity Premium Forecasts: Applying Machine Learning and Deep Learning Techniques

Korbinian Gabriel

43473

Directed research carried out under the supervision of:

Prof. Nicholas H. Hirschey

17-12-2021

**Abstract:**

This study re-investigates the relationship between the equity premium and variables, that have been either examined for predictive qualities by the literature or are used by active fund managers at the time of the study, with state-of-the-art machine learning models to uncover potential non-linear relationships. Results show that within the U.S. market, no variable, whether suggested by the literature or by active fund managers, is able to forecast the equity premium. However, it appears as though there is a predictive relationship between European long-term yields and the European equity premium.

**Keywords:**

*Financial Markets, Equity Premia, Machine Learning, Deep Learning, Financial Variables, Forecast.*

**Acknowledgements:**

I would like to thank my supervisor Professor Nicholas H. Hirschey. Thank you for your encouragement, patience and kind support through this process.

I would also like to express my deepest gratitude to my parents for their support through six university years.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209)

# 1 Introduction

A longstanding debate persists among observers and participants in financial markets about whether markets are reliably predictable and whether participants, indicators, or models actually exist that can repeatedly verify forecasting ability over the long term. Predicting the equity premium is vital for asset managers of all forms as it allows them to structure portfolios and asset allocations in the most optimal way. The basis of this prediction discussion is the Efficient Market Hypothesis, which postulates in three forms - strong, semi-strong and weak - that prices in capital markets always fully price in all accessible information (Bertoneche 1979; Fama 1970, 1995; Samuelson 1965, 1973), making it impossible to outperform the market consistently and thereby ruling out consistent forecasting. Opponents of the Efficient Market Hypothesis point to statistical properties of capital markets (Lo and MacKinlay 1988; Osborne 1962), as well as behavioral irrationality of markets themselves (Jegadeesh and Titman 1993). The literature is too vast to be thoroughly listed here. As there seem to be examples of funds outperforming the market over a longer period of time (Bali, Brown, and Demirtas 2013) [1] , attempts to further develop the EMH, for example the Adaptive Market Hypothesis (Lo 2004) which combines the EMH with behavioral biases such as loss aversion (Kahneman and Tversky 1979) or overreaction (Bondt and Thaler 1985), allow for temporary existence of forecasting methods of asset prices under the EMH. It has been frequently discovered that certain mispricings and biases tend to arise only under certain market conditions (Ammann and Verhofen 2009; Asness, Moskowitz, and Pedersen 2013). Moreover, and with special regard to systematic mispricings and forecasting, theories of crowding out and alpha decay have emerged (Baltas 2019; McLean and Pontiff 2016), serving explanations for vanishing performance of arbitrage and forecasting models.

1. Research on this topic is prone to the issue that successful closed hedge funds have no incentive to report their performance (e.g. Renaissance Technologies Medallion Fund). Additionally, out-performance can partly be explained through other risk premia such as illiquidity risk. Survivorship bias is often included in general statistics over the industry.

In their paper *"A comprehensive look at the empirical performance of equity premium prediction"*, Ivo Welch and Amit Goyal examine variables that previous literature has signed with good forecasting power of the equity premium in the US market (Welch and Goyal 2008). They find that most variables lost their predictive power within their respective models to the extent that those models are explained as being unstable or were spurious findings. New sources of information, financial innovation and technological progress, however, could find new inefficiencies and exploit them, as for example evident in high frequency arbitrage trading (Aquilina, Budish, and O'Neill 2020; Poutré, Dionne, and Yergeau 2021) or through the adoption of data mining and machine learning (Dash and Dash 2016; Enke and Thawornwong 2005; Prado 2016) [2] . The idea is that while inefficiencies and predictive relations have not been found in one market, possibly could be in other markets, uncovered utilizing different data or through the utilization of more advanced technology.

In this study, three experiments are undertaken to further investigate the contemporary predictive power on the equity premium. First, the research of Goyal & Welch is replicated and extended into 2020 to update the study. Next, the study is expanded to Europe, using European proxies for the variables. Finally, new predictors of the equity premium are examined, selected on the basis of interviews with at the time of this study active fund managers. All variables are examined using both the original in-sample (IS) and rolling out-of-sample (OOS) regression approach, as well as employing state-of-the art machine learning and deep learning to probe whether complex models are able to discover relations more accurately. This study contributes to the current body of research by not only testing the robustness of Goyal & Welch's results with 15 years of additional data across more complex models, but also by further investigating other markets and more variables. It is found that none of the variables have gained significance with time and even sophisticated models are not able to find significant relationships between the equity premium and individual variables.

2. Although many publicized approaches have been proven unprofitable out of sample.

However, a relationship could be found between european yields and the respective equity premium.

The chapters in this study are organized as follows: Chapter 2 provides a background to Machine Learning and Deep Learning. It is encouraged to use it like a lookup source if terms are unknown and skip it otherwise. References are made in the main text. Chapter 3 reviews previous studies and refers to results presented by other researchers. It also introduces the main hypotheses investigated in this study. Chapter 4 states the selected data and explains the applied methodology. Chapters 5 is the primary chapter concerned with the findings of this study. It presents the results and provides a discussion of those in detail. Finally, Chapter 6 concludes this study with a summary and suggests future research opportunities.

# 2 Background

The goal of this chapter is to briefly discuss multiple Machine Learning approaches. Models with a neural network architecture are discussed separately.

## 2.1 Machine Learning

Machine Learning is a sub-field of Artificial Intelligence [3] and Computer Science [4] (see Figure 7). It describes the learning from data and the performance of tasks without the underlying rules being explicitly programmed. The general concept of machine learning can be divided into supervised and unsupervised learning. Unsupervised learning algorithms seek the internal structure within data. In doing so, they respond to similarities. Examples of such algorithms would be the clustering algorithms which try to make classifications based on groups. Supervised learning algorithms, in

3. The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings (Copeland 1998).

4. The study of computers and computing, including their theoretical and algorithmic foundations, hardware and software, and their uses for processing information (Belford 1999).

contrast to regular programming where the combination of data and rules leads to outputs, seek to infer the rules by looking at the results and the data. If a general rule is found, it can then be applied to new, unseen data. Optimally, should the underlying relationship not have changed, this should lead to an accurate prediction of the output. Examples of such algorithms are, among others, regressions, support vector machines and decision trees.

### 2.1.1 Ordinary Least Squares

Ordinary Least Squares estimates the parameters to solve between the input vector $X^T = (X_1, X_2, X_3, \ldots)$ and the output $Y$. The underlying assumption is that the relationship between $Y$ and $X^T$ is of linear nature. Here the model has the general form

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \tag{1}$$

where the residual sum of squares is then minimized during estimation (Hastie, Tibshirani, and Friedman 2009)

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 \tag{2}$$

In the specific case of the problem studied here, $f(X)$ would represent the equity premium while $X_j \beta_j$ would be the representation of the individual indicators.

### 2.1.2 Shrinkage Methods: LASSO and Ridge Regression

Shrinkage methods are implicit feature selection models. By selecting a subset of the features they try to result in more interpretable models with lower prediction error, essentially reducing noise. Ridge regression (Hoerl and Kennard 1970b, 1970a; Hilt and Seegrist 1977), just like OLS, minimizes the residual sum of squares, but the coefficients are subjected to a penalty that decreases their magnitude.

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{3}$$

LASSO (Santosa and Symes 1986; Tibshirani 1996) - least absolute shrinkage and selection operator - has the same properties as the Ridge regression, but the regularization function of LASSO is subject to $L_1 := \lambda \sum p_{j=1}|\beta_j|$, so it tends to set coefficients to zero, unlike Ridge, which never sets coefficients entirely to zero (Hastie, Tibshirani, and Friedman 2009).

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta| \right\} \tag{4}$$

Lambda thereby represents the strength of the allowed regularization effect. In essence, the LASSO regularization, in contrast to the Ridge regularization, ensures that the model is built more quickly on a smaller feature set and is therefore more easily interpreted and suffers from less noise. However, this can also lead to the omission of important features and thus to a decrease in the explanatory power of the model. Since in this study every variable is tested on its own, coefficients can be forced to zero if the penalty is too large, leading to no model at all for the variable.

### 2.1.3 Support Vector Machine

Support Vector Machines (Boser, Guyon, and Vapnik 1992) are based on the idea of using lines (or hyperplanes) to divide data into groups. They provide a technique for classifying overlapping classes with non-linear boundaries based on a transformed version of the feature space (Hastie, Tibshirani, and Friedman 2009). Support vectors thereby are the data points which, if removed, would change the viewpoint or orientation of the hyperplane. Hyperplanes can become multidimensional and then divide a space using, for example, a three-dimensional space. SVMs can be adapted for regression problems. In doing so, the regression model is extended by dimensions via the kernel method (Patle and Chouhan 2013), essentially ”bending it in space” and thus enabling a new spatial separation via lines or hyperplanes. The regression estimates are:

$$f(x) = \sum_{t=1}^{T} (\alpha_t^* - \alpha_t)\Gamma(x_t, x_s) + b^* \tag{5}$$

$$\text{with} \quad \Gamma(x_t, x_s) = \begin{cases} x^T, & \text{linear function} \\ exp\Big( - \frac{||x-x'||^2}{2\sigma^2} \Big), & \text{radial basis function} \\ (x^T y + c)^d, & \text{polynomial function} \end{cases} \tag{6}$$

with $\alpha_t^* - \alpha_t$ being the models non-zero constant and $\Gamma(x_t, x_s)$ representing the selected kernel function. Variables $y$ and $c$ are kernel parameters, $d$ the polynomial degree (Nalbantov, Bauer, and Sprinkhuizen-Kuyper 2006; Iworiso 2020).

### 2.1.4 Random Forest

Random Forests (Breiman 2001) are large collections of decision trees and belong to the category of ensemble learners. A decision tree is a theoretical concept (or algorithm) whereby a node represents an outcome and each branch away a decision with an outcome at the end. Random Forests can be applied in both classification problems and in regressions. For a regression, the trees are trained many times with bootstrapped versions of the dataset and then averaged. Bootstrapping itself is a re-sampling technique that creates sample datasets with replacement (also referred to as bagging method). This method counteracts the overfitting behavior of many decision trees. Random forests include the special feature of calculating a feature importance, based on the improvement of the split criterion over all decision trees for the respective variable, i.e. how well the prediction improves.

## 2.2 Deep Learning

Deep Learning is a subfield of Machine Learning itself and is based on the use of Artificial Neural Networks consisting of neurons modeled after the biological brain, called artificial neurons [5] Thereby successive layers of more and more meaningful representations of the data are created.

5. The literature is not consistent in the term. The name ”Perceptron” is used interchangeably with neuron, but is also referring to a certain type of neuron or network. The definition by Rosenblatt refers to Perceptron as a classifier algorithm (Rosenblatt 1958). A neuron in its simplest form is a mathematical function sending activation signals (Figure 8).

The depth of the network indicates the number of these (hidden) layers. Depending on the design of the network, the properties of the neurons and their connections, different network architectures can be created (Liu et al. 2017). Neural Networks are good at finding patterns in raw data without much prior feature engineering compared to more traditional machine learning models. The drawback is that the required computing power is considerably higher, which is why the networks have become popular only in recent years.

### 2.2.1 Deep Neural Network

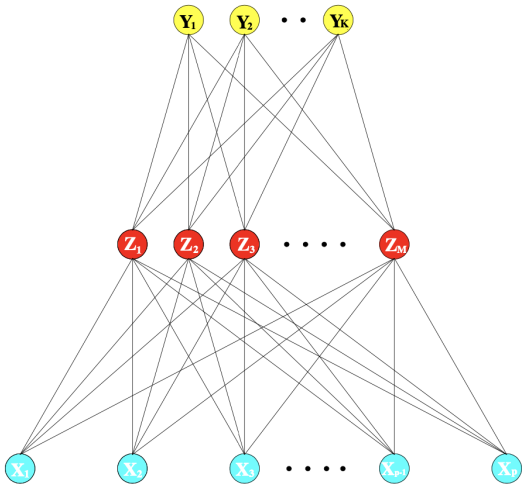


**Figure 1:** *The basic structure of a feed-forward neural network. Z represents the hidden layer, which can be multiple, increasing the models depth and with that the amount of feature representations (Figure by Hastie, Tibshirani, and Friedman 2009 - p.393).*

The central idea of Deep Neural Networks is to create linear combinations of the inputs and then model the target variable as a non-linear function to this (Hastie, Tibshirani, and Friedman 2009). Each neural network consists of layers of nodes that can be individually weighted, controlled by an activation function (see Figure 8), and which pass information to other layers (in this case feed-forward). The depth of the network, the size of the layers, the weighting, the activation function, the information forwarding mechanism, and more, can be freely adjusted. The individual layers are

to be defined as follows:

$$
\begin{aligned}
h_1 &= H_1(w_1 Z + B_1) \\
h_2 &= H_2(w_2 h_1 + B_2) \\
&\vdots \\
r &= H_L(w_L h_{L-1} + B_L)
\end{aligned}
\tag{7}
$$

where $Z$ is $L \times 1$ vector of input covariates; $L$ is the number of layers. Each network optimizes an objective function, and models may be evaluated using the loss function or a cost function such as mean squared error. Most networks get their weights of the neuron connections or learning rates, the speed of optimization, adjusted via optimizers during training to keep the loss low, for example via Stochastic Gradient Descent or Adaptive Moment Estimation (ADAM - Kingma and Ba 2014). In the training process of feed-forward Deep Neural Networks, so-called backpropagation (Rumelhart, Hinton, and Williams 1986) is applied. Backpropagation is an algorithm that updates the weights of each layer by iterating the gradient of the loss function backwards over the network, which makes the model more accurate.

### 2.2.2 Recurrent Neural Network: Long Short-Term Memory

Recurrent Neural Networks (RNN) are types of neural networks that are capable of learning temporal relationships, i.e., working efficiently with time-series problems. They accomplish this by retaining information about previously processed data within the network. In this way, they are able to perform well with for example with linguistic tasks. In finance they have become popular as they promise to deliver good results with time-series data (see section 3). However, they suffer from a vanishing-gradient problem that arises during the backpropagation operation for instance, when weights become of such small size that they do not change meaningfully even when they are updated (Chollet 2017). This problem is handled by introducing Long Short-Term Memory (LSTM) units (see Figure 9, Hochreiter and Schmidhuber 1997). These units can keep information in the network for a long time and establish gates to decide if and how a cell in the network is updated.

# 3 Literature Review

The prediction of financial markets has been studied scientifically for over 100 years. Charles Dow explored the influence of dividend ratios on the equity premium (Dow and Selden 1920). Fama and French were studying the predictive power of dividend yields on stock returns. They report that dividend yields increase in predictive power the longer the return horizon and attribute this to the influence of autocorrelation on return variance, as well as to a discounting effect on expected returns due to shocks (Fama and French 1988). Continuing, in a previous publication to their 2008 paper, Goyal and Welch examined dividend ratios in relation to the equity premium and showed that they lacked consistent predictive power on the equity premium. They derived the supposed ability from outlier years (Goyal and Welch 2003). However, in subsequent studies, the same methodology applied in different markets has been found to exhibit some predictive power (Kellard, Nankervis, and Papadimitriou 2010). Due to the different results, the usefulness of financial ratios for forecasting future returns was further investigated on a large scale. Lewellen analyzed the Dividend Yield over the period 1946-2000 in US markets and suggests that it has predictive power over the whole period as well as in sub-samples. Also, in direct response to the work of Goyal and Welch (Welch and Goyal 2008), Campbell and Thompson show that under the introduction of weak restrictions on the regression coefficients and forecasts, out-of-sample explanatory power is evident (Campbell and Thompson 2008). On the other hand, Hjalmarsson shows that such ratios do not hold up consistently in international comparisons, while other near-market variables do (Hjalmarsson 2010).

***Hypothesis 1****: The models examined by Goyal and Welch are still not robust in the true out-of-sample test.*

In extensive research, more and more novel methods have been sought to improve the equity premium predictability. For example, the influence of physical gold demand as a proxy for the equity risk premium has been investigated and a positive relationship between gold demand and future stock returns can be observed, thereby adding explanatory power to a models consisting of dividend yield and other variables (Baur and Löffler 2015). Nonejad finds that oil prices as a selec-

tor for choosing among regression models increases prediction accuracy by 10 percent as compared to individual models (Nonejad 2021). With the entrance of technical analysis into the repertoire of financial markets, studies on this topic also find relations to the prediction of the equity premium. In their study, Baetje and Menkhoff uncover the ability of technical indicators to predict the equity premium out-of-sample over a long period of time, while their selection of economic indicators start to lose their predictive power after the 1970s (Baetje and Menkhoff 2016). Neely, Rapach, and Zhou independently do find similar results (Neely et al. 2014).

***Hypothesis 2****: The use of other data utilized by currently active fund managers or looking for variables in other markets can successfully predict the equity premium.*

With the advancement to more powerful computers, it became possible to introduce more sophisticated models capable of recognizing non-linear relationships in data which were previously not easily observed. Ensemble models (Tsai et al. 2011) and deep learning (Abe and Nakayama 2018; Feng, He, and Polson 2018) have been tested again and again to predict asset returns and the equity premium. However, such models, the latter in particular, are susceptible to overfitting, random behavior and crowding out effects, as many models start to loose validity after 2010 (Fischer and Krauss 2018). This is partly due to the fact that many are based on, widely available and noisy, historical price data. The literature is still open, however, applying such new models to financial ratios and macroeconomic data in order to forecast the equity premium.

***Hypothesis 3****: Complex models are able to detect possible non-linear relations in the data and use it to predict the equity premium more accurately.*

# 4 Methodology

In the following the methodology of the analysis is presented. Appendix A Figure 10 presents the program structure.

## 4.1 Data

The dependent variable throughout the study is the equity premium. It is calculated from the total rate of return minus the short term interest rate applicable at the time, as employed in Welch and Goyal 2008.

### 4.1.1 Dataset 1 & 2: Welch and Goyal 2008

To investigate the first hypothesis, the original data of Goyal and Welch made available by Amit Goyal on his website is used (Goyal). Among others, the data includes aggregate earnings and aggregate dividends of the SP500 index, as well as return variance and yields. The individual time series are listed in Table 1. The data is available both in the original dataset through 2005 and continuing through the end of 2020. In the following, this data is referred to as *"GW"*. It is reported in monthly, quarterly, and annual frequencies since at least May 1937.

| Ticker | Name & Description | Availability since | Used since | Frequency |
|---|---|---|---|---|
| *Index* | S&P 500 index | 01.1871 | 01.2002 | Monthly |
| *E12* | S&P 500 aggregate earnings. | 01.1871 | 01.2002 | Monthly |
| *D12* | S&P 500 aggregate dividends. | 01.1871 | 01.2002 | Monthly |
| *b/m* | S&P 500 book to market ratio. Book Value of Index Components / Market Value of Index Components | 03.1921 | 01.2002 | Monthly |
| *tbl* | Treasury Bills Rate. U.S. Yields On Short-Term United States Securities, Three-Six Month Treasury Notes and Certificates, Three Month Treasury (from 1920 to 1933). The 3- Month Treasury Bill: Secondary Market Rate (1933 to 2005/2020) | 02.1920 | 01.2002 | Monthly |
| *AAA* | Corporate Bond Yields - AAA Rated | 01.1919 | 01.2002 | Monthly |
| *BAA* | Corporate Bond Yields - BAA Rated | 01.1919 | 01.2002 | Monthly |
| *lty* | Long Term Yield: U.S. Yield On Long-Term United States Bonds | 01.1919 | 01.2002 | Monthly |
| *ntis* | Net Equity Expansion: Ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks. | 12.1926 | 01.2002 | Monthly |
| *Rfree* | Risk-free rate: Treasury-bill rate from 1920 to 2005/2020 | 01.1871 | 01.2002 | Monthly |
| *infl* | Inflation: Consumer Price Index (All Urban Consumers) | 02.1913 | 01.2002 | Monthly |
| *ltr* | Long Term Rate of Returns | 01.1926 | 01.2002 | Monthly |
| *corpr* | Corporate Bond Returns | 01.1926 | 01.2002 | Monthly |
| *svar* | S&P 500 Stock Variance: Sum of squared daily returns | 02.1885 | 01.2002 | Monthly |
| *csp* | Cross-sectional premium: Relative valuations of high- and low-beta stocks | 05.1937 | 01.2002 | Monthly |

**Table 1:** ***Goyal and Welch Data****:*
*The data consists only of US data. It has been collected by Welch and Goyal 2008 from various sources. Note that "csp" is being dropped due to the fact that it is not collected after 2002 anymore. The descriptions of the variables are partly taken from Welch and Goyal 2008.*

### 4.1.2 Dataset 3: Europe Proxies for Welch and Goyal 2008

Data was also collected from Europe on proxies for the variables used in (Welch and Goyal 2008). In the following, this data will be referred to as *"GW-EU"*. Not all time series are available or of sufficient length. For example, proxies for AAA and BAA corporate bond yields are not included because they were not reasonably long enough to be reliable. The data was compiled through the provider Bloomberg and the individual time series are described in Table 2. All data from this dataset has been available since at least January 2002 and is available in monthly frequency. The European EuroStoxx 600 was taken as a comparison to the United States' S&P 500, as the number of assets is similar. However, they differ with respect to their composition, as the S&P 500 includes the largest U.S. companies, while the EuroStoxx 600 is constructed to include large cap, medium cap and small cap companies, thus covering 90 percent of the free-float market capitalization in the European market. This could effect predictive outcomes of the models.

| *Ticker* | *Name & Description (Bloomberg Ticker)* | *Available since* | *Used since* | *Frequency* |
|---|---|---|---|---|
| *Index* | EuroStoxx 600 Last Price. (SXXP Index PX LAST) | 12.1986 | 01.2002 | Monthly |
| *E12* | EuroStoxx 600 Trailing aggregate Earnings. Sum of the most recent 12 months, four quarters, two semiannuals or annual earnings per share (EPS) (SXXP Index Trail 12M EPS) | 01.2002 | 01.2002 | Monthly |
| *D12* | EuroStoxx 600 Dividend per Share 12 Months (Gross). Calculated by adding the gross dividend amounts for all dividend types that have gone ex over the past 12 months based on teh dividend frequency (SXXP Index DVD SH 12M) | 01.2002 | 01.2002 | Monthly |
| *bm* | Book to Market Ratio. Is calculated based on the EuroStoxx 600 Price to Book Ratio. The inverse yields the book to market ratio. (SXXP Index PX TO BOOK RATIO) | 01.2002 | 01.2002 | Monthly |
| *tbl* | ICE LIBOR EUR 3M (EE0003M Index) | 11.1989 | 01.2002 | Monthly |
| *lty* | Euro Generic Govt Bond 10 Year (GECU10YR Index) | 01.1993 | 01.2002 | Monthly |
| *Rfree* | EURUSD 3 Month Forward Implied Yield. Implied yields are annualized interest rates for a given currency and tenor (EURI3M Curncy) | 12.1998 | 01.2002 | Monthly |
| *infl* | Harmonized index of consumer prices (HICP), is a measure of prices paid by consumers for a market basket of goods and services (ECCPEMUY Index) | 01.1997 | 01.2002 | Monthly |
| *svar* | Stock Variance. Is calculated by taking the variance on a 12 month rolling window of the Index variable. | 12.1986 | 01.2002 | Monthly |
| *ntis* | Net Equity Expansion. Is calculated by taking the ratio of ECB Euro Area Quoted Shares Net Issues Total (EUQSNITL Index) and the EuroStoxx 600 Current Market Cap (SXXP Index MKT CAP) | 05.2000 | 01.2002 | Monthly |
| *eqis* | Percent Equity Issuing. Is calculated by taking the ratio of the ECB Euro Area Quoted Shares Non Financial Corp Net Issues (EUQSNFIN Index) and the EuroStoxx 600 Current Market Cap (SXXP Index MKT CAP) | 05.2000 | 01.2002 | Monthly |

**Table 2:** ***Europe proxy data****:*
*Most data could be obtained, however, not all data was sufficiently available. The ICE LIBOR 3 month rate was used as the proxy to the 3 month treasury bill. This is discussable, as LIBOR is an interbank rate and therefore prices a certain risk premium. However, a 3 month bill for the European Union was not available until this point and using Germany as a proxy would be problematic as it would underweight the risks exhibited from the other European states. As the LIBOR is highly correlated and adds a risk premium, this is deemed a good proxy for the European bill rate. The same reasoning counts for the risk free rate, which is the 3 month forward implied yield by the EURUSD. While not the rate itself, it is highly correlated and shows similar rates to the german 3 month bills. The equity percent issuing metric disregards issuing by Financials, it is considered a good enough proxy. All data is collected via the Bloomberg Terminal.*

### 4.1.3 Dataset 4: Macro Dataset

To investigate the second hypothesis, whether other variables that current investors pay attention to have predictive power, data was collected based on interviews with fund managers from a variety of asset management firms. Care was taken to focus on the U.S. market during the interviews, as this is where the data is most obtainable in the most complete fashion, making the results easy to compare with previous papers and quickly replicable. All economic data is available through public APIs. Futures price data was retrieved from Bloomberg. Spreads are directly calculated from the data in order to for example map the yield curve. In the following, this data is referred to *"MAC"*. The variables most frequently selected by the Feature Selector (see 4.2.1) are described in Table 3. While all data is available since 2010, care was taken in the selection process described later to select only time series available since at least 2002 (see Table 3). Since the data is publicly available, an update function for the economic data was included to allow for the verification of results at a later date with a longer data set.

| Ticker | Name & Description | Available since | Used since | Frequency |
|---|---|---|---|---|
| NU_P_TY1_Comdty | 10-Year US Treasury Note Futures, Generic 1st future | 03-05-1982 | 01.2002 | Daily |
| PCE | Personal Consumption Expenditures | 01-01-1970 | 01.2002 | Monthly |
| PERMIT | New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units | 01-01-1970 | 01.2002 | Monthly |
| NU_P_GC1_Comdty | Gold Futures, Generic 1st future | 02-01-1975 | 01.2002 | Daily |
| CPALTT01USM657N | Consumer Price Index: Total All Items for the United States | 01-01-1970 | 01.2002 | Monthly |
| ISRATIO | Total Business: Inventories to Sales Ratio | 01-01-1992 | 01.2002 | Monthly |
| ACOGNO | Manufacturers' New Orders: Consumer Goods | 01-02-1992 | 01.2002 | Monthly |
| NU_P_CL1_Comdty | Crude Oil Futures, Generic 1st future | 30-03-1983 | 01.2002 | Daily |
| ICSA | Initial Claims | 03-01-1970 | 01.2002 | Weekly |
| M2SL | M2 | 01-01-1970 | 01.2002 | Monthly |
| DGS30_DGS10 | 30-10 US Yield Spread. Calculated from the difference between Market Yield on U.S. Treasury Securities at 30-Year Constant Maturity and Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity | 15-02-1977 | 01.2002 | Daily |
| NU_P_TU1_Comdty | 2-Year US Treasury Note Futures, Generic 1st future | 25-06-1990 | 01.2002 | Daily |
| NU_P_US1_Comdty | U.S. Treasury Long Bond Futures, Generic 1st future | 22-08-1977 | 01.2002 | Daily |
| NEWORDER | Manufacturers' New Orders: Nondefense Capital Goods Excluding Aircraft | 01-02-1992 | 01.2002 | Monthly |
| NU_P_DX1_Curncy | U.S. Dollar Index Future, Generic 1st future | 07-04-1986 | 01.2002 | Daily |
| RSXFS | Advance Retail Sales: Retail Trade | 01-01-1992 | 01.2002 | Monthly |
| IC4WSA | 4-Week Moving Average of Initial Claims | 03-01-1970 | 01.2002 | Weekly |
| DSPIC96 | Real Disposable Personal Income | 01-01-1970 | 01.2002 | Monthly |
| CCSA | Continued Claims (Insured Unemployment) | 03-01-1970 | 01.2002 | Weekly |

**Table 3:** ***Macro data****:*
*The MAC dataset consists of a combination of economic indicators, leading and lagging, and market prices. It also contains calculated spreads. This table represents the chosen indicators which are used by the models. The analysis employs a feature selection algorithm and thus the indicators are selected based on statistical relations to the predictor variable.*

## 4.2 Experimental Procedure

An underlying assumption of Hypothesis 3 is that there is a non-linear relationship between the variables and the equity premium. For this reason, several models are tested that can incorporate this relationship on each dataset. There are 56 tests performed on each variable to be able to work out model differences and 1316 individual tests are produced. The tests vary

1. in model selection between an OLS, a LASSO, a Ridge Regression, a Support Vector Regressor, a Random Forest Regressor, a Deep Neural Network and an LSTM Network (see 2),

2. in datasets between GW (4.1.1) to 2005, GW to 2020, GW-EU (4.1.2), and MAC (4.1.3) (see 4.1),

3. and in frequency between monthly sample rate and annual sample rate. This study particularly focuses on the monthly sample rate, as it incorporates the annual data and leaves more impressions per test for the models to train.

### 4.2.1 Collecting Data and Feature Preparation

All data are loaded into the program and checked first for non-valid values, all time series have a common start and end time and values, which are missing in the time span between, are inserted over the forward-fill method. Values, which are not present at a time point, are copied forward from the last time point at which they were present. Care is taken that this is not done for variables that are based on any derivative of a time series (e.g. returns or rate of change in inflation), where a zero value is inserted.

In the next step, several features are engineered. For the Welch and Goyal 2008 replication, several new features are calculated from the GW (4.1.1) data, such as the Dividend Price Ratio or the Term Spread (see Table 4). For the European GW-EU (4.1.2) data, the same process is undertaken. In the MAC (4.1.3) dataset, the price data is converted to historical log-returns. Prior

to this, these are rebased as certain futures, for example Natural Gas, have different volatility when the historical price data is roll-adjusted (see Figure 11).

| Ticker | Name & Description | Frequency |
|---|---|---|
| index_div | Adds the rolling dividends (D12) back to the index (Index) | Monthly |
| dp | Dividend Price Ratio. Difference of the logarithm of the dividends (D12) and the log of prices (Index) | Monthly |
| ep | Earnings Price Ratio. Difference of the logarithm of the earnings (E12) and the log of prices (Index) | Monthly |
| de | Dividend Earnings Ratio. Difference of the logarithm of the dividends (D12) and the log of earnings (E12) | Monthly |
| dy | Dividend Yield. Difference of the logarithm of the dividends (D12) and the log of lagged prices (Index) | Monthly |
| tms | Term Spread. Difference of the Long Term Yield (lty) and the short term yield (tbl) | Monthly |

**Table 4:** ***Engineered Features****. Descriptions partly taken from Welch and Goyal 2008.*

To replicate the work of Goyal and Welch, all features from all datasets, GW (4.1.1), GW-EU (4.1.2), and MAC (4.1.3), are first inserted into models individually and tested. In this way, the impact of each variable on the equity premium is measured. In the MAC dataset, feature selection is done beforehand. This automated selection process combines univariate feature selection - variables are selected based on univariate statistical tests: SelectKBest (Pedregosa et al. 2011b) - with an extremely randomized tree regressor - an ensemble machine learning strategy that evaluates feature importance towards the dependent variable: ExtraTreeRegressor (Pedregosa et al. 2011a). Both processes select ten variables each and these are combined together, leading to the selection in table Table 3. This approach avoids that the models are later trained with already insufficiently diverse selected data and it is less likely that the models fall into local optima.

### 4.2.2 Models

Two different approaches are adopted to deal with the complexity of the models. For the regression and machine learning models, the rolling out-of-sample prediction approach from Goyal and Welch is adopted (Welch and Goyal 2008). For the neural networks, a time-series cross validation split is applied to reduce the training time. The next period of the equity premium is forecasted.

**Regression and ”non-deep” Models:** In the first step, the regression model of Goyal and Welch is used. The same procedure is then used to test LASSO, Ridge regressions, Support Vector Ma-

chines, and Random Forest regressors.

Goyal and Welch perform a procedure that lets them compare the insample performance (IS) of the model with the rolling out-of-sample performance (OOS). Thereby they subtract the prediction of the ALTERNATIVE from the NULL. NULL is the average equity premium of the full period in the IS period and the average equity premium of the expanding period in the OOS period. ALTERNATIVE is the residual of the actual expressions in the IS period and the rolling regression prediction in the OOS period. For performance, IS and OOS, the difference is the respective null hypothesis, the variable minus its average, and the respective alternative hypothesis, the residuals of the model selected in this case. The resulting graph demonstrates when the forecast of a regression outperforms the mean-prediction and can be used successfully (see Figure 13). The exogenous variables are subject to a lag of one period. For the out-of-sample performance, a varying 5 or 20 period warm-up time is integrated, depending on the length of the frequency. Each period, both the null hypothesis and the alternative hypothesis are recalculated. In the next step, the key performance indicators, MSE, R2 and dRMSE, are calculated for the whole model with $\Delta RMSE = \sqrt{MSE_N} - \sqrt{MSE_A}$.

**Deep Neural Network and LSTM Model:** The equity premium is investigated using a fully connected deep neural network and an LSTM model. To avoid long and computationally intensive training times of these models, a 5-fold time series cross validation is implemented, compared to the rolling prediction of Goyal and Welch (see Appendix A 12)[6]. MSE, R2, and RMSE are also calculated for these models.

The fully connected deep neural network consists of an input layer with 128 neurons, and a ReLU activation function (see section 2). This is followed by 3 hidden layers with 128 neurons each and also a ReLU activation function (see Figure 8). Finally, the model ends in an output layer with 1 neuron and a linear activation function. All layers are densely connected. The network is

6. For a true out of sample test, normally a holdout set should be separated. However, as Goyal and Welch did not opt for this approach, this was not implemented to make this study comparable.

configured with the mean squared error loss function and the ADAM optimizer. It is trained on 50 epochs, a batch size of 32 and a 20 percent validation split. A checkpoint system is built in to select and test the best model from each cross-validation episode. The LSTM model is a shallow model consisting of an LSTM layer with 64 input neurons leading into a densely connected output layer with 1 neuron and a linear activation function. The model is also configured with a mean squared error loss function and the ADAM optimizer. To make the DNN and LSTM comparable in performance, the model is also trained with 50 epochs, batch-size of 32, and a 20 percent validation set. The training set was configured in such a way, that 3 previous periods of the independent variable are predicting the next dependent impression.

# 5 Results and Discussion

In the following the results of the models are shown and discussed. GW is shown in blue, GW-EU is shown in yellow, MAC is shown in green.

## 5.1 Results

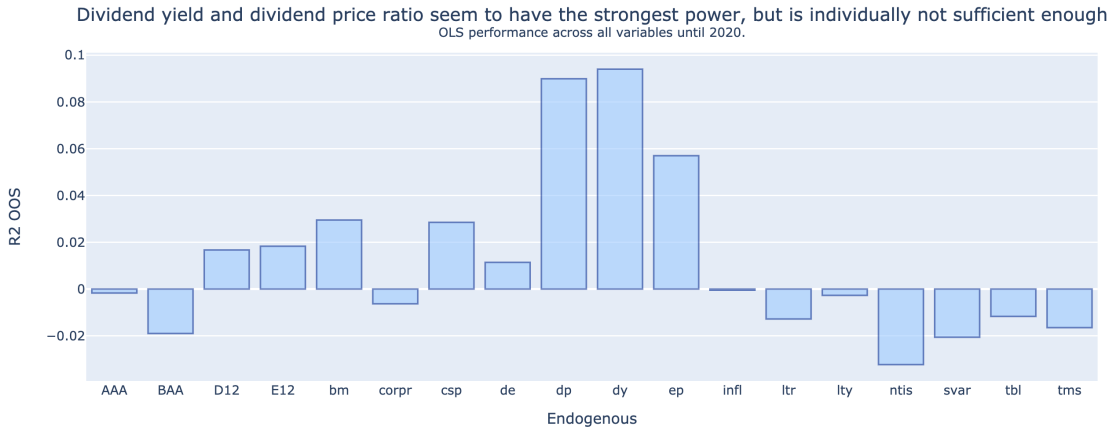


**Figure 2:** ***OLS performance over out-of-sample period:*** *The graph shows the $R^2$ performance of each variable in an OLS model against the equity premium. While dividends seem to yield better linear results than other variables to the equity premium, it is not majorly explanatory for performance.*

Over 18 years, in the out-of-sample period of Goyal & Welch, only dividend price ratio and

dividend yield, as well as earnings-price ratio prove to be somewhat predictive of the equity premium. However, all are below 10 percent $R^2$ and thus not really meaningful as linear predictors. Hypothesis 1 can thus be confirmed, as the models still lack robust out-of-sample predictive power.


**Figure 3:** ***OLS performance for other datasets:*** *The upper graph shows the performance of each variable from the european dataset in an OLS model. Interestingly the market can largely be explained by interest rates, inflation and dividends. The lower graph shows the performance of each variable from the macro dataset in an OLS model. None of these variables seem to be able to forecast the equity premium linearly.*

Unlike the U.S. data, these results suggest that the European variables do seem to be able to predict the European equity premium. In particular, the four variables of 12-month dividend, HCIP (infl), long-term yield (lty), and treasury bill (tbl) seem to be able to explain some of the variance in the market. Looking at the macro variables, we can see that none of them is able to predict the


**(a)** *OLS lty* **(b)** *OLS tbl*

**(c)** *OLS infl* **(d)** *OLS bm*

**(e)** *OLS de* **(f)** *OLS dy*

**Figure 4:** ***Insample and out-of-sample performance of GW-EU variables between 2002 and 2020:*** *These charts are made in the same style as Figure 13. The lines are calculated by taking the difference of the NULL and the ALTERNATIVE, as explained in 4.2.2. A rising line means the ALTERNATIVE is has a smaller error and therefore outperforms the NULL meaning the model is performing over time. When OOS starts to outperform IS, this hints towards unstable, very time-specific optimal fits of the regression where the smaller train set creates a better model than the full dataset. As such, all models are or start to be unstable.*

equity premium. Market data, i.e. yield curve, gold or dollar indicate the highest performance, albeit also with negative values. Hypothesis 2 can thus be partially accepted. While looking at modern alternative macro data may not have linear predictive power for the equity premium, there are indications that the equity premium may be predictable in other geographies.

Basically, one can first state that OLS and Ridge constitute good base models. If these do not

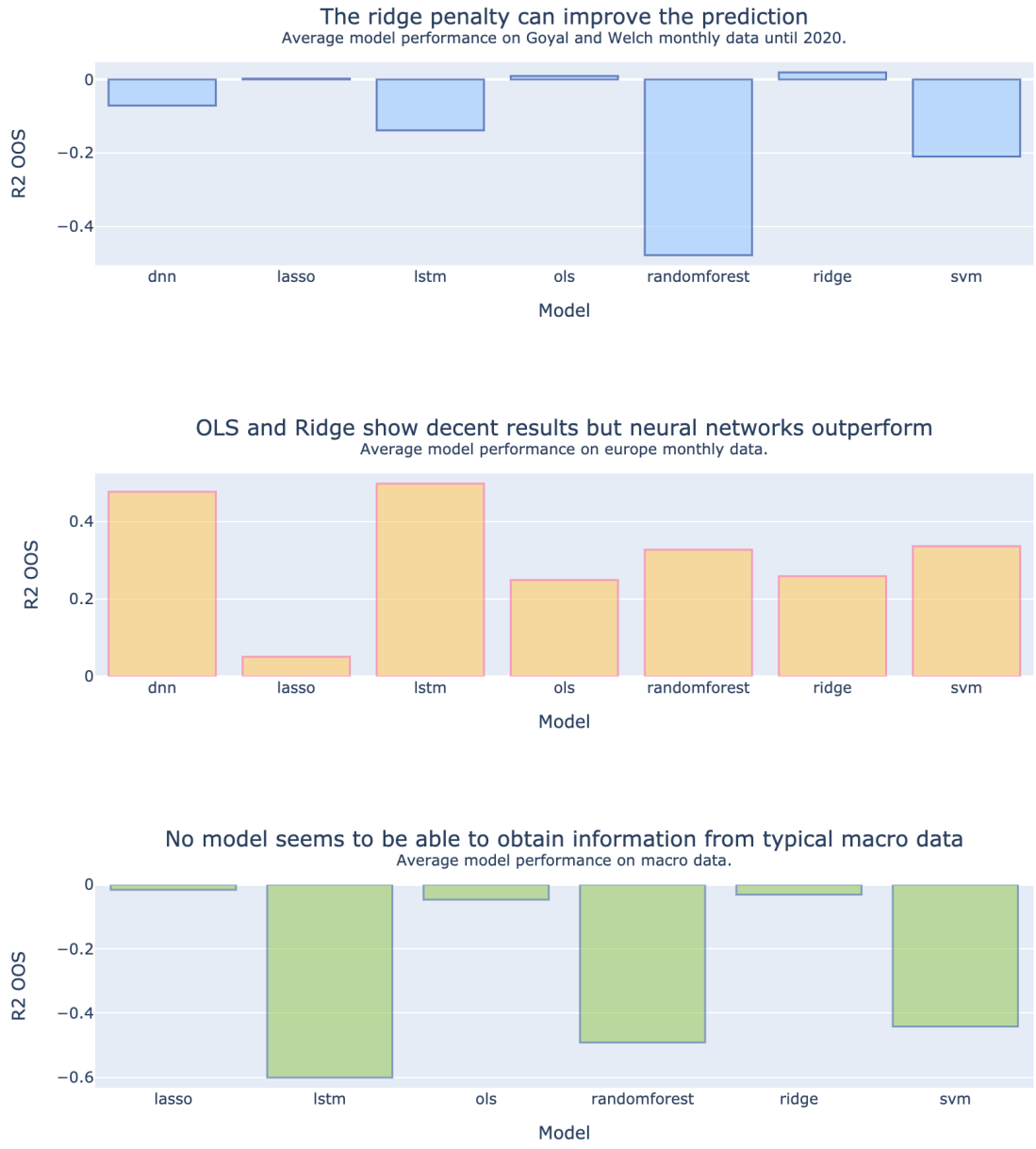


**Figure 5:** ***Model performance in each dataset:*** *These graphs picture the average out-of-sample $R^2$ of the model-architectures across all variables in each dataset. The first graph shows the GW data until 2020 on all different models. A slightly better performance of the ridge regression is noticable. Complex models do not pick up on any of the underlying relations. The second graph shows the GW-EU data and a clear improvement with complex models is noticable. Notice also, how LSTM is still outperforming the DNN model, although by a slide margin. The third graph shows the MAC data and underscores that no model was able to find relations in the data, linear or non-linear. The DNN was to be taken out to make the graph more comparable.*

show any results at all, more complex models are unlikely to be successful either. Worth mentioning is the tendency of Random Forest models to perform poorly, presumably because they do not find the optimal number of splits in the trees due to the non-categorized data. Hypothese 3 can be rejected

to a certain degree. While more complex models are able to take advantage of more non-linear relationships in data and be more accurate, somewhat of a statistical relationship is necessary to be present and this can already be assessed using simple models. From these results it can be concluded that complex models are able to generalize better on different data distributions compared to simple models.

| | **dnn** | **lasso** | **lstm** | **ols** | **randomforest** | **ridge** | **svm** | **Average** |
|---|---|---|---|---|---|---|---|---|
| **D12** | 0.0562 | 0.0527 | 0.2971 | 0.4031 | 0.4531 | 0.3541 | 0.3986 | 0.2878 |
| **E12** | 0.1011 | 0.0281 | 0.2237 | -0.0014 | 0.4820 | 0.0648 | 0.0976 | 0.1423 |
| **bm** | 0.4808 | 0.0000 | 0.5477 | 0.0607 | 0.2699 | 0.0398 | 0.4243 | 0.2605 |
| **de** | 0.4682 | 0.0000 | 0.5289 | 0.0601 | 0.5183 | 0.2330 | 0.4907 | 0.3285 |
| **dp** | 0.3372 | 0.0000 | 0.3859 | 0.0217 | 0.2072 | 0.0192 | 0.2601 | 0.1759 |
| **dy** | 0.3712 | 0.0000 | 0.3908 | 0.0396 | 0.2501 | 0.0283 | 0.2522 | 0.1903 |
| **ep** | 0.3492 | 0.0000 | 0.4033 | 0.0201 | 0.4092 | 0.1741 | 0.3907 | 0.2495 |
| **eqis** | 0.4843 | 0.0000 | 0.5327 | -0.0180 | -0.0884 | 0.0000 | -0.1209 | 0.1128 |
| **infl** | 0.5171 | 0.0000 | 0.5600 | 0.2920 | 0.3947 | 0.3372 | 0.2816 | 0.3404 |
| **lty** | *0.6743* | *0.1653* | *0.6744* | *0.8630* | *0.8456* | *0.8580* | *0.8419* | *0.7032* |
| **ntis** | 0.4978 | 0.0000 | 0.5363 | -0.0181 | 0.0530 | 0.0000 | 0.1099 | 0.1684 |
| **svar** | 0.4842 | 0.0000 | 0.5412 | -0.0036 | 0.0702 | 0.0000 | -0.0950 | 0.1424 |
| **tbl** | *0.7538* | *0.1284* | *0.7685* | *0.8201* | *0.9211* | *0.7998* | *0.8837* | *0.7251* |
| **tms** | 0.3592 | 0.0000 | 0.2829 | 0.0287 | 0.2051 | 0.0144 | 0.2548 | 0.1636 |
| **Average** | 0.4239 | 0.0267 | *0.4767* | 0.1834 | 0.3565 | 0.2088 | 0.3193 | - |

**Table 5:** ***Europe Model Results****:*
*Showing the $R^2$ OOS of all models on the GW-EU data set. The interest rates are marked in italic. Not only does LSTM seem to work some relation out, it is relatively stable across all models.*

Considering Table 5, it can be seen that neural networks seem to be able to predict the equity premium more accurately. Focusing on the individual models, it is noticeable that neural networks perform better on average for the individual variables. For example, with a $R^2$ of 0.5477, the LSTM is better at predicting the equity premium using book-to-market ratios than an OLS model. It also beats the Ridge regression. Interestingly, however, the LSTM underperforms both models in terms of interest rates. This suggests that book-to-market ratios may have a more non-linear relationship with the equity premium, while interest rates may have a more linear relationship (see Figure 5.2). LASSO models often get to zero, which could be due to the fact that $L_1$ forces the coefficient to

zero very quickly and thus no model emerges. Ridge models just push the coefficients power lower, leading to more smoothed results (see Figure 15).

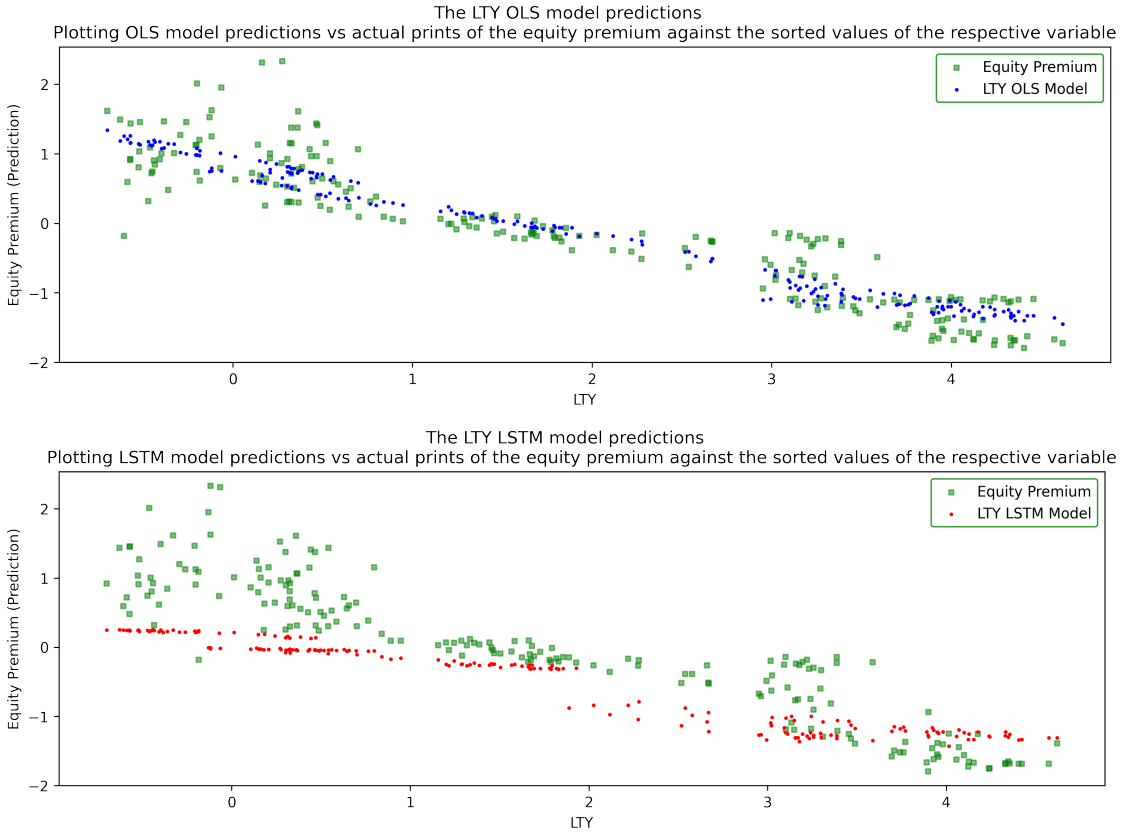


**Figure 6:** ***OLS and LSTM models on long-term yield:*** *The charts show the variable on the x-axis and the real equity premium (in green) and the model (in respective color) on top of it. Both models seem to be able to figure out the relation between equity premium and variable.*

Figure 6 displays the already mentioned performance of the models. The outperformance of the OLS is evident in this case. Appendix A Figure 14 et seq. present the other models in comparison. It is obvious from the figures that all models are more or less able to capture the relationship, as it appears to show a relatively linear nature.

## 5.2 Discussion

In a way, it is surprising that the U.S. equity premium cannot be predicted by any of the variables in GW or MAC, yet Europe exhibits this strong relation. No inconsistency has been found after multiple testing, hyper-parameter tuning and modification of the models, nevertheless the results

should be taken with some skepticism. Kellard, Nankervis, and Papadimitriou 2010 were able to find positive results on UK markets, but not as strong as in this analysis but the approach and the performance measuring employed by Welch and Goyal 2008 is specific, it is not a strategy backtest factoring in speed of information, immediate market reaction, commission, slippage, spreads and other trading costs, which can have a significant impact on profitability.

With these findings a predictive relationship between long-term yields and the equity premium is not excludable in the European market. This cannot be explained by the fact that the equity premium is calculated on the basis of short-term interest rates, otherwise this relationship would also have to exist in the U.S. markets. Also the relation persists to long-term yields (both in the U.S. and Europe 10 year yields but with different results), not only the 3 month bills. An attempt to explain the performance may be that European markets have been supplied with low interest rates since the financial crisis and the European debt crisis, which has pushed investors out of the risk curve and made equity markets more attractive. At the same time, the European market tends to be dominated by interest rate sensitive sectors such as Manufacturing or Utilities. One has to keep in mind specifically that the underlying assets S&P500 and EuroStoxx 600 are of different composition, namely the 500 largest companies versus a mixture of large-cap, mid-cap and small-cap companies. Interest rates are also related to inflation, thus providing an explanation for the generally high relationship between this variable and the equity premium.

Implications of these findings are manifold. For one generally looking at the variables on their own is not valuable to assess the equity premium. With specific look at variables like Earnings or Dividends as well, they seem not accurately tell whether a market is too expensive or not as often deemed by market commentators, at least not over the short-term prediction. Looking at the predictive power of 10 year yields in the European market of the equity premium there, this should impact asset allocations. Asset allocation models like the Black-litterman model (Black and Litterman 1991) can be used more accurately, portfolios Europe exposure tailored more to investors needs. For Monetary Policy Makers at the European Central Bank, this means that if they switch over to Yield Curve Control, a concept by which a certain yield is targeted at selected maturities

through unlimited buying or selling of the securities, they could control the equity premium even more effectively than they already do on the short end of the yield curve.

The results should be viewed critically, because they examine a relatively short period of time for macroeconomic relationships. The data covers the period from 2002 to 2020 and thus has not experienced many different market regimes during this time. Furthermore, such limited data makes complex models more susceptible to under- or overfitting. This was controlled by loss functions and prevented by cross-validation, but it cannot be completely ruled out. The low significance of the MAC variables can most likely be attributed to a high noise-to-signal ratio in the data. It can also be assumed that fund managers use these data to form an expectation about the future and not to make mechanical investment decisions based on these data.

# 6 Conclusion

This study investigated the relationship between variables already known to the research community and the equity premium out of sample, and was aimed at predicting the equity premium with the help of modern models and new data sources. The results of Welch and Goyal 2008 can still be confirmed out of sample until 2020 and modern models are not able to make a more accurate prediction using common macroeconomic data. While examining the same variables on the European market leads to the conclusion that interest rates and related inflation have an impact on the European risk premium, data selected from active fund managers could not show any relation. Future research can look deeper into the link between interest rates in the European market and the equity premium. It would also be interesting to look more closely at the difference, why Europe and the United States exhibit different relations. It is also encouraged to use other models, such as ensemble models or reinforcement learning models (Markov models), as these may be able to handle the data in a more robust style. Finally, it could be promising to investigate how exactly fund managers use macroeconomic data in their investment decisions. If this could be modelled, then a more accurate judgement about manager skill would be possible.

# A Appendix

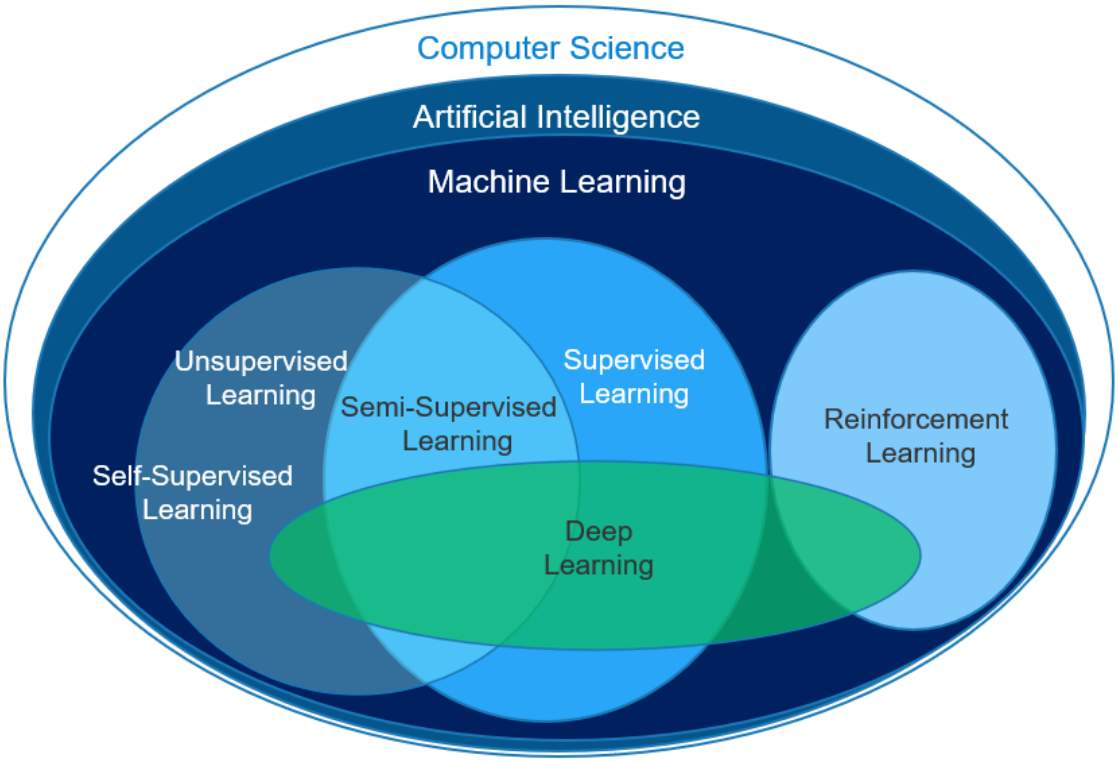


**Figure 7:** ***Terminology:*** *Machine Learning is the overarching term for data-learning techniques. Supervised Learning refers to the approach of learning the rules from the results and the data. Unsupervised Learning is finding the structure within data. Semi-Supervised learning is a mixture of the two. Self-supervised learning refers to techniques that do not require annotated data and is able to predict other parts of the data, while then generating labels in the next step. Reinforcement Learning is usually an Agent-Policy-Environment algorithm, where an "Agent" learns a "Policy" in an "Environment" by trying out policies in the environment many times (similar to learning to play a game). All of these use Deep Learning techniques, which consist of neural network structures (Figure by Mysla 2020).*

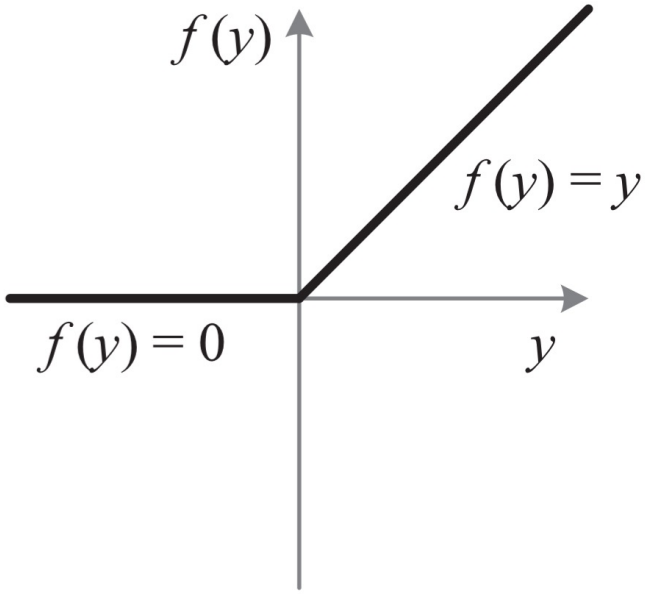


**Figure 8:** ***Rectified Linear Unit (ReLU):*** *ReLU is a type of activation function being zero if x is negative and going positive linear otherwise. Other activation functions are for example purely linear as the linear activation funciton (Figure by Colyer 2017).*

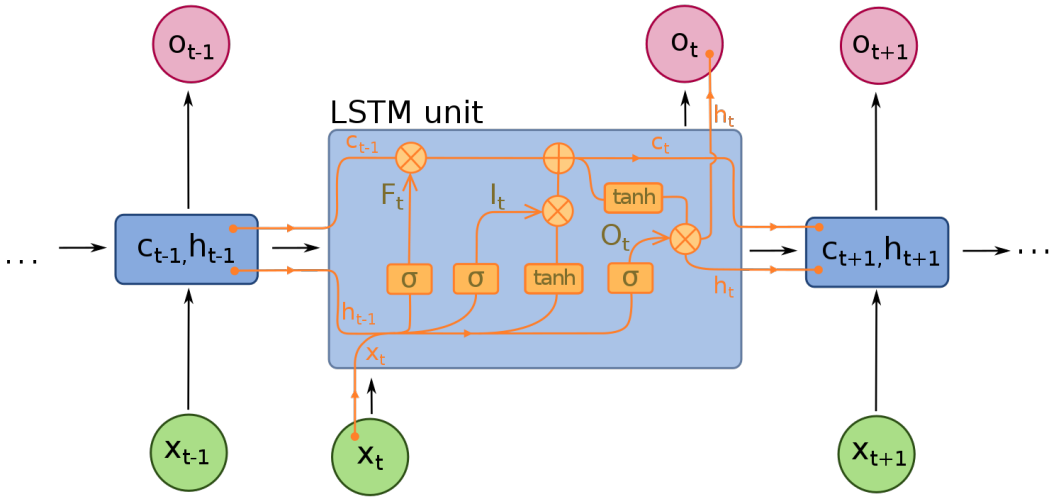


**Figure 9:** ***The structure of an LSTM Unit:*** *The gate decides on how to keep the information in the system, and is therefore able to work with sequential relationships (Figure by Foundation 2017).*

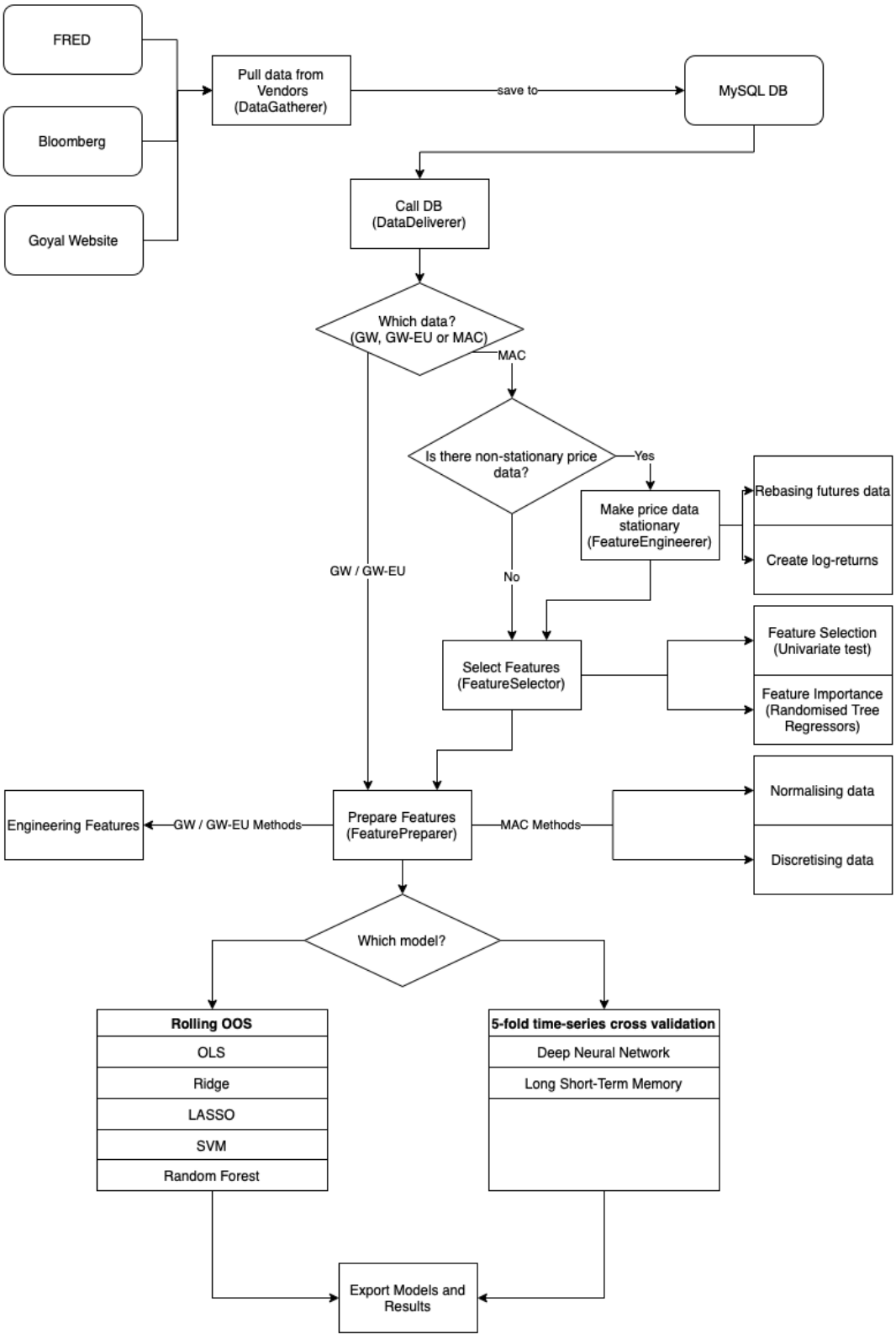


**Figure 10:** *Presenting the program structure for the analysis.*

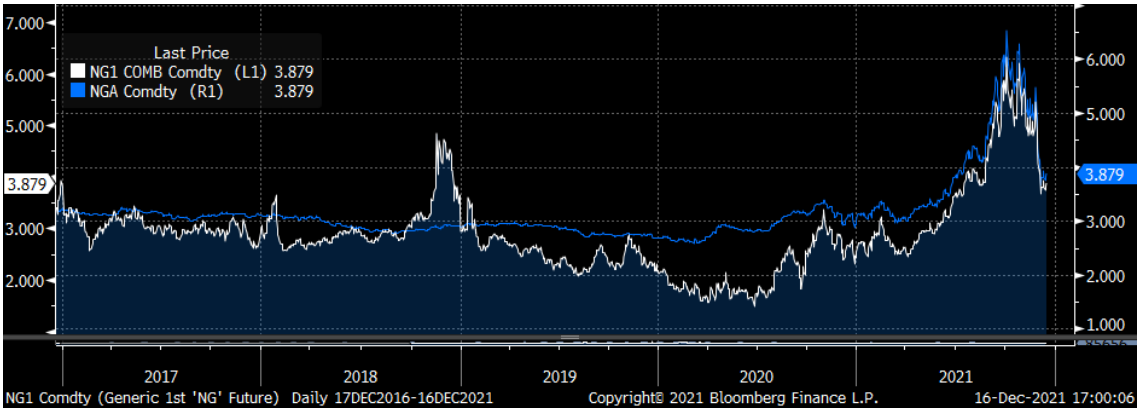


**Figure 11:** ***Difference in un-adjusted data:*** *Rebasing data when it is roll-adjusted is necessary as otherwise the volatility is distorted in the data.*

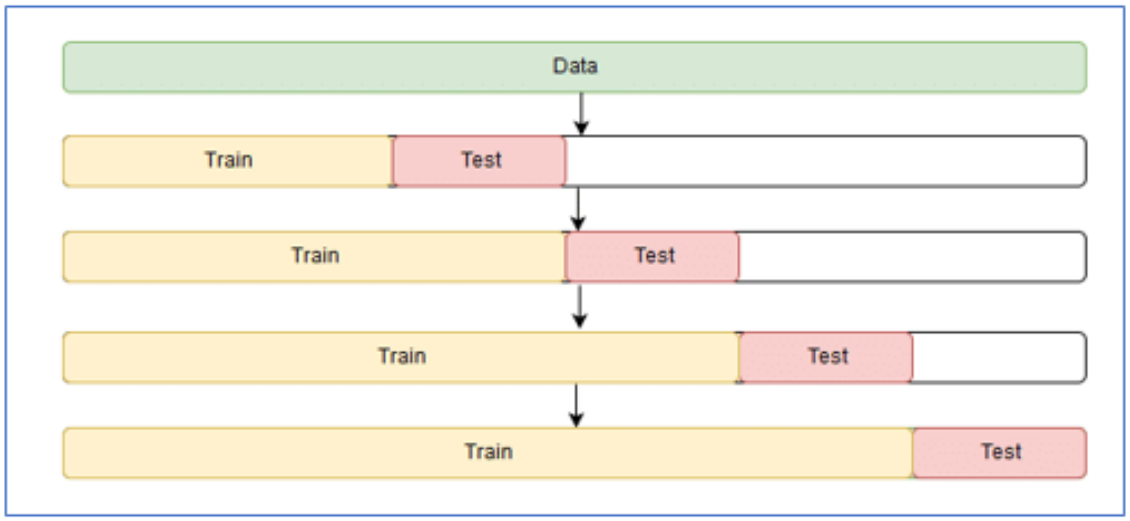


**Figure 12:** ***Time Series Cross Validation:*** *Time Series Cross Validation starts by taking a small dataset, then testing on the next couple of periods. Then in the next fold the test set is added to the training set and the next test set is selected. This process runs n-folds. Important is, that there is no shuffle and no randomizer on the dataset, which would remove time series dependencies (Figure by Vineeth, Kusetogullari, and Boone 2020).*

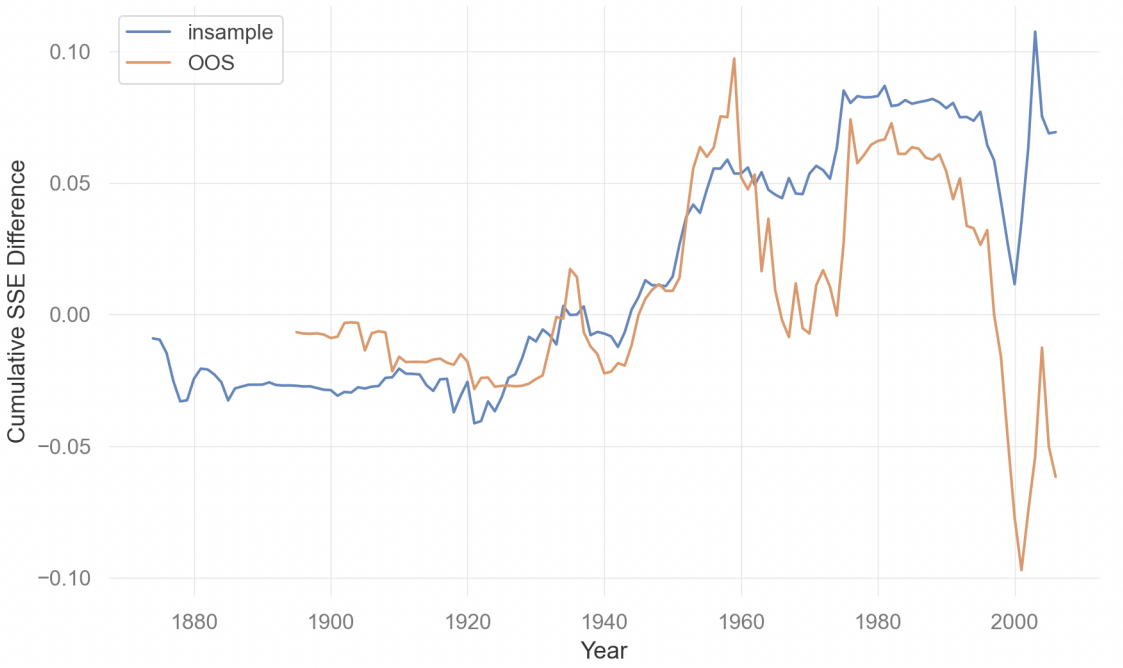


**Figure 13:** ***Annual performance of IS insignificant dividend yield****: This is the replication of the chart in Welch and Goyal 2008. It shows how the performance insample and out-of-sample differs. The lines are calculated as explained in 4.2.2. An increase in a line indicates a better performance by the model versus the NULL, which is the average equity premium.*

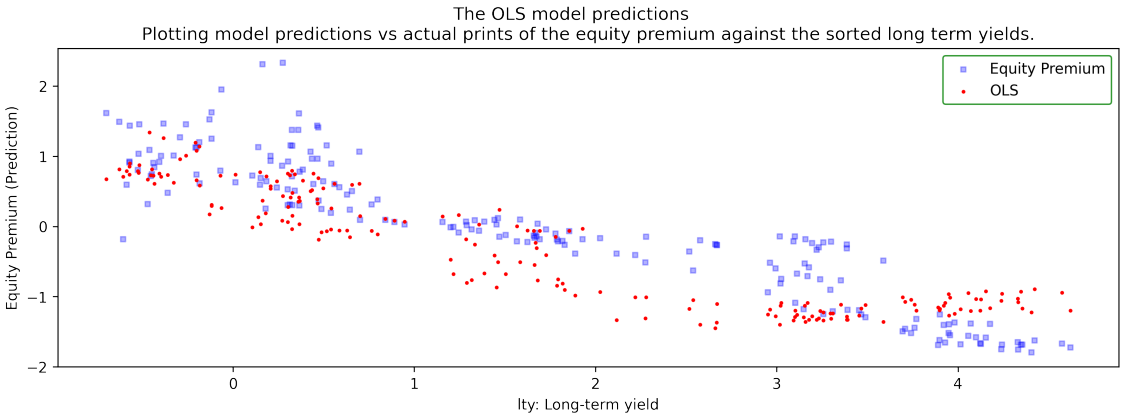


**Figure 14:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the OLS model.*

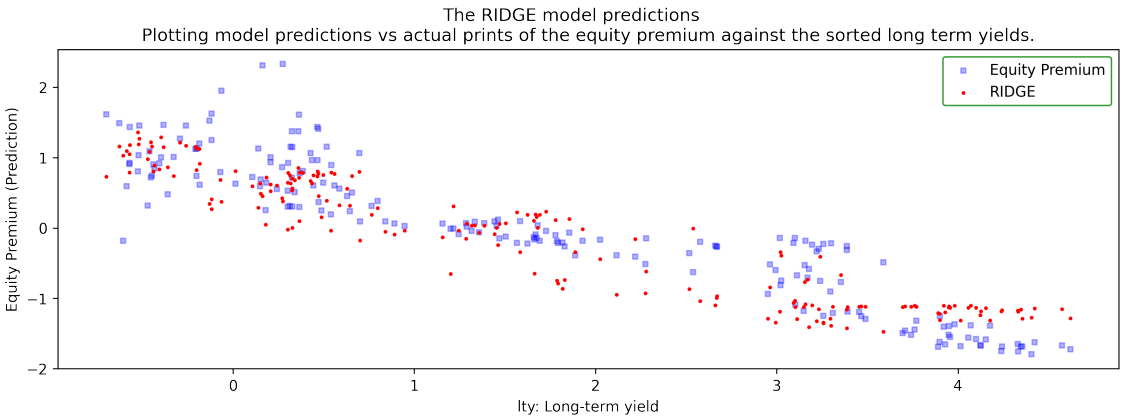


**Figure 15:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the Ridge model.*

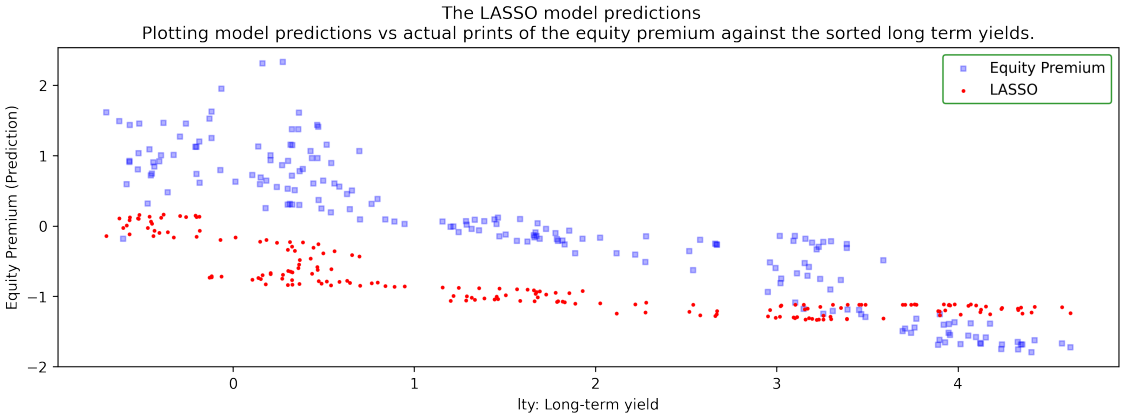


**Figure 16:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the LASSO model.*

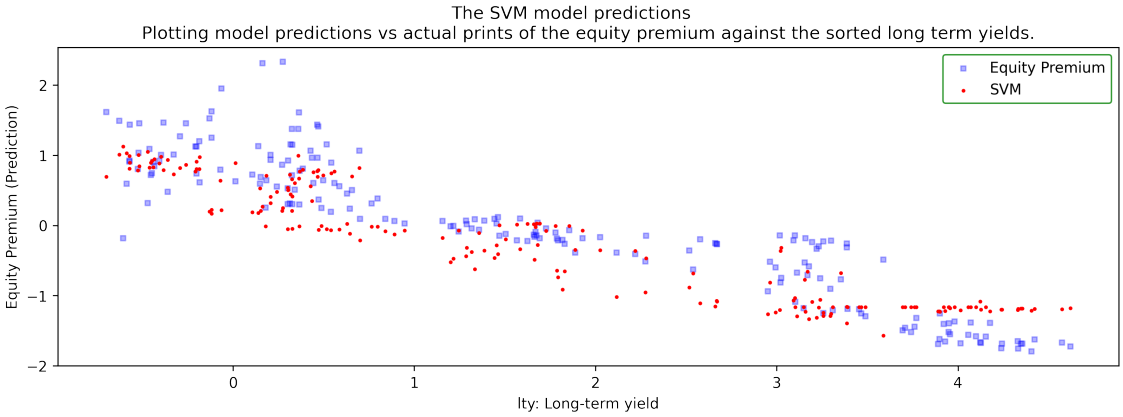


**Figure 17:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the SVM model.*

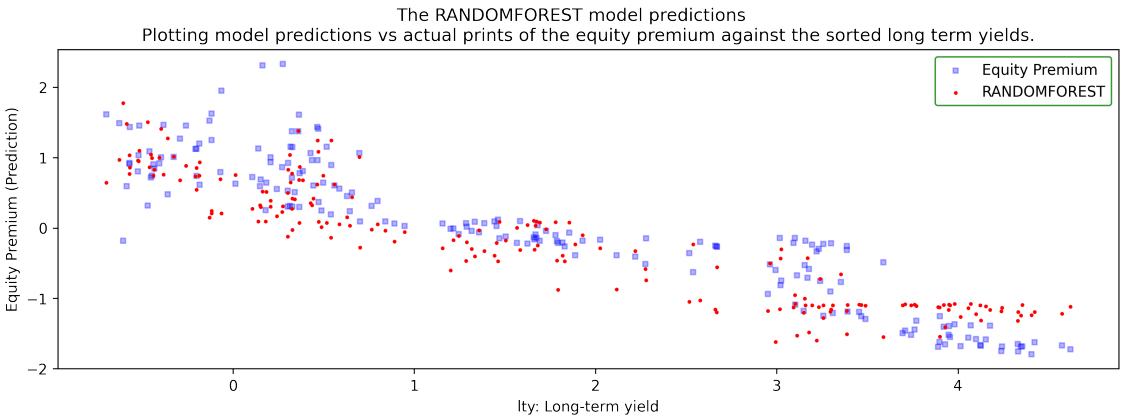


**Figure 18:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the Random Forest model.*

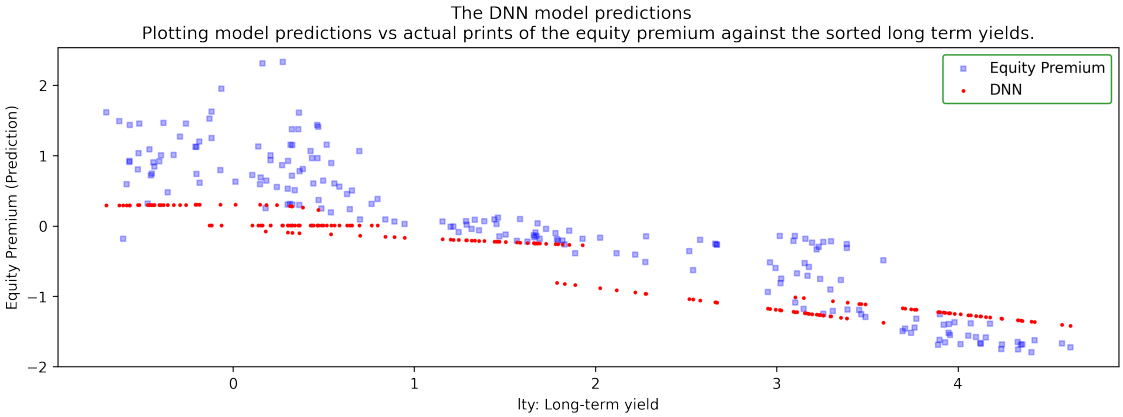


**Figure 19:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the DNN model.*

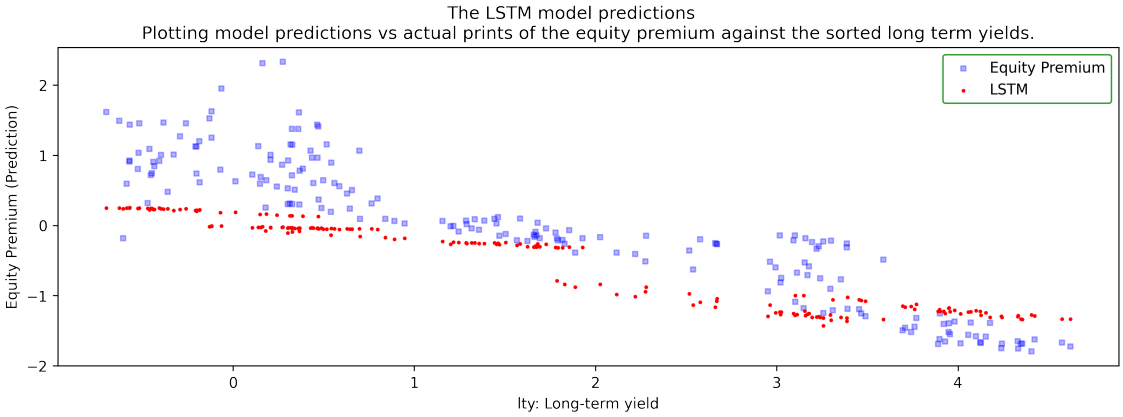


**Figure 20:** *Relation between long-term yield and equity premium prediction to the actual equity premium by the LSTM model.*

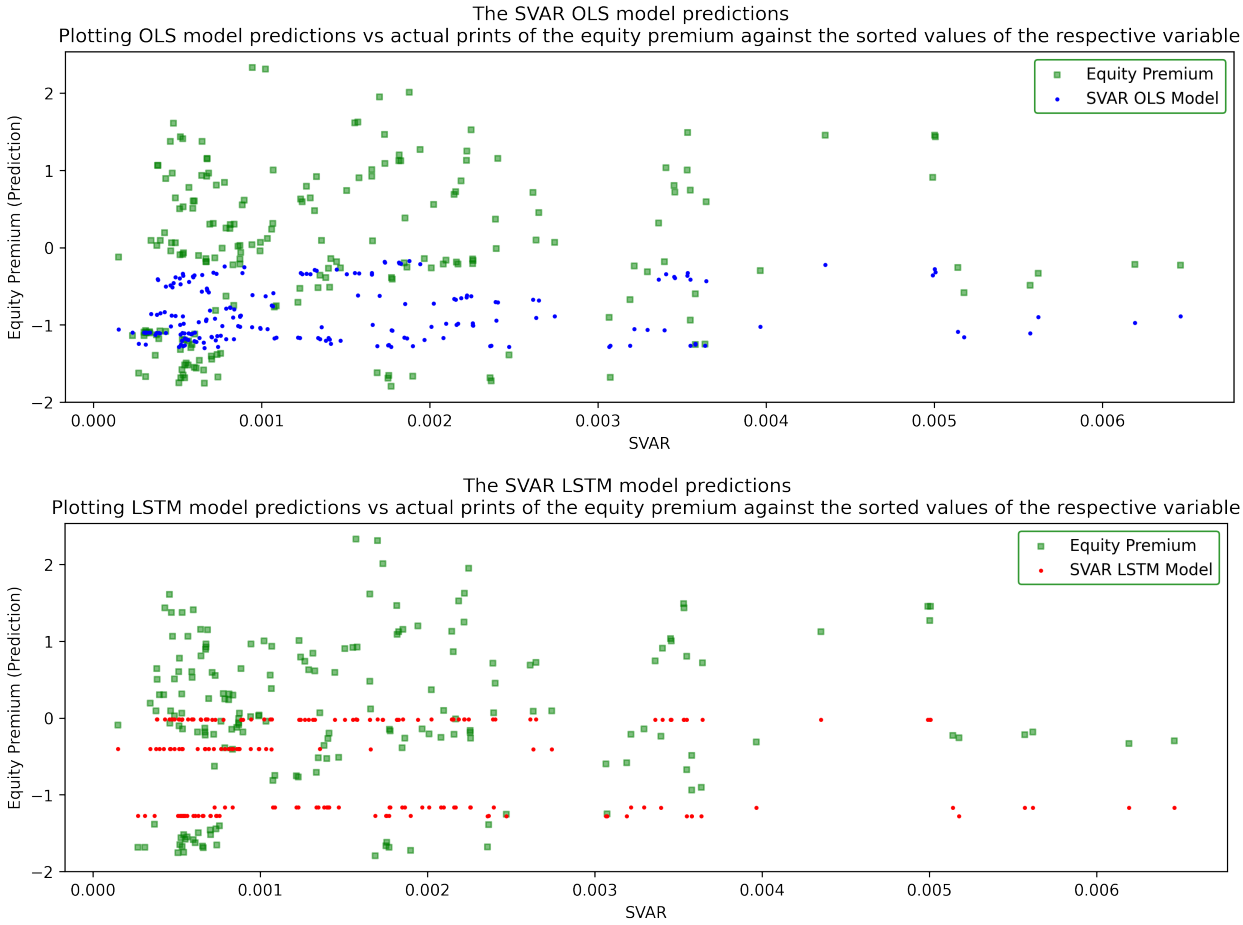


**Figure 21:** *Comparing the prediction of the equity premium between the OLS and the LSTM model based on the SVAR.*

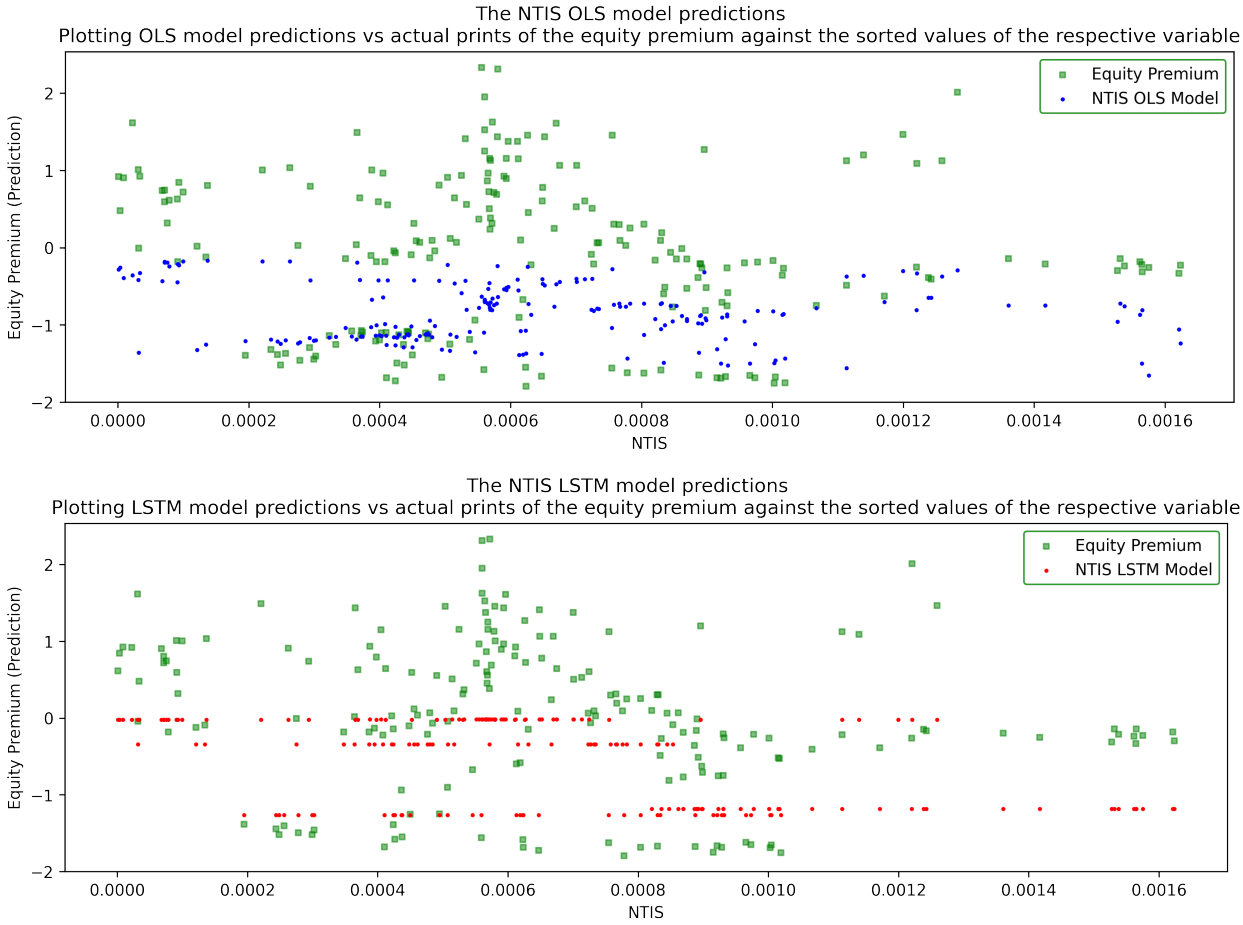


**Figure 22:** *Comparing the prediction of the equity premium between the OLS and the LSTM model based on ntis.*

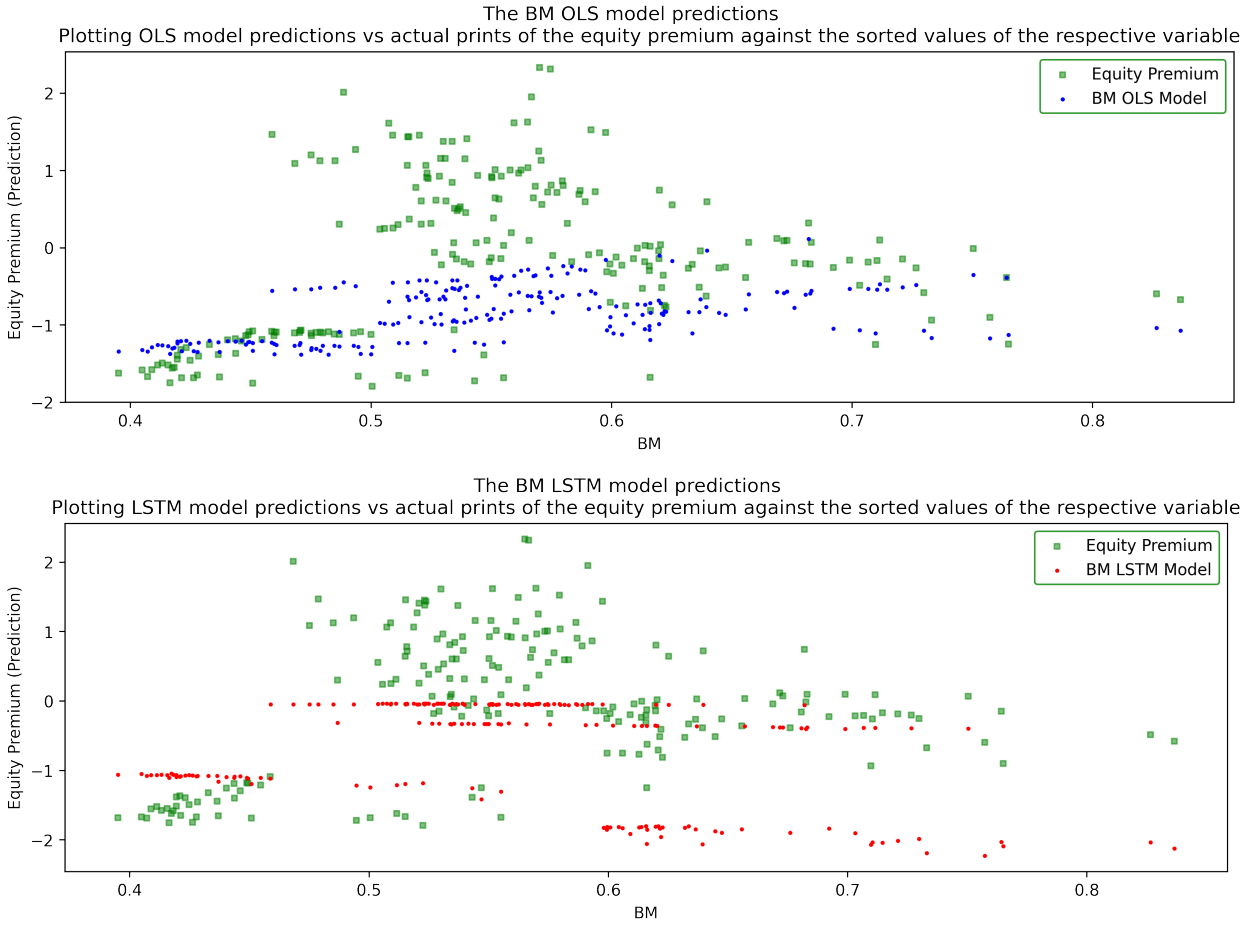


**Figure 23:** *Comparing the prediction of the equity premium between the OLS and the LSTM model based on bm.*

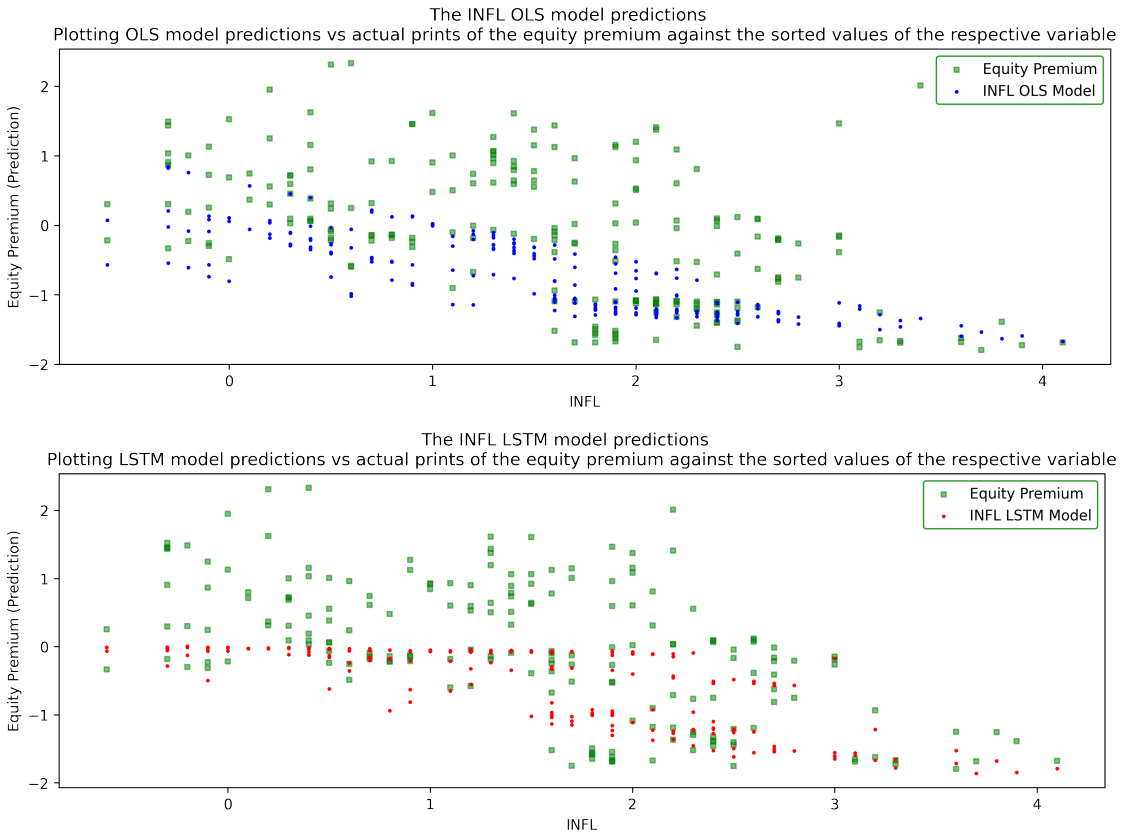


**Figure 24:** *Comparing the prediction of the equity premium between the OLS and the LSTM model based on infl.*

# References


Ammann, Manuel, and Michael Verhofen. 2009. "The Effect of Market Regimes on Style Allocation." *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.1322278. http://www.ssrn.com/abstract=1322278.

Aquilina, Matteo, Eric B. Budish, and Peter O'Neill. 2020. "Quantifying the High-Frequency Trading "Arms Race": A Simple New Methodology and Estimates." *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.3636323.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *Journal of Finance* 68 (3): 929–985. ISSN: 00221082. https://doi.org/10.1111/jofi.12021.

Baetje, Fabian, and Lukas Menkhoff. 2016. "Equity premium prediction: Are economic and technical indicators unstable?" *International Journal of Forecasting* 32 (4): 1193–1207. ISSN: 01692070. https://doi.org/10.1016/j.ijforecast.2016.02.006.

Bali, Turan G., Stephen J. Brown, and K. Ozgur Demirtas. 2013. "Do Hedge Funds Outperform Stocks and Bonds?" *Management Science* 59 (8): 1887–1903. ISSN: 0025-1909. https://doi.org/10.1287/mnsc.1120.1689.

Baltas, Nick. 2019. "The Impact of Crowding in Alternative Risk Premia Investing." *Financial Analysts Journal* 75 (3): 89–104. ISSN: 0015-198X. https://doi.org/10.1080/0015198X.2019.1600955.

Baur, Dirk G., and Gunter Löffler. 2015. "Predicting the equity premium with the demand for gold coins and bars." *Finance Research Letters* 13 (May): 172–178. ISSN: 15446123. https://doi.org/10.1016/j.frl.2015.01.007.

Belford, Geneva. 1999. *Computer science,* July. https://www.britannica.com/science/computer-science.

Bertoneche, Marc L. 1979. "Spectral analysis of stock market prices." *Journal of Banking Finance* 3 (2): 201–208. ISSN: 03784266. https://doi.org/10.1016/0378-4266(79)90015-3.

Black, Fischer, and Robert B Litterman. 1991. "Asset Allocation." *The Journal of Fixed Income* 1 (2): 7–18. ISSN: 1059-8596. https://doi.org/10.3905/jfi.1991.408013.

Bondt, Werner F. M. De, and Richard Thaler. 1985. "Does the Stock Market Overreact?" *The Journal of Finance* 40 (3): 793. ISSN: 00221082. https://doi.org/10.2307/2327804.

Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. "A training algorithm for optimal margin classifiers," 144–152. ACM Press. ISBN: 089791497X. https://doi.org/10.1145/130385.130401.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. ISSN: 08856125. https://doi.org/10.1023/A:1010933404324.

Campbell, John Y., and Samuel B. Thompson. 2008. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies* 21 (4): 1509–1531. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhm055. https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhm055.

Chollet, Francois. 2017. *Deep Learning with Python.* 1st. USA: Manning Publications Co. ISBN: 1617294438.

Colyer, Adrian. 2017. *Optimisation and training techniques for deep learning,* March. https://blog.acolyer.org/2017/03/01/optimisation-and-training-techniques-for-deep-learning/.

Copeland, B.J. 1998. *Artificial Intelligence,* July. https://www.britannica.com/technology/artificial-intelligence.

Dash, Rajashree, and Pradipta Kishore Dash. 2016. “A hybrid stock trading framework integrating technical analysis with machine learning techniques.” *The Journal of Finance and Data Science* 2 (1): 42–57. ISSN: 24059188. https://doi.org/10.1016/j.jfds.2016.03.002.

Dow, Charles, and George Charles Selden. 1920. *Scientific Stock Speculation.* 2016th ed. Creative Media Partners, LLC.

Enke, David, and Suraphan Thawornwong. 2005. “The use of data mining and neural networks for forecasting stock market returns.” *Expert Systems with Applications* 29 (4): 927–940. ISSN: 09574174. https://doi.org/10.1016/j.eswa.2005.06.024.

Fama, Eugene F. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *The Journal of Finance* 25 (2): 383. ISSN: 00221082. https://doi.org/10.2307/2325486.

———. 1995. “Random Walks in Stock Market Prices.” *Financial Analysts Journal* 51 (1): 75–80. ISSN: 0015-198X. https://doi.org/10.2469/faj.v51.n1.1861. https://www.tandfonline.com/doi/full/10.2469/faj.v51.n1.1861.

Fama, Eugene F., and Kenneth R. French. 1988. “Dividend yields and expected stock returns.” *Journal of Financial Economics* 22 (1): 3–25. ISSN: 0304405X. https://doi.org/10.1016/0304-405X(88)90020-7.

Feng, Guanhao, Jingyu He, and Nicholas G. Polson. 2018. “Deep Learning for Predicting Asset Returns” (April).

Fischer, Thomas, and Christopher Krauss. 2018. “Deep learning with long short-term memory networks for financial market predictions.” *European Journal of Operational Research* 270 (2): 654–669. ISSN: 03772217. https://doi.org/10.1016/j.ejor.2017.11.054.

Foundation, Wikimedia. 2017. *Long Short Term Memory Unit,* June. https://commons.wikimedia.org/wiki/File:Long_Short-Term_Memory.svg#metadata.

Goyal, Amit. *Amit Goyal.* https://sites.google.com/view/agoyal145.

Goyal, Amit, and Ivo Welch. 2003. "Predicting the Equity Premium with Dividend Ratios." *Management Science* 49 (5): 639–654. ISSN: 0025-1909. https://doi.org/10.1287/mnsc.49.5.639.15149.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning.* Springer New York. ISBN: 978-0-387-84857-0. https://doi.org/10.1007/978-0-387-84858-7.

Hilt, Donald E., and Donald W. Seegrist. 1977. *Ridge, a computer program for calculating ridge regression estimates /.* Dept. of Agriculture, Forest Service, Northeastern Forest Experiment Station, https://doi.org/10.5962/bhl.title.68934.

Hjalmarsson, Erik. 2010. "Predicting Global Stock Returns." *Journal of Financial and Quantitative Analysis* 45 (1): 49–80. ISSN: 0022-1090. https://doi.org/10.1017/S0022109009990469.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–1780. ISSN: 0899-7667. https://doi.org/10.1162/neco.1997.9.8.1735.

Hoerl, Arthur E., and Robert W. Kennard. 1970a. "Ridge Regression: Applications to Nonorthogonal Problems." *Technometrics* 12 (1): 69. ISSN: 00401706. https://doi.org/10.2307/1267352.

———. 1970b. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55. ISSN: 00401706. https://doi.org/10.2307/1267351.

Iworiso, Jonathan. 2020. *ON THE PREDICTABILITY OF U.S. STOCK MARKET USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES.*

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *The Journal of Finance* 48 (1): 65. ISSN: 00221082. https://doi.org/10.2307/2328882.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47 (2): 263. ISSN: 00129682. https://doi.org/10.2307/1914185.

Kellard, Neil M., John C. Nankervis, and Fotios I. Papadimitriou. 2010. “Predicting the equity premium with dividend ratios: Reconciling the evidence.” *Journal of Empirical Finance* 17 (4): 539–551. ISSN: 09275398. https://doi.org/10.1016/j.jempfin.2010.04.002.

Kingma, Diederik P., and Jimmy Ba. 2014. “Adam: A Method for Stochastic Optimization” (December).

Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. 2017. “A survey of deep neural network architectures and their applications.” *Neurocomputing* 234 (April): 11–26. ISSN: 09252312. https://doi.org/10.1016/j.neucom.2016.12.038.

Lo, Andrew W. 2004. “The Adaptive Markets Hypothesis.” *The Journal of Portfolio Management* 30 (5): 15–29. ISSN: 0095-4918. https://doi.org/10.3905/jpm.2004.442611.

Lo, Andrew W., and A. Craig MacKinlay. 1988. “Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test.” *Review of Financial Studies* 1 (1): 41–66. ISSN: 0893-9454. https://doi.org/10.1093/rfs/1.1.41.

McLean, R. David, and Jeffrey Pontiff. 2016. “Does Academic Research Destroy Stock Return Predictability?” *The Journal of Finance* 71 (1): 5–32. ISSN: 00221082. https://doi.org/10.1111/jofi.12365.

Mysla, Vlad. 2020. *Machine learning in 10 minutes,* April. https://medium.datadriveninvestor.com/machine-learning-in-10-minutes-354d83e5922e.

Nalbantov, Georgi, Rob Bauer, and Ida Sprinkhuizen-Kuyper. 2006. “Equity style timing using support vector regressions.” *Applied Financial Economics* 16 (15): 1095–1111. ISSN: 09603107. https://doi.org/10.1080/09603100500426556.

Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou. 2014. "Forecasting the Equity Risk Premium: The Role of Technical Indicators." *Management Science* 60 (7): 1772–1791. ISSN: 0025-1909. https://doi.org/10.1287/mnsc.2013.1838.

Nonejad, Nima. 2021. "Predicting equity premium by conditioning on macroeconomic variables: A prediction selection strategy using the price of crude oil." *Finance Research Letters* 41 (July): 101792. ISSN: 15446123. https://doi.org/10.1016/j.frl.2020.101792.

Osborne, M. F. M. 1962. "Periodic Structure in the Brownian Motion of Stock Prices." *Operations Research* 10 (3): 345–379. ISSN: 0030-364X. https://doi.org/10.1287/opre.10.3.345.

Patle, A., and D. S. Chouhan. 2013. "SVM kernel functions for classification," 1–9. IEEE, January. ISBN: 978-1-4673-5619-0. https://doi.org/10.1109/ICAdTE.2013.6524743.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011a. "Scikit-learn: Machine Learning in Python - ExtraTreeRegressor." *Journal of Machine Learning Research* 12:2825–2830. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html.

———. 2011b. "Scikit-learn: Machine Learning in Python - SelectKBest." *Journal of Machine Learning Research* 12:2825–2830. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html.

Poutré, Cédric, Georges Dionne, and Gabriel Yergeau. 2021. "International High-Frequency Arbitrage for Cross-Listed Stocks." *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.3890433.

Prado, Marcos Lopez de. 2016. "Hierarchical Portfolio Construction: Dispelling Markowitz's Curse." *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.2708678.

Rosenblatt, F. 1958. “The perceptron: A probabilistic model for information storage and organization in the brain.” *Psychological Review* 65 (6): 386–408. ISSN: 1939-1471. https://doi.org/10.1037/h0042519.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams. 1986. “Learning Internal Representations Error Propagation.” *Cognitive Science* 1 (V): 318–362. https://apps.dtic.mil/docs/citations/ADA164453.

Samuelson, Paul A. 1965. *A Theory of Induced Innovation along Kennedy-Weisäcker Lines.* https://www.jstor.org/stable/1927763.

———. 1973. *Proof That Properly Discounted Present Values of Assets Vibrate Randomly.* https://www.jstor.org/stable/3003046.

Santosa, Fadil, and William W. Symes. 1986. “Linear Inversion of Band-Limited Reflection Seismograms.” *SIAM Journal on Scientific and Statistical Computing* 7 (4): 1307–1330. ISSN: 0196-5204. https://doi.org/10.1137/0907087.

Tibshirani, Robert. 1996. *Regression Shrinkage and Selection via the Lasso.*

Tsai, Chih-Fong, Yuah-Chiao Lin, David C. Yen, and Yan-Min Chen. 2011. “Predicting stock returns by classifier ensembles.” *Applied Soft Computing* 11 (2): 2452–2459. ISSN: 15684946. https://doi.org/10.1016/j.asoc.2010.10.001.

Vineeth, Venishetty, Huseyin Kusetogullari, and Alain Boone. 2020. “Forecasting Sales of Truck Components: A Machine Learning Approach.” May. https://doi.org/10.1109/IS48319.2020.9200128.

Welch, Ivo, and Amit Goyal. 2008. “A comprehensive look at the empirical performance of equity premium prediction.” *Review of Financial Studies* 21 (4): 1455–1508. ISSN: 08939454. https://doi.org/10.1093/rfs/hhm014.