International Journal of Econometrics and Financial Management, 2023, Vol. 11, No. 1, 1-12
Available online at http://pubs.sciepub.com/ijefm/11/1/1
Published by Science and Education Publishing
DOI:10.12691/ijefm-11-1-1

Forecasting Stock Market Out-of-Sample with Regularised Regression Training Techniques

Jonathan Iworiso*

School of Computing and Digital Media, London Metropolitan University
*Corresponding author: j.iworiso@londonmet.ac.uk

Received March 22, 2023; Revised April 25, 2023; Accepted May 07, 2023

Abstract Forecasting stock market out-of-sample is a major concern to researchers in finance and emerging markets. This research focuses mainly on the application of regularised Regression Training (RT) techniques to forecast monthly equity premium out-of-sample recursively with an expanding window method. A broad category of sophisticated regularised RT models involving model complexity were employed. The regularised RT models which include Ridge, Forward-Backward (FOBA) Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), Relaxed LASSO, Elastic Net and Least Angle Regression were trained and used to forecast the equity premium out-of-sample. In this study, the empirical investigation of the Regularised RT models demonstrate significant evidence of equity premium predictability both statistically and economically relative to the benchmark historical average, delivering significant utility gains. Overall, the Ridge gives the best statistical performance evaluation results while the LASSO appeared to be most economical meaningful. They seek to provide meaningful economic information on mean-variance portfolio investment for investors who are timing the market to earn future gains at minimal risk. Thus, the forecasting models appeared to guarantee an investor in a market setting who optimally reallocates a monthly portfolio between equities and risk-free treasury bills using equity premium forecasts at minimal risk.

Keywords: *regression training, out-of-sample, expanding window, statistical predictability, economic significance, utility gains*

Cite This Article: Jonathan Iworiso, “Forecasting Stock Market Out-of-sample with Regularised Regression Training Techniques.” *International Journal of Econometrics and Financial Management*, vol. 11, no. 1 (2023): 1-12. doi: 10.12691/ijefm-11-1-1.

1. Introduction

The out-of-sample predictability of stock market is a research problem in empirical finance. The quest on stock market delivery to mean-variance investors above the treasury bill rate led to an “estimate of the equity premium”. The historical average model is an old-fashioned efficient market approach for forecasting the equity premium. Existing literature claimed that financial variables used as potential predictors can only forecast the equity premium in-sample but are unable to deliver significantly superior out-of-sample forecasts relative to the benchmark historical average. This led to the research question: can anything consistently beat the historical average out-of-sample? [4]. The historical average is used as a benchmark for comparing the performance of any model whose forecasts are estimated out-of-sample via expanding or rolling window [10,11]. Thus, any model whose statistical measures outperformed those from the benchmark historical average is said to beat the historical average.

This research proposes an application of regularised Regression Training (RT) techniques to forecast monthly equity premium out-of-sample recursively with expanding window.

In finance, the statistical predictability does not necessarily guarantee investor's profit from the trading strategy. Thus, the statistical predictability and economic significance are comparatively considered in the performance evaluation metrics in this paper.

The equity premium or excess stock return is the difference between the expected return on the market portfolio (SP500) and the risk-free treasury bill rate. It is the return that investors can expect from holding the market portfolio in excess of the return on the risk-free rate.

Mathematically, it is defined as:

$$EquityPrem_t = \log\left(\frac{P_t}{P_{t-1}}\right) - rf_{t-1} \qquad (1)$$

where P_t is the price of the stock index at period t; rf_{t-1} is the risk-free interest rate at $t-1$.

The Regression Training comes from the “***caret***” package, developed by [16] for Classification And REgression. The caret package (caret for R and RStudio, PyCaret for python) aimed to automate the main steps for evaluating and comparing machine learning algorithms.

The RT techniques presumed that all predictor variables are useful before preprocessing when training the model and the trained model decides variable importance associated with the final model. Thus, the RT model is resampled and fine-tuned iteratively, and the best tuning parameters are used to run the out-of-sample forecasts.

The remaining structure of the paper is laid out as follows: Section 2 described the research methodology; Section 3 present the variables, the empirical results and discussion; Section 4 concludes the paper.

2. Methodology

2.1. The Historical Average

Given a univariate time series $\{y_t\}_{t=1}^{T}$, with y_t denoting the monthly equity premium. The historical average (HA) model is defined as follows:

$$y_{t+1} = \beta + \epsilon_{t+1} \tag{2}$$

where β is a parameter representing the intercept; ϵ_t is a zero mean disturbance term; $t = 1, 2, ..., T$ [4,17]. The least squares estimator (LSE) of the historical average is as follows:

$$\hat{\beta}_{LSE}^{HA} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

which implies that the forecast for $\hat{y}_{T+1}$ is given by:

$$\hat{y}_{T+1|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

where $\hat{\beta}_{LSE}^{HA}$ is the parametric estimator of β.

2.2. The Least Squares Regression Training

Given a training dataset $\{y_t, X_{t,1}, X_{t,2}, ..., X_{t,k}\}_{t=1}^{T}$ of T statistical units, then a kitchen sink predictive linear model takes the form:

$$y_{t+1} = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \cdots + \beta_k X_{t,k} + \epsilon_{t+1} \tag{3}$$

where y_{t+1} is the equity premium at $t+1$; $X_{t,1}, X_{t,2}, ..., X_{t,k}$ are the predictor variables available at the end of t used to predict y_{t+1}; β_0 is a constant term representing the intercept; $\beta_1, \beta_2, ..., \beta_k$ are the model coefficients; ϵ_{t+1} is a zero mean disturbance term [11,22].

The above model can be represented in matrix form, as follows:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon \tag{4}$$

where $\mathbf{y}$ is a $T \times 1$ vector of observed values; $\mathbf{X}$ is $T \times (k+1)$ matrix of predictor variables; β is $(k+1) \times 1$ dimensional parameter vector; ϵ is $T \times 1$ zero mean vector of disturbances.

If the parameters $\beta = (\beta_0, \beta_1, ..., \beta_k)$ are estimated by OLS, then the linear model (LM) forecasts can be obtained from the resulting kitchen sink predictive model:

$$\hat{y}_{T+1}(\hat{\beta}^{OLS}) = \mathbf{X}_T' \hat{\beta}^{OLS} \tag{5}$$

where $\hat{\beta}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the OLS estimate of β.

2.3. The Regularised or Penalised Regression Training

2.3.1. The Ridge

Using the training set

$$\{(X_{t,1}, y_1), (X_{t,2}, y_2), ..., (X_{T,k}, y_T)\},$$

and by imposition of ridge constraints, the model parameter estimates will be obtained by minimizing the objective function

$$\sum_{t=1}^{T-1}(y_{t+1} - \beta_0 - \sum_{j=1}^{k}\beta_j X_{t,j})^2$$

$$\text{subject to } \sum_{j=1}^{k}\beta_j^2 \le s_1 \ \ for\ s_1 \in \lambda_1$$

which is a convex optimization problem, hence the solution has a closed form [12,15].

The ridge model parameter estimates will be:

$$\hat{\beta}^{Ridge} = \underset{\beta \in \mathbb{R}^{k+1}}{argmin}\left\{\sum_{t=1}^{T-1}(y_{t+1} - \beta_0 - \sum_{j=1}^{k}\beta_j X_{t,j})^2 + \lambda_1 \sum_{j=1}^{k}\beta_j^2\right\}$$

$$\Rightarrow \hat{\beta}^{Ridge} = (\mathbf{X}'\mathbf{X} + \lambda_1 \mathbf{I}_k)^{-1}\mathbf{X}'\mathbf{y}$$

which is always invertible, and hence non-singular [1]; where $\mathbf{X}$ is the $T \times k$ matrix of covariates; $\lambda_1 \sum_{j=1}^{k}\beta_j^2 = \lambda_1 \|\beta\|_2^2$ is the shrinkage penalty; $\sum_{j=1}^{k}|\beta_j^2| = \|\beta\|_2^2$ is the ℓ_2 – norm of the vector β; $\lambda_1 > 0$ is the ridge tuning parameter; β_0 is the intercept; $\beta_1, \beta_2, ..., \beta_k$ are the ridge coefficients; I_k is a $k \times k$ identity matrix; k is the number of parameters to be estimated; T is the sample size; $j = 1, 2, ..., k$.

Thus, the ridge forecasts are obtained from the resulting forecasting model:

$$\hat{\mathbf{y}}_{t+1|T}^{Ridge}(\hat{\beta}^{Ridge}) = \mathbf{X}_t' \hat{\beta}^{Ridge} \tag{7}$$

where

$$\hat{\beta}^{Ridge} = \underset{\beta \in \mathbb{R}^{k+1}}{argmin}\left\{\sum_{t=1}^{T-1}(y_{t+1} - \mathbf{X}_t'\beta)^2 + \lambda_1 \sum_{j=1}^{k}\beta_j^2\right\}$$

β is $(k+1) \times 1$ vector of unknown parameters, including the intercept; T is the sample size.

The ridge forecasts converges to the sample mean for large values of tuning parameter, λ_1:

$$\hat{\mathbf{y}}_{t+1|T}^{Ridge} \rightarrow \frac{1}{T-1}\sum_{t=2}^{T} y_t \text{ as } \lambda_1 \rightarrow \infty.$$

2.3.2. The Forward - Backward Ridge

The forward-backward (FOBA) ridge is an extension of the ridge model. It implements the forward and backward sparse learning algorithms for the ridge regression model. In this case, $\boldsymbol{\varrho} \in (\mathbf{0}, \mathbf{1})$ controls how likely the steps are to be taken, determining either addition or deletion of a variable in the ridge model. The FOBA method takes a backward step when the ridge penalised risk increase is less than $\boldsymbol{\varrho}$ times the ridge penalised risk reduction in the corresponding forward step, and vice versa.

2.3.3. The Least Absolute Shrinkage and Selection Operator

The LASSO model parameter estimates are obtained by minimizing the objective function:

$$\begin{gathered} \sum_{t=1}^{T-1}(y_{t+1} - \beta_0 - \sum_{j=1}^{k}\beta_j X_{t,j})^2 \\ \text{subject to } \sum_{j=1}^{k}|\beta_j| \leq s_2 \;\; for\; s_2 \in \lambda_2 \end{gathered} \tag{8}$$

where $\lambda_2 > 0$ is the LASSO tuning parameter; β_0 is the intercept; $\beta_1, \beta_2, ..., \beta_k$ are the LASSO coefficients; $\sum_{j=1}^{k}|\beta_j| = \|\beta\|_1$ is the ℓ_1 – norm of the vector β [23,25].

The LASSO model parameter estimates will be:

$$\hat{\beta}^{LASSO} = \underset{\beta\in\mathbb{R}^{k+1}}{argmin}\left\{\sum_{t=1}^{T-1}(y_{t+1} - \beta_0 - \sum_{j=1}^{k}\beta_j X_{t,j})^2 + \lambda_2\sum_{j=1}^{k}|\beta_j|\right\}$$

where $\lambda_2\sum_{j=1}^{k}|\beta_j| = \lambda_2\|\beta\|_1$ is the shrinkage penalty; and $\hat{\beta}^{LASSO} \rightarrow \hat{\beta}^{OLS}$ as $\lambda_2 \rightarrow \infty$.

Thus, the LASSO forecasts are obtained from the resulting LASSO forecasting model:

$$\hat{\mathbf{y}}_{t+1|T}^{LASSO}(\hat{\beta}^{LASSO}) = \mathbf{X}_t'\hat{\beta}^{LASSO} \tag{9}$$

where

$$\hat{\beta}^{LASSO} = \underset{\beta\in\mathbb{R}^{k+1}}{argmin}\left\{\sum_{t=1}^{T-1}(y_{t+1} - \mathbf{X}_t'\beta)^2 + \lambda_2\sum_{j=1}^{k}|\beta_j|\right\}$$

where $\mathbf{X}$ is the $T \times k$ matrix of covariates; β is $(k+1)\times 1$ vector of unknown parameters, including the intercept; T is the sample size; λ_2 controls the amount of shrinkage [9,19].

2.3.4. The Relaxed Least Absolute Shrinkage and Selection Operator

The relaxed least absolute shrinkage and selection operator (RELAXO) is a generalisation of the LASSO for linear regression.

Let λ and α be two separate parameters for controlling model selection and shrinkage estimation. The RELAXO estimator can be defined for $\lambda \in [0,\infty)$ and $\alpha \in (0,\infty]$ as follows:

$$\hat{\beta}^{RELAXO} = \underset{\beta\in\mathbb{R}^{k+1}}{argmin}\left\{\sum_{t=1}^{T-1}(y_{t+1} - \mathbf{X}_t'\{\beta\cdot\mathbf{1}_{\mathcal{S}_\lambda}\})^2 + \alpha\lambda\|\beta\|_1\right\}$$

where $\mathcal{S}_\lambda$ is the set of predictor variables selected by LASSO estimator; $\mathbf{1}_{\mathcal{S}_\lambda}$ is the indicator function on the set of predictor variables; $\alpha\lambda\|\beta\|_1$ is the shrinkage penalty for the RELAXO [20]. It can be expressed as follows:

$$\{\mathbf{1}_{\mathcal{S}_\lambda}\}_k = \begin{cases} 0, & k \notin \mathcal{S}_\lambda \\ \beta_k, & k \in \mathcal{S}_\lambda \end{cases}$$

Let $\mathcal{L}(\beta)$ be the negative log-likelihood under the parameter β, then the generalized RELAXO estimator takes the form [20]:

$$\hat{\beta}^{RELAXO} = \underset{\beta\in\mathcal{S}_\lambda}{argmin}\left\{\mathcal{L}(\beta) + \alpha\lambda\|\beta\|_1\right\} \tag{10}$$

Thus, the RELAXO forecasts are obtained from the resulting RELAXO forecasting model:

$$\hat{\mathbf{y}}_{t+1|T}^{RELAXO}(\hat{\beta}^{RELAXO}) = \mathbf{X}_t'\hat{\beta}^{RELAXO} \tag{11}$$

where

$$\hat{\beta}^{RELAXO} = \underset{\beta\in\mathbb{R}^{k+1}}{argmin}\left\{\sum_{t=1}^{T-1}(y_{t+1} - \mathbf{X}_t'\{\beta\cdot\mathbf{1}_{\mathcal{S}_\lambda}\})^2 + \alpha\lambda\|\beta\|_1\right\} \tag{12}$$

where β is $(k+1)\times 1$ vector of unknown parameters, including the intercept; $\mathcal{S}_\lambda$ is the set of predictor variables selected by LASSO estimator; $\mathbf{1}_{\mathcal{S}_\lambda}$ is the indicator function on the set of predictor variables.

2.3.5. The Elastic Net

The elastic net, as proposed by [28] combines both the ℓ_1 and ℓ_2 penalty vector norms, and tends to eliminate extreme solutions. Thus the elastic net model parameter estimates are obtained by minimizing the objective function that includes the ridge and LASSO shrinkage penalties subject to both constraints, as follows:

$$\hat{\beta}^{ElNet} = \underset{\beta\in\mathbb{R}^{k+1}}{argmin}\left\{\begin{aligned} &\sum_{t=1}^{T-1}(y_{t+1} - \beta_0 - \sum_{j=1}^{k}\beta_j X_{t,j})^2 \\ &+\lambda_2\Big(\sum_{j=1}^{k}(1-\lambda_1)\beta_j^2 + \lambda_1\|\beta_j\|\Big)\end{aligned}\right\}$$

where λ_1 is the ridge tuning parameter; λ_2 is the LASSO tuning parameter [27].

It is worth noting that the Elastic Net is Ridge if $\lambda_1 = 0$; it is LASSO if $\lambda_2 = 1$ and it is strictly convex if $\frac{\lambda_2}{\lambda_1 + \lambda_2} > 0$ [3].

Therefore, the elastic net forecasts are obtained from the elastic net forecasting model:

$$\hat{\mathbf{y}}^{ElNet}_{t+1|T}(\hat{\beta}^{ElNet}) = \mathbf{X}'_t \hat{\beta}^{ElNet} \quad (12)$$

where

$$\hat{\beta}^{ElNet} = \underset{\beta \in \mathbb{R}^{k+1}}{argmin} \left\{ \begin{array}{l} \sum_{t=1}^{T-1} (y_{t+1} - \mathbf{X}'_t \beta)^2 \\ +\lambda_2 \left(\sum_{j=1}^{k} (1-\lambda_1)\beta_j^2 + \lambda_1 \left\| \beta_j \right\| \right) \end{array} \right\}$$

β is $(k+1)\times 1$ vector of unknown parameters including the intercept; and $\mathbf{X}$ is the $T \times k$ matrix of covariates.

2.3.6. The Least Angle Regression

The least angle regression (LARS), introduced by [8] is a machine learning model selection algorithm for fitting linear regression models to high dimensional data. In the LARS algorithm, the parameter estimates are increasing in an equiangular direction to each of the corresponding correlations associated with the model residuals.

The LARS algorithm adapted from [2] and [8] is summarized as follows:

The LARS Algorithm
1. Initialise all coefficients $\beta = 0$;
2. Search for the predictor $X_{j,t}$ most correlated with the response variable y_t;
3. Increase the coefficient β_j in the direction of its correlation sign;
4. Obtain residuals $\epsilon_t = y_t - \hat{y}_t$. Stop if another predictor $X_{k,t}$ has as much correlation with ϵ_t as $X_{j,t}$;
5. Increase (β_j, β_k) in their joint LS direction until another predictor $X_{j,m}$ has as much correlation with ϵ_t;
6. Increase $(\beta_j, \beta_k, \beta_m)$ in their LS direction until another predictor $X_{l,t}$ has as much correlation with ϵ_t
7. Continue until all predictors are in the model;
8. End

The LARS2 is a special improved case of the LARS that uses **step** as the tuning parameter instead of **fraction**.

2.4. Statistical and Economic Performance Evaluation

2.4.1. Mean Squared Forecast Error

The mean squared forecast error (*MSFE*) is computed as follows:

$$MSFE = \frac{1}{\mathbb{T}} \sum_{t=1}^{\mathbb{T}} (y_t - \hat{y}_t)^2$$

where $\mathbb{T}$ is the out-of-sample forecasting period; y_t is the actual value at specific time t; $\hat{y}_t$ is the forecast value at specific time t.

2.4.2. Out-of-Sample Forecast Evaluation: The R^2_{OOS} Statistic

The out-of-sample statistical goodness of fit used to measure the performance of individual equity premium forecasting model, suggested by [5] for evaluating the overall performance of any competing model forecasts in terms of proportional error minimization, relative to the benchmark historical average forecast is defined as follows:

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{\mathbb{T}} (y_{\mathbb{T}_0+t} - \hat{y}_{\mathbb{T}_0+t})^2}{\sum_{t=1}^{\mathbb{T}} (y_{\mathbb{T}_0+t} - \bar{y}_{\mathbb{T}_0+t})^2}$$

where $R^2_{OOS} > 0$ implies that the MSE of the forecasting model is less than the MSE of the benchmark forecasts based on historical average; $\hat{y}_{\mathbb{T}_0+t}$ represents an equity premium forecast based on a specific competing model.

2.4.3. Diebold-Mariano Test

The assumptions of Diebold-Mariano (DM) test rely on the forecast error loss differential function [6,7]. Let $\epsilon_{1,t}$ and $\epsilon_{2,t}$ denote the forecast errors associated with the loss functions $L(\epsilon_{1,t})$ and $L(\epsilon_{2,t})$ for forecasts 1 and 2 respectively. The time-t loss differential between forecasts 1 and 2 is defined as follows:

$$d_{1,2(t)} = L(\epsilon_{1,t}) - L(\epsilon_{2,t})$$

The *DM* hypothesis of equal forecast accuracy, also known as equal expected loss, corresponds to the zero mean assumption of $d_{1,2(t)}$, i.e., $E(d_{1,2(t)}) = 0$; where $E(\cdot)$ denotes the mean value. Thus, the null hypothesis of equal forecast accuracy against the alternative hypothesis of unequal forecast accuracy between forecasts 1 and 2, based on monthly forecast horizon $h = 1$, can be tested using the *DM* test statistic as follows [6]:

$$DM_{1,2} = \frac{\bar{d}_{1,2}}{\hat{\sigma}_{\bar{d}_{1,2}}} \overset{asymptotically}{\sim} N(0,1)$$

where $\bar{d}_{1,2} = \frac{1}{\mathbb{T}} \sum_{t=1}^{\mathbb{T}} d_{1,2(t)}$ is the sample mean loss differential and $\hat{\sigma}_{\bar{d}_{1,2}}$ is a consistent estimate of the standard deviation of $\bar{d}_{1,2}$. Thus, the *DM* test statistic has the asymptotic standard normal distribution under the null hypothesis of equal forecast accuracy. In this study, the forecast errors of each RT model are compared with the forecast errors from the benchmark historical average.

2.4.4. Sharpe Ratio

[24] employed the Sharp Ratio (SR) as a measure of excess return per unit of risk in an investment asset or trading strategy. In this study, the SR standardizes the realized returns with the risk of the portfolios, and it is computed as follows:

$$SR_p = \frac{E(R_p) - E(R_f)}{\sqrt{Var(R_p)}}$$

where $E(R_p)$ is the average realized return of the portfolio over the out-of-sample period; $E(R_f)$ is the average risk-free treasury bill rate; $Var(R_p)$ is the variance of the portfolio over the out-of-sample period.

2.4.5. Cumulative Return

The cumulative return (CR) of the portfolio, is computed as follows:

$$CR = \sum_{t=1}^{\mathbb{T}} R_t \qquad (13)$$

where R_t is the return on month t; $\mathbb{T}$ is the number of months in the out-of-sample periods.

2.4.6. Utility Gain

A mean-variance investor who forecasts the monthly equity premium using the HA will decide at the end of time t to allocate risky weights as share of her portfolio to equities in time $t+1$, in the form:

$$\omega_{0,t} = \gamma^{-1}\left(\frac{\bar{y}_{t+1}}{\hat{\sigma}^2_{R,t+1}}\right)$$

where the portfolio risky weights $\omega_{0,t}$ are constrained to lie between 0% and 150%, (i.e., $\omega_{0,t} = 0$ if $\omega_{0,t} < 0$ and $\omega_{0,t} = 1.5$ if $\omega_{0,t} > 1.5$; γ is the risk aversion parameter; $\bar{y}_{t+1}$ is the equity premium forecasts based on HA; $\hat{\sigma}^2_{R,t+1}$ is the variance of stock returns [5,11,12].

The investor realizes an average utility from the HA, given by:

$$\hat{U}_0 = \hat{\mu}_{0,p} - \frac{1}{2}\gamma\hat{\sigma}^2_{0,p}$$

where $\hat{\mu}_{0,p}$ is the sample mean over the out-of-sample period; $\hat{\sigma}^2_{0,p}$ is the sample variance over the out-of-sample period.

The weight risky equity share can be chosen by the following:

$$\omega_{j,t} = \gamma^{-1}\left(\frac{\bar{y}_{j,t+1}}{\hat{\sigma}^2_{R,t+1}}\right)$$

Then the investor realizes an average utility from an individual RT model, defined by:

$$\hat{U}_j = \hat{\mu}_{j,p} - \frac{1}{2}\gamma\hat{\sigma}^2_{j,p}$$

where $\hat{\mu}_{j,p}$ is the sample mean over the out-of-sample period; $\hat{\sigma}^2_{j,p}$ is the sample variance over the out-of-sample period.

Thus, the utility gain (UG) can be computed as follows:

$$UG = \hat{U}_j - \hat{U}_0$$

for each of the RT out-of-sample forecasting models.

3. The Empirical Results and Discussion

3.1. Data, Variables and Forecasting Method

The dataset with financial variables used in this paper are obtained from [26], Amit Goyal's website and Robert Shiller's website, covering monthly observations from January 1960 to December 2019. The stock indices are obtained from the CRSP's month-end values of the $S\&P500$ monthly index, and the stock returns are the continuously compounded returns on the $S\&P500$ index. All out-of-sample forecasts are obtained by expanding window; and the out-of-sample period is from January 1994 to December 2019. The parameters of the forecasting models are estimated recursively using an expanding window of observations, with data point from the start date to the present time and obtain a one month-period-ahead forecast. The forecast horizon is one month ahead, and the procedure is repeated until the last forecast is obtained.

3.2. Results and Discussion

In this paper, the empirical results for the RT models are summarised in two panels, displayed in Table 1 and Table 2 respectively. Following the benchmark statistical performance evaluation metrics in [4,5,10], any model which gives a positive out-of-sample R^2_{OOS} (i.e., $R^2_{OOS} > 0$) using expanding or rolling window is said to have outperformed or consistently beat the historical average. In the Kitchen Sink RT Model panel, the Linear Model (*LM*) gives a negative R^2_{OOS} (i.e., $R^2_{OOS} < 0$), which indicates underperformance relative to the benchmark historical average. It corroborates previous findings in empirical literature in which the ordinary linear regression cannot consistently beat the benchmark historical average out-of-sample. Thus, the introduction of model training with fine-tuning of parameters recursively in the *LM* does not improve the statistical predictive task of the *LM* in this direction.

Table 1. Data & Description of Time Series Variables - January 1960 to December 2019

Variable	Description
Stock Index, SP_t	The Standard & Poor 500 U.S stock index.
Dividend Price Ratio (log), DPR_t	The dividends over the past year divided by the current stock index value.
Dividend Yield (log), DY_t	The difference between the log of dividends and the log of lagged prices.
Earnings Price Ratio (log), EPR_t	The earnings over the past year divided by the current stock index value.
Realized Stock Variance, RSV_t	The sum of squared daily returns on the $S\&P500$ index within one month.
Book to Market Value, BMV_t	The ratio of book value to market value for the Dow Jones Industrial Average.
Net Equity Expansion, NEE_t	The ratio of 12-month moving sums of net issues by New York Stock Exchange (NYSE) listed stocks to total end of year market capitalization of the NYSE stocks.
Treasury Bill Rate, TBR_t	The interest rate on a 3-month treasury bill, secondary market.
Long Term Yield, LTY_t	The long term government bond yield, constant maturity.
Long Term Return, LTR_t	The return on long term government bonds.
Term Spread, TS_t	The difference between the long term yield (LTY_t) and the treasury bill rate (TBR_t).
Default Yield Spread, DYS_t	The difference between the *BAA* and *AAA* rated corporate bond yields.
Default Return Spread, DRS_t	The difference between the long term corporate bond and long term government bond returns.
Inflation, INF_t	Is computed from the consumer price index (CPI) for all urban consumers.

Table 2. The Statistical Performance Evaluation Results

RT Model	**Package**	**Method Value**	**MSFE**	**DM Stat**	**DM pValue**	R^2_{OOS} (%)
Kitchen Sink RT Model						
Linear Model	stats	lm	0.13512	0.6686	0.2519	-0.0389
Regularised RT Models						
Ridge	**elasticnet**	**ridge**	**0.00449**	**2.4286**	**0.0076**	**8.3047**
FOBA	elasticnet	foba	0.00554	1.6761	0.0469	6.6870
LASSO	elasticnet	lasso	0.00686	1.6601	0.0484	6.6724
RELAXO	elasticnet	relaxo	0.00459	1.8669	0.0310	7.9053
Elastic Net	elasticnet	enet	0.00996	1.3766	0.0843	4.9447
LARS	lars	lars	0.00516	1.7558	0.0396	6.9657
LARS2	lars	lars2	0.00476	1.7824	0.0373	7.3695

In the regularised RT Models panel, each of the models produced a positive R^2_{OOS} (i.e., $R^2_{OOS} > 0$), which indicates statistical evidence of outperformance over the benchmark historical average. In this paper, the Diebold-Mariano *DM* test is introduced as an additional statistical performance evaluation measure to compare the forecast accuracy of each RT model with those obtained from historical average. Interestingly, the regularised RT Models demonstrate statistically significant evidence of producing better forecasts than those obtained from historical average at 5% significance level, except for the Elastic Net. Also, the *LM* in the Kitchen Sink Model panel could not give any statistically significant evidence of producing unequal forecast accuracy relative to the historical average, as judged by the *DM* test. In the regularised RT Models panel, the Ridge gives the highest R^2_{OOS} with corresponding minimum *MSFE* and *DM pValue* among all the RT models tested, in terms of statistical predictive power. Thus, the presence of the ℓ_2 -vector norm in the Ridge model seems to improve the statistical predictive task of the Ridge model. The *FOBA* underperformed the Ridge model while the relaxed LASSO (*RELAXO*) outperformed the LASSO. The combination of both ℓ_1 and ℓ_2 vector norms in the Elastic Net does not improve the statistical predictive task of the Elastic Net model, as compared to their individual forms, as in the Ridge and LASSO. The *step* as a tuning parameter in the *LARS2* algorithm seems to improve the predictive task of the model, as compared to the *LARS* algorithm which uses *fraction* as a tuning parameter. Thus, the concept of bias-variance trade off in the sophisticated regularised RT Models is a more useful approach for forecasting the U.S. monthly equity premium out-of-sample with significant predictive power, relative to the benchmark historical average.

Turning to the economic performance evaluation measures (Table 3), it is important to note that the statistical predictive power of a model relative to the benchmark historical average does not necessarily guarantee economic significance in real market setting. The R^2_{OOS} and *MSFE* alone cannot explicitly account for an investor's risk over the out of sample period. In this paper, the useful economic performance evaluation metrics which includes the Cumulative Return *CR*, Sharpe Ratio *SR* and Utility Gains *UG* based on the out-of-sample periods were employed. The study seeks to reconcile the statistical and economic evidence in an attempt to guarantee the future expectation of a mean-variance portfolio investor. In this paper, the average risk-free treasury bill rate is $\bar{R}_{free} = 0.78\%$ and the risk aversion parameter is $\lambda = 3$. A mean-variance investor can increase her monthly portfolio return by computing a proportional factor $\frac{R^2_{OOS}}{(SR)^2}$, where *SR* is the Sharpe ratio. In [21] and [22], the *UG* is expressed in the form of average annualised percentage returns, also known as certainty equivalent returns. The *UG* is important in a real market setting in that it provides useful economic information on the portfolio management fee that an investor would be willing to pay in order to have access to the additional available information in the forecasting model relative to the sole information in the historical equity premium. For a mean-variance portfolio investor, a model that produced a higher *UG* based on the out-of-sample periods than the average risk-free treasury bill is preferable to its counterpart. Whereas, if risk is equal, then it is more profitable to invest in the treasury bills than in the portfolio based on the forecasting model.

[5] argued that even very low positive R^2_{OOS} values for monthly data can produce a meaningful economic evidence of equity premium predictability in terms of increased annual portfolio returns for a mean-variance investor. In agreement with [5], the *LM* in the Kitchen Sink Model panel gives an economically meaningful evidence, preferable to the average risk-free treasury bill, as judged by the *UG* and *SR*. In spite of the weak statistical predictive power of the *LM*, it seems to provide useful economic information to a mean-variance portfolio investor.

Table 3. The Economic Performance Evaluation Results

$\bar{R}_{free} = 0.78\%$	Risk Aversion Parameter = 3		
RT Model	**CR**	**SR**	**UG (%)**
Kitchen Sink RT Model			
Linear Model	1.9085	0.3382	0.8373
Regularised RT Models			
Ridge	2.7272	0.6004	2.1169
FOBA	2.5901	0.5720	1.9705
LASSO	**3.2634**	**0.8152**	**3.2459**
RELAXO	2.0084	0.3933	0.9297
Elastic Net	2.7789	0.6591	2.4606
LARS	2.9309	0.6609	2.4733
LARS2	2.7414	0.6123	2.1903

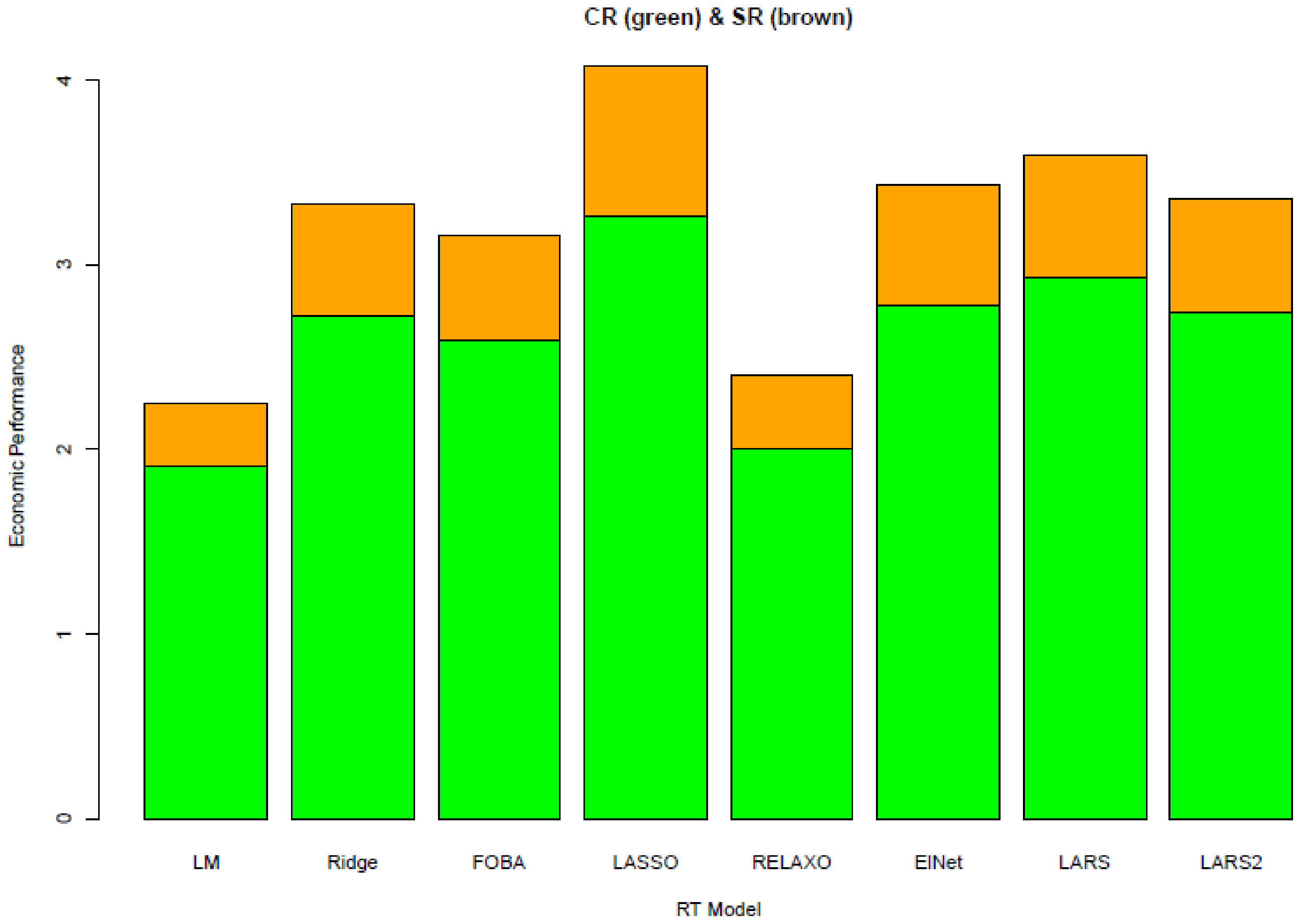

Figure 1. Stacked Bar Chart showing CR and SR for the RT Models

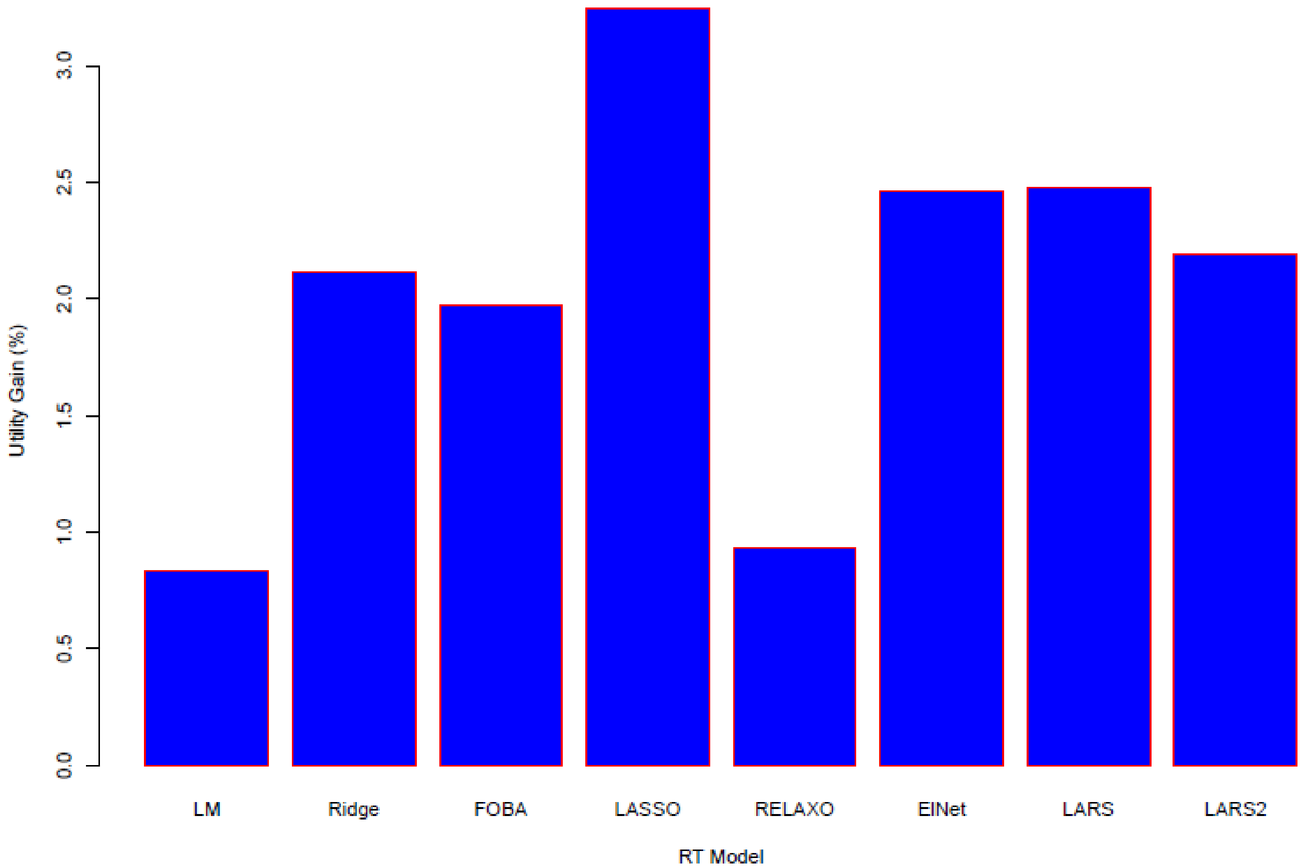

Figure 2. The Utility Gain (%) for the RT Models

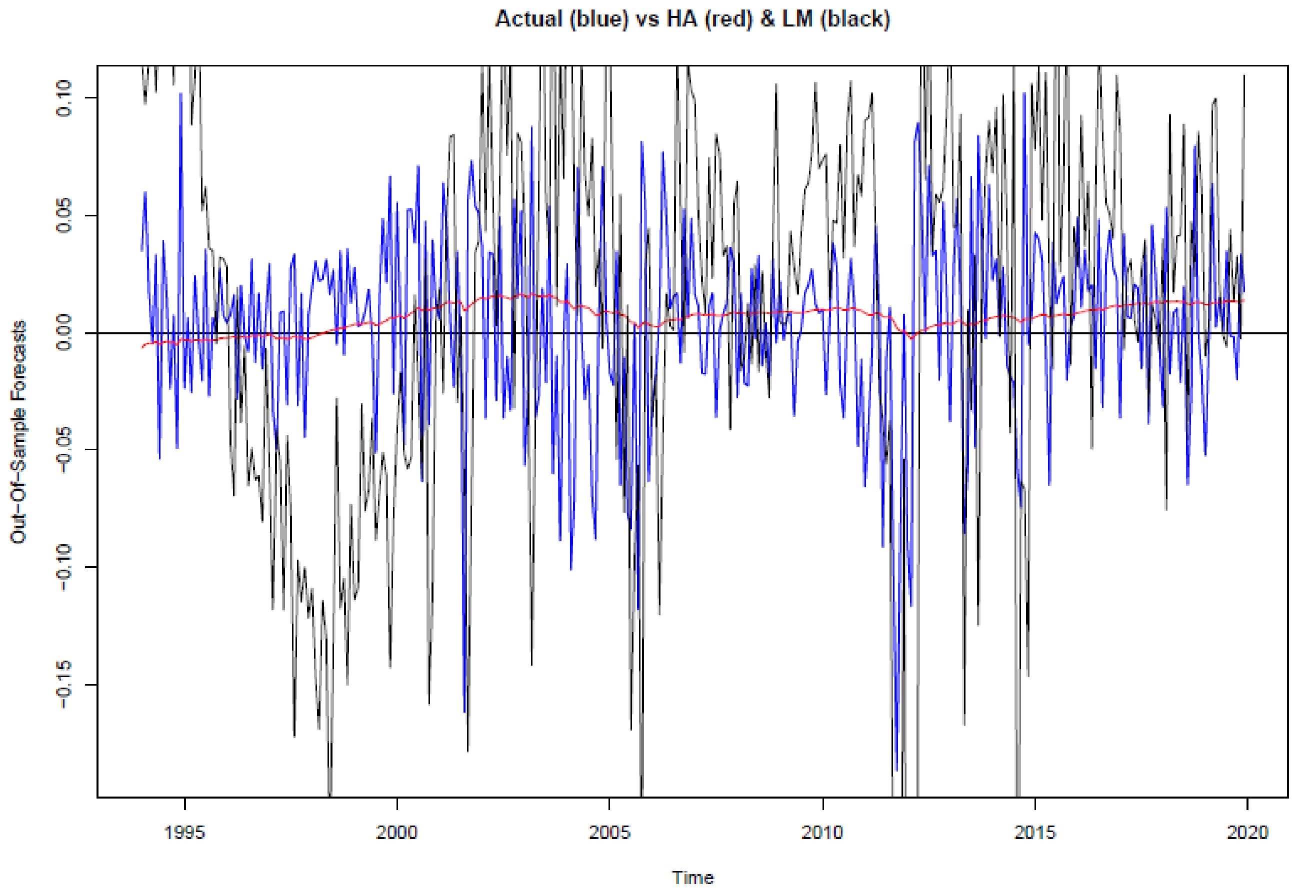

Figure 3. Comparing Actual (Equity Prem) with HA and LM Forecasts

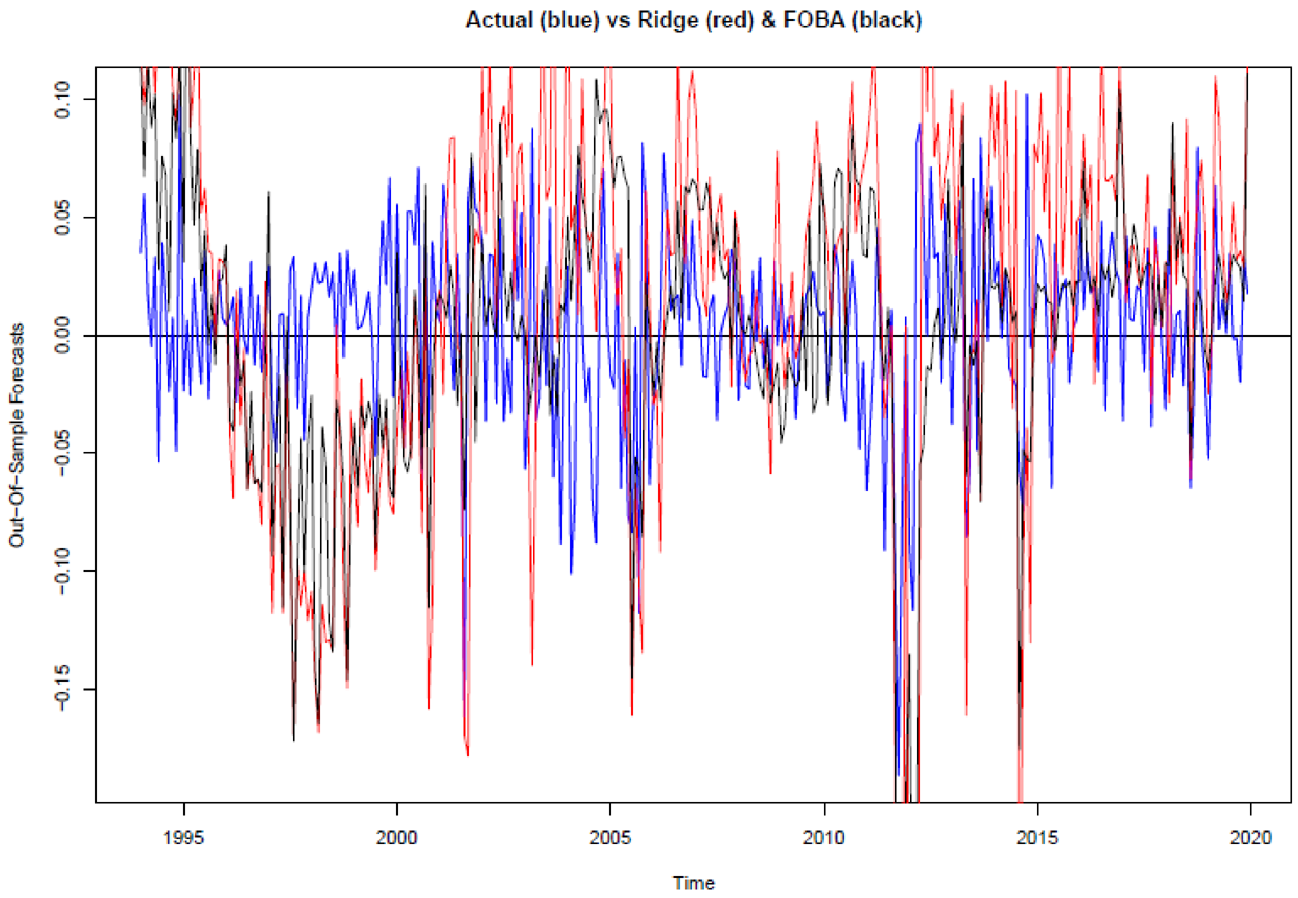

Figure 4. Comparing Actual with Ridge and FOBA Forecasts

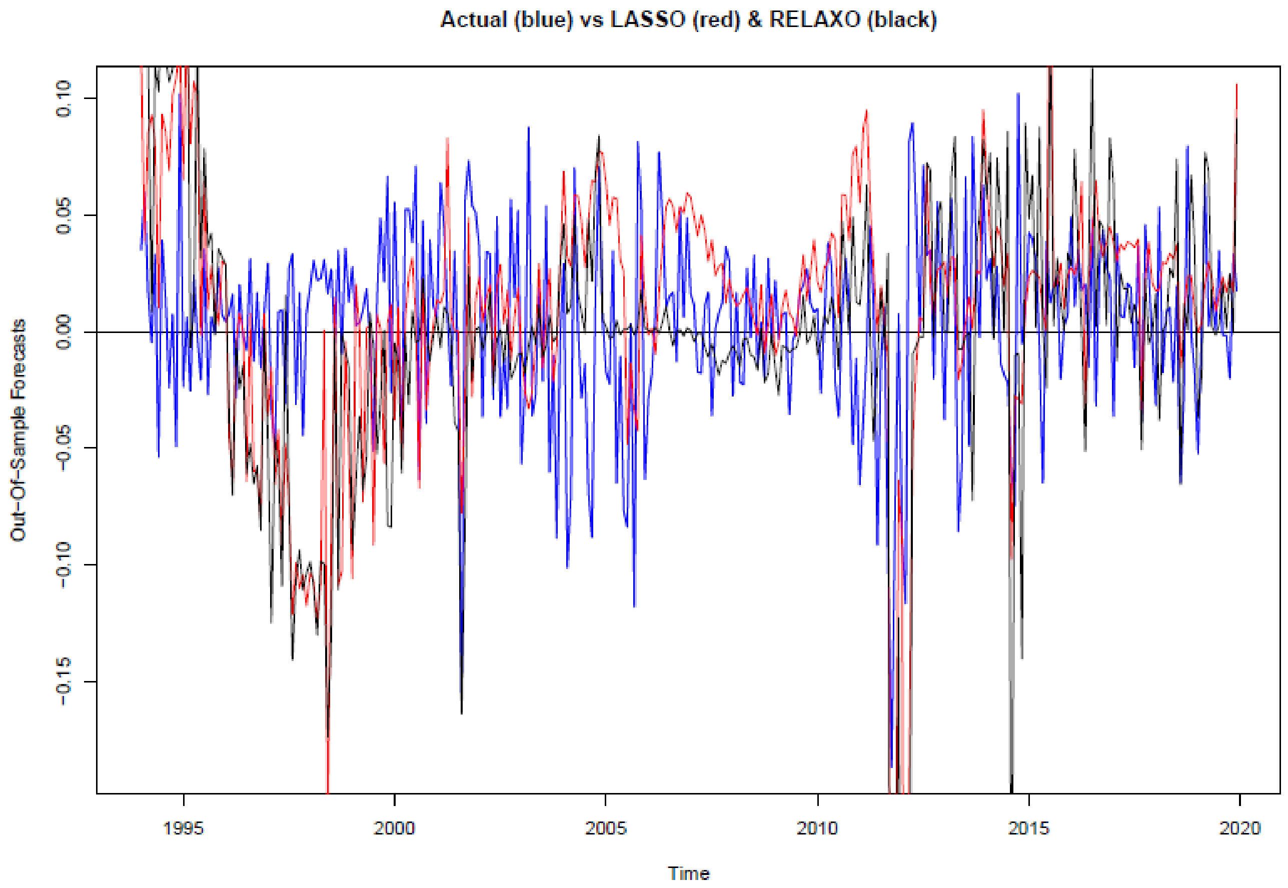

Figure 5. Comparing Actual with LASSO and Relaxed LASSO Forecasts

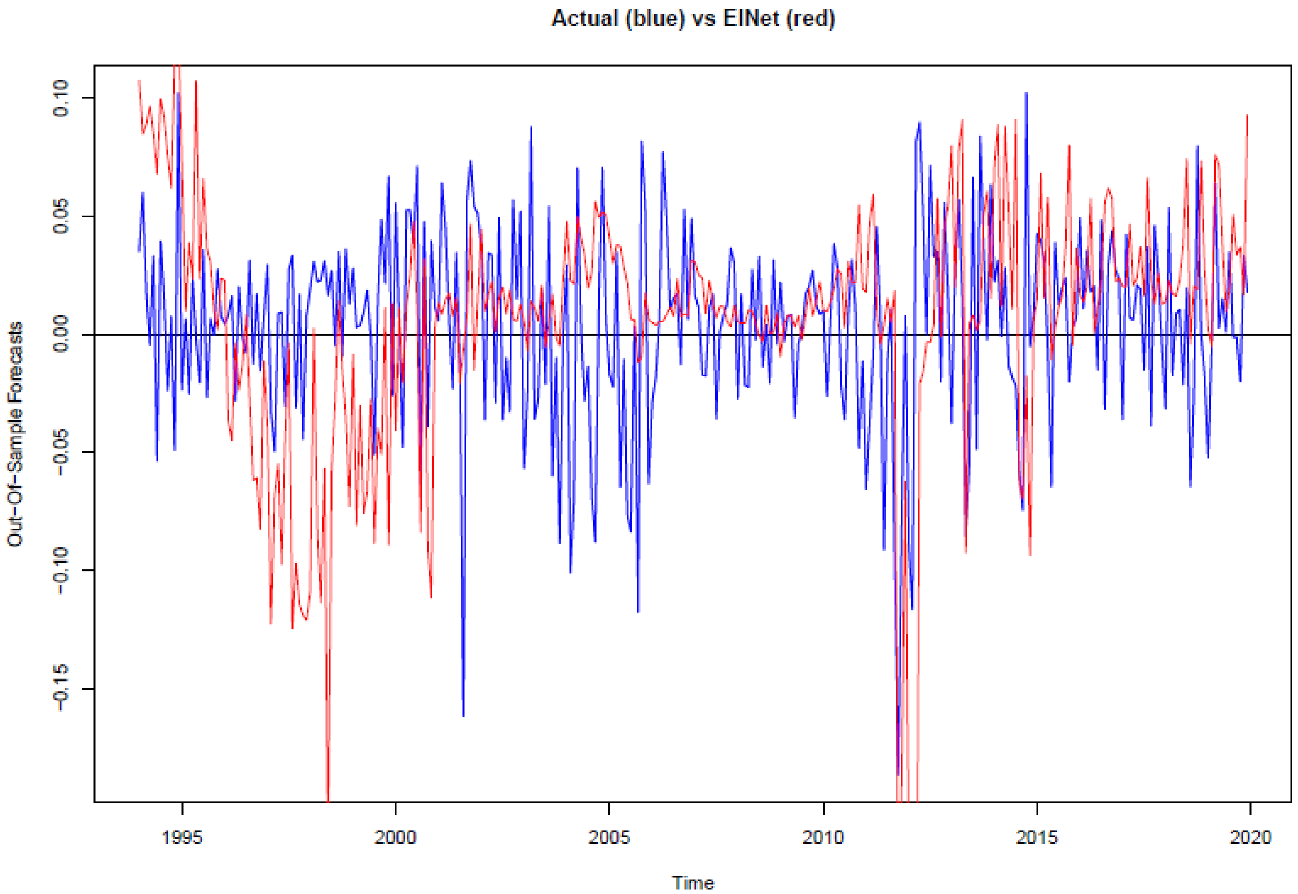

Figure 6. Comparing Actual with Elastic Net Forecasts

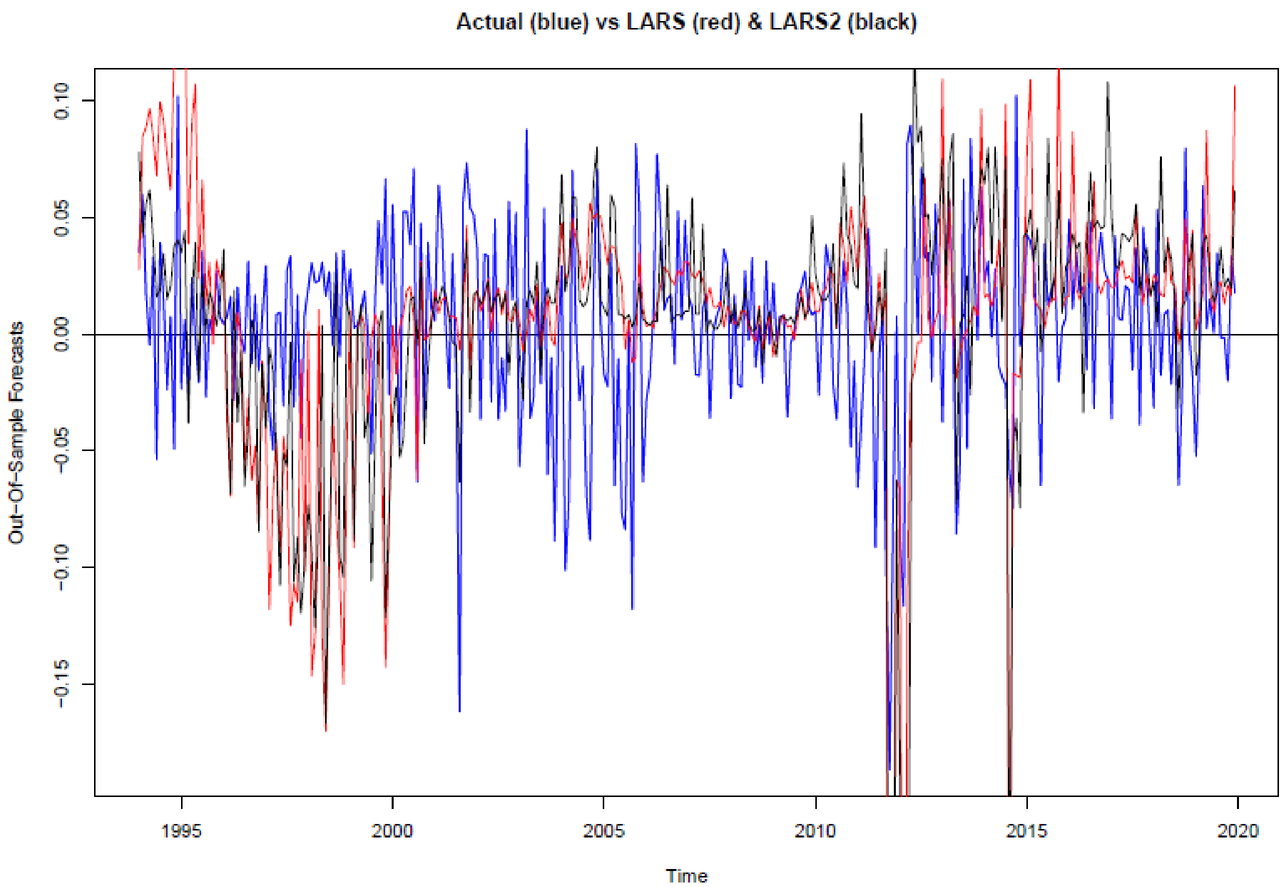

Figure 7. Comparing Actual with LARS and LARS2 Forecasts

In the regularised RT Models panel, all the models provide strong significant evidence of economic predictability and outperformance over the treasury bill. It is worth noting that the superiority of a forecasting model in terms of statistical predictability does not correspondingly guarantee superiority in economic significance. In the statistical performance evaluation metrics, the Ridge gives the best results. Whereas, in the economic performance evaluation metrics, the LASSO produced the best results. Thus, the ℓ_1-vector norm in the LASSO forecasting model seems to be more economically powerful than the ℓ_2-vector norm in the Ridge forecasting model. Like in [13] in which the penalised binary probit models used as classifiers for sign or directional forecasting, and the application of deep learning in [14], the training and fine-tuning approach of the regularised regression models in this paper also provides statistically significant evidence of equity premium predictability with significant economic gains. Figure 1 and Figure 2 depict the graphical analysis of the out-of-sample RT forecasting models. Figure 1 is a stacked bar chart while Figure 2 is bar chart, showing the cumulative returns (*CRs*), Sharpe ratios (*SRs*) and utility gains (*UGs*). The time series graphical representation of actual versus forecasts for the various RT forecasting models are depicted in Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7 respectively. As in [18], the regularised RT forecasting models in this paper provide significant evidence of equity premium predictability over the benchmark historical average with useful economic gains, and suggesting better alternatives to mean-variance investors.

The empirical analysis in this paper revealed that the sophisticated regularised RT forecasting models consistently beat the benchmark historical average out-of-sample, both statistically and economically. Thus, the regularised RT forecasting models used in this study appeared to guarantee a mean-variance portfolio investor in a real-time market setting who optimally reallocates a monthly portfolio between equities and risk-free treasury bill using equity premium forecasts at minimal risk.

4. Conclusion

This paper has answered the research question in [4,5], demonstrating the superiority of regularised RT forecasting models over the benchmark historical average out-of-sample with significant economic gains. Interestingly, all the regularised RT forecasting models consistently beat the benchmark historical average out-of-sample, both statistically and economically.

Overall, the Ridge gives the best statistical performance evaluation results while the LASSO appeared to be most economically meaningful. The regularised RT forecasting models provide useful economic information on mean-variance portfolio investment for investors who are timing the market to earn future gains at minimal risk. Thus, the regularised RT forecasting models appeared to guarantee a mean-variance investor in a real-time setting who optimally reallocates a monthly portfolio between equities and risk-free treasury bill using equity premium forecasts at minimal risk.

References

[1] Ahn, J. J., Byun, H. W., Oh, K. J., and Kim, T. Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications*, 39(9): 8369-8379.

[2] Alfons, A., Croux, C., and Gelper, S. (2016). Robust groupwise least angle regression. *Computational Statistics & Data Analysis*, 93: 421-435.

[3] Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2): 304-317.

[4] Campbell, J. Y. and Thompson, S. B. (2005). Predicting the equity premium out of sample: can anything beat the historical average? Technical report, National Bureau of Economic Research.

[5] Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4): 1509-1531.

[6] Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1): 1-1.

[7] Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1): 134-144.

[8] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of Statistics*, 32(2): 407-499.

[9] Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2): 357-373.

[10] Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5): 639-654.

[11] Goyal, A. and Welch, I. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4): 1455-1508.

[12] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55-67.

[13] Iworiso, J. and Vrontos, S. (2020). On the directional predictability of equity premium using machine learning techniques. *Journal of Forecasting*, 39(3): 449-469.

[14] Iworiso, J. and Vrontos, S. (2021). On the predictability of the equity premium using deep learning techniques. *The Journal of Financial Data Science*, 3(1): 74-92.

[15] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

[16] Kuhn, M. et al. (2008). Caret package. *Journal of statistical software*, 28(5): 1-26.

[17] Lee, T.-H., Tu, Y., and Ullah, A. (2015). Forecasting equity premium: global historical average versus local historical average and constraints. *Journal of Business & Economic Statistics*, 33(3): 393-402.

[18] Li, J. and Tsiakas, I. (2017). Equity premium prediction: The role of economic and statistical constraints. *Journal of Financial Markets*, 36: 56-75.

[19] MartíNez-MartíNez, J. M., Escandell-Montero, P., Soria-Olivas, E., Mart´ıN-Guerrero, J. D., Magdalena-Benedito, R., and G´oMez-Sanchis, J. (2011). Regularized extreme learning machine for regression problems. *Neurocomputing*, 74(17):3716-3721.

[20] Meinshausen, N. (2007). Relaxed lasso. Computational Statistics & Data Analysis, 52(1): 374-393.

[21] Rapach, D. E., Strauss, J. K., and Zhou, G. (2007). Out-of-sample equity premium prediction: Consistently beating the historical average. Review of Financial Studies.

[22] Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2): 821-862.

[23] Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., and Desmet, B. (2018). Tactical sales forecasting using a very large set of macroeconomic indicators. *European Journal of Operational Research*, 264(2):558-569.

[24] Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management*, 21(1): 49-58.

[25] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267-288.

[26] Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4): 1455-1508.

[27] Wu, L. and Yang, Y. (2014). Nonnegative elastic net and application in index tracking. *Applied Mathematics and Computation*, 227: 541-552.

[28] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 301-320.