ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

BACHELOR THESIS ECONOMETRICS AND OPERATIONAL RESEARCH

Investigating the use of Ridge, LASSO and Elastic Net for predicting the direction of the US stock market

Author
RUBEN SCHORNO 545096

Supervisor
BRAM VAN OS

Second assesor
PROF.DR. PHBF FRANSES

Abstract

This paper investigates the use of binary probit models for predicting the direction of the US stock market. Specifically, the setup of Nyberg (2011) is extended by implementing Ridge, LASSO and Elastic Net versions of the probit models. Monthly S&P500 index data from January 1986 to December 2021 is used to model the US stock market and several explanatory variables are considered. From which the recession forecast, introduced by Nyberg (2011), improves the forecasts the most. The results show that it is possible to improve the forecasting abilities of the models in-sample by implementing the Ridge, LASSO and Elastic Net techniques, when looking at the number of correctly forecasted periods and return on investment. However, out-of-sample it is found that the models without the machine learning techniques have the best forecasting ability.

July 3, 2022

ERASMUS UNIVERSITEIT ROTTERDAM

Contents

1 Introduction

Investors on the stock markets tend to have one main goal, namely maximizing their overall returns. Consequently, this aim has therefore over the years led to a numerous amount of research on modeling and -perhaps more importantly- forecasting of returns. However, papers such as Welch and Goyal (2008) show that the overall level of excess return on stocks is not significantly predictable. Fortunately, it is possible to predict the sign of excess stock returns to some extend (see, Christoffersen and Diebold (2006) and Breen et al. (1989)). That the sign is predictable offers an opportunity to assess whether an investor needs to allocate their resources into stocks or Treasury Bills for the following period.

This research aims at improving upon the results found in the paper of Nyberg (2011) by incorporating machine learning techniques, which would lead to models more capable of successfully predicting the direction of the stock market. Throughout this paper, models which do not incorporate machine learning techniques are referred to as standard models. In his paper, Nyberg extends the dynamic autoregressive probit model introduced by Kauppi and Saikkonen (2008) by imposing a restriction based on the efficient market theory. This results in the new ’error correction model’, which has better out-of-sample performance concerning forecasting the direction of the US stock market than other standard probit models. Adding to this, Nyberg also investigates the use of a recession forecast as explanatory variable. The use of this variable in return improves his forecasts.

To improve the results, three machine learning techniques are used both to make the forecasts of the recession variable as well as the direction of the US stock market. These are the so-called Ridge, LASSO and Elasitc Net probit model. Iworiso and Vrontos (2020) and Vrontos et al. (2021) find that machine learning techniques yields superior forecasts of recession periods and the direction of the US stock market, when compared to the standard static and dynamic probit model also used in Nyberg (2011). This is due to the regularization term which the Ridge, LASSO and Elastic Net techniques add. These terms penalize the parameters in the models, leading to lower coefficient values(in absolute sense) and therefore avoiding overfitting. However, Iworiso and Vrontos (2020) and Vrontos et al. (2021) only use two out of the five probit models of Nyberg (2011). Therefore, this paper investigates the following question: ”Can machine learning techniques improve the forecasting ability of standard probit models for predicting the direction of the US stock market?”

To formulate an answer to this question, financial and economic data are used. The US stock market as a whole is investigated as closely as possible by using the S&P 500 index, but also small and large size firm are used. Various often used explanatory variables are explored, such as the interest rates or volatility measures. Using these data, an in-sample analysis is conducted to see the performances of the models and determine explanatory variables and tunning parameters which would be used out-of-sample. In the out-of-sample part one-month ahead forecasts are made using an expanding window.

The in-sample analysis shows that using the Ridge, LASSO or Elastic Net probit models it is possible to improve the performances of all the standard models used in Nyberg (2011). However, these improvements are not transmitted into the out-of-sample forecasts. Out-of-sample show the Ridge, LASSO and Elastic Net probit models similar or worse forecasting abilities. Only in the case of large size firms is a LASSO probit model able to outperform the standard probit models. The worse performance of the extended models could be caused by not optimizing their tunning parameter for every forecast, due to time limitations.

The remainder of this paper is structured in the following manner. Section 2 provides an overview of the literature related to this research. Section 3 gives the precise sources of the used data and all variable used. Section 4 explains the methodology used in this research, consisting of all the models and evaluation tools. Section 5 shows the results. Finally, Section 6 is dedicated to the conclusions which can be drawn following this research.

2 Literature

A reason why the overall level of stock returns may not be predictable is that observed returns contain too much noise for accurate forecasting of the total amount of excess return, as is argued by Nyberg (2011) among others. However, the amount of research which is conducted concerning predicting the direction of stock markets is sparse when compared to the amount on the overall level of excess return. This Section provides the most important findings so far.

As mentioned in the introduction, research shows that the sign of excess stock return is to some extend predictable and that the results can be of economic significance (see, among others, Pönkä (2017), Chevapatrakul (2013)). The previously done research mostly employed different econometric methods, but papers by Christoffersen and Diebold (2006) and Christoffersen et al. (2006) show some theoretical frameworks. They investigate the relation between asset return volatility and the predictability of the asset return sign. In their papers they show that volatility indeed has explanatory powers. However, most commonly used are binary dependent time series models such as the logit and probit models. Leung et al. (2000) used in their paper a simple static logit and probit model, whereas for instance Anatolyev and Gospodinov (2010a) used an autologistic model specification.

The paper of Nyberg (2011) is an extension of the research done by Kauppi and Saikkonen (2008). They introduce the dynamic autoregressive probit model and Nyberg (2011) extends this model by including a restriction and including a recession forecast as explanatory variable. The idea of incorporating the recession forecast comes from previously done research. For instance, Fama and French (1989) indicate that business conditions have a high level of indication for future stock returns. Because of that result, incorporating the recession forecast as an explanatory variable is very useful as the results show in Nyberg (2011). The introduction of the restriction is based on the efficient market theory and led to the new 'error correction model'. The complete model specifications are shown in the methodology section. Nyberg (2011) found that the 'error correction model' performs better in forecasting the direction of the US stock market than the previously found probit models.

Over the past years, researchers have more and more found their way into using machine learning techniques. Ridge regression, LASSO and Elastic Net, introduced by Hoerl and Kennard (1970),Tibshirani (1996) and Zou and Hastie (2005) respectively, are three of the most famous examples of machine learning techniques. Iworiso and Vrontos (2020) and Vrontos et al.

(2021) show that using these methods improves the predictability of recessions and the direction of stock markets, when compared to the standard static and dynamic binary probit models also used by Nyberg (2011).

3 Data

This section indicates how the data is collected. By using similar data to Nyberg (2011), this paper aims at having the results best suited for comparing to his newly introduced models. The data used is provided by the Federal Reserve Economic Data (FRED), Wharton Research Data Service (WRDS), or the Kenneth R. French library. The data has a monthly frequency starting from January 1968 until December 2021. Table 1 shows the used variables and their description. The Standard&Poor's 500 US stock index[1] is used to model the US stock market. Data concerning size-sorted CRSP indices[2] are used to compare the predictive abilities of the models for small and large size firms. The domestic and foreign interest rates[3]are used as explanatory variables. The nominal stock return of the S&P500 index is used in two ways, namely as an explanatory variable and to use in investment simulations. This nominal return is a simple return. The term spreads used are defined as the difference between the short-term interest rate and the long-term interest rate. Finally, the dividend and earnings data[4] are used as explanatory variables.

Variable	Description
P_t	Standard&Poor's 500 US stock index
${P_t}^S$	CRSP small size firms index, first decile
${P_t}^L$	CRSP large size firms index, tenth decile
$r_t, {r_t}^S, {r_t}^L$	One-month excess return over the risk-free return (see Section 4.1)
${r_t}^n$	One-month nominal stock return from the S&P500 index
y_t	US recession periods (NBER)
i_t	Three-month US Treasury Bill rate, secondary market
R_t	10-year US Treasury Bond rate, constant maturity
$\Delta i_t, \Delta R_t$	First differences of i_t and R_t
${SP_T}^{US}$	US term spread between R_t and i_t
${SP_t}^{GE}$	German term spread between German long- and short-term interest rates
σ_t	Sum of squared daily stock returns in the S&P500 index within one month
DP_t	Dividends over the past year divided by the current stock index value, $DP_t = D_t/P_t$
EP_t	Earnings over the past year divided by the current stock index value, $EP_t = E_t/P_t$
N = 648	

Table 1: Dependent and explanatory variables

[1]Retrieved from WRDS: https://wrds-www.wharton.upenn.edu/

[2]Retrieved from the Kenneth R. French library: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[3]Retrieved form the FRED: https://fred.stlouisfed.org/

[4]Retrieved from the homepage of Robert Shiller's book Irrational exuberance:http://www.econ.yale.edu/~shiller/data.htm

Figure 1 shows the excess stock returns from January 1968 until December 2021 together with the recession periods. The graph shows that in times of recession, as indicated by the National Bureau of Economic research, the excess returns firstly often are negative. When the recession period comes to an end, excess stock returns typically are more often positive.

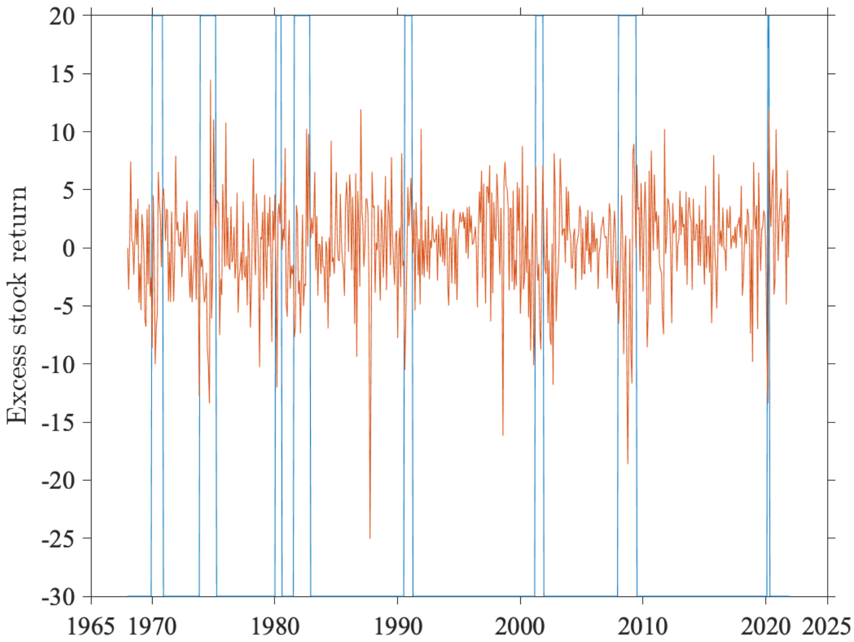

Figure 1: US excess stock returns r_t (%) and the NBER recession periods y_t (small areas between the blue lines) for the sample period from 1968M1 to 2021M12

4 Methodology

This Section provides an overview of the econometric models and methods. Firstly, the methods used by Nyberg (2011) are explained in detail. Thereafter the added machine learning techniques are provided and a description of how the recession forecast is made. Finally, statistical and economic measures are provided that give insights in the predictive ability of the models in the research.

4.1 Standard probit models

One of the characteristics of probit models is that the dependent variable is binary. Therefore, define $r_t = 100 \log \left(\frac{P_t}{P_{t-1}} \right) - rf_t$ as the continuously compounded excess stock return over the risk-free interest rate and let I_t be a sign indicator:

$$I_t = \begin{cases} 1, & \text{if } r_t > 0 \\ 0, & \text{if } r_t \leq 0 \end{cases} \tag{1}$$

This indicator takes on value one if the excess stock return is positive and zero otherwise. This means that I_t is a binary stochastic process. Then conditional on the information available up to $t-1$, Ω_{t-1}, the sign indicator has a Bernoulli distribution with probability p_t^I. Meaning $I_t \mid \Omega_{t-1} \sim B\left(p_t^I\right)$. Define E_{t-1} as the conditional expectation given the information set Ω_{t-1}.

Then the conditional probability of a positive excess stock return ($I_t = 1$) in a probit model needs to satisfy the condition:

$$p_t^I = E_{t-1}(I_t) = P_{t-1}(I_t = 1) = P_{t-1}(r_t > 0) = \Phi\left(\pi_t^I\right), \tag{2}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. This choice of $\Phi(.)$ defines a probit model, to implement a logit model this function should be changed to a logistic distribution function. Following equation (2), the conditional probability then is modeled by introducing a function of π_t^I which includes explanatory variables included in the information set. The benchmark forecasting model is defined as:

$$\pi_t^I = \omega + \boldsymbol{x}'_{t-1}\boldsymbol{\beta}, \tag{3}$$

where $\boldsymbol{x}_{t-1}$ is a vector containing the explanatory variables, ω is the constant and $\boldsymbol{\beta}$ is a vector containing the parameters of the p explanatory variables. This model is called static as it does not contain any lagged value of I_t. This model might be adequate as the literature indicates that there is not much correlation between two successive excess stock returns. The specification in equation (3) is extended to a dynamic model by including the lagged value of the sign indicator:

$$\pi_t^I = \omega + \delta_1 I_{t-1} + \boldsymbol{x}'_{t-1}\boldsymbol{\beta}. \tag{4}$$

When the value of δ_1 turns out statistically significant it indicates that, although there may not be much correlation, the lagged value still has useful predictive powers.

Kauppi and Saikkonen (2008) introduced a new variant of the static and dynamic probit models by including the lagged value π_{t-1}^I. This in return made the models autoregressive. The original static and dynamic models in equations (3) and (4) therefore now become an autoregressive model:

$$\pi_t^I = \omega + \alpha_1 \pi_{t-1}^I + \boldsymbol{x}'_{t-1}\boldsymbol{\beta}, \tag{5}$$

and a dynamic autoregressive model:

$$\pi_t^I = \omega + \alpha_1 \pi_{t-1}^I + \delta_1 I_{t-1} + \boldsymbol{x}'_{t-1}\boldsymbol{\beta}. \tag{6}$$

As Nyberg (2011) notes in his paper, by recursive substitution and assuming $|\alpha_1| < 1$ the model in equation (6) can be rewritten as:

$$\pi_t^I = \sum_{i=1}^{\infty} \alpha_1^{i-1}\omega + \delta_1 \sum_{i=1}^{\infty} \alpha_1^{i-1} I_{t-i} + \sum_{i=1}^{\infty} \alpha_1^{i-1}\boldsymbol{x}'_{t-i}\boldsymbol{\beta}. \tag{7}$$

Meaning that when several lagged values of the sign indicator or of the explanatory variable in $\boldsymbol{x}_t$ have meaningful predictive abilities, the two specifications in equations (5) and (6) are parsimonious forecast models.

The following step is to estimate the parameters of the probit models in equations (3)-(6). This is done by means of maximum likelihood estimation as described in Kauppi and Saikkonen (2008), de Jong and Woutersen (2011) and Pesaran (2016). The maximum likelihood estimator of θ, including all unknown parameters, is defined as follows:

$$\hat{\theta}_{ML} = argmax_{\theta}\, l(\theta), \tag{8}$$

where the log-likelihood function is defined as:

$$l(\theta) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} \left[I_t \log \Phi\left(\pi_t(\theta)\right) + (1 - I_t) \log\left(1 - \Phi\left(\pi_t(\theta)\right)\right)\right]. \tag{9}$$

The parameter estimates are found by maximizing the log-likelihood function using Matlab and the fmincon or fminunc function, which incorporates Quasi-Newton. More details on the code can be found in the appendix.

4.2 Error correction model

Nyberg (2011) extends the number of available models by incorporating the restriction $\delta_1 = 1 - \alpha_1$ and the assumption of $|\alpha_1| < 1$ in equation (6). This leads to the new "restricted" dynamic autoregressive model:

$$\pi_t^I = \omega + \alpha_1 \pi_{t-1}^I + (1 - \alpha_1) I_{t-1} + \boldsymbol{x}_{t-1}' \boldsymbol{\beta}. \tag{10}$$

The idea of restricting δ_1 derives its origin from the efficient market theory. The theory implies that lagged values of stock indicators should not help forecasting the direction the market takes in the future. This model is called an error correction model by Nyberg, as adding $-\pi_{t-1}^I$ to both sides of equation (10) leads to an error correction form. A more detailed description of the model can be found in Section 2.2 of Nyberg (2011).

4.3 Ridge, LASSO and Elastic Net probit models

By incorporating the structures of Ridge regression, LASSO and Elastic Net into the probit models, they are transformed into so called penalized likelihood binary probit models. The main effect of these structures is that they try shrinking the coefficients estimates towards zero. This can result in smaller variances and therefore lower prediction errors. The papers of Iworiso and Vrontos (2020) and Vrontos et al. (2021) provide a more detailed description of the extended models.

The maximum likelihood estimator of the Ridge probit model is defined as follows:

$$\hat{\theta}_R = \operatorname{argmax}_\theta \left\{ \sum_{t=1}^{T} \left[I_t \log \Phi\left(\pi_t(\theta)\right) + (1 - I_t) \log\left(1 - \Phi\left(\pi_t(\theta)\right)\right)\right] - \lambda_R \sum_{j=1}^{p} \theta_j^2 \right\}. \tag{11}$$

Equation (11) is in essence an extension on equation (9), where the shrinkage penalty ℓ_2-norm of θ and ridge tuning parameter λ_R, with $\lambda_R > 0$, are added.

The maximum likelihood estimator of the LASSO probit model is defined as follows:

$$\hat{\theta}_L = \operatorname{argmax}_\theta \left\{ \sum_{t=1}^{T} \left[I_t \log \Phi\left(\pi_t(\theta)\right) + (1 - I_t) \log\left(1 - \Phi\left(\pi_t(\theta)\right)\right)\right] - \lambda_L \sum_{j=1}^{p} |\theta_j| \right\}. \tag{12}$$

Again, is equation (12) in essence an extension on equation (9), where the shrinkage penalty ℓ_1-norm of θ and LASSO tuning parameter λ_L, with $\lambda_L > 0$, are added. One of the characteristics of the LASSO probit model is that the ℓ_1-norm can force some parameters towards zero when the λ_L is large enough. This is because the constraint region of the LASSO model is shaped in

the form of a diamond and therefore has "corners", compared to the round constraint region of Ridge due to its quadratic form. When the optimization process reaches one of these corners on the axis, it shrinks that parameter to zero. This is explained in more detail in Hastie et al. (2020).
The maximum likelihood estimator of the Elatic Net probit model is defined as follows:

$$\hat{\theta}_{EN} = \operatorname{argmax}_{\theta} \left\{ \sum_{t=1}^{T} \left[I_t \log \Phi\left(\pi_t(\theta)\right) + (1 - I_t) \log\left(1 - \Phi\left(\pi_t(\theta)\right)\right) \right] - \lambda_L \left[(1-\alpha) \sum_{j=1}^{p} \frac{\theta_j^2}{2} + \alpha \sum_{j=1}^{p} \left|\theta_j\right| \right] \right\}. \quad (13)$$

Again, is equation (13) in essence an extension on equation (9), where the shrinkage penalties ℓ_1-norm and ℓ_2-norm of θ and Elastic Net tuning parameters λ_{EN}, with $\lambda_{EN} > 0$, and $\alpha = 0.5$ are added.

The use of Ridge, LASSO and Elastic Net requires a value of λ for the machine learning techniques, which needs to be chosen beforehand. Iworiso and Vrontos (2020) optimizes this λ specifically for every forecast they make out-of-sample. However, this process needs a lot of computing time which was not possible given the available time. Therefore, the λ is only optimized ones using the in-sample period, and used for all out-of-sample forecasts. The choice of λ is made based on the CR value, meaning that the λ which provides the best CR value in-sample is used. The choice of using the CR value as the criteria is made because it measures the accuracy of the models. For all models and the recession forecast, λ's in the range of 10^{-5} up to 10^{-1} are investigated, where the steps are of magnitude 10^{-4} as also done by Vrontos et al. (2021).

4.4 Recession forecast

The second innovation Nyberg introduced was the implementation of a recession forecast as one of the explanatory variables. Nyberg (2011) used the recession indicator of the NBER. This indicator is defined as follows:

$$y_t = \begin{cases} 1, & \text{if the economy is in a recession at month t;} \\ 0, & \text{if the economy is in an expansion at month t.} \end{cases} \quad (14)$$

As this is a binary variable, it is possible to model it by means of a probit model and has the same structure as described in Section 5.1. The forecast which will be used is a six-month ahead forecast of the recession indicator. The conditional probability of the recession indicator in period $t + 5$, based on the information up to $t - 1$, is defined as follows:

$$E_{t-1}\left(y_{t+5}\right) = P_{t-1}\left(y_{t+5} = 1\right) = \Phi\left(\pi_{t+5}^{y}\right) = p_{t+5}^{y}. \quad (15)$$

To forecast p_{t+5}^{y}, the autoregressive probit model as defined in Nyberg (2011) is used:

$$\pi_{t+5}^{y} = c + \phi \pi_{t+4}^{y} + z_{t-1}' \boldsymbol{b} \quad (16)$$

where $z_{t-1} = (SP_{t-1}^{US}\ r_{t-1}^{n}\ SP_{t-1}^{GE})'$ is a vector containing the lagged term spread of the US and Germany and the nominal stock market return.

4.5 Forecast evaluation

When the different models are estimated and the forecasts are made, they can be evaluated using various goodness-of-fit and statistical predictability measures. The first measure is the pseudo-R^2 of Estrella (1998): ps$R^2 = 1 - (\hat{l}_u/\hat{l}_c)^{-(2/T)\hat{l}_c}$. This formula contains the maximum values of the estimated log-likelihood functions $\hat{l}_u$ and $\hat{l}_c$, where the unconstrained model and a constrained model with only a constant term are used. Using the maximum value of the log-likelihood function of the unconstrained model, it is also possible to compute the Bayesian information criterion of Schwarz (1978): BIC= $-2\hat{l}_u + k\,log(n)$. Here the k is the number of parameters in the used model and n is the number of observations.

The models in this research aim at forecasting the direction of the US stock market, a first measure which focuses on the number of correctly made forecasts is the CR. This measure provides the ratio of correctly made predictions (i.e., when the realized excess stock return was positive or negative and the model predicted a positive or negative excess return respectively). The value of this ratio is then compared to a threshold value to examine how the model performs. As the dependent variable is the excess stock return sign, the hypothesis of no predictability leads to a threshold value of 0.5. Therefore, this value of 0.5 is used.

To put the forecasts in an economical context, it is important to compute the return on investment which these models would have. Here the same trading strategy is used as in Nyberg (2011). This strategy consists of the following rules. Firstly, the investor has two options to invest in: risk-free Treasury Bills or stocks(in this case the S&P500 index). At the start of each month an investor makes their choice based on the made forecast. When ${p^I}_t \leq 0.5$ the investor choices the Treasury Bills, when ${p^I}_t > 0.5$ they choice to invest in stocks. Whenever an investor decides to change their investment from Treasury Bills to stocks or the other way around, they must pay transaction costs as found by Granger and Pesaran (2000). The "low-cost scenario" of Pesaran and Timmermann (1995) is used as in Nyberg (2011). Meaning that the transaction costs are $\zeta_s = 0.5\%$ of the total investment when moving from Treasury Bills to stocks and $\zeta_b = 0.1\%$ vice versa.

Following this trading strategy, the overall return on investment can be calculated. This return is denoted as the average annualized return RET. Next to the overall return on investment, it is also notable to look at the Sharpe ratio of a specific model found by Sharpe (1994). As this ratio takes the risk into account. The formula of this ratio is : SR= $\frac{\overline{RET}^k - \overline{RET}^{rf}}{\hat{\sigma}^k}$, where $\overline{RET}^k$ is the average portfolio return of the used model, $\overline{RET}^{rf}$ is the average Treasury Bill return, and $\hat{\sigma}_k$ is the sample standard deviation of portfolio returns RET^k. Portfolios with a high Scharpe ratio are preferred to portfolios with low values.

Finally, some statistical measures are used to compare the models. When looking at the CR, it cannot easily be seen if the value is significantly different from 0.5. In which case the predictions do not perform better than a random prediction. Therefore, a test proposed by Pesaran and Timmermann (1992) is used. Granger and Pesaran (2000) show that the test statistic is defined as follows:

$$PT = \frac{\sqrt{m}KS}{\left(\frac{\bar{P}_I(1-\bar{P}_I)}{\bar{I}(1-\bar{I})}\right)^{1/2}}. \tag{17}$$

Where $KS = HR - FR$ is the Kuipers score between the "hit rate" $HR = \frac{\hat{I}^{uu}}{\hat{I}^{uu}+\hat{I}^{du}}$ and the "false rate" $FR = \frac{\hat{I}^{ud}}{\hat{I}^{ud}+\hat{I}^{dd}}$, where the forecast classification is denoted by $\hat{I}^{uu} = \sum_{t=1}^{m} \mathbf{1}\left(I_t^f = 1, I_t = 1\right)$; $\hat{I}^{ud} = \sum_{t=1}^{m} \mathbf{1}\left(I_t^f = 1, I_t = 0\right)$; $\hat{I}^{du} = \sum_{t=1}^{m} \mathbf{1}\left(I_t^f = 0, I_t = 1\right)$; $\hat{I}^{dd} = \sum_{t=1}^{m} \mathbf{1}\left(I_t^f = 0, I_t = 0\right)$. In these cases, the f indicates that these are forecasts. Then finally to compute the PT statistic some other values are needed. The sample average $\bar{I}$ of the indicator function I_t over the m-month sample period and $\bar{P}_I = \bar{I}HR + (1 - \bar{I})FR$, where the statistic PT has an asymptotic standard normal distribution. The last evaluation tool is the test of Diebold and Mariano (1995). This test is used to statistically compare the investment returns between different models. As Nyberg (2011) notes, because the time-horizon is one month the statistic is as follows:

$$DM = \frac{\sqrt{m}\bar{d}}{\sqrt{\mathrm{var}(\bar{d})}} \tag{18}$$

and has a asymptotic standard normal distribution. In this formula the $\bar{d}$ is the average difference between the predicted excess stock returns. All models are then compared to a normal buy-and-hold strategy.

5 Results

This section provides the results and their interpretation. Firstly, the performance of the models is shown in-sample and thereafter the out-of-sample. Finally, the results are extended by comparing the best performing probit models to ARMAX models and investigating how the probit models perform when only small or large companies are considered.

5.1 In-sample

Investigating the results of the in-sample period is done in two steps. Firstly, only the standard probit models found in Nyberg (2011) are used and compared. Then in Section 5.1.2 the in-sample results of the extended models using machine learning techniques are shown.

5.1.1 Standard probit models

The in-sample period is the same as used in Nyberg (2011) (i.e., January 1969 until December 1988). This period is chosen because this provides an even comparison between results. The first twelve observations from 1968 are used as initial values. The objective of investigating the in-sample performance of the models is mainly to find the best performing explanatory variables to use in the out-of-sample research and to compare the findings to the paper of Nyberg (2011).

All variables in Table 1 from the three-month US Treasury Bill rate downwards (except the German term spread) are used as explanatory variables in the models, as well as the recession forecast. When the models with a single explanatory variable are compared based on the values of the psR^2, CR and the returns on investment, it is shown that the first differences of the short-term interest rate and the recession forecast have the best predictive ability. The first difference of the long-term interest rate and the dividend-price ratio also seem to have some predictive power, but the dividend-price ratio does not perform well when used in the static or autoregressive static model. The term-spread also does perform well when looking at the psR^2 and the CR but has a low return on investment and therefore would not be useful to use in practice. When two explanatory variables are incorporated in the model, it can be seen that a model with the recession forecast and the first difference of the short-term interest rate performs the best. As more parsimonious models are easier to use in practice, this model will be used for the out-of-sample part and no more variables are added.

Table 2 shows the estimation results of the different standard probit models when the recession forecast and first-difference of the interest rate are used as explanatory variables. In general, are the results found in-sample similar to those found in Nyberg (2011), with some exceptions such as RET, the sharpe ratio's, BIC and overall a little higher p-values which could be caused by using a slightly different data source. An important note to make is that the results in Table 2 show that the ECM model has the best return on investment in-sample, instead of the worst which was found by Nyberg (2011). Although that the ECM again has the lowest Sharpe ratio and the worst p-values for the statistical tests, the high RET could indicate better in-sample performance than is found by Nyberg (2011).

Table 2: Estimation results of in-sample standard probit models

	Static model	**Dynamic model**	**Static auto. model**	**Dynamic auto. model**	**ECM model**
Constant	0.059	0.015	0.056	0.014	-0.065
	(0.096)	(0.135)	(0.091)	(0.139)	(0.034)
π^I_{t-1}			0.061	-0.016	0.868
			(0.393)	(0.365)	(0.066)
I_{t-1}		0.078		0.083	
		(0.172)		(0.189)	
Δi_{t-1}	-0.304	-0.291	-0.298	-0.291	-0.153
	(0.129)	(0.132)	(0.134)	(0.133)	(0.065)
p^y_{t+5}	-0.497	-0.468	-0.470	-0.474	-0.015
	(0.255)	(0.263)	(0.297)	(0.297)	(0.048)
Log-L	-161.49	-161.38	-161.48	-161.38	-162.64
psR^2	0.039	0.040	0.039	0.040	0.030
BIC	339.43	344.69	344.88	350.17	347.20
CR	0.579	0.583	0.579	0.579	0.575
RET	8.8%	9.7%	8.8%	9.7%	9.8%
SR	0.163	0.158	0.131	0.158	0.086
PT	0.013	0.009	0.014	0.013	0.022
DM	0.025	0.017	0.017	0.018	0.038
DM_{ra}	0.004	0.003	0.003	0.003	0.022
$DM^{I_t=0}$	0.000	0.000	0.000	0.000	0.000
$DM^{I_t=0}_{ra}$	0.000	0.000	0.000	0.000	0.000

Notes: Standard errors are given in parentheses. The p-values of the PT and Diebold and Mariano (1995) tests are given. Finally, **ra** implies risk-adjusted returns, meaning that average return is standardized by the standard deviation of returns. And $I_t = 0$ implies that only periods with negative excess stock returns are investigated.

5.1.2 Ridge, LASSO and Elatic Net probit models

Firstly, only the recession forecasts are estimated using Ridge, LASSO or Elastic Net probit. The explanatory variables used are the recession forecast and the first difference of the short-term interest rate, as these are found to have the best predictive power in the standard models. Table 3 shows that using the more advanced recession forecasts causes the CR-value, psR^2 and the

RET to be generally similar for the models, and sometimes higher.

Table 3: psR^2, CR and RET values of models using standard, Ridge, LASSO or Elasitc Net computed recession forecasts

			standard			Ridge			LASSO			EN	
Model	x_{t-1}	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**
Static	p^y_{t+5}, Δi_{t-1}	0.039	0.579	8.8%	0.040	0.575	9.2%	0.040	0.575	9.2%	0.040	0.575	9.2%
Dynamic	p^y_{t+5}, Δi_{t-1}	0.040	0.583	9.7%	0.041	0.583	9.6%	0.041	0.579	9.4%	0.041	0.583	9.6%
Static A.	p^y_{t+5}, Δi_{t-1}	0.039	0.579	8.8%	0.040	0.579	9.4%	0.040	0.583	9.7%	0.040	0.583	9.7%
Dynamic A.	p^y_{t+5}, Δi_{t-1}	0.040	0.579	9.7%	0.041	0.583	9.6%	0.041	0.579	9.4%	0.041	0.583	9.6%
ECM	p^y_{t+5}, Δi_{t-1}	0.030	0.575	9.8%	0.030	0.579	10%	0.030	0.579	10%	0.030	0.579	10%

Secondly, only the models are estimated using Ridge, LASSO or Elastic Net probit. Table 4 shows that these more advanced probit models provide higher CR and RET values compared to the standard probit models. This means that these models are better at indicating the sign of the excess stock returns in-sample, although having lower pseudo-R^2 values due to the regularization terms in their likelihood functions.

Table 4: In-sample values of psR^2, CR and RET when only the model is estimated with standard, Ridge, LASSO or Elastic Net probit

			standard			Ridge			LASSO			EN	
Model	x_{t-1}	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**
Static	p^y_{t+5}, Δi_{t-1}	0.039	0.579	8.8%	0.031	0.608	10.6%	0.026	0.608	10.6%	0.025	0.608	10.6%
Dynamic	p^y_{t+5}, Δi_{t-1}	0.040	0.583	9.7%	0.017	0.596	9.7%	neg.	0.608	10.6%	0.010	0.583	10%
Static A.	p^y_{t+5}, Δi_{t-1}	0.039	0.579	8.8%	0.028	0.608	10.4%	0.026	0.608	10.6%	0.025	0.608	10.6%
Dynamic A.	p^y_{t+5}, Δi_{t-1}	0.040	0.579	9.7%	0.017	0.604	10.4%	0.021	0.596	10.7%	0.010	0.592	10%
ECM	p^y_{t+5}, Δi_{t-1}	0.030	0.575	9.8%	neg.	0.592	10.7%	neg.	0.592	10.8%	neg.	0.592	10.7%

Notes: "neg." indicates that the pseudo-R^2 is negative in that model.

As the results show that using the Ridge, LASSO or Elastic Net probit models to forecast the recession variable leads to similar and sometimes better forecasts of the sign of the excess stock returns, these will continue to be used. Using Ridge, LASSO or Elastic Net probit when only estimating the model itself, also shows promising results. When both the recession forecast and the model are computed using the Ridge, LASSO or Elastic Net probit models, the results of Tables 5-7 are found. As these models put a penalty on the parameters, in general are the parameters closer to zero than in the models estimated using the standard probit model.

Comparing the values of the CR and RET between the models using standard probit and the ones using Ridge probit, LASSO probit or Elastic Net probit, shows that all values of the models using machine learning techniques are higher. This result indicates that using the Ridge, LASSO and Elastic Net probit models it is possible to increase the in-sample predictive performance of all models, as these models can correctly predict the direction of the US stock market in more periods. This result may be unanticipated at first glance, as in most cases models which use regularization terms such as Ridge, LASSO and Elastic Net perform worse in-sample compared to their normal specification. However, the reason for this result is that the λ of these models are chosen based on the CR value, meaning that the models are partly optimized to provide good CR values while in the standard probit models this is not necessarily the case. That the Ridge, LASSO and Elastic Net probit models normally perform worse in-sample is also shown by the pseudo-R^2. The values of the models using machine learning techniques are lower in all cases, as these models are prevented from over-fitting and therefore explain less of the variance of the dependent variable. The results of the statistical measures show a very even picture between the techniques used, with nearly all models being significant at the 1% level. The ECM model has the highest p-values in general, but the differences are marginal.

Lastly, Table 6 shows sparsity results for the autoregressive variable π^I_{t-1} and the recession forecast in the ECM model. The variables already were insignificant in the models using the standard probit technique and the ℓ_1-norm then consequently eliminated them from the model.

Table 5: Estimation results of in-sample Ridge probit models

	Static model	Dynamic model	Static auto. model	Dynamic auto. model	ECM model
Constant	0.024	-0.05	0.013	-0.049	-0.109
	(0.091)	(0.098)	(0.083)	(0.103)	(0.062)
π^I_{t-1}			0.072	0.019	0.774
			(0.224)	(0.155)	(0.122)
I_{t-1}		0.067		0.076	
		(0.1020)		(0.116)	
Δi_{t-1}	-0.261	-0.147	-0.234	-0.176	-0.154
	(0.118)	(0.089)	(0.111)	(0.098)	(0.071)
p^y_{t+5}	-0.319	-0.098	-0.248	-0.136	-0.038
	(0.203)	(0.116)	(0.172)	(0.137)	(0.073)
Log-L	-162.53	-164.14	-162.96	-163.66	-168.18
psR^2	0.031	0.017	0.027	0.021	neg.
BIC	341.50	350.19	347.85	354.73	358.28
CR	0.604	0.600	0.604	0.604	0.596
RET	10.5%	9.9%	10.2%	10.5%	10.8%
SR	0.175	0.155	0.159	0.180	0.119
PT	0.001	0.002	0.001	0.001	0.003
DM	0.009	0.017	0.017	0.025	0.023
DM_{ra}	0.000	0.002	0.002	0.005	0.016
$DM^{I_t=0}$	0.000	0.000	0.000	0.000	0.000
$DM^{I_t=0}_{ra}$	0.000	0.000	0.000	0.000	0.000
λ_R	0.01961	0.12421	0.03411	0.08071	0.03221

Notes: $\lambda_R > 0.1$ because this gave the optimal CR-value and is only just larger than 0.1

Table 6: Estimation results of in-sample LASSO probit models

	Static model	Dynamic model	Static auto. model	Dynamic auto. model	ECM model
Constant	0.027	0.001	0.026	0.031	-0.134
	(0.096)	(0.135)	(0.096)	(0.140)	(0.130)
π^I_{t-1}			0.000	0.000	0.739
			(0.408)	(0.416)	(0.242)
I_{t-1}		0.027		0.045	
		(0.171)		(0.194)	
Δi_{t-1}	-0.258	-0.010	-0.254	-0.267	-0.091
	(0.127)	(0.122)	(0.128)	(0.131)	(0.096)
p^y_{t+5}	-0.338	-0.011	-0.337	-0.435	0.000
	(0.259)	(0.262)	(0.264)	(0.289)	(0.076)
Log-L	-163.31	-166.50	-162.63	-163.35	-175.02
psR^2	0.026	neg.	0.026	0.032	neg.
BIC	343.05	354.93	348.62	352.67	371.96
CR	0.608	0.600	0.608	0.588	0.588
RET	10.6%	9.8%	10.6%	9.7%	10.8%
SR	0.176	0.150	0.176	0.160	0.115
PT	0.001	0.002	0.001	0.006	0.008
DM	0.009	0.014	0.014	0.015	0.014
DM_{ra}	0.000	0.001	0.001	0.002	0.006
$DM^{I_t=0}$	0.000	0.000	0.000	0.000	0.000
$DM^{I_t=0}_{ra}$	0.000	0.000	0.000	0.000	0.000
λ_L	0.01131	0.08451	0.01161	0.00711	0.05301

Table 7: Estimation results of in-sample Elastic Net probit models

	Static model	Dynamic model	Static auto. model	Dynamic auto. model	ECM model
Constant	0.029	-0.022	0.027	-0.022	-0.098
	(0.094)	(0.126)	(0.091)	(0.128)	(0.059)
π^I_{t-1}			0.000	0.000	0.807
			(0.160)	(0.004)	(0.112)
I_{t-1}		0.020		0.020	
		(0.160)		(0.145)	
Δi_{t-1}	-0.262	-0.201	-0.258	-0.202	-0.132
	(0.125)	(0.123)	(0.125)	(0.131)	(0.056)
p^y_{t+5}	-0.345	-0.141	-0.339	-0.141	0.008
	(0.244)	(0.229)	(0.245)	(0.230)	(0.060)
Log-L	-163.03	-164.53	-163.09	-164.53	-170.89
psR^2	0.027	0.014	0.026	0.014	neg.
BIC	342.49	350.99	348.11	356.47	363.70
CR	0.608	0.583	0.608	0.583	0.592
RET	10.6%	10.1%	10.6%	9.8%	10.7%
SR	0.176	0.142	0.176	0.147	0.110
PT	0.001	0.010	0.001	0.007	0.005
DM	0.009	0.004	0.004	0.004	0.003
DM_{ra}	0.000	0.000	0.000	0.000	0.001
$DM^{I_t=0}$	0.000	0.000	0.000	0.000	0.000
$DM^{I_t=0}_{ra}$	0.000	0.000	0.000	0.000	0.000
λ_{EN}	0.01631	0.04011	0.01711	0.04011	0.05021

5.2 Out-of-sample

The most important part of evaluating the forecasting abilities of the models, is the out-of-sample forecasting. The performance out-of-sample indicates whether a model can be useful in practice. Previously done research finds that although models can have good in-sample performance, this does not imply good out-of sample performance as found by Han (2007). However, even if the predictive ability is small out-of-sample, Anatolyev and Gospodinov (2010b) still mention that using these models can lead to economically good performance.

The out-of-sample period in this research starts in January 1989 and ends in December 2021. The parameters are estimated using an expanding window, meaning that a one-month ahead forecast is made for every month using parameters estimated with all data available up to that month. The construction of the forecasts is done in two steps, namely first all recession forecasts are computed and thereafter are the models estimated and the out-of-sample forecasts computed. The computation of the pseudo-R^2 out-of-sample is more complicated than in-sample, the log-likelihood values out-of-sample are computed using the estimated parameters of each one-month ahead forecast separately.

In-sample it is shown that using the Ridge, LASSO and Elastic Net probit models improves the performance of the models, therefore these models are compared to the standard probit models to see whether this improvement remains out of sample. In case of the Ridge, LASSO and Elastic Net probit models, the recession forecast and the total model estimation is made using these types of probit models and using the $\lambda's$ found in-sample while using the given set of explanatory variables. The explanatory variables used are the recession forecast and the first-differences of the interest rates. These are chosen following the in-sample results and because Nyberg (2011) found these to have the best out-of-sample predictive ability. The results are shown in Table 8.

Firstly, the results of the standard probit models are compared. The results show that it

Table 8: Out-of-sample results of the standard, Ridge and LASSO probit models

		standard			Ridge			LASSO			EN		
Model	x_{t-1}	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**	psR^2	**CR**	**RET**
Static	p^y_{t+5}	0.013	0.621	10.9%	0.011	0.624	10.9%	neg.	0.624	9.4%	neg.	0.614	8.9%
Static	$p^y_{t+5}, \Delta R_{t-1}$	0.015	0.619	5.1%	0.009	0.619	3.9%	0.009	0.619	4.1%	0.010	0.614	4.0%
Static	$p^y_{t+5}, \Delta i_{t-1}$	0.016	0.614	4.7%	0.010	0.611	3.9%	0.009	0.616	3.7%	0.012	0.611	3.9%
Dynamic	p^y_{t+5}	0.013	0.614	10.0%	0.003	0.601	8.1%	neg.	0.596	8.1%	neg.	0.601	8.1%
Dynamic	$p^y_{t+5}, \Delta R_{t-1}$	0.017	0.621	10.6%	0.006	0.609	9.1%	neg.	0.609	3.4%	neg.	0.609	3.3%
Dynamic	$p^y_{t+5}, \Delta i_{t-1}$	0.018	0.621	10.6%	0.004	0.599	2.6%	neg.	0.611	3.5%	neg.	0.601	2.9%
Static A.	p^y_{t+5}	0.018	0.626	10.7%	0.011	0.624	10.7%	0.005	0.604	9.2%	0.007	0.619	10.6%
Static A.	$p^y_{t+5}, \Delta R_{t-1}$	0.022	0.601	3.5%	0.009	0.619	3.9%	0.002	0.619	4.1%	0.005	0.614	4.0%
Static A.	$p^y_{t+5}, \Delta i_{t-1}$	0.017	0.609	4.0%	0.009	0.609	3.6%	0.004	0.614	3.2%	0.008	0.609	3.9%
Dynamic A.	p^y_{t+5}	0.021	0.588	8.6%	0.003	0.601	8.1%	neg.	0.593	8.1%	neg.	0.601	8.0%
Dynamic A.	$p^y_{t+5}, \Delta R_{t-1}$	0.022	0.606	9.3%	0.007	0.611	9.2%	neg.	0.596	2.3%	neg.	0.596	2.4%
Dynamic A.	$p^y_{t+5}, \Delta i_{t-1}$	0.019	0.619	10.4%	0.006	0.609	3.2%	0.007	0.614	4.3%	neg.	0.599	2.7%
ECM	p^y_{t+5}	0.035	0.621	10.1%	0.004	0.599	9.3%	0.019	0.604	10.1%	0.001	0.596	9.4%
ECM	$p^y_{t+5}, \Delta R_{t-1}$	0.031	0.626	10.6%	neg.	0.556	7.4%	0.007	0.596	10.1%	0.007	0.586	9.0%
ECM	$p^y_{t+5}, \Delta i_{t-1}$	0.015	0.609	4.5%	0.007	0.588	4.5%	neg.	0.576	8.4%	0.011	0.583	4.3%
B&H				8.8%	0.139								

indeed seems possible to predict the sign of the excess stock market returns and therefore yield higher returns on investment than the B&H strategy, however it depends on the combination of model and explanatory variables. The predictive ability of the standard probit models is lower than in-sample, as is shown by the values of the pseudo-R^2. All Sharpe ratios of the standard probit models are higher than the ratio of the Buy-and-Hold strategy. A possible explanation for this could be that the investment in the models is able to change between the stock index and the risk-free interest rate, with the risk-free interest rate having a much smaller volatility.

Comparing the performances of the models, shows that on average the dynamic model has the best predictive ability out-of-sample of the standard probit models. This result is similar to what was found in-sample, but in-sample the ECM model has a little higher return on investment. When the results of the ECM model are investigated out-of-sample it still performs well, but not as good as the dynamic model. The ECM model yields higher returns than the B&H strategy two out of three times. The ECM model also seems to outperform the dynamic autoregressive model, as Nyberg (2011) also found. Noting that Nyberg (2011) uses a smaller out-of-sample period up to December 2006. The ECM model is in fact a dynamic autoregressive model, but with a restriction put on the dynamic parameter (Section 4.2). The worst performing models out-of-sample seem to be the static and static autoregressive models when using the recession forecast and another variable as explanatory variables. These models have especially low RET values, but the CR an SR are not very low compared to the other models. These low RET values over the entire period are caused by low returns in the beginning of the forecasting period. The best performing models on the other hand are the static and static autoregressive models using only the recession forecast as explanatory variable. These two models yield the highest CR, RET and SR values of all standard probit models. Which type of standard probit model performs the best is difficult to compare to the results found in Nyberg (2011) as the dynamic type models on average outperform the static type models, but the best two models are of the static type. Where Nyberg (2011) found that the static type models outperformed the dynamic type models, excluding the ECM model.

Investigating the results of the Ridge probit models out-of-sample shows a very different picture compared to the results found in-sample. The only similarity is that the Ridge probit models have lower pseudo-R^2 values. The improvements which the Ridge probit models shows in-sample compared to the standard probit models completely disappeared. Only in the case of the static autoregressive model using the recession forecast and the first-difference of the long-term interest rate yields the Ridge probit model a higher return on investment. When looking at the CR values of the Ridge probit models, the values are higher than the values of the standard probit models when the RET values are similar. For instance, in the case of the dynamic autoregressive model with the recession forecast as explanatory variable. This could be the result of the fact that the λ_R tuning parameter is chosen based on the highest CR value in-sample. This could imply that in-sample the higher CR value translates into higher returns on investment, but out-of-sample this is not the case. This fact is nevertheless strange as it therefore would be suspected that all CR values should be higher, but they are not. Although the results are not better than the standard probit models, the Ridge probit models are still in some models able to outperform the B&H strategy in terms of RET and the Sharpe ratios are nearly always higher. The best performing Ridge probit models are the same as found using the standard probit models, namely the static and static autoregressive model using only the recession forecast as explanatory variable. Giving also very similar results.

The results using the LASSO probit model are similar to those using the Ridge probit model. The LASSO probit model is also not able to outperform the standard probit model out-of-sample, while this was the case in-sample. The LASSO probit model is only able to yield higher return on investment using the static autoregressive model with the recession forecast and the long-term interest rate and in case of the ECM model with the recession forecast and the short-term interest rate. The LASSO probit model also seems to have the same characteristic concerning the CR values as the Ridge probit model when the RET values are similar with the standard probit model. Finally, the LASSO probit model seems to be the only technique which is able to yield relatively high RET values for all ECM models. The best performing models using LASSO Probit are therefore the ECM models using only the recession forecast or also using the first-difference of the long-term interest rate.

The last model to investigate is the Elastic Net probit. The regularization term of this model is a combination of the ones used by Ridge and LASSO probit and the results also show this. The out-of-sample performance of the Elastic Net probit model is nearly always worse than at least one of the other two models using machine learning techniques. Therefore, performs the Elastic Net probit model in general the worst out of the four probit models considered. However, the Elastic Net probit model is still sometimes able to outperform the B&H strategy and the best performing model of the Elastic Net type is the static autoregressive model using only the recession forecast as an explanatory variable.

Figure 2 shows the out-of-sample probability forecasts of the best performing standard, Ridge, LASSO and Elastic Net model. An enlarged version can be found in the Appendix. The graph shows that the static autoregressive Elastic Net model (green line) and static Ridge model (red line) provide almost the same probabilities. The standard static model differs slightly from those two models as the standard static model does not include the regularization terms. This fact is visible as the probabilities of the static Ridge and static autoregressive Elastic Net model are less extreme since their parameters are closer to zero. A very visible difference between the use of the static models and the ECM model is the volatility of the probabilities over time. The ECM model shows very large spikes and changes aggressively between investing in stocks or the risk-free interest rate, while the static model moves very "static" and only during economical bad times falls below the 0.5 threshold. For instance, during the dot-com bubble 2001-2003, the economic crisis in 2008 or at the beginning of the COVID-19 pandemic in 2020. It should also be noted that all models are above the 0.5 threshold most of the time, meaning that in most periods the investment in stocks is chosen. These characteristics are model specific and not caused by using the Ridge or LASSO probit, as Nyberg (2011) also describes these behaviors on the shorter forecasting period.

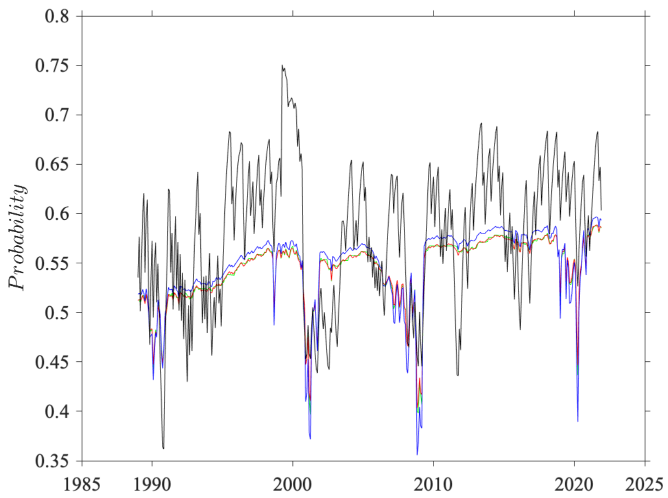

Figure 2: Out-of-sample probability forecasts of the static model with p^y_{t+5}(blue), static Ridge model with p^y_{t+5}(red), ECM LASSO model with p^y_{t+5} and ΔR_{t-1}(black) and the Elastic Net model with p^y_{t+5}(green)

Table 9 shows the results of the statistical tests belonging to the best performing standard, Ridge, LASSO and Elastic Net model. As the standard static and static Ridge model are similar, they also give similar results. The only difference is that when only periods of negative excess returns and normalized returns are considered, the returns of the standard static model are at a 5% significantly different than those of the B&H strategy. Where the returns of the static Ridge model are significantly different at a 10% level. The difference between the p-values of the static Ridge and static autoregressive Elastic Net model is remarkable. These two models provide nearly equal probability forecast, shown in Table 2, but the test results are not. The static autoregressive Elastic Net model performs worse and is not significant at the 10% level in the case of the PT and $DM^{I_t=0}_{ra}$ tests, while the static Ridge model is. This indicates that the static autoregressive Elastic Net model is not able to significantly predict the sign of the excess stock return. When the results of the ECM LASSO model are investigated, it is outperformed by the static (autoregressive) models. Only in the case of $DM^{I_t=0}_{ra}$ does it have a better p-value.

Table 9: Statistical test for the best performing standard, Ridge, LASSO and Elastic Net probit model

Model	x_{t-1}	**CR**	**RET**	**SR**	**PT**	**DM**	DM_{ra}	$DM^{I_t=0}$	$DM_{ra}^{I_t=0}$
Static	p^y_{t+5}	0.621	10.9%	0.199	0.096	0.154	0.017	0.000	0.029
Static Ridge	p^y_{t+5}	0.624	10.9%	0.195	0.092	0.141	0.019	0.000	0.088
ECM LASSO	p^y_{t+5}, ΔR_{t-1}	0.596	10.1%	0.186	0.503	0.381	0.071	0.000	0.002
Static Auto. EN	p^y_{t+5}	0.619	10.6%	0.189	0.189	0.231	0.040	0.000	0.111

Notes: The p-values of the PT and DM tests are shown, where the base trading strategy is the Buy-and-Hold strategy

5.3 Comparison probit and ARMAX models

Up to this point, only probit models have been investigated and compared. This Section introduces the ARMAX model and compares the out-of-sample performances of the previous probit models and ARMAX type models. It is possible to compare the models, as both have the excess stock return as their dependent variable. Noting that in the probit models the excess stock return is made binary, meaning a value of 1 if the excess stock return is positive and o otherwise. The ARMAX models used to compare to the probit models are the same as used in Section 4.4 of Nyberg (2011), but now run on a longer time period. These models were chosen based on their BIC. The results of these models are shown in Table 10.

Table 10: Out-of-sample results of ARMAX models

Model	x_{t-1}	**CR**	**RET**	**SR**
B&H			8.8%	0.139
ARMAX(1,0)	p^y_{t+5}	0.586	9.6%	0.184
ARMAX(1,0)	p^y_{t+5}, SP^{US}_{t-1}	0.576	4.2%	0.168
ARMAX(1,0)	p^y_{t+5}, ΔR_{t-1}	0.576	3.7%	0.138
ARMAX(1,0)	p^y_{t+5}, Δi_{t-1}	0.583	4.2%	0.164
ARMAX(2,0)	p^y_{t+5}	0.588	8.9%	0.169
ARMAX(2,0)	p^y_{t+5}, SP^{US}_{t-1}	0.578	8.7%	0.170
ARMAX(2,0)	p^y_{t+5}, ΔR_{t-1}	0.571	8.2%	0.152
ARMAX(2,0)	p^y_{t+5}, Δi_{t-1}	0.588	9.4%	0.173

Notes: The ARMAX(p,0) models used are defined as follows: $r_t = a + \sum_{i=1}^{p} b_i r_{t-i} + x'_{t-1}d$. To estimate the ARMAX models in Matlab, The Oxford MFE Toolbox is used.

Investigating the results in Table 3 shows that nearly all probit models are able to give higher CR values than the ARMAX models, as the ARMAX give values around the 0.58 and the probit models are more around the 0.60 and above. When the returns on investment are compared it shows that the ARMAX(2,0) models are performing well and in two out of four times give higher returns than the 8.8% of the B&H strategy. However, the best performing model of the ARMAX type is the ARMAX(1,0) model using only the recession forecast as an explanatory variable. The predicted excess stock returns of the best performing ARMAX(1,0) and ARMAX(2,0) model are shown in Figure 3. When this best performing ARMAX model is compared to the probit models, it has lower CR, RET and SR values than the best performing standard, Ridge, LASSO and Elastic Net probit models. The fact that the probit models outperform the ARMAX models is in line with the results found on the shorter time period by Nyberg

(2011).

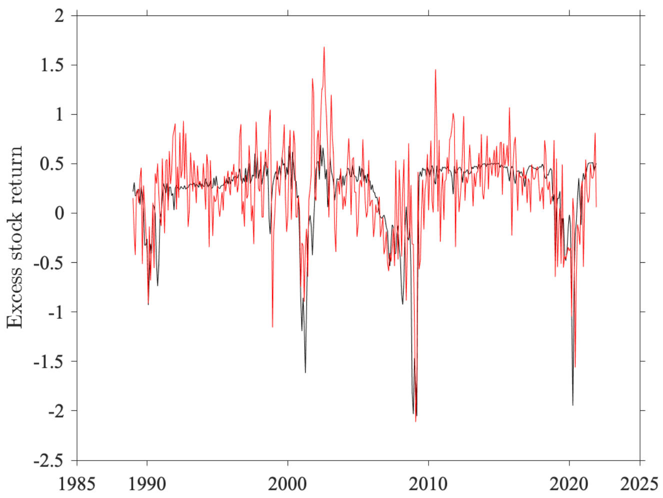

Figure 3: Out-of-sample excess stock return forecasts of the ARMAX(1,0) model with p^y_{t+5}(black) and of the ARMAX(2,0) model with p^y_{t+5} and Δi_{t-1}(red)

5.4 Small and large size firms

In the final part of this research, the forecasting abilities of the best performing models are investigated using data concerning small or large size firms separately instead of the US stock market as a whole. Research has shown that small size firms can react differently to market news than large size firms, as for example was shown by Chan and Chen (1991).

For the out-of-sample forecasting the best standard, Ridge, LASSO and Elatic Net model are used to calculate the one-month ahead forecasts of the small and large size firms in the same manner as done in Section 5.2. The excess stock returns of the small and large size firms ${r_t}^S$ and ${r_t}^L$ are made using the size-sorted CRSP decile portfolios, where the small size firms belong to the first decile and the large size firms are the tenth decile. As Nyberg (2011) also mentioned concerning the shorter time period, there is a large similarity between the values of the excess stock return indicator of the S&P 500 index and the indicators of the small and large size firms. The large size firms have in 96.8% of the months the same value and the small size firms in 79.8% of the months. The results of the out-of-sample forecasts of the small and large size firms are shown in Table 11.

Table 11: Out-of-sample results of small and large size firms' returns

Firm size	Model	x_{t-1}	psR^2	CR	RET	SR	PT	DM	DM_{ra}	$DM^{I_t=0}$	$DM_{ra}^{I_t=0}$
Small	**B&H**				11.3%	0.141					
	Static	p^y_{t+5}	0.011	0.596	13.2%	0.174	0.247	0.360	0.104	0.002	0.491
	Static Ridge	p^y_{t+5}	0.006	0.599	12.5%	0.158	0.207	0.168	0.093	0.024	0.030
	ECM LASSO	p^y_{t+5}, ΔR_{t-1}	0.008	0.583	11.0%	0.209	0.030	0.624	0.052	0.000	0.000
	Static Auto. EN	p^y_{t+5}	0.002	0.591	11.5%	0.146	0.679	0.996	0.674	0.030	0.030
Large	**B&H**				11.3%	0.182					
	Static	p^y_{t+5}	0.015	0.644	13.0%	0.232	0.013	0.190	0.020	0.000	0.023
	Static Ridge	p^y_{t+5}	0.005	0.634	11.3%	0.181	0.190	0.317	0.301	0.315	0.000
	ECM LASSO	p^y_{t+5}, ΔR_{t-1}	0.020	0.636	13.3%	0.242	0.020	0.275	0.019	0.000	0.001
	Static Auto. EN	p^y_{t+5}	neg.	0.631	11.3%	0.181	0.699	0.334	0.324	0.359	0.000

The results of the small size firms show that only the ECM LASSO model is able to provide a p-value below the 5% level for the PT test, showing that it is able to predict the sign of the excess stock returns. However, all models have low pseudo-R^2 values. Furthermore, the DM tests indicate that the static Ridge and ECM model are able to provide significantly different returns to the Buy-and-Hold strategy when there is adjusted for the risk, at the 10% level for all returns and at the 5% or even 1% level in case of only months with negative excess stock returns. Investigating the CR values shows that these are all lower than when forecasts are made for the S&P 500 index. Lastly, only the ECM LASSO model is not able to yield higher RET values than the B&H strategy, but all returns were higher than those found for the S&P 500 index.

When the large size firms are used, the models are able to give higher CR values than for small firms and the S&P 500 index. The B&H strategy yields the same RET as for the small firms but has a larger Sharpe ratio due to the fact that large firms tend to have less volatile returns. This is also shown by the Sharpe ratios of the models, which are higher than those in case of the small firms. While even having lower returns on investment in two out of three models. The results of the statistical tests show again that the ECM LASSO model has many significant p-values, as well as the static model this time. Only the normal DM test is not significant at the 5% level. Lastly, the ECM LASSO model provides the best return on investment of all models.

Figure 4 shows the probability forecasts of the best model when the small or large size firms are investigated. The characteristics of the static and ECM type model are again very visible.

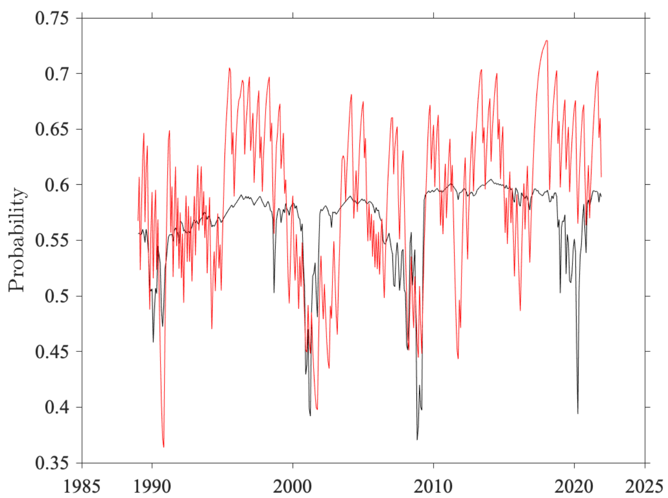

Figure 4: Out-of-sample probability forecasts of the static model with p^y_{t+5} for small firms(black) and of the ECM LASSO model with p^y_{t+5} and ΔR_{t-1} for large firms(red)

6 Conclusion

This paper investigates the possibility of improving the standard probit models for predicting the direction of the US stock markets, following the general setup of Nyberg (2011). This is done by implementing the machine learning techniques Ridge, LASSO and Elastic Net. Therefore, the following research question is considered: "Can machine learning techniques improve the forecasting ability of standard probit models for predicting the direction of the US stock market?"

The results of the Ridge, LASSO and Elastic Net probit models show that it is possible to improve the performance of the standard probit models in-sample, as they yield higher returns on investment and have higher CR values which indicate the amount of correctly forecasted periods. However, the Ridge, LASSO and Elastic Net probit models do not improve the performance of forecasting the direction of the US stock market out-of-sample as a whole or when only small size firms are considered. Only an ECM LASSO model was able to outperform the standard probit model, when only large size firms were considered. The results show that the best Ridge, LASSO and Elastic Net probit models compare favorably against sign forecasts made by ARMAX models.

The results found in this paper show that it is possible to improve the forecasting ability of standard probit models in-sample, but not out-of-sample. This is shown by out-of-sample analysis when the tunning parameter of the machine learning technique is optimized once, based on the in-sample period. Repeatedly re-optimizing the tunning parameter every time period before making a new forecast was not possible due to time limitations. This perhaps leads to the worse out-of-sample performance, as forecasts made in 2020 are using a parameter optimized for data more than 30 years before. Therefore, an interesting extension on this paper would be an out-of-sample analysis where the tunning parameter is repeatedly re-optimized every time period. Another part of the analysis which could be extended, is the number of machine learning techniques used. As more advanced methods are available.

References

Anatolyev, S. and Gospodinov, N. (2010a). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, 28(2):232–245.

Anatolyev, S. and Gospodinov, N. (2010b). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, 28(2):232–245.

Breen, W., Glosten, L. R., and Jagannathan, R. (1989). Economic significance of predictable variations in stock index returns. *The Journal of Finance*, 44(5):1177–1189.

Chan, K. C. and Chen, N.-F. (1991). Structural and return characteristics of small and large firms. *The Journal of Finance*, 46(4):1467–1484.

Chevapatrakul, T. (2013). Return sign forecasts based on conditional risk: Evidence from the uk stock market index. *Journal of Banking Finance*, 37:2342–2353.

Christoffersen, P., Diebold, F., Mariano, R., Tay, A. S., and Tse, Y. K. (2006). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: International evidence. Pier working paper archive, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Christoffersen, P. F. and Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52(8):1273–1287.

de Jong, R. M. and Woutersen, T. (2011). Dynamic time series binary choice. *Econometric Theory*, 27(4):673–702.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business Economic Statistics*, 13(3):253–63.

Estrella, A. (1998). A new measure of fit for equations with dichotomous dependent variables. *Journal of Business & Economic Statistics*, 16(2):198–205.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Granger, C. W. and Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7):537–560.

Han, Y. (2007). Return predictability, economic profits, and model mis-specification: How important are the better specified models? *SSRN Electronic Journal*.

Hastie, T., Tibshirani, R., and Wainwright, M. (2020). *Statistical learning with sparsity: The lasso and generalizations*. CRC Press, Taylor amp; Francis Group.

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

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

Kauppi, H. and Saikkonen, P. (2008). Predicting u.s. recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791.

Leung, M. T., Daouk, H., and Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, 16(2):173–190.

Nyberg, H. (2011). Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting*, 27(2):561–578.

Pesaran, M. H. (2016). *Time series and panel data econometrics*. Oxford University Press.

Pesaran, M. H. and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business Economic Statistics*, 10(4):461–465.

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance*, 50(4):1201–1228.

Pönkä, H. (2017). Predicting the direction of us stock markets using industry returns. *Empirical Economics*, 52(4):1451–1480.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461 – 464.

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58.

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

Vrontos, S. D., Galakis, J., and Vrontos, I. D. (2021). Modeling and predicting u.s. recessions using machine learning techniques. *International Journal of Forecasting*, 37(2):647–671.

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.

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.

Appendix

Code description:

All code and data used in for paper can be found in the zip-file `Code&Data\_BachelorThesis\_RubenSchorno\_545096.zip` . The sources of the data are described in the data Section of this paper. The code is programmed and run in version 9.9 of Matlab. The ReadMe.txt file in the zip-file explains how to run the code in order to find the results of this paper. The most important part of the code is the optimization process of the log-likelihood functions to find the maximum likelihood estimates. This is done by programming the log-likelihood functions as separate files, which returns the log-likelihood value with a (-) sign in front. The transformation is made to make it a minimization problem instead of a maximization problem, because Matlab is better at minimizing. Then the fmincon or fminunc function of Matlab is used to solve the minimization problem and return the MLE parameters, together with other results such as the maximum likelihood value and hessian. The fmincon is used when the model considered has an autoregressive parameter or a constraint such as the ECM model. Using the parameter estimates, the forecasts can be made. The results are then used to compute all tests and performances measures using mostly simple for-loops.

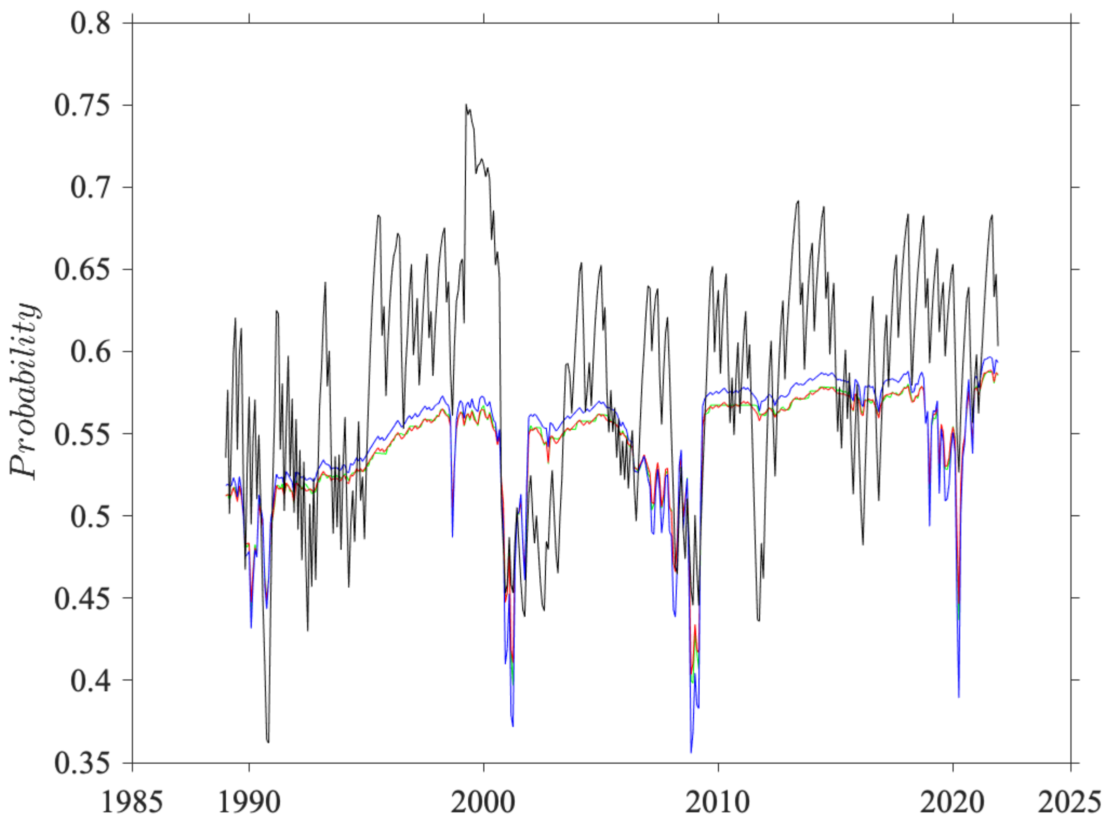

Figure 5: Out-of-sample probability forecasts of the static model with p^y_{t+5}(blue), static Ridge model with p^y_{t+5}(red), ECM LASSO model with p^y_{t+5} and ΔR_{t-1}(black) and the Elastic Net model with p^y_{t+5}(green)