A Research of Combination between Pricing Strategy and Forecasting Performance

Yiren Liu[1,a,*], Yuli Bian[1,b]
[1] *Viterbi School of Engineering, University of Southern California, 3335 S Figueroa, Los Angeles, CA, U.S.A.*
[2]*College of Economics and Management, South China Agricultural University, 483 Wushan Rd, Tianhe Dist, Guangzhou City, Guangdong Province, P.R.C.*
a. yirenliu@usc.edu, b. bylsysu@foxmail.com
**Yuli Bian*

Keywords: S&P500, Economic Indicators, Technical Indicators, Forecast, Regression.

Abstract: Using assessing pricing model and mathematical method, this paper find both return timing and volatility timing are shown to be profitable and their result has proven to be statistically sound. Given the promising performance, we are thinking about exploring ways to combine these two timing strategies. Gradient boost machine learning algorithm are also used to predict SPY return with 20 variables including some constructed variables and develop trading strategies using the forecasted return and stock volatility. After getting the forecast return, using the forecast return to determine the weight of buy and hold S&P500 to create the portfolio. The simulated strategy shows a significant increase in Sharpe ratio compared to the return of just hold S&P500. We also extend our strategy from daily rebalancing to monthly trading which also shows significant Sharpe ratio.

1. Introduction

Market timing is one of the most easy-to-implement strategies. While we are astounded by the result of volatility timing and return timing, we have not seen a combination of these two. However, whenever we discuss the performance of a trading strategy, often the first measure we ask about is the Sharpe ratio, which is a combination of return and volatility. Inspired by the formation of the Sharpe ratio, we decide to build strategy based on the forecasted return over volatility.

It is well-known that equity return is hard to predict. In response to the problem, and some variable we choose maybe not connect with the return, and adding those variables will disturb the predicted travel. So we introduce machine learning algorithms to implement. We consider two major machine learning algorithms, Lasso regression and gradient boost. We then compare the correlation of the forecasting return and SPY daily return and choose the algorithm that produced better result. Finally, we combine the chosen forecasting return and volatility to time the market. Detailed data processing and strategy implementation are included in the rest of the sections.

We use three methods to choose variables and compare the performance. Although the correlation screening method used by Hull and Qiao(2018)[1] has proven to be effective, we use different variables that considered to be useful in predicting market returns. And we choose three different methods to predict market returns, which are correlation screen, PCA LASSO and Gradient Boost.

After getting the forecast return, we build the weight function using the forecast return. Let m(t) denote the monthly forecast excess return of S&P500 on month t, then set portfolio weight equal to $w(t)= \min\{1.5, \max\{-.5, 100\times m(t)\}\}$, which makes me set between -0.5 to 1.5 weight of money to short or long the S&P500. So if my forecast return in next month is 0.01(a 1% return in the next month or 12% annualized), then I will put all the money in market index. Then my portfolio real return will equal to w(t)*return(t). So if my forecast return is positive I will set the weight positive, if my forecast return is negative I will set the weight negative(short). Then no matter the index go up or go down I will earn money if my prediction works well.

We develop the correlation screening method from Hull and Qiao(2018). Considering data from two regimes, economic and technical, and collect them as many as we can. For our return prediction, we used 30 different variables which will be explained in detail in the following section. Furthermore, we also construct further variables such like momentum. And in each time t, we choose the variables whose correlation with market return are larger than 0.087(different numerical values are used, but 0.087 gets the best Sharpe ratio).

In the PCA&LASSO regression method, we use the package inside python and set the window to 120. And there is no need to set correlation to choose variables.

In the gradient boost method, we choose the variables inside XGboost package and we get a feature importance, which is a factor that could reflect the predict effectiveness. And then we choose the variables of 7, 6 and 5 importance in predicting. These variables could help us better predict returns and avoid the relatively irrelevant variables.

After that, we draw the line of daily actual return line and the daily return line of each methods. In this way, we could directly compare their performance in different durations. It shows that correlation screen more in line with actual situation. And then we draw the line of cumulative return and actual return and the three methods.

The paper proceeds as follows. Section 1 outlines our data and variables. Section2 show the return forecasting implementation details, using Correlation Screen, PCA+LASSO and XGBoost . Section 3 show the differences of forecast results and strategy comparison, and Section 4 reveal the strategy implementation detail including transaction cost and consider about the interest rates of interest and comprehensive portfolio comparison. Section 5 analyzing the results and discussing the implications of our findings. Section 6 conclusions.

2. Data and variables

This section describes the forecasting variables and data source. As mentioned, we considered two different kinds of data, economic indicators and technical indicators.

2.1. Economic Indicators

Economic research and analysis were performed to determine which indicators —measures of different aspects of the economy—affected market return the most. Indicators with high data report rates and economic significance were selected, and any relationships or overlaps between the indicators were noted, as indicator relationships would be important for later regression analysis the indicators were then grouped according to a collection of overarching attributes.

Count is the number of each indicator, and for some indicators who has smaller number, that is because they were founded later. Standard deviation is annualized, plus all the other indicators during the entire length of time.

Table 1: Economic Indicators.

	CPI	UR	M1	Pi	Cs	Cl_Loans	Recession	GDP	VIX	Oil	Di	Tbill_s	GOLD
Count	462	462	462	462	462	462	462	462	341	389	461	461	461
Mean	168.07	5.62	1364.3	32172.5	86.61	965.4	0.1212	10155.16	19.25	43.62	92.89	1.023	632.30
Std	49.14	1.57	826.60	6482.82	12.52	491.1	0.3267	5091.71	7.66	29.80	14.91	0.903	418.08
Min	77.80	3.10	377.80	21288.0	51.70	281.85	0	2789.84	9.51	11.13	68.76	-2.010	254.05
25%	124.70	4.50	787.73	26964.8	77.05	594.09	0	5695.37	13.56	19.61	83.39	0.330	350.70
50%	165.60	5.20	1129.6	32166.5	90.20	885.22	0	9470.71	17.42	29.43	90.49	0.980	404.40
75%	214.58	6.50	1500.9	38044.8	94.88	1242.90	0	14628.02	22.55	65.09	98.60	1.730	899.50
Max	251.99	10.40	3693.1	43614.0	112.00	2202.84	1.0000	20411.92	68.51	141.06	145.74	2.870	1854.00

- Consumer Price Index(CPI)

Consumer Price Index(CPI) is a measure of the average monthly change in the price for goods and service. Changes in CPI are used to assess price changes associated with the cost of living. CPI is one of the most frequently used statistics for identifying period of inflation or deflation. There a negative correlation between monetary inflation and dividend yield. (Jeffrey Oxman, 2011[2]). So we consider that there is some correlation between CPI and market returns.

- Unemployment Rate(UR)

Unemployment rate is an indicator of economic and stock market health, generally interpreting a drop in the unemployment rate as bad for stocks. Sometimes, falling unemployment rate is a trigger for increases in the Federal Reserve target interest rate and adverse stock market reactions. There is some support for belief that a strong increase (decrease) in the U.S. unemployment rate is good (bad) for the U.S. stock market over the next year or so.(Steve LeCompte, 2018[3])

- Real Disposable Personal Income (Pi)

Real Disposable Personal Income refers to the income of individuals after paying direct taxes and government fees. Actual disposable income refers to the after-tax and welfare income available to families after adjusting for price changes. Disposable personal income is often monitored as one of the many key economic indicators used to gauge the overall state of the economy. In theory, the impact that disposable income has on the stock market is that a widespread increase in disposable income leads to increases in stock valuations and, therefore, increases the overall value of the stock market.(EVAN TARVER, 2018[4])

- Real Gross Domestic Product (GDP)

The Real Gross Domestic Product (GDP) is a measure of total economic output, which is the aggregate value of all goods and services produced per year of a nation. This is the indicator most commonly used to compare economic progress between nations. It is reported quarterly which is not as frequent as more specific indicators. GDP would lead to increase in stock market performance. (Crispin Ochieng Ogutu, 2011[5]) And in this paper, the author suggest that Investors should buy shares when GDP is performing poorly and sell when the GDP performs well for them to have valuable investments.

- 10-Year Treasury Constant Maturity Minus 2-Year Treasury

The 10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity is the difference between the ten and two-year measurements of constant maturity. As opposed to ten-year maturity alone, this rate increases sharply when there is an increase in volatility, as a result of the fluctuating differences between the two rates.

- Money supply

Money supply is one of the most basic parameters in an economy and measures the abundance or scarcity of money. Stock prices tend to move higher when the money supply in an economy is high. Plenty of money circulating in the economy both makes more money available to invest in stocks and also makes alternative investment instruments, such as bonds less attractive. Based on vector error correlation Granger causality tests, findings show evidence of strong and statistically significant inverse causal relationship between money market interest and stock market returns.(Trust Kganyago, Victor Gumbo, 2015[6])

- Cl loans(BUSINESS LOAN)

Business Loans are commercial and industry loans, all commercial loans. Business loans refer to debts that a company is obliged to repay in accordance with the terms and conditions of the loan. According to the U.S. Small Business Administration, before applying for a loan from a lender, the owner of the business must understand how the loan works and what the lender wants to see from the owner. The development of the market is inseparable from the development of enterprises. And enterprises need business loans to make them thrive. To a certain extent, the market is connected with commercial loans.

- VIX

VIX is a popular measure of the stock market's expectation of volatility implied by S&P 500 index options and it is referred to as the fear index or the fear gauge. There is a lot study about the relationship between the VIX and market returns. The empirical analysis of this study has important practical implications for financial market practitioners, as it shows that they can use the VIX futures term structure not only as a proxy of market expectations on forward volatility, but also as a stock market predictive tool.(Nikolaos L. Hourvouliades, 2018[7])

- Consumer

University of Michigan Consumer Sentiment Index as an indicator of U.S. economic and stock market health, generally interpreting a jump (drop) in sentiment as good (bad) for future consumption and stocks. SPY may increase with Consumer Sentiment Increasing. Consumer confidence goes up and down with stock returns.(Kenneth L. Fisher, 2002[8])

- Gold

Gold has an inverse relationship with the dollar, stocks markets also have a deep connection to the Gold. Investors commonly perceive gold as a haven in the event of a severe stock market downturn. when we experience a global market decline, stocks and SPY may also move downward. The study finds that there is no causal relationship exist in between Gold Price and Stock market price in the short run. However Gold price and Stock market price are co-integrated indicating long-run equilibrium relationship between them, and they move together.(Naliniprava Tripathy, 2016[9])

- Recession

Recession is a slowdown or halt to the economic growth of the country. This can lead to unemployment and lower spending by individuals and companies, and leading the whole stock market lower, as well as SPY.

- US Dollar Index

The Trade Weighted US Dollar Index measures the relative value of the dollar (USD) compared to the value of other world currencies. It is a weighted average of the foreign exchange value of the U.S. dollar against a subset of the broad index currencies that circulate widely outside the country of issue. Major currencies index includes the Euro Area, Canada, Japan, United Kingdom, Switzerland, Australia, and Sweden.

Depending on the global impact of economic volatility, this quantity can increase or decrease. Since the 2008 crisis deeply affected many other nations, the relative value of the dollar increased. A weak dollar usually coincides with an upward trend in the equity markets and vice versa.(Chris Markouizos, 2018[10])

2.2. Technical Indicators

Table 2: Technical Indicators.

	High	Low	Adj Close	3MA	6MA	12MA	Ret	Momentym
Count	466	466	466	466	463	457	465	455
Mean	976.27	915.43	951.62	945.97	942.55	936.10	0.0078	4.21
Std	698.80	662.85	684.83	680.23	671.39	654.27	0.0425	3.46
Min	106.79	94.23	102.09	106.54	109.23	116.41	-0.2176	-0.11
25%	336.54	314.41	329.26	328.21	330.69	334.61	-0.0168	1.28
50%	993.56	900.04	960.44	949.23	962.74	968.17	0.0107	3.99
75%	1381.64	1287.57	1335.68	1342.52	1330.12	1328.80	0.0350	5.75
max	2940.91	2864.12	2913.98	2877.26	2803.68	2752.29	0.1318	15.23

Count is the number of each indicator. Standard deviation is annualized, plus all the other indicators during the entire length of time.

- High, Low and Adj Close

High is the highest price the stock reached that month and then that's Low. The adjusted close is usually the after hours price and the true open price adjusted from the close price posted.

- Moving Average (rolling average or running average)

Moving Average is a calculation to analyze data points by creating series of averages of different subsets of the full data set. It is a type of finite impulse response filter.

- Exponential Moving Average MACD: (3-month EMA - 6-month EMA)

EMA is a first-order infinite impulse response filter that applies weighting factors which decrease exponentially.

- Volume moving average

A Volume Moving Average is the simplest volume-based technical indicator. It is used to smooth and describe a volume trend by filtering short term spikes and gaps.

- Momentum

Momentum is the rate of acceleration of a security's price or volume – that is, the speed at which the price is changing. Simply put, it refers to the rate of change on price movements for a particular asset and is usually defined as a rate. In technical analysis, momentum is considered an oscillator and is used to help identify trend lines. a multiplicative combination of sentiment and momentum can help predict the S&P 500 stock return over the next month.(Kevin J. Lansing and Michael Tubbs, 2018[11])

3. Return forecasting implementation detail

3.1. Method 1: PCA + Lasso regression

The first method we tried is Principle Component Analysis and Lasso regression. PCA is an unsupervised learning algorithm that reduces dimensionality of a data set. It finds the hyperplanes that separate the original data set with the largest variance and project the data onto the hyperplanes. The reason we use PCA is that we introduced 47 variables to predict future return and we are trying to mitigate the curse of dimensionality. In addition, many technical indicators have high collinearity, for example moving averages with different time window. Therefore, we apply PCA before running regression to help us get a better result. Due to the nature of PCA algorithm, we also normalized our data before applying PCA. We then apply Lasso regression on a rolling basis with 120-day window. The forecasting return is produced by the rolling regressions meaning every regression only produces one prediction.

3.1.1. PCA

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components.

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Consider a data matrix, X, with column-wise zero empirical mean (the sample mean of each column has been shifted to zero), where each of the n rows represents a different repetition of the experiment, and each of the p columns gives a particular kind of feature (say, the results from a particular sensor).

Mathematically, the transformation is defined by a set of p-dimensional vectors of weights or coefficients $W_{(k)} = (W_1, \dots, W_p)_{(k)}$ that map each row vector $X_{(i)}$ of X to a new vector of principal component scores $T_{(i)} = (T_1, \dots, T_l)_{(i)}$ given by

$$T_{k(i)} = X_{(i)} * W_{(k)} \quad for\ i = 1, \dots, n \quad k = 1, \dots l \tag{1}$$

in such a way that the individual variables $T_1, \dots, T_l$ of T considered over the data set successively inherit the maximum possible variance from X, with each coefficient vector W constrained to be a unit vector.

Since there are 47 variables all involved prices and many of them are highly correlated, to reduce the dimensionality of the data set, we only take the largest principals components of these variables, and use them as return predictors in our models. This step helps to make the model more stable.

3.1.2. Least Absolute Shrinkage and Selection Operator (LASSO) Regression

LASSO regression is able to perform feature selection by setting the coefficients of many irrelevant variables to 0 and disregarding them. LASSO regression has a cost function similar to both OLS and Ridge regression, which is shown by Equation

$$J = \sum_{i=1}^{N} ||y_i - y_i||^2 + \lambda ||\beta|| \tag{2}$$

The regularization term here is the l_1 term which represented as the sum of the magnitudes of the weight vector's components rather than the magnitude of the weight vector itself.

The resulting cost function serves a purpose of constraining the absolute sum of the variable coefficients. However, because the regularization term is defined by the weight vector's components, it is often best for a component to be entirely eliminated for the cost function to be minimized.

LASSO regression is characterized by variable selection and regularization while fitting the generalized linear model. Therefore, whether the dependent / response variable is continuous or binary or discrete, LASSO regression can be used to model and predict. Variable selection here means that all variables are not put into the model for fitting, but are selectively put into the model to obtain better performance parameters. Complexity adjustment is to control the complexity of the model through a series of parameters, so as to avoid over-fitting.

For linear models, the complexity is directly related to the number of variables in the model. The more variables, the higher the complexity of the model. More variables can often give a seemingly better model when fitting, but they also face the risk of over-fitting. When validation is validated with new data, it usually works poorly. So when using LASSO regression we could avoid it.

However, with LASSO regression, instead of over-fitting the data by utilizing too many variables, there is the possibility for under-fitting to occur, in which too many variables are eliminated for an accurate model to be generated.

3.2. Method 2: Gradient Boost

For our second predicting method, we choose gradient boost which is a decision tree based model so that it will be different from the linear regression method. Algorithm flow chart is as follows：

$$
\begin{aligned}
& Algorithm\ 1: Gradient\ Boost \\
& F_0(x) = \arg Min_\rho \sum_{i=1}^{N} L(y_i, \rho) \\
& For\ \ m = 1\ to\ M\ \ do: \\
& \qquad \tilde{y}_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, i = 1, N \\
& \qquad a_m = \arg Min_{a,\beta} \sum_{i=1}^{N} [\tilde{y}_i - \beta h(x_i; a)]^2 \\
& \qquad \rho_m = \arg Min_\rho \sum_{i=1}^{N} L(y_i - F_{m-1}(x_i) + \rho h(x_i; a_m)) \\
& \qquad F_m(x) = F_{m-1}(x) + \rho_m h(x; a_m) \\
& endFor \\
& end\ \ Algorithm
\end{aligned} \tag{3}
$$

We used the python package, XGBoost. The reason we choose this package is that it is fast and has high performance. Similar to the Lasso regression, we apply gradient boost on a rolling basis with 60-month window。

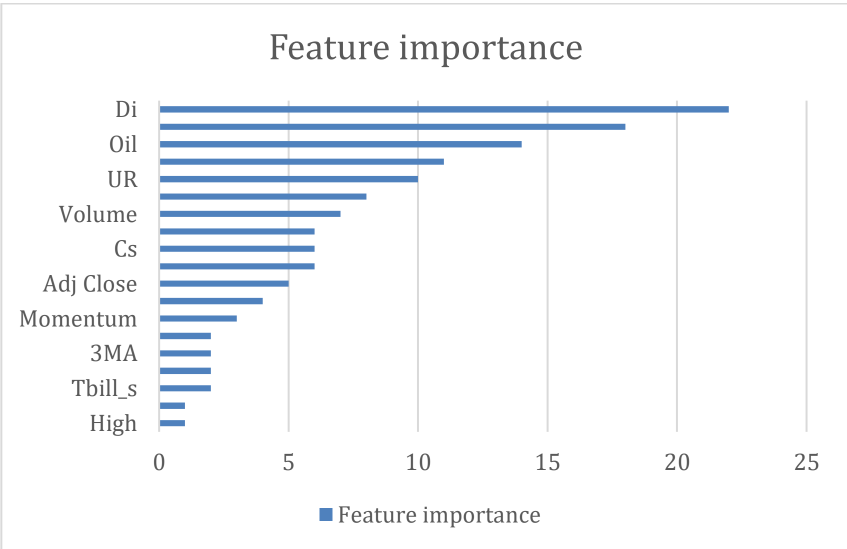

Figure 1: Feature importance.

Through XGBoost Algorithm, each indicator is given a feature importance. After getting the feature importance, we sort them in descending order. The biggest one is 22 and the smallest one is 1.

From the feature importance chart, we could see Di is the most important feature to forecast the return of S&P500. And other features like CPI, Oil, VIX and UR are all very important to forecast.

Finally, in order to predict the returns accurately, we select features with more than 5 importance to predict, which are Di, CPI, Oil, VIX, UR, GOLD, Volume, Pi, Cs, M1 and Adj Close.

The most important features are Di, CPI, Oil and VIX for the following reasons.

Di: When we see the chart of US Dollar index (DXY) and S&P 500, we could find that US Dollar index (DXY) and the S&P500 show a positive correlation sometimes, and then at other times it flips to an inverse correlation. And this kind of relationship has going on for years. But the thing is that we the we are likely to end up disappointed as the correlation flips. So finding the periodic rule plays an important role in our prediction.

CPI: We augment our information set by including a transformation of CPI as a measure of inflation. Of course, the real economy and inflation are not independent, so it is possible CPI may contain information about the macroeconomy as well, and the macro variables contain information about inflation. Variables that contain direct information about financial markets will enlarge the forecasting information set. Plus, Campbell and Vuolteenaho (2004) [12] argue that stock mispricing can be explained by inflation. They find the level of inflation explains 80% of the time-series variation in stock-market mispricing. We use the change in CPI over the last 60 months as the measure of inflation.

VIX: The S&P 500 Volatility Index (VIX) reflects volatility expectations on the benchmark US equity index, and high values are most often associated with fear of significant declines. From the regression of past 60 month, we could always find a high negative correlation between the S&P 500 and the VIX. And we calculate the correlation is equal to -0.79 from 1990 to 2016. However, when the market is nervous or in a panic mode, the VIX/S&P500 relationship can break down, and the indices start to move out of whack.(Rvarb, December 31, 2016[13])

Oil: Oil is an important driver of economic activities. Energy consuming constitute a large portion in US consumer spending. Excessive oil price will put pressure on consumer spending, in addition it will strain on economic growth. On the other hand, a sudden fall in oil price could indicate a slowdown in economic activities. What's more, excessive low oil price will affect the industries that support it, employment in the sector, and the financial sector that lends to it.

Historically, crude oil price and S&P 500 index could influence each other. From the joint chart of Oil price and S&P 500 index since June 2014, we could find that the correlation of them reach to the highest when oil price dropped below $30, and lowest when oil price was at around $100. (Robert Scott, April 2018[14])

3.3. Method 3: Correlation Screen

3.3.1. Correlation Screen

We also try to replicate the method used by Hull and Qiao(2018). However, since our indicators are very different, the regression result after correlation screening has negative correlation with the real return. We believe that Hull and Qiao(2018) did a much better data preprocessing, for example some of their indicators are calculated as log of raw value minus exponential moving average. With such calibrated data processing and broader data source, Hull and Qiao(2018) were able to deliver satisfying forecasting result. But in our case, we decide to abandon the forecasting result due to lack of accuracy.

Correlation screening is another way to remove some of the noise in the forecasting variables. It is used in selecting individual forecasting variables: Using a look-back period of 10 years, we only keep those variables that have at least a 10% correlation with the upcoming 130-day returns. The correlation screening model is according to hull and Qiao(2018).

$$R^{e}_{m,t\to t+130} = \alpha_{CS} + \beta^{'}_{CS}\tilde{X}_t + \varepsilon_{CS,t\to t+130} \tag{4}$$

Where

$$\tilde{X}_t = \begin{bmatrix} x_{1,t} I_{|\rho_1,m|>0.1} \\ x_{2,t} I_{|\rho_2,m|>0.1} \\ \cdots \\ x_{16,t} I_{|\rho_{16},m|>0.1} \end{bmatrix} \tag{5}$$

$$\rho_{i,m} = Corr(x_{i,t}, R^e_{m,t\to t+130}) \tag{6}$$

3.3.2. Implement

In this part, I set different correlation from 0-0.1 and find 0.08 get the biggest portfolio return till now. After that I set it from 0.085-0.095, then 0.088 get the biggest cumulative return and Sharpe ratio. So I set it to 0.088.

Table 3: Cumulative return and portfolio Sharpe ratio at different correlation for Correlation Screen.

Correlation	Cumulative return	Sharpe ratio	Correlation	Cumulative return	Sharpe ratio
0.00	1.07	0.16	0.085	3.08	0.30
0.01	2.51	0.26	0.086	3.52	0.32
0.02	2.40	0.26	0.087	3.54	0.32
0.03	3.52	0.32	0.088	4.16	0.34
0.04	3.16	0.30	0.089	3.97	0.34
0.05	3.00	0.29	0.090	3.99	0.34
0.06	3.30	0.31	0.091	2.70	0.28
0.07	2.18	0.24	0.092	2.93	0.29
0.08	2.72	0.28	0.093	3.01	0.30
0.09	3.99	0.34	0.094	3.06	0.30
0.10	1.88	0.22	0.095	3.10	0.30

On the left side, we choose the correlation from 0.00 to 0.10, because when we set the value bigger than 0.1, there will be not enough indicators selected. And we just need to find the interval which the peak exists, which is 0.085 to 0.095 on the right side.

3.3.3. Correlation screen selecting chart

In this part, I drawn all the variables that selected from each time to the regression to forecast the return.

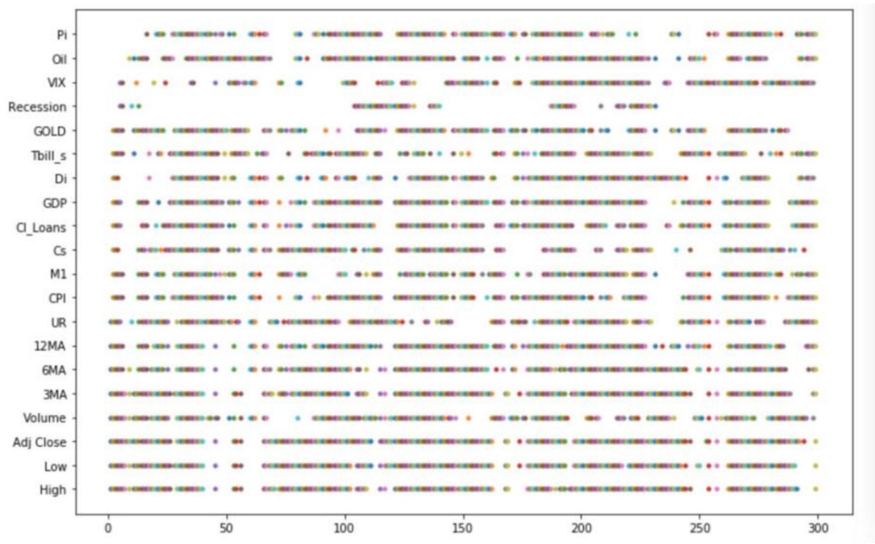

Figure 2: Correlation Screen indicator selecting chart from September 1995-November 2018.

In this table, we could see which indicators do we choose every time when we do the regression.

4. Strategy Comparison

4.1. Forecast return comparison

In the following chart, I draw all the line from S&P 500 and PCA forecast and Correlation screen forecast monthly return from 2011 to 2018. Here, we can recall the weight equation which is $w(t) = \min\{1.5, \max\{-.5, 100\times m(t)\}\}$ where $m(t)$ is the forecast return at time t. And our portfolio real return is $w(t)*r(t)$, where $r(t)$ is the S&P500 return at time t.

We could find as long as out expected forecast return $m(t)$ have the same sign with $r(t)$ at time t, then we could have positive return no matter $r(t)$ is positive or negative. But if they have opposite sign, our portfolio will lose money. Therefore, a good fitting between forecast return and real return is needed.

Figure 3: Comparison between 3 forecast return and S&P 500 return from August 2010-November 2018.

From this chart, the blue line which is PCA forecast return tend to have larger distance from 0. Then if the forecast return of PCA and Correlation screen both have same sign with real return, the portfolio return using PCA forecast more likely to have larger real return, which we could infer from the equation.

At each time point, we could see the predicted return by each strategy and compare them with the return S&P 500.

4.2. Strategy comparison

In this section, I analyze the advantage of PCA & Lasso regression and XGBoost.

Table 4: Advantage of PCA&Lasso Regression and XGBoost Regression.

Method	Purpose
PCA & Lasso Regression	PCA: • Dimensionality reduction • Address Multicollinearity Lasso : • Prevent overfitting • Variable Selection
XGBoost	• Machine learning algorithm with high speed and good performance • Perform automatic feature selection

In this table, the purpose and advantages of each strategy are giving.

5. Strategy implementation detail

5.1. Transaction cost and financing cost

When changing the weight, we need to pay for the transaction cost, assuming two thousandths of the transaction amount.

After getting the forecast return, we start to use our strategy. To buy or short the S&P500 index between [-0.5, 1.5]. We could set the amount of money to put on this portfolio, and this range allow us to adjust at most short 0.5 of the set amount of money on S&P500, and at most long 1.5 of the set amount of money on S&P500. If we don’t have enough money to invest 1.5 times of set money on S&P 500, we need to borrow money. In this case, we must pay for the interest of financing and we will talk about this situation later. The cumulative return of 3 strategies without interest of financing and just hold S&P 500 are showing below:

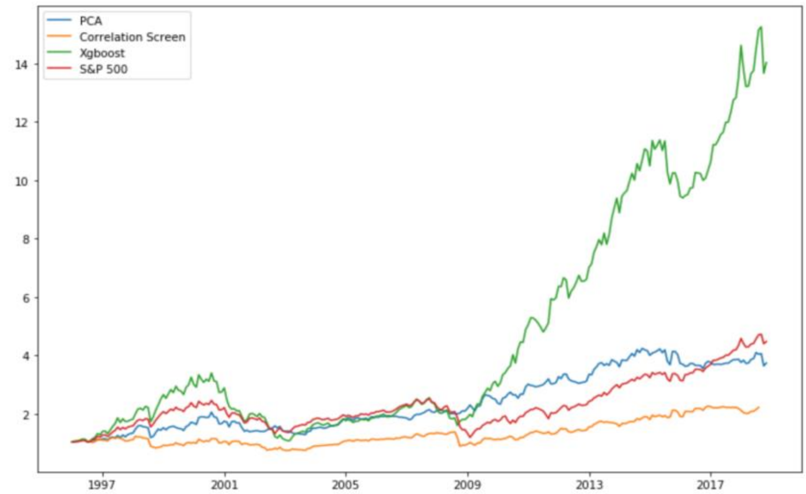

Figure 4: Cumulative return of 3 strategies without paying interest and S&P 500 from January 1996-Octorber 2018.

From the chart we could see:

1. When we look closer at the chart, we find that all of the three strategies experienced a big draw down during the 2008 financial crisis. To further analysis of why this happened, I think this is due to the limitation of our strategy.

2. We can see the obvious advantages of XGBoost portfolio from 1997 to 2001 and from 2009 to the present. But during 2002 to 2008, it has roughly the same trend as S&P 500 and also better than the portfolio of PCA and Correlation Screen.

For the situation which we need to borrow money when the weight is larger than 1, we choose to pay for the interest every month and we set the annual interest rates at 4.2%. So monthly interest rates equal to 0.327%, if the adjusted weight is larger than 1. And we assume that the financing will be repaid at the end of each month. Considering transaction cost and borrowing cost, result of strategies from 1995 to 2018 shows as below:

Figure 5: Cumulative return of 3 strategies with paying interest and S&P 500 from January 1996-Octorber 2018.

In this situation, each strategy shows worse cumulative return than buy and hold S&P 500, not to speak of Sharpe ratio. This is because each strategy always predicts returns greater than 0.1, and make us to enlarge the weight and beyond the presupposed investment which we set to put on S&P 500. Then we need to pay for the interest at the end of month compare just buy-and-hold S&P 500.

5.2. Comprehensive Portfolio Comparison

Now that we have a list of forecasted returns, we have to calculate the Annual return, Volatility, Sharpe ratio and Max drawdown of each strategy and buy & hold S&P 500. We compare the performance of different strategies in Table 10. The best strategy is XGB portfolio shows annual returns that are about 550 basic points higher compare to buy-and-hold in this period, but the Correlation Screen portfolio and PCA portfolio have lower returns compared to both the XGB portfolio and buy-and-hold returns. Annual market standard deviation from January 1996-Octorber 2018 was 4.24%, whereas XGB portfolio was 5.54%, Correlation Screen portfolio was 4.46%, and PCA portfolio has the lowest standard deviation which is 4.16%. Combining the volatility and return, we use Sharpe ratio to compare strategies synthetically. The XGB portfolio has the largest Shape ratio, which is 0.71. And the Sharpe ratio of PCA portfolio and Correlation Screen portfolio both are lower than buy-and-hold S&P 500.

For buy-and-hold, the maximum drawdown was 55.2% during the Global Financial Crisis in late

2008. In comparison, all the three strategies offer greatly reduced drawdowns: The Correlation Screen portfolio has a maximum drawdown of 34.9%, the PCA portfolio has a maximum drawdown of 29.3% and XGB portfolio has a maximum drawdown of 30.1%. All the strategies avoid the Global Financial Crisis, which can attribute the forecast return that make us lower the weight put on S&P 500 or short the index.

Table 5: Strategy Summary Statistics from January 1996-Octorber 2018.

	Correlation Screen Portfolio	PCA Portfolio	XGB Portfolio	Buy & Hold S&P 500
Annual Return	3.38%	5.75%	12.2%	6.76%
Std	4.46%	4.16%	5.54%	4.24%
Sharp ratio	0.32	0.32	0.71	0.52
Max Drawdown	34.9%	29.3%	30.1%	55.2%

Annualized Return is the average annual gross return of three strategies over the period from January 1996-Octorber 2018. Standard Deviation and Sharpe ratio are also annualized. Maximum drawdown is calculated from peak to trough of the cumulative return series.

6. Conclusion

In this paper, we do the similar job as Hull and Qiao(2018) return predictability. Searching nearly 20 prominent return predictors, combining them using three different ways and then doing the regressions. And we assume people don't to borrow money from bank to invest in portfolio when the weight is larger than 1, because we invest only part of our total asset. Finally, we get following results:

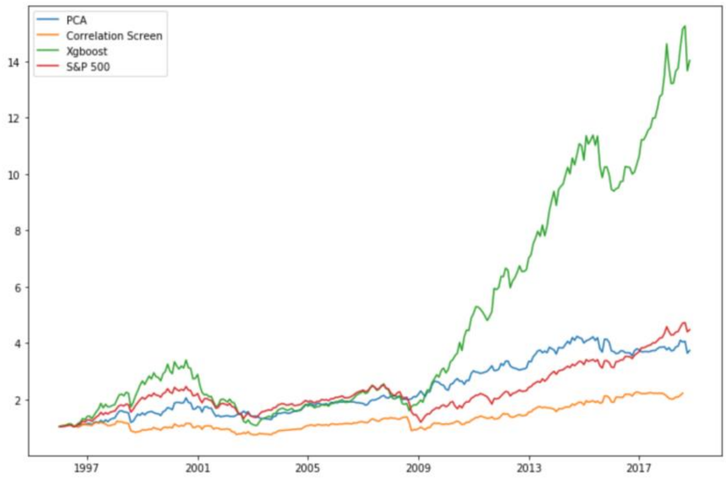

Figure 6: Cumulative return of 3 strategies with paying interest and S&P 500 from January 1996-Octorber 2018.

Because until 09/01/1990 can we firstly get all the indicator and we need 60 months' data to do the regression, so the first forecast return begin at 10/01/1995.

Finally, in the end date on 11/01/2018, the cumulative return for S&P500, PCA portfolio, Correlation Screen portfolio and XGBoost portfolio are 4.72, 4.05, 2.40 and 15.17 respectively. So we could draw the conclusion that XGBoost has made the most accurate prediction and made the strategy get the biggest rate of return.

References

[1] Hull, B. and Qiao, X. (2017) A Practitioner's Defense of Return Predictability, Journal of Portfolio Management, 43, 60-76.

[2] Oxman, Jeffrey(2012) Price inflation and stock returns, Working Paper, University of St. Thomas, Minnesota.

[3] Steve, LeCompte(2018) Unemployment Claims Reports and Near-term Stock Market Returns, Economic Indicators, Equity Premium, 2018(08).

[4] Tarver Evan(2018) The Impact of Disposable Income on the Stock Market, Investopedia, 2018(11).

[5] Ogutu, Crispin O(2011) The relationship between the GDP and share price movement at the Nairobi Stocks Exchange, University of Nairobi, Kenya(2011).

[6] Trust Kganyago1, Victor Gumbo(2015) An Empirical Study of the Relationship between Money Market Interest Rates and Stock Market Performance: Evidence from Zimbabwe (2009-2013), International Journal of Economics and Financial Issues, 2015, 5(3), 638-646.

[7] Fassas, Athanasios and Hourvouliades, Nikolaos L.(2018) VIX Futures As a Market Timing Indicator (June 3, 2018)

[8] Kenneth L. Fisher and Meir Statman(2003), Consumer Confidence and Stock Returns, The Journal of Portfolio Management Fall 2003, 30 (1) 115-127.

[9] Naliniprava Tripathy and Amit Tripathy(2016) A Study on Dynamic Relationship between Gold Price and Stock Market Price in India, European Journal of Economics, Finance and Administrative Sciences, Issue 88, 23-35.

[10] Markouizos, Chris(2018) U.S. Dollar Index - A Contradictory View - Major Inflection Point, Seeking Alpha, 2018(01).

[11] Lansing J. Kevin and Tubbs Michael, Using Sentiment And Momentum To Predict Stock Returns, Economic Letters,2018(12).

[12] Campbell, John Y. and Tuomo Vuolteenaho. "Inflation Illusion And Stock Prices," American Economic Review, 2004, v94,19-23.

[13] Rvarb(2016) Relationship Between the VIX and SP500 Revisited, Journal of Relative Value Arbitrage.

[14] Robert Scott(2018) How Are Crude Oil Prices and the S&P 500 Correlated?, YAHOO FINANCE, 2018(03).