

TDK dolgozat

2023

NEUMANN JÁNOS EGYETEM
MNB INTÉZET: FENNTARTHATÓ PÉNZÜGYEK KÖZPONT
NEMZETKÖZI GAZDASÁG ÉS GAZDÁLKODÁS MESTERSZAK



MNB INTÉZET
FENNTARTHATÓ PÉNZÜGYEK KÖZPONT

**A RÉSZVÉNYHOZAMOKBAN TÖRTÉNŐ
REZSIMVÁLTÁSOK KEZELÉSE DINAMIKUS
MODELLVÁLASZTÁSSAL**

**CAPTURING REGIME CHANGES IN STOCK RETURNS
WITH DYNAMICAL MODEL SELECTION**

Csillag Balázs József

Konzulens: Neszveda Gábor

Kézirat lezárásának dátuma: 2023. április 7.

Abstract

The S&P 500 is widely recognized as one of the most prominent stock market indices globally, and forecasting its returns is a question widely researched in the literature. Previous research found that incorporating the regime-switching behaviour of the stock returns into predictive models might increase accuracy. This study proposed a dynamical model selection framework to predict daily stock returns of the S&P 500 index ETF. In an attempt to assess regime switches, the study generates an extended number of features using historical return and trading volume data and creates numerous simple OLS models using different random subsets of these features. Before every one-step-ahead prediction, an ensemble prediction is created based on the signal of those models which could predict the recent past most accurately. A trading strategy based on the predicted signals could achieve a yield significantly different from zero in 2020 and 2021, mainly driven by the general overprediction of up-days. The proposed prediction framework could achieve an average accuracy of 51.72% on the period between 2010 and 2022, achieving the highest (lowest) accuracy in 2020 (2015) with 58.5% (46.43%). Despite the proposed ensemble predictions could outperform the baseline models occasionally, based on the overall results we can conclude that our prediction system is inferior to both of the selected benchmarks. Although the trading simulation did not yield successful results for the SPY index, stock level experiments show that it has the potential to generate significantly positive returns for individual stocks.

Contents

1	Introduction	4
1.1	Previous results	7
1.2	Regime shifts	12
2	Data	13
2.1	Data source	13
2.2	Data preparation	15
2.3	Summary statistics	15
3	Methodology	21
4	Results	26
4.1	Framework outputs	26
4.2	Market simulation	31
4.3	Comparison to baselines	33
5	Robustness checks	35
6	Conclusion	39

List of Figures

1	Closing price of SPY for period between 5th of January 2004 and 30th of December 2022, data in \$	14
2	Trading volume of SPY for period between 5th of January 2004 and 30th of December 2022	14
3	Daily returns of SPY for period between 6th of January 2004 and 30th of December 2022	15
4	Distribution of daily returns between 2004 and 2022	16
5	Distribution of daily returns between 2004 and 2022 by year	16
6	Q-Q plot of the daily returns between 2004 and 2022	17
7	Yearly trading days with negative and positive returns	17
8	Autocorrelation and partial autocorrelation of the daily returns between 2004 and 2022	18
9	Rolling windowed autocorrelation of the daily returns between 2004 and 2022	20
10	Rolling windowed time lagged cross-correlation of the daily returns between and volume 2004 and 2022	21
11	Proposed dynamical model selection framework	25
12	Prediction and fact daily returns of model 654	27
13	Prediction and fact daily return directions of model 654	27
14	Rolling accuracy	28
15	Ensemble predictions based on the top 3 model for year 2021	28
16	Ensemble predictions based on the top 3 model for year 2022	29
17	Different ensemble predictions' accuracy	29

List of Tables

1	Return prediction accuracy achieved by previous studies.	9
2	Prediction accuracy achieved on S&P 500 by previous studies.	11
3	Descriptive statistics of daily returns and volumes	16
4	Confusion matrix based on the top 3 model based ensemble prediction . . .	30

5	Confusion matrix based on the top 10 model based ensemble prediction . .	30
6	Confusion matrix based on the top 50 model based ensemble prediction . .	31
7	Average yields achieved based on the different ensemble predictions	31
8	Average yields achieved based on the ensembe predictions of the top three models	32
9	Fact and predicted ratio of days with positive return	33
10	Comparison to baselines using period between 22/02/2013 and 22/11/2013 .	34
11	Accuracy achieved by the top three model based ensemble prediction com- pared to two baseline models	35
12	Accuracy achieved by ensemble predictions for each ticker	36
13	Yield achieved by ensemble predictions for each ticker	38

1 Introduction

The S&P 500 index is a stock market index that measures the performance of 500 largest companies listed on stock exchanges in the United States.¹ The index is widely recognized as one of the most prominent stock market indices globally, and forecasting its returns has piqued the interest of both researchers and practitioners. (Kumbure et al. 2022)

Market participants often seek to achieve the goal of consistently earning higher risk-adjusted returns than the market. This has led to the use of machine learning techniques to develop accurate models for stock market prediction. However, consistently beating the market may violate the efficient market hypothesis. (Fama 1970) The theory suggests that the market price follows a random walk, making it impossible to predict future changes using existing information. There are three forms of market efficiency - weak-form, semi-strong form, and the strong-form. Weak-form efficiency assumes that information absorbed by past prices are already reflected in the current stock price. Semi-strong form efficiency states that stock prices fully reflect all publicly available information, including fundamental information. Lastly, strong-form efficiency implies that all information, including insider information, is reflected in stock prices, making it impossible for any investor, even those with insider information, to consistently achieve higher expected returns than the market.

The efficient market hypothesis has faced growing challenges over time. Asset pricing anomalies are widely documented in the literature. These include, but are not limited to the following (the grouping is based on (Jacobs 2015):

- Law of one price: the prices of similar (dual-listed) instruments that provide identical payouts can differ significantly from each other during certain periods (Baker et al. 2012; Rosenthal and Young 1990)
- Momentum: the average past return significantly predicts the next month's return positively. (Jegadeesh and Titman 1993) Several studies have found that the effect

¹It is often considered a benchmark for the U.S. stock market. The index is weighted by float-adjusted market capitalization (FMC) meaning that larger companies have a greater influence on the index's performance. Float-adjusted market capitalization is a measure of a company's worth based on the total value of its shares which are available for public trading. See more in S&P Dow Jones Indices: Index Methodology.

Link: <https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf>

of momentum is amplified when considering the interaction of returns with certain stock-level variables (e.g., company age, book-to-market ratio, market capitalization). (Csillag and Neszveda 2020; Da et al. 2014; Mérő et al. 2019; Zhang 2006)

- Short-term return reversal: previous research has found negative autocorrelation between returns (Jegadeesh 1990; Lehmann 1990)
- Long-term return reversal: there is a negative relationship between a stock's past 3-5 year returns and its future returns. (De Bondt and R. Thaler 1985; McLean 2010)
- Calendar-based anomalies: price movements related to recurring events or certain months can be predicted. (Hartzmark and Solomon 2013; Heston and Sadka 2008)
- Lead-lag effects: the price of economically related (e.g., a buyer and supplier in a supply chain) stocks can be predicted. (Cohen and Lou 2012)
- Pairs trading: stocks that have performed similarly in the past are likely to perform similarly in the future. (Jacobs and Weber 2015)
- Distress risk anomalies: contrary to the conclusions of classical financial models, research finds that higher risk does not lead to higher returns. (Campbell et al. 2008; Shumway 2001)
- Skewness anomalies: lottery-type stocks typically underperform. (Bali et al. 2011)
- Differences of opinion: stocks that are highly contested by the market generally underperform. (Ang, Hodrick, et al. 2006; Diether et al. 2002)
- Anomalies related to industry effects: stocks related to the alcohol, tobacco, and gambling industries outperform. (Hong and Kacperczyk 2009)
- Net stock and financing anomalies: managers successfully exploit investor sentiment in the timing of their financial decisions. (Greenwood and Hanson 2012)
- Capital investment and growth anomalies: previous research found a negative relationship between corporate growth or capital investment and future returns. (Hirshleifer, Hou, et al. 2004; Lou 2014)

- Anomalies related to innovation: investors underreact to or misinterpret a company's innovation activities. (Fairfield et al. 2003; Hirshleifer, Hsu, et al. 2013)
- Dividend anomalies: companies that initiate their first dividend payment generally outperform. (Boehme and Sorescu 2002; Michaely et al. 1995)
- Earnings surprises: after the release of earnings reports, the price follows a rising (falling) trend for months if the company announces better (worse) than expected results. (Ball and Brown 1968; Bernard and Thomas 1989)

A substantial number of published studies have attempted to forecast stock markets precisely by creating sophisticated forecasting systems. (Sedighi et al. 2019; Song et al. 2019) Despite that some studies have obtained results that challenge the weak form of the efficient market hypothesis by demonstrating accurate predictions of short term market trends (Atsalakis and Valavanis 2009; Weng et al. 2017), stock market prediction is still widely regarded as one of the most challenging tasks in financial research. (Y. Chen and Hao 2017) Through an examination of the available literature, it has been revealed that the most extensively researched financial instruments are the indices and stocks in the United States, with a primary focus on the S&P 500 index. (Kumbure et al. 2022)

Recent advancements in stock market prediction have been extensively using deep learning methods. This is because it is capable of detecting latent nonlinear relationships and extracting features from noisy data without relying on human expertise or economic presumptions. (Chong et al. 2017) Consequently, deep neural networks, convolutional neural networks, and long short-term memory networks have been extensively employed in the forecasting of stock market price and return.

This study aims to predict the daily stock returns of S&P 500 index ETF (SPY) by employing a dynamical model selection framework that captures the time-varying behavior of stock markets. (Hamilton 1989) Our results show, that the proposed ensemble predictions average accuracy was 51.72% for the period between 2010 and 2022, with the lowest accuracy in 2015 (46.43%) and the highest in 2020 (58.5%). A trading simulation based on the predictions could achieve an average yield of 0.0114%. However, the yield is not significantly greater than zero. Comparing our accuracy results to benchmarks we can conclude that our prediction system is inferior to the selected baselines for the S&P 500

index ETF. However, stock level experiments show that, the proposed prediction system has the potential to generate significantly positive returns for individual stocks.

Section 1 briefly presents the results achieved by the literature so far. Section 1.2 summarizes the findings of previous researches focusing on the regime shift behaviour of stock returns. Section 2 describes the data used in the analysis, data preparation and provides summary statistics. Section 3 explains the proposed ensemble prediction framework, whilst Section 4 displays the results achieved with this approach, and compares it to several benchmarks. In order to have a more nuanced picture on the performance of the proposed prediction framework, we also conducted experiments with individual stocks: Section 5 shows the results of these experiments.

1.1 Previous results

Despite the belief that a model grounded in investor rationality was adequate, concerns lingered about its ability to account for observed anomalies. (Daniel et al. 1998) This realization prompted a shift towards the development of models that incorporate human psychology, ultimately leading to the emergence of behavioral finance. (De Bondt and R. H. Thaler 1990) Behavioral finance challenges the assumption of perfect rationality among investors, recognizing that behavioral biases such as loss aversion, overreaction, and underreaction can significantly impact investment decisions. (K.-P. Lim and Brooks 2011)

Research that incorporates investor behavior has reported promising results. Li et al. (J. Li et al. 2017) introduced a novel approach for forecasting stock market performance, utilizing a Long Short-Term Memory (LSTM) neural network ² that incorporates both investor sentiment and market factors. The inclusion of investor sentiment, extracted from forum posts, allows for the analysis of the irrational component of stock prices. Empirical

²Long Short-Term Memory (LSTM) neural networks are a type of recurrent neural network that uses specialized memory cells to address the vanishing gradient problem in traditional recurrent neural networks. The problem refers to the case, when during backpropagation the gradient will be vanishingly small, effectively preventing the weight from changing its value. (Basodi et al. 2020) LSTM memory cells have the ability to selectively store, retrieve and erase information over time, making them suitable for learning long-term dependencies in sequential data. Due to their effectiveness in processing time series data, LSTMs have become increasingly popular in applications such as speech recognition, natural language processing and financial forecasting. (Y. Yu et al. 2019)

research on the CSI300 index demonstrates that their approach outperforms other benchmark models with a prediction accuracy of 87.86%.

Lien Minh et al. (Minh et al. 2018) tried to predict stock market trends utilizing both historical stock prices, financial news and sentiment dictionary. Their approach could predict the S&P 500 index stock price directions with an accuracy of 66.32%. A similar approach was used by Shi et al. (Shi et al. 2018) They used stock-related tweets together with historical price data and financial news articles from Reuters and Bloomberg regarding the US stock market. The paper reported an average out-of-sample prediction accuracy of 55.9% over the stock price movement of nine companies in S&P 500.

The extensive use of the internet has led to the development of new forms of collective intelligence, such as Google and Wikipedia, which represent a new era of "crowd-sourced" knowledge bases. These platforms collate information on publicly traded companies and capture web traffic statistics to reflect the public's interest. Combining these online data sources with traditional indicators can create a more effective daily trading expert system. Weng et al. (Weng et al. 2017) achieved an 85% accuracy in predicting the next-day stock movement of Apple using these information. According to their findings, financial expert systems can enhance their knowledge base by incorporating data obtained from nontraditional sources such as Google and Wikipedia. Also, research demonstrated that Google Trends data can aid in predicting the direction of the stock market index. (H. Hu et al. 2018)

Other researches also report that taking investors expectations into account might increase prediction accuracy. Rosillo et al. (Rosillo et al. 2014) analyse the use of the Chicago Board Options Exchange Market Volatility Index (VIX) in forecasting the weekly change in the S&P 500 index. They obtained a significantly better results with the VIX index relative to a model without VIX.

Another part of the literature focuses on incorporating news data into the prediction process. Ding et al. (Ding et al. 2015) proposes a deep learning approach for event-driven stock market prediction, which involves extracting events from news titles. The proposed method was evaluated using the S&P 500 index and individual stock prediction; their reported maximum accuracy was 65%.

Previous research even tried to improve forecast accuracy by applying methodology used for visual imagery to stock price prediction. Sim et al (Sim et al. 2019) proposed a new

model based on convolutional neural network (CNN).³ They converted 9 technical indicators time series to images to verify the applicability of the CNN method in the stock market. Other papers applying image processing algorithms reported promising results. Using 82 different technical indicators Hoseinzade et al. (Hoseinzade and Haratizadeh 2019) could improve the predictive performance of other CNN based approaches by about 3% to 11%.

To exploit the latent interconnections between different markets and market indices several studies predicted the S&P 500 market index using data of other indices. In their studies Zhong and Enke (Zhong and Enke 2017) proposes a complete and efficient data mining process to forecast the daily direction of the S&P 500 Index ETF return using 60 financial and economic features. They reported a testing accuracy of 56.8% to 59.9%.

Table 1 summarizes the return prediction accuracy reported in the literature. The table is constructed based on the literature review paper of Hu et al. (Zexin Hu et al. 2021) There are several papers which used the S&P 500 index as the target variable applying different novel neural-network based models. These could achieve an accuracy in the range of 55.44% to 66.32% using different test periods. Large scale experiments show that the accuracy of S&P 500 index prediction is around 60%, whilst predicting individual stocks the accuracy can increase up to around 70%. (Ding et al. 2014)

In order to have a broader overview of the results of literature, other influential studies which experimented with other indexes or individual stocks have been listed in the summary table. Their prediction accuracy varies greatly but some papers report quite ambitious accuracies. Note, that although these studies are listed here, their results have not been validated by us. Thus for example it is possible, that the reported accuracy applies to other than daily granularities, or the paper did not reported the accuracy on the test set.

Table 1: Return prediction accuracy achieved by previous studies.

<i>Reference</i>	<i>Dataset</i>	<i>Methodology</i>	<i>Accuracy</i>
Minh et al. 2018	S&P 500, VN-index	Documents preprocessing- Documents labeling -Stock2Vec embedding- BGRU	66.32%
Ding et al. 2015	S&P 500 index	NTN-CNN	66.08%

³A convolutional neural network (CNN) is a class of deep neural networks which has been applied extensively in the field of image processing, speech recognition, natural language processing. (Gu et al. 2018)

Continuation of Table 1			
<i>Reference</i>	<i>Dataset</i>	<i>Methodology</i>	<i>Accuracy</i>
Santos Pinheiro and Dras 2017	S&P 500 index	NLP+LSTM	63.34%
Yang et al. 2018	S&P 500 Index ETF (SPY)	CNN	60.02%
Ding et al. 2014	S&P 500 index	NTN-CNN	58.94%
P. Yu and Yan 2020	S&P 500, DJIA, Nikkei 225, HSI, CSI 300 and the ChiNext index	DNN with phasespace reconstruction (PSR) and LSTM	58.07%
Wen et al. 2019	S&P 500	CNN	56.14%
Y. Liu et al. 2018	S&P 500 index (S&P 500)	TransE-CNN-LSTM	55.44%
Rana et al. 2019	Spanish stock company Acciona	LSTM	98.49%
J. Li et al. 2017	Chinese stock market	LSTM with Naïve Bayes	87.86%
Naik and Mohan 2019	NSE, ICICI Bank, SBI Bank, Kotak Bank, Yes Bank	DNN	84.5%
Zhao et al. 2017	SSE Composite Index, SSE 50, CSI 500, CSI 300 and SZSE Composite Index	Time-weighted LSTM	83.91%
Shi et al. 2018	AAPL, GOOG, BAC, WMT, GM, XOM, BA, GSPC, T	Hybrid of RNN, LSTM and CNN	max. 79.7%
Gudelek et al. 2017	SPDR ETF	CNN	78.46%
Long et al. 2020	Chinese Stock market	LSTM with CNN	75.89%
M.-Y. Chen et al. 2019	Chinese Stock market	LSTM	75.58%
S. Chen and He 2018	Chinese stock price	CNN	75.2%
Gudelek et al. 2017	Exchange Traded Funds(ETF)	2D-CNN	71.72%
C. Li et al. 2019	CSI300, CSI200 and CSI500	Multi-task RNN with MRFs	68.95%
W. Chen et al. 2018	HS300	RNN-Boost	66.54%
Zeng and X. Liu 2018	SSE50 index	LSTM	65%
Song et al. 2019	KOSPI and KOSDAQ	DNN	61.9%
Shah et al. 2018	Tech Mahindra (NSE: TECHM) BSESensex	LSTM	60.60%
J.-F. Chen et al. 2016	Taiwan index futures	GAF+CNN	57.88%
Nelson et al. 2017	IBovespa index	LSTM	55.9%

Continuation of Table 1			
<i>Reference</i>	<i>Dataset</i>	<i>Methodology</i>	<i>Accuracy</i>
Xu and Keselj 2019	Microsoft (MSFT), PO logistics (XPO) and AMD Daily stocks	Attention-based LSTM	54.58%
Pang et al. 2020	Chinese stock	DeepLSTM with encoder	53.2%
Ziniu Hu et al. 2018	Chinese stock price	HAN with SPL	47.8%
K. Chen et al. 2015	China stock market in Shanghai and Shenzhen	LSTM	27.2%

The results of the papers which used the S&P 500 index for their experiments is summarized in Table 2. The test period column of the table shows the period on which their models accuracy was tested and the *Accuracy on the test set* column shows the related value. They all chose a backtest period of couple of months, only Yu et al.’s (P. Yu and Yan 2020) study report the prediction accuracy for a longer period. For this reason we will use their yearly accuracy as one of the baseline results. As it will be clear in Section 4 the chosen period on which the accuracy is reported greatly influences the prediction performance.

Table 2: Prediction accuracy achieved on S&P 500 by previous studies.

<i>Reference</i>	<i>Test period</i>	<i>Accuracy on the test set</i>
Ding et al. 2014	22/02/2013 - 21/11/2013	58.94%
Ding et al. 2015	22/02/2013 - 21/11/2013	66.08%
Santos Pinheiro and Dras 2017	22/02/2013 - 21/11/2013	63.34%
Yang et al. 2018	Unknown	60.02%
Y. Liu et al. 2018	01/01/2017—07/31/2017	55.44%
Minh et al. 2018	16/06/2013 - 31/12/2013	66.32%
Wen et al. 2019	Not clear, presumably: 21/05/2017 - 15/09/2017	56.14%
P. Yu and Yan 2020	2010-2017	average: 58.07%

1.2 Regime shifts

As the above tables shows, stock returns contain a sizable unpredictable component, so that the best forecasting models can explain only a relatively small part of stock returns. The “best” model specification or its corresponding parameter values are unknown. Furthermore, due to parameter instability, the best model can change over time. This uncertainty and instability are relevant to stock return forecasting due to the connection between business-cycle fluctuations and stock return predictability. The difficulty of predicting out-of-sample returns arises from the substantial model uncertainty and parameter instability in the data-generating process for stock returns. A possible approach for improving equity premium forecastability centers on regime shifts. (Rapach and Zhou 2013) In this section we summarize the findings of previous research regarding evidence of distinct regimes in stock market returns.

Hamilton (Hamilton 1989) proposed the use of Markov switching techniques to model non-stationary time series. The Hamilton approach involves interpreting parameters as the result of a discrete-state Markov process. For instance, returns in the stock market may undergo occasional, distinct shifts.

Regime switching econometric models can be defined as a scenario in which stock market returns are drawn from distinct distributions, with a specific stochastic process determining the probability that each return originates from a particular distribution. (Schaller and Norden 1997) Regime-switching models are useful in accounting for sudden changes in financial markets and the persistence of such changes. These models correspond to different periods in regulation, policy, and other secular changes. The empirical estimates of these models show that asset returns differ in means, volatilities, autocorrelations, and cross-covariances across regimes, which allows them to capture the characteristics of financial series. (Ang and Timmermann 2012) In the equity market, these distinct regimes are characterized by periods of high and low volatility as well as extended bull and bear markets. (Pagan and Sossounov 2003) Several novel research found empirical evidence of parameter instability in predictive regression models of U.S. aggregate stock returns. (D. Kim et al. 2014; Paye and Timmermann 2006)

Previous research found that incorporating the regime-switching behaviour of the stock returns into predictive models might increase accuracy. (Rapach and Zhou 2013) Henkel

(Henkel et al. 2011) reports robust proof of predictability of U.S. stock returns using regime-switching regressions, while documenting less compelling evidence for stock returns of other G-7 countries. Not regression based research also found that responding to regime switching between “bull” and “bear” markets substantially improves portfolio performance for an investor allocating across the 25 Fama-French portfolios. (Tu 2010) Deep learning methods have been proposed to asses the problem of sudden changes in temporal patterns. Lim et al. (B. Lim et al. 2021) applied a Temporal Fusion Transformer (TFT) – an attention-based architecture to stock return forecasting. By selecting relevant features and implementing several gate layers to eliminate redundant components, TFT can achieve high performance even for regime shifting time series.

This study aims to predict stock returns by employing a dynamical model selection framework that captures time-varying behavior. The proposed method chooses the model that predicts most accurately in the recent past before every one-step ahead prediction, allowing for dynamically changing models. A large number of simple OLS models are created using different random subsets of features, and ensemble predictions are generated based on the top-performing models. The models are evaluated daily, allowing for regime changes and dynamical model and implicitly feature selection to improve accuracy.

2 Data

In this section we present in detail the database used in the analysis. First we describe the obtained stock market index data and the preprocessing steps then report the results of the exploratory data analysis.

2.1 Data source

In this paper we aim to predict the future daily returns of the SPDR S&P 500 ETF Trust, also known as the SPY. The SPY ticker is one of the most popular exchange-traded funds (ETF) that tracks the S&P 500 index, meaning that it attempts to replicate the performance of the index by holding the same stocks in the same proportion as the index. The fund is traded on the New York Stock Exchange (NYSE).

This paper uses daily data on the SPY ticker, downloaded from the database of Yahoo!Finance. The following variables were obtained through Yahoo!Finance API:

- **Closing Price:** price quoted at the the last transaction before the market officially closes for normal trading, data in \$. See in Figure 1.
- **Volume:** number of shares traded on a given day. See in Figure 2

The initial time series contains 4.782 observations covering the period between 5th of January 2004 and 31st of December 2022. Thus, on average the we have 265 trading days per year which is in line with the trading days reported by NYSE.⁴



Figure 1: Closing price of SPY for period between 5th of January 2004 and 30th of December 2022, data in \$

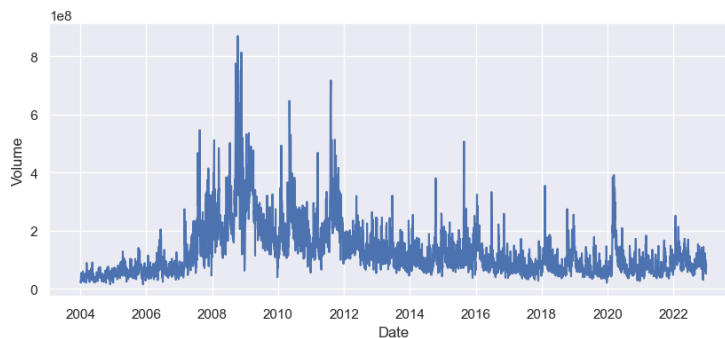


Figure 2: Trading volume of SPY for period between 5th of January 2004 and 30th of December 2022

⁴<https://www.nyse.com/markets/hours-calendars>

2.2 Data preparation

Using these variables the following data preparation steps were executed:

- Calculate daily returns of SPY:

$$r_t = \frac{price_t - price_{t-1}}{price_{t-1}} \quad (1)$$

Where r_t is the return on the asset on day t ; $price_t$ is the closing price of the asset on day t , and $price_{t-1}$ is the closing price on the day before. The NAs in the dataframe were forward filled. This results, that if $price_{t-1}$ was not available in the dataset (e.g.: because it was not a trading day), the last available price before day t was used to calculate r_t . During the return calculation the first day of the original time series was dropped. The daily returns of SPY are displayed on Figure 3. The daily returns vary between -5% and +5% with more volatile periods around September of 2009 (Financial Crisis of 2007-08), August of 2011 (Stock Markets Fall), and early 2020 (Coronavirus Stock Market Crash). The augmented Dickey-Fuller test reveals that the time series is stationary. (Test statistic: -17.24, p-value: 0.0000)



Figure 3: Daily returns of SPY for period between 6th of January 2004 and 30th of December 2022

2.3 Summary statistics

This section provides summary statistics on our data.

As shown in Table 3 the distribution of daily returns seems to be centered around zero, with most of the observations falling between the range of -5% and +5%. This is in line

with our expectations based on Figure 3. The light tails of the distribution reflect that daily returns may have extreme values, but these events happen only occasionally thus there are only a few of them in our sample. This is also represented in Figure 6: the daily return's distribution is light tailed relative to the normal distribution: less observation fall on the tails relative to the theoretical quantiles.

Table 3: Descriptive statistics of daily returns and volumes

<i>Variable</i>	<i>Mean</i>	<i>Median</i>	<i>St. Dev.</i>	<i>Minimum</i>	<i>Maximum</i>
Daily return	0.000330	0.000609	0.012153	-0.109424	0.145198
Volume	127 182 197.84	96 878 000	93 580 512.05	15 270,000	871 026 300

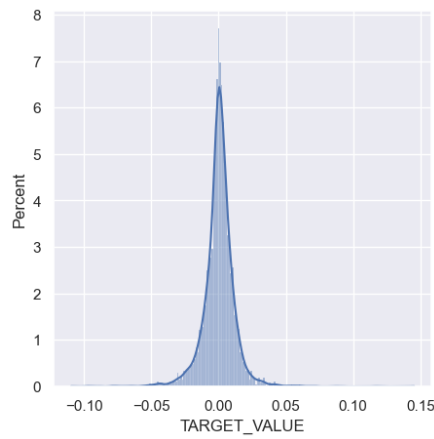


Figure 4: Distribution of daily returns between 2004 and 2022

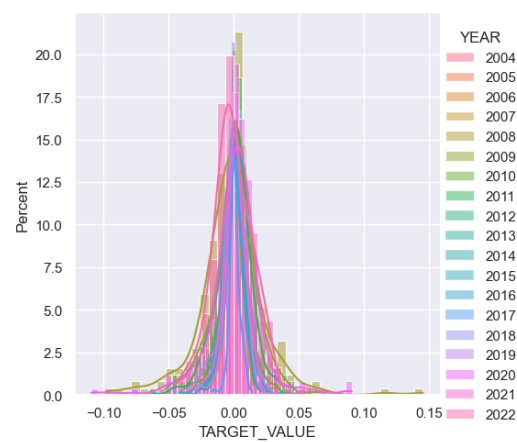


Figure 5: Distribution of daily returns between 2004 and 2022 by year

The yearly distribution of daily returns in Figure 5 reflect the increased volatility around the 2008 financial crisis. The distribution in this year (displayed with a greenish line on the plot) has the fattest tails relative to the return distribution of other years.

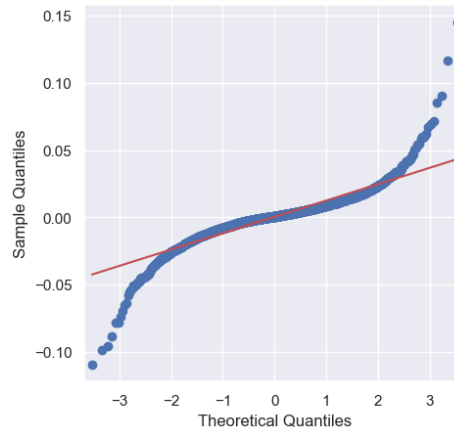


Figure 6: Q-Q plot of the daily returns between 2004 and 2022

Figure 7 displays the overall positive trend in the S&P index's prices: from 2004 to 2022 there were only three years (2008, 2015 and 2022) in which the number of trading days with negative return exceeded the number of days with positive return.

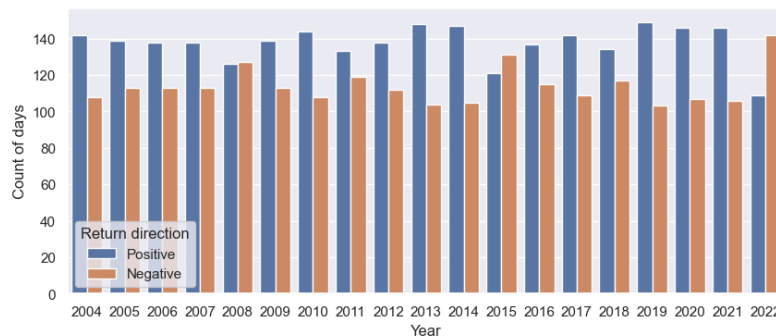


Figure 7: Yearly trading days with negative and positive returns

As Figure 8 shows there are several autocorrelations that are significantly non-zero. The autocorrelation values at lag 1, lag 7, lag 8, lag 9 and at lag 15 and 16 in both ACF and PACF plots are higher than the 5% confidence bound. Hence, it is quite unlikely (only 5% chance) that the underlying process has zero ACF/PACF. That is, it is quite likely (95% chance) that the ACF/PACF is real and not just a matter of coincidence. That is, there is almost surely a pattern in the data and thus we can conclude that time series is not white

noise. The Ljung-Box Q test confirms this finding. Test statistic using lags from 0 to 10 was 99.921645 with a corresponding p-value of 0.0000.

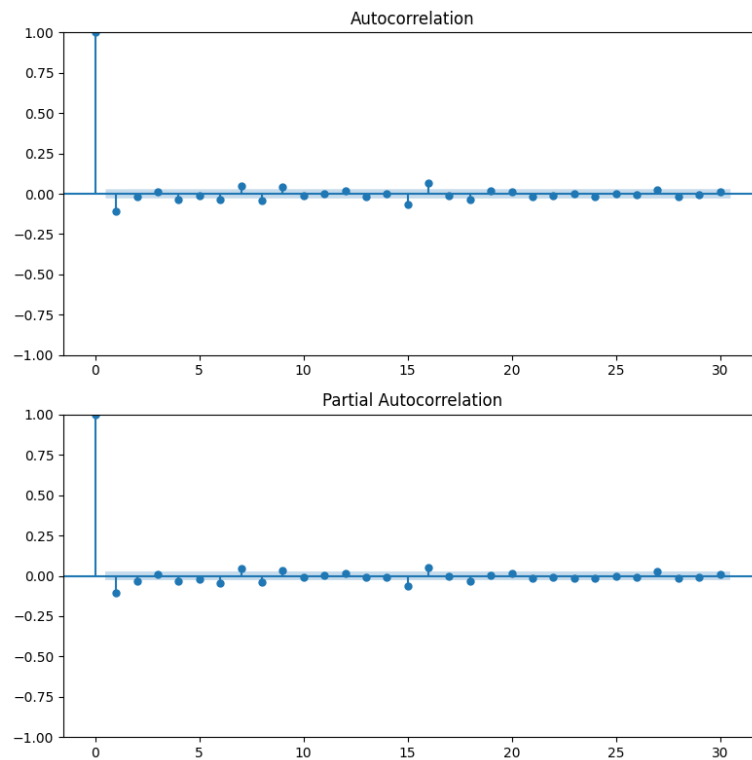


Figure 8: Autocorrelation and partial autocorrelation of the daily returns between 2004 and 2022

To see how the overall autocorrelation changes with time we also calculated a rolling windowed autocorrelation for the daily returns using 30 day windows. The results are plotted in Figure 9. The y-axis shows the windows, where each "line" represents a 30 day window. The start date of these windows are displayed on the axis. The x-axis shows the lags starting from -1 going to -30. The plot shows the measured autocorrelation in the given window at the given lag. The time series are not wrapped, meaning the data on which the correlation is calculated is getting shorter and shorter as we increase lags. The darker red (darker blue) the tile is the higher (the lower) the correlation is at the given lag in the given window.

The plot suggest that the negative autocorrelation with yesterday's return - also depicted in Figure 8 - exist regardless of time. However, Figure 8 showed significant positive autocorrelations at lag 7, lag, 9 and 16, and negative autocorrelation at lag 8 and lag 15. The rolling windowed autocorrelation suggest that these connections are not a general pattern but

rather the result of relatively strong but temporary autocorrelations with the above-listed lags. For example the significant positive autocorrelation with lag 7 and lag 9 are mainly driven by the period between end of 2018 and early 2020.

Figure 9 can also be evaluated as a weak evidence for the regime switching behaviour of stock returns. The autocorrelations found earlier can not be evaluated as a generally true relationship, as they might go weaker or stronger with time and might also change signs. For instance lag 14 is negatively autocorrelated in the period between end of 2017 and end of 2018, but changes its sign and becomes positively autocorrelated in the period between early 2019 and early 2020. This pattern can be seen to happen several times suggesting that returns in the stock market may undergo occasional, distinct shifts.

These switches in the autocorrelation patterns might be swift (e.g.: lag 2 at the end of 2018) or might develop gradually (e.g.: lag 7 between mid 2016 and mid 2018). By looking at the plot one might form the hypothesis that the former one happens more often than the latter. However these regimes can end just as quickly as they started. Also, it seems that these switches affect lags at the same time. For instance, lag 1, lag 2, lag 3 and lag 4 begin to develop a stronger autocorrelation starting at the same time around mid 2010. Another great example is the switch around early 2019 making several lags to form a stronger autocorrelation. Also, the autocorrelation patterns seem to be stable. They might even last for as long as a year: a strong autocorrelation pattern developed around early 2019 until it ended rapidly around early 2020.

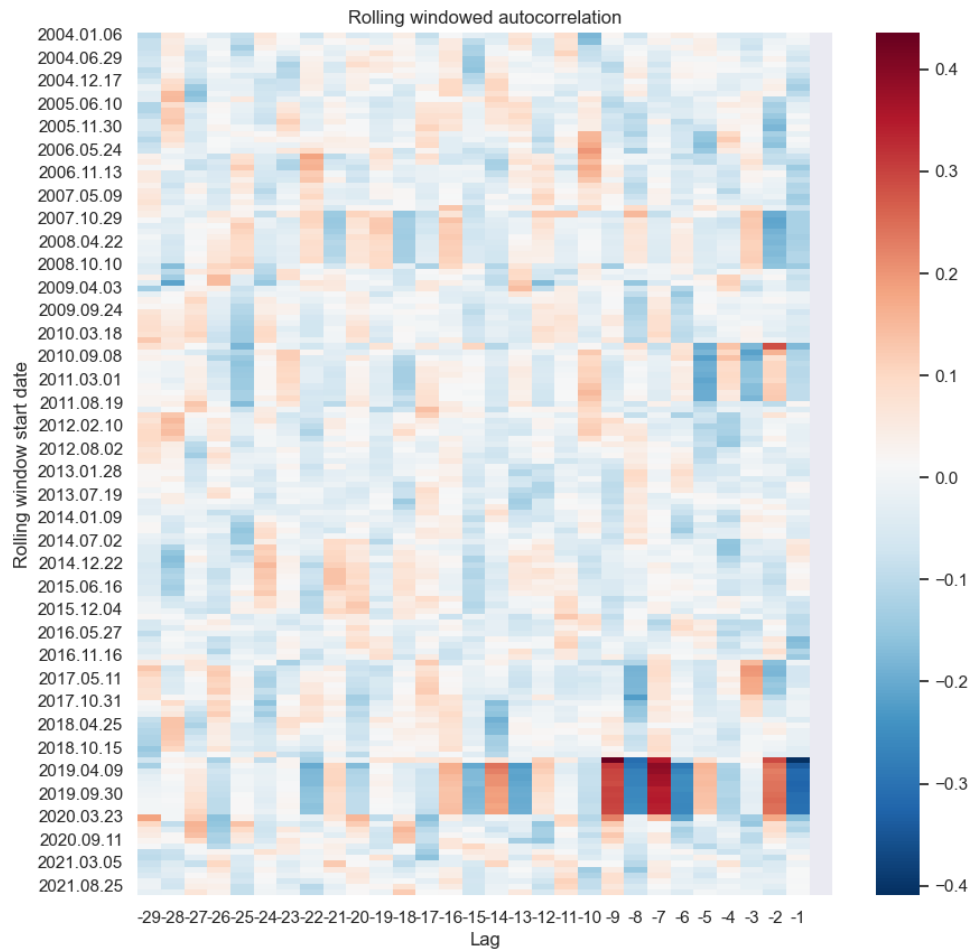


Figure 9: Rolling windowed autocorrelation of the daily returns between 2004 and 2022

Figure 10 shows the rolling windowed cross correlation between daily returns and trading volume. In the y-axis each "line" represents a 30 day window. The x-axis shows the volume lags/forwards starting from +30 going to -30.⁵ The plot shows the measured cross-correlation between returns and volume in the given window at the given lag. The darker red (darker blue) the tile is the higher (the lower) the correlation is at the given lag in the given window.

A clear pattern emerges from the plot: today's returns are generally negatively correlated with trading volumes in the preceding days, while positively correlated with the trading volumes in the following days. This is represented by the bluish colors - dominating lags between -1 to around -7 - turning into redish as forwarded daily volumes are used for correlation calculations. Despite these plots raise several interesting research question, answering these is beyond the scope of this project.

⁵Positive lags in this case means forwards.

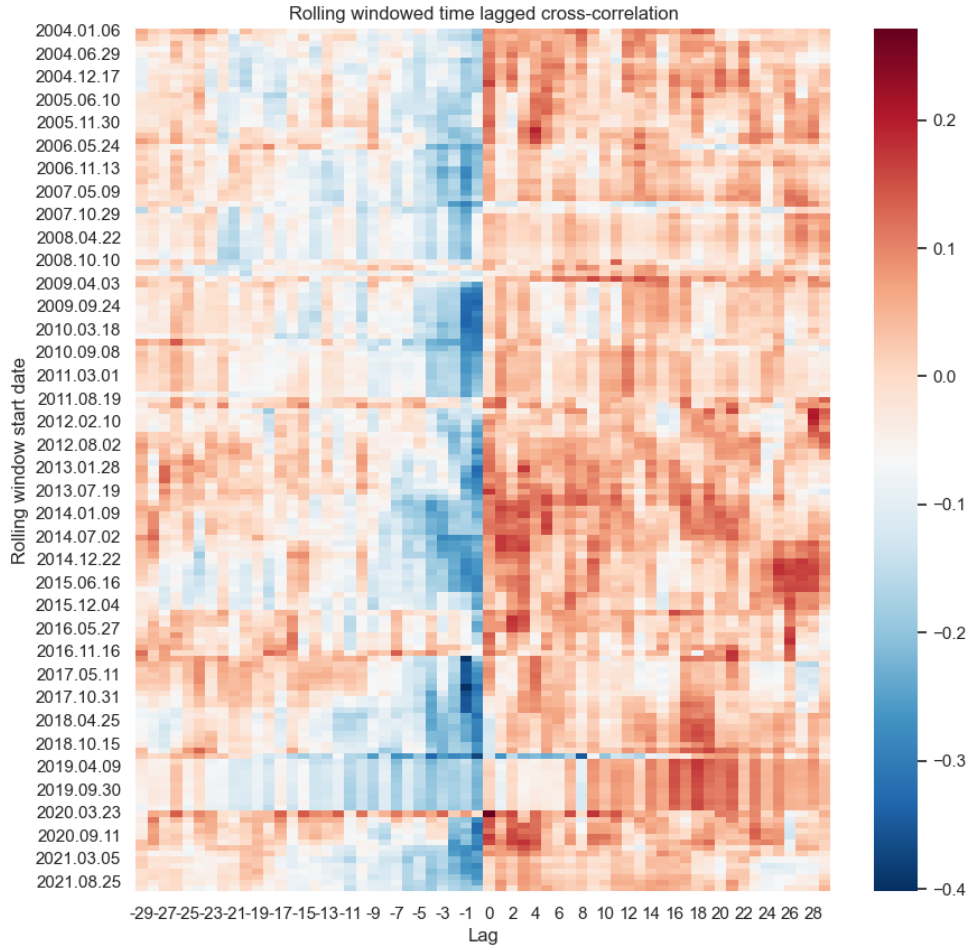


Figure 10: Rolling windowed time lagged cross-correlation of the daily returns between and volume 2004 and 2022

3 Methodology

In an attempt to capture the time-varying behaviour of stock returns, this study employs a dynamical model selection framework to predict stock returns. We propose to asses the regime switching behaviour by choosing the model predicting most accurately in the recent past before every one-step ahead prediction. We hypothesise that dynamically changing the model for predictions might lead to an improvement in prediction accuracy because of the changing underlying data generation process. To analyse this hypothesis we compare the achieved prediction accuracy with proposed method to several benchmarks.

The general idea behind this study is that we create a wide range of different features

(lags, polynoms and their interactions) and create a large number of simple models using different random subsets of these features. Next, we select the models which could predict the recent past most accurately. After that, ensemble predictions are created based on the individual predictions of these top models. The evaluation of the models current accuracy and the selection of the current best models is done daily. Doing so we allow regimes to form the models we use for ensemble predictions and thus implicitly allow to change the features which are used to predict the given day's return direction. If a new regime evolves and lasts for enough periods the model selection will assure that the ensemble predictions are driven by those features which are important in the given regime. .

The actual model giving the predictions is a simple Ordinary Least Squares model. We chose this methodology because it is sufficient for validating the general idea, whilst it is relatively cheap to compute. Also as its' coefficients are easy to interpret, the patterns found by the model can be revealed simply. The primary objective of this study is to determine whether the proposed extension could yield reasonable results with the simple OLS approach. Also results obtained through the proposed method could be compared with those reported in the literature, which extensively uses novel machine-learning techniques. If current results suggest a more complex method could be used instead.

The implementation of the dynamical model selection framework is summarized in Figure 11. The proposed prediction process includes the model evaluation and model selection: we always choose the model predicting most accurately in the recent past, thus accounting for occasional regime changes. The prediction process consists of the following steps:

- **Create a "feature lake".** Generate a wide range of features including lags, polynomial terms, interactions. Lags between 1 day and 21 days were created for trading volume and daily returns variables. Then 2nd, 3rd and 4th degree polynoms were created for the lagged variables. Interactions between the variables defined in the preceding two steps were created up until lags of 10 days. As a result we had a "feature lake" with 10,628 features.⁶
- **Assign features to models randomly.** Create an extensive number of models using different random subsets of the "feature lake" as predictor. The number of features per

⁶To avoid data leakage the actual days trading volume was not included in the "feature lake".

model was defined as follows:

$$\text{number of features per model} = \lfloor \frac{\text{number of features}}{\text{number of models}} \rfloor = \lfloor \frac{10628}{966} \rfloor = 11 \quad (2)$$

We created 966 models, thus 11 feature were randomly assigned to each model (without replacement), with the last model containing 13 features. Note that results can change from experiment to experiment as the assignment of features to the models is random.

- **Run rolling window backtests for all models.** For one model the rolling backtest consists of three steps. The backtest period was between 1st of January 2010 and 31st of December 2022.
 - Model training: train models using daily returns using data from $t-1$ to $t-90$. When training models we used data from the last 90 days. Thus the training set was also rolled during the rolling backtest.
 - Model prediction: predict the daily return of t with the trained model.
 - Shift windows: After the *Model prediction* step the training window and the predicted day is also shifted with one day.

These steps are repeated until t reaches the end of the given backtest period end. The rolling backtest is run for all models each model having different features. Thus for a given day the number of predictions is equal to the number of models. For instance if we have 100 models, than for a given day in the backtest period we would have 100 predictions.

- **Create return sign predictions.** Turn the actual forecasted values and daily returns to show the daily direction (up or down) of the closing price of the SPDR S&P 500 ETF. This is done based on the below-detailed rules, where $\hat{rd}_t$ is the predicted return direction on day t , rd_t is the factual return direction, $\hat{r}_t$ is the return on day t and r_t is the actual return.

$$\hat{rd}_t = \begin{cases} 1, & \text{if } \hat{r}_t \geq 0 \\ -1, & \text{otherwise} \end{cases}$$

$$rd_t = \begin{cases} 1, & \text{if } r_t \geq 0 \\ -1, & \text{otherwise} \end{cases}$$

- **Find the current best model.** To find the model/models which currently has the best performance we employed a rolling evaluation approach. This means that at a given day t we calculated model i 's accuracy $accuracy_{t,i}$ based on the following equation, where TP is the number of true positive predictions, TN is the number of true negative predictions, FN and FP are the number of false negative and positive predictions and ew is the selected evaluation window. Note, that the length of evaluation window controls how fastly the process should react to regime changes. The longer the evaluation window is the slower the new regime's features are incorporated into ensemble predictions. In this experiment we used an evaluation window of 14 observations.

$$accuracy_{t,i} = \frac{TP_{t-ew:t-1} + TN_{t-ew:t-1}}{TP_{t-ew:t-1} + TN_{t-ew:t-1} + FP_{t-ew:t-1} + FN_{t-ew:t-1}} \quad (3)$$

The aforementioned process was conducted iteratively on a daily basis, and continued until the evaluation window spanned the backtest period. Consequently, the resulting table provides daily information on the models' performance on the horizon of the selected evaluation window.

- **Generate ensemble prediction.** The subsequent step of the current process involved conducting a voting based on the different models' predictions to obtain an ensemble forecast of the next day's returns' direction. Several voting method has been implemented, thus the resultant ensemble forecasts for the backtesting period were dependent on the chosen methodology of voting between the models. The following voting methods were included:

- Top n model's direction based ensemble predictions: Based on the rolling accuracy we selected the top n models for every day. For instance, if $n = 10$, than only those models are included in the voting which were among the top 10 models on the given day based on their rolling accuracy. The ensemble prediction is positive (negative) for the next day if the the number of models that predicted a positive return for that day exceeds the number of models with negative return prediction. When creating ensemble predictions we used the best 3, the best 10 and the best

50 models predictions.

$$\hat{rd}_{t, \text{top } n \text{ model's direction based}} = \begin{cases} 1, & \text{if } |TP_t| \geq n - |P_t| \\ -1, & \text{otherwise} \end{cases}$$

where n is the number of top models and $n \leq I$, and $|TP_t|$ is the number of models from the top n model with positive predictions for day t .

- **Calculate ensemble predictions accuracy.** Based on Equation 3 calculate the accuracy of the above described ensemble predictions.
- **Calculate yields.** We buy or short the asset based on the ensemble signals. The yield for a given day ($yield_{t,E}$) is calculated using the following equation:

$$yield_{t,E} = \hat{rd}_{t,E} \times r_t \quad (4)$$

where $\hat{rd}_{t,E}$ is the ensemble prediction and r_t is the actual return for day t .

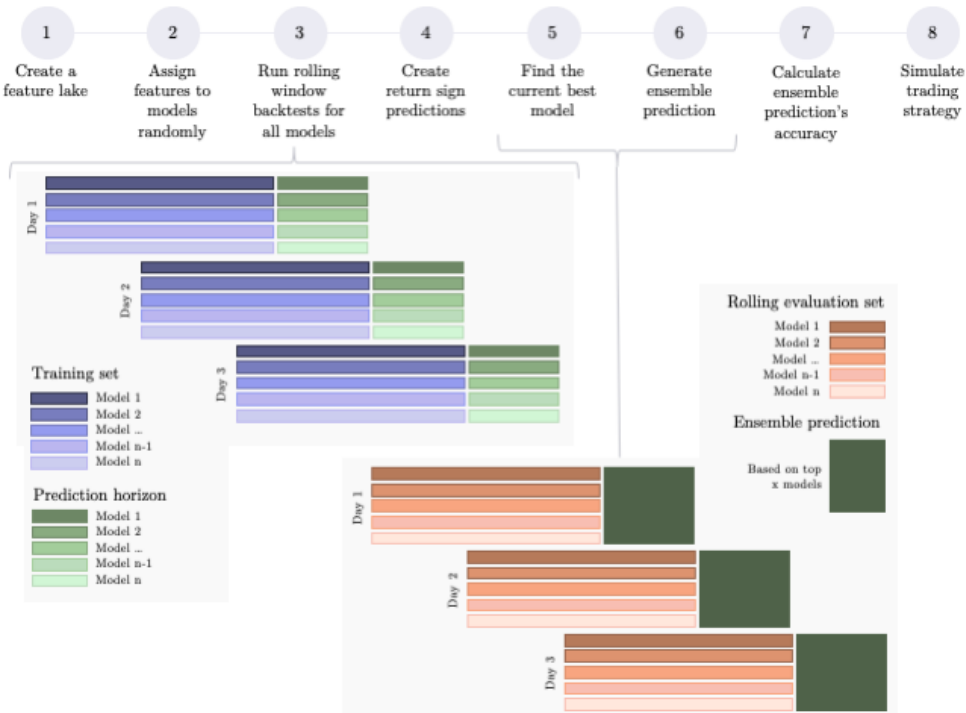


Figure 11: Proposed dynamical model selection framework

4 Results

In this section first we summarize the step-by-step output of the above detailed process: from creating the features to calculating the accuracy of the ensembled predictions. Then we create a trading simulation based on our predictions and test whether we could achieve yields significantly greater than zero. Lastly, we compare our prediction accuracy results to those of multiple baseline models.

4.1 Framework outputs

In this section we present the backtesting results with the above-detailed dynamical model selection framework step-by-step. For the experiment we created a "feature lake" with 10,628 features, including lags, polynomial terms, and interactions of daily returns and trading volume. These features are randomly assigned to create 966 models, with 11 features in each model.⁷ Rolling window backtests are run for all models for the period between 1st of January 2010 and 31st of December 2022, consisting of model training, prediction, and shifting training and prediction windows. Each model generates one return direction prediction per day, resulting in 966 predictions per day. The study creates ensemble predictions based on the current best-performing models for predicting daily return directions.

First we created lags up until 21 days from the daily returns and volume variables, resulting in 42 features. After that, for lags up until 10 days we created 2nd, 3rd and 4th degree polynomials and all of their possible interactions, reaching the final size of the "feature lake" with 10,628 features. We took a random 11 feature sample from the feature lake without replacement and assign the drawn sample to the first model. This step was repeated until before the feature assignment of the last model. After the feature assignment was done for 965 models, the "feature lake" only consisted 13 features. For this reason we assigned the remaining 13 features to the last model.

The rolling backtest was executed as described in Section 3. During the period from January 1st, 2010, to December 31st, 2022, rolling window backtests were conducted for all models, which include the processes of model training, prediction, and the shifting of training and prediction windows. Each model generates a single return direction prediction for each

⁷The last model contains 13 features.

day, resulting in a total of 966 predictions per day. Figure 12 shows the backtest results of a model (model ID = 3), whilst Figure 13 shows according return direction predictions of the given model for January and February of 2022.

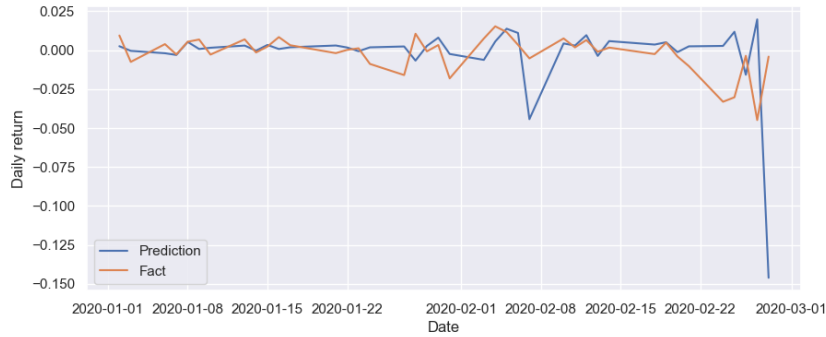


Figure 12: Prediction and fact daily returns of model 654

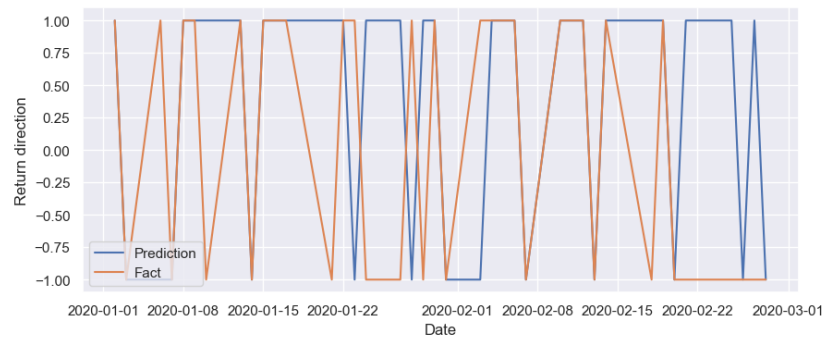


Figure 13: Prediction and fact daily return directions of model 654

As described in Section 3 a two week rolling accuracy was calculated for every day of the backtest period to find the current best model. The results of the rolling evaluation is displayed in Figure 14. The x-axis shows the dates going from 13th of January 2010 to 30th of December 2022. As we use a two-week rolling evaluation window the first two weeks of 2010 is omitted from the analysis. At each day the plot displays the rolling accuracy of the 966 available models, thus the quite noisy plot. However, the plot shows that at every day there is a model which could reach an accuracy above 80% on the last 14 observation. On the other side, there are also models which had a rolling accuracy around 0% meaning that they failed to hit the yield direction once in the last two weeks. The general idea of this paper can be greatly exhibited using this plot: we create ensemble predictions using the models

which are at the top of the rolling accuracy plot, hoping that their great performance will extend to the next period.

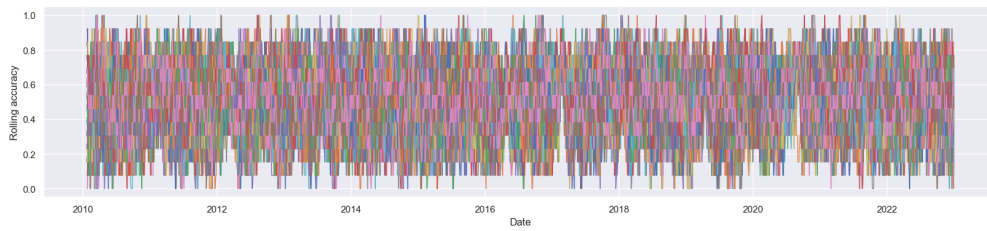


Figure 14: Rolling accuracy

Based on the top 3, top 10 and top 50 models' return direction prediction we created ensemble predictions, signaling positive (negative) return if the majority of the top models predicted positive (negative) return. An example of the ensemble predictions based on the top 3 models are depicted for year 2021 on Figure 15 and for year 2022 on Figure 16. As it can be seen from the plots there are distinct periods when the prediction process always generates positive signals (e.g.: between end of October 2021 and end of November 2021) and periods when the prediction signal is consistently negative (eg.: between mid March 2022 and early May 2022). This suggest that the framework is unable to catch short-term patterns, but rather creates buy and sell periods which can last up to even a couple of months.

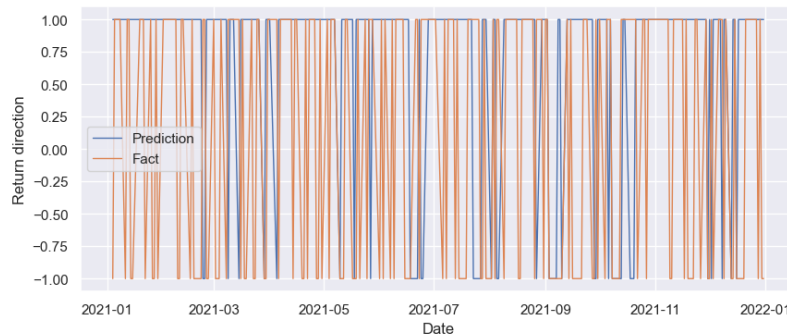


Figure 15: Ensemble predictions based on the top 3 model for year 2021

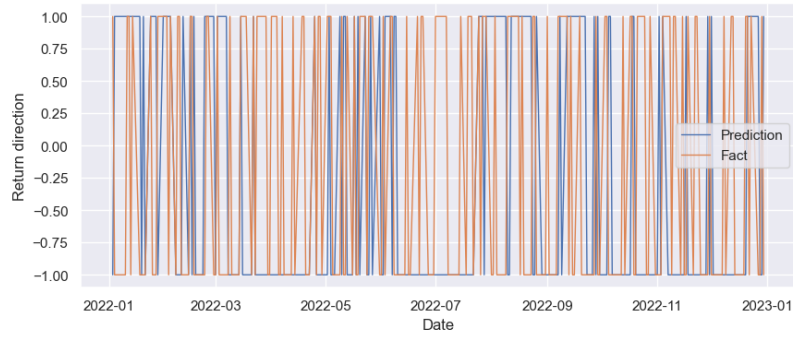


Figure 16: Ensemble predictions based on the top 3 model for year 2022

The three different ensemble prediction's yearly accuracy is displayed in Figure 17. The average accuracy for the period between 2010 and 2022 is 51.72% for the prediction based on the top 3 model, 51.87% for predictions based on the top 10 models, and 51.9% for the those based on the top 50 models. The three voting methods' performance generally moves together, however there might be deviations: for instance the top 3 models based prediction could achieve a substantially higher accuracy in years 2010 and 2020.

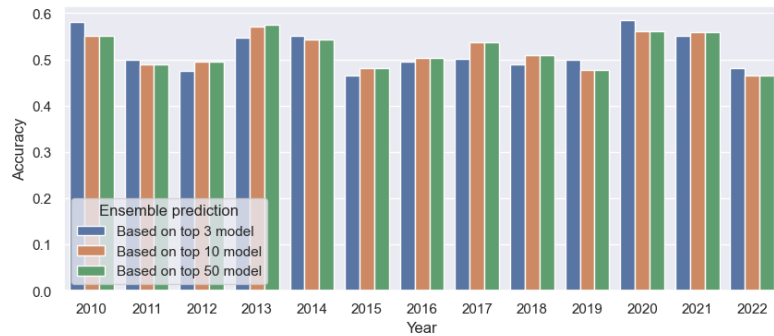


Figure 17: Different ensemble predictions' accuracy

Lastly we calculated the confusion matrix of the different ensemble models, which are shown in Tables 4, 5 and Table 6. The average accuracy on the backtest period was 51.72%, with the lowest accuracy in 2015 with 46.43% and the highest accuracy in 2020 with 58.5%. The confusion matrices reveal that there are only slight differences - which might be random - in the accuracy of the different voting processes.

Table 4: Confusion matrix based on the top 3 model based ensemble prediction

		Prediction outcome		total
		p	n	
actual	p'	True Positive = 1,199	False Negative = 586	P' = 1,785
	n'	False Positive = 987	True Negative = 486	N' = 1,473
total		P = 2,186	N = 1,072	

Table 5: Confusion matrix based on the top 10 model based ensemble prediction

		Prediction outcome		total
		p	n	
actual	p'	True Positive = 1,199	False Negative = 586	P' = 1,785
	n'	False Positive = 982	True Negative = 491	N' = 1,473
total		P = 2,181	N = 1,077	

Table 6: Confusion matrix based on the top 50 model based ensemble prediction

		Prediction outcome		total
		p	n	
actual value	p'	True Positive = 1,200	False Negative = 585	P' = 1,785
	n'	False Positive = 982	True Negative = 491	N' = 1,473
total		P = 2,182	N = 1,076	

4.2 Market simulation

As the comparing accuracy of the different ensemble predictions was inconclusive we created a trading simulation based on the ensemble prediction signals as described in Section 3. The trading logic is designed to "go long" (buy) at the closing price the day before if the signal states that an up-day will occur and then sell at the close. Conversely, the trading is designed "go short" (sell) if the signal states that a down-day will occur and subsequently close out at the closing price. We test that the ensemble model's yields' equals to zero. The hypotheses were defined as follows:

- H_0 : The yielded returns' mean is equal to zero.
- H_1 : The yielded returns' mean is greater than zero.

Table 7: Average yields achieved based on the different ensemble predictions

<i>Voting technique</i>	<i>Average daily yield</i>	<i>T-statistic</i>	<i>p-value</i>	<i>Accuracy</i>
Based on top 3 models	0.000114	0.58	0.28	51.72%
Based on top 10 models	-0.000189	-0.97	0.8322	51.87%
Based on top 50 models	-0.000187	-0.96	0.8314	51.9%

Only the trading strategy based on the top 3 model based ensemble predictions could achieve a positive average daily yield. The average yield based on this trading is 0.000114. Testing that the mean is greater than zero reveals that we can not reject the null hypothesis

(statistic=0.58, p-value=0.28). For this reason, from this point on we are focusing only on the ensemble predictions based on the voting of the current-best top three models. However, it is important to note, that it is quite suspicious that while the accuracy of the different voting methods were basically the same, only one could achieve an average yield greater than zero. As the p-value shows this yield is also indifferent from zero, raising the question that the top 3 model based approach could achieve positive return only by chance.

To get a more nuanced picture we also took the average yields achieved by the top 3 model based approach by year and tested whether they are significantly greater than zero. This is summarized in Table 8. The table contains the yearly average yields, the according statistic and p-value, and also the framework's accuracy in the given year. From the 13 years included in the backtest period the framework could achieve an average positive yield in 8 years, whilst producing negative yields in 5 years. The yield usually insignificant from zero. In 2020 the framework produced an average daily yield of 0.2204%, which is significantly greater than zero on a 5% significance level (p-value: 0.00188). The 0.0784% average yield in 2021 is significantly greater than zero on a 10% significance level, with an according p-value of 0.0671.

Table 8: Average yields achieved based on the ensemble predictions of the top three models

<i>Year</i>	<i>Mean yield</i>	<i>T-statistic</i>	<i>p-value</i>	<i>Accuracy</i>
2010	0.000206	0.27	0.39	57.98%
2011	-0.000156	-0.17	0.5675	50.00%
2012	0.000135	0.26	0.396	47.6%
2013	0.000439	0.99	0.1628	54.76%
2014	0.000221	0.49	0.3112	55.16%
2015	-0.000682	-1.11	0.8651	46.43%
2016	-0.000477	-0.92	0.8198	49.6%
2017	0.000021	0.07	0.4696	50.12%
2018	0.000103	0.15	0.4401	49%
2019	-0.000166	-0.33	0.6291	50%
2020	0.002204	1.67	0.0482**	58.5%
2021	0.000784	1.5	0.0671*	55.16%
2022	-0.001161	-1.21	0.8854	48.21%

Generally, the predictions contain more positive returns than would be justified based on the factual number of positive returns. The number of predicted days with positive return is 2,186, which is 67.1% percent of all trading days in the backtest period. In contrast there

were only 1,785 days where the actual return was positive, which means a ratio of 54.79% which is much lower than predicted.

The general overprediction of up-days raises questions regarding the significant yields in 2020 and 2021. Table 9 shows the ratio of trading days with positive return and the also the ratio of predicted up-days. In 2017, 2020 and 2021 the system produced buy signals over 80% of the trading days. In 2020 and 2021 this overall positive view of the prediction system was matched with relatively high number of days where the actual return was positive. The factual positive return ratio was 57.71% in 2020 and 57.94% in 2021. So a simple trading strategy which takes a buy position every day could reach an accuracy equal to these ratios. In these years the proposed prediction system could reach an accuracy of 58.5% and 55.16%. This suggest that our prediction system in these 2021 is inferior to a simple always buy strategy. This very simple strategy would also outperform our prediction accuracy in 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019 and 2021. Our prediction system's accuracy could achieve better results only in three years: in 2010, 2020 and 2022.

Table 9: Fact and predicted ratio of days with positive return

<i>Year</i>	<i>Positive return ratio (fact)</i>	<i>Positive return ratio (prediction)</i>
2010	56.33%	65.13%
2011	52.78%	59.92%
2012	55.2%	69.2%
2013	58.73%	72.22%
2014	58.33%	75.4%
2015	48.02%	47.62%
2016	54.37%	62.7%
2017	56.57%	84.06%
2018	53.39%	55.78%
2019	59.13%	77.38%
2020	57.71%	81.03%
2021	57.94%	84.52%
2022	43.42%	37.05%

4.3 Comparison to baselines

In this section we compare our results to several benchmarks. The first part consists of a comparison made on a smaller period, whilst the second part is a more comprehensive comparison to test how our prediction system performs on the long run.

To see how our proposed method performs on a shorter period, we compared our results to those reported in the literature using the test set between and of February 2013 and 21st of November 2013. Three different research (Ding et al. 2014, 2015; Santos Pinheiro and Dras 2017) used this period as a test period, thus it is ideal to compare our results to previous works. Their reported accuracy and the accuracy of the different ensemble predictions of our prediction system is summarized in Table 10. Our model is beaten by the models of Ding et al. (Ding et al. 2015) and Santos et al. (Santos Pinheiro and Dras 2017) in term of prediction accuracy. However, the ensemble model based on the top 10 and top 50 models could slightly outperform the earlier work of Ding et al. (Ding et al. 2014)

Table 10: Comparison to baselines using period between 22/02/2013 and 22/11/2013

<i>Model reference</i>	<i>Reported accuracy</i>
(Ding et al. 2015)	66.08%
(Santos Pinheiro and Dras 2017)	63.34%
Ensemble prediction based on top 10 models	59.16%
Ensemble prediction based on top 50 models	59.16%
(Ding et al. 2014)	58.94%
Ensemble prediction based on top 3 models	55.5%

Lastly, we compared the accuracy of the proposed framework to the results of two baseline models using a longer period. As seen from Section 4.2 a simple trading strategy which produces a positive signal for every day generally beats our proposed method. For this reason, we have used this *always-buy strategy* as one of the baselines. Yu et al.’s (P. Yu and Yan 2020) study was used as a second baseline given that it is a fairly new but acknowledged research, and also they report the prediction accuracy for the period between 2010 and 2017, allowing us to conduct a more robust comparisons.

As summarized in Table 11, generally Yu et al.’s deep neural network (DNN) based prediction could reach the highest accuracy. However, it is interesting to note that the very simple always-buy-strategy could beat their sophisticated model’s performance in years 2014 and 2017. Our prediction framework could beat the DNN only in year 2014, thus we can conclude that our OLS based system is inferior relative the the results reported by Yu et al.

Comparing our accuracy to the simple always-buy strategy reveals a slightly less clear picture. Our prediction system could beat this baseline only in years 2010, 2020 and 2022. However, in the remaining nine years this baseline, which does not require any computational

costs could achieve higher accuracy. Based on these results, we can conclude that our prediction system is inferior to both of these baselines.

Table 11: Accuracy achieved by the top three model based ensemble prediction compared to two baseline models

<i>Year</i>	<i>Baseline 1: Always-buy strategy</i>	<i>Baseline 2: P. Yu and Yan 2020</i>	<i>Our accuracy</i>
2010	56.33%	62.76%	57.98%
2011	52.78%	57.08%	50%
2012	55.2%	62.87%	47.6%
2013	58.73%	60.18%	54.76%
2014	58.33%	49.37%	55.16%
2015	48.02%	55.65%	46.43%
2016	54.37%	62.39%	49.6%
2017	56.57%	54.31%	50.2%
2018	53.39%	-	49.00%
2019	59.13%	-	50.00%
2020	57.71%	-	58.5%
2021	57.94%	-	55.16%
2022	43.42%	-	48.21%
Average (2010 - 2017)	55.04%	58.07%	51.47%
Average (2010 - 2022)	54.76%	-	51.72%

5 Robustness checks

In this section we briefly present the results of stock level experiments. The trading simulation with SPY based on the predictions yielded insignificant returns. In order to have a more nuanced picture on the performance of the proposed prediction framework, we predicted the return direction of individual stocks. Large scale experiments show that the accuracy of S&P 500 index prediction is around 60%, whilst predicting individual stocks the accuracy can increase up to around 70%. (Ding et al. 2014) Our results confirm that trading based on individual shares can yield a significantly positive return, however this is not reflected in more accurate forecasts. In this section we aimed to predict the daily stock returns of three different stocks listed in Nasdaq by employing the same dynamical model selection framework detailed

in Section 3. The selected stocks in the experiment were First Solar (FSLR), Monster Beverage Corporation (MNST) and Apple (AAPL). The first two is generally considered a less liquid stock, unlike the latter one. The results are summarized in Table 12 and Table 13.

Table 12: Accuracy achieved by ensemble predictions for each ticker

	AAPL			MNST			FSLR		
	Top 3	Top 10	Top 50	Top 3	Top 10	Top 50	Top 3	Top 10	Top 50
2010	52.94%	50.84%	51.68%	51.68%	52.52%	52.52%	49.58%	47.9%	47.9%
2011	52.38%	51.19%	51.19%	54.37%	50.79%	50.79%	55.56%	52.78%	52.78%
2012	55.6%	54.00%	54.00%	52.00%	54.00%	54.00%	54.00%	57.2%	57.2%
2013	54.76%	53.17%	53.17%	48.41%	50.00%	50.00%	42.06%	43.25%	43.25%
2014	51.59%	51.98%	51.98%	49.60%	50.00%	50.00%	43.65%	46.03%	46.03%
2015	50.00%	47.22%	47.22%	53.57%	54.76%	54.76%	55.16%	53.57%	53.17%
2016	49.21%	46.03%	45.63%	47.62%	49.60%	49.60%	48.02%	50.79%	50.79%
2017	55.38%	55.38%	55.38%	50.6%	54.98%	54.98%	49.81%	52.99%	52.99%
2018	52.19%	52.19%	51.8%	51.00%	51.39%	51.79%	47.81%	49.80%	49.80%
2019	54.76%	53.57%	53.57%	51.59%	52.78%	52.78%	48.41%	50.00%	50.00%
2020	54.55%	54.15%	54.15%	61.26%	60.08%	60.47%	53.36%	53.36%	53.36%
2021	50.4%	50.79%	50.79%	45.24%	42.86%	42.86%	50.4%	52.78%	52.38%
2022	47.41%	45.42%	45.42%	47.41%	47.41%	47.41%	48.61%	51.79%	51.79%
Mean	52.5%	51.2267%	51.2301%	51.10%	51.63%	51.69%	49.72%	50.94%	50.88%

Table 12 summarizes the yearly prediction accuracy for each ticker. For each ticker three different ensemble prediction were calculated based on the top three models' (Top 3), top 10 models' (Top 10) and top 50 models' (Top 50) predictions. The figure displays that the system could achieve a similar average accuracy for individual stocks and the SPY index over the period of 2010 and 2022. This is somewhat contradicts the findings of previous research, stating that individual stock predictions' accuracy is generally higher then prediction accuracy of the S&P 500 index. The general yearly accuracy is around 50%. The highest accuracy was achieved in 2020 for the MNST ticker (60.08%), and the lowest in 2013 for FSLR ticker (42.06%). Moreover we can not draw conclusions regarding the accuracy on liquid and less liquid stocks, as the overall prediction accuracy for AAPL, MNST and FSLR stocks do not differ greatly. Lastly, we can not generalize our earlier finding, that the top 3 model based ensemble prediction outperform the top 10 and top 50 model based predictions' accuracy, as this pattern do not emerge in stock predictions.

The next table (Table 12) summarizes the trading simulations based on the predictions for each individual ticker. Similarly the three different ensemble prediction is shown in columns Top 3, Top 10 and Top 50 respectively. For each year the first line shows the average daily yield achieved in the given year, whilst the second row shows the significance level. The next section of the table displays the average daily yield achieved throughout the whole period (Row: Mean), the corresponding p-value (Row: P-value) and the significance level. The last two rows of the table shows the number of years with positive yield, and the number of years with significantly positive yield respectively.

In general we found, that whilst the trading simulation was not succesfull for the SPY index, it can achieve significantly positive returns for stocks. The average daily yield for Apple was 0.0791% (top 3 model based), for Monster was 0.0716% (top 3 model based) and 0.1414% (top 10 model based) for First Solar's stocks. The average yields achieved is quite similar in magnitude for the AAPL ticker and for the MNST ticker, whilst the yields are much more higher for the First Solar stock. From the 13 years in focus, the trading strategy would have achieved positive returns in 11 years for AAPL, 8 years for MNST and 10 years for FSLR.

Table 13: Yield achieved by ensemble predictions for each ticker

	AAPL			MNST			FSLR		
	Top 3	Top 10	Top 50	Top 3	Top 10	Top 50	Top 3	Top 10	Top 50
2010	0.000439 <i>ns</i>	0.000159 <i>ns</i>	0.000324 <i>ns</i>	0.000443 <i>ns</i>	0.000853 <i>ns</i>	0.000853 <i>ns</i>	-0.001032 <i>ns</i>	-0.002615 <i>ns</i>	-0.002615 <i>ns</i>
2011	0.000388 <i>ns</i>	0.000107 <i>ns</i>	0.000107 <i>ns</i>	0.00287 **	0.001683 <i>ns</i>	0.001683 <i>ns</i>	0.0048 **	0.002853 <i>ns</i>	0.002853 <i>ns</i>
2012	0.001947 **	0.001593 *	0.001593 *	0.000622 <i>ns</i>	0.00087 <i>ns</i>	0.00087 <i>ns</i>	0.000827 <i>ns</i>	0.00217 <i>ns</i>	0.00217 <i>ns</i>
2013	0.001764 *	0.001804 *	0.001804 *	0.000464 <i>ns</i>	0.000554 <i>ns</i>	0.000554 <i>ns</i>	-0.005019 <i>ns</i>	-0.001287 <i>ns</i>	-0.001287 <i>ns</i>
2014	-0.00006 <i>ns</i>	0.000206 <i>ns</i>	0.000206 <i>ns</i>	0.002408 *	0.001503 <i>ns</i>	0.001503 <i>ns</i>	-0.000506 <i>ns</i>	0.001248 <i>ns</i>	0.001248 <i>ns</i>
2015	-0.001018 <i>ns</i>	-0.00079 <i>ns</i>	-0.00079 <i>ns</i>	-0.000403 <i>ns</i>	-0.001132 <i>ns</i>	-0.001132 <i>ns</i>	0.001369 <i>ns</i>	0.001859 <i>ns</i>	0.001714 <i>ns</i>
2016	0.000219 <i>ns</i>	-0.00082 <i>ns</i>	-0.000877 <i>ns</i>	-0.000858 <i>ns</i>	-0.000372 <i>ns</i>	-0.000372 <i>ns</i>	-0.000517 <i>ns</i>	0.00272 *	0.00272 *
2017	0.001887 ***	0.001383 **	0.001383 **	0.000736 <i>ns</i>	0.001939 ***	0.001939 ***	0.000211 <i>ns</i>	0.003312 **	0.003312 **
2018	0.000068 <i>ns</i>	0.000827 <i>ns</i>	0.000716 <i>ns</i>	0.000754 <i>ns</i>	0.000557 <i>ns</i>	0.000631 <i>ns</i>	-0.001446 <i>ns</i>	-0.000239 <i>ns</i>	-0.000239 <i>ns</i>
2019	0.000688 <i>ns</i>	0.001303 <i>ns</i>	0.001303 <i>ns</i>	-0.000172 <i>ns</i>	0.000394 <i>ns</i>	0.000394 <i>ns</i>	0.000814 <i>ns</i>	0.000644 <i>ns</i>	0.000644 <i>ns</i>
2020	0.003677 **	0.001237 <i>ns</i>	0.001237 <i>ns</i>	0.003682 ***	0.002085 *	0.002144 *	0.003344 *	0.002907 <i>ns</i>	0.002907 <i>ns</i>
2021	0.000153 <i>ns</i>	0.000048 <i>ns</i>	0.000048 <i>ns</i>	-0.000398 <i>ns</i>	-0.000592 <i>ns</i>	-0.000592 <i>ns</i>	0.001204 <i>ns</i>	0.001353 <i>ns</i>	0.001296 <i>ns</i>
2022	0.000112 <i>ns</i>	-0.000874 <i>ns</i>	-0.000874 <i>ns</i>	-0.000871 <i>ns</i>	-0.001006 <i>ns</i>	-0.001006 <i>ns</i>	0.000479 <i>ns</i>	0.003237 *	0.003237 *
Mean	0.000791	0.000477	0.000476	0.000716	0.000563	0.000574	0.000355	0.001414	0.001398
P-value	0.0063	0.0665	0.0669	0.0216	0.0559	0.05275	0.2736	0.0083	0.0089
Sign.	***	*	*	**	*	*	<i>ns</i>	***	***
Pos. years	11	10	10	8	9	9	8	10	10
Prof. years	4	3	3	3	2	2	2	3	3

ns: not significant, *: p-value < 0.1, **: p-value < 0.05, ***: p-value < 0.01

Abbreviations in table:

Sign.: signficancy level

Pos.years: number of years with positive returns

Prof.years: number of years with significantly positive yield.

6 Conclusion

This study aimed to predict the daily stock returns of S&P 500 index ETF (SPY) by employing a dynamical model selection framework that captures the time-varying behavior of stock markets. We propose to assess regime switches by

- creating a wide range of different features (lags, polynoms and their interactions)
- creating a large number of simple models using different random subsets of these features
- evaluating the models accuracy on the last two-weeks on a daily basis
- choosing the top three current best model every day (thus allowing for features used for prediction to change dynamically)
- creating ensemble predictions based on the top three models signals.

We hypothesised that dynamically choosing the model predicting most accurately in the recent past before every one-step ahead prediction might lead to an improvement in prediction accuracy because it explicitly allows the framework to use those features which are important in the given regime. The daily predictions were tested on the period between 2010 and 2022. The proposed framework's results are mixed, but generally, they suggest that our original hypothesis was wrong: our complicated prediction system could not deliver a sharp improvement in accuracy.

We tested our hypothesis from two approaches: first, by running market simulations and testing that the generated returns are greater than zero, secondly comparing our accuracy to several baseline models from previous researches.

The trading strategy based on the top 3 model based ensemble predictions could achieve an average daily yield of 0.0114%, which is not greater than zero on a 5% significance level. However, the strategy could achieve an average positive yield in 8 years from the 13 years included in the backtest period. In year 2020 the strategy yielded an average daily return of 0.2204% which is significantly greater than zero, with a corresponding p-value of 0.0482. Similarly, in 2021 the average daily yield of this market simulation was 0.784%, which is

greater than zero on a 10% significance level. However, these results should be evaluated with caution as in these years the system produced buy signals over 80% of the trading days.

To see how our proposed method performs on a shorter period, we compared our results to those reported in the literature using the test set between and of February 2013 and 21st of November 2013. Our ensemble prediction accuracy based on the top 10 models could achieve better performance than the deep neural network based model of Ding et al. (Ding et al. 2014). However, the achieved 59.16% accuracy during this period still falls short of other researches reporting over 60% accuracy.

Lastly, we compared the accuracy of the proposed framework to the results of a simple always-buy prediction system and to those reported in Yu et al's (P. Yu and Yan 2020) study. Despite the proposed ensemble predictions could outperform Yu et al.'s deep neural network (DNN) model in 2014, and the always-buy strategy baseline in years 2010, 2020 and 2022, in the remaining years our frameworks' accuracy falls short of the benchmarks. Based on the results on the period between 2010 and 2022 we can conclude that our prediction system is inferior to both of these baselines.

To gain a more comprehensive evaluation of the proposed prediction framework, the study also conducted an experiment predicting the return direction of individual stocks. Section 5 utilized the proposed method to predict daily stock returns of three different Nasdaq-listed stocks: First Solar (FSLR), Monster Beverage Corporation (MNST), and Apple (AAPL). The experiment included less liquid stocks, FSLR and MNST, and a more liquid stock, AAPL. Results indicated that while the trading simulation was not successful for the SPY index, it has the potential to generate significantly positive returns for individual stocks.

Despite the limited accuracy of the proposed framework, this study might help further research. Based on Figure 9, one might attempt to create a variable that tries to indicate the likelihood that we are in a stable regime and the found relationships will hold for the next period as well. For instance, the Figure greatly displays that there is a switch around early 2019 making several lags to form a stable, stronger autocorrelation. This strong autocorrelation pattern developed around early 2019 until it ended rapidly around early 2020. In this case, as we move forward from early 2019, the probability of being in a stable regime would increase from day to day. Then, when the regime breaks around early 2020, this probability would drop close to zero. The trading strategy introduced in this study could be

extended with this: trading is only executed when the probability of being in a regime is greater than a given threshold. Also, another good extension to the proposed system could be the addition of new variables. This might include newly calculated features based on the current data, such as moving averages, and also price and volume data on other market indices, currency rates and commodities to exploit the interconnections between different markets.

References

- Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang (2006). “The cross-section of volatility and expected returns”. In: *The journal of finance* 61.1, pp. 259–299.
- Ang, Andrew and Allan Timmermann (2012). “Regime changes and financial markets”. In: *Annu. Rev. Financ. Econ.* 4.1, pp. 313–337.
- Atsalakis, George S and Kimon P Valavanis (2009). “Forecasting stock market short-term trends using a neuro-fuzzy based methodology”. In: *Expert systems with Applications* 36.7, pp. 10696–10707.
- Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan (2012). “Global, local, and contagious investor sentiment”. In: *Journal of financial economics* 104.2, pp. 272–287.
- Bali, Turan G, Nusret Cakici, and Robert F Whitelaw (2011). “Maxing out: Stocks as lotteries and the cross-section of expected returns”. In: *Journal of financial economics* 99.2, pp. 427–446.
- Ball, Ray and Philip Brown (1968). “An empirical evaluation of accounting income numbers”. In: *Journal of accounting research*, pp. 159–178.
- Basodi, Sunitha, Chunyan Ji, Haiping Zhang, and Yi Pan (2020). “Gradient amplification: An efficient way to train deep neural networks”. In: *Big Data Mining and Analytics* 3.3, pp. 196–207.
- Bernard, Victor L and Jacob K Thomas (1989). “Post-earnings-announcement drift: delayed price response or risk premium?” In: *Journal of Accounting research* 27, pp. 1–36.

- Boehme, Rodney D and Sorin M Sorescu (2002). “The long-run performance following dividend initiations and resumptions: underreaction or product of chance?” In: *the Journal of Finance* 57.2, pp. 871–900.
- Campbell, John Y, Jens Hilscher, and Jan Szilagyi (2008). “In search of distress risk”. In: *The Journal of finance* 63.6, pp. 2899–2939.
- Chen, Jou-Fan, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, and An-Pin Chen (2016). “Financial time-series data analysis using deep convolutional neural networks”. In: *2016 7th International conference on cloud computing and big data (CCBD)*. IEEE, pp. 87–92.
- Chen, Kai, Yi Zhou, and Fangyan Dai (2015). “A LSTM-based method for stock returns prediction: A case study of China stock market”. In: *2015 IEEE international conference on big data (big data)*. IEEE, pp. 2823–2824.
- Chen, Sheng and Hongxiang He (2018). “Stock prediction using convolutional neural network”. In: *IOP Conference series: materials science and engineering*. Vol. 435. 1. IOP Publishing, p. 012026.
- Chen, Weiling, Chai Kiat Yeo, Chiew Tong Lau, and Bu Sung Lee (2018). “Leveraging social media news to predict stock index movement using RNN-boost”. In: *Data & Knowledge Engineering* 118, pp. 14–24.
- Chen, Mu-Yen, Chien-Hsiang Liao, and Ren-Pao Hsieh (2019). “Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach”. In: *Computers in Human Behavior* 101, pp. 402–408.
- Chen, Yingjun and Yongtao Hao (2017). “A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction”. In: *Expert Systems with Applications* 80, pp. 340–355.
- Chong, Eunsuk, Chulwoo Han, and Frank C Park (2017). “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies”. In: *Expert Systems with Applications* 83, pp. 187–205.
- Cohen, Lauren and Dong Lou (2012). “Complicated firms”. In: *Journal of financial economics* 104.2, pp. 383–400.
- Csillag, Balázs and Gábor Neszveda (2020). “A gazdasági várakozások hatása a tőzsdei momentumstratégiára”. In: *Közgazdasági Szemle* 67.11, pp. 1093–1111.

- Da, Zhi, Umit G Gurun, and Mitch Warachka (2014). “Frog in the pan: Continuous information and momentum”. In: *The review of financial studies* 27.7, pp. 2171–2218.
- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998). “Investor psychology and security market under-and overreactions”. In: *the Journal of Finance* 53.6, pp. 1839–1885.
- De Bondt, Werner FM and Richard Thaler (1985). “Does the stock market overreact?” In: *The Journal of finance* 40.3, pp. 793–805.
- De Bondt, Werner FM and Richard H Thaler (1990). “Do security analysts overreact?” In: *The American economic review*, pp. 52–57.
- Diether, Karl B, Christopher J Malloy, and Anna Scherbina (2002). “Differences of opinion and the cross section of stock returns”. In: *The journal of finance* 57.5, pp. 2113–2141.
- Ding, Xiao, Yue Zhang, Ting Liu, and Junwen Duan (2014). “Using structured events to predict stock price movement: An empirical investigation”. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1415–1425.
- (2015). “Deep learning for event-driven stock prediction”. In: *Twenty-fourth international joint conference on artificial intelligence*.
- Fairfield, Patricia M, J Scott Whisenant, and Teri Lombardi Yohn (2003). “Accrued earnings and growth: Implications for future profitability and market mispricing”. In: *The accounting review* 78.1, pp. 353–371.
- Fama, Eugene F (1970). “Efficient capital markets: A review of theory and empirical work”. In: *The journal of Finance* 25.2, pp. 383–417.
- Greenwood, Robin and Samuel G Hanson (2012). “Share issuance and factor timing”. In: *The Journal of Finance* 67.2, pp. 761–798.
- Gu, Jiuxiang et al. (2018). “Recent advances in convolutional neural networks”. In: *Pattern recognition* 77, pp. 354–377.
- Gudelek, M Ugur, S Arda Boluk, and A Murat Ozbayoglu (2017). “A deep learning based stock trading model with 2-D CNN trend detection”. In: *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE, pp. 1–8.

- Hamilton, James D (1989). “A new approach to the economic analysis of nonstationary time series and the business cycle”. In: *Econometrica: Journal of the econometric society*, pp. 357–384.
- Hartzmark, Samuel M and David H Solomon (2013). “The dividend month premium”. In: *Journal of Financial Economics* 109.3, pp. 640–660.
- Henkel, Sam James, J Spencer Martin, and Federico Nardari (2011). “Time-varying short-horizon predictability”. In: *Journal of financial economics* 99.3, pp. 560–580.
- Heston, Steven L and Ronnie Sadka (2008). “Seasonality in the cross-section of stock returns”. In: *Journal of Financial Economics* 87.2, pp. 418–445.
- Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang (2004). “Do investors overvalue firms with bloated balance sheets?” In: *Journal of Accounting and Economics* 38, pp. 297–331.
- Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li (2013). “Innovative efficiency and stock returns”. In: *Journal of financial economics* 107.3, pp. 632–654.
- Hong, Harrison and Marcin Kacperczyk (2009). “The price of sin: The effects of social norms on markets”. In: *Journal of financial economics* 93.1, pp. 15–36.
- Hoseinzade, Ehsan and Saman Haratizadeh (2019). “CNNpred: CNN-based stock market prediction using a diverse set of variables”. In: *Expert Systems with Applications* 129, pp. 273–285.
- Hu, Hongping, Li Tang, Shuhua Zhang, and Haiyan Wang (2018). “Predicting the direction of stock markets using optimized neural networks with Google Trends”. In: *Neurocomputing* 285, pp. 188–195.
- Hu, Zexin, Yiqi Zhao, and Matloob Khushi (2021). “A survey of forex and stock price prediction using deep learning”. In: *Applied System Innovation* 4.1, p. 9.
- Hu, Ziniu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu (2018). “Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction”. In: *Proceedings of the eleventh ACM international conference on web search and data mining*, pp. 261–269.
- Jacobs, Heiko (2015). “What explains the dynamics of 100 anomalies?” In: *Journal of Banking & Finance* 57, pp. 65–85.

- Jacobs, Heiko and Martin Weber (2015). “On the determinants of pairs trading profitability”. In: *Journal of financial markets* 23, pp. 75–97.
- Jegadeesh, Narasimhan (1990). “Evidence of predictable behavior of security returns”. In: *The Journal of finance* 45.3, pp. 881–898.
- Jegadeesh, Narasimhan and Sheridan Titman (1993). “Returns to buying winners and selling losers: Implications for stock market efficiency”. In: *The Journal of finance* 48.1, pp. 65–91.
- Kim, Dongcheol, Tai-Yong Roh, Byoung-Kyu Min, and Suk-Joon Byun (2014). “Time-varying expected momentum profits”. In: *Journal of Banking & Finance* 49, pp. 191–215.
- Kumbure, Mahinda Mailagaha, Christoph Lohrmann, Pasi Luukka, and Jari Porras (2022). “Machine learning techniques and data for stock market forecasting: A literature review”. In: *Expert Systems with Applications*, p. 116659.
- Lehmann, Bruce N (1990). “Fads, martingales, and market efficiency”. In: *The Quarterly Journal of Economics* 105.1, pp. 1–28.
- Li, Chang, Dongjin Song, and Dacheng Tao (2019). “Multi-task recurrent neural networks and higher-order Markov random fields for stock price movement prediction: Multi-task RNN and higher-order MRFs for stock price classification”. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1141–1151.
- Li, Jiahong, Hui Bu, and Junjie Wu (2017). “Sentiment-aware stock market prediction: A deep learning method”. In: *2017 international conference on service systems and service management*. IEEE, pp. 1–6.
- Lim, Bryan, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister (2021). “Temporal fusion transformers for interpretable multi-horizon time series forecasting”. In: *International Journal of Forecasting* 37.4, pp. 1748–1764.
- Lim, Kian-Ping and Robert Brooks (2011). “The evolution of stock market efficiency over time: A survey of the empirical literature”. In: *Journal of Economic Surveys* 25.1, pp. 69–108.
- Liu, Yang, Qingguo Zeng, Huanrui Yang, and Adrian Carrio (2018). “Stock price movement prediction from financial news with deep learning and knowledge graph embedding”.

- In: *Knowledge Management and Acquisition for Intelligent Systems: 15th Pacific Rim Knowledge Acquisition Workshop, PKAW 2018, Nanjing, China, August 28-29, 2018, Proceedings 15*. Springer, pp. 102–113.
- Long, Jiawei, Zhaopeng Chen, Weibing He, Taiyu Wu, and Jiangtao Ren (2020). “An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market”. In: *Applied Soft Computing* 91, p. 106205.
- Lou, Dong (2014). “Attracting investor attention through advertising”. In: *The Review of Financial Studies* 27.6, pp. 1797–1829.
- McLean, R David (2010). “Idiosyncratic risk, long-term reversal, and momentum”. In: *Journal of Financial and Quantitative Analysis* 45.4, pp. 883–906.
- Mérő, Balázs, Olivér Nagy, and Gábor Neszveda (2019). “Új faktorok tesztelése az empirikus eszközárakban”. In: *SZIGMA Matematikai-közgazdasági folyóirat* 50.4, pp. 263–281.
- Michael, Roni, Richard H Thaler, and Kent L Womack (1995). “Price reactions to dividend initiations and omissions: Overreaction or drift?” In: *the Journal of Finance* 50.2, pp. 573–608.
- Minh, Dang Lien, Abolghasem Sadeghi-Niaraki, Huynh Duc Huy, Kyungbok Min, and Hyeonjoon Moon (2018). “Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network”. In: *Ieee Access* 6, pp. 55392–55404.
- Naik, Nagaraj and Biju R Mohan (2019). “Stock price movements classification using machine and deep learning techniques-the case study of indian stock market”. In: *Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24-26, 2019, Proceedings 20*. Springer, pp. 445–452.
- Nelson, David MQ, Adriano CM Pereira, and Renato A De Oliveira (2017). “Stock market’s price movement prediction with LSTM neural networks”. In: *2017 International joint conference on neural networks (IJCNN)*. Ieee, pp. 1419–1426.
- Pagan, Adrian R and Kirill A Sossounov (2003). “A simple framework for analysing bull and bear markets”. In: *Journal of applied econometrics* 18.1, pp. 23–46.
- Pang, Xiongwen, Yanqiang Zhou, Pan Wang, Weiwei Lin, and Victor Chang (2020). “An innovative neural network approach for stock market prediction”. In: *The Journal of Supercomputing* 76, pp. 2098–2118.

- Paye, Bradley S and Allan Timmermann (2006). “Instability of return prediction models”. In: *Journal of Empirical Finance* 13.3, pp. 274–315.
- Rana, Masud, Md Mohsin Uddin, and Md Mohaimnul Hoque (2019). “Effects of activation functions and optimizers on stock price prediction using LSTM recurrent networks”. In: *Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence*, pp. 354–358.
- Rapach, David and Guofu Zhou (2013). “Forecasting stock returns”. In: *Handbook of economic forecasting*. Vol. 2. Elsevier, pp. 328–383.
- Rosenthal, Leonard and Colin Young (1990). “The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever NV/PLC”. In: *Journal of Financial Economics* 26.1, pp. 123–141.
- Rosillo, Rafael, Javier Giner, and David de la Fuente (2014). “The effectiveness of the combined use of VIX and support vector machines on the prediction of S&P 500”. In: *Neural Computing and Applications* 25, pp. 321–332.
- Santos Pinheiro, Leonardo dos and Mark Dras (2017). “Stock market prediction with deep learning: A character-based neural language model for event-based trading”. In: *Proceedings of the Australasian Language Technology Association Workshop 2017*, pp. 6–15.
- Schaller, Huntley and Simon Van Norden (1997). “Regime switching in stock market returns”. In: *Applied Financial Economics* 7.2, pp. 177–191.
- Sedighi, Mojtaba, Hossein Jahangirnia, Mohsen Gharakhani, and Saeed Farahani Fard (2019). “A novel hybrid model for stock price forecasting based on metaheuristics and support vector machine”. In: *Data* 4.2, p. 75.
- Shah, Dev, Wesley Campbell, and Farhana H Zulkernine (2018). “A comparative study of LSTM and DNN for stock market forecasting”. In: *2018 IEEE international conference on big data (big data)*. IEEE, pp. 4148–4155.
- Shi, Lei, Zhiyang Teng, Le Wang, Yue Zhang, and Alexander Binder (2018). “DeepClue: visual interpretation of text-based deep stock prediction”. In: *IEEE Transactions on Knowledge and Data Engineering* 31.6, pp. 1094–1108.
- Shumway, Tyler (2001). “Forecasting bankruptcy more accurately: A simple hazard model”. In: *The journal of business* 74.1, pp. 101–124.

- Sim, Hyun Sik, Hae In Kim, Jae Joon Ahn, et al. (2019). “Is deep learning for image recognition applicable to stock market prediction?” In: *Complexity* 2019.
- Song, Yoojeong, Jae Won Lee, and Jongwoo Lee (2019). “A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction”. In: *Applied Intelligence* 49, pp. 897–911.
- Tu, Jun (2010). “Is regime switching in stock returns important in portfolio decisions?” In: *Management Science* 56.7, pp. 1198–1215.
- Wen, Min, Ping Li, Lingfei Zhang, and Yan Chen (2019). “Stock market trend prediction using high-order information of time series”. In: *Ieee Access* 7, pp. 28299–28308.
- Weng, Bin, Mohamed A Ahmed, and Fadel M Megahed (2017). “Stock market one-day ahead movement prediction using disparate data sources”. In: *Expert Systems with Applications* 79, pp. 153–163.
- Xu, Yichuan and Vlado Keselj (2019). “Stock prediction using deep learning and sentiment analysis”. In: *2019 IEEE international conference on big data (big data)*. IEEE, pp. 5573–5580.
- Yang, Hui, Yingying Zhu, and Qiang Huang (2018). “A multi-indicator feature selection for CNN-driven stock index prediction”. In: *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V* 25. Springer, pp. 35–46.
- Yu, Pengfei and Xuesong Yan (2020). “Stock price prediction based on deep neural networks”. In: *Neural Computing and Applications* 32, pp. 1609–1628.
- Yu, Yong, Xiaosheng Si, Changhua Hu, and Jianxun Zhang (2019). “A review of recurrent neural networks: LSTM cells and network architectures”. In: *Neural computation* 31.7, pp. 1235–1270.
- Zeng, Yawen and Xing Liu (2018). “A-stock price fluctuation forecast model based on LSTM”. In: *2018 14th International Conference on Semantics, Knowledge and Grids (SKG)*. IEEE, pp. 261–264.
- Zhang, X Frank (2006). “Information uncertainty and stock returns”. In: *The journal of Finance* 61.1, pp. 105–137.

- Zhao, Zhiyong, Ruonan Rao, Shaoxiong Tu, and Jun Shi (2017). “Time-weighted LSTM model with redefined labeling for stock trend prediction”. In: *2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI)*. IEEE, pp. 1210–1217.
- Zhong, Xiao and David Enke (2017). “Forecasting daily stock market return using dimensionality reduction”. In: *Expert Systems with Applications* 67, pp. 126–139.