Machine Learning Methods for Equity Time Series Forecasting: A Compendium

Alberto Matuozzo[1], Paul D. Yoo[1], Alessandro Provetti[1] and Maria H. Kim[2]

[1]Birkbeck College, University of London, Malet St, London, WC1E 7HX, United Kingdom
[2]University of Wollongong, Wollongong, NSW 2522, Australia

Abstract
Machine learning is a method of building predictive models using a vast amount of data from different sources, capturing non-linear relationships between different variables. As a result, financial markets in general and stock markets in particular, offer a promising ground for the application of such method. This survey examines machine learning methods for equity market forecasting, identifying gaps in current knowledge and suggesting potential avenues to pursue further research. Computer science-centred quantitative studies have focused mainly on algorithms, testing results mostly on US data on short time-frames, yet, feature engineering, and testing findings on different markets and different time horizons, appear to be under-explored. This study thus introduces the financial context for non-experts and moves to review different models and tools in the realm of statistical learning, and deep learning. We believe that this approach will prove to be effective in financial practice to an interested reader without much prior knowledge of the finance literature. We survey the end-to-end deployment of machine learning to help readers from industry and academia to understand the peculiarities of applying these methods to equity market forecasting.

Keywords
Machine Learning, Deep Learning, Time Series Forecasting, Equity market forecasting

1. Introduction

Devising Equity markets forecasting is relevant not only for the parties directly involved in price formation (e.g., companies and investors), but also for policymakers and regulators. Central banks have shown a growing interest in modelling equities to decide macro prudential policy, assess investors' attitude towards risk, and in some cases deploy capital in the market directly. For example, according to the 30th of June 2022 13F SEC filings [1] , the Swiss National Bank owns about $11.11 billions of Apple and $7.49 billions of Microsoft.

This study is driven by the need to review machine learning techniques applied to equity markets time series forecasting problems, with the objective to provide an overview for practitioners, highlight areas under explored and warranting further research.

This study explores research from finance and computer science, from industry and academia. As such, the selection of papers for this literature review is not purely bibliometric. We aim to cover a selection of high impact publications and original contributions from the field for the critical steps of the machine learning deployment process, from pre-processing to algorithm selection. We cite papers that are now part of history of financial theory for the benefit of non-experts. Research from other domains is included where it is reasonable to assume that the related techniques could be ported to financial time series forecasting.

The contributions of this study are as follows:

- We review, with the caveat that the field is evolving at a significant speed, the main machine learning solutions for equity market forecasting deployed in terms of features and algorithms. We observe that most findings have been corroborated on specific markets or on specific periods. Data sampling and forecasting horizon are mostly daily.
- We highlight potential directions for future research, emphasising the adoption of a more data–centric approach. Ensemble methods should be further researched as architectures to leverage the peculiarities of financial data.

The rest of the paper is organised as follows: Section 2 outlines financial background knowledge for non-experts. Section 3 examines features, main algorithms deployed and related financial applications, followed by a discussion on gaps in the current knowledge. The final parts of this study are dedicated to areas of future work and how the gaps could be addressed.

CIKM'22: Workshop on Applied Machine Learning Methods for Time Series Forecasting (AMLTS), October 21, 2022, Atlanta, GA, USA
*Corresponding author.
amatuo01@mail.bbk.ac.uk (A. Matuozzo); p.yoo@bbk.ac.uk (P. D. Yoo); a.provetti@bbk.ac.uk (A. Provetti); mhykim@uow.edu.au (M. H. Kim)
0000-0002-5614-3129 (A. Matuozzo); 0000-0001-7665-8616 (P. D. Yoo); 0000-0001-9542-4110 (A. Provetti); 0000-0002-6279-5836 (M. H. Kim)

[1]Please see https://www.sec.gov/edgar.shtml

2. Framing the problem

The non-stationarity and non-linearity of financial variables are the primary attributes to estimate a forecasting model for equity markets.

2.1. The "Efficient Markets" Hypothesis

Under the Efficient Markets Hypothesis (EMH) developed by Fama [1], the sequence of price changes must be unpredictable if they fully incorporate information and expectations of market participants.

Yet, in the real world, analysing past returns and processing publicly available information are instrumental to build a forecasting model. Jagadeesh and Titman [2], showed that it is possible to generate statistically significant abnormal returns, by buying the stocks that exhibited top decile returns in the past (i.e., winners) and selling stocks in the bottom decile, losers. Malkiel [3] documented several market anomalies that increase predictability; however, these inefficiencies do not persist.

If market participants firmly believed in EMH, it would be irrational to trade, hence there would not be a financial market. Grossman and Stiglitz [4] offer the following solution: the information derived by skilled investors is not entirely reflected in the market. As a result, investment research is compensated. Bauer et al. [5] showed skill persistence amongst retail option traders: top decile performers over one year outperform individuals in the bottom decile the following year.

2.2. The "Adaptive Markets" Hypothesis

Traditional financial theories assume that market participants are rational and with the same utility function. When these assumptions are violated, behavioural biases such as dissimilar risk appetite and inconsistent probability beliefs play a significant role to account for abnormal profit opportunities. Grinblatt et al. [6] showed that individuals' personal traits and sentiment tend to affect trading behavior. Bernstein [7] points out how financial markets undergo regime shifts through an evolutionary process, which explains the market dynamics.

The Adaptive Market Hypothesis (AMH) [8] offers an alternative framework: different agents interact and adapt in response to ever changing market environment, competing to capture returns opportunities.

Machine learning, being adaptive in nature, has the set of tools to conduct research in the new paradigm: models are non-parametric in nature and have the capability to approximate complex non-linear, non-continuous functions.

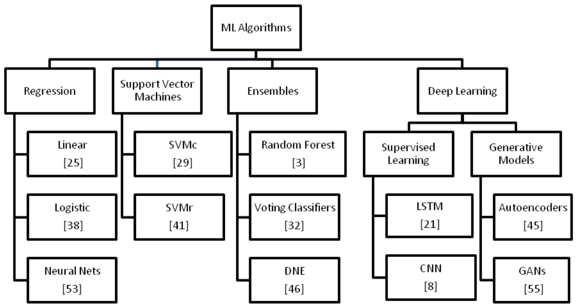

Figure 1: Machine learning methods for equity market forecasting; algorithms and reference papers.

2.3. Machine Learning in Finance

The emergence of big data gives rise to a paradigm shift in the field of financial engineering where the state-of-the-art techniques and data processing capacity are of utmost importance. Not only newsflow, economic and stock market data are widely available in copious size, however, unstructured data (e.g., geo-location), can be accessed at any time, processed in different ways. As a result, investors equipped with the latest technological advancement can make use of an extensive set of features. Machine learning algorithms do not require data transformations to make a time series stationary, being able to learn complex patterns in high dimensions.

Quian [9] showed the outperformance of several machine learning algorithms (logistic regression, Neural Network (NN), Support Vector Machines (SVM), denoising Autoencoders) compared to ARIMA. Gu et al. [10] compared ordinary least squared regression to tree-based methods and NNs on the task of predicting equity excess returns for US stocks. The authors ascribe the superior performance of machine learning techniques to the ability to accommodate a large number of predictors and learn non-linear relationships between covariates.

3. Main Algorithm Deployed and Features utilised

In this section, different machine learning algorithms and their applications to stock market time series are presented. Figure 1 depicts the ML algorithms and their related references. Here the literature is reviewed focusing on the main steps required for an end-to-end machine learning application, rather than looking, almost exclusively, at the different algorithms deployed.

3.1. Peculiarities in Financial Timeseries

Framing time series forecasting problems as regression might present an issue in the financial domain: the prediction might be very close to the actual value and yet induce the wrong action. For example, given a value of the S&P 500 of 4200, assume the 1-day forward forecast generated by two different models to be 4210 and 4150 respectively, with the actual future level being 4190: while the first model is more accurate, it will suggest entering a long position causing a loss. The second model, despite being further away from the observed value, would trigger the correct (profitable) action, establishing a short position.

In several time series classification studies, the labeling is conditional on a gating value g: an instance is labelled +1 if the response variable $y > g$, -1 otherwise. For example, Ballings et al. [11] compared the performance of ensemble methods against single classifiers on the task to group European shares based on whether they will be up at least 25% in a year. Dixon et al. [12] framed securities time series problems as multiclass attributing +1 to a positive movement, -1 to a downward direction and 0 as flat market, setting a value to determine 'no action' in order to balance the classes. Labelling examples, using an arbitrary fixed threshold does not take into account that returns are realised in a specific market regime. Defining for example a threshold of $\pm 5\%$ to determine the bounds of the 0 class, might be appropriate for a quiet market environment, however, it could be too small in a high volatility phase. Research should be conducted to scale raw data in ways that would result in predictors that carry information content related to the respective market regime.

Accuracy in time series classification problems might not reflect profitability if a model fails to capture the correct direction for large moves. For example, consider the sequence of returns +0.8%, +1%, +0.5%, -3%, and two binary classifiers (+1, -1): the first predicts four upwards movements, it is therefore 75% accurate. The second classifier outputs -1, -1, +1, -1: it is 50% accurate. If the outputs were to be fed into a trading system, the first classifier would generate a 0.7% loss while the second would yield a profit of 1.7%. An approach to mitigate this issue, suggested by industry practitioners [13] is to train a model to predict returns and then value it based on direction accuracy (also called hit ratio).

3.2. Features

Despite the potential of machine learning algorithms to extract knowledge in many dimensions, several studies do not go beyond using lags of the target series as predictors. For example, Fischer and Krauss [14] used only percentage changes of adjusted prices of the S&P 500 index constituents.

An interesting feature engineering technique that tries to capture both past observations and market regime can be seen in [15]. The authors implement a relative change transformation of the predictors in each data window: each element in a sample is subtracted and divided by the first observation of the sequence. As a result, the network will be learning from a price change that is put in the context of the short-term market environment in which it was generated. Possible avenues of research would be to devise ad hoc methods to rescale data dynamically.

The overwhelming majority of the papers surveyed in this study employs technical indicators [16] as additional features. Patel et al. [17] ran experiments on Indian equities using different representations of technical indicators: in their original form and in discrete form. The latter, called "trend-deterministic," is obtained by discretising continuous values of oscillators (e.g the Relative Strength Index, RSI) in two categories: +1, -1. Their study shows that the trend deterministic representation improves results irrespective of the algorithm.

Sentiment analysis is performed to take the pulse of the market participants' emotional state. Extracting sentiment from social media, particularly for short time frames applications has gained paramount importance given the rise of retail traders, whose decisions combined can have a major impact on price formation. Sentiment indicators can be the direct result of investors polls like for example, the American Association of Individual Investors (AAII) survey. The Investopedia portal[2] computes the Investopedia Anxiety Index (IAI) based on their readers' interests for topics such as macroeconomics, negative market and credit disruptions.

Schumaker and Chen [18] proposed a model to predict the intraday share price of the companies after a piece of news was released. They used as predictors the price forecast using a linear regression model and features derived from text analysis of news articles. Kazemian et al. [19] deployed an SVM to classify the sentiment extracted from financial news on US stocks. Bollen et al. [20] showed that the emotional state, particularly the 'Calm' dimension, of the broad population has predictive power on the daily changes in the DJIA. Othan et al. [21] proposed a model to predict the direction of Turkish stocks based exclusively on sentiment extracted from Twitter posts.

It is noteworthy that these sentiment centred techniques aim to output short term predictions. It would be interesting to extend this research over longer time horizons, using texts from commentators with domain knowledge, for example analysing blog posts from the likes of "Seeking Alpha".

Cho et al. [22] simulated an investment strategy based

[2]Please see www.investopedia.com

on the text characteristics of equity research reports published by brokerage houses. The authors deployed Part of Speech tagging to engineer features such as: numbers of nouns, adjectives, number of sentences. Classifiers, using the text extracted features, were trained on the task of recognising successful buy recommendations on Korean equities.

Fundamental analysis does much more than assessing the intrinsic value of an asset or company given that it includes both macro-econometrics and micro-econometrics approaches. As such it focuses on financial statements and macroeconomic data. Specific attention is warranted to avoid look-ahead bias: new data releases and revisions should enter the time series only when the updated information is disclosed. For example, quarterly GDP data, after the initial release, are later revised twice to get to the final figure.

Only a limited number of studies in the field adopt fundamental data; this is probably due, on one hand to the aforementioned issues, and on the other to the low frequency nature. Olson and Mossman [23] studied predictive modelling of Canadian stock excess returns and direction using as inputs 61 accounting ratios. Tsai et al. [24] studied the application of ensemble learning methods to Taiwanese stock market quarterly direction forecast. Predictors included company financial ratios and macroeconomic indicators.

Fundamental reported data are backward-looking. Alberg et al. [25] and Chauhan et al. [26] while studying neural network applications to factor investing [27] in US stocks, built a predictive model of companies' fundamental metrics. Stocks were then ranked and picked to form portfolios based on the forecasted fundamental values. These studies show that investing based on predicted company data outperforms stock selection based on reported information. An alternative solution could be using, as predictors consensus forecast figures from the financial analyst community. This approach would also allow sampling data with a frequency higher than the quarterly earnings cycle.

As summarised in Table 1, technical features can be easily retrieved from data vendors or calculated from price series. They are mostly used by practitioners to have short term timing indications, albeit with mixed results. Studying sentiment allows to read the emotional state of market participants. Sentiment data is mostly unstructured therefore, requires specific Natural Language Processing (NLP) tasks. Fundamental data are low frequency and potentially subject to the dangers of look-ahead bias. However, this kind of data reveals the state of the economy. A predictive model aiming to generate a medium- or long-term forecast should take fundamental variables in consideration.

The vast majority of research consulted for this study used only one or two types of features, as summarised in Figure 2. In light of this survey more attention should be devoted to expanding the feature space, via both feature engineering based on domain knowledge, and combining predictors from different sources of data (technical, fundamental and sentiment). The literature surveyed appears to be centred on short term forecasting as shown in Figure 3. Furthermore, tests are mostly related to US markets (Figure 4).

Feature type	Examples	Pros	Cons
Technical	RSI Momentum	Easy to compute	Mixed results
Sentiment	AAII Twitter Data	Capture emotional st.	Requires NLP
Fundamental	GDP P/E	Medium- or Long-term	Look-ahead bias

Table 1
Pros and Cons of different kinds of features.

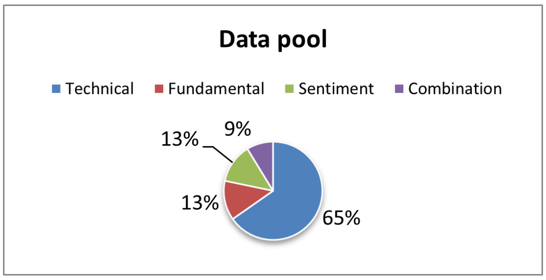

Figure 2: Surveyed papers grouped by kinds of features adopted. The combination is composed by studies using at least two sources of inputs, e.g., technical & sentiment.

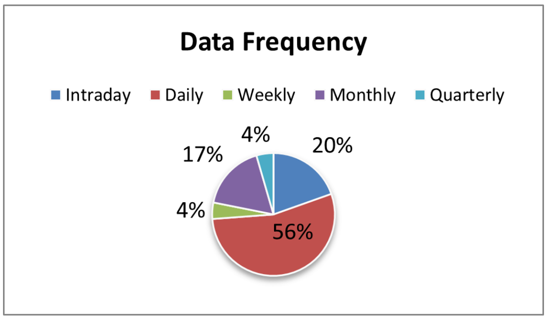

Figure 3: Surveyed papers grouped by Sampling frequency.

3.3. Ensemble Learning

Ensemble methods prescribe combing multiple algorithms to achieve better performance than inducing only the best member. Ensemble architectures are determined by how a population of diverse models is deployed and by the choice of aggregation mechanism to derive the

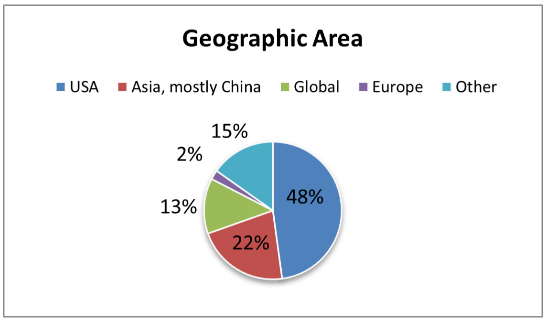

Figure 4: Markets used for experiments. Global encompasses research using securities from multiple geographic areas. Others refer to securities of different asset classes, e.g., futures on indices and commodities.

final prediction. The Random Forest (RF) algorithm [28] represents a prototypical example.

In Krauss et al. [29], a statistical arbitrage strategy is simulated after an ensemble architecture with majority voting (composed by a deep neural netwowk, gradient boosted tree and RF) has classified S&P 500 constituents according to the probability to outperform the index. Yuan et al. [30], deployed NN, SVM and RF, to classify Chinese stocks based on belonging to the top 30% of companies in generating excess return over one month.

The authors highlight an issue arising adopting standard k–folds cross validation for time series problems: future data might be used to predict past observations, overstating the performance. They evaluate algorithms using a sliding window train and test set instead. Evaluating a model on periods prior to the training set might be desirable in certain cases: for example, for stress testing purposes.

3.4. Support Vector Machines

The Support Vector Machine (SVM) [31] is a very effective algorithm able to solve non-linearly separable problems by enlarging the feature space thanks to the "kernel trick". Using the kernel functions allow to solve the problem in the original dimension without having to transform the data in a more complex space.

SVMs were compared in [32] to linear and quadratic discrimination analysis, and Elman network to forecast the weekly direction of the Nikkei 225 index. Zbikowski [33] proposed to consider volume data by plugging this information directly in the SVM formulation, multiplying the hyper-parameter representing the cost of margin violation, C, by a coefficient v based on the transaction volume in a given security over the input window. Patel et al. [34] deployed a 2 stages model with the task of predicting forward closing prices. Their architecture comprises a Support Vector Regressor (SVR) to predict the features' values, and then feed this output into a different model to produce the estimated index close price.

SVM incurs in high computational costs with a large dataset. The adoption of deep learning has grown in popularity thanks to the capability to overcome SVM scalability problems without sacrificing performance.

3.5. Deep Learning

Feed forward neural networks extend linear models by composing different functions that might carry an element of non-linearity. As a result, this is equivalent to expanding the linear problem $y = wx + b$ to $y = w(g(x, \theta)) + b$. An example of activation function is the well known Rectified Linear Unit (ReLU). In this study the terms Deep Learning and Deep Neural Network (DNN) are used for architectures encompassing more than 3 layers.

Chong et al. [35] run experiments on stocks listed on the Korean market to make intraday returns prediction (5 minutes). The authors considered a network to be deep with just three hidden layers. Over the subsequent few years, structures significantly more complex will be developed. The impact of the network depth on a financial time series classification task performance was studied in [36].

3.5.1. Recurrent Neural Networks

A Recurrent Neural Network (RNN) processes at any given stage as inputs information from that time step, and a hidden state derived from the previous one. Dixon et al. [37] show that a RNN, with linear activation, can be assimilated to a Auto Regressive time series model. The output of a given time step can be written as: $Y_t = f(W_x x_t + W_y y_{t-1} + b)$.

The Long Short Term Memory (LSTM) cell [38] overcomes the RNNs limitations (exploding or vanishing gradient). A LSTM neuron adds to the recurrent unit hidden state (short term memory), a long-term cell state controlled by different gates. LSTM network can therefore handle long sequences.

A comparison of LSTM to RF, NNs can be seen in [39]: the task in question was the prediction of the direction of 5 components of the BOVESPA index in a high frequency setting. Fischer and Krauss [14] used as feature only the return series of the past 240 observations to classify 1 day forward returns, showing the ability of LSTM cells to learn long sequences. Alonso et al. [13] tackled predicting the 1 day ahead return and the related direction of 50 constituents of the S&P 500. This study showed that increasing the number of time steps to train the model from 1 to 10 improves the performance of the LSTM.

3.5.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been particularly effective in computer vision. Convolution layers exhibit equivariance to translation; in time series data this property allows the algorithm, to detect patterns within a sample and recognise them, notwithstanding at which point in time they appear. Convolution filters over 1-dimension (1D) can be deployed to slide across time, deriving hidden patterns within samples. 1-dimensional convolution layers can be deployed as feature extractors before LSTMs, forming a CNN-LSTM architecture proposed in [15].

The dilation rate d of a convolution layer indicates the frequency of the number of elements in the input to which apply a convolution. Researchers from DeepMind conceived an architecture called WaveNet [40], able to achieve state-of-the-art performance in several text to speech tasks without the use of RNNs. It is obtained by stacking several 1D convolution layers, doubling the dilation rate at every layer. Thus, the receptive field is larger, albeit without additional parameters. The initial layers learn short term patterns, while as data progresses towards the output, sequences of longer term are extracted.

Borovykh et al. [41] adapted the Wavenet architecture, using ReLU as activation function, for multivariate financial time series forecasting: 1D dilated causal convolutions run independently for each input time series to be combined before making the prediction. Experiments were run on several financial instruments, on 1 day forward return prediction tasks. Wavenet and LSTM performed equally in terms of forecasting the direction, however, Wavenet outputs more accurate point estimation. Borjesson et al. [42] proposed a WaveNet inspired model, using as activation function the Scaled exponential Linear Unit (SeLU) for the convolution layers. They ran experiments with the goal of predicting the next day price and trend of the S&P 500.

Convolutions over two dimensions (2D) could be used to extract the most salient features and the most significant sub sequence in a sample. This approach, recently emerging in the literature, offers the advantage to learn jointly patterns within each predictor time series, and the relationship between features. This could be particularly advantageous with economic time series, where the correlation between variables changes over time, albeit in a recurrent fashion.

Gudelek et al. [43] used 2D convolutions simulating trading strategies on several Exchange Traded Fund (ETFs). Prices were differenced once to mitigate the issues of non-stationarity and transformed to be in a range between +1 and −1. The authors interpret the prediction in this range as confidence values. It would be interesting to conduct further research applying to financial time series tasks, architectures centred on 2D convolutions that have proven to be successful in other domains.

A recent stream of research leverages the success achieved in computer vision directly: time series data is transformed in order to assume the characteristics of pixel values and intensity. Cohen et al. [44] converts a time series classification problem (spotting technical patterns) as an image recognition task; Zeng et al. [45] ported a multivariate time series forecasting problem in US equities to the video prediction domain.

3.5.3. Autoencoders

Autoencoders (AE) are models conceived to replicate their inputs without supervision. The design usually prescribes an input layer, an encoding layer with a smaller number of units and a decoding layer of the same size of the input, providing a reconstructed representation. Autoencoders can be applied to financial time series prediction problems to reduce dimensions or noise. Troiano et al. [46] focused on the impact of the feature reduction stage (starting from 40 technical indicators). Denoising Autoencoders were used in [9] in order to create a latent feature representation, adding a noise component that randomly alters the raw data, forcing the algorithm to reconstruct robust inputs. Bao et al. [47] stacked 5 autoencoders connecting the final output to a LSTM network to predict the one step ahead level of several stock market indices.

3.5.4. Generative Adversarial Networks

Generative Adversarial Networks (GANs) [48] are composed by 2 competing models: the generator aims to approximate the data distribution and outputs synthetic samples; the discriminator takes either actual or synthetic data and estimates the probability that the sample in question is genuine. The generator aims to maximise the probability that the discriminator will make a mistake. Zhou et al. [49] applied GANs to make one step ahead predictions in 42 Chinese stocks within a high frequency framework. The generator is an LSTM network, while a 1D-CNN plays the role of the discriminator.

Back testing is a common approach to measure the profitability of a model against past trends in the precise order in which they occurred. GANs could be employed to provide synthetic test sets, as if engineers were forward testing algorithms devised by researchers. A recent architecture fit for this purpose is Time−series Generative Adversarial Network (TimeGAN) [50]. This framework was conceived to capture the time dependent conditional distribution of data. In addition to the unsupervised adversarial loss, the model prescribes the use of a supervised loss based on the original data.

3.5.5. Attention

Attention mechanisms, pioneered by Bahdanau et al. [51], allow a decoder to focus at each step of a sequence on the most relevant (encoded) input. The decoder computes a weighted sum of the output of the encoder; the weights are learned by an attention layer, using as inputs the encoder output concatenated with the decoder previous hidden state. These techniques, developed for neural machine translation problems, have been applied to time series forecasting: attention mechanism weights differently each time step of a sequence, this is then fed to a forecasting model to derive a prediction. Zhang et al. [52]proposed the AT-LSTM architecture: an attention mechanism using LSTM as encoder, assigns different weights to input features (time steps). Attention weighted sequences are than used as inputs into an LSTM network to output the prediction. A more complex version of this architecture can be seen in [53]: here the input series is first encoded with an LSTM and an attention layer assigns weights to the features at each time step, obtaining an attention weighted features matrix. In the next stage another LSTM based attention mechanism weights the different hidden states across time steps. The final stage prescribes an LSTM bloc to make a prediction using as inputs the output of the previous stage and the target time series. This Dual-Stage Attention RNN (DA-RNN) aims to capture the most important features while learning time dependencies.

3.6. Underexplored Areas

So far we have discussed researchers tackling the problems of non-stationarity of financial data, different market regimes and low signal to noise ratio, deploying more and more complex algorithms. Further study should be conducted focusing on feature engineering: rescaling and labelling examples considering the related market environment, therefore avoiding a fixed arbitrary threshold to define classes, would put the data in relation to the context in which they were generated.

Framing the prediction problem in terms of classification is vulnerable to not identifying large movements. On the other hand, developing a model to predict future securities prices could have the pitfall that a fairly accurate point estimate, albeit in the wrong direction, could result in a loss-making course of action.

Research could be pursued in developing alternative approaches in terms of training, for example, conceiving models learning jointly regression and classification tasks.

In contrast with many of the studies reviewed that use as input the target series and technical indicators, further research could be conducted combining features from different domains. A diverse data pool could point towards unexplored avenues in economics research.

The literature reviewed in this study adopts input time series exclusively in tabular form, however, financial professional place considerable value in extracting knowledge from the relationship and interaction amongst different variables. Graph Neural Networks (GNNs) [54] are able to extract knowledge from the interplay between different nodes in a graph, therefore could represent a novel approach, worthy of further research in our field. An exploratory study related to Japanese equities can be seen in [55]. The authors motivated the use of GNNs to leverage inter-market and inter-company relationships.

The main paradigm adopted by both industry and academia when devising a machine learning solution has been to consider a dataset as fixed, focusing on algorithm development. This survey shows that there is potential for progress, keeping the algorithm fixed, placing the data at the centre of the research process instead.

Ensemble learning approaches combining neural networks have been the winning architecture in image recognition competitions. Similar ideas could be explored, applying diversified network ensembles to financial data. Different models could be trained on different market regimes. One way to deal with a changing environment, is to constantly discard old data and retrain the model with more recent examples. Nevertheless, often, market dynamics observed in the past occur in a similar way at a later stage. Conceiving a way to use past, however, relevant information is an interesting avenue to pursue further study.

While the value of machine learning methods to equity time series forecasting has been shown in the short term, it would be interesting to test further these techniques sampling data with lower frequency. Extending findings to monthly data would graduate machine learning methods to applications beyond the realm of trading.

This survey shows how the playground for machine learning experiment in equities is the S&P 500. Few studies try to corroborate findings extending the research to different countries. European Equities in particular, emerge as an area which remains rather under-explored.

Moreover, given that different algorithms are simulated on different data, it is difficult to assess what could be considered the state of the art. It would be very fruitful for the field if, perhaps as a cooperative project between industry and academia, different financial multivariate datasets (e.g., one for US equities and one for European shares) would be engineered as standard, in order to provide a more objective common ground to conduct research.

4. Concluding Remarks

Researchers tackled the problem of equity market forecasting, initially deploying statistical time-series forecasting techniques and then experimenting with complex deep learning architectures.

In particular, LSTM has been proven to be useful in solving the vanishing/exploding gradient problem while offering the advantage of modelling non-linear time series data. Dilated convolutions, on their own, or as feature extractors, constitute an effective technique when dealing with long sequences. C2D centred architectures are an emerging method capable of extracting local knowledge about the interaction of different features.

With this review, we advocate the pursuit of research on every component of the machine learning value chain, rather than focusing exclusively on the algorithmic core. Most studies show results only on a specific market or related to a specific period, thus, the general robustness of findings could be improved. Applying machine learning to financial time series is a challenging, however, rewarding endeavour. Given the importance of the decisions based on equity market forecasts, even a small improvement in model performance can have a major impact.

References

[1] E. F. Fama, Efficient capital markets: A review of theory and empirical work, The Journal of Finance 25 (1970) 383–417. URL: http://www.jstor.org/stable/2325486.

[2] N. Jegadeesh, S. Titman, Returns to buying winners and selling losers: Implications for stock market efficiency, The Journal of Finance 48 (1993) 65–91. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1993.tb04702.x. doi:https://doi.org/10.1111/j.1540-6261.1993.tb04702.x.

[3] B. G. Malkiel, The efficient market hypothesis and its critics, The Journal of Economic Perspectives 17 (2003) 59–82. URL: http://www.jstor.org/stable/3216840.

[4] S. J. Grossman, J. E. Stiglitz, On the impossibility of informationally efficient markets, American Economic Review 72 (1982) 393–408. doi:https://doi.org/10.7916/D8765R99.

[5] R. Bauer, M. Cosemans, P. Eichholtz, Option trading and individual investor performance, Journal of Banking & Finance 33 (2009) 731–746. URL: https://www.sciencedirect.com/science/article/pii/S0378426608002720. doi:https://doi.org/10.1016/j.jbankfin.2008.11.005.

[6] M. Grinblatt, M. Keloharju, Sensation seeking, overconfidence, and trading activity, The Journal of Finance 64 (2009) 549–578. URL: http://www.jstor.org/stable/20487979.

[7] P. L. Bernstein, Why the efficient market offers hope to active management, Journal of Applied Corporate Finance 12 (1999) 129–136.

[8] A. W. Lo, Adaptive Markets: Financial Evolution at the Speed of Thought, Princeton University Press, 2017. URL: http://www.jstor.org/stable/j.ctvc77k3n.

[9] X.-Y. Qian, S. Gao, Financial series prediction: Comparison between precision of time series models and machine learning methods, 2017.

[10] S. Gu, B. T. Kelly, D. Xiu, Empirical asset pricing via machine learning, 2018. URL: https://doi.org/10.2139%2Fssrn.3281018. doi:10.2139/ssrn.3281018.

[11] M. Ballings, D. V. den Poel, N. Hespeels, R. Gryp, Evaluating multiple classifiers for stock price direction prediction, Expert Syst. Appl. 42 (2015) 7046–7056. URL: https://doi.org/10.1016/j.eswa.2015.05.013. doi:10.1016/j.eswa.2015.05.013.

[12] M. Dixon, D. Klabjan, J. H. Bang, Classiffication-based financial markets prediction using deep neural networks, CoRR abs/1603.08604 (2016). URL: http://arxiv.org/abs/1603.08604. arXiv:1603.08604.

[13] T. Guida, Big Data and Machine Learning in Quantitative Investment, Wiley, 2018.

[14] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, Eur. J. Oper. Res. 270 (2018) 654–669. URL: https://doi.org/10.1016/j.ejor.2017.11.054. doi:10.1016/j.ejor.2017.11.054.

[15] J. Eapen, D. Bein, A. Verma, Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction, in: IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019, Las Vegas, NV, USA, January 7-9, 2019, IEEE, 2019, pp. 264–270. URL: https://doi.org/10.1109/CCWC.2019.8666592. doi:10.1109/CCWC.2019.8666592.

[16] J. J. Murphy, Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications, Prentice Hall, 1986.

[17] J. Patel, S. Shah, P. Thakkar, K. Kotecha, Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques, Expert Syst. Appl. 42 (2015) 259–268. URL: https://doi.org/10.1016/j.eswa.2014.07.040. doi:10.1016/j.eswa.2014.07.040.

[18] R. P. Schumaker, H. Chen, Textual analysis of stock market prediction using breaking financial news: The azfin text system, ACM Trans. Inf. Syst. 27 (2009) 12:1–12:19. URL: https://doi.org/10.1145/1462198.1462204. doi:10.1145/1462198.

1462204.

[19] S. Kazemian, S. Zhao, G. Penn, Evaluating sentiment analysis in the context of securities trading, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers, The Association for Computer Linguistics, 2016. URL: https://doi.org/10.18653/v1/p16-1197. doi:10.18653/v1/p16-1197.

[20] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (2011) 1–8. URL: https://doi.org/10.1016/j.jocs.2010.12.007. doi:10.1016/j.jocs.2010.12.007.

[21] D. Othan, Z. H. Kilimci, M. Uysal, Financial sentiment analysis for predicting direction of stocks using bidirectional encoder representations from transformers (bert) and deep learning models, in: Proc. Int. Conf. Innov. Intell. Technol., volume 2019, 2019, pp. 30–35.

[22] P. Cho, J. H. Park, J. W. Song, Equity research report-driven investment strategy in korea using binary classification on stock price direction, IEEE Access 9 (2021) 46364–46373.

[23] D. Olson, C. Mossman, Neural network forecasts of canadian stock returns using accounting ratios, International Journal of Forecasting 19 (2003) 453–465.

[24] C. Tsai, Y. Lin, D. C. Yen, Y. Chen, Predicting stock returns by classifier ensembles, Appl. Soft Comput. 11 (2011) 2452–2459. URL: https://doi.org/10.1016/j.asoc.2010.10.001. doi:10.1016/j.asoc.2010.10.001.

[25] J. Alberg, Z. C. Lipton, Improving factor-based quantitative investing by forecasting company fundamentals, CoRR abs/1711.04837 (2017). URL: http://arxiv.org/abs/1711.04837. arXiv:1711.04837.

[26] L. Chauhan, J. Alberg, Z. C. Lipton, Uncertainty-aware lookahead factor models for quantitative investing, in: Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, volume 119 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 1489–1499. URL: http://proceedings.mlr.press/v119/chauhan20a.html.

[27] E. F. Fama, K. R. French, A five-factor asset pricing model, Journal of financial economics 116 (2015) 1–22.

[28] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32. URL: https://doi.org/10.1023/A:1010933404324. doi:10.1023/A:1010933404324.

[29] C. Krauss, X. A. Do, N. Huck, Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500, Eur. J. Oper. Res. 259 (2017) 689–702. URL: https://doi.org/10.1016/j.ejor.2016.10.031. doi:10.1016/j.ejor.2016.10.031.

[30] X. Yuan, J. Yuan, T. Jiang, Q. U. Ain, Integrated long-term stock selection models based on feature selection and machine learning algorithms for china stock market, IEEE Access 8 (2020) 22672–22685. URL: https://doi.org/10.1109/ACCESS.2020.2969293. doi:10.1109/ACCESS.2020.2969293.

[31] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297. URL: https://doi.org/10.1007/BF00994018. doi:10.1007/BF00994018.

[32] W. Huang, Y. Nakamori, S.-Y. Wang, Forecasting stock market movement direction with support vector machine, Computers & operations research 32 (2005) 2513–2522.

[33] K. Zbikowski, Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy, Expert Syst. Appl. 42 (2015) 1797–1805. URL: https://doi.org/10.1016/j.eswa.2014.10.001. doi:10.1016/j.eswa.2014.10.001.

[34] J. Patel, S. Shah, P. Thakkar, K. Kotecha, Predicting stock market index using fusion of machine learning techniques, Expert Syst. Appl. 42 (2015) 2162–2172. URL: https://doi.org/10.1016/j.eswa.2014.10.031. doi:10.1016/j.eswa.2014.10.031.

[35] E. Chong, C. Han, F. C. Park, Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, Expert Systems with Applications 83 (2017) 187–205.

[36] X. Zhong, D. Enke, Predicting the daily return direction of the stock market using hybrid machine learning algorithms, Financial Innovation 5 (2019) 1–20.

[37] M. F. Dixon, I. Halperin, P. Bilokon, Machine learning in Finance, volume 1406, Springer, 2020.

[38] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. URL: https://doi.org/10.1162/neco.1997.9.8.1735. doi:10.1162/neco.1997.9.8.1735.

[39] D. M. Nelson, A. C. Pereira, R. A. De Oliveira, Stock market's price movement prediction with lstm neural networks, in: 2017 International joint conference on neural networks (IJCNN), Ieee, 2017, pp. 1419–1426.

[40] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, K. Kavukcuoglu, Wavenet: A generative model for raw audio, in: The 9th ISCA Speech Synthesis Workshop, Sunnyvale, CA, USA, 13-15 September 2016, ISCA, 2016, p. 125. URL: http://www.isca-speech.org/archive/SSW_2016/abstracts/ssw9_DS-4_van_den_Oord.html.

[41] A. Borovykh, S. M. Bohté, C. W. Oosterlee, Condi-

tional time series forecasting with convolutional neural networks, 2017.

[42] L. Börjesson, M. Singull, Forecasting financial time series through causal and dilated convolutional neural networks, Entropy 22 (2020) 1094. URL: https://doi.org/10.3390/e22101094. doi:10.3390/e22101094.

[43] M. U. Gudelek, S. A. Boluk, A. M. Ozbayoglu, A deep learning based stock trading model with 2-d CNN trend detection, in: 2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017, Honolulu, HI, USA, November 27 - Dec. 1, 2017, IEEE, 2017, pp. 1–8. URL: https://doi.org/10.1109/SSCI.2017.8285188. doi:10.1109/SSCI.2017.8285188.

[44] N. Cohen, T. Balch, M. Veloso, Trading via image classification, in: T. Balch (Ed.), ICAIF '20: The First ACM International Conference on AI in Finance, New York, NY, USA, October 15-16, 2020, ACM, 2020, pp. 53:1–53:6. URL: https://doi.org/10.1145/3383455.3422544. doi:10.1145/3383455.3422544.

[45] Z. Zeng, T. Balch, M. Veloso, Deep video prediction for time series forecasting, in: A. Calinescu, L. Szpruch (Eds.), ICAIF'21: 2nd ACM International Conference on AI in Finance, Virtual Event, November 3 - 5, 2021, ACM, 2021, pp. 39:1–39:7. URL: https://doi.org/10.1145/3490354.3494404. doi:10.1145/3490354.3494404.

[46] L. Troiano, E. Mejuto, P. Kriplani, On feature reduction using deep learning for trend prediction in finance, CoRR abs/1704.03205 (2017). URL: http://arxiv.org/abs/1704.03205. arXiv:1704.03205.

[47] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PloS one 12 (2017) e0180944.

[48] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, 2014, pp. 2672–2680. URL: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

[49] X. Zhou, Z. Pan, G. Hu, S. Tang, C. Zhao, Stock market prediction on high-frequency data using generative adversarial nets., Mathematical Problems in Engineering (2018).

[50] J. Yoon, D. Jarrett, M. van der Schaar, Time-series generative adversarial networks, in: H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, 2019, pp. 5509–5519. URL: https://proceedings.neurips.cc/paper/2019/hash/c9efe5f26cd17ba6216bbe2a7d26d490-Abstract.html.

[51] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015. URL: http://arxiv.org/abs/1409.0473.

[52] X. Zhang, X. Liang, A. Zhiyuli, S. Zhang, R. Xu, B. Wu, At-lstm: An attention-based lstm model for financial time series prediction, in: IOP Conference Series: Materials Science and Engineering, volume 569, IOP Publishing, 2019, p. 052037.

[53] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, G. W. Cottrell, A dual-stage attention-based recurrent neural network for time series prediction, in: C. Sierra (Ed.), Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, ijcai.org, 2017, pp. 2627–2633. URL: https://doi.org/10.24963/ijcai.2017/366. doi:10.24963/ijcai.2017/366.

[54] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, AI Open 1 (2020) 57–81.

[55] D. Matsunaga, T. Suzumura, T. Takahashi, Exploring graph neural networks for stock market predictions with rolling window analysis, arXiv preprint arXiv:1909.10660 (2019).