Financial time series forecasting with machine learning techniques: A survey

Bjoern Krollner, Bond University
Bruce Vanstone, Bond University
Gavin Finnie, Bond University

Bjoern Krollner, Bruce Vanstone, Gavin Finnie

School of Information Technology, Bond University
Gold Coast, Queensland, Australia

Abstract. Stock index forecasting is vital for making informed investment decisions. This paper surveys recent literature in the domain of machine learning techniques and artificial intelligence used to forecast stock market movements. The publications are categorised according to the machine learning technique used, the forecasting timeframe, the input variables used, and the evaluation techniques employed. It is found that there is a consensus between researchers stressing the importance of stock index forecasting. Artificial Neural Networks (ANNs) are identified to be the dominant machine learning technique in this area. We conclude with possible future research directions.

1 Introduction

Stock index prediction is an important challenge in financial time series prediction. The stock market is subject to large price volatility which translates to high risks for holders of common shares. Portfolio diversification permits the reduction of company specific risk, but the 2007/2008 financial crises highlighted the enormous effects of systematic market risk on portfolio returns. Derivative trading vehicles based on stock indices provide an effective means to hedge against systematic risk. In addition, they offer profit making opportunities for speculators. Determining more effective ways of stock index prediction is important for market participants in order to make more informed and accurate investment decisions.

This paper surveys recent literature in the domain of machine learning techniques and artificial intelligence used to forecast stock market movements. The main contribution of this survey is to provide researchers with a cohesive overview of recent developments in stock index forecasting and to identify possible opportunities for future research.

2 Technologies Used

Machine learning techniques aim to automatically learn and recognise patterns in large amounts of data. There is a great variety of machine learning techniques within the literature which makes the classification difficult. This paper divides the literature into artificial neural network (ANN) based and evolutionary & optimisation based techniques.

Table 1 shows that variations of ANNs and hybrid systems are very popular in the recent literature. There is a clear trend to use established ANN models and enhance them with new training algorithms or combine ANNs with emerging technologies into hybrid systems.
Table 1: Reviewed papers classified by machine learning technique

<table>
<thead>
<tr>
<th>Technology</th>
<th>Number</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN based</td>
<td>21</td>
<td>[1], [4], [5], [8], [13], [15], [16], [20], [24], [25], [27], [31], [33], [35], [36], [37], [38], [39], [41], [43], [46]</td>
</tr>
<tr>
<td>Evolutionary &amp; optimisation techniques</td>
<td>4</td>
<td>[23], [29], [30], [45]</td>
</tr>
<tr>
<td>Multiple / hybrid</td>
<td>15</td>
<td>[2], [3], [6], [7], [11], [14], [17], [18], [21], [22], [26], [32], [34], [40], [42]</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>[9], [10], [12], [19], [28], [44]</td>
</tr>
</tbody>
</table>

Table 2: Reviewed papers classified by forecasting time-frame

<table>
<thead>
<tr>
<th>Time-frame</th>
<th>Number</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>31</td>
<td>[1], [2], [3], [4], [6], [7], [8], [9], [10], [13], [14], [17], [19], [20], [21], [22], [24], [27], [28], [31], [32], [33], [34], [35], [36], [37], [40], [41], [42], [43], [44]</td>
</tr>
<tr>
<td>Week</td>
<td>3</td>
<td>[18], [23], [43]</td>
</tr>
<tr>
<td>Month</td>
<td>3</td>
<td>[26], [38], [39]</td>
</tr>
<tr>
<td>Multiple / Other</td>
<td>9</td>
<td>[5], [11], [12], [15], [16], [25], [29], [30], [46]</td>
</tr>
</tbody>
</table>

Table 3: Input Variables

Selecting the right input variables is very important for machine learning techniques. Even the best machine learning technique can only learn from an input if there is actually some kind of correlation between input and output variable.

Table 3 shows that over 75% of the reviewed papers rely in some form on lagged index data. The most commonly used parameters are daily opening, high, low and close prices. Also used often are technical indicators which are mathematical transformations of lagged index data. The most common technical indicators found in the surveyed literature are the simple moving average (SMA), exponential moving average (EMA), relative strength index (RSI), rate of change (ROC), moving average convergence / divergence (MACD), William’s oscillator and average true range (ATR).
5 Evaluation Methods

In order to determine the effectiveness of a machine learning technique, a benchmark model is needed. A variety of evaluation methods is used in the literature. This survey categorises the evaluation models into the categories buy & hold, random walk, statistical techniques, other machines learning techniques, and no benchmark model.

Table 4 shows that the majority of authors use other machine learning techniques as a benchmark. This category consists of publications which perform a comparative analysis between two different models or use an established model and propose an improvement to that model. The proposed improved version is then compared to the original version.

Over 80% of the papers report that their model outperformed the benchmark model. However, most analysed studies do not consider real world constraints like trading costs and slippage. 31 out of 46 studies use the forecast error as an evaluation metric. This is a surprising finding since a smaller forecast error does not necessarily translate into increased trading profits.
6 Conclusion

This paper has examined recent literature in the domain of machine learning techniques and artificial intelligence used to forecast stock market movements. The reviewed papers have been categorised according to the machine learning technique used, the forecasting time-frame, the input variables used, and the evaluation techniques employed.

In regards to the employed machine learning technique, there seems to be a trend to use existing artificial neural network models which are enhanced with new training algorithms or combined with emerging technologies into hybrid systems. This finding indicates that neural network based technologies are accepted and suitable in the domain of stock index forecasting.

The surveyed forecasting time-frames revealed that the majority of publications tries to make one day ahead predictions using stock index data. It has been pointed out that for an investor it will be difficult to take advantage of this information, especially since the analysed literature does hardly examine any data of actually tradable derivatives.

Lagged index data and derived technical indicators have been identified as the most popular input parameters in the literature.

In summary, there seems to be a consensus between researchers stressing the importance of stock index forecasting and that the reported results are predominantly positive. Artificial Neural Networks (ANNs) have been identified as the dominant machine learning technique in this area.

The main finding of this survey is that there is a lack of literature examining if machine learning techniques can improve an investors’ risk-return tradeoff under real world constraints.

References


