**References**

|  |  |  |
| --- | --- | --- |
| 2013.10.24 | Kercheval and Zhang | Modeling high-frequency limit order book dynamics  with support vector machines |
| 2018.xx.xx | Ntakaris | Mid-price Prediction Based on Machine Learning Methods with Technical and Quantitative Indicators |
| 2018.xx.xx | Tsantekidis | Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks |
| 2017.xx.xx | Tsantekidis | Using Deep Learning to Detect Price Change  Indications in Financial Markets |
| 2018.08.23 | Ntakaris | Benchmark Dataset for Mid-Price Forecasting of Limit Order Book Data with Machine Learning Methods |
| 2018.08.31 | Zhang | DeepLOB: Deep Convolutional Neural Networks for Limit Order Books |
| 2018.09.19 | Nousi | Machine Learning for Forecasting Mid Price Movement using Limit Order Book Data |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**2013.10.24, Kercheval and Zhang, Modeling high-frequency limit order book dynamics with support vector machines**

We propose a machine learning framework to capture the dynamics of high frequency limit order books in financial equity markets and automate real-time prediction of metrics such as mid-price movement and price spread crossing. By characterizing each entry in a limit order book with a vector of attributes such as price and volume at different levels, the proposed framework builds a learning model for each metric with the help of multi-class support vector machines (SVMs). Experiments with real data establish that features selected by the proposed framework are predictive for short term price movement forecasts.

**2017.xx.xx, Ntakaris, Mid-price Prediction Based on Machine Learning Methods with Technical and Quantitative Indicators**

Stock price prediction is a challenging task, but machine learning methods have recently been used successfully for this purpose. In this paper, we extract over 270 hand-crafted features (factors) inspired by technical and quantitative analysis and tested their validity on short-term mid-price movement prediction. We focus on a wrapper feature selection method using entropy, least-mean squares, and linear discriminant analysis. We also build a quantitative feature based on adaptive logistic regression for online learning, which is constantly selected first among the majority of the proposed feature selection methods. This study examines the best combination of features using high frequency limit order book data from Nasdaq Nordic. Our results suggest that sorting methods and classifiers can be used in such a way that one can reach the best performance with a combination of only very few informative features. This paper opens avenues for developing more advanced features combined with more sophisticated feature selection methods. It also provides helpful insight to market makers and traders in general by providing useful results that can be used to gain an information edge in trading.

**2017.xx.xx Tsantekidis, Using Deep Learning to Detect Price Change Indications in Financial Markets**

Forecasting financial time-series has long been among the most challenging problems in financial market analysis. In order to recognize the correct circumstances to enter or exit the markets investors usually employ statistical models (or even simple qualitative methods). However, the inherently noisy and stochastic nature of markets severely limits the forecasting accuracy of the used models. The introduction of electronic trading and the availability of large amounts of data allow for developing novel machine learning techniques that address some of the difficulties faced by the aforementioned methods. In this work we propose a deep learning methodology, based on recurrent neural networks, that can be used for predicting future price movements from large-scale high-frequency time series data on Limit Order Books. The proposed method is evaluated using a large-scale dataset of limit order book events.

**2018.xx.xx, Tsantekedis, Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks**

In today’s financial markets, where most trades are performed in their entirety by electronic means and the largest fraction of them is completely automated, an opportunity has risen from analyzing this vast amount of transactions. Since all the transactions are recorded in great detail, investors can analyze all the generated data and detect repeated patterns of the price movements. Being able to detect them in advance, allows them to take profitable positions or avoid anomalous events in the financial markets. In this work we proposed a deep learning methodology, based on Convolutional Neural Networks (CNNs), that predicts the price movements of stocks, using as input large-scale, high-frequency time-series derived from the order book of financial exchanges. The dataset that we use contains more than 4 million limit order events and our comparison with other methods, like Multilayer Neural Networks and Support Vector Machines, shows that CNNs are better suited for this kind of task.

**2018.08.23, Ntakaris, Benchmark Dataset for Mid-Price Forecasting of Limit Order Book Data with Machine Learning Methods**

Managing the prediction of metrics in high-frequency metrics in financial markets is a challenging task. An efficient way is by monitoring the dynamics of a limit order book to identify the information edge. This paper describes the first publicly available benchmark dataset of high-frequency limit order markets for mid-price prediction. We extracted normalized data representations of time series data for five stocks from the NASDAQ Nordic stock market for a time period of ten consecutive days, leading to a dataset of 4,000,000 time series samples in total. A day-based anchored cross-validation experimental protocol is also provided that can be used as a benchmark for comparing the performance of state-of-the-art methodologies. Performance of baseline approaches are also provided to facilitate experimental comparisons. We expect that such a largescale dataset can serve as a testbed for devising novel solutions of expert systems for high-frequency limit order book data analysis.

**2018.08.31, Zhang, DeepLOB: Deep Convolutional Neural Networks for Limit Order Books**

We develop a large-scale deep learning model to predict price movements from limit order book (LOB) data of cash equities. The architecture utilises convolutional filters to capture the spatial structure of the limit order books as well as LSTM modules to capture longer time dependencies. The model is trained using electronic market quotes from the London Stock Exchange. Our model delivers a remarkably stable out-of-sample prediction accuracy for a variety of instruments and outperforms existing methods such as Support Vector Machines, standard Multilayer Perceptrons, as well as other previously proposed convolutional neural network (CNN) architectures. The results obtained lead to good profits in a simple trading simulation, especially when compared with the baseline models. Importantly, our model translates well to instruments which were not part of the training set, indicating the model’s ability to extract universal features. In order to better understand these features and to go beyond a “black box” model, we perform a sensitivity analysis to understand the rationale behind the model predictions and reveal the components of LOBs that are most relevant. The ability to extract robust features which translate well to other instruments is an important property of our model which has many other applications.

**2018.09.19, Nousi, Machine Learning for Forecasting Mid Price Movement using Limit Order Book Data**

Forecasting the movements of stock prices is one the most challenging problems in financial markets analysis. In this paper, we use Machine Learning (ML) algorithms for the prediction of future price movements using limit order book data. Two diff erent sets of features are combined and evaluated: handcrafted features based on the raw order book data and features extracted by ML algorithms, resulting in feature vectors with highly variant dimensionalities. Three classifiers are evaluated using combinations of these sets of features on two different evaluation setups and three prediction scenarios. Even though the large scale and high frequency nature of the limit order book poses several challenges, the scope of the conducted experiments and the signicance of the experimental results indicate that Machine Learning highly befits this task carving the path towards future research in this field.