Enhancing Financial Time Series Forecasting: A Comparative Study of Discrete Wavelet Transform and LSTM Models for Selected Global Indices

Krzysztof Ziółkowski[1] Paweł Wieczyński[2]

[1]WSB Merito University Gdansk
kziolkowski@wsb.gda.pl

[2]Gdańsk University of Technology
pawel.wieczynski@pg.edu.pl

Inwestycje Alternatywne, 2024

Background

Forecasting stock index prices is a complex process that may involve various methods and techniques:

- Technical analysis
- Fundamental analysis
- Econometric models
- Artificial intelligence and machine learning
- Sentiment analysis
- Leading indicators

Objectives

This study underscore the potential and challenges of using *machine learning* (ML) and *deep learning* (DL) models for stock indices market prediction.

This study aims to analyze the predictive capabilities using the:

- *discrete wavelet transform* (DWT) and *long short-term memory* (LSTM) methods
- raw LSTM models as a benchmark
- ARIMA models as a benchmark.

Dataset

Indices analyzed: WIG, WIG20, and S&P 500 (closing prices, source stooq.pl).

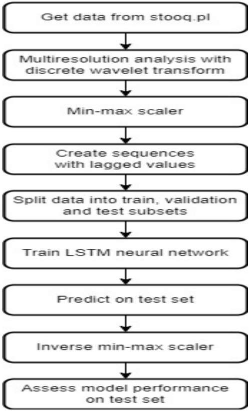
Data period: Daily observations over 15 years, from July 1, 2009, to June 30, 2024.

Total observations:

- WIG/WIG20: 3753
- S&P 500: 3774

Methodology

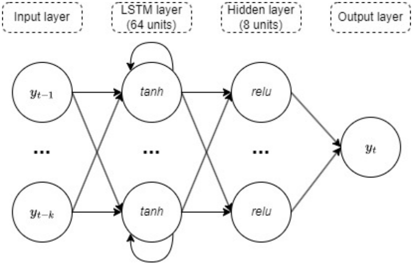
- Wavelet bases used: *haar*, *db1*, *db2*, *db3*, *db4*, *db5*, *db15*, *db30*.
- Levels of wavelet decomposition: 1-8.
- Data splitting: 70% for training, 15% for validation, 15% for testing.
- Lag intervals (k): k = 5 days (weekly lag), k = 21 days (monthly lag), k = 63 days (quarterly lag).



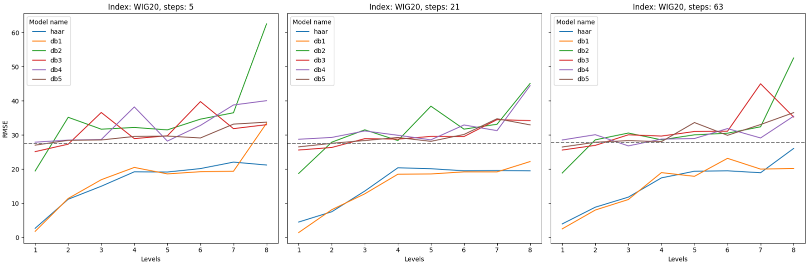
Network architecture

Artificial neural network with the following architecture was implemented:

- Input layer - shape determined by time-series length and number of regressors.
- LSTM layer with 64 units and tanh activation.
- Hidden layer with 8 units and ReLU activation.
- Single unit output layer with linear activation.



Key results


Index: WIG20, steps: 5
Index: WIG20, steps: 21
Index: WIG20, steps: 63
Model name
haar
db1
db2
db3
db4
db5
RMSE
Levels

Findings and challenges

The DWT+LSTM combination provides more accurate predictions than LSTM alone or ARIMA.

Challenges:

- Higher levels of MRA with wavelets can lead to boundary effects, increasing errors.
- Increased computational demand for recalibrating models when new data is available.
- The choice of wavelet type and network architecture is critical for achieving optimal results.

Future work

Future work:

- Explore other wavelet types beyond Daubechies for potential improvement.
- Incorporate feature selection methods like XGBoost for further refining prediction models - alike *Vuong et al. (2022)*.
- Investigate the potential for real-time forecasting in investment strategies.