Quantitative Trading Strategies using Deep Learning: Pairs Trading
Simerjot Kaur (sk3391)

Introduction

- Pairs Trading is a widely used quantitative trading strategy.
- It involves two key aspects:
  - Identify co-integrated pairs of financial assets.
  - **Forecast spread** between the paired assets.
- Forecasted spread is then used to make trading decisions.
- However, forecasting spread between the paired assets is a non-trivial exercise, and requires deep domain expertise to create predictive models.

Problem Statement

- Current state-of-the-art spread forecasting methods include statistical methods such as Kalman Filters, ARIMA models, and VAR.
- We aim to explore LSTM/RNNs as a means to forecast spread.
- Formally, given a time series of spread ($S_1 \ldots\ldots S_t$) predict the value of spread 1-time step (S_{t+1}), 5-time step (S_{t+5}), and 15-time steps (S_{t+15}) in the future.

Dataset and Features

- Dataset Used :
  - 10 year (10.01.2008 - 10.01.2018) daily 'close', 'open', 'high' ,'low' price and 'volume traded' for various exchange traded funds (ETFs) were extracted from Yahoo! Finance and used as dataset.

Snapshot : Sample Data

	Close	High	Low	Open	Volume
2008-10-01	35.860001	37.200001	34.939999	35.410000	569600.0
2008-10-02	33.700001	34.990002	33.529999	34.860001	751900.0
2008-10-03	33.299999	35.080002	33.200001	34.099998	853100.0

- Data Preprocessing Steps
  - Data Cleaning
    - Pre-processed data to discard any dead ETFs
  - Wavelet Denoising
    - Multi-resolution discrete wavelet transform to perform de-noising.
  - Feature Extraction:
    - Extracted various technical indicators (TI) such as 'Momentum', 'Volatility' and 'Trend' indicators from raw time series data that can be used as feature inputs for LSTM .

End-to-end System Pipeline
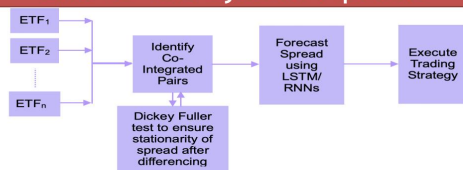
Proposed Deep Learning Model
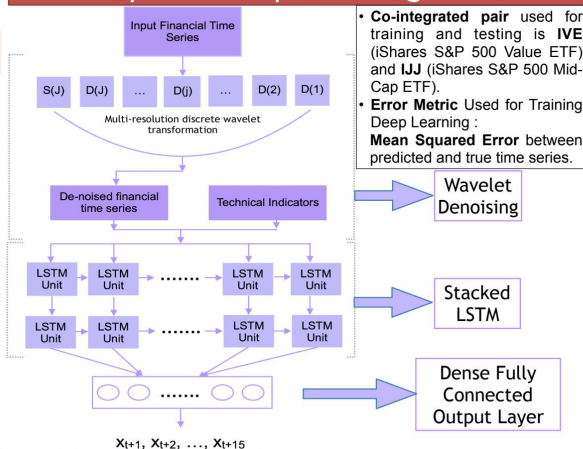
- **Co-integrated pair** used for training and testing is **IVE** (iShares S&P 500 Value ETF) and **IJJ** (iShares S&P 500 Mid-Cap ETF).
- **Error Metric** Used for Training Deep Learning : **Mean Squared Error** between predicted and true time series.

Discussions and Conclusions

- Deep learning based methods outperform statistical techniques in predicting spread for pairs trading.
- Additional features, like technical indicators, and Wavelet de-noising that eliminates market data noise, improve performance significantly.
- Prediction of S_{t+1} is performing significantly better than [1]. Additionally, we have been able to predict spread at 15-time steps ahead.
- Profits made using deep learning based methods are significantly more than the traditional statistical techniques.

References

[1] Wei Bao, Jun Yue2, Yulei Rao. A deep learning framework for financial time series using stacked autoencoders and long short term memory.

Results

Model Comparison: Training Size: 95%, Dev Set: 5%, Test Set 5%, Total Data: 10 Years

| *TI = Technical Indicators	1-Time Step Ahead Prediction		5-Time Step Ahead Prediction		15-Time Step Ahead Prediction	
	Training Error (MSE)	Test Error (MSE)	Training Error (MSE)	Test Error (MSE)	Training Error (MSE)	Test Error (MSE)
Statistical Methods						
Kalman Filter	N/A	1.9347	N/A	N/A	N/A	N/A
VAR	N/A	0.5792	N/A	0.8156	N/A	1.1932
ARIMA	N/A	0.5797	N/A	0.8182	N/A	1.1936
Facebook Prophet	N/A	2.1973	N/A	2.2625	N/A	2.3586
Deep Learning Techniques						
LSTM	0.1599	0.4342	0.2586	0.7112	0.4293	1.009
LSTM + TI	0.1499	0.4183	0.2213	0.6682	0.2451	0.9439
Stacked LSTM + TI	0.1449	0.4055	0.2197	0.6552	0.2439	0.9342
Stacked LSTM + Wavelet Denoising + TI	0.1394	0.4024	0.2174	0.6436	0.2427	0.8945
Stacked LSTM + Wavelet Denoising + TI + Optimized Regularization	0.1392	0.4002	0.2137	0.5997	0.2305	0.8091

1-Time Step Prediction: Actual vs Predicted Spread
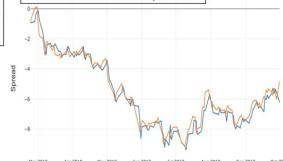
15-Time Step Prediction: Actual vs Predicted Spread
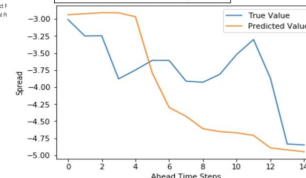
Hyperparameter Tuning
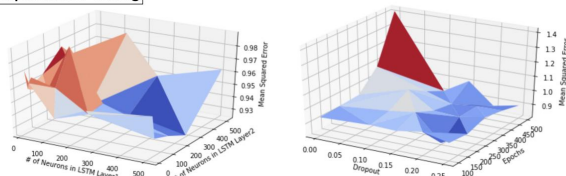
Profits made using Pairs Trading Strategy

	Ground Truth	Kalman Filter	VAR	ARIMA	Facebook Prophet	Deep Learning Method
Profit Made	$91,965	$29,960	$64,448	$64,580	$49,317	$75,689

Future Work

- Add more information from other sources, such as current news, twitter feeds etc. via NLP to help improve long term predictive power of model.
- Replace current statistical methods with deep learning methods for identifying co-integrated pairs.

YouTube Link: https://youtu.be/KMWc5jAcnOA