Trend Prediction for S&P 500 ETF Trust using LSTM Model

Yulin Zhao

This report focuses on trend prediction of S&P 500 ETF Trust (SPY) using long short-term memory (LSTM) model.

Table of Contents

1. Feature Engineering

Features fall into five categories: technical indicators, time-based factors, rolling statistics, fundamental factors, and factors based on cointegration analysis. Features are listed as follows:

Category	Feature	Notation
Technical Indicators	Moving Average (period 5,10,20,65)	MA
	Exponential Moving Average (period 5,10,20,65)	EMA
	Bollinger Bands	UBol, LBol
	Relative Strength Index	RSI
	Moving Average Convergence Divergence	MACD
	Average True Range	ATR
Time-Based Factors	Lagged Values of Return (lagged period 1,2,3,4,5,6,7)	Lagret
	The ratio of High Price to Low Price	HC
	The ratio of Close Price to Open Price	CO
	Price Change over Period (Period 5,10,20,65)	PCHG
	Log Return	RET
Rolling Statistics	Rolling Standard Deviation (rolling 5,10,20,65)	STD
	Rolling Mean (rolling 5,10,20,65)	RET
Fundamental Factors	Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity	DGS10
	5-Year Forward Inflation Expectation Rate	T5YIFR
	CBOE Volatility Index	VIX
Factors Based on Cointegration Analysis	Invesco QQQ Trust, tracking Nasdaq-100 Index	QQQ
	SPDR Dow Jones Industrial Average ETF Trust	DIA

		iPath Series B S&P 500 VIX Short-Term Futures ETN, tracking the CBOE Volatility Index (VIX), which measures market volatility	VXX

1.1 Technical Indicators, Time-based Factors, Rolling Statistics

As for technical indicators, time-based factors, and rolling statistics, there are total of 37 features. The date range is from 2020-03-11 to 2024-12-03. The feature shape is 1192*37. By using correlation analysis, there are 18 features kept when the threshold is 0.8, and 12 features left when the threshold is 0.6. It remains 11 features when the date range is updated, from 2020-01-08 to 2025-01-03. The following results show that the key factors are lag terms of asset returns, even though the coefficient is from 0.8 to 0.6. In the model building sessions, the report also does the binary transformation for the lag term of returns.

	Correlation=0.8	Correlation= 0.6
CO	√	√
HC	√	√
RET	√	
Lagret1	√	√
Lagret2	√	√
Lagret3	√	√
Lagret4	√	√
Lagret5	√	√
Lagret6	√	√
Lagret7	√	√
PCHG5	√	√
MA5	√	√
EMA5		
RET5		
STD5	√	
PCHG10	√	
MA10		
EMA10		
RET10		

STD10		
PCHG20	√	
MA20		
EMA20		
RET20		
STD20		
PCHG65	√	√
MA65		
EMA65		
RET65		
STD65	√	
Ubol		
Lbol		
RSI	√	
MACD		
MACD_signal		
true_range		
ATR14		

Notation: PCHG65 is excluded from the analysis on the updated data range.


1.2 Fundamental Factors

For fundamental factors, this report chooses the 10-year market yield on U.S. treasury securities, since it has the daily data, thereby can be matched to the close price of S&P on trading day. Also, the report uses a 5-year forward inflation expectation rate rather than the CPI index, which only has monthly data. The same situations are for the unemployment rate and federal funds rates. In addition, it is reasonable that the forward inflation expectation rate as the perspective indicator is used for prediction.

For missing values of the inflation rate and the market yield, the report uses the average of adjacent values to fill them.

Moreover, based on Augmented Dickey-Fuller unit root test, the close price of VIX and forward inflation expectation rate are stationary, whereas the market yield on U.S. treasury securities has unit root. Therefore, the report uses the first difference of market yield data.

A step further to test the unit root of residuals derived from linear regression of the closing price of S&P and 10-year market yield on U.S. treasury securities. The result shows it is nonstationary. Therefore, there is no cointegration.

Factor	Unit Root Test	Result
DGS10	Unit root	Use first-difference
T5YIFR	Stationary	
VIX	Stationary	
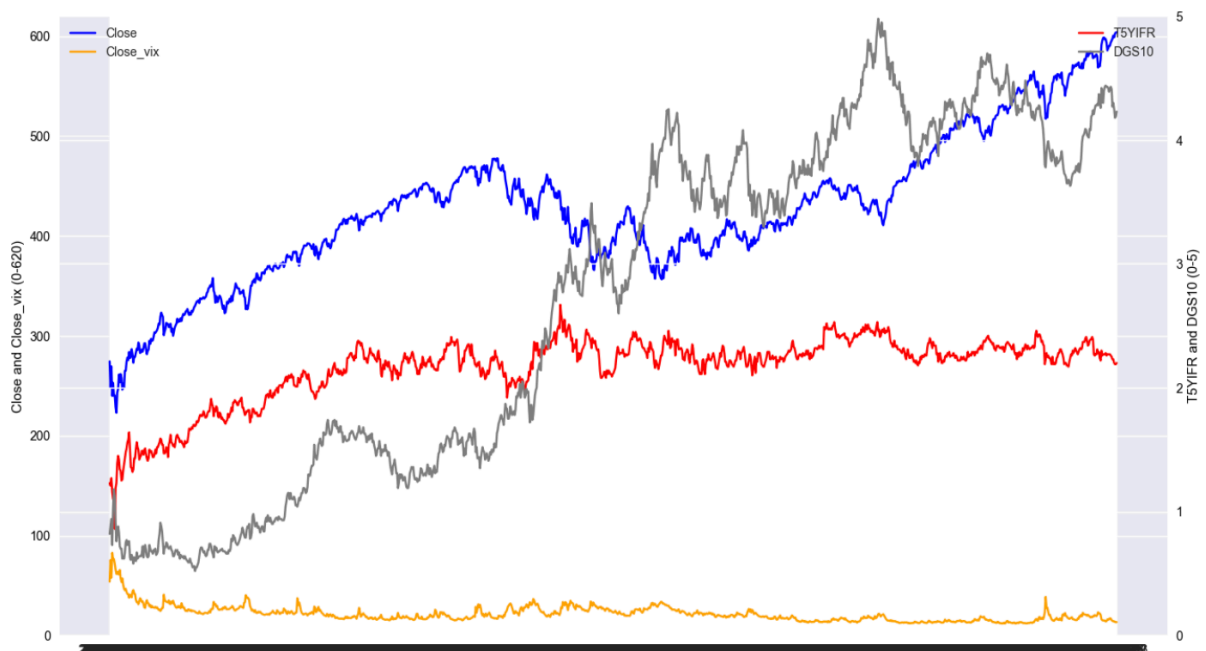
The test results are summarized as follows:

Augmented Dicky-Fuller Test Unit Root Test

Close Price of VIX

	Estimation	Std. Error	t value	Pr(>|t|)	
Intercept	0.852825	0.158603	5.377	9.11E-08	***
z.lag.1	-0.042447	0.006897	-6.155	1.03E-09	***
z.diff.lag	-0.248387	0.026372	-9.419	<2e-16	***

Market Yield (DGS10)

	Estimation	Std. Error	t value	Pr(>|t|)

	Estimation	Std. Error	t value	Pr(>|t|)	
Intercept	0.006146	0.003961	1.551	0.121	
z.lag.1	-0.001227	0.001301	-0.943	0.346	
z.diff.lag	-0.002769	0.029012	-0.095	0.924	

Forward Inflation Expectation Rate(T5YIFR)

	Estimation	Std. Error	t value	Pr(>|t|)	
Intercept	0.035648	0.009357	3.81	0.000146	***
z.lag.1	-0.016098	0.004301	-3.743	0.000191	***
z.diff.lag	-0.066414	0.028804	-2.306	0.021298	*

Residuals from Linear Regression of Closing price and Market Yield

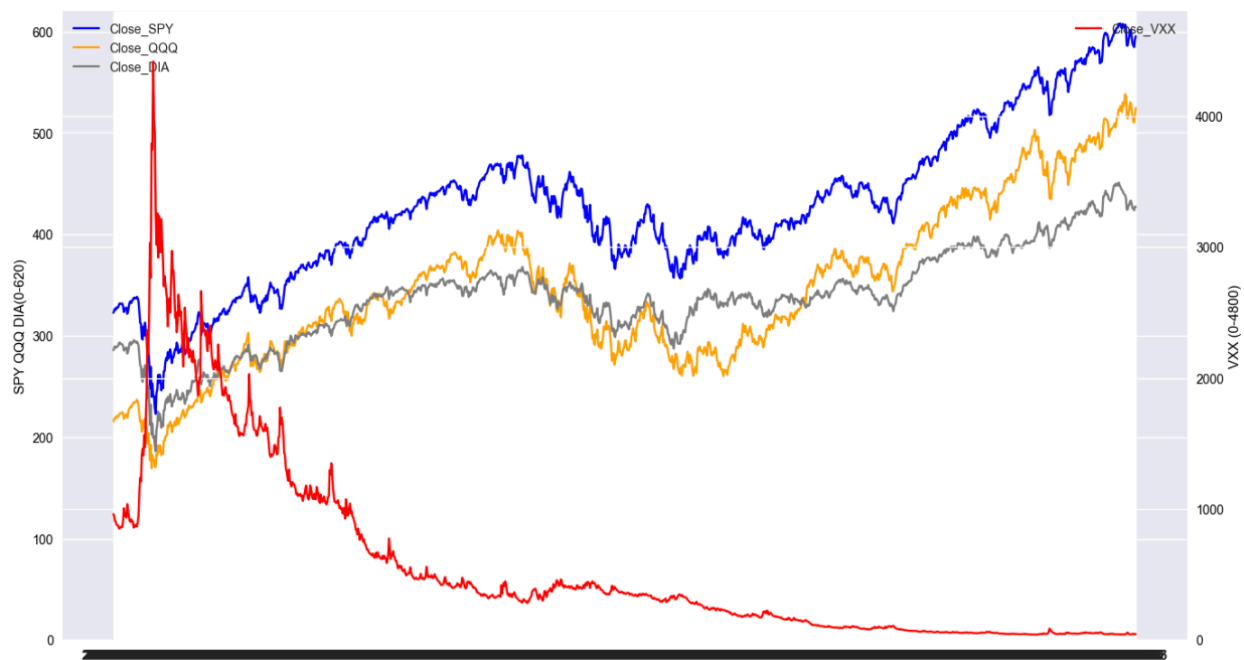|	Estimation	Std. Error	t value	Pr(>|t|)
z.lag.1	-0.003790	0.002657	-1.426	0.154
z.diff.lag	-0.013855	0.028667	-0.483	0.629

1.3 Cointegration Analysis

For cointegration analysis, unit root tests show that QQQ, DIA, and VXX all have unit roots. Residuals from the linear regression of SPY and QQQ is stationary, as well as residuals from the linear regression of SPY and DIA. However, residuals from the linear regression of SPY and VXX has unit root. Therefore, VXX is not considered in further test.

Based on error correction model (ECM), the first difference of QQQ and the lag term of residuals are statistically significant. Also, the first difference of DIA and lag term of residuals are of significance. However, according to the results from argument error correction model, lag term of first difference of SPY is not significant for either QQQ or DIA.

Finally, four factors are included in the model, the first difference of QQQ, and the lag term of its residuals, the first difference of DIA, and the lag term of its residuals.

Factors		
First difference of QQQ	✓	
Residual lag (SPY~QQQ)	✓	
First difference of DIA	✓	
Residual lag (SPY~DIA)	✓	
Lag term of first difference of SPY close price	✗	No statistical significance
VXX	✗	No cointegrated relationship

The results of the cointegration analysis are summarized as follows:

Augmented Dicky-Fuller Test Unit Root Test

QQQ

	Estimation	Std. Error	t value	Pr(>|t|)	
Intercept	0.5538394	0.6361097	0.871	0.3841	
z.lag.1	-0.0008575	0.0017963	-0.477	0.6332	
z.diff.lag	-0.0526885	0.0282508	-1.865	0.0624	

DIA

	Estimation	Std. Error	t value	Pr(>|t|)	
Intercept	0.684166	0.752895	0.909	0.363676	
z.lag.1	-0.001668	0.002214	-0.753	0.451311	
z.diff.lag	-0.096779	0.028145	-3.439	0.000604	***

VXX

	Estimation	Std. Error	t value	Pr(>|t|)	
Intercept	1.207396	2.40237	0.503	0.615	
z.lag.1	-0.003551	0.002626	-1.352	0.177	
z.diff.lag	-0.171303	0.027836	-6.154	1.01E-09	***

Residuals from SPY~QQQ

	Estimation	Std. Error	t value	Pr(>|t|)

	Estimation	Std. Error	t value	Pr(>|t|)	
z.lag.1	-0.008925	0.003593	-2.484	0.0131	*
z.diff.lag	-0.010729	0.028229	-0.38	0.7039	

Residuals from SPY~DIA

	Estimation	Std. Error	t value	Pr(>|t|)	
z.lag.1	-0.01473	0.00501	-2.941	0.00333	**
z.diff.lag	-0.04383	0.02824	-1.552	0.12094	

Residuals from SPY~VXX

	Estimation	Std. Error	t value	Pr(>|t|)	
z.lag.1	-0.004318	0.003058	-1.412	0.158	
z.diff.lag	0.037487	0.028272	1.326	0.185	

Error Correction Model

	Estimation	Std. Error	t value	Pr(>|t|)	
QQQ_diff	0.898009	0.010761	83.447	< 2e-16	***
residualQQQ_lag	-0.009739	0.003532	-2.758	0.00591	**
	Estimation	Std. Error	t value	Pr(>|t|)	
DIA_diff	1.217613	0.014109	86.302	<2e-16	***
residualDIA_lag	-0.009594	0.003919	-2.448	0.0145	*

Augmented Error Correction Model

	Estimation	Std. Error	t value	Pr(>|t|)	
SPY_diff_lag	-0.003712	0.011058	-0.336	0.73714	
QQQ_diff	0.897667	0.010802	83.102	< 2e-16	***
residualQQQ_lag	-0.009883	0.003544	-2.788	0.00538	**
	Estimation	Std. Error	t value	Pr(>|t|)	
SPY_diff_lag	0.003101	0.010766	0.288	0.773	
DIA_diff	1.21807	0.014169	85.97	<2e-16	***
residualDIA_lag	-0.009676	0.003932	-2.461	0.014	*

1.4 Dependent Variable

The dependent variable is labeled 0,1 according to the percentage change of the closing price one-day forward. For the data range from 2020-03-11 to 2024-12-03, the threshold is set to 0.1%, and 0:1 is 598:594, which is pretty close to the balance situation. Therefore, it does not apply the imbalance weight.

However, when the data set is updated from 2020-01-08 to 2025-01-03, 0:1 is 624:622. Imbalance weight is applied, and the weights are 0: 0.99053627, 1: 1.00964630.

2. Building Model and Fine-tuning Hyperparameters

The report does trials on five sets of features:

1) 18 features, correlation threshold = 0.8, no fundamental factors, no factors based on cointegration analysis

2) 12 features, correlation threshold = 0.6, no fundamental factors, no factors based on cointegration analysis

3) 12 features, correlation threshold = 0.6, binary transformation for the lag term of returns, no fundamental factors, no factors based on cointegration analysis

4) 3 features, fundamental factors as features

5) 18 features, correlation threshold = 0.6, fundamental factors and factors based on cointegration analysis included

The results of fine-tuning hyperparameters are listed as follows:

2.1 Base Model

Features No.: 18
Correlation Threshold = 0.8
Optimizer: Adam
Early Stopping: validation accuracy, max

Tuning Method	Random Search	Hyperband	Bayesian
Unit1	8	16	32
Unit2	20	32	16
Unit3	28	28	32
Dropout Rate	0.2	0.3	0.3
Learning Rate	3.714E-05	3.631E-03	2.851E-04
Sequence Length	5	20	130
Activation Function	eLU	eLU	eLU
Best Val_accuracy	0.526636223	0.543379009	0.526636223

Features No.: 12
Correlation Threshold = 0.6
Optimizer: Adam

Early Stopping: validation accuracy, max

Tuning Method	Random Search	Hyperband	Bayesian
Unit1	16	36	64
Unit2	20	4	32
Unit3	24	8	32
Dropout Rate	0.4	0.4	0.4
Learning Rate	1.81E-04	8.89E-04	1.00E-02
Sequence Length	20	65	130
Activation Function	eLU	ReLU	eLU
Best Val_accuracy	0.529680371	0.57077628	0.532724500

Optimizer: RMSprop

Early Stopping: validation accuracy, max

Tuning Method	Random Search	Hyperband	Bayesian
Unit1	16	36	8
Unit2	20	16	12
Unit3	12	28	28
Dropout Rate	0.1	0	0.2
Learning Rate	5.22E-05	4.08E-04	1.88E-03
Sequence Length	20	20	65
Activation Function	eLU	eLU	eLU
Rho	0.931873347	0.90663511	0.891656157
Best Val_accuracy	0.528158287	0.56164384	0.528158307

When adjusting correlation coefficient threshold from 0.8 to 0.6, the model achieves results by applying two optimization methods, Adam and RMSprop. However, the results show that Adam is a glimmer of better. Therefore, further model fine-tuning use Adam as the optimization method. In addition, comparing the model with correlation = 0.6 to that with correlation = 0.8, the result of the model with correlation = 0.6 is subtle better.

Features No.: 16
Correlation Threshold = 0.6
Binary transformation for lag return
Optimizer: Adam
Early Stopping: validation accuracy, max

Tuning Method	Random Search	Hyperband	Bayesian

		44	56	56
Unit1		44	56	56
Unit2		4	28	20
Unit3		12	20	20
Dropout Rate		0.3	0.1	0.2
Learning Rate		9.098E-05	3.717E-04	3.030E-04
Sequence Length		5	20	20
Activation Function		eLU	ReLU	ReLU
Best Val_accuracy		0.529680371	0.557077646	0.525114159

When transferring series of the lag term of returns to binary values, the performance of the model does not show the significant improvement.

2.2 Model with Fundamental Factors

Features No.: 3
Fundamental factors as features
Optimizer: Adam
Early Stopping: validation accuracy, max

Tuning Method	Random Search	Hyperband	Bayesian
Unit1	128	100	36
Unit2	4	4	20
Dropout Rate	0	0.2	0.2
Learning Rate	8.517E-03	5.785E-05	5.783E-03
Sequence Length	5	20	65
Activation Function	eLU	eLU	ReLU
Best Val_accuracy	0.522070030	0.529680371	0.53272452
Tuning Method	Random Search	Hyperband	Bayesian
Unit1	16	116	88
Unit2	44	52	56
Unit3	36	44	12
Unit4	24	4	24
Dropout Rate	0.4	0.3	0.1
Learning Rate	1.568E-05	3.683E-05	1.631E-03
Sequence Length	65	20	20
Activation Function	ReLU	eLU	ReLU
Best Val_accuracy	0.528158287	0.520547926	0.517503798

The report shows the model result with isolated fundamental factors as features, since these factors have clear and straightforward economic interpretations for the prediction of S&P 500 ETF trust. The results show the performance which is comparable to those from previous models. In addition, the number of layers does not have remarkable matters.

2.3 Model with Features of All Categories

Finally, the model uses all the features stated above to do the trend prediction. For fine-tuning the hyperparameters of this model, it just applies Bayesian Optimization. Given the time-consuming and the tuning result from previous trials, Bayesian Optimization is the reasonable choice.

The results are showed as follows:

Features No.: 18
Correlation Threshold = 0.6
include fundamental factors
include factors from cointegration analysis
Optimizer: Adam
Early Stopping: validation accuracy, max
Threshold for label = 0.1%

Tuning Method	Bayesian
Unit1	56
Unit2	44
Unit3	4
Unit4	12
Dropout Rate	0.3
Learning Rate	9.098E-05
Sequence Length	5
Activation Function	elu
Best Val_accuracy	0.530172408

Threshold for label = 0.09%

Tuning Method	Bayesian
Unit1	32
Unit2	12
Unit3	48

Unit4	24
Dropout Rate	0.3
Learning Rate	8.890E-05
Sequence Length	65
Activation Function	elu
Best Val_accuracy	0.524425268

The trials set a different threshold for the label, since the indicator of precision, recall and F1-score are quite sensitive to the nuance of this threshold.

3. Model Evaluation

When the threshold for the label is 0.1%, the model only identifies 31% of actual class 1 cases, whereas it can identify 100% of actual class 1 cases, when the threshold for the label becomes 0.09%. But it totally cannot identify class 0 cases.
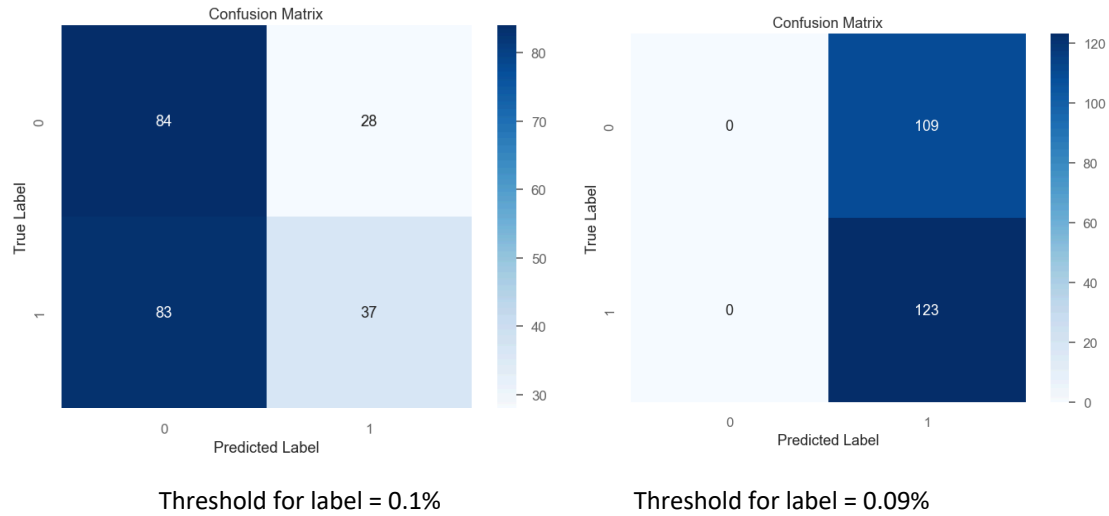
Classification Report

Threshold for label = 0.1%

	Precision	Recall	F1-score
0	0.5	0.75	0.6
1	0.57	0.31	0.4

Threshold for label = 0.09%

	Precision	Recall	F1-score
0	0	0	0
1	0.53	1	0.69

	Confusion Matrix	
Threshold for label = 0.1%		

	Confusion Matrix	
Threshold for label = 0.09%		

4. Further Work

Based on the model fine-tuning above, results are unsatisfying. One consideration is to add more relevant features, which better distinguish classes.

4.1 Wavelet Transform

This report does not use Self-organizing Maps (SOM) to do dimensionality reduction, since SOM cannot capture temporal information. It treats each time series data point as an independent input vector, so the temporal sequence or order of the data points is not considered during the mapping process. To solve the distortions, Dynamic Time Warping (DTW) can be used within the SOM framework. Also, advanced algorithms SOMTimeS can be used to optimize the process [Jav+24].

However, some research focus on the wavelet transform. It is an extension of the Fourier transform, which is its time domain and frequency domain transform, and is designed to isolate the periodicity [Tho21, Lia+19]. Wavelet transform may decompose a signal directly according to the frequency and represent it in the frequency domain distribution state in the time domain, therefore both time and frequency information of the signal are retained [Rhi+19].

13

Wavelet transform has the capability of capturing long-term movements and high-frequency details, which is useful to deal with non-stationary and complex functions [Sch02]. The low frequency part of wavelet decomposition reports the general tendency of the series, whereas the high-frequency part reflects the short-term stochastic disturbance [Tan+21].

Some researchers use the discrete wavelet transform (DWT) with the Daubechies wavelet (db4) to do denoised process. They apply soft threshold to detail coefficients, and reconstruct the denoised signal by setting high-frequency coefficients to zero [LDM24]. Also, some researchers select the db4 wavelet as the parent wavelet, and set five layers of decomposition. After autocorrelation analysis, they kept the wavelet coefficients D5 and D3, and the scale coefficient S[PCL21]. In addition, some authors use a Sym wavelet. It is a symmetric orthogonal function of a db wavelet.

He decomposes the time-series into 4 detail components and a low-frequency layer, and then uses the low-frequency layer to recompose the time-series data. It eliminates the high frequency noise from the data [Tho21].

According to the literature review, it is feasible to do db4 wavelet transform on the close price, and then the soft threshold-denoising method is applied. Wavelet decomposition can separate short-term fluctuation, which is reflected by detail coefficient, and long-term trend showed by approximation coefficient. It is beneficial to provide more useful features.

4.2 Factors from Cointegration Analysis

This report does the bivariate cointegration, using the first difference term and the lag term of residuals as features. Further research may try ratios as features, such as SPY/QQQ, and SPY/DIA.

Reference

[Cha+24] Chang, Victor, Qianwen Ariel Xu, Anyamele Chidozie, and Hai Wang. "Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced Machine Learning Techniques." *Electronics* 13, no. 17 (August 26, 2024): 3396. https://doi.org/10.3390/electronics13173396.

[Tho21] Thomas, Matthew W. "Noise Reduction for LSTM Using Wavelet Transform and Singular Spectrum Analysis," *Stanford University*, 2021.

[Tan+21] Tang, Qi, Ruchen Shi, Tongmei Fan, Yidan Ma, and Jingyan Huang. "Prediction of Financial Time Series Based on LSTM Using Wavelet Transform and Singular Spectrum Analysis." Edited by Jie Chen. *Mathematical Problems in Engineering* 2021 (June 8, 2021): 1–13. https://doi.org/10.1155/2021/9942410.

[Rhi+19] Rhif, Manel, Ali Ben Abbes, Imed Riadh Farah, Beatriz Martínez, and Yanfang Sang. "Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review." *Applied Sciences* 9, no. 7 (March 30, 2019): 1345. https://doi.org/10.3390/app9071345.

[PCL21] Peng, Lifang, Kefu Chen, and Ning Li. "Predicting Stock Movements: Using Multiresolution Wavelet Reconstruction and Deep Learning in Neural Networks." *Information* 12, no. 10 (September 22, 2021): 388. https://doi.org/10.3390/info12100388.

[Lia+19] Liang, Xiaodan, Zhaodi Ge, Liling Sun, Maowei He, and Hanning Chen. "LSTM with Wavelet Transform Based Data Preprocessing for Stock Price Prediction." Edited by Sylwester Samborski. *Mathematical Problems in Engineering* 2019, no. 1 (January 2019): 1340174. https://doi.org/10.1155/2019/1340174.

[LDM24] Gil, Gonzalo Lopez, Paul Duhamel-Sebline, and Andrew McCarren. "An Evaluation of Deep Learning Models for Stock Market Trend Prediction." *arXiv*, August 22, 2024. https://doi.org/10.48550/arXiv.2408.12408.

[ZSI24] Zhang, Cheng, Nilam Nur Amir Sjarif, and Roslina Ibrahim. "Deep Learning Models for Price Forecasting of Financial Time Series: A Review of Recent Advancements: 2020–2022." *WIREs Data Mining and Knowledge Discovery* 14, no. 1 (January 2024): e1519. https://doi.org/10.1002/widm.1519.

[Col21] Colah. Understanding LSTM Networks – colah's blog. [Online; accessed 29. Oct. 2021]. Oct. 2021. URL: https://colah.github.io/posts/2015-08-Understanding- LSTMs.

[Wu+23] Wu, Jimmy Ming-Tai, Zhongcui Li, Norbert Herencsar, Bay Vo, and Jerry Chun-Wei Lin. "A Graph-Based CNN-LSTM Stock Price Prediction Algorithm with Leading Indicators." *Multimedia Systems* 29, no. 3 (June 2023): 1751–70. https://doi.org/10.1007/s00530-021-00758-w.

[SS20] Shen, Jingyi, and M. Omair Shafiq. "Short-Term Stock Market Price Trend Prediction Using a Comprehensive Deep Learning System." *Journal of Big Data* 7, no. 1 (December 2020): 66. https://doi.org/10.1186/s40537-020-00333-6.

[Sch02] Schleicher, Christoph. "An Introduction to Wavelets for Economists," *Bank of Canada,* working paper 2002-3.

[Jav+24] Javed, Ali, Donna M. Rizzo, Byung Suk Lee, and Robert Gramling. "Somtimes: Self Organizing Maps for Time Series Clustering and Its Application to Serious Illness Conversations." *Data Mining and Knowledge Discovery* 38, no. 3 (May 2024): 813–39. https://doi.org/10.1007/s10618-023-00979-9.