Optimizing the Transformer Model Architecture for trading in Equity and FX

- FYP Group RO4
(SAXENA Srijan, GOEL Kalpa, DYCHENGBENG, Matthew, CHRISTANTO, Nicholas)

Overview

Presentation Agenda

1. Introduction & Literature Review
2. Objectives
3. Design
4. Implementation & Testing
5. Evaluation & Discussion
6. Conclusion

1 Introduction

Motivation
Lit. Review
Objectives

2 Methodology

Design
Implementation
Testing

3 Evaluation

Equity Models
FX Models
Discussion

4 Conclusion

Technical
Accomplishments

Why Machine Learning in Finance?

72% of existing institutional investor base utilize Machine Learning in enhancing trading lifecycle

Industry Drivers

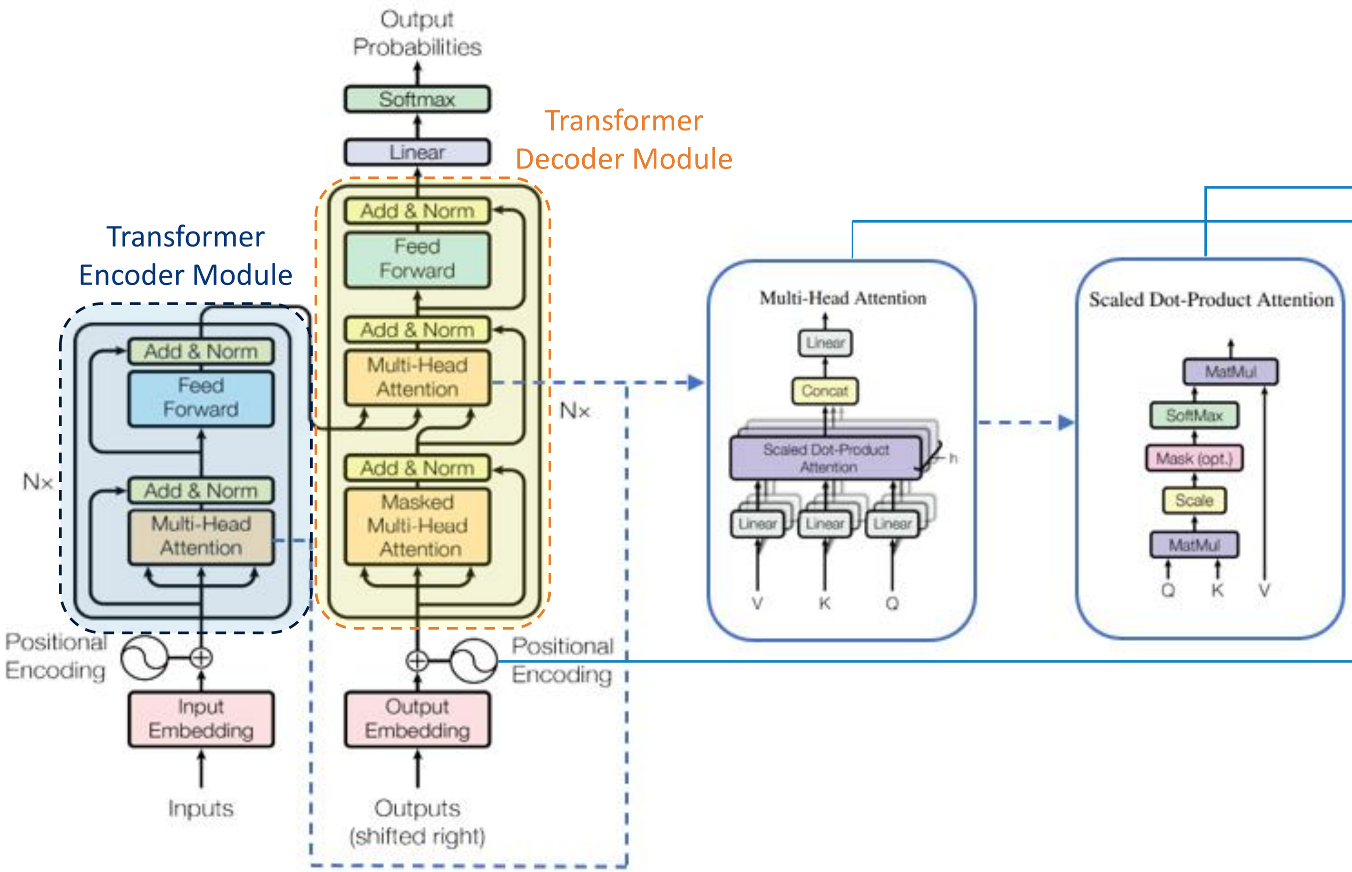
80% of institutional investors making significant investment

27% of institutional investors utilize AI/ML in trade execution

80% of institutional investors utilize AI/ML in risk management

Source: Refinitv

Literature Review – "Attention is All You Need"



Key Features

Self-Attention

- Allows each time point in the sequence to understand how it relates with every other time point.

Multi-Headed Attention

- Multiple self-attention modules can learn different types of relationships.

Positional Encoding

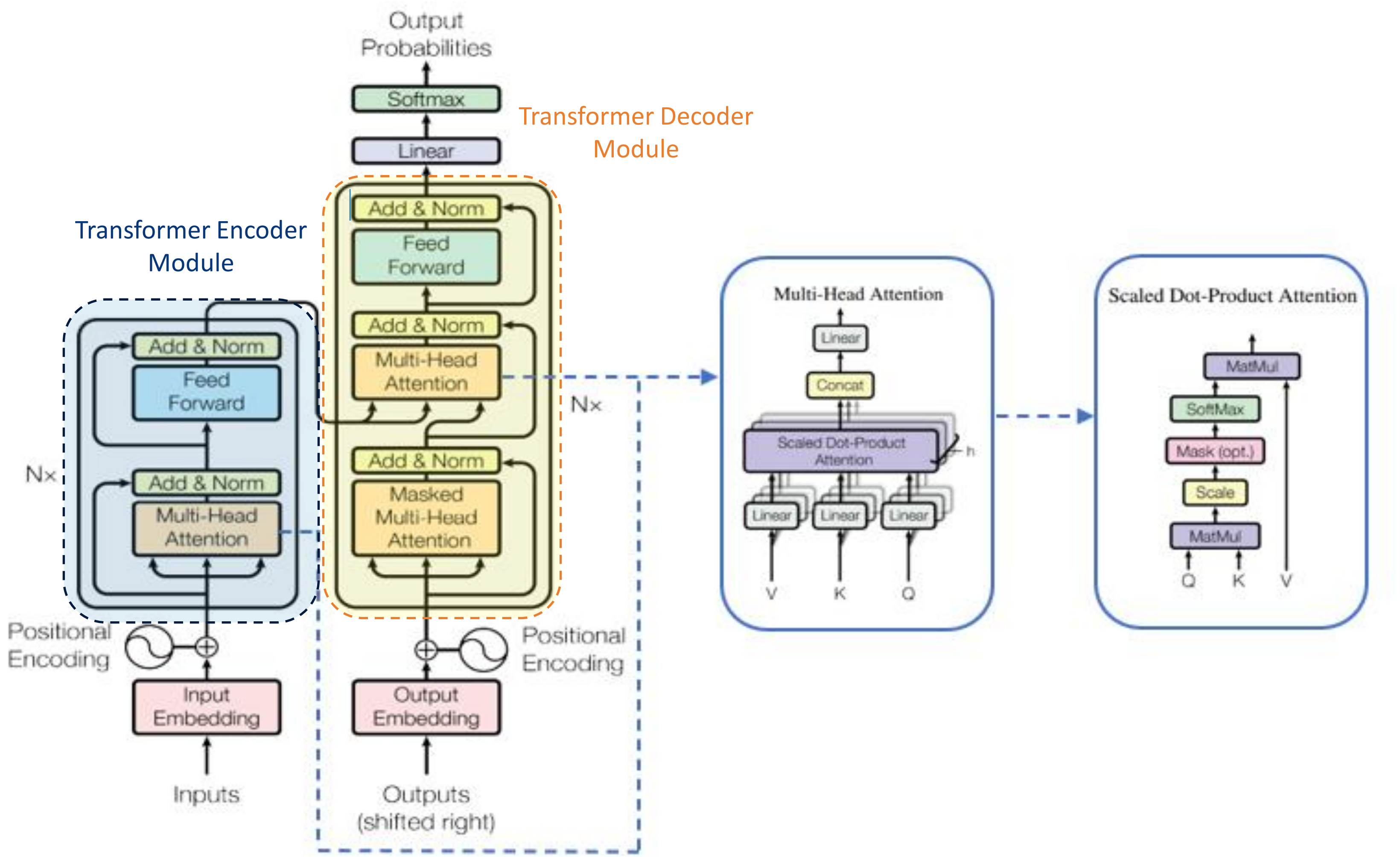
- Allows the model to understand the sequential nature of data.

Parallelization

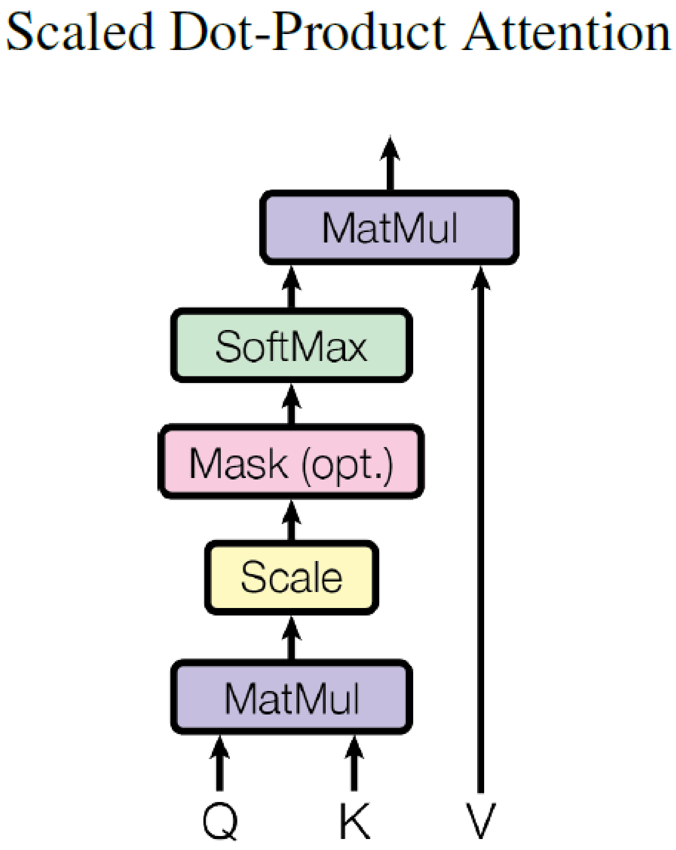
- Models can be efficiently trained with large datasets.

Key Takeaway	**Transformer Models are very effective at pattern recognition within a sequential context**

Literature Review – "Attention is All You Need"



Literature Review – "Attention is All You Need"

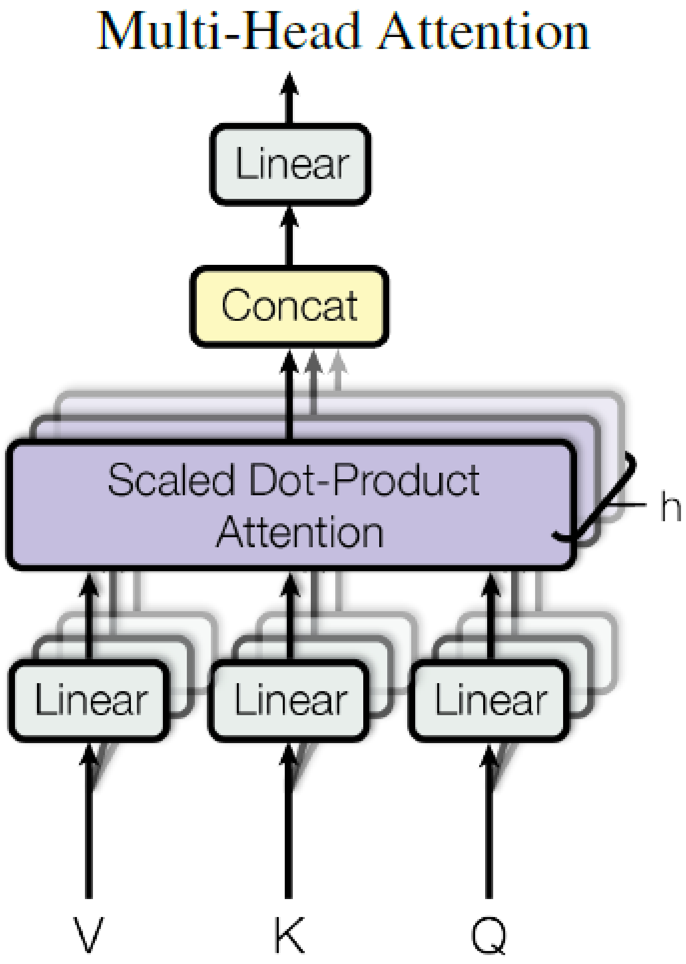


Key Features

Self-Attention

- Allows each time point in the sequence to understand how it relates with every other time point.

Literature Review – "Attention is All You Need"



Key Features

Self-Attention

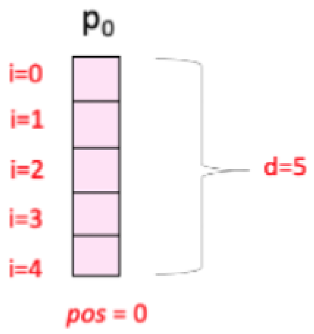
- Allows each time point in the sequence to understand how it relates with every other time point.

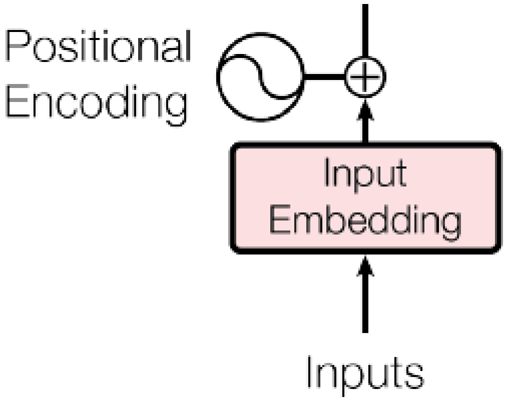
Multi-Headed Attention

- Multiple self-attention modules can learn different types of relationships.

Literature Review – "Attention is All You Need"

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$





Key Features

Self-Attention

- Allows each time point in the sequence to understand how it relates with every other time point.

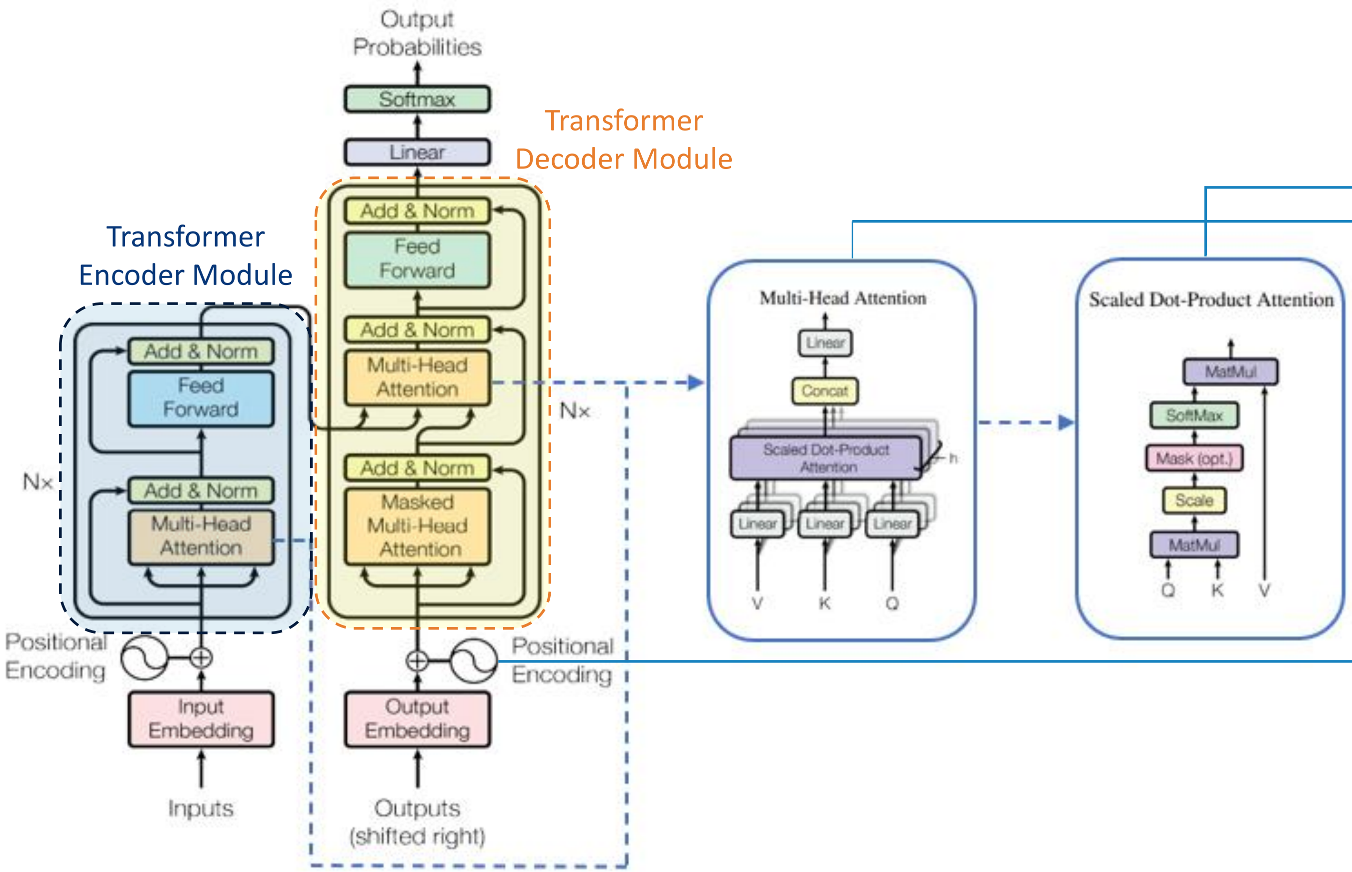
Multi-Headed Attention

- Multiple self-attention modules can learn different types of relationships.

Positional Encoding

- Allows the model to understand the sequential nature of data.

Literature Review – “Attention is All You Need”



Key Features

Self-Attention

- Allows each time point in the sequence to understand how it relates with every other time point.

Multi-Headed Attention

- Multiple self-attention modules can learn different types of relationships.

Positional Encoding

- Allows the model to understand the sequential nature of data.

Parallelization

- Models can be efficiently trained with large datasets.

Key Takeaway | **Transformer Models are very effective at pattern recognition within a sequential context**

Literature Review – The Success of the Transformer Model Architecture

Natural Language Processing (Machine Translation)

Table 2: The Transformer achieves better BLEU scores than previous state-of-the-art models on the English-to-German and English-to-French newstest2014 tests at a fraction of the training cost.

Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	**41.29**	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$\mathbf{3.3 \cdot 10^{18}}$	
Transformer (big)	**28.4**	**41.8**	$2.3 \cdot 10^{19}$	

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," arXiv, 2017.

Video Processing (Panoptic Segmentation)

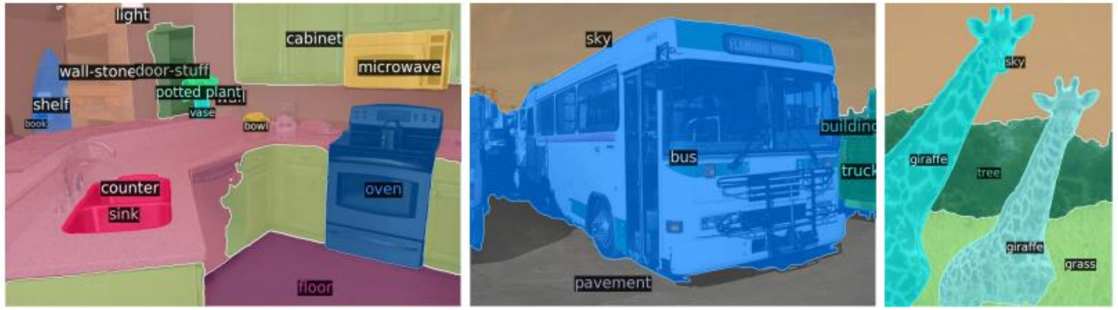


Fig. 9: Qualitative results for panoptic segmentation generated by DETR-R101. DETR produces aligned mask predictions in a unified manner for things and stuff.

Model	Backbone	PQ	SQ	RQ	PQ^{th}	SQ^{th}	RQ^{th}	PQ^{st}	SQ^{st}	RQ^{st}	AP
PanopticFPN++	R50	42.4	79.3	51.6	49.2	82.4	58.8	32.3	74.8	40.6	37.7
UPSnet	R50	42.5	78.0	52.5	48.6	79.4	59.6	33.4	75.9	41.7	34.3
UPSnet-M	R50	43.0	79.1	52.8	48.9	79.7	59.7	34.1	78.2	42.3	34.3
PanopticFPN++	R101	44.1	79.5	53.3	**51.0**	**83.2**	60.6	33.6	74.0	42.1	**39.7**
DETR	R50	43.4	79.3	53.8	48.2	79.8	59.5	36.3	78.5	45.3	31.1
DETR-DC5	R50	44.6	79.8	55.0	49.4	80.5	60.6	**37.3**	**78.7**	**46.5**	31.9
DETR-R101	R101	**45.1**	**79.9**	**55.5**	50.5	80.9	**61.7**	37.0	78.5	46.0	33.0

N. Carion, F. Massa "End-to-End Object Detection with Transformers," arXiv, 2020.

Key Takeaway | **Transformer Models have surpassed the state-of-the-art in various time-series domains**

1

Conduct preliminary statistical data analysis to identify the most important factors to act as input for the model.

2

Design and train a transformer-based model and develop a supplementary profitable trading strategy, that accounts for market friction in its decision-making process.

3

Systematically evaluate the performance of each factor pillar model to understand the benefits of each factor pillar and to determine the best performing model.

1

Introduction

Motivation
Lit. Review
Objectives

2

Methodology

Design
Implementation
Testing

3

Evaluation

Equity Models
FX Models
Discussion

4

Conclusion

Technical Accomplishments

Design – Scope of Securities

Equity Market

S&P Dow Jones Indices
A Division of S&P Global

S&P500 (SPY)

Apple Inc. (AAPL)

Amazon.com, Inc. (AMZN)

Microsoft Corp (MSFT)

Foreign Exchange (FX) Market

EUR/USD

USD/CAD

AUD/USD

USD/CHF

GBP/USD

USD/JPY

CNY/USD

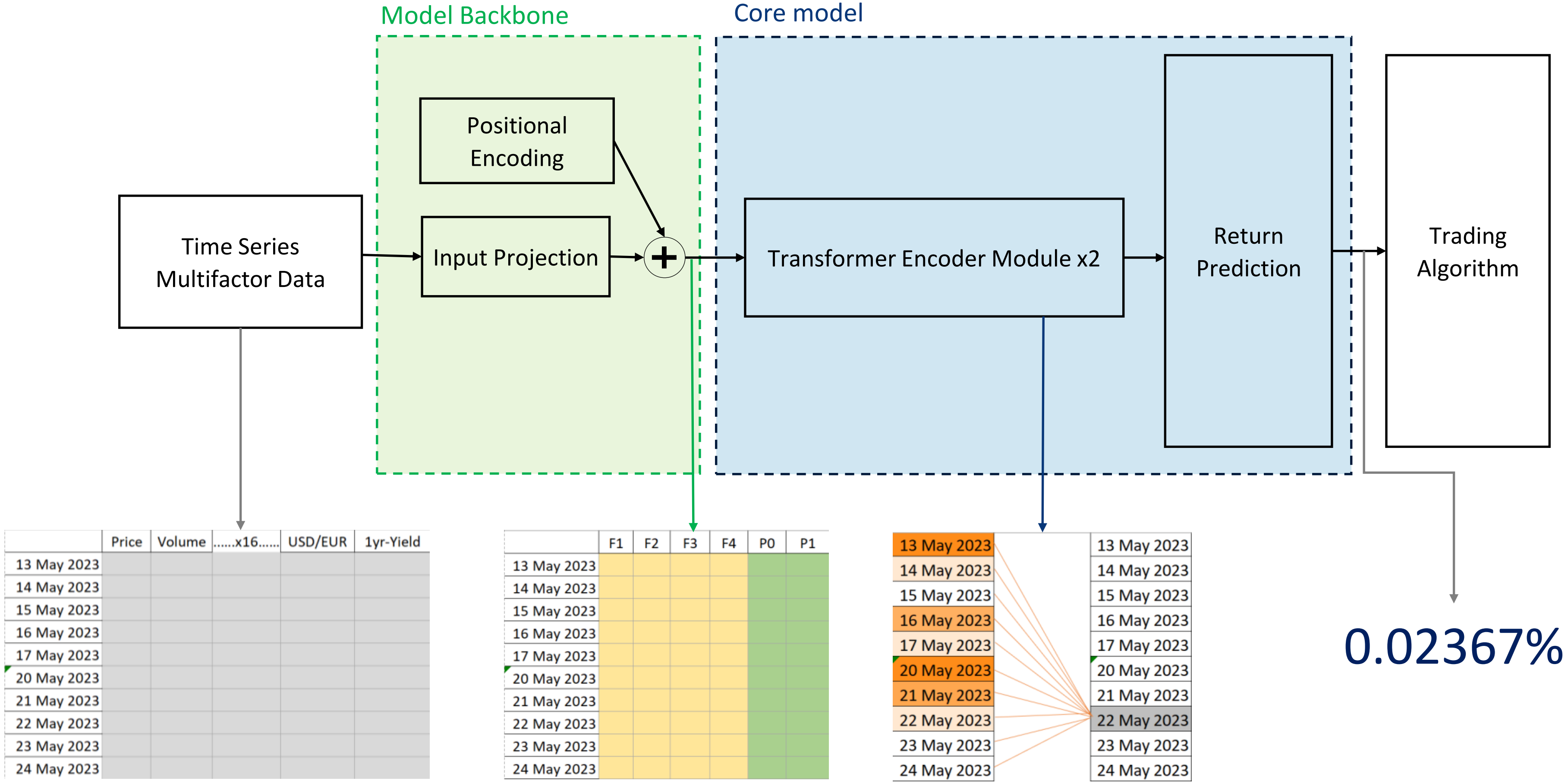
Design – Multifactor Approach

Factor Pillars	
Technical Factors	Price & Trading Data
Fundamental Factors	Numbers from Accounting Statements
Macro-economic Factors	External Factors that describe prevailing macroeconomic conditions
Value Factors	Metrics that describe how under/overvalued the security is.

T	Technical Factors Model	E	F
T+F	Technical + Fundamental Factors Model	E*	
T+M	Technical + Macroeconomic Factors Model	E	
T+V	Technical + Value Factors Model	E*	
T+F+M+V	All Factors Model	E*	

E – For All Equity Securities
E* - For All Equity Securities except SPY
F – For All FX Securities

Design – High Level System Architecture



Input Data: Selected factors over 10 days

Model Backbone Output: Learned Features + Positional Encoding

Transformer Encoder: Self-attention Visualization

Core Model Output: Predicted Normalized Average Return

Design – Evaluation Criteria

Metric	Definition	Why?	Desired Value
Cumulative Annual Growth Rate (CAGR)	***calculates the annually compounded equivalent rate of return over a period.***	Instead of using overall return, using CAGR is more objective since the timeframe is standardized into annual growth.	**GREATER** than CAGR of Buy & Hold Strategy
Sharpe Ratio	***calculated by dividing the average returns over the standard deviation of returns.***	It measures risk-adjusted returns and shows that the trading strategy can produce higher returns while taking on less risk.	**GREATER** than Sharpe Ratio of Buy & Hold Strategy
Maximum Drawdown	***measures the maximum decline (in %) of the trading balance from its peak to its trough.***	High maximum drawdown despite high returns indicate that the trading strategy is inconsistent.	**LESSER** than Maximum Drawdown of Buy & Hold Strategy

Data Sourcing

1 yahoo! finance

- Technical Data
- Fundamental Data
- Macroeconomic Data

2 EOD HISTORICAL DATA

- Fundamental Data

3 Bloomberg

- Macroeconomic Data
- Value Data

Security	Start Date	End Date	Number of Trading Day Data Points
AAPL	02-Jan-04	21-Sep-21	4410
AMZN	02-Jan-04	29-Sep-20	4163
MSFT	02-Jan-04	29-Sep-20	4163
SPY	02-Jan-04	17-Feb-23	4766
All FX	17-Sep-03	17-Feb-23	4998

To make **accurate comparisons between securities**, we **standardized** the time period for all factors

Due to **limited Fundamental Data**, the dataset within Equities was limited to September 2020

Factor Selection

Factors Used		Why?
Technical	**Daily data** • Opening Price • Closing Price • Highest Price • Lowest Price • Daily Trading Volume	Standard technical Indicators
Value	**Ratios** • Price to Earnings ratio • Price to Sales ratio • EV to EBITDA ratio	First three indicate underlying valuation relative to the firms' earnings and revenue
	Put Call ratio	Put Call Ratio summarizes the stock options trading flow and highlights investor sentiment on value

Factor Selection

Factors Used		Why?
Macroeconomic	**Treasury bills/bonds** • 13-week treasury bill • 10-year treasury bond • 30-year treasury bond	Treasury bills/bonds represent the overall interest rate conditions
	Crude oil	Crude Oil prices reflect all major industries and consumption across US industries
	Gold	Gold portray broader macroeconomic and business conditions influencing stock price movements.
	Currency pairs • CAD to USD exchange rate • JPY to USD exchange rate • EUR to USD exchange rate • CNY to USD exchange rate	Currency pairs represent the largest trading partners of the US

Fundamental Factor Selection

Stage 1: Identifying Key Fundamental Factors

Stage 2: Evaluate Profitability the Factors

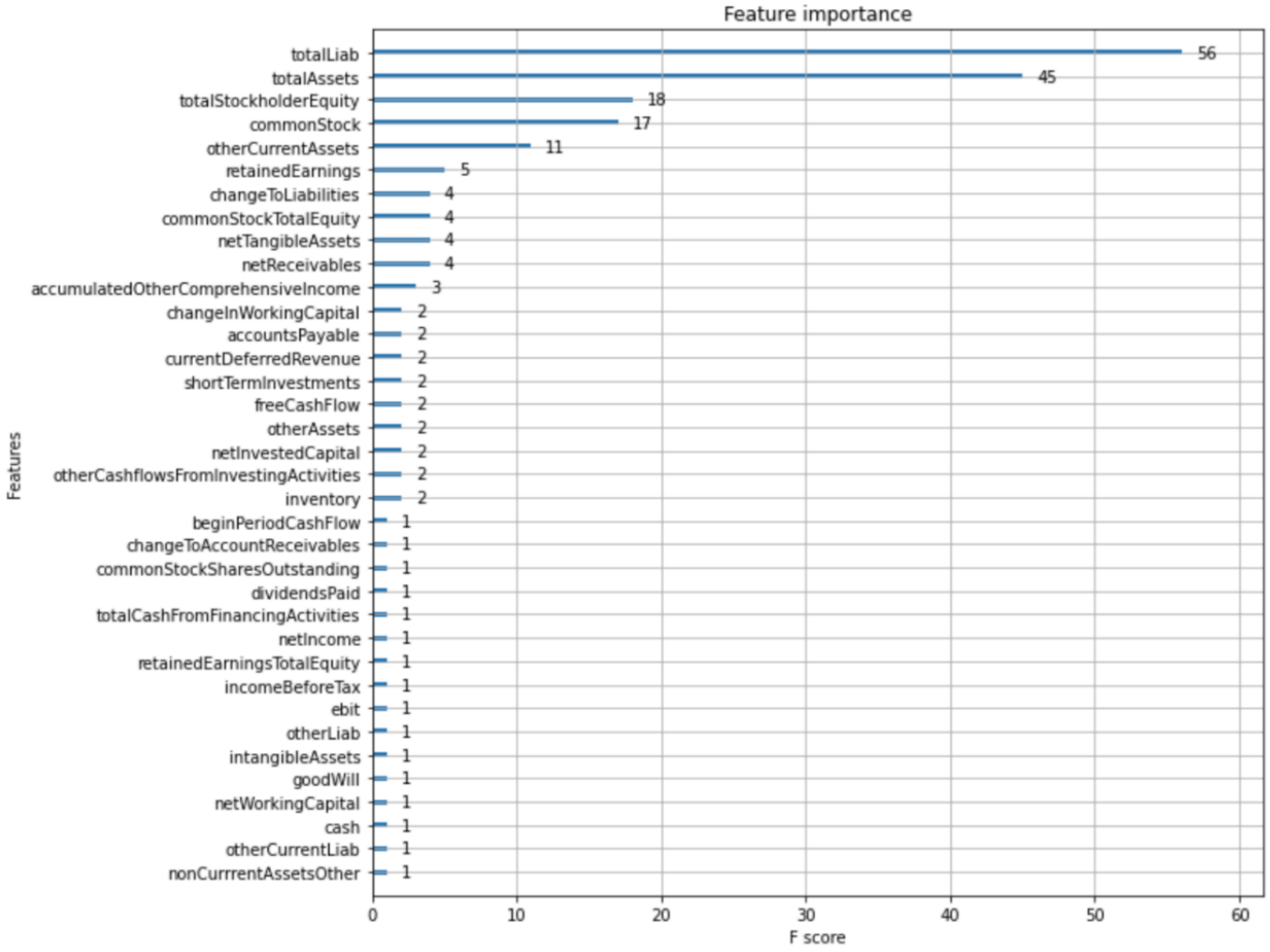
XGBoost Regressor + Extratree Regressor

1. XGBoost Regressor returned **5 key fundamental factors** based on feature importance
2. Extratree Regressor returned **16 key fundamental factors** based on feature importance
3. Selected the **16 key features for evaluation**

Mock Factor Portfolio

1. For each key feature, we **rank 100+ securities based on the security's respective value**
2. We create **a portfolio** for each key feature that **longs the top 5 securities and shorts the bottom 5 securities** per quarter
3. We **calculate the profit** for each feature and **rank the 16 stocks**
4. **Top 5 stocks** from this ranking are the key fundamental features used for the pillar

Fundamental Factor Selection



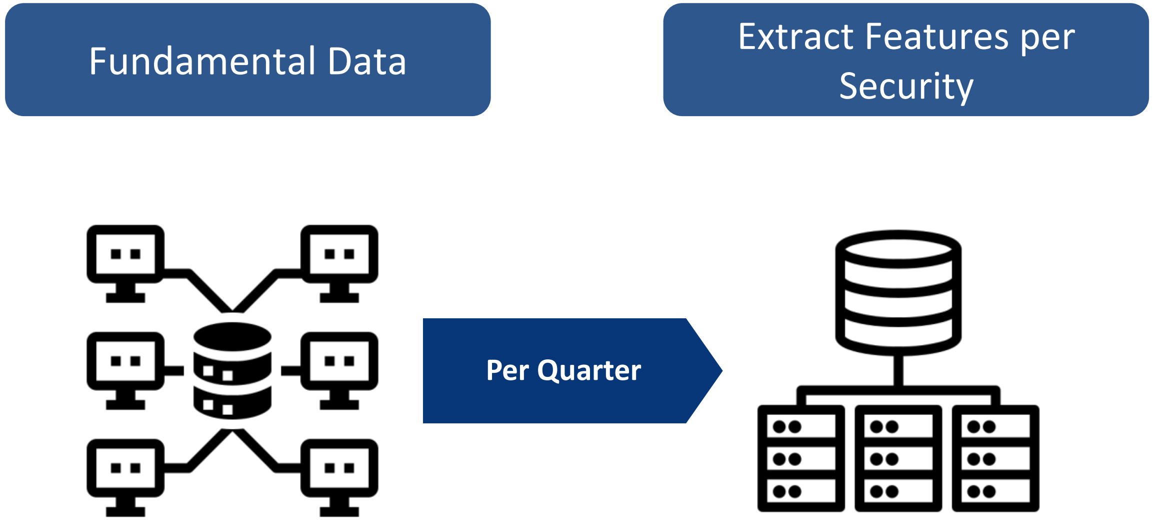
Key Features:

- Total Liabilities
- Total Assets
- Total Stockholder Equity
- Common Stock
- Other Current Assets

Fundamental Factor Selection

Mock Factor Portfolio

1. For each key feature, we rank 100+ securities based on the security's respective value
2. We create a portfolio for each key feature that longs the top 5 securities and shorts the bottom 5 securities per quarter
3. We calculate the profit for each feature and rank the 16 stocks
4. Top 5 stocks from this ranking are the key fundamental features used for the pillar



Fundamental Factor Selection

Mock Factor Portfolio

1. For each key feature, we rank 100+ securities based on the security's respective value
2. We create a portfolio for each key feature that longs the top 5 securities and shorts the bottom 5 securities per quarter
3. We calculate the profit for each feature and rank the 16 stocks
4. Top 5 stocks from this ranking are the key fundamental features used for the pillar

Retrieve Best and Worst 5 Securities

Total Assets

GE
WMT
XOM
VZ
PFE

.
.
.

NFLX
VRTX
EQIX
LRCX
REGN

"Total Assets" Mock Factor Portfolio

Fundamental Factor Selection

Mock Factor Portfolio

1. For each key feature, we rank 100+ securities based on the security's respective value
2. We create a portfolio for each key feature that longs the top 5 securities and shorts the bottom 5 securities per quarter
3. We calculate the profit for each feature and rank the 16 stocks
4. Top 5 stocks from this ranking are the key fundamental features used for the pillar

Retrieve Best and Worst 5 Securities

	time	long	short
0	2004-01	[GE, WMT, XOM, VZ, PFE]	[NFLX, VRTX, EQIX, LRCX, REGN]
1	2004-04	[GE, XOM, VZ, WMT, CVX]	[NFLX, LRCX, REGN, EQIX, VRTX]
2	2004-07	[GE, XOM, VZ, CVX, MSFT]	[EQIX, CRM, REGN, VRTX, DHR]
3	2004-10	[CVX, XOM, GE, VZ, JPM]	[LRCX, EQIX, CRM, VRTX, REGN]
4	2005-01	[GE, XOM, VZ, WMT, CVX]	[BLK, EQIX, REGN, CRM, VRTX]
...	...	...	...
62	2019-07	[T, MSFT, VZ, XOM, WMT]	[AMD, INTU, NOW, REGN, MMC]
63	2019-10	[T, AAPL, VZ, WMT, XOM]	[SLB, REGN, AXP, BSX, FISV]
64	2020-01	[T, AAPL, CVX, AMZN, MSFT]	[REGN, NOW, VRTX, JPM, C]
65	2020-04	[T, VZ, WMT, MSFT, AAPL]	[AMD, NOW, SLB, DHR, USB]
66	2020-07	[T, WMT, AMZN, MSFT, VZ]	[EL, NOW, NVDA, EOG, MS]

Fundamental Factor Selection

Rank	Feature
1	Total Non-Current Assets
2	Total Liabilities
3	Net Debt
4	Total Assets
5	Intangible Assets
6	Other Non Cash Items
7	Dividends Paid
8	Other Cashflows From Investing Activities
9	Gross Profit
10	Total Stockholder Equity
11	Net Working Capital
12	Common Stock
13	EBITDA

Key Features:

- Total Assets
- Total Liabilities
- Net Debt
- Intangible Assets
- Total non-current Assets.

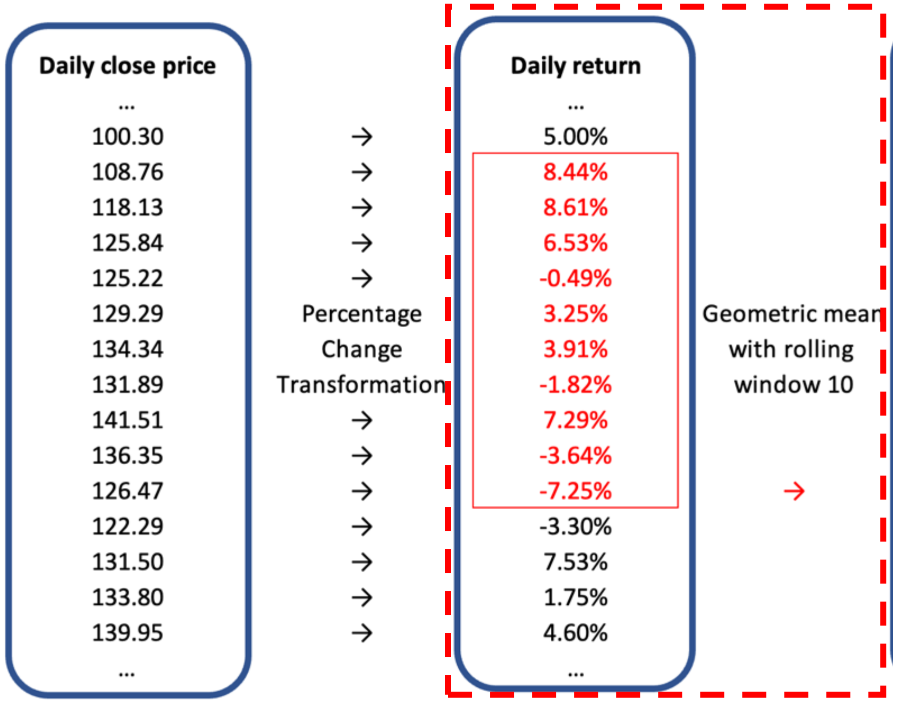
Pre-processing and Transformation

Daily close price		Daily return
...		...
100.30	→	5.00%
108.76	→	8.44%
118.13	→	8.61%
125.84	→	6.53%
125.22	→	-0.49%
129.29	Percentage	3.25%
134.34	Change	3.91%
131.89	Transformation	-1.82%
141.51	→	7.29%
136.35	→	-3.64%
126.47	→	-7.25%
122.29	→	-3.30%
131.50	→	7.53%
133.80	→	1.75%
139.95	→	4.60%
...		...

Stationarity

- The condition where the statistical properties of the data remain unchanged over time
- Use percentage change which removes trends and frame the data into rate of change

Pre-processing and Transformation



Transformation

- Rolling geometric mean transformation was used to smooth data and reduce noise
- Rolling mean window is set to 10 to cover an average of two weeks in our daily data.

Implementation

Pre-processing and Transformation

Daily close price		Daily return		Averaged data
...		...		...
100.30	→	5.00%		2.29%
108.76	→	8.44%		0.23%
118.13	→	8.61%		0.35%
125.84	→	6.53%		0.10%
125.22	→	-0.49%		0.83%
129.29	Percentage	3.25%	Geometric mean	0.84%
134.34	Change	3.91%	with rolling	0.47%
131.89	Transformation	-1.82%	window 10	0.84%
141.51	→	7.29%		0.44%
136.35	→	-3.64%		0.33%
126.47	→	-7.25%	→	2.48%
122.29	→	-3.30%		1.31%
131.50	→	7.53%		1.20%
133.80	→	1.75%		0.72%
139.95	→	4.60%		1.23%
...		...		...

Outlier Selection

- Remove data points where the factor values were higher than 10 times the Inter Quartile Range of that factor

Pre-processing and Transformation

Daily close price		Daily return		Averaged data	
...		...		...	
100.30	→	5.00%		2.29%	→
108.76	→	8.44%		0.23%	→
118.13	→	8.61%		0.35%	→
125.84	→	6.53%		0.10%	→
125.22	→	-0.49%		0.83%	→
129.29	Percentage	3.25%	Geometric mean	0.84%	MaxAbs
134.34	Change	3.91%	with rolling	0.47%	and
131.89	Transformation	-1.82%	window 10	0.84%	MinMax
141.51	→	7.29%		0.44%	scaling
136.35	→	-3.64%		0.33%	→
126.47	→	-7.25%	→	2.48%	→
122.29	→	-3.30%		1.31%	→
131.50	→	7.53%		1.20%	→
133.80	→	1.75%		0.72%	→
139.95	→	4.60%		1.23%	→
...		...		...	

Normalization

- Performed max absolute scaling for features that includes negative values
- Takes the values in each factor and divides it with the maximum absolute value of that feature

Pre-processing and Transformation

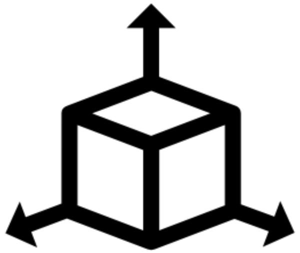
Daily close price		Daily return		Averaged data		Normalized data
...		...		...		...
100.30	→	5.00%		2.29%	→	0.93
108.76	→	8.44%		0.23%	→	0.92
118.13	→	8.61%		0.35%	→	0.41
125.84	→	6.53%		0.10%	→	0.56
125.22	→	-0.49%		0.83%	→	0.63
129.29	Percentage	3.25%	Geometric mean	0.84%	MaxAbs	0.69
134.34	Change	3.91%	with rolling	0.47%	and	0.62
131.89	Transformation	-1.82%	window 10	0.84%	MinMax	0.68
141.51	→	7.29%		0.44%	scaling	0.00
136.35	→	-3.64%		0.33%	→	0.12
126.47	→	-7.25%	→	2.48%	→	0.58
122.29	→	-3.30%		1.31%	→	0.53
131.50	→	7.53%		1.20%	→	0.55
133.80	→	1.75%		0.72%	→	0.05
139.95	→	4.60%		1.23%	→	0.79
...		...		...		...

Normalization

- Performed min-max scaling for that only include positive values
- Takes the minimum and maximum values of the feature and scale the data into the range from 0.0 to 1.0

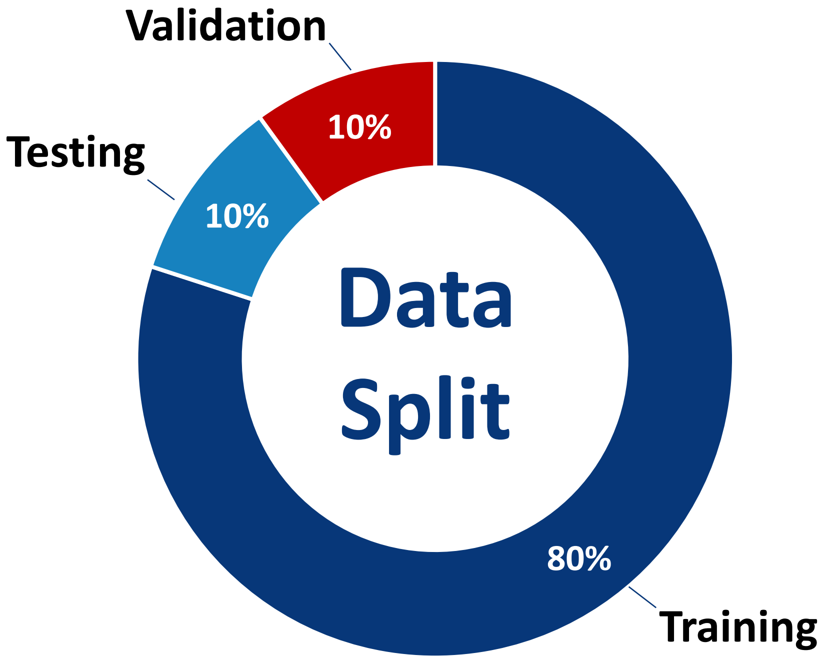
Pre-processing and Transformation

Principal Component Analysis

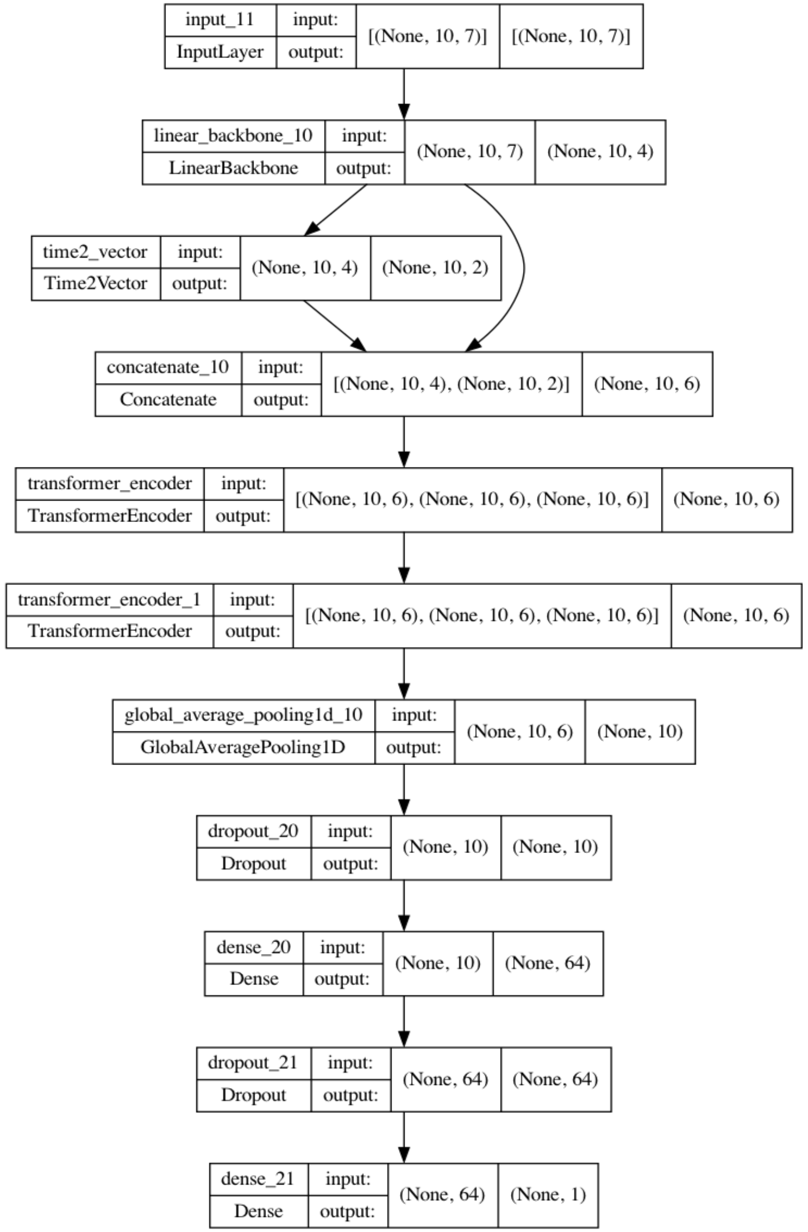


- A dimensionality reduction method that is often used to reduce the dimensionality of large data sets
- Chose the components that cumulatively explained greater than 99% of covariance between the factors

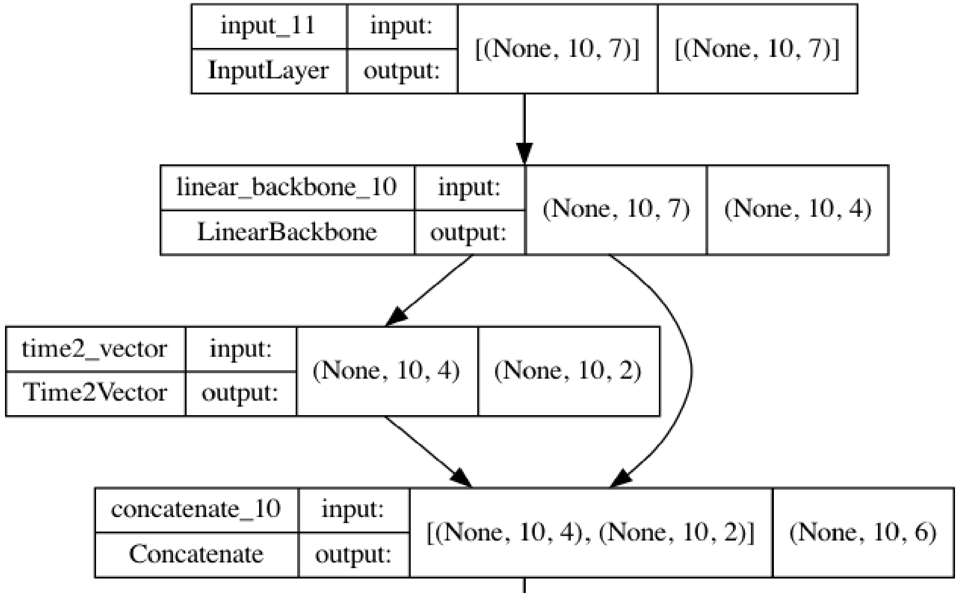
Dataset Split



Model Design and Training



Model Design and Training

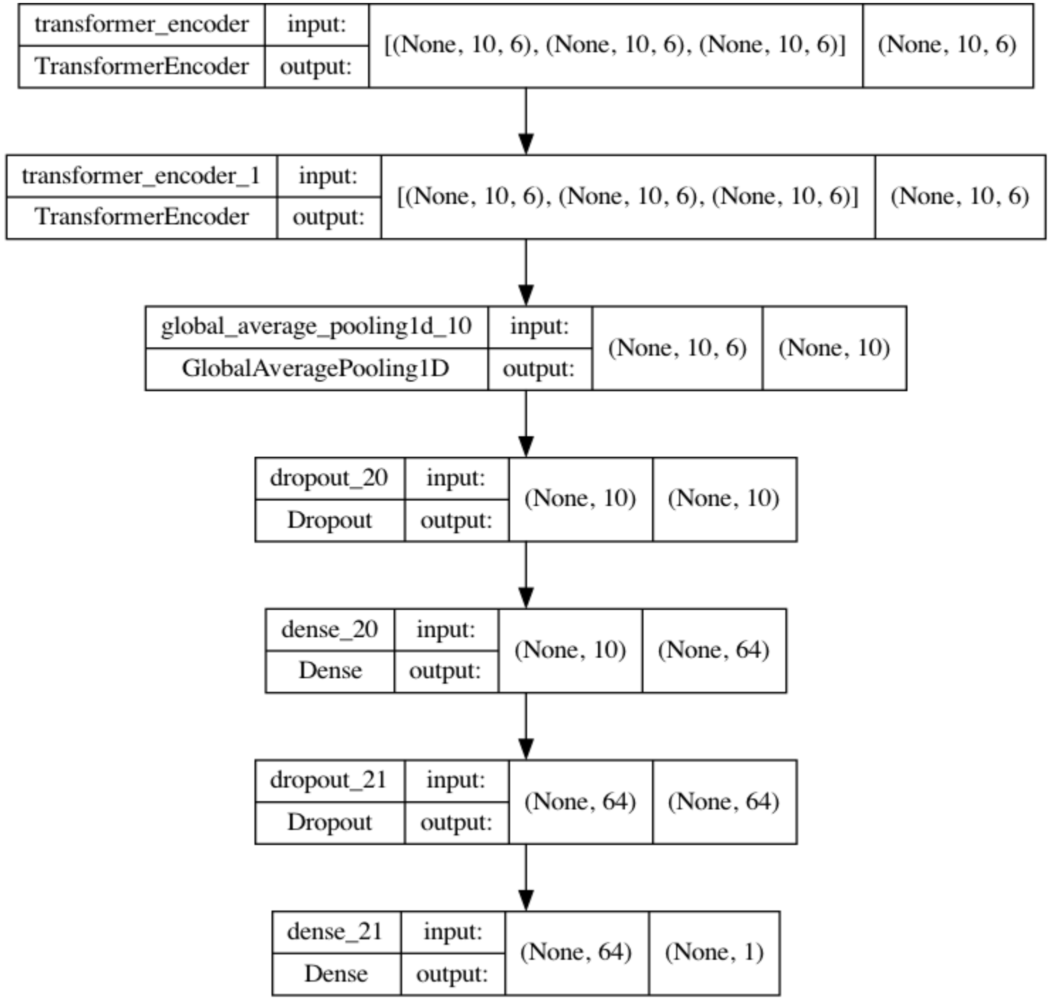


Time2Vec Layer for positional encoding

- Designed to use 10 days of continuous daily data to predict the following day's closing returns
- Utilizes a feedforward layer as the input projection layer
- Use Time2Vec to apply positional encoding on the predicted data

Implementation

Model Design and Training



Transformer Encoders

- Data is passed through two consecutive transformer encoder modules
- Data is then followed by global average pooling to turn the data into one dimension
- Two dropout-enabled feedforward layers are used to output a single value

Trading strategy

How to generate buy or sell instructions? → Based on target variable: **Close Price**

Pre-processing of Close Price:

i) Daily raw close price

↓

ii) % change in close price (x_i)

iii) Processed labels ($Close_i$) for Close Price [sample calculation]:

For time t_4

$$Close_4 = [(1 + x_0) \times (1 + x_1) \times (1 + x_2) \times (1 + x_3) \times (1 + x_4)]^{\frac{1}{5}} - 1$$

For time t_5

$$Close_5 = [(1 + x_1) \times (1 + x_2) \times (1 + x_3) \times (1 + x_4) \times (1 + x_5)]^{\frac{1}{5}} - 1$$

Notes on actual implementation:

1) 10 day rolling window utilized for geometric mean
2) Max absolute normalization after label generation ($Close_i$) does not impact buy/sell instruction logic

Trading strategy

Processed data labels for two consecutive days

For time t_4

$$Close_4 = [(1+x_0)*(1+x_1)*(1+x_2)*\ (1+x_3)*\ (1+x_4)]^{\frac{1}{5}}-1$$

For time t_5

$$Close_5 = [(1+x_1)*(1+x_2)*\ (1+x_3)*\ (1+x_4)*\ (1+x_5)]^{\frac{1}{5}}-1$$

What is the difference?

$(1+x_0)$ vs $(1+x_5)$

How does the difference help generate trading signals?

Buy Signal

$Close_5 > Close_4\ \rightarrow x_5 > x_0$

What if x_5 & x_0 are negative?

→ Narrowed down buy condition:
$\boldsymbol{Close_5 > Close_4}$ **and** $\boldsymbol{x_0 > 0 \rightarrow x_5 > 0}$

Sell Signal

$Close_5 < Close_4\ \rightarrow x_5 < x_0$

What if x_5 & x_0 are positive?

→ Narrowed down sell condition:
$\boldsymbol{Close_5 < Close_4}$ **and** $\boldsymbol{x_0 < 0 \rightarrow x_5 < 0}$

Back-testing

Stop Loss Implementation

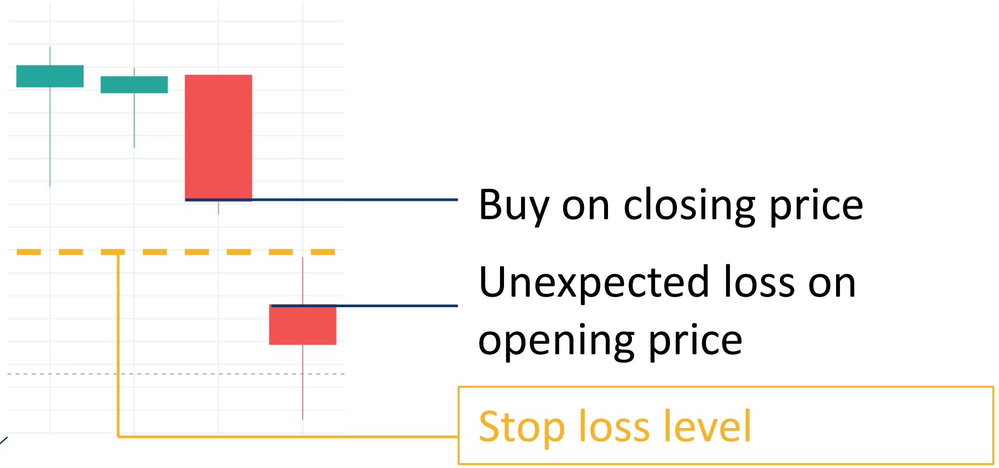
i) Tested stop-loss levels on validation set:
0.0001% to 10% at every 10x multiple

ii) Utilized the optimal on testing data set

iii) Stop-loss trigger mechanism:
→ Measured as % change in daily close price

iv) Limitation on daily frequency data:
→ If $\%\Delta(Open_t - Close_{t-1}) > \%\Delta(Close_t - Close_{t-1})$
→ Loss set at $\Delta(Open_t - Close_{t-1})$

Example of unexpected loss



Back-testing

Stop Loss Implementation

i) Tested stop-loss levels on validation set:
0.0001% to 10% at every 10x multiple

ii) Utilized the optimal on testing data set

iii) Stop-loss trigger mechanism:
→ Measured as % change in daily close price

iv) Limitation on daily frequency data:
→ If $\%\Delta(Open_t - Close_{t-1}) > \%\Delta(Close_t - Close_{t-1})$
→ Loss set at $\Delta(Open_t - Close_{t-1})$

Trading Simulation

i) Set 10,000 USD initial trading balance

InteractiveBrokers

ii) Used Interactive Brokers' rates for commission simulation as one of Hong Kong's largest retail brokers

iii) Rates breakdown:
→ Equities: 0.05 USD/share (min. 1 USD – max. 1% notional)
→ FX: 0.2 bps or 0.02% of notional trade

Data and Pre-processing Testing

Data Testing

Cross referencing collected data across multiple sources.

EODHD APIs

Bloomberg

yahoo! finance

Investing.com

Pre-processing Testing

Inversion of processed data to retrieve pre-processed data in the inverted 3 step process.

Percentage Change transformation

Geometric Mean transformation

Min-Max or Max Abs. Scaling transformation

Model Testing

To evaluate the learning ability of the transformer model:

Self generated data set 1:

$$C_t = 1.001 \times C_{t-1} - 1.0009 \times C_{t-2} + 1.0008 \times C_{t-3} - 1.0007 \times C_{t-4}$$
$$+ 1.0006 \times C_{t-5} - 1.0005 \times C_{t-6} + 1.0004 \times C_{t-7} - 1.0003 \times C_{t-8}$$
$$+ 1.0002 \times C_{t-9} - 1.0001 \times C_{t-10}$$

i) C_t stands for closing price on day t

ii) Simulates the relational trend of a stock price with historical context

iii) Choice of co-efficient and linear function was arbitrary

Self generated data set 2:

$$C_t = \sin(x)\,, where\ x\ is\ a\ random\ generated\ positive\ integer$$

i) C_t stands for closing price on day t

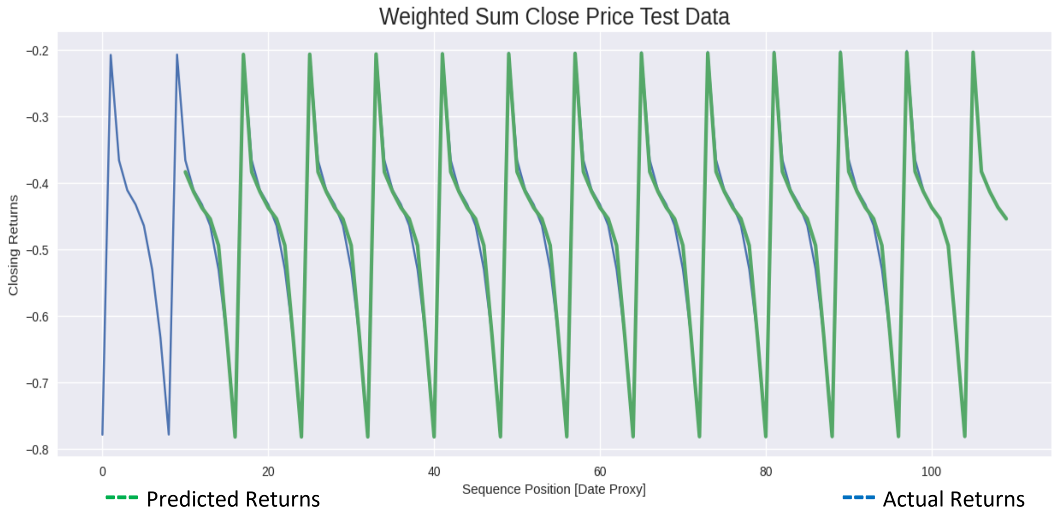
ii) Devoid of any relational trends and historical context

iii) Creates a random path for closing price

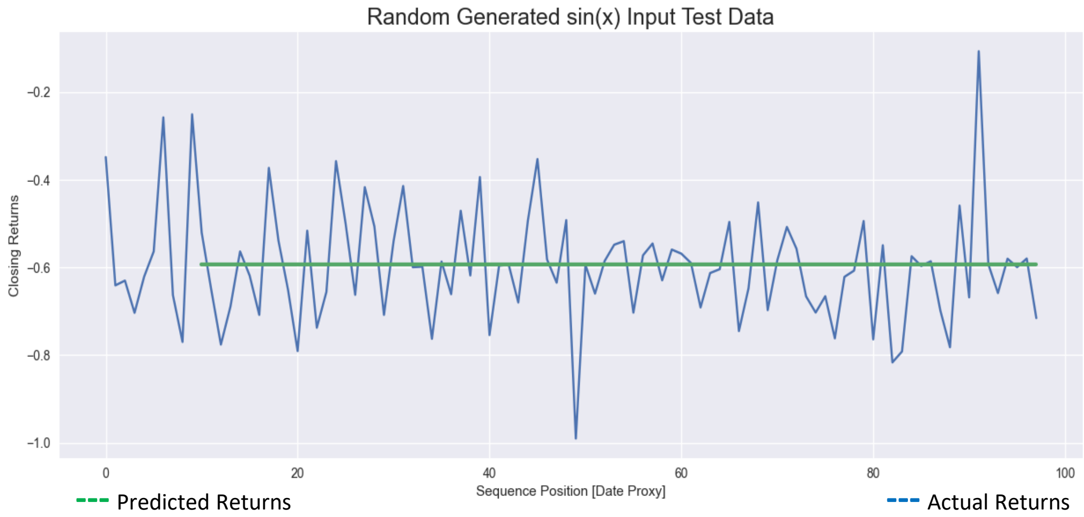
Model Testing

To evaluate the learning ability of the transformer model:

Self generated data set 1:



Self generated data set 2:



Key Takeaway:

1) Transformer is able to learn trends and underlying patterns from a sequential and relational data set.
2) Transformer does not learn under randomness without relational/sequential trends.

1

Introduction

Motivation
Lit. Review
Objectives

2

Methodology

Design
Implementation
Testing

3

Evaluation

Equity Models
FX Models
Discussion

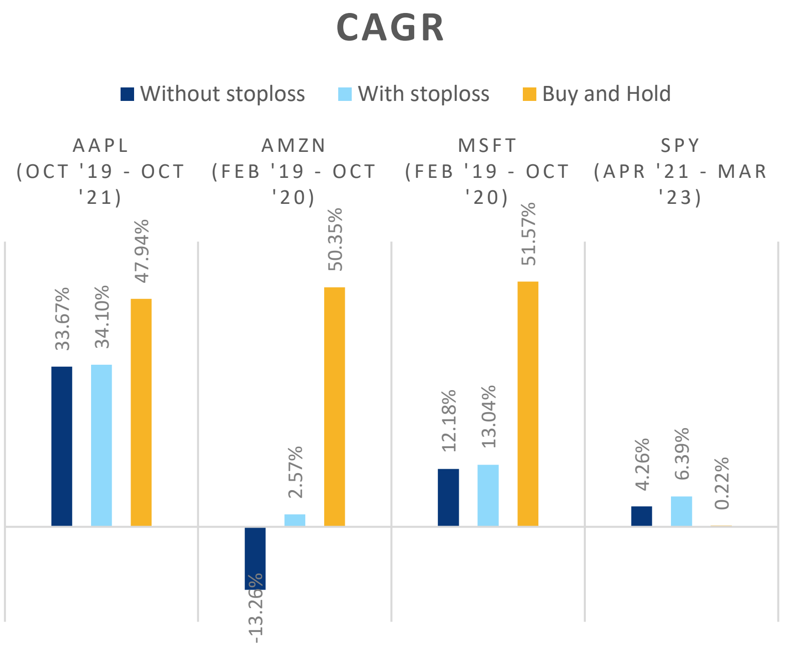
4

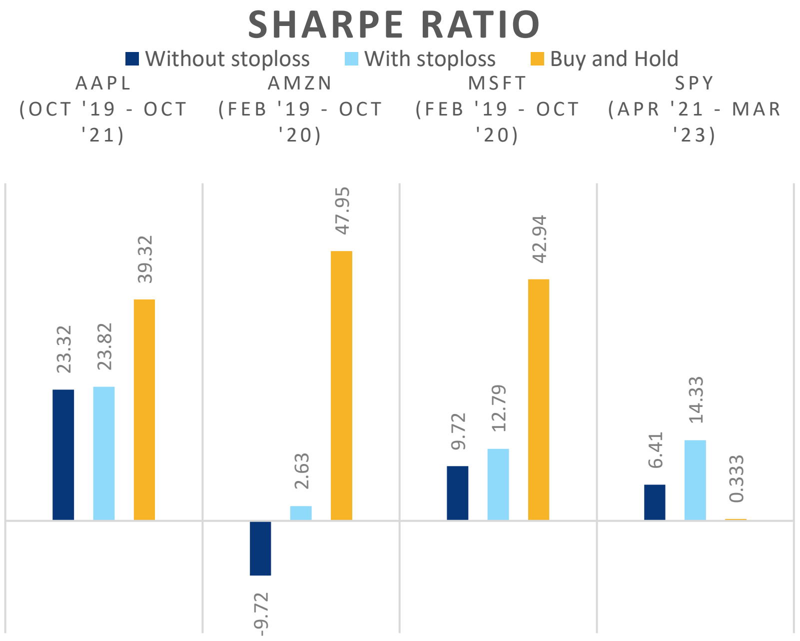
Conclusion

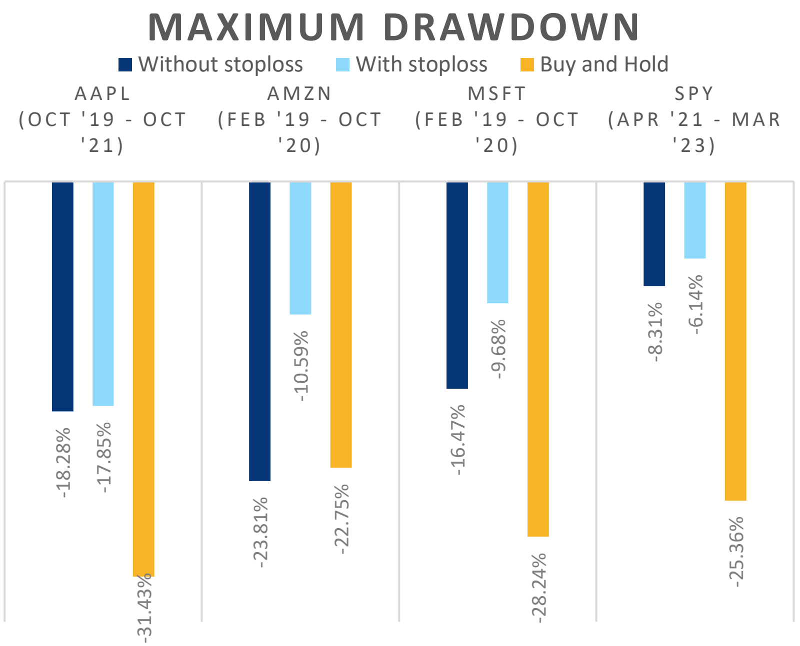
Technical Accomplishments

Evaluation of model trading performance on equity

Equity: Technical model







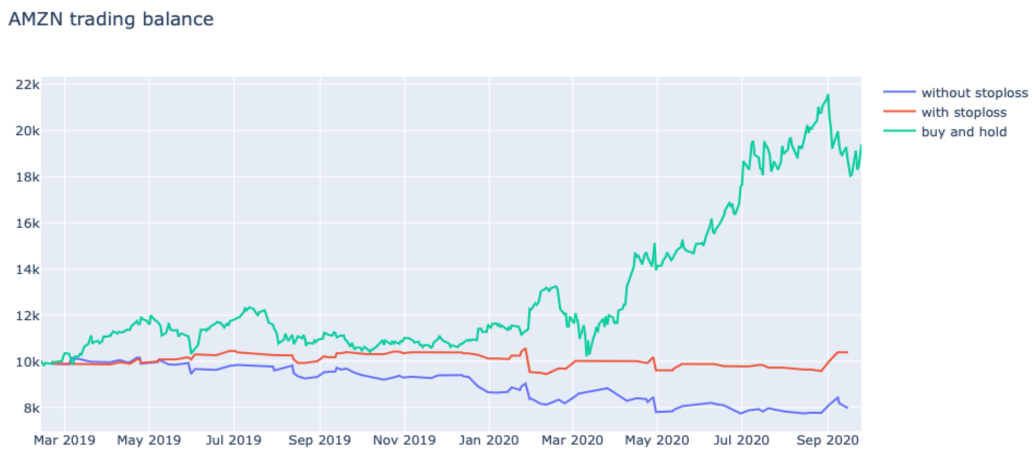
Key takeaway

Transformer Model is **profitable but not better** than buy-and-hold strategy

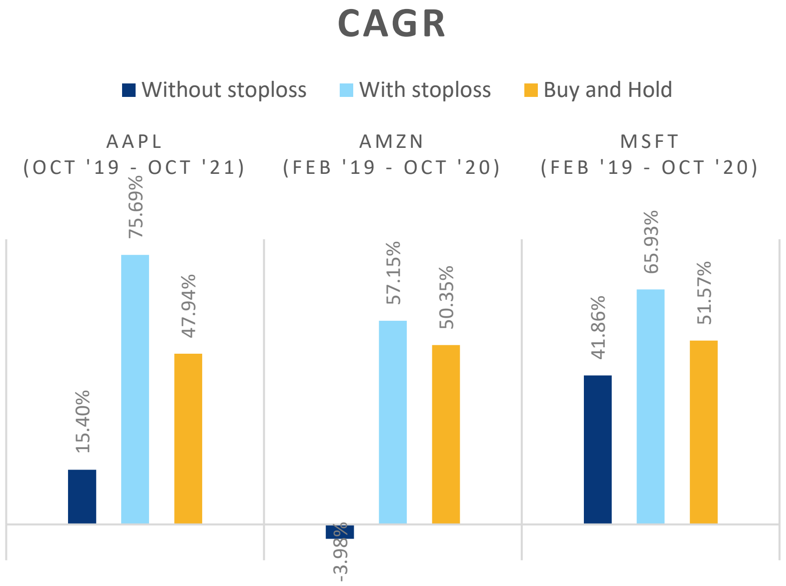
Transformer model has **lower risk-adjusted returns**

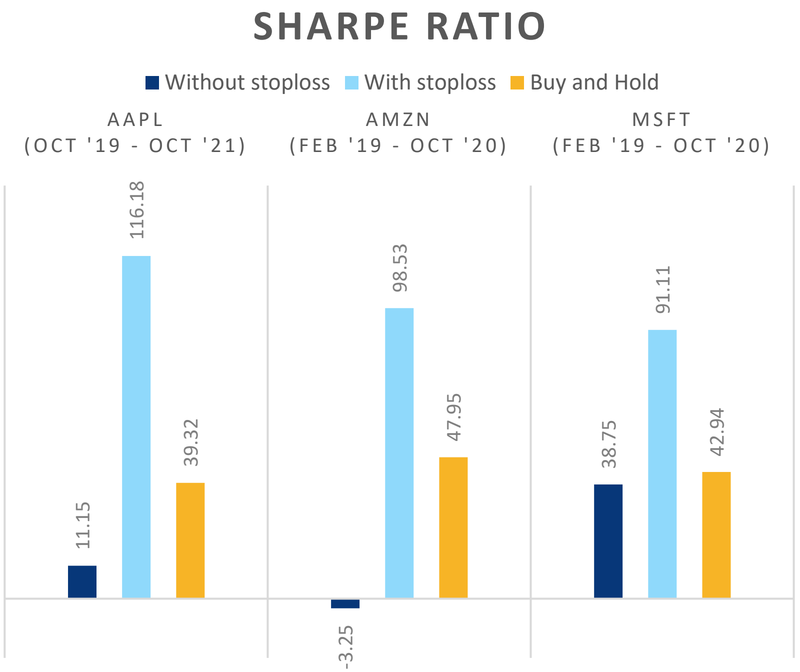
Transformer model was able to **reduce maximum drawdown** by taking less trades, especially losing trades

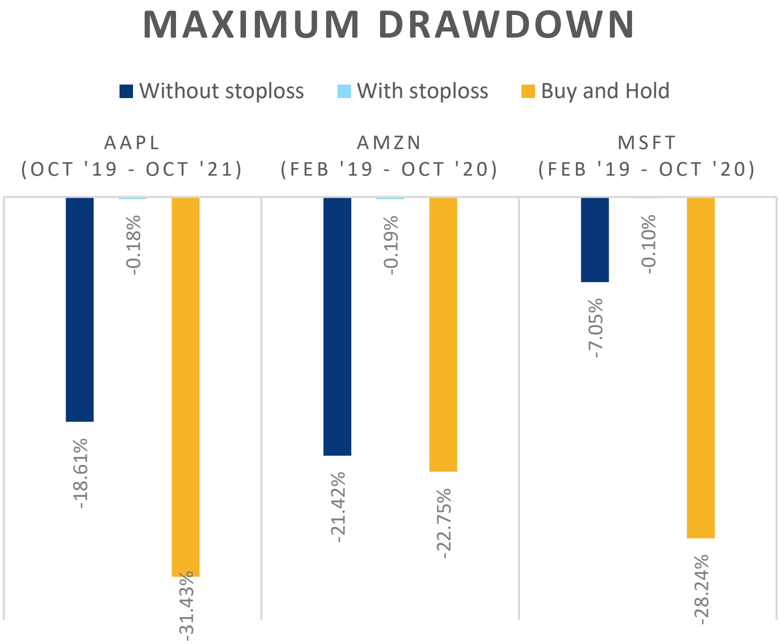
Equity: Technical model



Equity: Technical + Fundamental model







Key takeaway

Transformer model was **able to outperform** buy-and-hold strategy CAGR when using stoploss

Transformer model with stoploss mechanism produce **higher risk-adjusted returns**

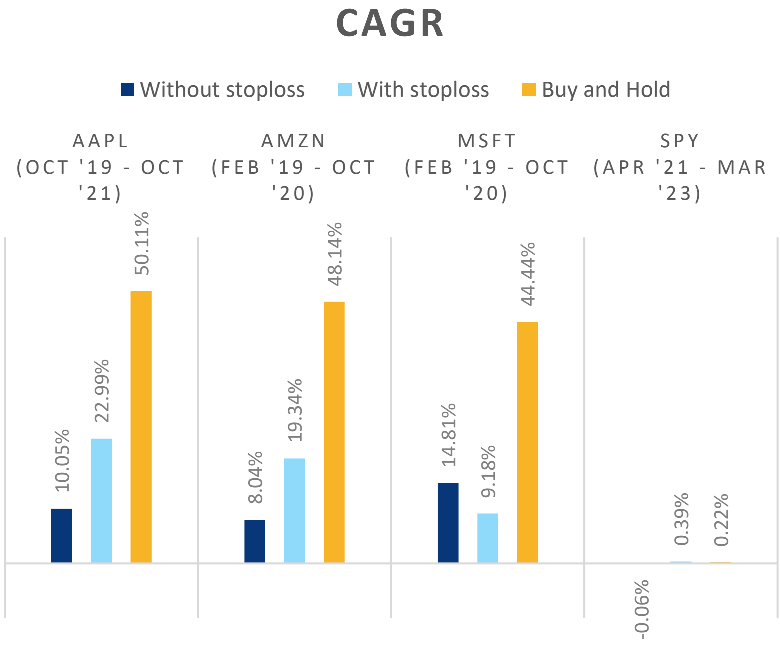
Transformer model has **lower maximum drawdown** than buy-and-hold strategy

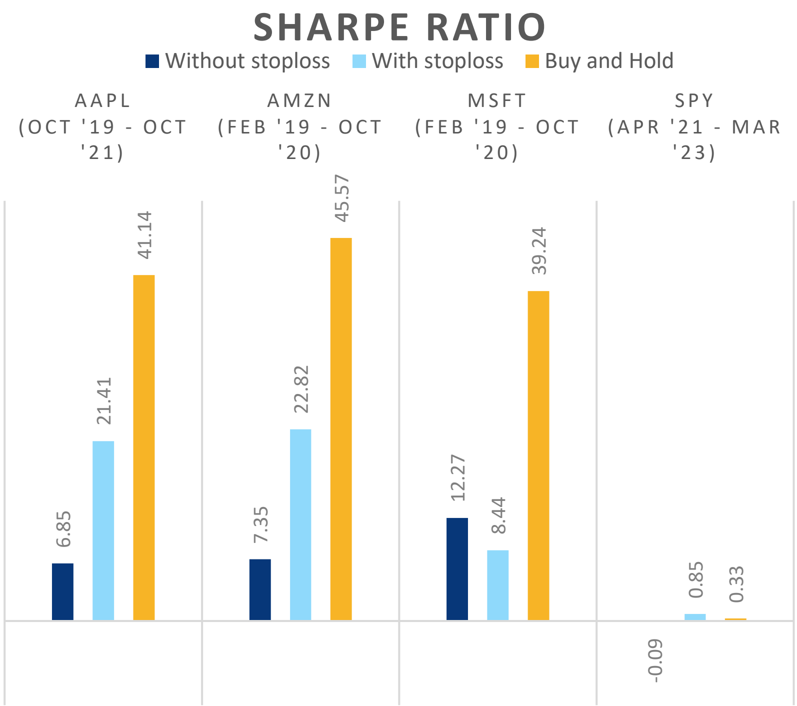
Evaluation of model trading performance on equity

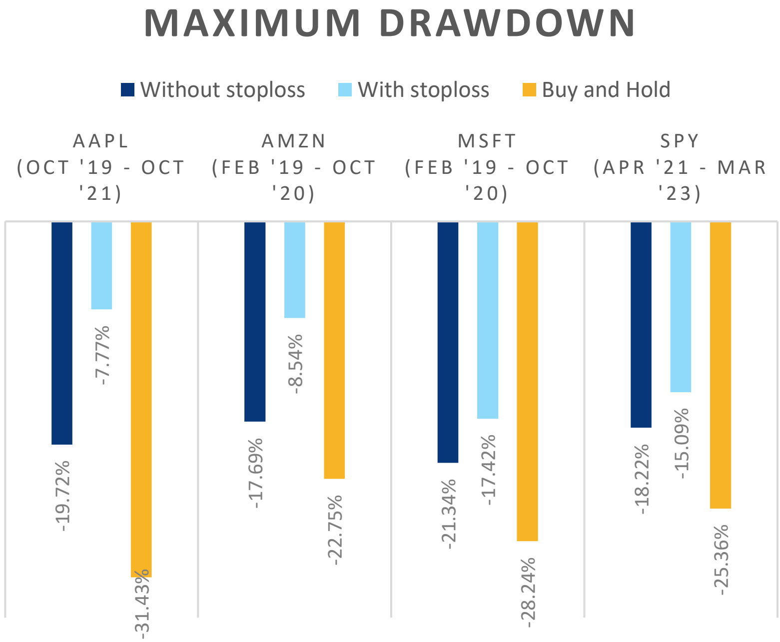
Equity: Technical + Fundamental model

Evaluation of model trading performance on equity

Equity: Technical + Macroeconomic model







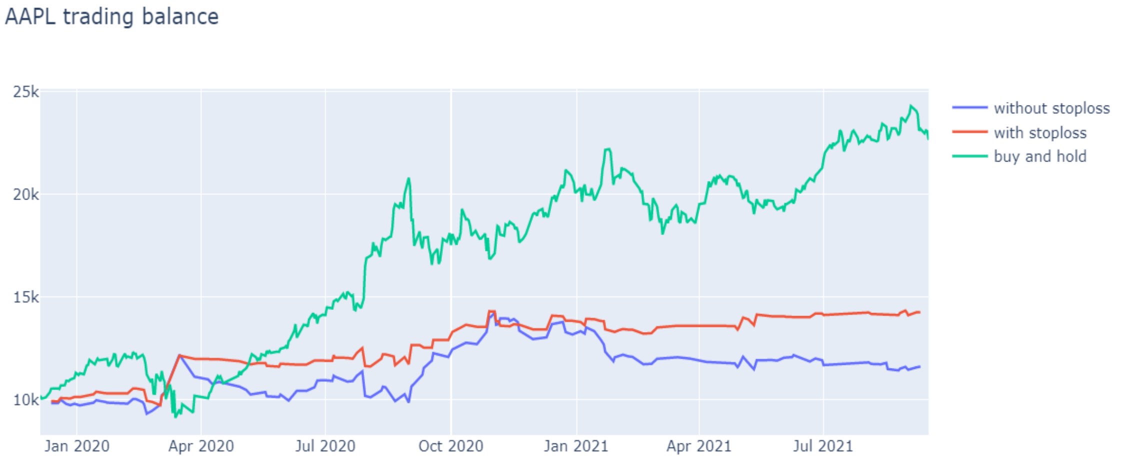
Key takeaway

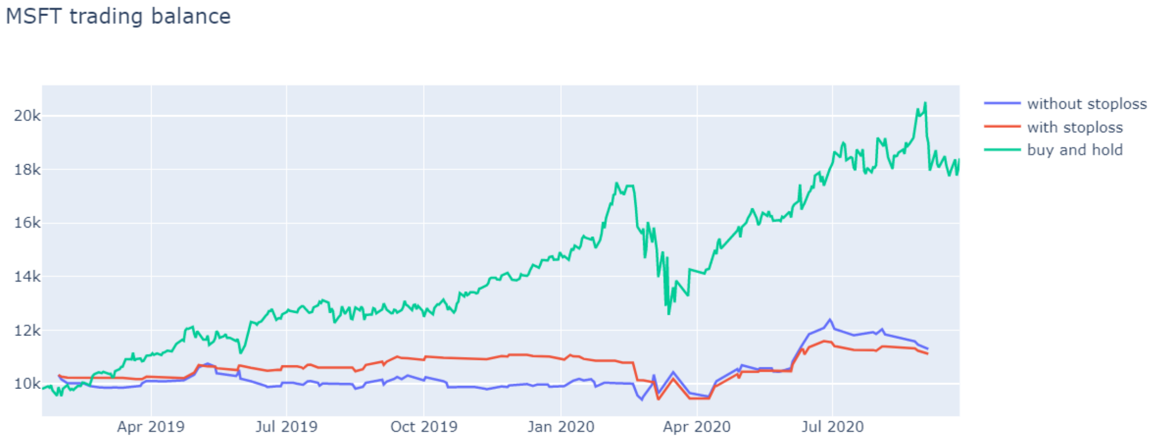
Transformer model was **not** able to outperform buy-and-hold strategy CAGR when using stoploss

Transformer model with stoploss mechanism produce **lower** risk-adjusted returns

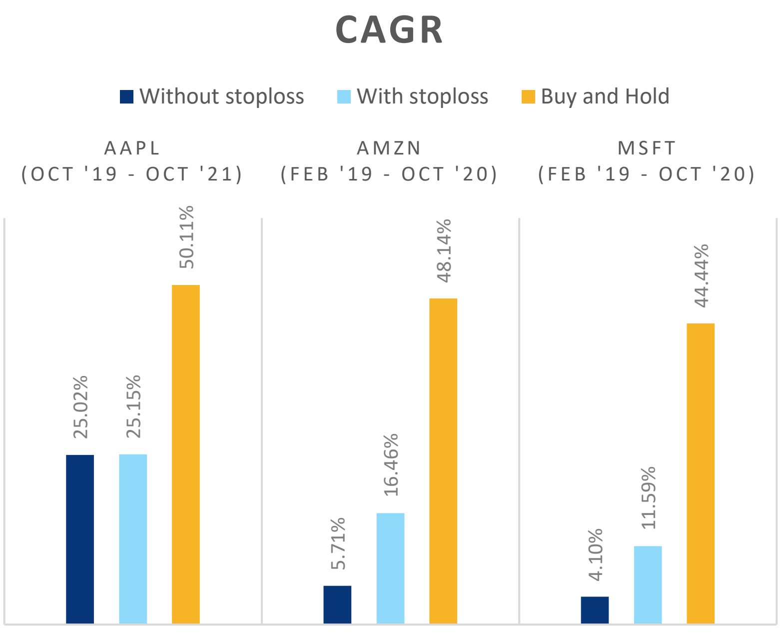
Transformer model has **lower maximum drawdown** than buy-and-hold strategy

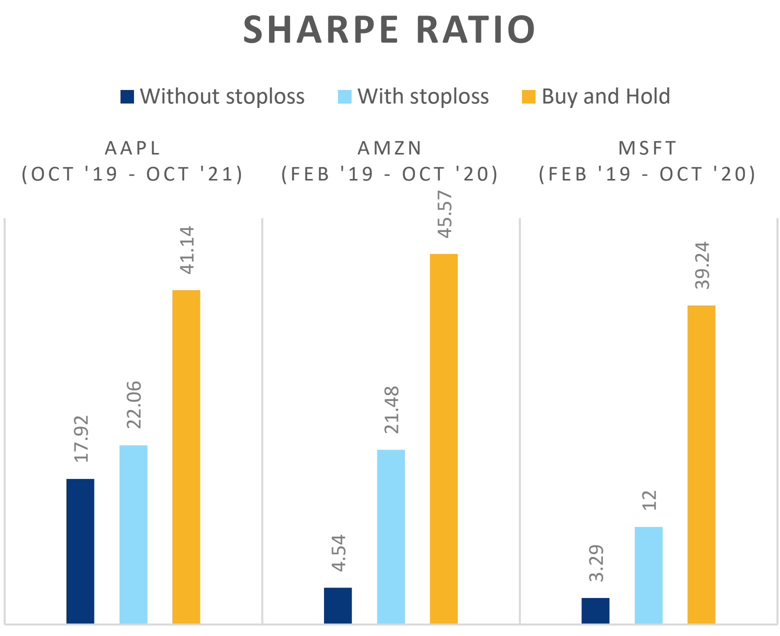
Equity: Technical + Macroeconomic model

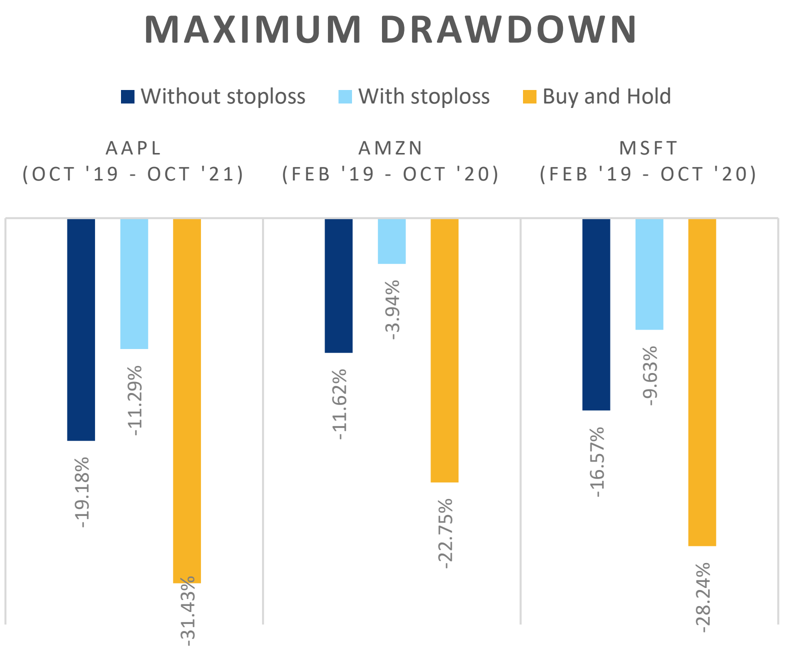




Equity: Technical + Value model







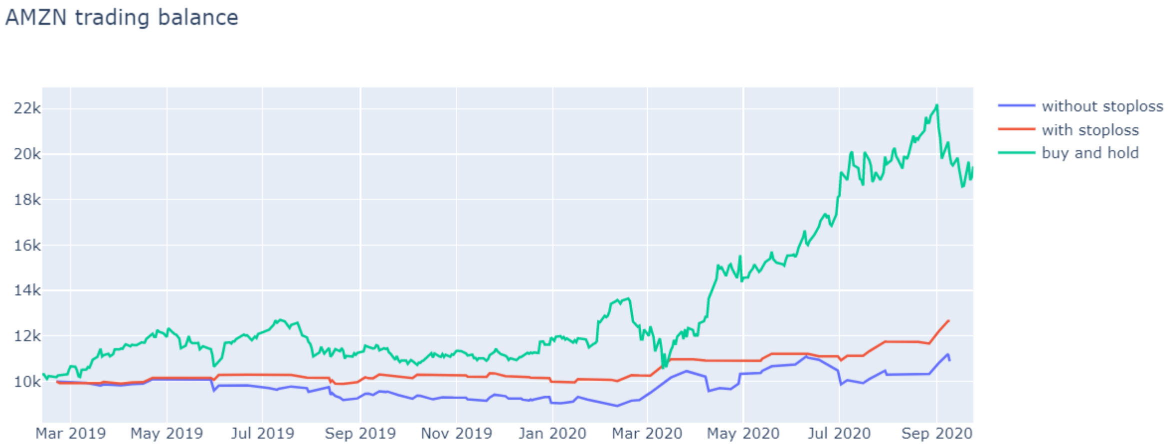
Key takeaway

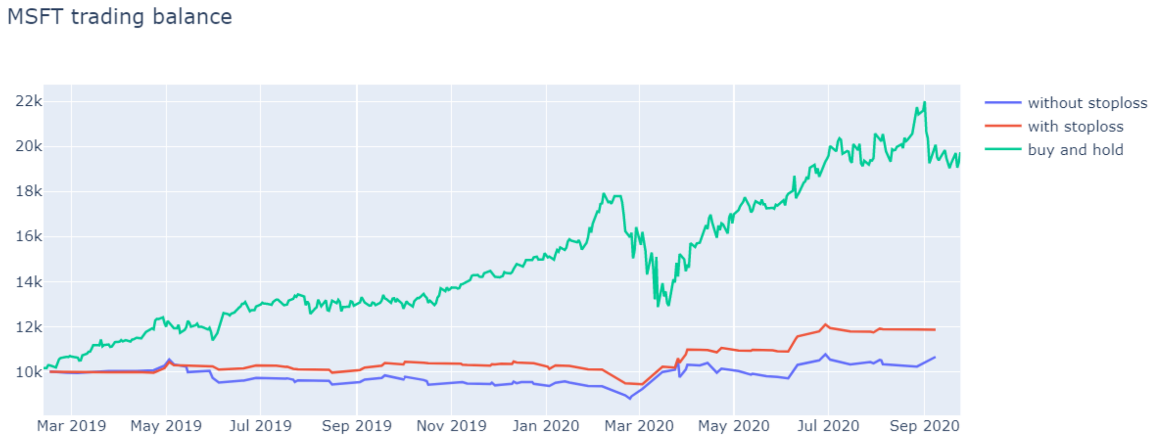
Transformer model was **not** able to outperform buy-and-hold strategy CAGR when using stoploss

Transformer model with stoploss mechanism produce **lower** risk-adjusted returns

Transformer model has **lower maximum drawdown** than buy-and-hold strategy

Equity: Technical + Value model

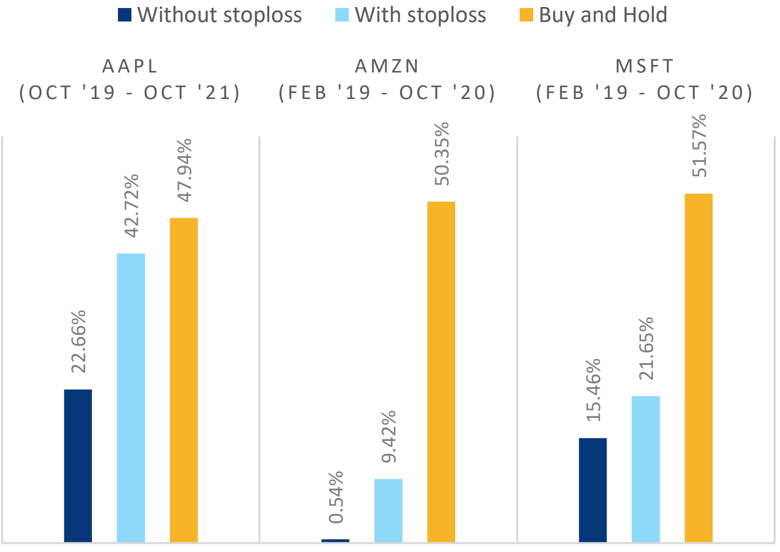




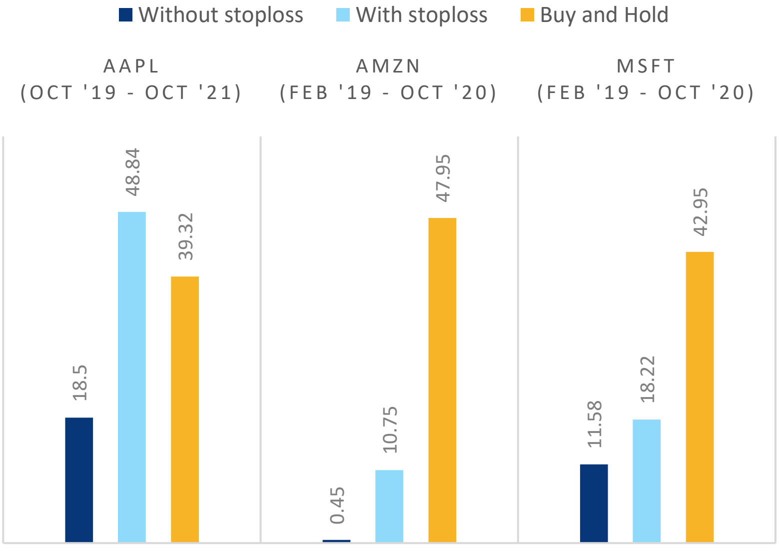
Evaluation of model trading performance on equity

Equity: Technical + Fundamental + Macroeconomic + Value model

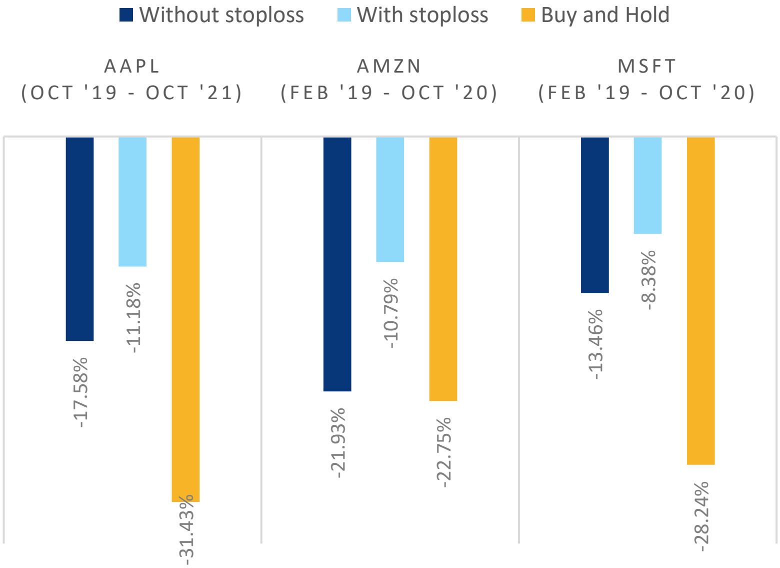
CAGR



SHARPE RATIO



MAXIMUM DRAWDOWN



Key takeaway

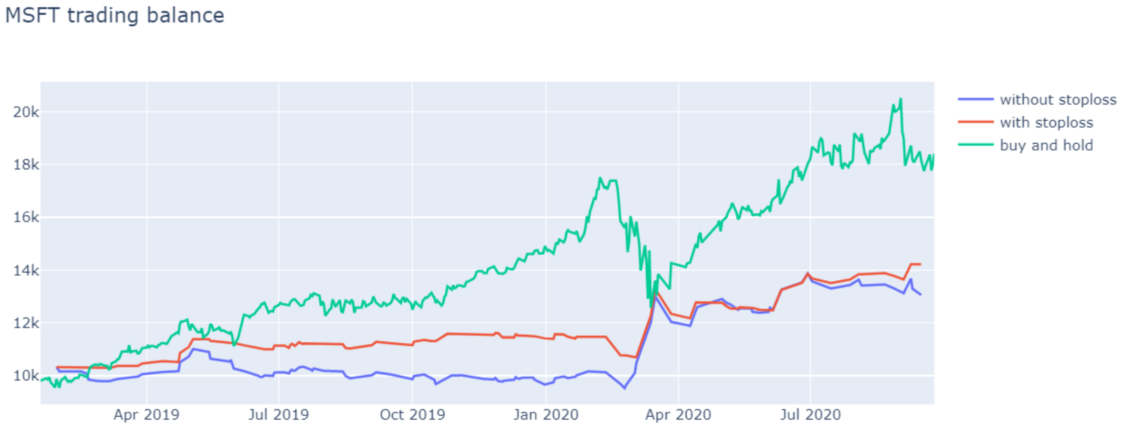
Transformer model was **not** able to outperform buy-and-hold strategy CAGR when using stoploss

Transformer model with stoploss mechanism produce **higher** risk-adjusted returns for AAPL but are fairly lower for the other equities

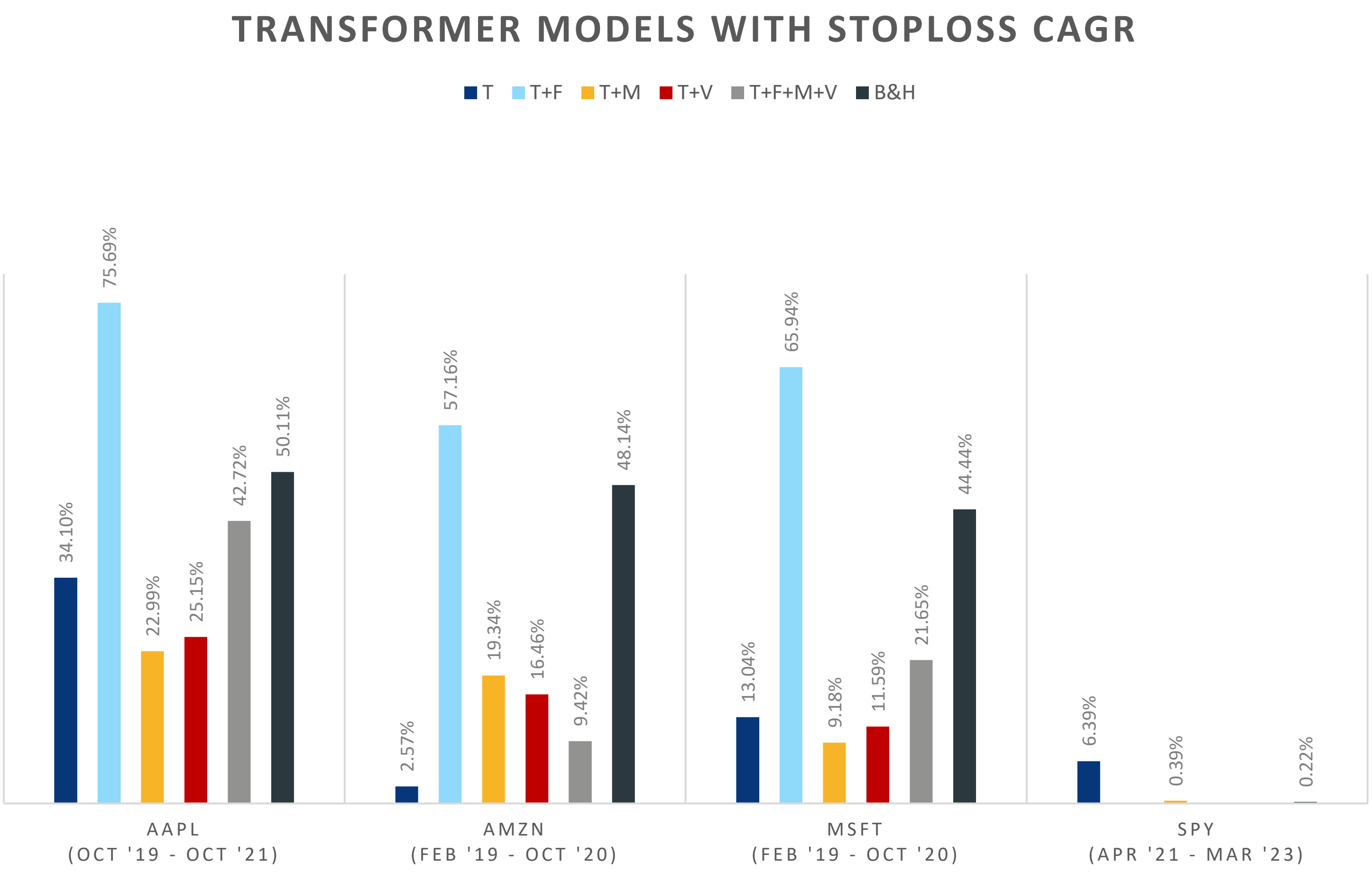
Transformer model has **lower maximum drawdown** than buy-and-hold strategy

Evaluation of model trading performance on equity

Equity: Technical + Fundamental + Macroeconomic + Value model



Comparison Between Equity Model Performances



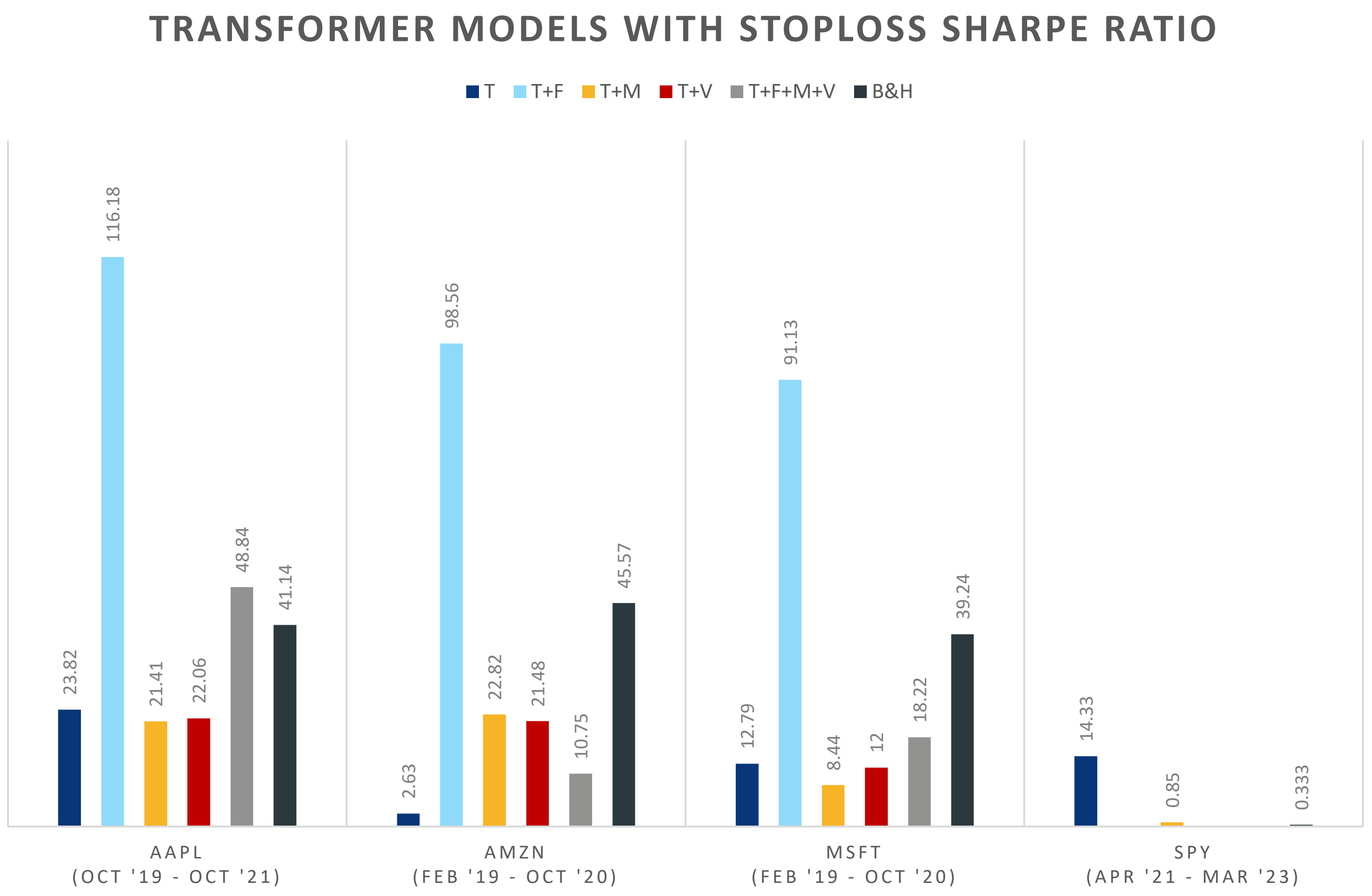
Key takeaway

Technical + Fundamental model is the only transformer model that is more profitable than buy-and-hold

Fundamental information introduce company specific intrinsic information that impacts the performance and investors' speculation

All models have lower maximum drawdown compared to the buy-and-hold

Comparison Between Equity Model Performances



Key takeaway

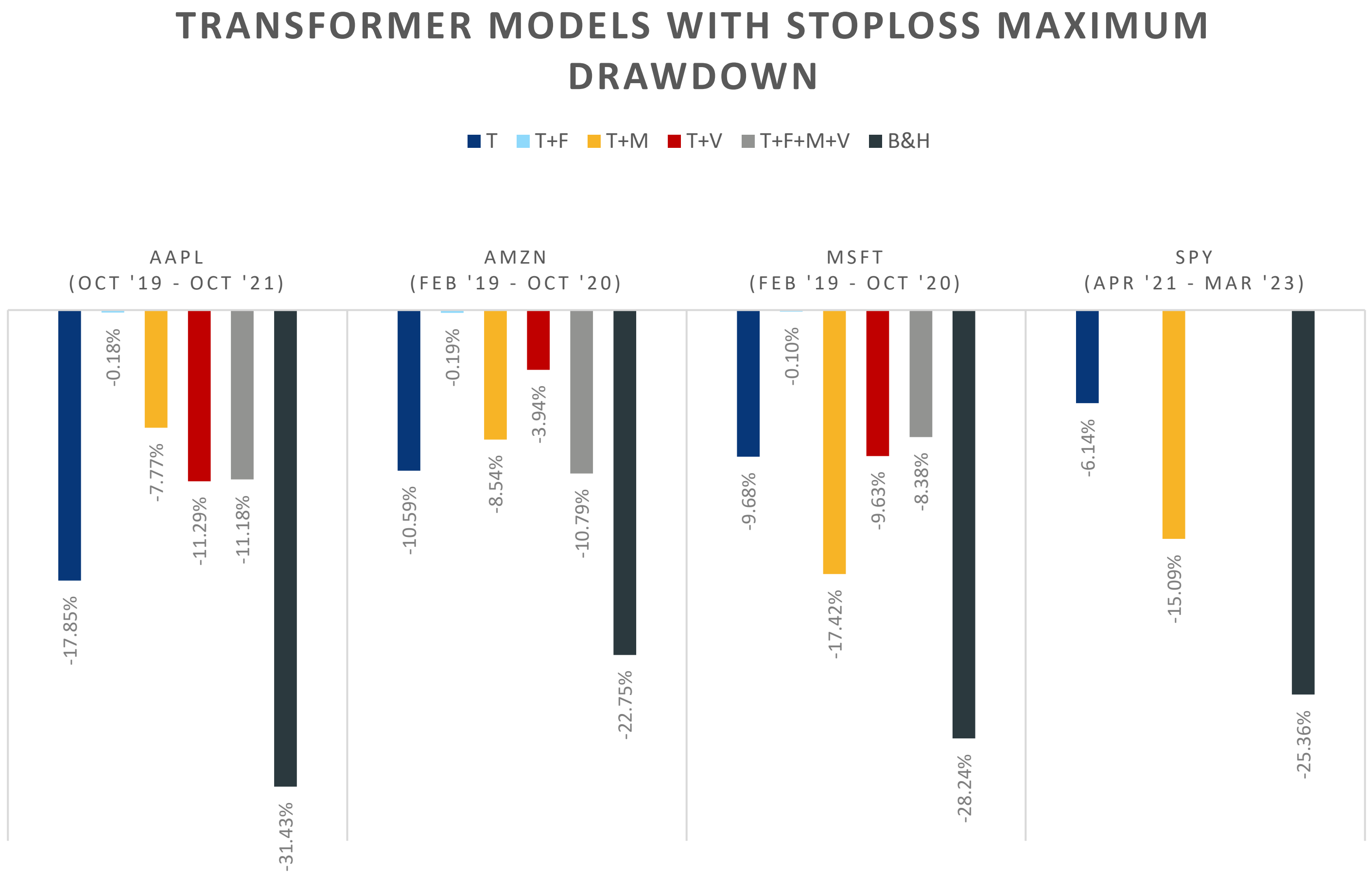
Technical + Fundamental model is the only transformer model that is more profitable than buy-and-hold

Fundamental information introduce company specific intrinsic information that impacts the performance and investors' speculation

All models have lower maximum drawdown compared to the buy-and-hold

Discussion

Comparison Between Equity Model Performances



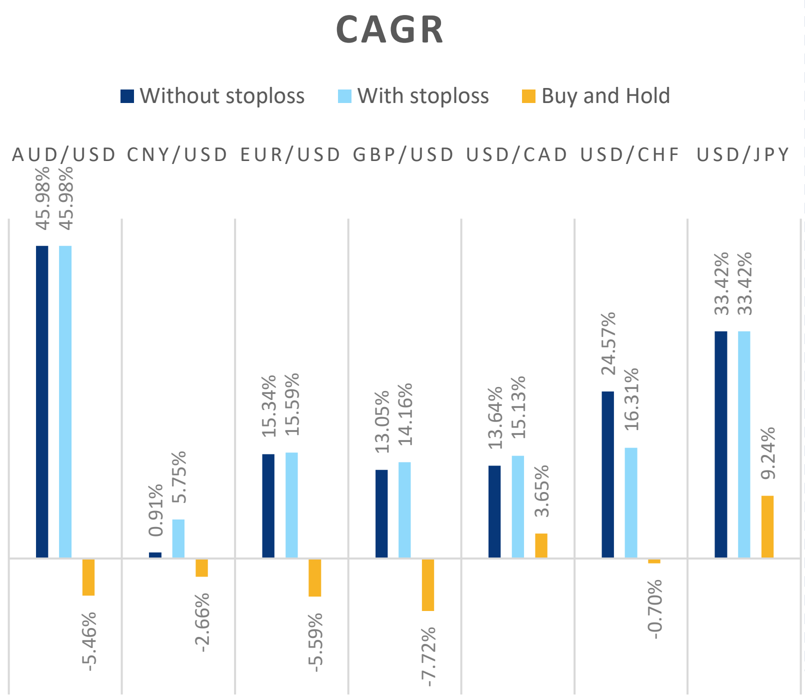
Key takeaway

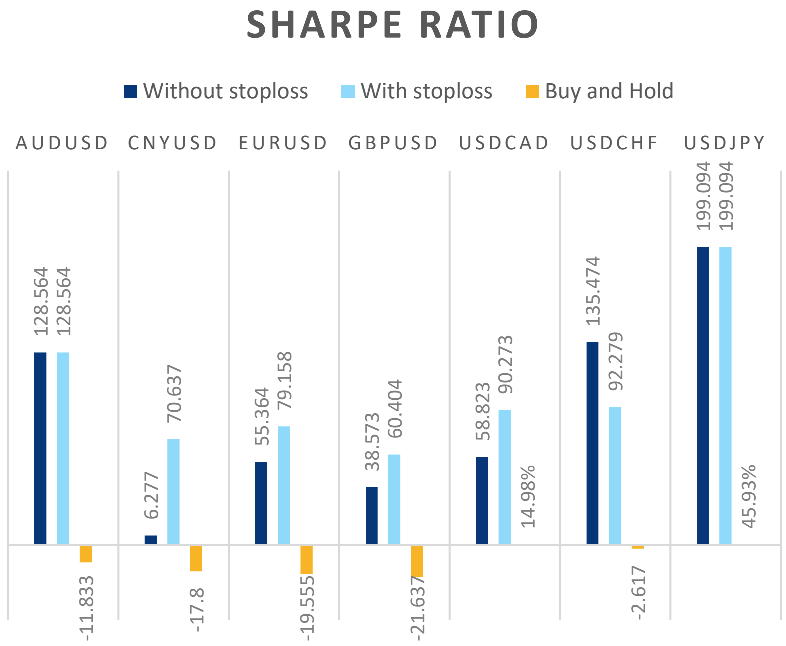
Technical + Fundamental model is the only transformer model that is more profitable than buy-and-hold

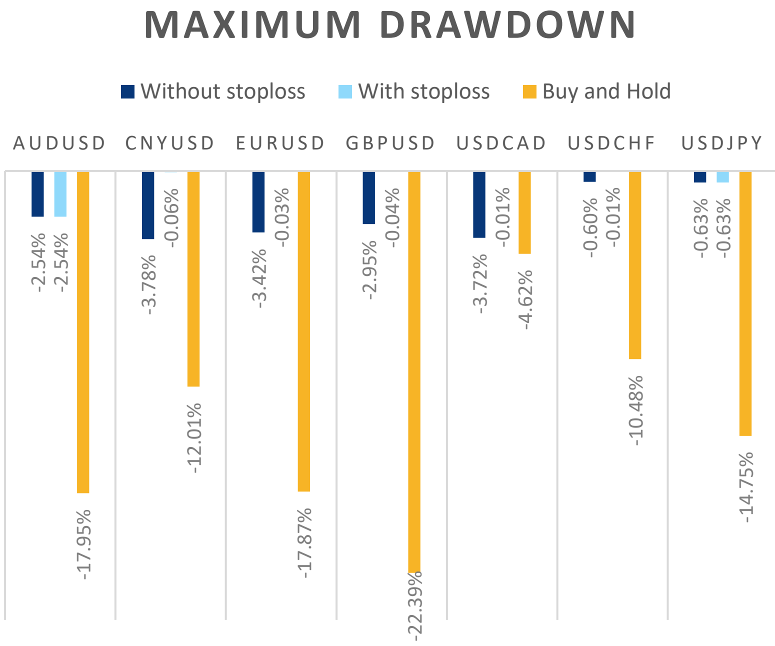
Fundamental information introduce company specific intrinsic information that impacts the performance and investors' speculation

All models have lower maximum drawdown compared to the buy-and-hold

FX: Technical model







Key takeaway

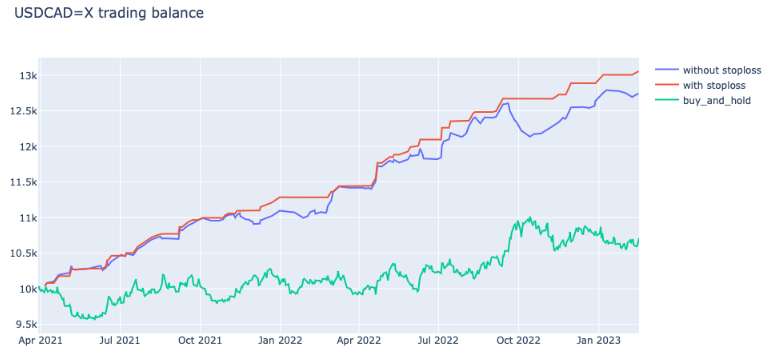
Transformer model **outperformed** buy-and-hold strategy CAGR **even without stoploss**

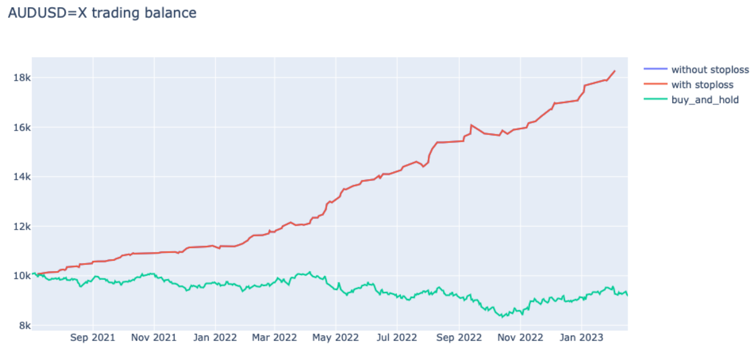
Transformer model was able to **produce higher risk-adjusted returns** across all currency pairs

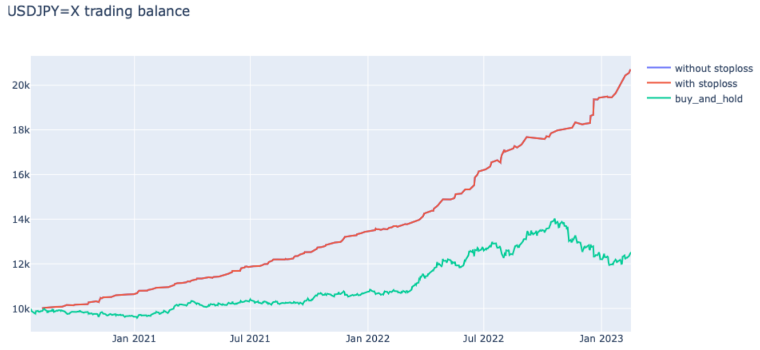
Transformer model has **lower maximum drawdown** than buy-and-hold strategy across all currency pairs

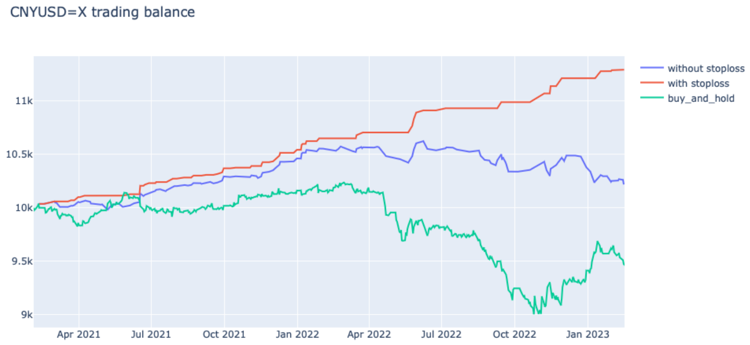
Evaluation of model trading performance on FX

FX: Technical model



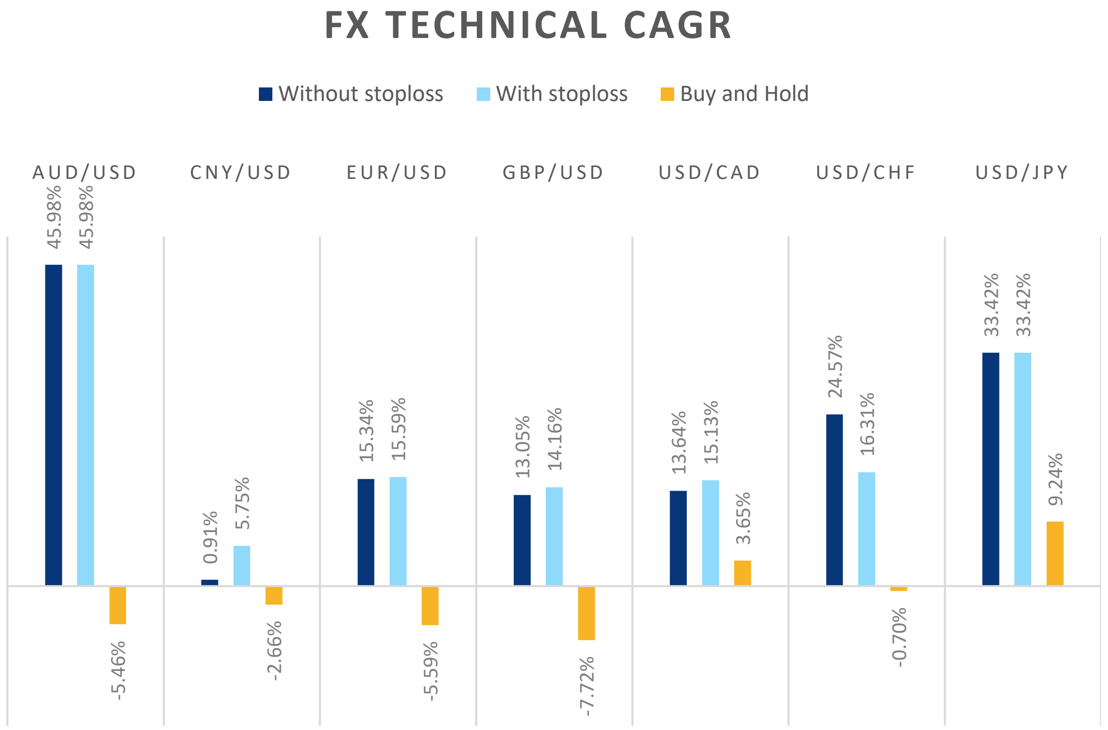


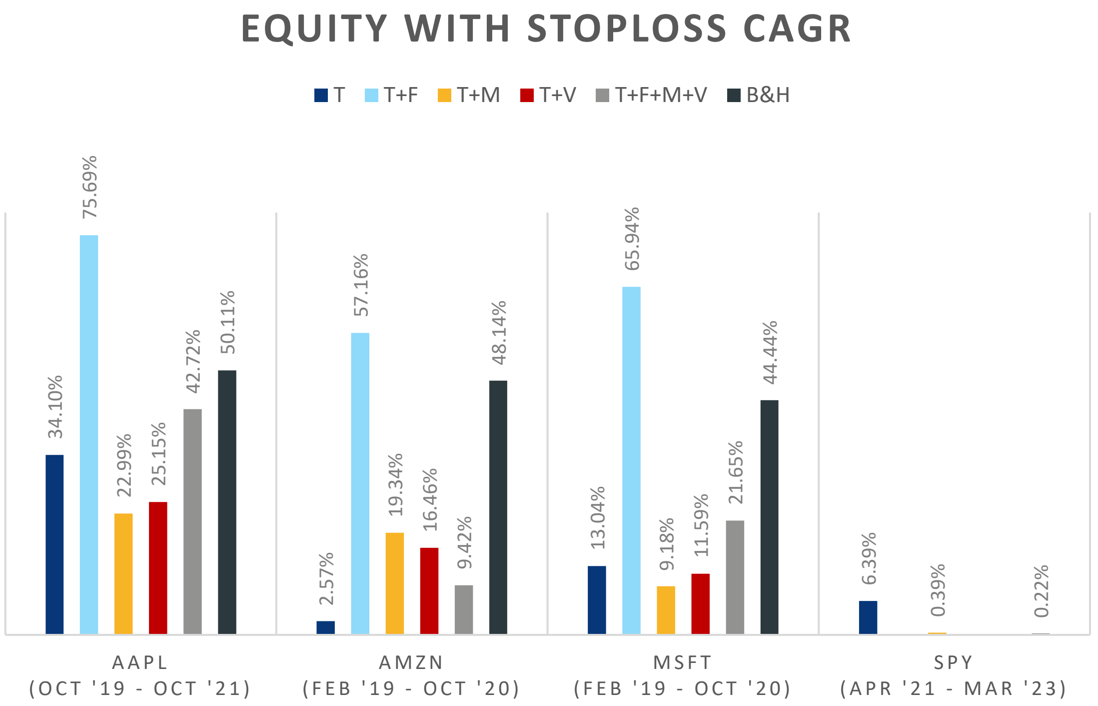




Discussion

Analysis of equities vs FX





Key question

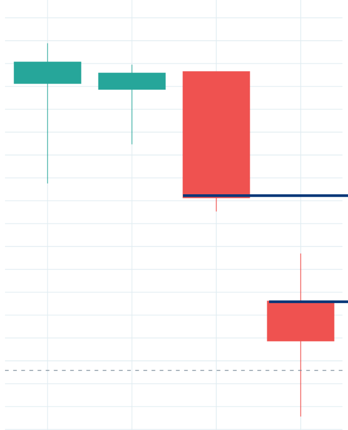
Why Transformer model on FX performs better than buy-and-hold consistently compared to equities?

Why Transformer model on FX performs better than buy-and-hold consistently compared to equities?

Different trading hours

FX	Equity
24 hours Monday to Friday	9 am to 4.30 pm Monday to Friday

Example of unexpected loss



Buy on closing price

Unexpected loss on opening price

External company-specific factors

Equity's prices are also subjected to idiosyncratic risk

Management

Supply chain disruption

Lawsuit

Change in regulations

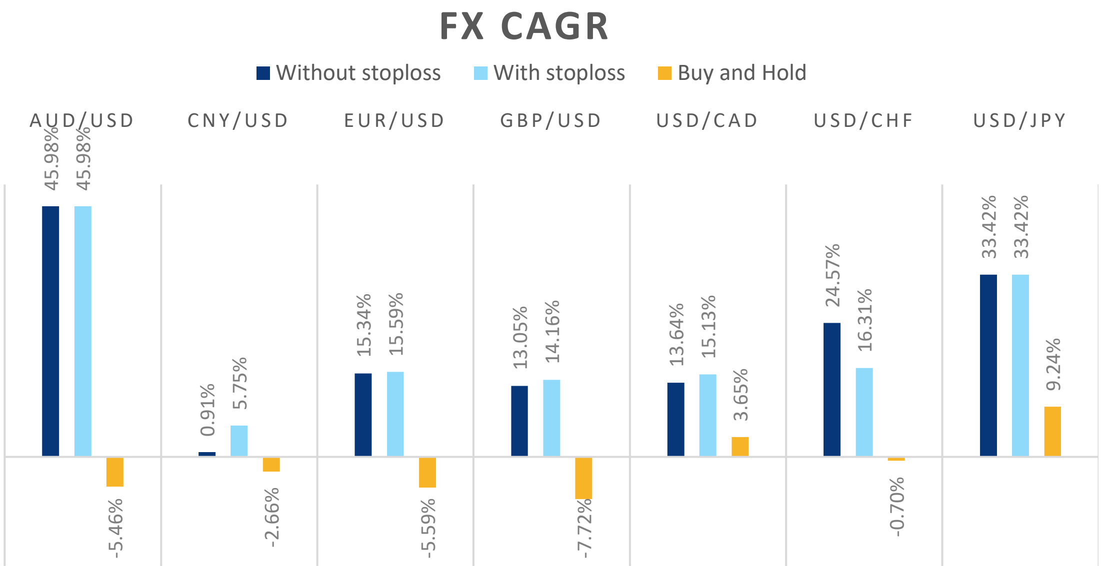
Why Transformer model on FX performs better than buy-and-hold consistently compared to equities?

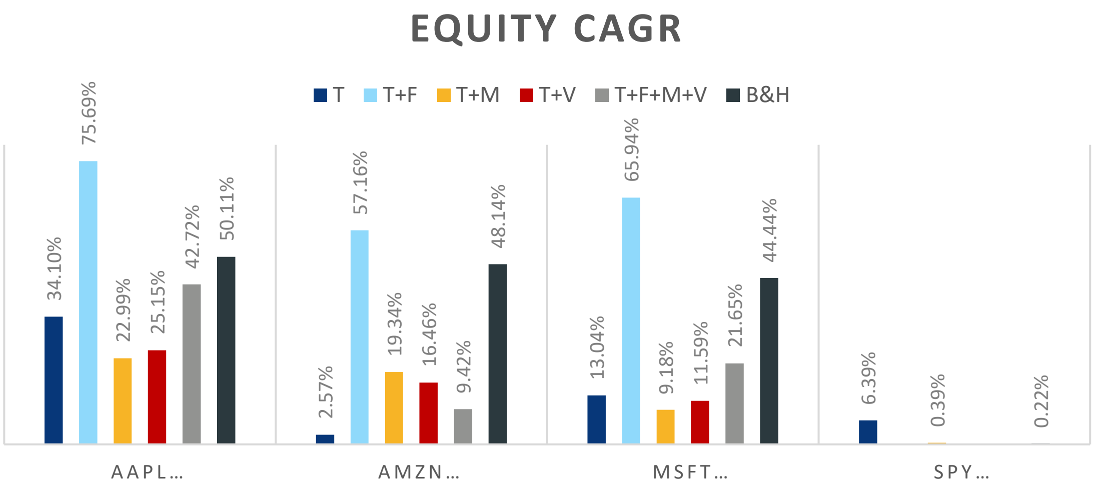
FX is characterized by stable prices

Price movements can induce arbitrage opportunities

FX is also less volatile than Equities. This can be seen from the:

- CAGR





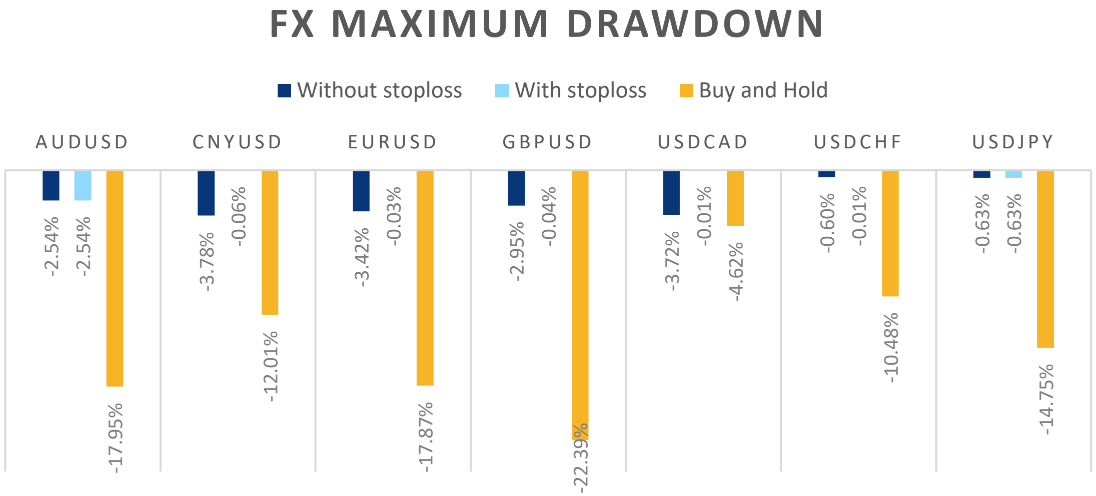
Why Transformer model on FX performs better than buy-and-hold consistently compared to equities?

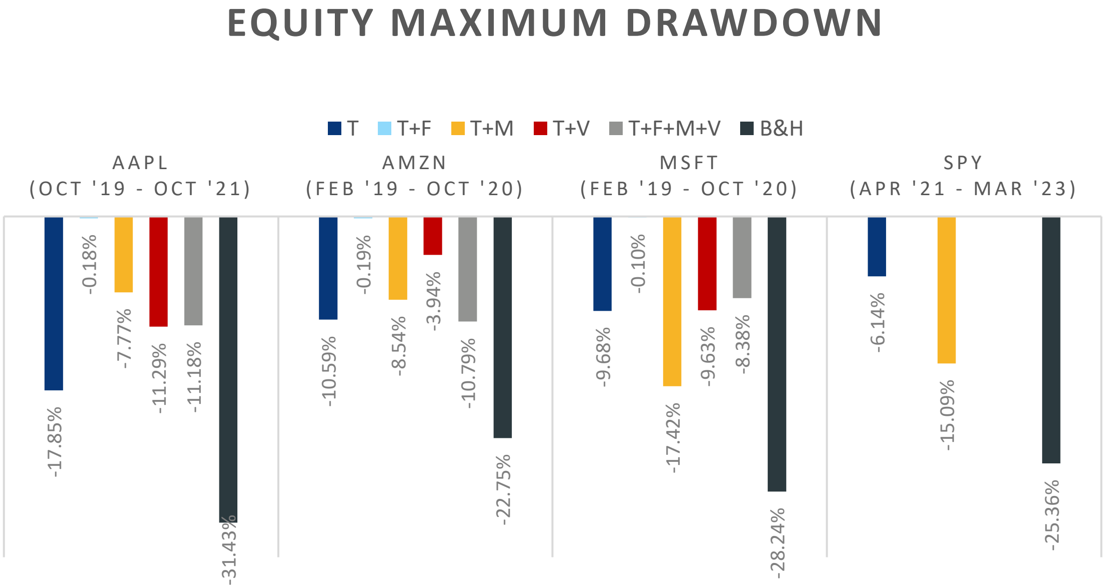
FX is characterized by stable prices

Price movements can induce arbitrage opportunities

FX is also less volatile than Equities. This can be seen from the:

- CAGR
- Maximum drawdown





1

Introduction

Motivation
Lit. Review
Objectives

2

Methodology

Design
Implementation
Testing

3

Evaluation

Equity Models
FX Models
Discussion

4

Conclusion

Technical Accomplishments

Technical accomplishments

Transformer model for trading

- Successfully developed a transformer model that can follow the return changes
- Formulated a trading strategy based on the model output

Optimizing transformer model for trading

Technical FX model
19.4% in excess from buy-and-hold CAGR

Technical + Fundamental Equity model
16.3% in excess from buy-and-hold CAGR

Appendix

References

[1] A. West, "New research reveals machine learning is maturing in finance," Refinitiv, 28 October 2020. [Online]. Available: https://www.refinitiv.com/perspectives/ai-digitalization/machine-learning-new-research-reveals-is-maturing-in-finance/. [Accessed 12 September 2022].

[2] "scalefocus," July 2022. [Online]. Available: https://www.scalefocus.com/blog/how-financial-institutions-can-leverage-the-power-of-machine-learning. [Accessed 12 September 2022].

[3] D. Conko, "Neural Networks: Forecasting Profits," Investopedia, 9 May 2022. [Online]. Available: https://www.investopedia.com/articles/trading/06/neuralnetworks.asp#:~:text=Neural%20networks%20do%20not%20make,using%20traditional%20technical%20analysis%20methods. [Accessed 12 Septmeber 2022].

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," arXiv, 2017.

[5] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko, "End-to-End Object Detection with Transformers," CoRR, 2020.

[6] P. Agrawal, H. O. Bansal, A. R. Gautam, O. P. Mahela and B. Khan, "Transformer-based time series prediction of the maximum power point for solar photovoltaic cells," Energy Science & Engineering, vol. 10, no. 9, pp. 3397-2410, 2022.

[7] H. Hamoudi and M. A. Elseif, "Stock Market Prediction using CNN and LSTM," Stanford University, 2021.

[8] Jane Street, "Jane Street Market Prediction | Kaggle," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/competitions/jane-street-market-prediction/overview. [Accessed 12 September 2022].

[9] Yahoo Finance, [Online]. Available: https://finance.yahoo.com/. [Accessed 12 September 2022].

[10] "Bloomberg Terminal," [Online]. Available: https://www.bloomberg.com/professional/solution/bloomberg-terminal/. [Accessed 12 September 2022].

[11] "EODHD APIs," [Online]. Available: https://eodhistoricaldata.com/. [Accessed 27 December 2022].

[12] Environmental Systems Research Institute, "How Extra trees classification and regression algorithm works," Environmental Systems Research Institute, [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/geoai/how-extra-tree-classification-and-regression-works.htm. [Accessed 18 April 2023].

References

[13] Environmental Systems Research Institute, "How XGBoost algorithm works," Environmental Systems Research Institute, [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/geoai/how-xgboost-works.htm. [Accessed 18 April 2023].

[14] NumPy, [Online]. Available: https://numpy.org/. [Accessed 12 September 2022].

[15] Pandas, [Online]. Available: https://pandas.pydata.org/. [Accessed 12 September 2022].

[16] Seaborn, [Online]. Available: https://seaborn.pydata.org/. [Accessed 12 September 2022].

[17] matplotlib, [Online]. Available: https://matplotlib.org/. [Accessed 12 September 2022].

[18] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller and O. Grisel, "{API} design for machine learning software: experiences from the scikit-learn," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013.

[19] F. Chollet, *Keras,* 2015.

[20] S. M. Kazemi, R. Goel, S. Eghbali, J. Ramanan, J. Sahota, S. Thakur, S. Wu, C. Smyth, P. Poupart and M. Brubaker, "Time2Vec: Learning a Vector Representation of Time," Ahsanullah University of Science and Technology Dhaka, Bangladesh, 11 July 2019. [Online]. Available: https://arxiv.org/abs/1907.05321. [Accessed 20 January 2023].

[21] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optmization," 2015. [Online]. Available: https://arxiv.org/pdf/1412.6980.pdf. [Accessed 25 January 2023].

[22] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics,* vol. 35, no. 1, pp. 73-101, 1964.

[23] Interactive Brokers, "Commissions | Interactive Brokers Hong Kong Limited," [Online]. Available: https://www.interactivebrokers.com.hk/en/index.php?f=1590. [Accessed 15 January 2023].

[24] Investing.com, [Online]. Available: https://www.investing.com/. [Accessed 2 January 2023].

References

[25] python, [Online]. Available: https://www.python.org/. [Accessed 12 September 2022].

[26] Visual Studio Code, [Online]. Available: https://code.visualstudio.com/. [Accessed 12 September 2022].

[27] Microsoft, [Online]. Available: https://www.office.com/. [Accessed 12 September 2022].

[28] Github, [Online]. Available: github.com. [Accessed 12 September 2022].

[30] Oracle, "What is Natural Language Processing (NLP)? | Oracle Hong Kong SAR, PRC," Oracle, [Online]. Available: https://www.oracle.com/hk/artificial-intelligence/what-is-natural-language-processing/#:~:text=Natural%20Language%20Processing%3F-,Natural%20Language%20Processing%20(NLP)%20Defined,natural%20language%20text%20or%20voice. [Accessed 12 September 2022].

[31] S. Goled, "Why Transformers Are Increasingly Becoming As Important As RNN And CNN?," Analytics India Mag, 17 March 2021. [Online]. Available: https://analyticsindiamag.com/why-transformers-are-increasingly-becoming-as-important-as-rnn-and-cnn/. [Accessed 12 September 2022].

[32] A. Lu, A. Parulekar and H. Xu, "Cluster-Based Statistical Arbitrage Strategy," 10 June 2018.

[33] J. Fernando and C. Potters, "Relative Strength Index (RSI) Indicator Explained With Formula," Investopedia, [Online]. Available: https://www.investopedia.com/terms/r/rsi.asp. [Accessed 25 January 2023].

[34] "Companies Market Cap," [Online]. Available: https://companiesmarketcap.com/usa/largest-companies-in-the-usa-by-market-cap/. [Accessed 10 February 2023].

[35] G. Siddhad, A. Gupta, D. P. Dogra and P. P. Roy, "Efficacy of Transformer Netwroks for Classification of Raw EEG Data," arXiv, 11 February 2022. [Online]. Available: https://arxiv.org/pdf/2202.05170.pdf. [Accessed 31 August 2022].

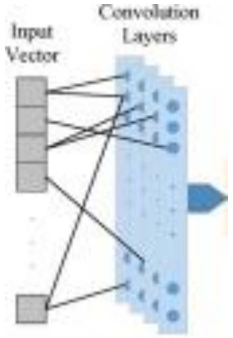
[36] B. Dolan, "MACD Indicator Explained, with Formula, Examples, and Limitations," Investopedia, [Online]. Available: https://www.investopedia.com/terms/m/macd.asp. [Accessed 31 August 2022].

[37] The Hong Kong University of Science and Technology Information Technology Services Centre, "X-GPU CLUSTER," [Online]. Available: https://itsc.hkust.edu.hk/services/academic-teaching-support/high-performance-computing/x-gpu-cluster. [Accessed 27 December 2022].

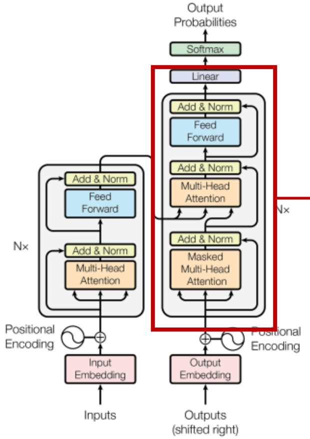
Future works

More data

86308 parameters for each encoder
But only,
4000 datapoints for each model



More complex input projection layer
Perform learned durational feature extraction



Adding decoder module
Predict future values for multiple days

Min-Max equation

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \rightarrow x = \frac{x_{scaled}\left(x_{max} - x_{min}\right)}{x - x_{min}}$$

Max-abs equation

$$x_{scaled} = \frac{x}{|\max(x)|} \rightarrow x = x_{scaled} \times |\max(x)|$$

Rolling geometric mean

$$gmean_0 = [(1 + x_0) \times (1 + x_{-1}) \times (1 + x_{-2}) \times (1 + x_{-3}) \times (1 + x_{-4})]^{\frac{1}{5}} - 1$$

$$x_0 = \frac{[g_{mean} + 1]^5}{(1 + x_{-1}) \times (1 + x_{-2}) \times (1 + x_{-3}) \times (1 + x_{-4})} - 1$$

Percentage Change

$$percentage\_change_0 = \frac{close_0 - close_{-1}}{close_{-1}}$$

$$close_0 = [percentage\_change_0 \times close_{-1}] + close_{-1}$$

GitHub link

https://github.com/antonioxav/FYP_model

Time2Vec

Time2Vec: We propose *Time2Vec*, a representation for time which has the three identified properties. For a given scalar notion of time τ, Time2Vec of τ, denoted as $\mathbf{t2v}(\tau)$, is a vector of size $k+1$ defined as follows:

$$\mathbf{t2v}(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & \text{if } i = 0. \\ \mathcal{F}(\omega_i \tau + \varphi_i), & \text{if } 1 \leq i \leq k. \end{cases} \tag{1}$$

where $\mathbf{t2v}(\tau)[i]$ is the i^{th} element of $\mathbf{t2v}(\tau)$, $\mathcal{F}$ is a periodic activation function, and ω_is and φ_is are learnable parameters. Given the prevalence of vector representations for different tasks, a vector representation for time makes it easily consumable by different architectures. We chose $\mathcal{F}$ to be the sine function in our experiments[2] but we do experiments with other periodic activations as well. When $\mathcal{F} = \sin$, for $1 \leq i \leq k$, ω_i and φ_i are the frequency and the phase-shift of the sine function.

The period of $\sin(\omega_i \tau + \varphi_i)$ is $\frac{2\pi}{\omega_i}$, *i.e.* it has the same value for τ and $\tau + \frac{2\pi}{\omega_i}$. Therefore, a sine function helps capture periodic behaviors without the need for feature engineering. For instance, a sine function $\sin(\omega\tau + \varphi)$ with $\omega = \frac{2\pi}{7}$ repeats every 7 days (assuming τ indicates days) and can be potentially used to model weekly patterns. Furthermore, unlike other basis functions which may show strange behaviors for extrapolation (see, *e.g.*, [49]), sine functions are expected to work well for extrapolating to future and out of sample data [57]. The linear term represents the progression of time and can be used for capturing non-periodic patterns in the input that depend on time. Proposition 1 establishes the invariance of Time2Vec to time rescaling. The proof is in Appendix D.

Source: https://arxiv.org/pdf/1907.05321.pdf

Equity Technical Model Result

Stock Name	Evaluation Metrics	Without stoploss	With stoploss	Buy and hold
AAPL	Win rate	65.38%	65.38%	52.99%
	CAGR	33.67%	34.10%	47.94%
	Sharpe Ratio	23.32	23.82	39.32
	Max. Drawdown	-18.28%	-17.85%	-31.43%
	Trade turnover	5.69 days	5.69 days	1.46 days
AMZN	Win rate	50.49%	25.74%	54.88%
	CAGR	-13.26%	2.57%	50.35%
	Sharpe Ratio	-9.72	2.63	47.95
	Max. Drawdown	-23.81%	-10.59%	-22.75%
	Trade turnover	5.673 days	5.673 days	1.44 days
MSFT	Win rate	50.00%	27.27%	58.58%
	CAGR	12.18%	13.04%	51.57%
	Sharpe Ratio	9.72	12.79	42.94
	Max. Drawdown	-16.47%	-9.68%	-28.24%
	Trade turnover	5.22 days	5.22 days	1.44 days
SPY	Win rate	53.33%	25.18%	50.213
	CAGR	4.26%	6.39%	0.223%
	Sharpe Ratio	6.41	14.33	0.333
	Max. Drawdown	-8.31%	-6.14%	-25.361%
	Trade turnover	4.837 days	4.837 days	1.46 days

Equity Technical + Fundamental Model Result

Stock Name	Evaluation Metrics	Without stoploss	With stoploss	Buy and hold
AAPL	Win rate	55.32%	34.75%	52.99%
	CAGR	15.40%	75.69%	47.94%
	Sharpe Ratio	11.15	116.18	39.32
	Max. Drawdown	-18.61%	-0.18%	-31.43%
	Trade turnover	4.37 days	4.37 days	1.46 days
AMZN	Win rate	46.59%	23.30%	54.88%
	CAGR	-3.98%	57.16%	50.35%
	Sharpe Ratio	-3.25	98.56	47.95
	Max. Drawdown	-21.42%	-0.19%	-22.75%
	Trade turnover	3.31 days	3.31 days	1.44 days
MSFT	Win rate	61.60%	37.60%	58.58%
	CAGR	41.87%	65.94%	51.57%
	Sharpe Ratio	38.75	91.13	42.94
	Max. Drawdown	-7.05%	-0.099%	-28.24%
	Trade turnover	4.64 days	4.64 days	1.44 days

Equity Technical + Macroeconomic Model Result

Stock Name	Evaluation Metrics	Without stoploss	With stoploss	Buy and hold
AAPL	Win rate	50.70%	29.58%	52.99%
	CAGR	10.05%	22.99%	47.94%
	Sharpe Ratio	6.85	21.41	39.32
	Max. Drawdown	-19.72%	-7.77%	-31.43%
	Trade turnover	5.69 days	5.69 days	1.46 days
AMZN	Win rate	52.89%	30.58%	54.88%
	CAGR	8.04%	19.34%	50.35%
	Sharpe Ratio	7.35	22.82	47.95
	Max. Drawdown	-17.69%	-8.54%	-22.75%
	Trade turnover	5.673 days	5.673 days	1.44 days
MSFT	Win rate	51.20%	21.08%	58.58%
	CAGR	14.81%	9.18%	51.57%
	Sharpe Ratio	12.27	8.44	42.94
	Max. Drawdown	-21.34%	-17.42%	-28.24%
	Trade turnover	3.86 days	3.86 days	1.44 days
SPY	Win rate	55.28%	23.60%	50.213
	CAGR	-0.06%	0.39%	0.22%
	Sharpe Ratio	-0.09	0.85	0.33
	Max. Drawdown	-18.22%	-15.09%	-25.36%
	Trade turnover	4.08 days	4.08 days	1.46 days

Equity Technical + Value Model Result

Stock Name	Evaluation Metrics	Without stoploss	With stoploss	Buy and hold
AAPL	Win rate	55.13%	30.77%	52.99%
	CAGR	25.02%	25.15%	47.94%
	Sharpe Ratio	17.92	22.06	39.32
	Max. Drawdown	-19.18%	-11.29%	-31.43%
	Trade turnover	3.95 days	3.95 days	1.46 days
AMZN	Win rate	50.00%	23.00%	54.88%
	CAGR	5.71%	16.46%	50.35%
	Sharpe Ratio	4.54	21.48	47.95
	Max. Drawdown	-11.62%	-3.94%	-22.75%
	Trade turnover	5.67 days	5.67 days	1.44 days
MSFT	Win rate	52.69%	33.33%	58.58%
	CAGR	4.11%	11.59%	51.57%
	Sharpe Ratio	3.29	12	42.94
	Max. Drawdown	-16.57%	-9.63%	-28.24%
	Trade turnover	6.14 days	6.14 days	1.44 days

Equity Technical + Fundamental + Macroeconomic + Value Model Result

Stock Name	Evaluation Metrics	Without stoploss	With stoploss	Buy and hold
AAPL	Win rate	56.21%	34.91%	52.99%
	CAGR	22.66%	42.72%	47.94%
	Sharpe Ratio	18.50	48.84	39.32
	Max. Drawdown	-17.58%	-11.18%	-31.43%
	Trade turnover	3.81 days	3.81 days	1.46 days
AMZN	Win rate	50.00%	28.76%	54.88%
	CAGR	0.54%	9.42%	50.35%
	Sharpe Ratio	0.45	10.75	47.95
	Max. Drawdown	-21.93%	-10.79%	-22.75%
	Trade turnover	4.10 days	4.10 days	1.44 days
MSFT	Win rate	55.37%	32.23%	58.58%
	CAGR	15.46%	21.65%	51.57%
	Sharpe Ratio	11.58	18.22	42.94
	Max. Drawdown	-13.46%	-8.38%	-28.24%
	Trade turnover	4.88 days	4.88 days	1.44 days

Foreign Exchange Technical Model Training Result

Currency pair name	Evaluation metrics	Without stoploss	With stoploss	Buy and hold
AUDUSD	Win rate	77.612%	77.612%	48.471%
	CAGR	45.979%	45.979%	-5.457%
	Sharpe Ratio	128.564	128.564	-11.833
	Max. Drawdown	-2.537%	-2.537%	-17.952%
	Trade turnover	4.291	4.291	1.398
CNYUSD	Win rate	53.896%	31.169%	50.487%
	CAGR	0.913%	5.753%	-2.658%
	Sharpe Ratio	6.277	70.637	-17.800
	Max. Drawdown	-3.781%	-0.055%	-12.011%
	Trade turnover	4.773	4.773	1.410
EURUSD	Win rate	70.161%	42.742%	47.942%
	CAGR	15.341%	15.592%	-5.590%
	Sharpe Ratio	55.364	79.158	-19.555
	Max. Drawdown	-3.424%	-0.030%	-17.866%
	Trade turnover	5.355	5.355	1.395
GBPUSD	Win rate	65.812%	32.479%	47.541%
	CAGR	13.049%	14.161%	-7.716%
	Sharpe Ratio	38.573	60.404	-21.637
	Max. Drawdown	-2.952%	-0.035%	-22.394%
	Trade turnover	5.701	5.701	1.395

Foreign Exchange Technical Model Training Result

Currency pair name	Evaluation metrics	Without stoploss	With stoploss	Buy and hold
USDCAD	Win rate	64.748%	35.971%	50.405%
	CAGR	13.640%	15.125%	3.654%
	Sharpe Ratio	58.823	90.273	14.982%
	Max. Drawdown	-3.718%	-0.011%	-4.618%
	Trade turnover	4.892	4.892	1.395
USDCHF	Win rate	85.714%	49.107%	51.822%
	CAGR	24.565%	16.306%	-0.704%
	Sharpe Ratio	135.474	92.279	-2.617
	Max. Drawdown	-0.595%	-0.008%	-10.475%
	Trade turnover	6.018	6.018	1.395
USDJPY	Win rate	85.106%	85.106%	54.627%
	CAGR	33.420%	33.420%	9.238%
	Sharpe Ratio	199.094	199.094	45.929%
	Max. Drawdown	-0.631 %	-0.631%	-14.749%
	Trade turnover	4.888	4.888	1.398