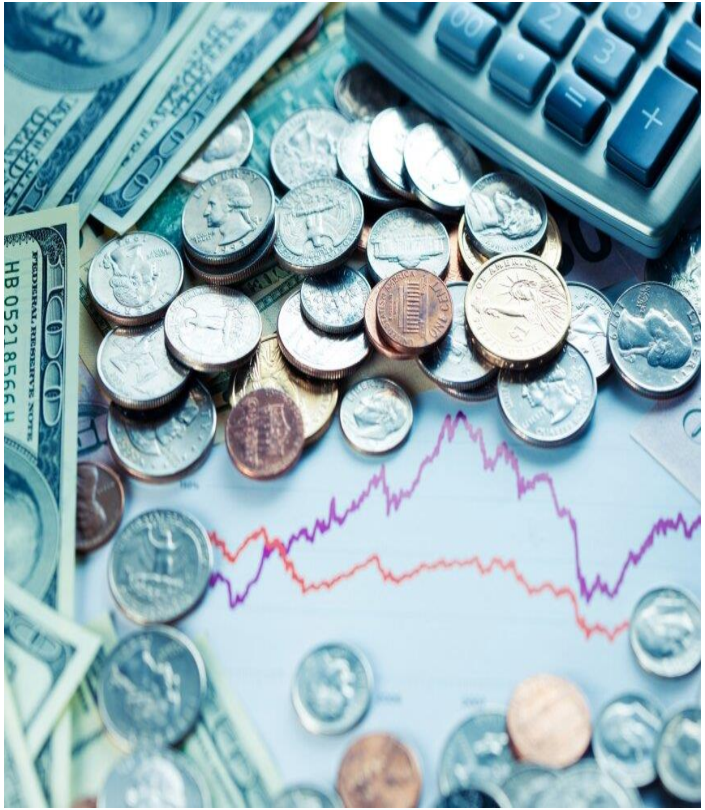


Stock Investment Customization

Team: Group 2

MA CHENGBIN A0261804J
XIE SITENG A0261982W
ZHOU JIECHENG A0261829W
CHEN LIUJUN A0261904H
LIN FANGZHOU A0261850H

ARCADIA
CAPITAL

Contents

1. Project Introduction
2. Data Understanding & Preparation
3. Predictive Analytics
4. Trading Strategies
5. Optimisation Analytics
6. Discussion

Project Introduction

1. Industry Overview

2. What we offer

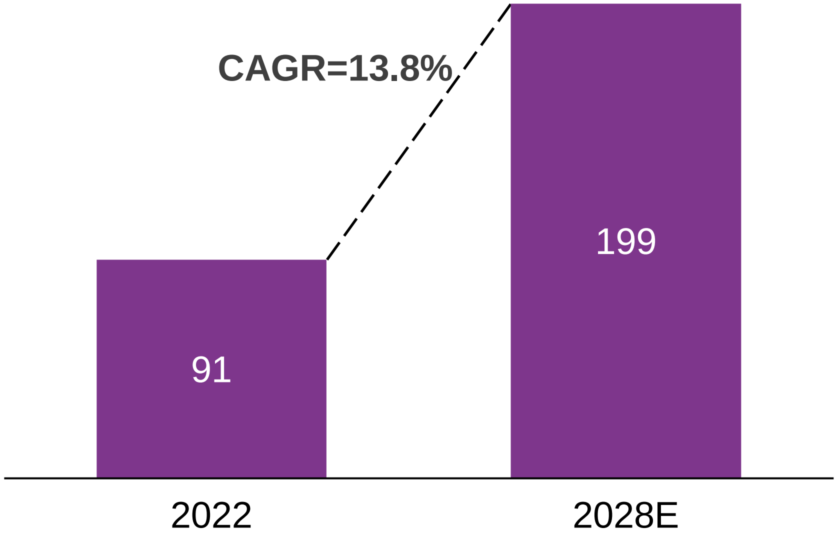
3. Business Objectives

4.Technical Objectives

1.1 Industry Overview

The development of the stock market played an important role in economic development. The market continues to increase.

Global Asset Management Market | Billion(USD)



Market Capitalization | Y2022

93.6 trillion (USD)

1.2 What We Offer?

ARCADIA
CAPITAL

Who we are

A regional leading asset management firm in Southeast Asia, working on stock investment.

What we offer

Make suitable portfolio which can maximize returns and minimize risks as well as making predictions on the stock price.

1.3 Business Objectives

1 **Make Informed Investment Decisions**

By Predicting stock return and Buy-and-sell-point

2 **Provide Customized Asset Allocation & Optimization Solutions**

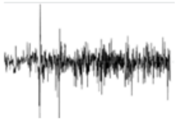
For clients with different investment preferences

We help them achieve long-term financial success!

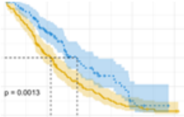
1.4 Technical Objectives

Prediction → **Trading Strategy** → **Product Selection(Filter)** → **Portfolio Optimization**

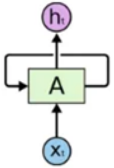
Prediction



ARMA GARCH



Survival Analysis



RNN-LSTM

Multiple Factor Regression

Trading Strategy

based on $Return_{pred}$

based on prediction of certain events

based on $Return_{pred}$

based on Alpha value

Product Selection(Filter)

Mature Return

Volatility

Number of transactions

Portfolio Optimization

Goal Programming

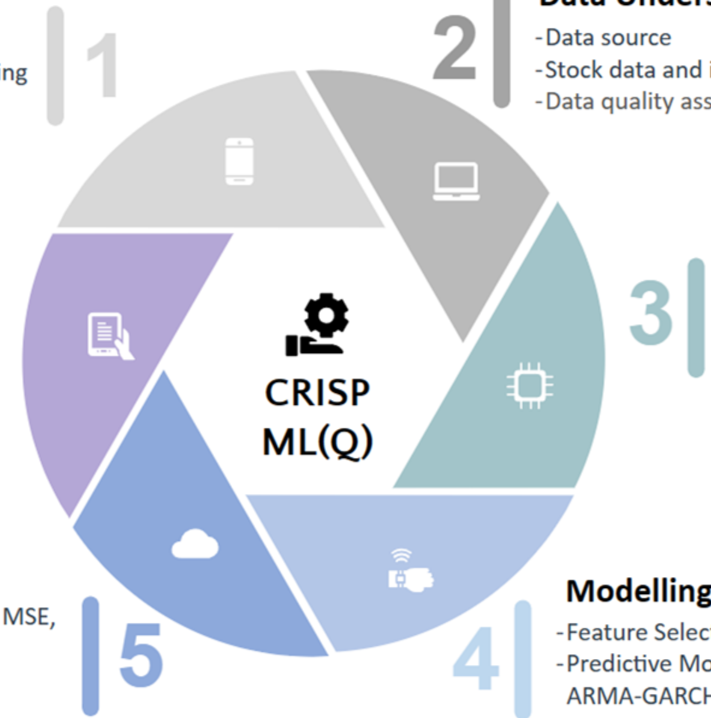
1.5 Project Management Plan

Business Understanding

- Stock market understanding
- Stock data and industry index understanding
- Clients' profile and preferences

Data Understanding

- Data source
- Stock data and industry stock index
- Data quality assurance



Data Preparation

- Data Cleaning and Transformation
- Data Dictionary

Deployment

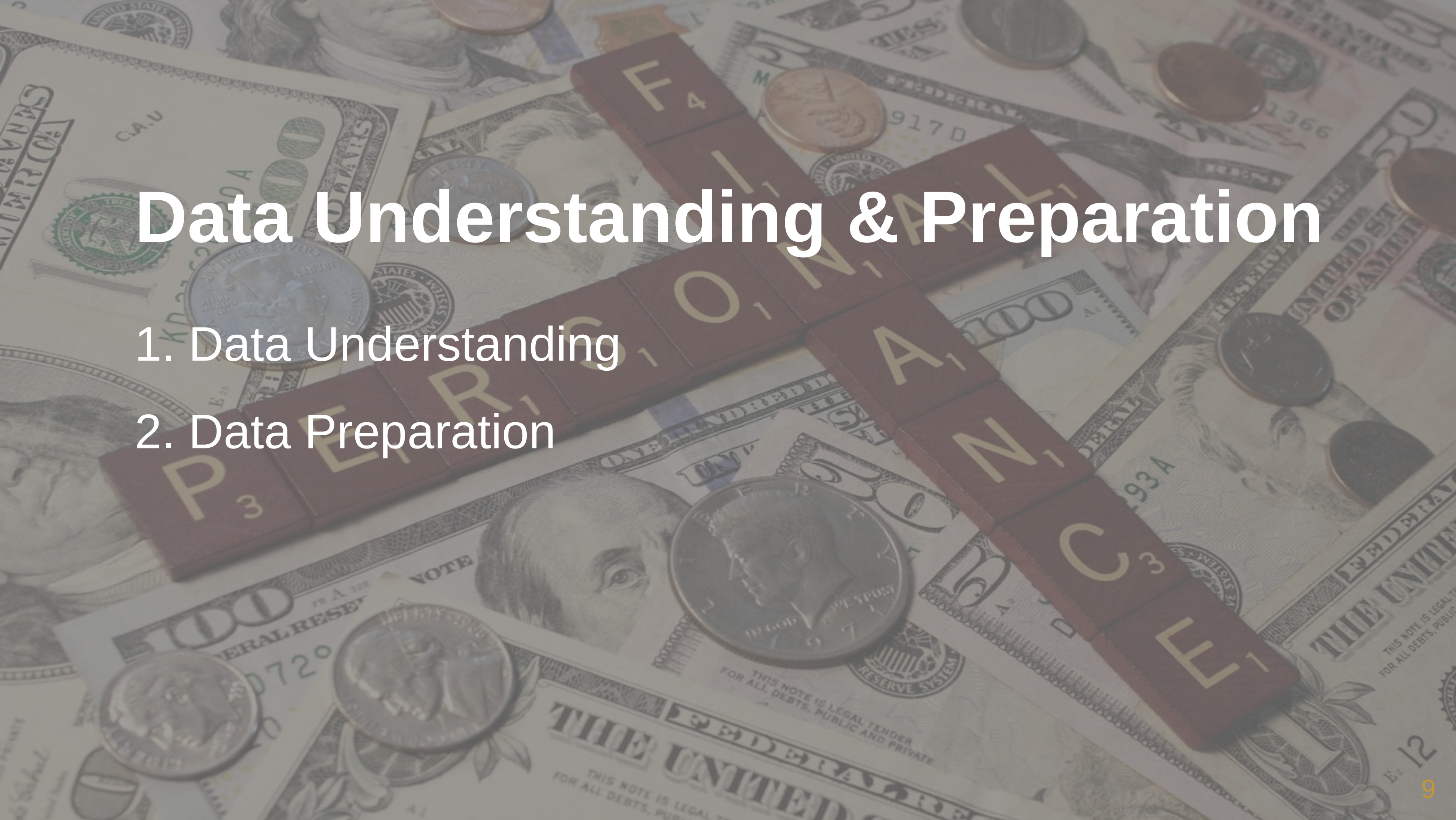
- Products with different returns and risks
- Various investment portfolios and Plans for clients with different investment preferences

Evaluation

- Predictive Model Comparisons & Evaluation: MSE, RMSE, MAPE, etc.
- Simulation and Deviance Minimization.
- Business Success Criteria

Modelling

- Feature Selection & Extraction
- Predictive Model Building: ARMA-GARCH, Survival Analysis, LSTM, etc.
- Optimisation Model Building: LP, Non LP & GP



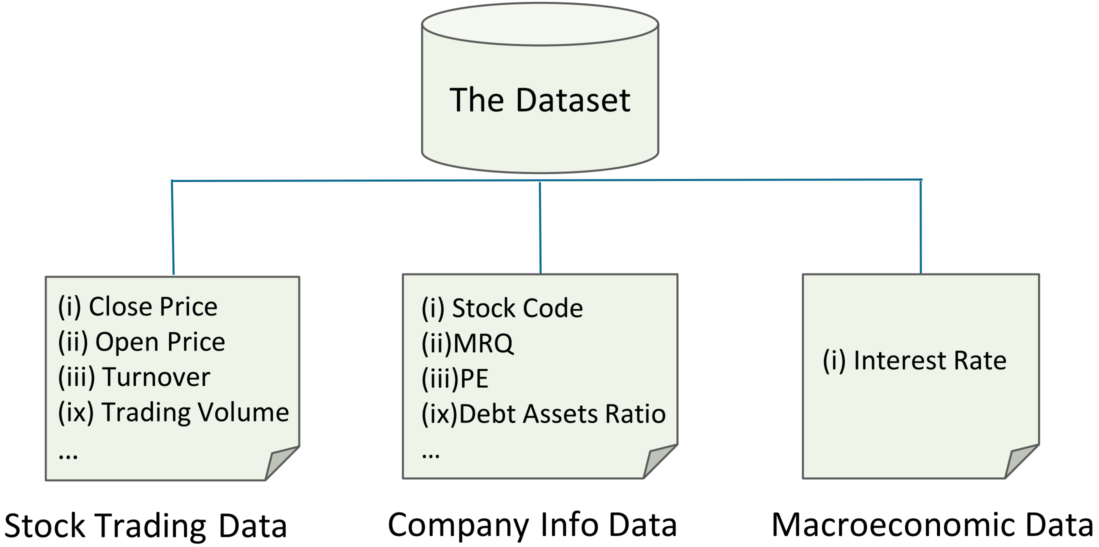
Data Understanding & Preparation

1. Data Understanding
2. Data Preparation

2.1 Data Understanding: Data Sources

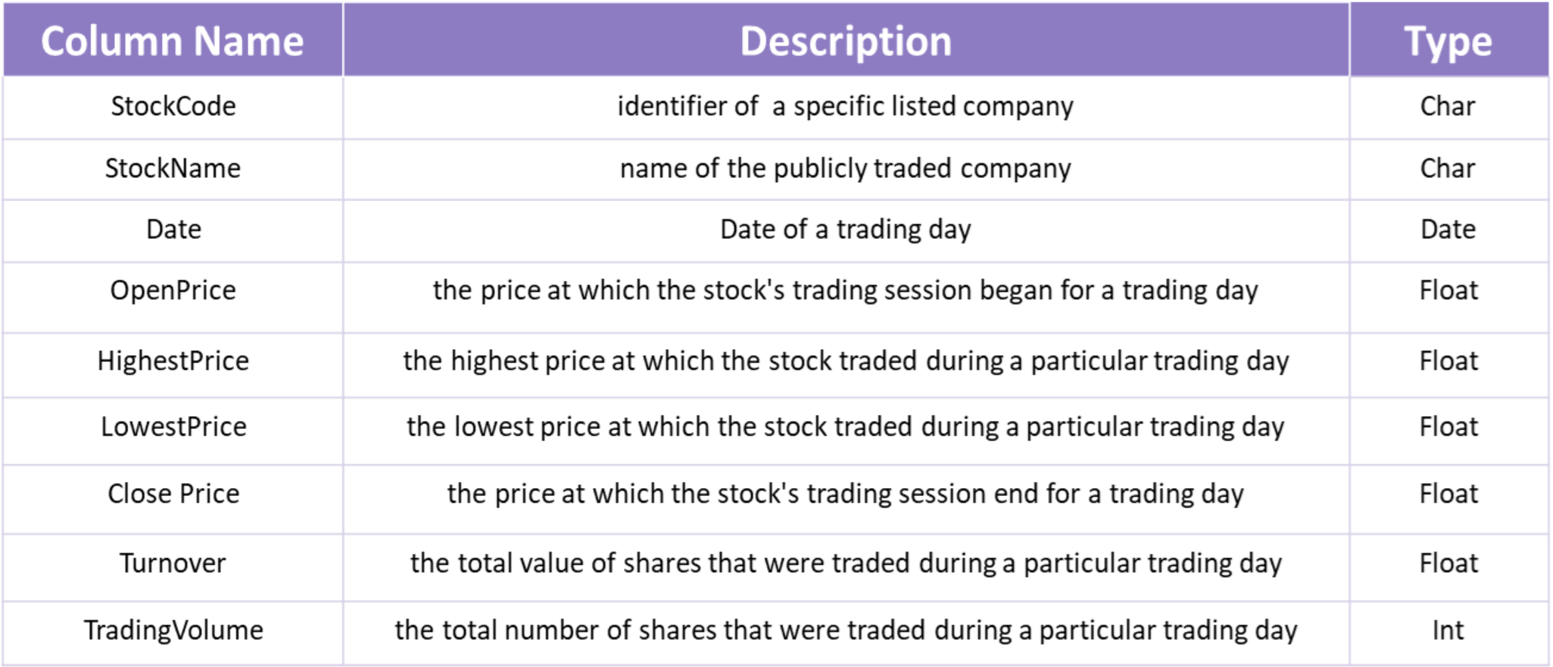
Data Source: Win.d yahoo! finance **Yahoo** (through API).

The Dataset: Top 20 listed companies(by market capitalization) in the new energy industry.
Time covered: From initial public offering (IPO) to the present day.



2.2 Data Understanding

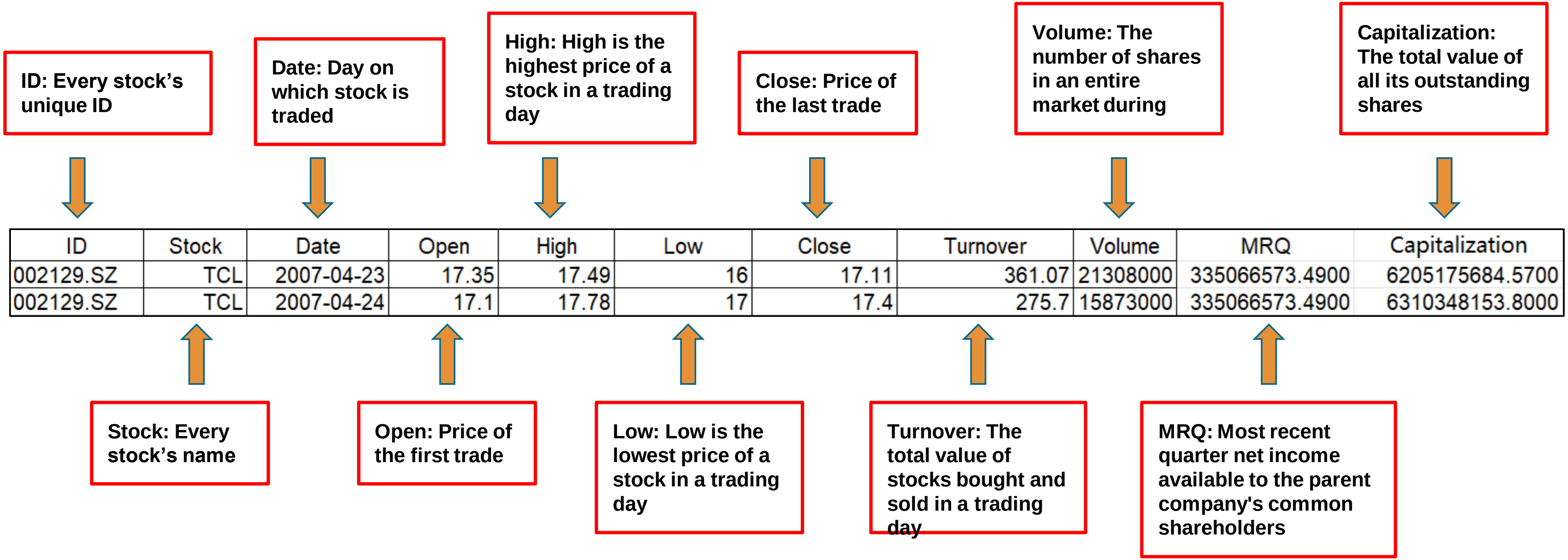
Data Dictionary



Column Name	Description	Type
StockCode	identifier of a specific listed company	Char
StockName	name of the publicly traded company	Char
Date	Date of a trading day	Date
OpenPrice	the price at which the stock's trading session began for a trading day	Float
HighestPrice	the highest price at which the stock traded during a particular trading day	Float
LowestPrice	the lowest price at which the stock traded during a particular trading day	Float
Close Price	the price at which the stock's trading session end for a trading day	Float
Turnover	the total value of shares that were traded during a particular trading day	Float
TradingVolume	the total number of shares that were traded during a particular trading day	Int

2.1 Data Understanding

Data Sample



ID: Every stock's unique ID

Date: Day on which stock is traded

High: High is the highest price of a stock in a trading day

Close: Price of the last trade

Volume: The number of shares in an entire market during

Capitalization: The total value of all its outstanding shares

ID	Stock	Date	Open	High	Low	Close	Turnover	Volume	MRQ	Capitalization
002129.SZ	TCL	2007-04-23	17.35	17.49	16	17.11	361.07	21308000	335066573.4900	6205175684.5700
002129.SZ	TCL	2007-04-24	17.1	17.78	17	17.4	275.7	15873000	335066573.4900	6310348153.8000

Stock: Every stock's name

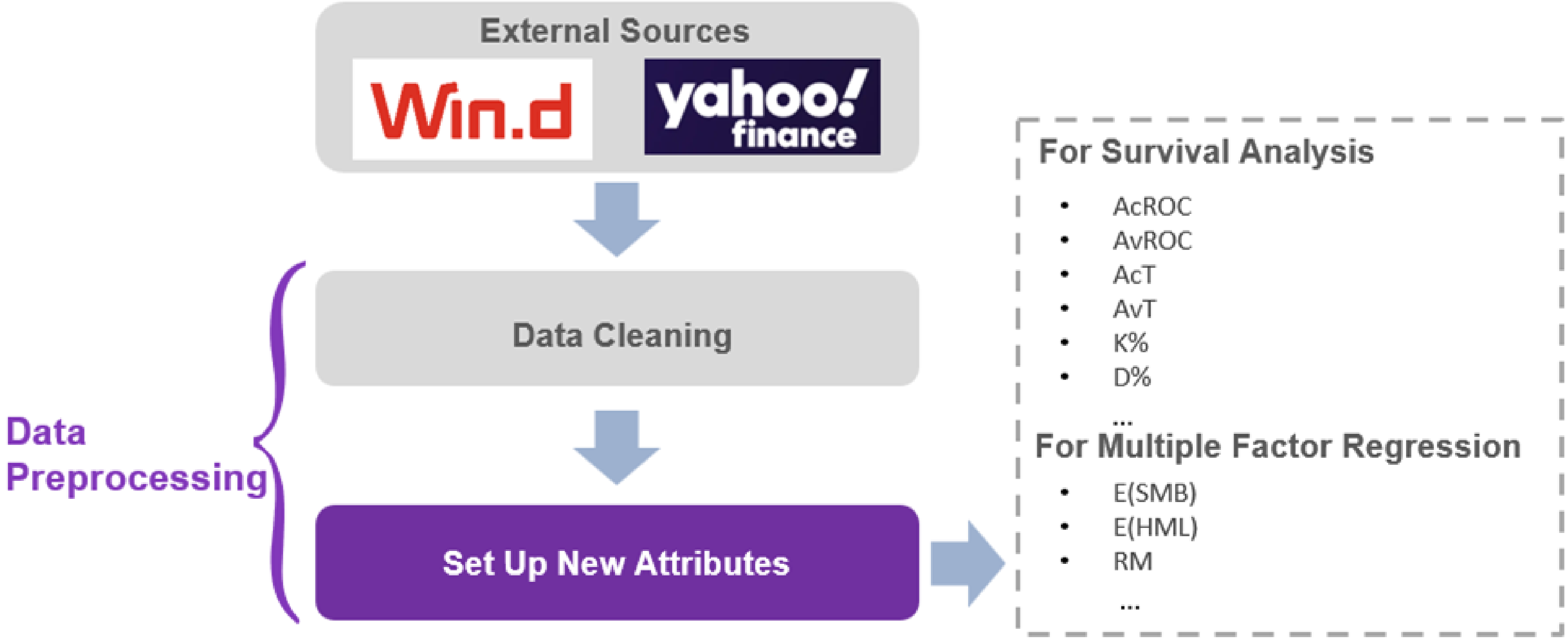
Open: Price of the first trade

Low: Low is the lowest price of a stock in a trading day

Turnover: The total value of stocks bought and sold in a trading day

MRQ: Most recent quarter net income available to the parent company's common shareholders

2.2 Data Preparation



For Survival Analysis

- AcROC
- AvROC
- AcT
- AvT
- K%
- D%

…

For Multiple Factor Regression

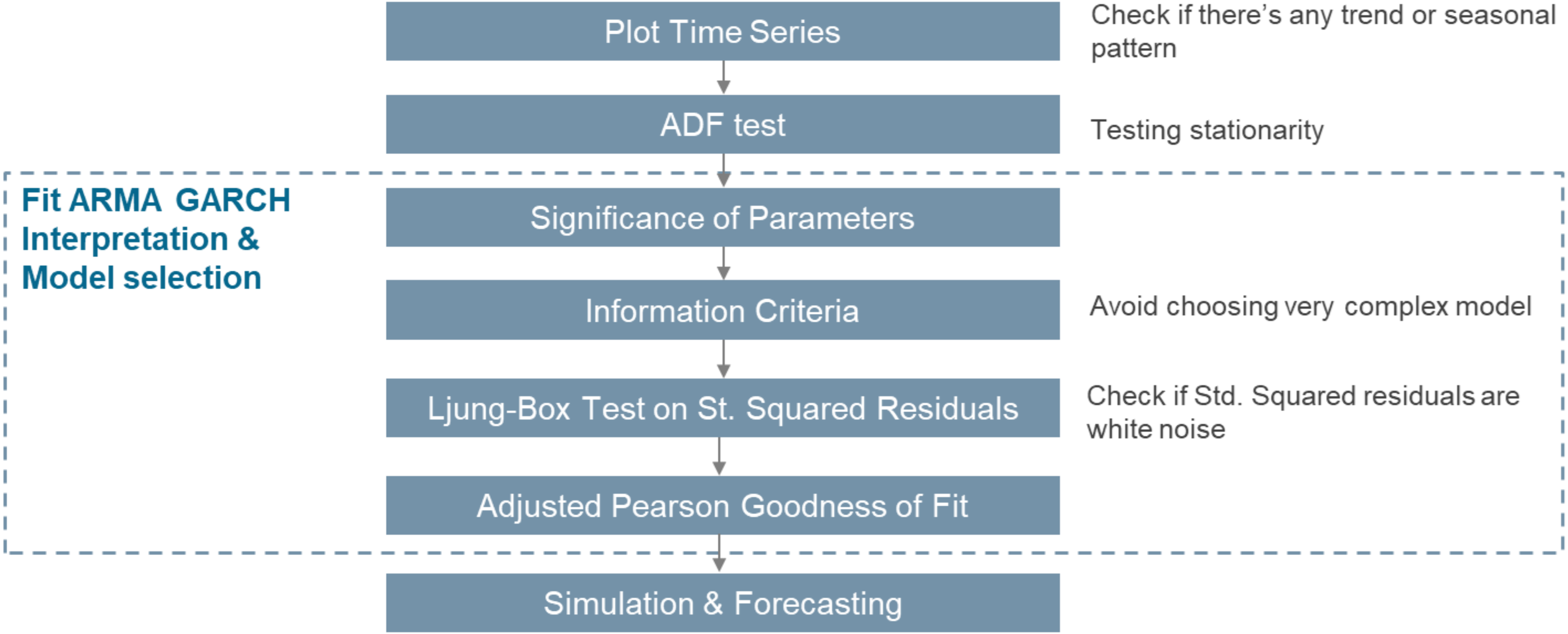
- E(SMB)
- E(HML)
- RM

…

Predictive Analytics

1. ARMA-GARCH
2. LSTM
3. Survival Analysis
4. Multiple Factor Regression

3.1 ARMA-GARCH



Plot Time Series

Check if there’s any trend or seasonal pattern

ADF test

Testing stationarity

Fit ARMA GARCH Interpretation & Model selection

Significance of Parameters

Information Criteria

Avoid choosing very complex model

Ljung-Box Test on St. Squared Residuals

Check if Std. Squared residuals are white noise

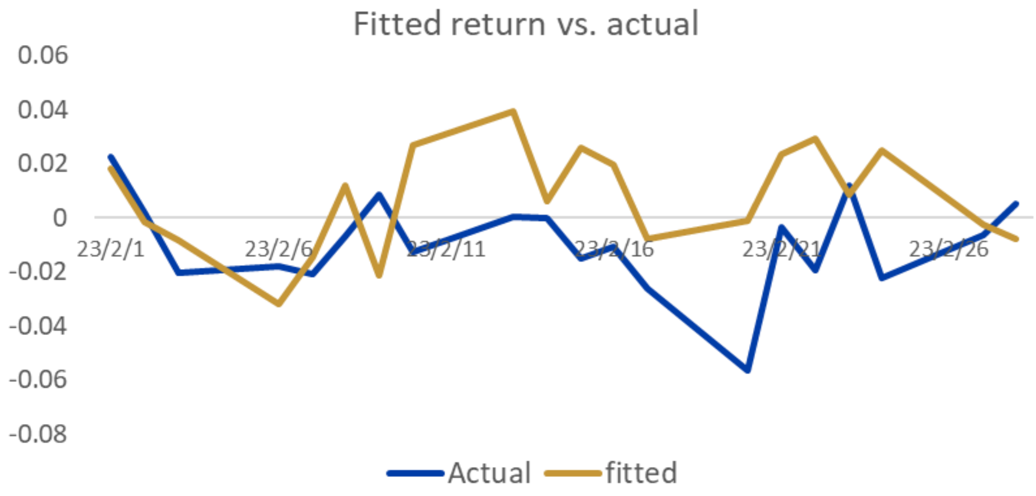
Adjusted Pearson Goodness of Fit

Simulation & Forecasting

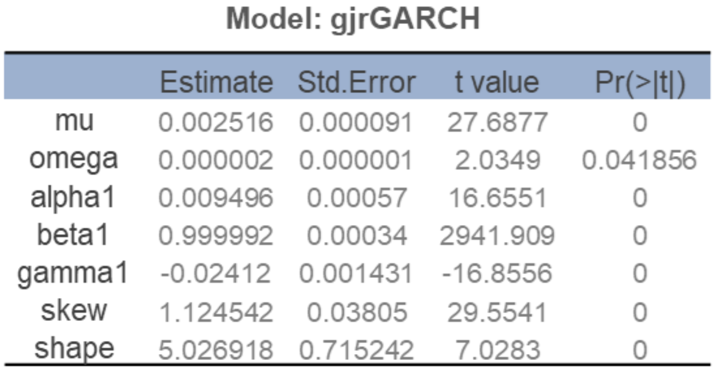
3.1 ARMA-GARCH

Prediction Sample

Stock Name： Evemall
Stock Code： 300014.SZ



Fitted return vs. actual



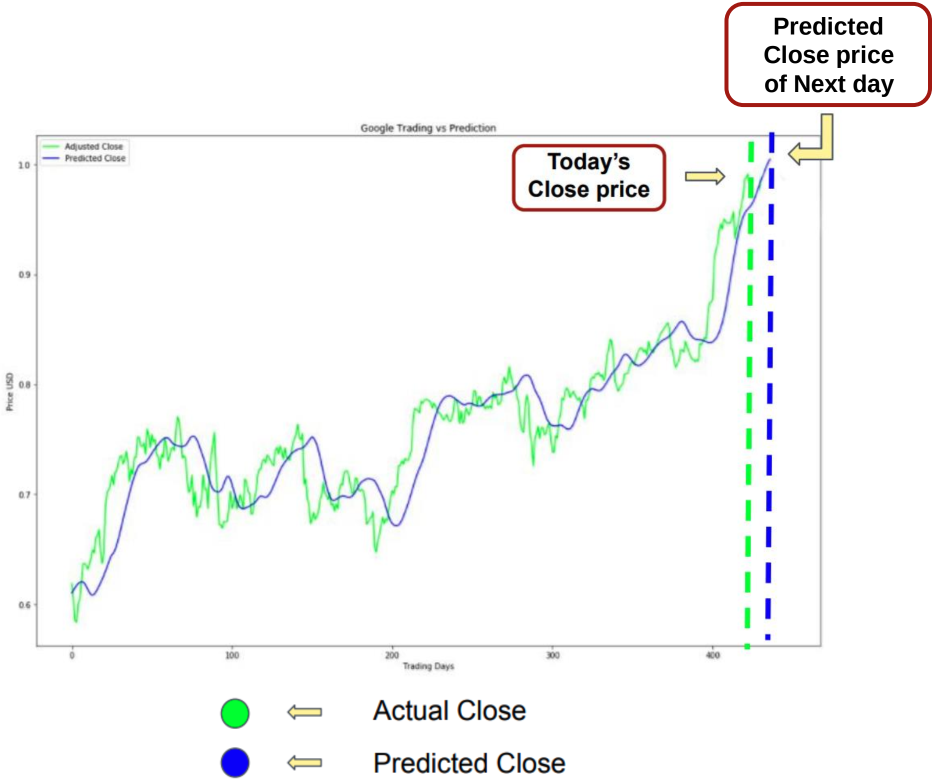
Model: gjrGARCH

	Estimate	Std.Error	t value	Pr(>|t|)
mu	0.002516	0.000091	27.6877	0
omega	0.000002	0.000001	2.0349	0.041856
alpha1	0.009496	0.00057	16.6551	0
beta1	0.999992	0.00034	2941.909	0
gamma1	-0.02412	0.001431	-16.8556	0
skew	1.124542	0.03805	29.5541	0
shape	5.026918	0.715242	7.0283	0

RMSE: 0.031

3.3 LSTM - Problem Statement

- To accurately predict the future closing value of a given stock across a given period of time in the future.



Predicted Close price of Next day

Today's Close price

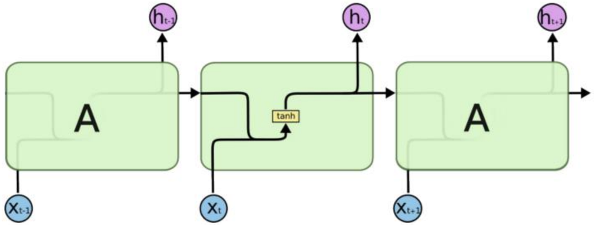
● Actual Close

● Predicted Close

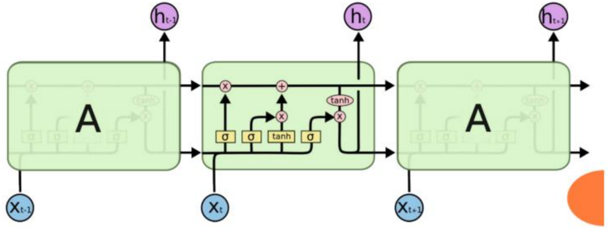
3.3 LSTM

- Long Short-Term Memory (LSTM) is one type of recurrent neural network which is used to learn order dependence in sequence prediction problems.

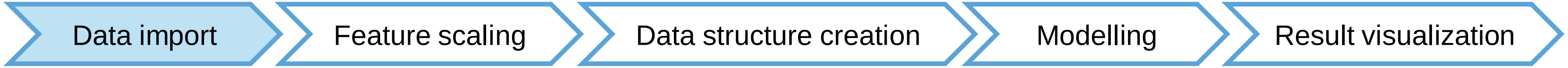
- Due to its capability of storing past information, LSTM is very useful in predicting stock prices. This is because the prediction of a future stock price is dependent on the previous prices.



Recurrent Neural Network

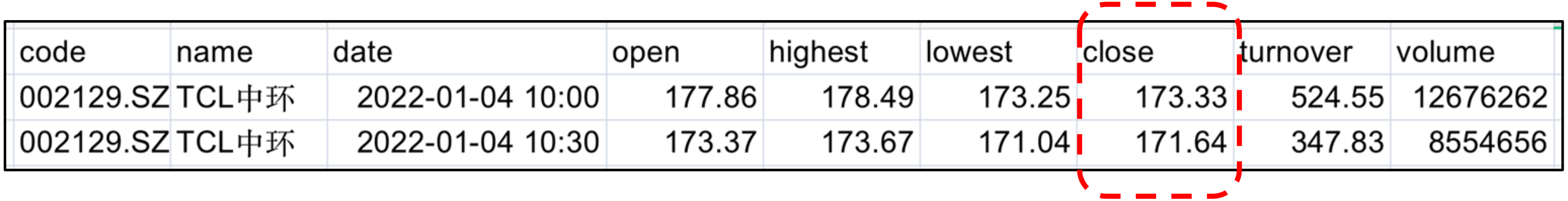


3.3 LSTM - Processing Flow



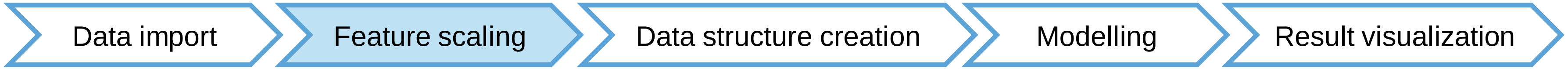
Data import → Feature scaling → Data structure creation → Modelling → Result visualization

- We will use close price for prediction

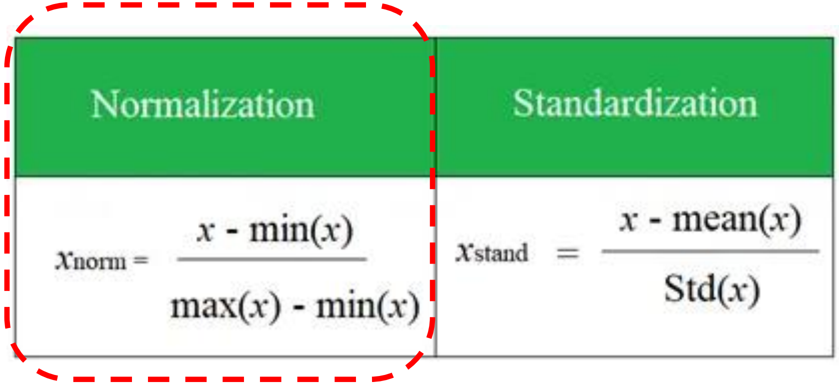


code	name	date	open	highest	lowest	close	turnover	volume
002129.SZ	TCL中环	2022-01-04 10:00	177.86	178.49	173.25	173.33	524.55	12676262
002129.SZ	TCL中环	2022-01-04 10:30	173.37	173.67	171.04	171.64	347.83	8554656

3.3 LSTM - Processing Flow



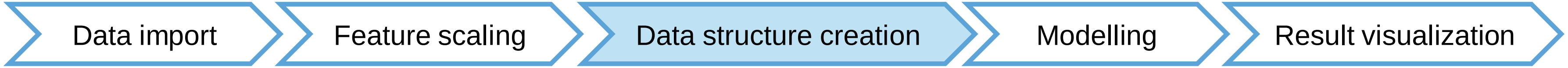
- The next step is to scale the stock prices between (0, 1) to avoid intensive computation. Common methods include Standardization and Normalization as shown in figure. It is recommended to take Normalization, particularly when working on RNN with a Sigmoid function in the output layer.



Normalization	Standardization
$x_{norm} = \dfrac{x - \min(x)}{\max(x) - \min(x)}$	$x_{stand} = \dfrac{x - \text{mean}(x)}{\text{Std}(x)}$

ARCADIA CAPITAL

3.3 LSTM - Processing Flow



Data import

Feature scaling

Data structure creation

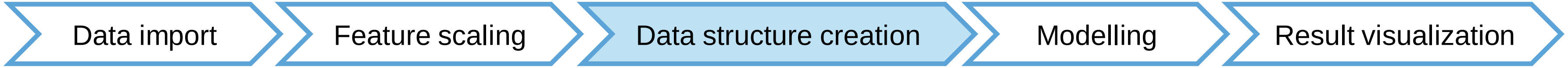
Modelling

Result visualization

- We use timestep prices to predict the the timestep+1 price (one price)

x = timestep days data	y= timestep+1 data

3.3 LSTM - Processing Flow



Data import

Feature scaling

Data structure creation

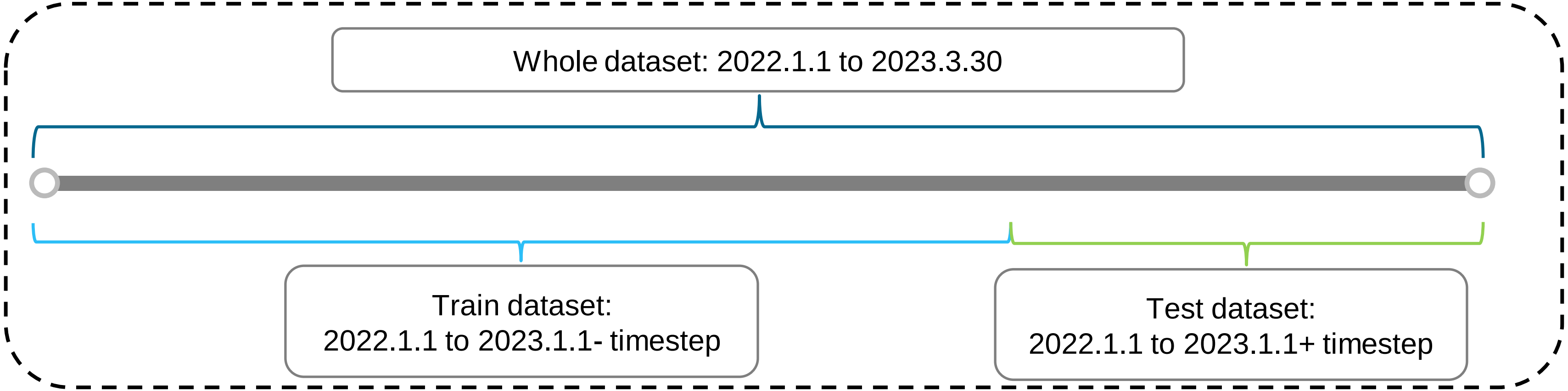
Modelling

Result visualization

- We use timestep prices to predict the the timestep+1 price (one price)

x = timestep days data	y= timestep+1 data

- Then split the whole dataset, we want to predict the price from 2023.1.1 to 2023.3.30

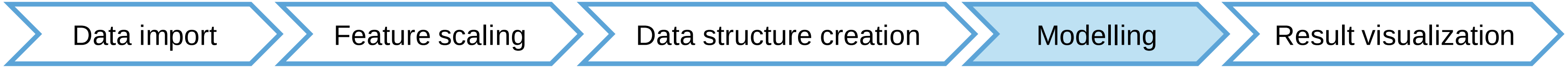


Whole dataset: 2022.1.1 to 2023.3.30

Train dataset:
2022.1.1 to 2023.1.1- timestep

Test dataset:
2022.1.1 to 2023.1.1+ timestep

3.3 LSTM - Processing Flow



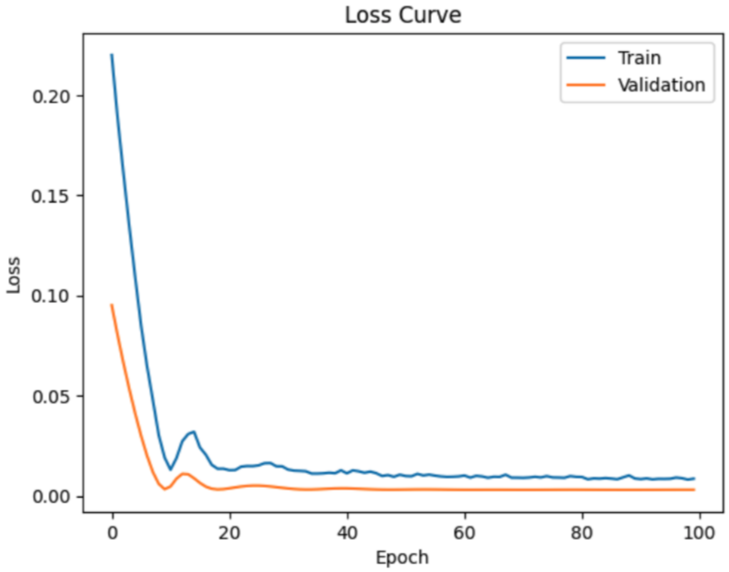
- By tuning hyperparameters to get the model with best performance for each stock

Hyperparameters	Pros	Cons
Time step ↑	remember historical information	computational complexity and training time.
Number of layers ↑	complexity and expressive power	the risk of overfitting and training time.
Hidden dimension of each layer ↑	complexity and expressive power	computational complexity and training time.
Batch size ↑	training speed and stability	consumes more memory resources and can lead to reduced generalization performance
Epochs ↑	fitting ability and performance	training time and the risk of overfitting.

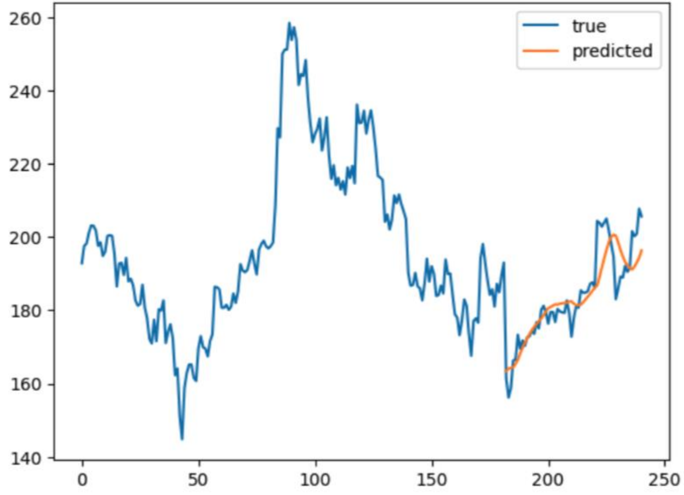
3.3 LSTM - Processing Flow

Data import
Feature scaling
Data structure creation
Modelling
Result visualization

Example Stock: Contemporary Amperex Technology Co.



```
Train RMSE: 10.026642312176302
Test RMSE: 6.281534922746179
```



3.3 Survival Analysis-Assumptions

Goal: find best **Buy-and-Sell-Point** by predicting return increase and decrease probability.

Two Events: A stock with at least α one-day rise (R_a) is the rise event; with at least β one-day drop (D_β) are is the drop event.

α: the rate of stock return increase

β: the rate of stock return decrease

γ_t: stock return

μ_t: the turnover rate at time t.

For Return Increase:

Event: When α> 0.01, event is triggered and the event status is 1 ; when α< 0.01, event is not triggered and event status is 0.

Duration(Tα): Each Event’s Living Time

For Return Decrease:

Event: When β< -0.01, event is triggered and the event status is 1 ; when β> -0.01, event is not triggered and event status is 0.

Duration(Tβ): Each Event’s Living Time

3.3 Survival Analysis-Covariates Calculation

The following covariates are constructed based on stock activities observed from the last “event” time to current observation time.

Accumulated Rate of Change (AcROC): $x_1(T)=\sum_{t=T_c-T}^{T_c-1}\gamma_t$

The cumulative gains from the time point when the latest "event" happened to the current time point

Average Rate of Change (AvROC): $x_2(T)=\frac{\sum_{t=T_c-T}^{T_c-1}\gamma_t}{T}$

The average gains from the time point when the latest "event" happened to the current time point.

Accumulated Turnover (AcT): $x_3(T)=\sum_{t=T_c-T}^{T_c-1}\mu_t$

The cumulative turnover rate from the time point when the latest "event" happened to the current time point.

Average Turnover (AvT): $x_4(T)=\frac{\sum_{t=T_c-T}^{T_c-1}\mu_t}{T}$

The average turnover rate from the time point when the latest "event" happened to the current time point.

3.3 Survival Analysis-Covariates Calculation

Stochastic K% (K%): $x_5(T)=\frac{P_{T_{c-1}}-LL_T}{HH_T-LL_T} * 100\%$

This covariate refers to the point of a current price in relation to its price range over a period of time; HHT and LLT mean lowest low and highest high in the last T days, respectively

Stochastic D% (D%): $x_6(T)=\frac{\sum_{t=T_c-T}^{T_c-1} * K_t\%}{T}$

This covariate measures the average K% over the last n days.

Stochastic J% (J%): $x_7(T) \ = 3K_{T_{c-1}}\%\text{-}2D_{c-1}\%$

This covariate is a derived form of the stochastic with the only difference being an extra line.

Relative Strength Index (RSI): $x_8(T) \ = 100\text{-}\frac{100}{1+RS}$ $\quad \text{RS}=\frac{\sum_{t=T_c-T\text{, } \gamma_t \geq 0}^{T_c-1} \gamma_t}{\sum_{t=T_c-T}^{T_c-1} \gamma_t}$

This covariate is intended to chart the current and historical strength or weakness of a stock on the closing prices of a recent trading period.

Psychological Line (PSY): $x_9(T) = \frac{\sum_{t=T_c-T}^{T_c-1} I(\gamma_t)}{T}$ Where $I(\gamma_t)$=1 if $\gamma_t \geq 0 \ and \ 0 \ otherwise$

This covariate measures the ratio of the number of rising periods over the total number of periods.

3.3 Survival Analysis-Cox Regression

Two thresholds(0.01, -0.01); two models：

Rise Model

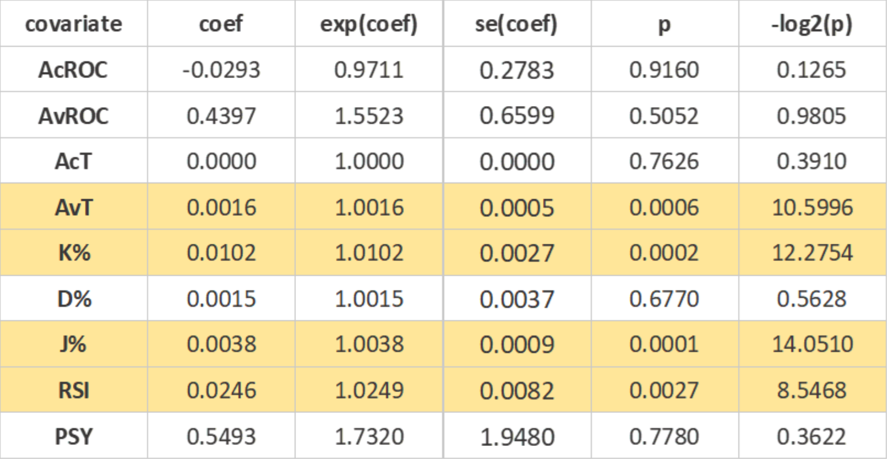
Drop Model

- Covariates Selection: **Wald Test**
- Parameter Selection(Penalizer): Parameter grid, choose parameter with highest **c-index**
- **C-index:** indicates accordance and accuracy of survival analysis
  (range from 0-1, 0.5 means random prediction, above 0.7 means good prediction)

3.3 Survival Analysis- Example

Take CGNPC(China Guangdong Nuclear Power)as an example

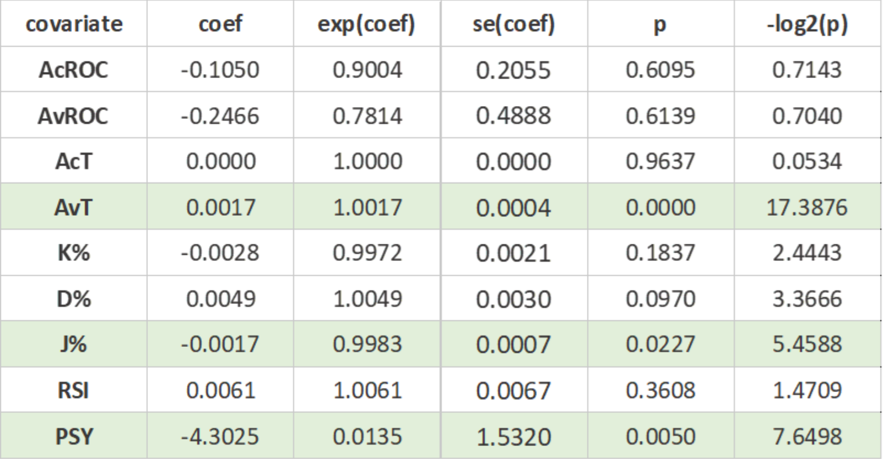
the Rise Model



covariate	coef	exp(coef)	se(coef)	p	-log2(p)
AcROC	-0.0293	0.9711	0.2783	0.9160	0.1265
AvROC	0.4397	1.5523	0.6599	0.5052	0.9805
AcT	0.0000	1.0000	0.0000	0.7626	0.3910
AvT	0.0016	1.0016	0.0005	0.0006	10.5996
K%	0.0102	1.0102	0.0027	0.0002	12.2754
D%	0.0015	1.0015	0.0037	0.6770	0.5628
J%	0.0038	1.0038	0.0009	0.0001	14.0510
RSI	0.0246	1.0249	0.0082	0.0027	8.5468
PSY	0.5493	1.7320	1.9480	0.7780	0.3622

Covariate importance for the rise model with α= 1%.

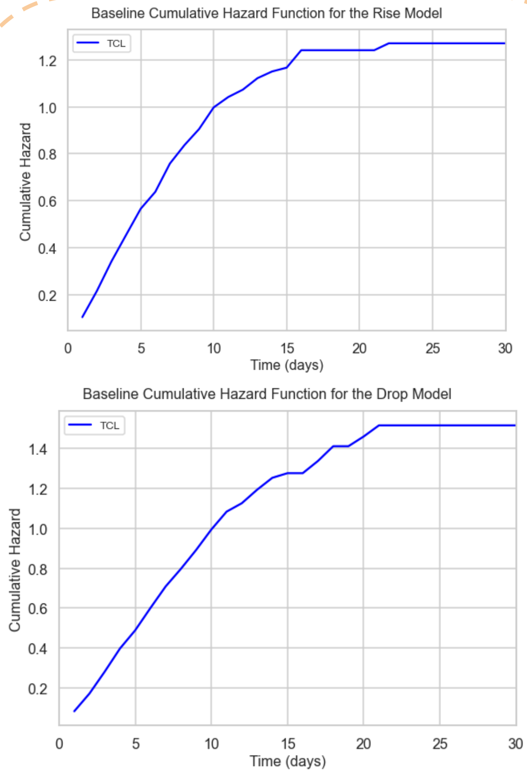
the Drop Model



covariate	coef	exp(coef)	se(coef)	p	-log2(p)
AcROC	-0.1050	0.9004	0.2055	0.6095	0.7143
AvROC	-0.2466	0.7814	0.4888	0.6139	0.7040
AcT	0.0000	1.0000	0.0000	0.9637	0.0534
AvT	0.0017	1.0017	0.0004	0.0000	17.3876
K%	-0.0028	0.9972	0.0021	0.1837	2.4443
D%	0.0049	1.0049	0.0030	0.0970	3.3666
J%	-0.0017	0.9983	0.0007	0.0227	5.4588
RSI	0.0061	1.0061	0.0067	0.3608	1.4709
PSY	-4.3025	0.0135	1.5320	0.0050	7.6498

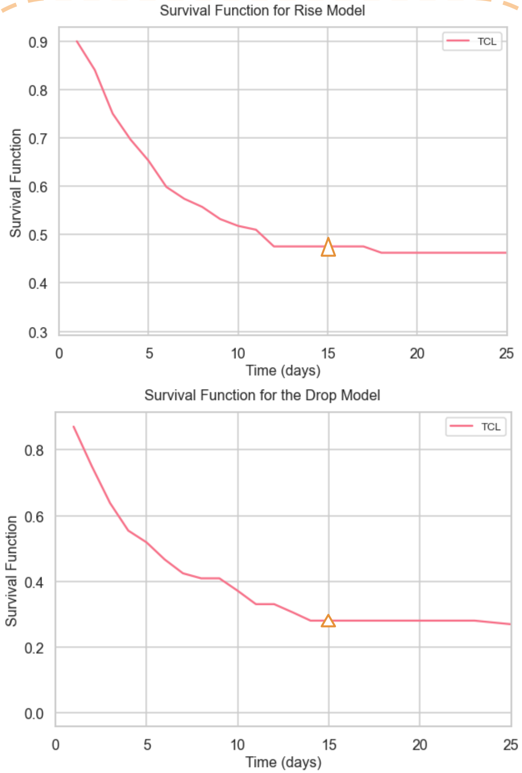
Covariate importance for the drop model with β = −1%.

3.3 Survival Analysis- Example


Baseline Cumulative Hazard Function for the Rise Model

Baseline Cumulative Hazard Function for the Drop Model

Baseline Cumulative Hazard Function


Survival Function for Rise Model

Survival Function for the Drop Model

Survival Function

the Rise Model

the Drop Model

3.4 Multiple Factor Regression-Select Model

- We utilized Fama-French Three Factor Model to do stock trading.
- This model introduces two additional factors on the basis of the Capital Asset Pricing Model (CAPM) to explain the variation of stock returns.
- Formula:

$$R_i = a_i + b_i R_M + s_i E(SMB) + h_i E(HML) + \varepsilon_i$$

- As shown above, stock return are influenced by three factors: **E(SMB), E(HML) and RM(Rm-Rf).**

3.4 Multiple Factor Regression-Factors

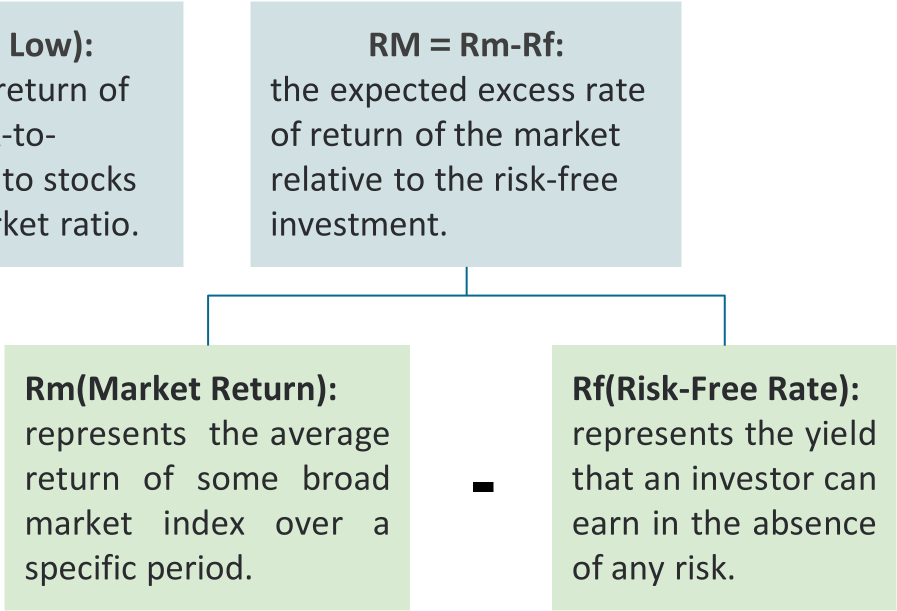
E(SMB)

SMB(Small Minus Big): measures the excess return of small-cap stocks relative to large-cap stocks.

E(HML)

HML(High Minus Low): measures the excess return of stocks with high Book-to-Market Ratio relative to stocks with low book-to-market ratio.

RM



RM = Rm-Rf: the expected excess rate of return of the market relative to the risk-free investment.

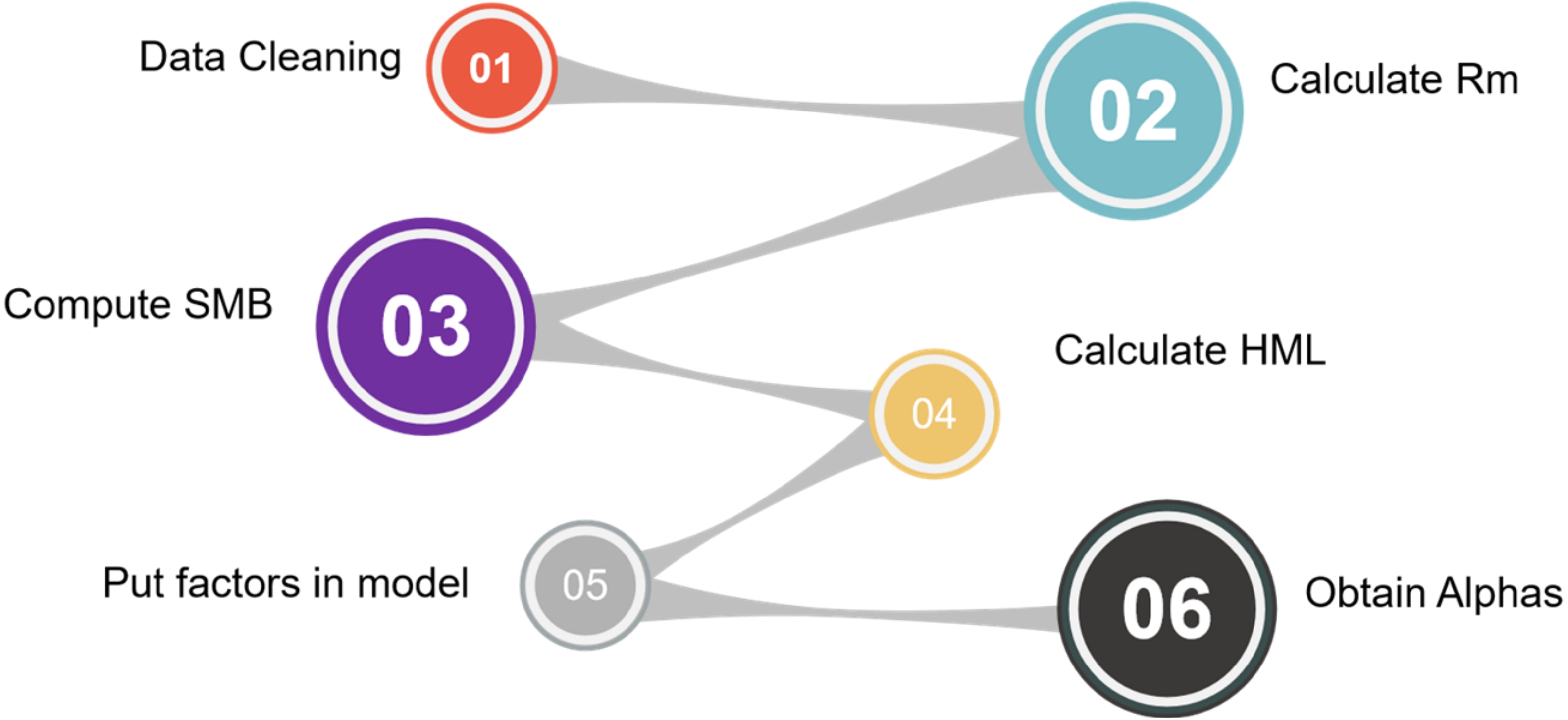
Rm(Market Return): represents the average return of some broad market index over a specific period.

-

Rf(Risk-Free Rate): represents the yield that an investor can earn in the absence of any risk.

Three Key Factors for Modeling

3.4 Multiple Factor Regression-Steps



3.4 Multiple Factor Regression-Benefits of Settings

- We Set the Rollback Day to 60 days

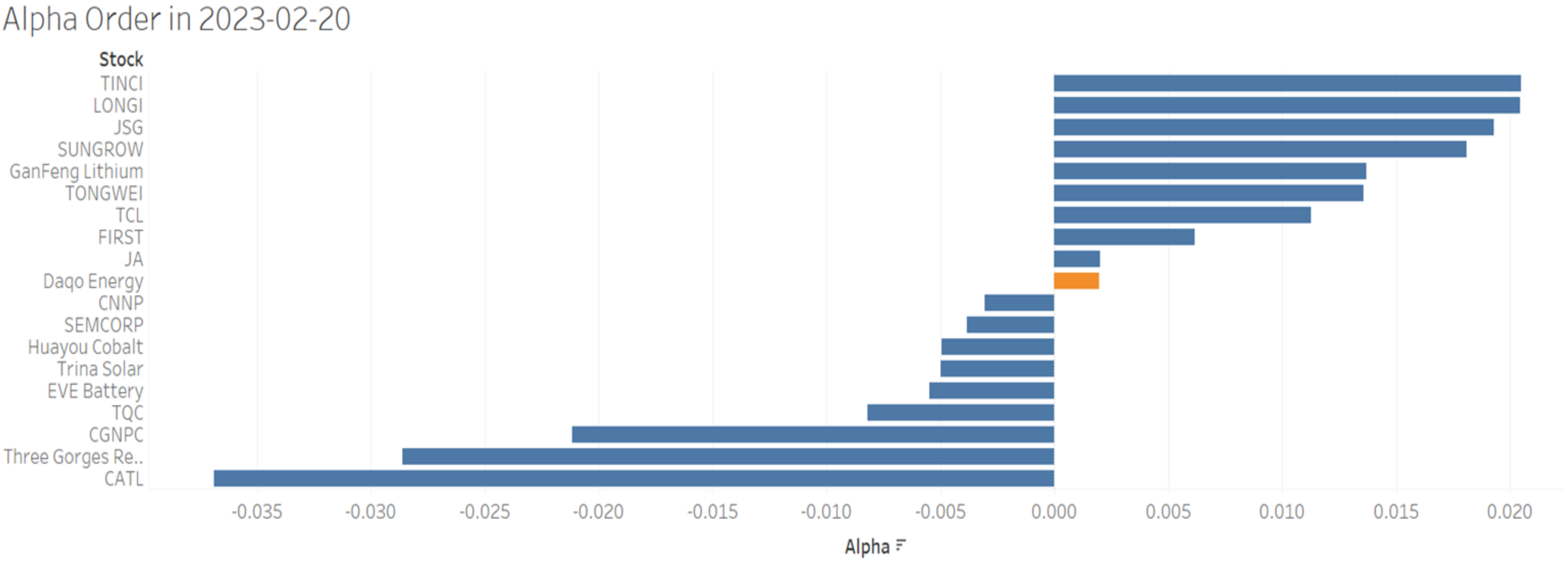
Data Smoothing

More Data Points

Adding Cyclical Considerations

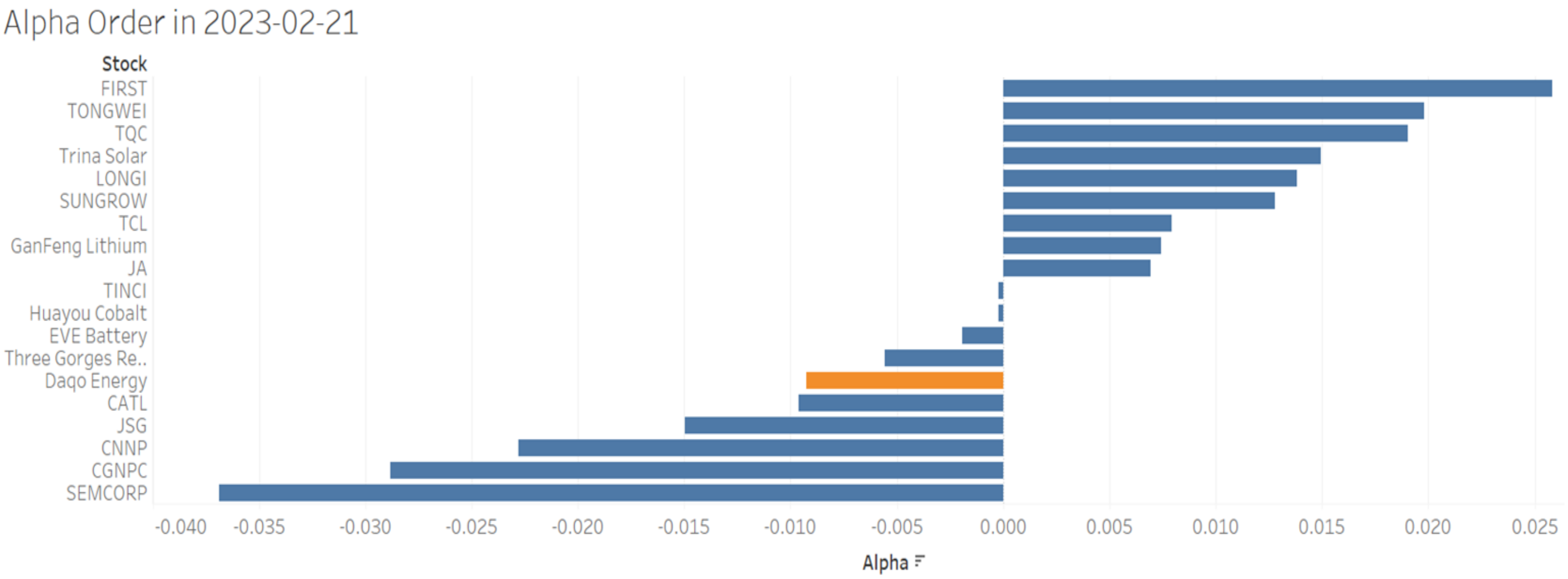
Better Generalization

3.4 Multiple Factor Regression-Alpha Order



ARCADIA CAPITAL

3.4 Multiple Factor Regression-Alpha Order



Financial Products

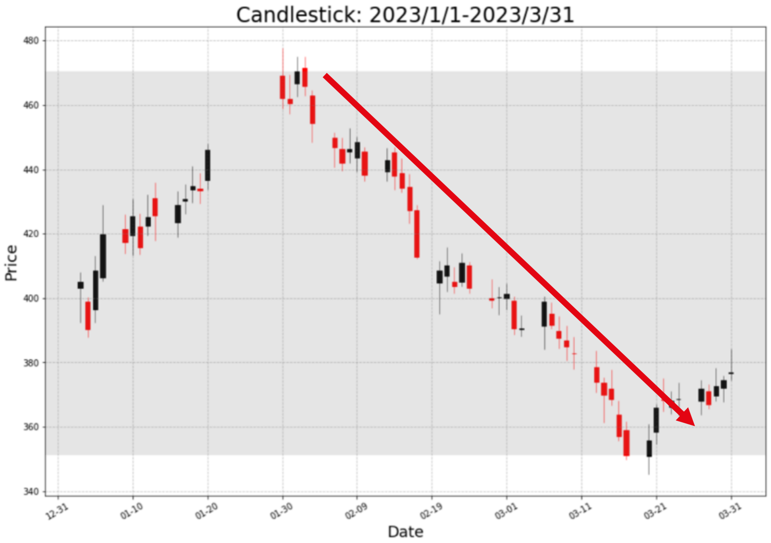
1. Trading Strategies for Each Predictive Methods
2. Filtering Conditions

4.1 Trading Strategy- Testing Period

NEW ENERGY

(Top 20 Market Value)

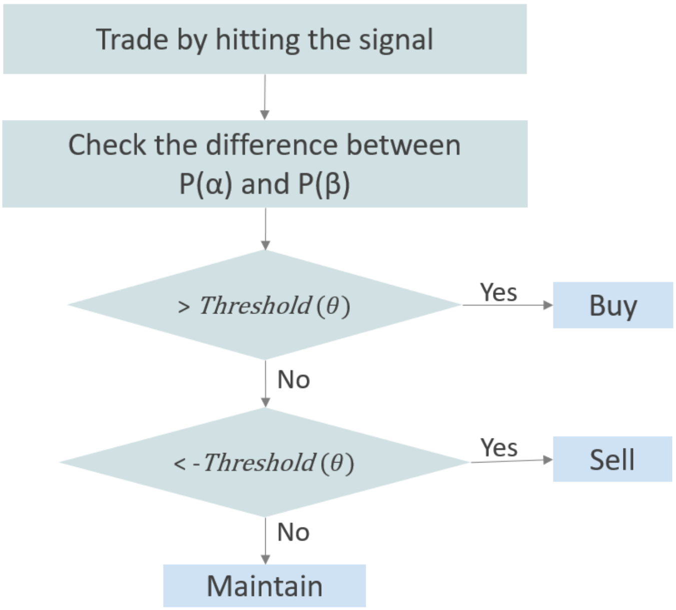
2023-01-01
2023-03-31



Short Position is not allowed!

4.1 Trading Strategy-Single Stock

Survival Function



Trade by hitting the signal

Check the difference between P(α) and P(β)

$> Threshold(\theta)$

Yes

Buy

No

$< \text{-}Threshold(\theta)$

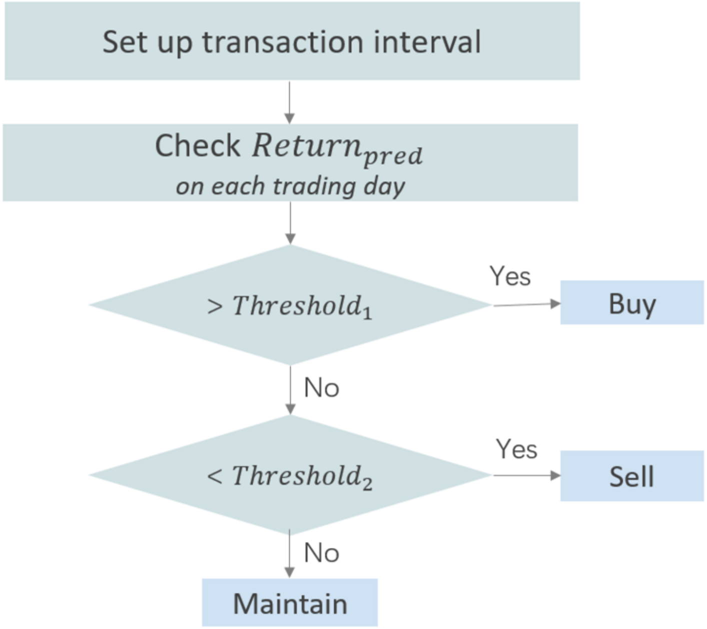
Yes

Sell

No

Maintain

ARMA GARCH & LSTM



Set up transaction interval

Check $Return_{pred}$

on each trading day

$> Threshold_1$

Yes

Buy

No

$< Threshold_2$

Yes

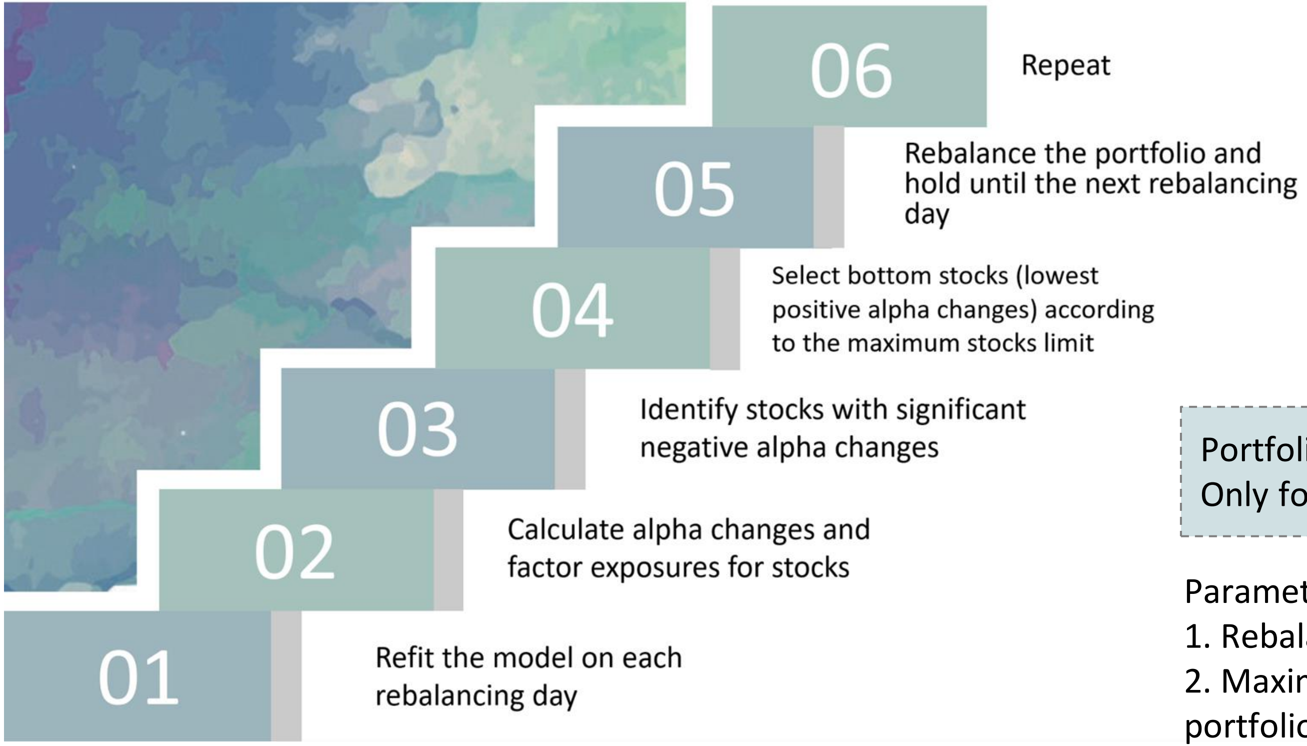
Sell

No

Maintain

- Thresholds and transaction interval is set by traversal

4.1Trading Strategy- For Multiple Stocks



01 Refit the model on each rebalancing day

02 Calculate alpha changes and factor exposures for stocks

03 Identify stocks with significant negative alpha changes

04 Select bottom stocks (lowest positive alpha changes) according to the maximum stocks limit

05 Rebalance the portfolio and hold until the next rebalancing day

06 Repeat

Portfolio(Multiple Stocks) Change: Only for Multi Factor Method

Parameters to consider:

1. Rebalancing interval
2. Maximum stocks in the portfolio

4.2 Filtering Conditions

	ARMA GARCH	LSTM	Survival Analysis	Multi Factor Method
Filter 1 Delete products that have **low number of transactions**	521	37	1434	141
Filter 2 Delete products that have **negative mature return**	133	30	131	0
Further Optimization				

Note: *The Multi Factor Method has not formed an effective product, mainly due to the overall downward trend of new energy stocks in the first three months of FY23*

Optimisation Analytics

1. Products Attributes
2. Customer Investment Persona
3. Optimisation Solution

5.1 Optimization - Efficient Frontier （NLP）

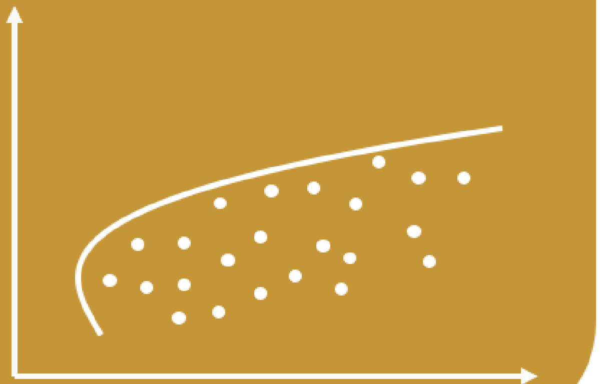
Higher Return

Lower Risk

$$S_a = \frac{E[R_a - R_b]}{\sigma_a}$$

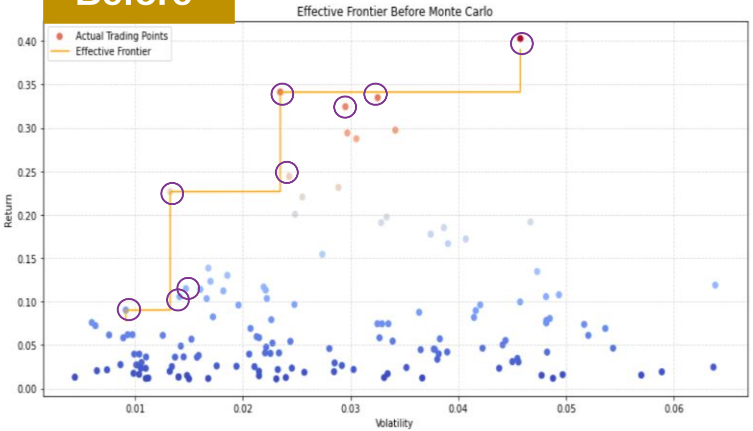
Sharpe Ratio

Markowitz Effective Frontier

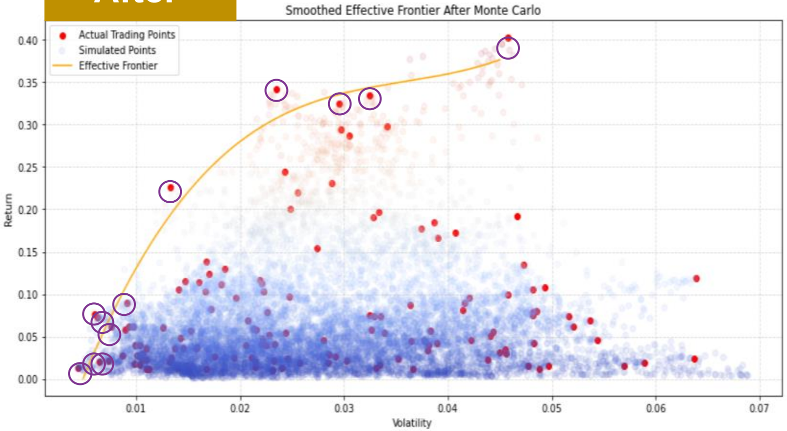


5.2 Optimization - Monte Carlo

Before



After



Imputation →

8 Strategy Detected

More than 10 Strategy Detected

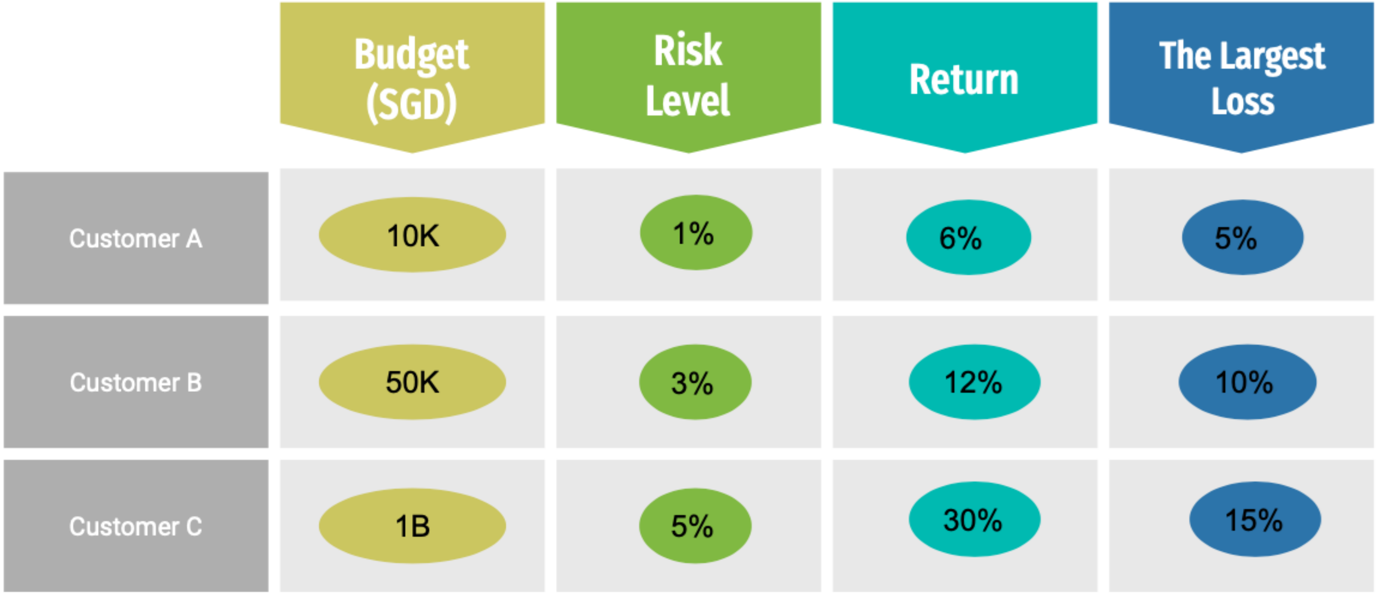
5.2 Optimization- Product

Stock	Return	Risk	Maximum Drawdown	Applied Model	
CGNPC	1.30%~9.02%	0.44%~0.91%	3.16%~4.16%	GARCH SURVIVAL LSTM	5 Products
CNNP	2.15%	0.74%	7.48%	GARCH	1 Products
TCL	11.48%~40.28%	1.47%~4.57%	7.10%~18.77%	GARCH LSTM	7 Products
JSG	13.86%	1.68%	6.85%	LSTM	1 Products
					14 Products

ARCADIA
CAPITAL

5.3 Optimization- Customer Risk Profile

- Different clients have different needs

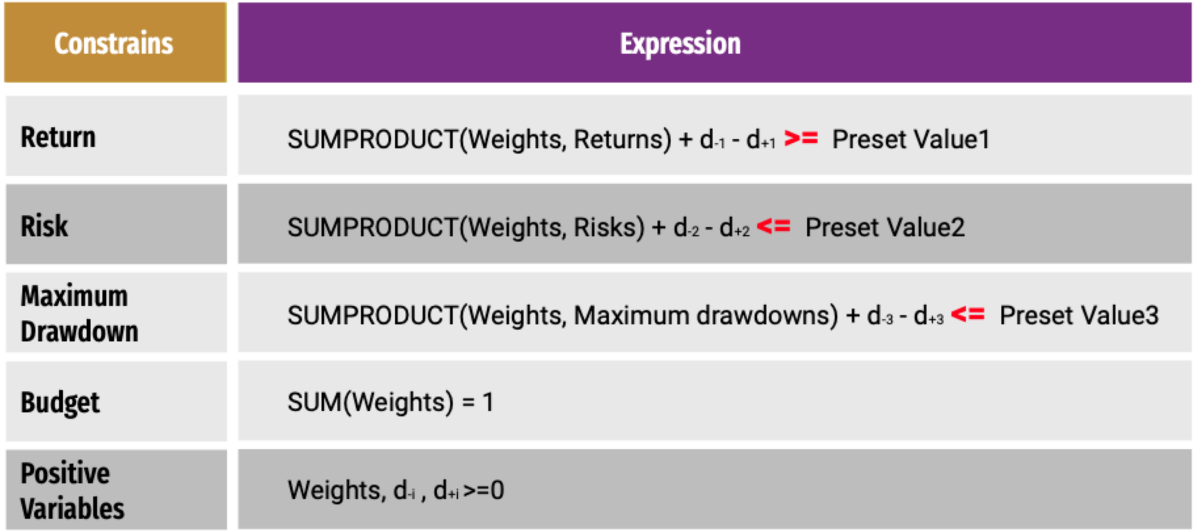


	Budget (SGD)	Risk Level	Return	The Largest Loss	
Customer A	10K	1%	6%	5%	Low
Customer B	50K	3%	12%	10%	Mid
Customer C	1B	5%	30%	15%	High

5.4 Optimization- Solution

- We utilize Goal Programming in Excel solver

Goals



Constrains	Expression
Return	SUMPRODUCT(Weights, Returns) + d_{-1} - d_{+1} **>=** Preset Value1
Risk	SUMPRODUCT(Weights, Risks) + d_{-2} - d_{+2} **<=** Preset Value2
Maximum Drawdown	SUMPRODUCT(Weights, Maximum drawdowns) + d_{-3} - d_{+3} **<=** Preset Value3
Budget	SUM(Weights) = 1
Positive Variables	Weights, d_{-i} , d_{+i} >=0

Variables

- ➢ **"Weights"** : the allocation of the total budget as a proportion for each individual stock.
- ➢ **“d_i”:** proportional deviation

The GP objective is to minimise the **objective function, *Z***

$$Z = \sum_{i=1}^{3} (d_{-i} + d_{+i})/t_i$$

ARCADIA CAPITAL

5.4 Optimization- Example

Middle Class Family

- Seek for normal return rate

1

Budget： 50k SGD

	Return	Risk	Maximum Drawdown
Level	Mid	Mid	Mid
Set Value	12%	3%	10%

⇩

Product	Return	Risk	Max down	Invest Amount
No.14	40.28%	4.57%	4.35%	12000
No.1	1.3%	0.44%	3.16%	3000
…	…	…	…	…
Total	21.4%	2.5%	8%	50k

5.4 Optimization- Example

Retired Seniors

- Seek for stable and long term return

2

Budget: 10k SGD

	Return	Risk	Maximum Drawdown
Level	Low	Low	Low
Set Value	6%	1%	5%

⇩

Product	Return	Risk	Max down	Invest Amount
No.14	40.28%	4.57%	4.35%	7300
No.1	1.3%	0.44%	3.16%	2600
…	…	…	…	…
Total	7.6%	0.87%	2.2%	10k

ARCADIA CAPITAL

5.4 Optimization- Example

Rich Guys

- Seek for maximum return, able to bear high risk.

3

Budget: 100k SGD

	Return	Risk	Maximum Drawdown
Level	High	High	High
Set Value	30%	5%	15%

Product	Return	Risk	Max down	Invest Amount
No.1	1.3%	0.44%	3.16%	12000
No.2	2.15%	0.74%	7.48%	6000
…	…	…	…	…
Total	36%	4.4%	15%	100k

Discussion

1. Outcome Discussion
2. Limitations & Further Prospects

6. Outcome Discussion

The products utilizing **Survival Analysis** have attributes of low return and low risk. e.g. it is more fit for current and short-term investment in small amount.

The **Multifactor Method** is not able to construct profitable products in our analytics as this method is highly influenced by the market performance and new energy industry is in loss in our testing period(Mar 2023).

6. Limitations & Further Prospects

The scope of stock is not wide enough.

e.g.,stratified sampling based on capitalization in an industry

stocks in multiple industries

The scope of test set is not wide enough

e.g., longer period as test set (not limited to 3 months)

No quantitative evaluation for customer risk, e.g. conjoint analysis

Thank you for listening!