SPINNAKER ANALYTICS

Capstone Project: Spinnaker

Identifying and Predicting Tradable Signals

Team B1
Carmen Cheung, Haiyuan Zhang, Linh To, Song Lin, Zichen Wang

Agenda

1 Introduction

1. Background
2. EDA
3. Hypotheses

2 Models

1. Logistic Regression
2. VAR
3. ARIMA
4. LSTM
5. SARIMAX
6. DES

3 Best Models

1. SARIMAX
2. DES
3. Results

4 Conclusion

1. Interpretations
2. Findings
3. Limitations

Introduction

Background | EDA | Hypotheses

Background Information

Who?

Analytical team of a FOF manager

What?

Define and identify tradable signals based on past performance to provide insight that supplements and aids in deciding investment strategies

How?

Build and evaluate models targeting the 20 assets in the dataset

Dataset Overview

Three Datasets	Columns	Description
US Sector Inst ETF	ReportDate	Weekly data from 2006 to 2017
	AssetClass	Name of portfolio
US Sector Inst MF	Flow	How much money ($M) is coming in or out
	AssetsEnd	Assets at the end of the week ($M)
US Sector Retail MF	FlowPct	Flow/Assets beginning of the week * 100%
	PortfolioChangePct	Percent change in overall portfolio during week

US Sector Inst ETF

Investments made/redeemed by institutional investors in Exchange Traded Funds.

US Sector Inst MF

Investments made/redeemed by institutional investors in Institutional Mutual Fund.

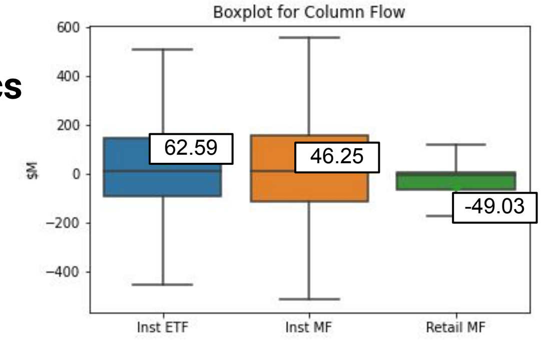
US Sector Retail MF

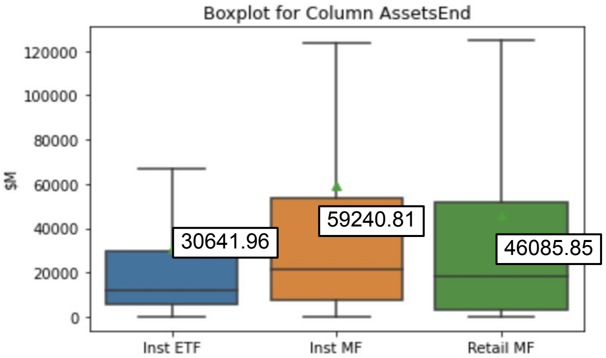
Investments made/redeemed by individual investors in their portfolios.

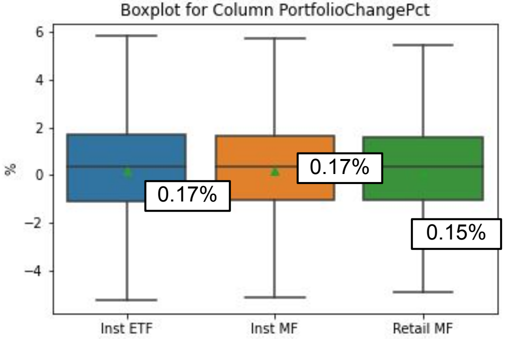
Recap of Exploratory Data Analysis

1. Basic Statistics

= Avg

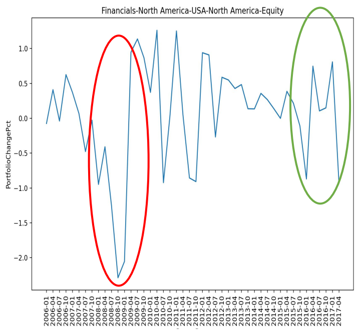


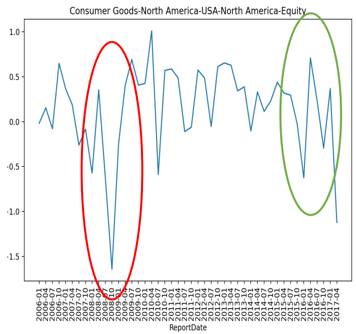


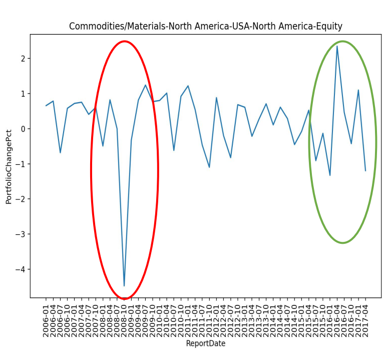


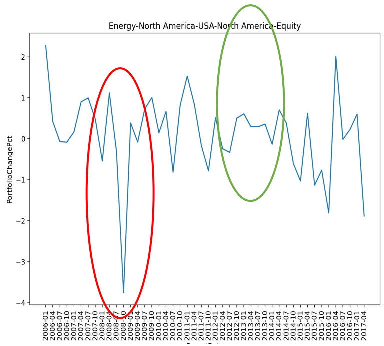
2. Avg Portfolio Change% Common Trend

(Sample of 4 industries)









An observed dive around Oct 2008 and upsurge around April 2016 across all industries → potentially due to systematic factors that impact the whole market

Recap of Exploratory Data Analysis

Example Industry:
Technology

3. Flow Trend

- ETF and Inst MF have similar trends
- Retail MF trend shrunk after 2012

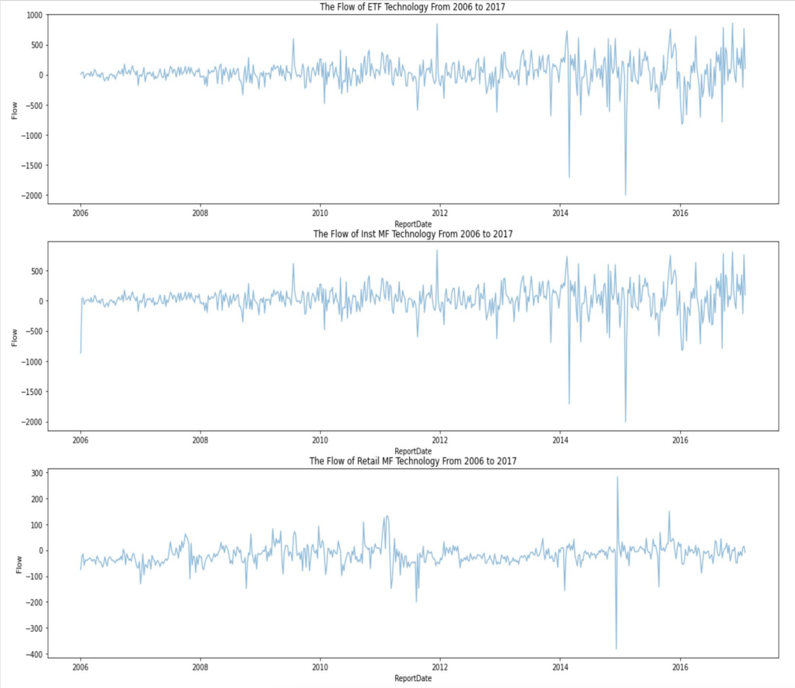
4. AssetEnd Trend

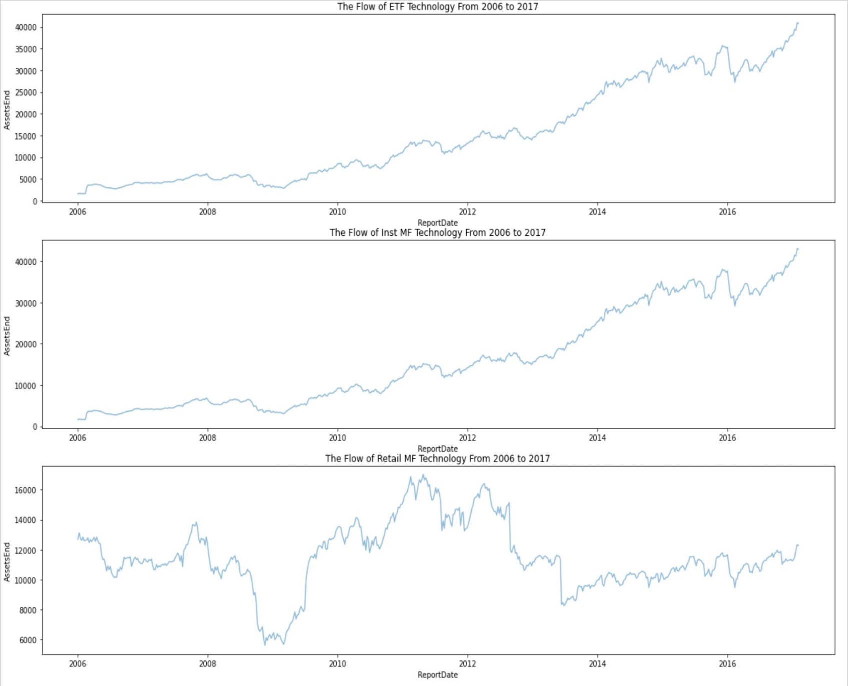
- ETF and Inst MF have similar trends
- Retail MF experience higher fluctuations

ETF

Inst MF

Retail MF





Years (2006 - 2017)

Hypotheses for Predicting a Tradeable Signal

01 Smoothing the Trend

Using moving average to smooth the trend can help indicate future patterns.

02 Effect of the Market

Removing systematic market factors using the S&P 500 as a proxy could indicate which industries are performing better or worse relative to the market

03 Time Series Models

Time-series models are expected to perform the best since the order of the data matters in predicting future values

04 Best Model per Industry Varies

Each industry may have unique inherent features that can be explained better with different models according to each model's strength.

Target Variable Label Generation

01

Rolling Average of “AssetEnd”

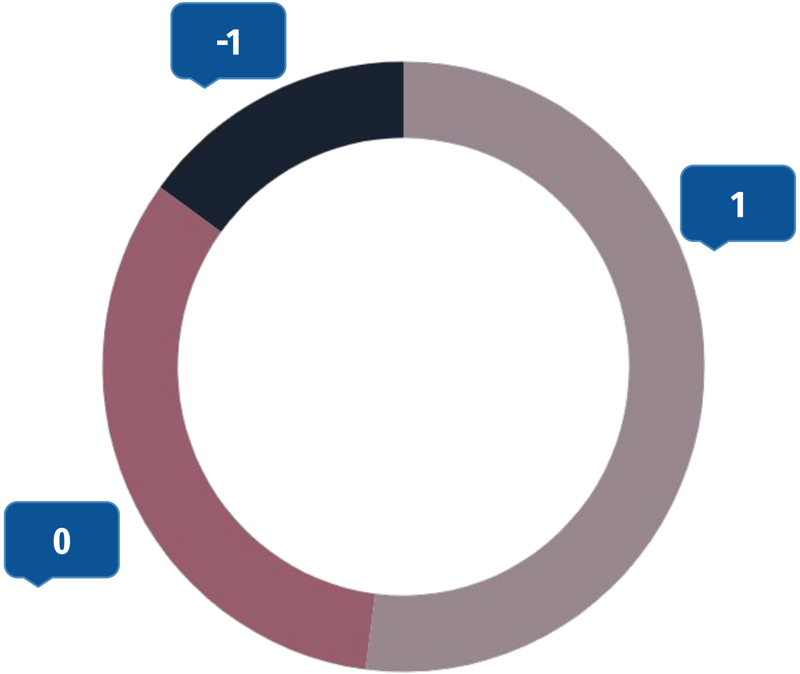
Method:
Measure the industry’s future performance based on the assetend variable
Label:
Percent change based on rolling average of asset/ Rolling average
Economic Meaning:
Measuring the overall risk of the portfolio

02

Removing Effect of the Market

Method:
Use CAPM model to get excess return related to idiosyncratic risk as the measure of future performance
Label:
Percent change of future rolling average for future excess return
OR
T-test result for future excess return
Economic Meaning:
Measuring the idiosyncratic risk

Breakdown of Balance in Classification Labels



52% **1** **Label = 1**

Change in “AssetEnd” > 5%

33% **0** **Label = 0**

-5% < Change in “AssetEnd” < 5%

15% **- 1** **Label = - 1**

Change in “AssetEnd” < -5%

MODELS

Logistic Regression | VAR | SARIMAX | LSTM | DES

Overview of Model Design

Category	Classification	Time - Series			Neural Net	Forecast
Model	Logistic Regression	VAR	ARIMA	SARIMAX	LSTM	DES
Train/Test Ratio	80/20	Test on last 10 weeks	90/10	80/20	80/20	NA
Target Variable	Label	AssetEnd	AssetEnd	AssetEnd	Label	AssetEnd
Predictors	Flow FlowPct AssetEnd PortfolioChange%	Flow FlowPct AssetEnd PortfolioChange%	AssetEnd (Train/Test)	Flow FlowPct AssetEnd PortfolioChange% Moving Average 16/50 week	Flow FlowPct AssetEnd PortfolioChange%	AssetEnd
Evaluation Metric	AUC	RMSE	RMSE	MSE AUC	Accuracy	Accuracy/AUC /MSE

Recap of Initial Models

Logistic Regression

- The underlying implication is to flatten out the values and measure the AUC by taking a threshold
- In our model we created a new column called “Rolling average” for “Asset Ends” and used the same method of MA15 to create trading signals

VAR

- Vector auto regression
- Each variable is a linear function of past lags of itself and past lags of the other variables
- VAR performed the best with Large Cap Growth and Small Cap Blend for Retail dataset with an MSE of 254 and 771, but not with other asset classes

ARIMA

- Autoregressive integrated moving average
- ARIMA model works well on most asset classes in three dataset except Large_Cap_Growth and Small_Cap_Growth in Retail dataset
- Has good MSE but doesn’t account for relationship between independent variables

MSE on VAR Model

	Etf	Institutional	Retail
model_Utilities	169229	220717	230438
model_Commodities/Materials	520807	532447	1603336
model_Consumer Goods	137509	137509	156514
model_Energy	7137668	12140902	143104
model_Financials	45174061	49626576	10004292
model_Health Care/Biotech	657868	1289787	522038
model_Industrials	2464983	2464983	44807894
model_Large Cap Blend	693291565	1201883001	7417725
model_Large Cap Growth	36367426	11874972	$\boxed{254}$
model_Telecom	69397	69397	733402
model_Mid Cap Blend	22428921	46015415	1236547
model_Technology	4179707	4336722	5741649
model_Small Cap Value	5296657	13461327	222554
model_Small Cap Growth	968943	1517035	561745
model_Large Cap Value	21254147	80403700	63447
model_Real Estate	2754912	2507222	110281
model_Mid Cap Value	1566044	7344058	1238
model_Mid Cap Growth	122381	27906	2289
model_Small Cap Blend	75708490	133372346	$\boxed{771}$

Results of VAR and ARIMA

MSE on ARIMA Model

	Etf	Institutional	Retail
model_Utilities	2.965695e+06	3.270422e+06	1.362161e+07
model_Commodities/Materials	1.689366e+06	2.048223e+06	2.331048e+08
model_Consumer Goods	1.902484e+07	1.902484e+07	1.169006e+05
model_Energy	4.155100e+07	9.000894e+07	7.182761e+06
model_Financials	1.814615e+07	2.029748e+07	8.775202e+08
model_Health Care/Biotech	4.758217e+07	5.656915e+07	2.954015e+08
model_Industrials	1.225051e+07	1.225051e+07	5.308930e+08
model_Large Cap Blend	1.170641e+09	4.386867e+09	6.536446e+07
model_Large Cap Growth	1.906099e+07	8.711199e+08	4.528610e+02
model_Telecom	1.317453e+05	1.317453e+05	4.137518e+07
model_Mid Cap Blend	2.636198e+07	4.618157e+07	1.797201e+07
model_Technology	1.123849e+07	1.359662e+07	1.309856e+08
model_Small Cap Value	8.424452e+06	3.668705e+07	3.710688e+08
model_Small Cap Growth	1.306339e+06	2.939469e+07	4.454934e+07
model_Large Cap Value	2.156147e+08	6.807123e+08	1.843846e+06
model_Real Estate	1.242387e+07	2.215344e+08	6.224666e+07
model_Mid Cap Value	7.153766e+07	2.762401e+08	6.177286e+03
model_Mid Cap Growth	3.179709e+06	1.965512e+08	2.975884e+03
model_Small Cap Blend	7.030059e+07	1.096481e+08	3.739333e+02

LSTM

Model1

- Train Test split: 80%-20%
- Independent variables:
  - Flow
  - FlowPct
  - PortfolioChangePct
- Target variable:
  - AssetsEnd
- Optimizer: Adam
- Metrics: mean absolute error

Results

- Training mean absolute error: 15310.3779
- Validation mean absolute error: 48351.1367

Caveats:

- LSTM need large scale of data for training
- For a single asset class, we only have 463 observations as training set

Model2

- Train Test split: 80%-20%
- Independent variables:
  - Flow
  - FlowPct
  - PortfolioChangePct
- Target variable:
  - Labels
- Optimizer: Rmsprop
- Metrics: accuracy

Results

- Model not efficiently learning
- When making predictions, the outcomes are all null values

Selecting Best Models

Advantages and Disadvantages of Each Model

Logistic Regression	VAR	ARIMA	SARIMAX	LSTM	DES
Pro: Easy to understand and implement **Con:** The model doesn't account for time-series data and could only be used for classification problem	**Pro:** The estimate is flexible and less demanding in information and time **Con:** Has lower accuracy than SARIMAX and DES	**Pro:** High accuracy time-series model **Con:** Can only use 1 independent variable to predict and doesn't account for the interactions between variables	**Pro:** Extension of ARIMA that explicitly supports univariate time series data with a seasonal component **Con:** Has lower accuracy than Pure ARIMA model because we are adding extra independent variables	**Pro:** Relative insensitivity to gap length **Con:** Needs a large dataset to be able to work, hard to explain algorithm	**Pro:** Good at capturing the overall trend of a time series **Con:** Not accurate for predicting single observation

SARIMAX

Model Setup

Target Variable | AssetEnd

Train-Test split: 80%-20%

Independent Variables | Feature Engineering | Flow | FlowPct | PortfolioChangePct

Feature Engineering

We performed feature engineering to generate additional features for the model. These features are derived from the given dataset itself. These features also highlight some common trends that helps the model to predict better:

- **M16** : Moving Average 16 M16 uses a window of 16 weeks and calculates the values based on the past 16 moving average of the AssetsEnd.
- **M50** : Moving Average 50 M50 uses a window of 50 weeks and calculates the values based on the past 50 moving average of the AssetsEnd.

SARIMAX

——Model Tuning

Trend Elements

There are three trend elements that require configuration.

They are the same as the ARIMA model; specifically:

p: Trend autoregression order. 2

d: Trend difference order. 0

q: Trend moving average order. 0

Seasonal Elements

There are four seasonal elements that are not part of ARIMA that must be configured; they are:

P: Seasonal autoregressive order. 0

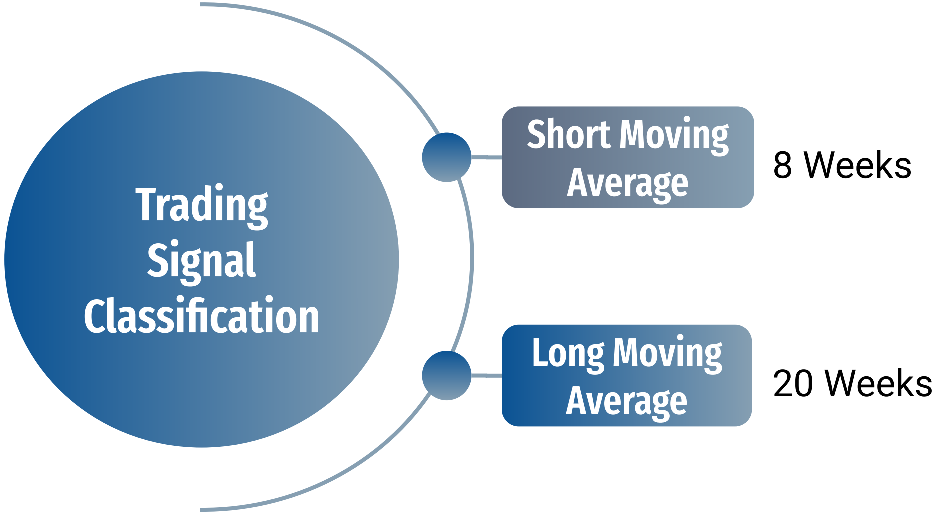
D: Seasonal difference order. 1

Q: Seasonal moving average order. 2

m: The number of time steps for a single seasonal period. 16

SARIMAX

——Trading Signal



A crossover occurs when a short moving average crosses a long moving average.

In stock trading, this meeting point can be used as a potential indicator to buy or sell an asset.

- When the short term moving average crosses above the long term moving average, this indicates a buy signal.
- When the short term moving average crosses below the long term moving average, it may be a good moment to sell.

* Reference:
https://towardsdatascience.com/making-a-trade-call-using-simple-moving-average-sma-crossover-strategy-python-implementation-29963326da7a

SARIMAX

——Function Building

Asset Class

- The 19 asset classes we have
- Financials, Energy, Industrials...

Dataset

- ETF
- Institutional
- Retail

Target

- AssetsEnd

Function

Prediction for AssetsEnd

<matplotlib.axes._subplots.AxesSubplot at 0x7fecaa37ba50>

SARIMAX

——Function Building

ETF - Technology Asset Class :AUC Score: 0.6911

Trading Signal for Test Data

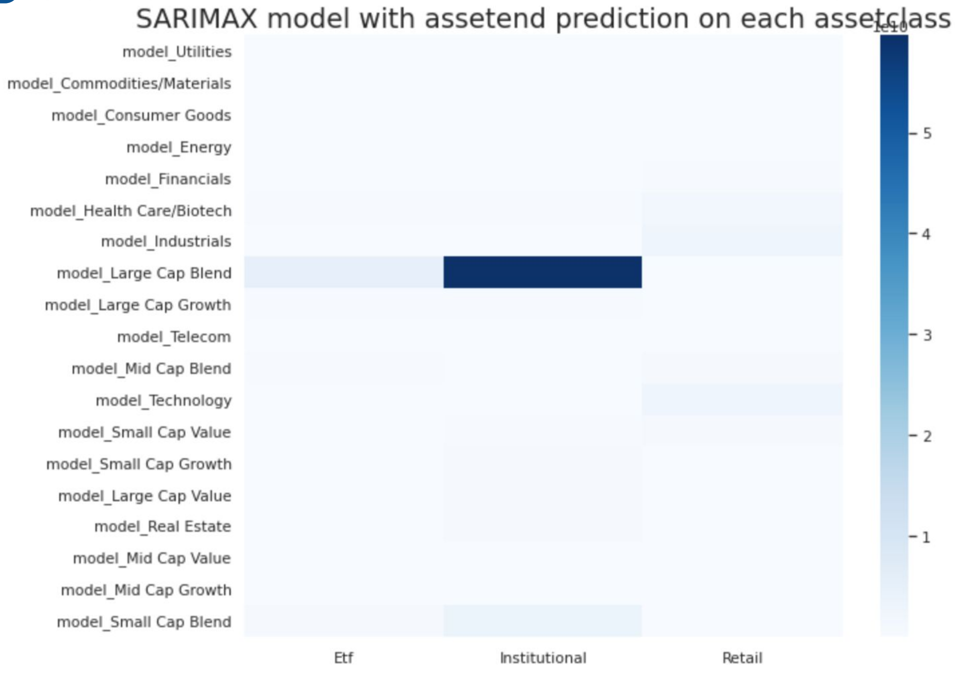
Trading Signal for Predicted Data

SARIMAX

——Combined Results

MSE Metric

	Etf	Institutional	Retail
model_Utilities	1.012354e+08	9.688173e+07	2.671870e+07
model_Commodities/Materials	1.726491e+08	1.777422e+08	1.506078e+08
model_Consumer Goods	3.050750e+07	5.274480e+07	3.482008e+06
model_Energy	9.074193e+07	2.618927e+07	1.295298e+07
model_Financials	3.170412e+07	2.714843e+07	4.386883e+08
model_Health Care/Biotech	4.328380e+08	4.397761e+08	1.860628e+09
model_Industrials	1.203487e+07	2.505699e+07	2.886196e+09
model_Large Cap Blend	4.957666e+09	5.961074e+10	1.519108e+07
model_Large Cap Growth	4.597751e+08	3.805982e+08	5.239626e+02
model_Telecom	7.881575e+04	1.513959e+05	9.743626e+06
model_Mid Cap Blend	2.732450e+08	2.310788e+08	8.586540e+08
model_Technology	7.983608e+06	4.582365e+07	2.820192e+09
model_Small Cap Value	3.840652e+07	4.176206e+08	8.553254e+08
model_Small Cap Growth	4.685072e+06	8.650262e+08	7.967155e+06
model_Large Cap Value	9.706309e+07	7.840565e+08	1.976894e+07
model_Real Estate	8.959693e+07	7.334740e+08	1.338696e+08
model_Mid Cap Value	1.494160e+07	2.159457e+07	1.947389e+05
model_Mid Cap Growth	7.571073e+06	7.878874e+07	1.721652e+04
model_Small Cap Blend	7.389203e+08	3.662471e+09	4.804297e+03



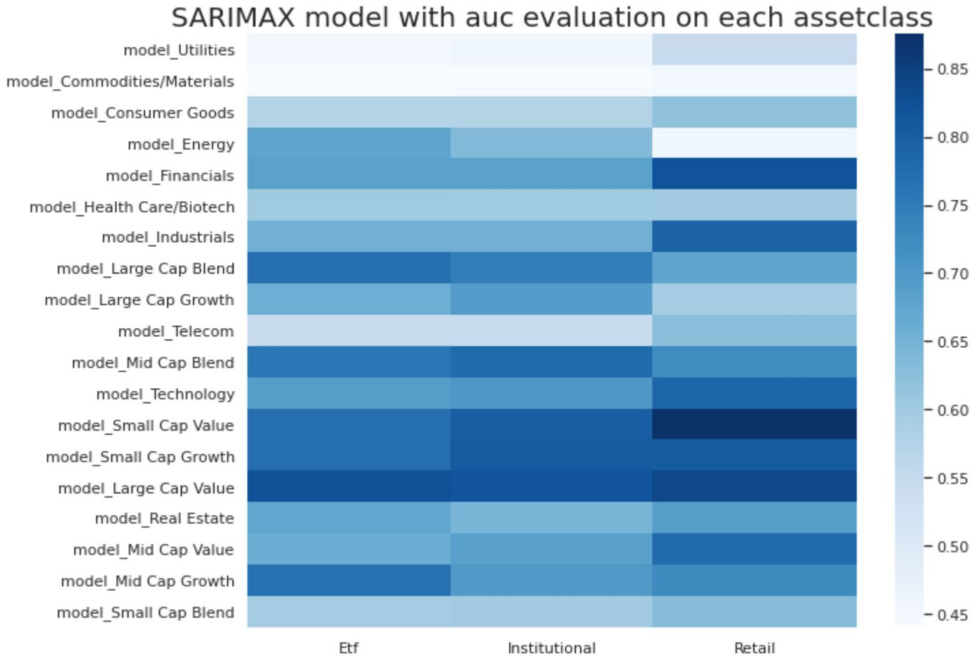
SARIMAX Model predicts AssetsEnd well for almost all asset class in three dataset except the Large Cap Growth class in institutional dataset

SARIMAX

——Combined Results

AUC Metric

	Etf	Institutional	Retail
model_Utilities	0.450820	0.458333	0.548513
model_Commodities/Materials	0.440000	0.444444	0.450000
model_Consumer Goods	0.575368	0.575368	0.620098
model_Energy	0.678922	0.634545	0.458333
model_Financials	0.684351	0.683817	0.820312
model_Health Care/Biotech	0.601852	0.600000	0.598214
model_Industrials	0.653639	0.653639	0.793103
model_Large Cap Blend	0.768293	0.744444	0.677419
model_Large Cap Growth	0.658570	0.692577	0.595987
model_Telecom	0.547181	0.547181	0.625632
model_Mid Cap Blend	0.756250	0.777027	0.721677
model_Technology	0.691106	0.702156	0.787500
model_Small Cap Value	0.768293	0.800956	0.875214
model_Small Cap Growth	0.770018	0.805735	0.804087
model_Large Cap Value	0.822874	0.818732	0.837209
model_Real Estate	0.675595	0.644720	0.688419
model_Mid Cap Value	0.662295	0.685152	0.778091
model_Mid Cap Growth	0.767875	0.698335	0.724405
model_Small Cap Blend	0.597561	0.598765	0.630824



SARIMAX Model predicts tradeable signal well for almost all asset class in three dataset except the Utilities, Commodities/Materials and Telecom class

SARIMAX

——Conclusion

- SARIMAX model did well with prediction of 116 week for all three datasets, but it does not capturing the Large Cap Blend industry that well compared with other industries we have.
- The model did well at predicting the “buy” signals than “sell” signals. The accuracy or AUC score of this method varies between different asset classes, but works the best for retail dataset compared with ETF and Institutional dataset, with an average of 60% accuracy
- When making actual investment decisions, one should consider whether the seasonal factor is significant enough to action, and we can always change the parameter to fit different prediction interval or situations. But the model only counts as a reference before making final investment decisions.

SARIMAX

——Business Interpretation

I would like to check the trading signals in the ETF portfolio for the next 10 weeks.
Should I buy or sell it?

Let's run the function with target class and ETF dataset

We found 5 trading signals for the future weeks.
Let's take this as a reference and discuss with finance team for next steps

Great!
The historic signal accuracy for this model is around 70%, it might be a good start for trading

Period 1

Period 2

Double Exponential Smoothing (DES) Forecast

Time Series Data | Randomness | Trend | Seasonality | Cycle

$$F_t = \boxed{\alpha} A_{t-1} + (1-\alpha) FIT_{t-1}$$

$$T_t = \boxed{\beta} (F_t - F_{t-1}) + (1-\beta)(T_{t-1})$$

$$FIT_{t+k-1} = F_t + kT_t$$

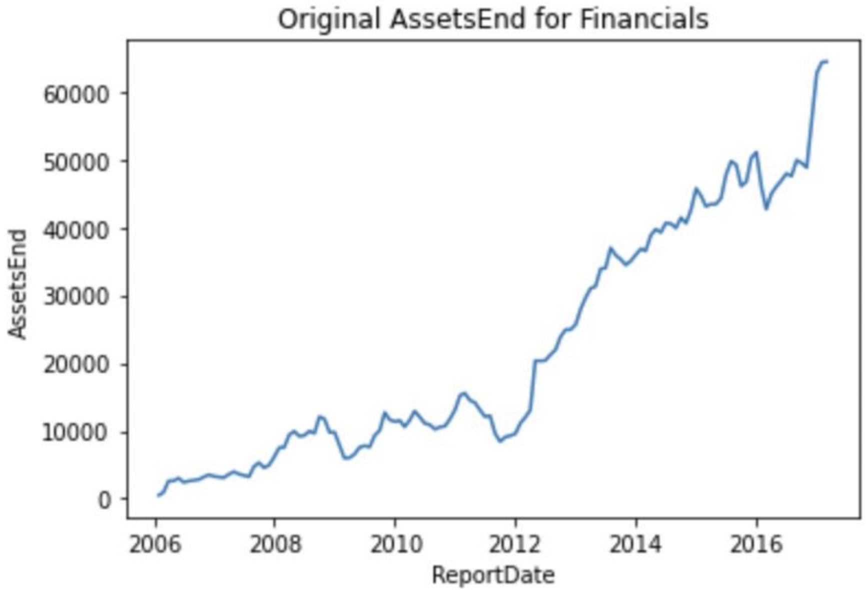
FIT : Forecast including trend
α : Base smoothing constant
β : Trend smoothing constant
k : Number of periods to forecast in the future
A : Actual observation

DES is good at capturing the trend of time series data and make prediction by past data.

Double Exponential Smoothing (DES) Forecast
——Example: Financials

AssetClass: Financial

Target Series: AssetEnd



	ReportDate	AssetsEnd
1737	2006-01-04	501.334569
1738	2006-01-11	492.598808
1739	2006-01-18	493.687562
1740	2006-01-25	496.072619
1741	2006-02-01	502.318241
...	...	...
2311	2017-01-04	64676.472421
2312	2017-01-11	65014.491385
2313	2017-01-18	63604.582397
2314	2017-01-25	63997.173726
2315	2017-02-01	64378.816729

579 rows × 2 columns

Double Exponential Smoothing (DES) Forecast
——Example: Financials

1. Data preprocessing

	ReportDate	AssetsEnd
1737	2006–01–04	501.334569
1738	2006–01–11	492.598808
1739	2006–01–18	493.687562
1740	2006–01–25	496.072619
1741	2006–02–01	502.318241
...	...	...
2311	2017–01–04	64676.472421
2312	2017–01–11	65014.491385
2313	2017–01–18	63604.582397
2314	2017–01–25	63997.173726
2315	2017–02–01	64378.816729

579 rows × 2 columns

Monthly average

	month	year	AssetsEnd
0	1	2006	495.923389
1	2	2006	929.448375
2	3	2006	2608.001460
3	4	2006	2619.976664
4	5	2006	3024.714756
...	...	...	...
129	10	2016	48797.631463
130	11	2016	55658.004831
131	12	2016	62735.054322
132	1	2017	64323.179982
133	2	2017	64378.816729

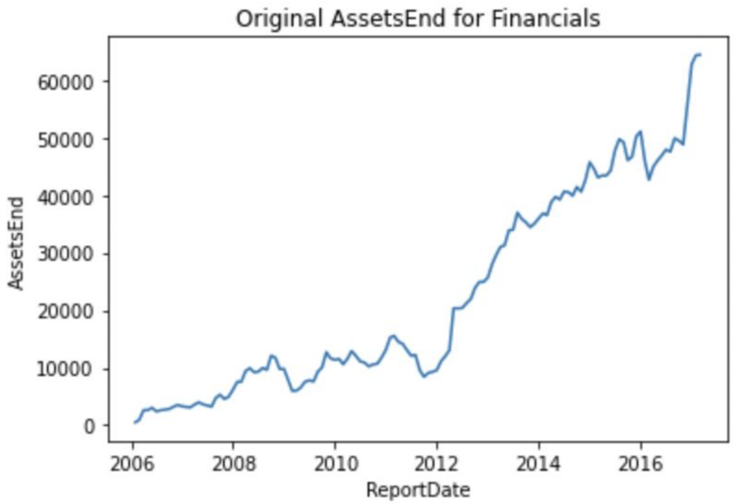
134 rows × 3 columns

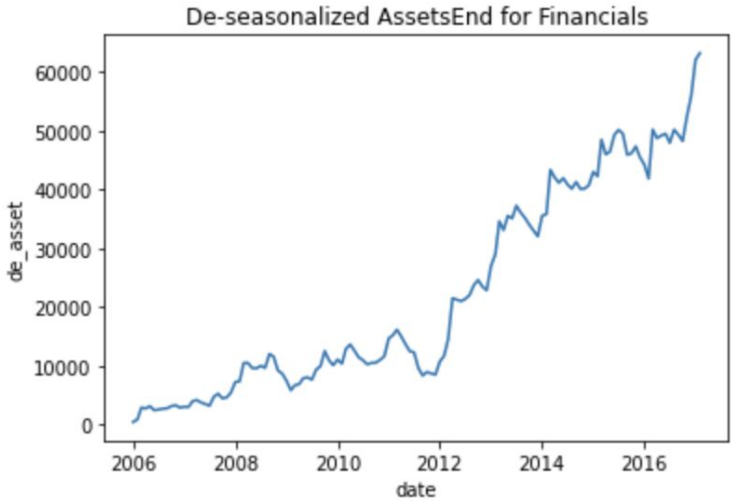
De-seasonalize

	month	year	de_asset
0	1	2006	478.704825
1	2	2006	912.376220
2	3	2006	2914.758547
3	4	2006	2776.409396
4	5	2006	3173.288024
...	...	...	...
129	10	2016	48254.079377
130	11	2016	52505.960326
131	12	2016	55918.818849
132	1	2017	62089.865638
133	2	2017	63196.303394

134 rows × 3 columns

Double Exponential Smoothing (DES) Forecast
——Example: Financials

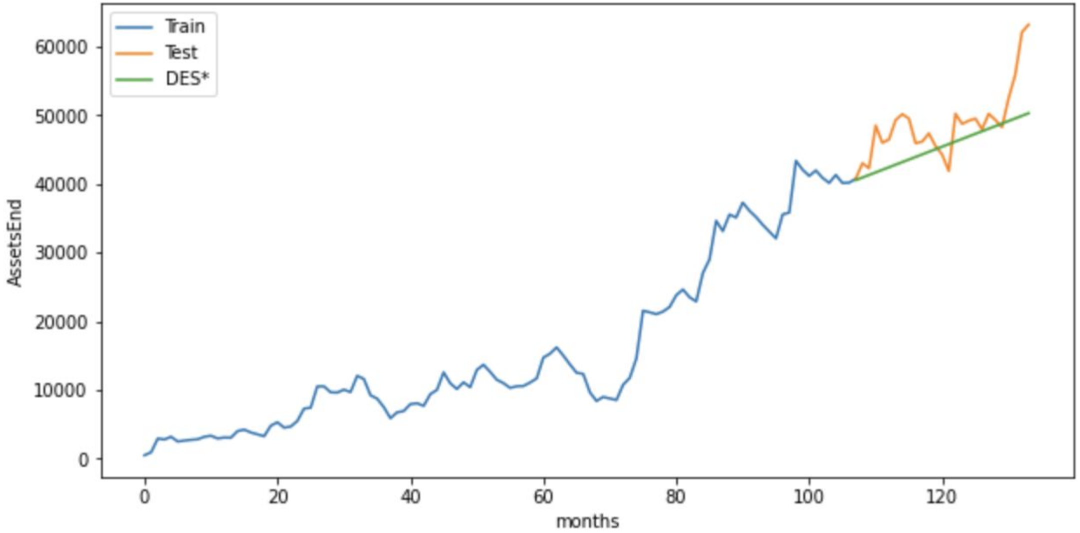




Capture the trend of the original time series

Double Exponential Smoothing (DES) Forecast
——Example: Financials

2. Train and predict



- Train-Test split: 80%-20%
- Burn-in period of 10 months
- α : Base Smoothing Constant (Try 0-1.0 with 0.05 step)
- β: Trend Smoothing Constant (Try 0-1.0 with 0.05 step)
- Evaluation metric: MSE
- Find the best combination of α and β by grid search

- Test MSE: 23931604.92(4892^2)
- DES is only good with capturing the smooth trend of one period after the train set

Double Exponential Smoothing (DES) Forecast
——Example: Financials

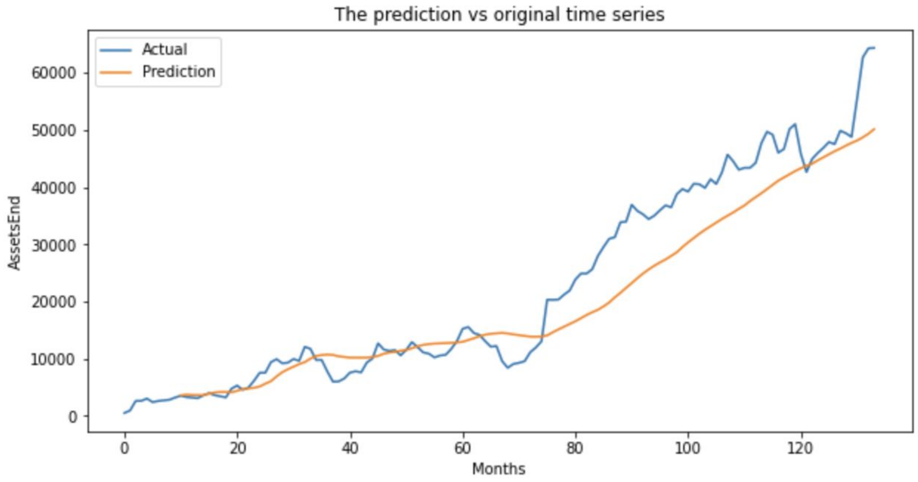
3. Extrapolate the method to the entire time series

First ten months → The eleventh month's prediction

First eleven months → The twelfth month's prediction

…

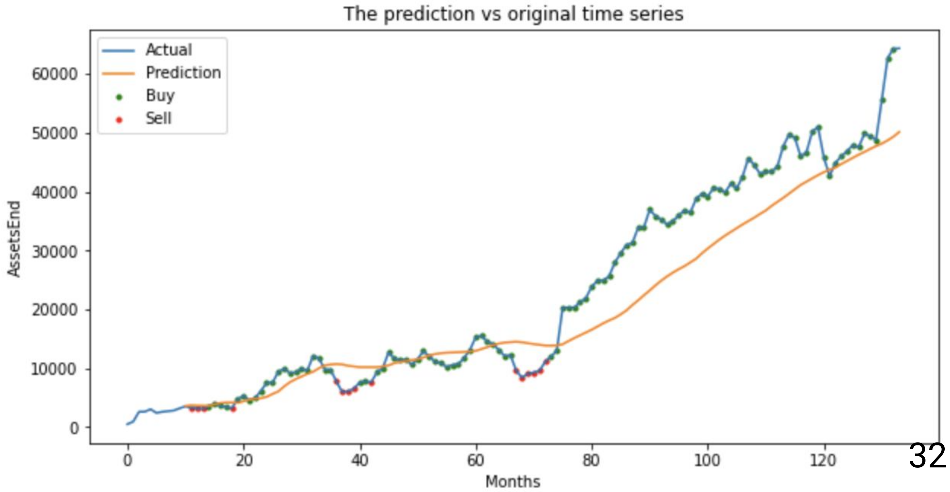
All months before the last one → The last month's prediction



4. Identify tradable signals

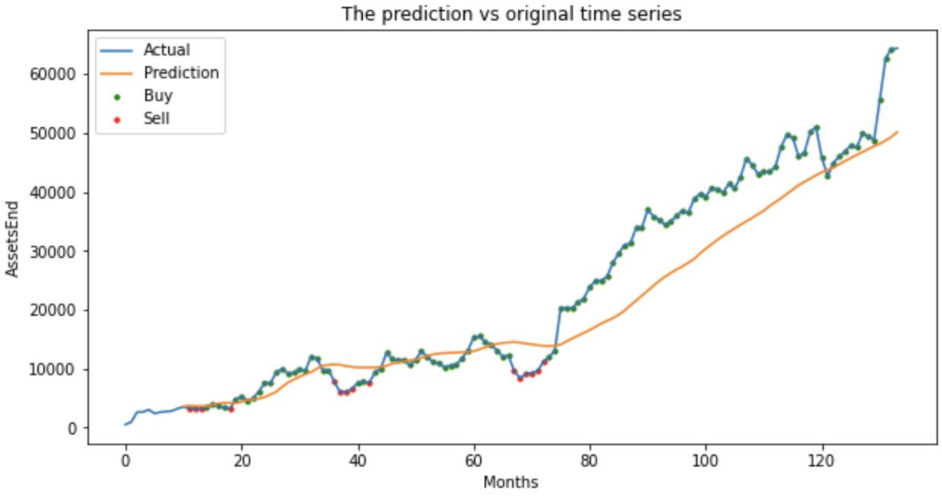
AssetsEnd	act_signal	prediction	pre_signal
3522.955867	Sell	3582.850000	Buy
3284.826647	Sell	3702.646531	Sell
3173.225080	Sell	3646.744123	Sell
3086.179979	Buy	3639.685322	Sell
3576.932134	Buy	3618.643203	Buy
3970.014798	Sell	3855.575593	Buy
3621.189334	Sell	4088.782423	Buy
3412.667455	Sell	4173.705374	Buy
3222.398201	Buy	4177.198997	Sell
4752.533635	Buy	4118.206820	Buy
5298.463066	Sell	4372.558293	Buy

- Standing at this month, we can get the prediction of next month.
- If we see the prediction of next month (3702.65 in this case) is higher than this month, we identify that we should "buy" this month.
- If lower, we should "sell" .
- If the same, we should "hold" (rarely the case).
- The same with "actual signal".



Double Exponential Smoothing (DES) Forecast ——Example: Financials

5. Evaluation



- Accuracy: 56.10%
- Sensitivity: 85.53%
  - How many of the "buy" signals were correctly identified
- Specificity: 8.51%
  - How many of the "sell" signals were correctly identified
- Precision: 60.19%
  - How many predictive "buy" signals were correct

Double Exponential Smoothing (DES) Forecast
——Example: Financials

6. Different intervals

AssetsEnd	act_signal3	prediction	pre_signal3
3284.826647	Buy	3702.646531	Sell
3173.225080	NaN	3646.744123	NaN
3086.179979	NaN	3639.685322	NaN
3576.932134	Sell	3618.643203	Buy
3970.014798	NaN	3855.575593	NaN
3621.189334	NaN	4088.782423	NaN
3412.667455	Buy	4173.705374	Buy
3222.398201	NaN	4177.198997	NaN
4752.533635	NaN	4118.206820	NaN
5298.463066	Buy	4372.558293	Buy
4542.328511	NaN	4677.027404	NaN

- Since DES is better at capturing the trend, if we enlarge the interval to identify the signals, we could get better results
- For every three months, identify one signal
- Three months interval accuracy: 63.41%
- Six months interval accuracy: 90%

Double Exponential Smoothing (DES) Forecast
——Function Building

Asset

- The 19 asset classes we have
- Financials, Energy, Industrials...

database

- ETF
- INS
- RE

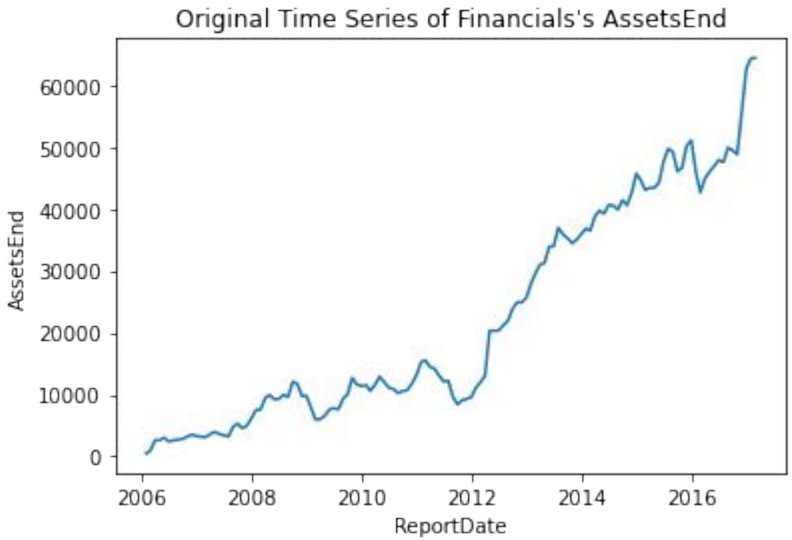
target

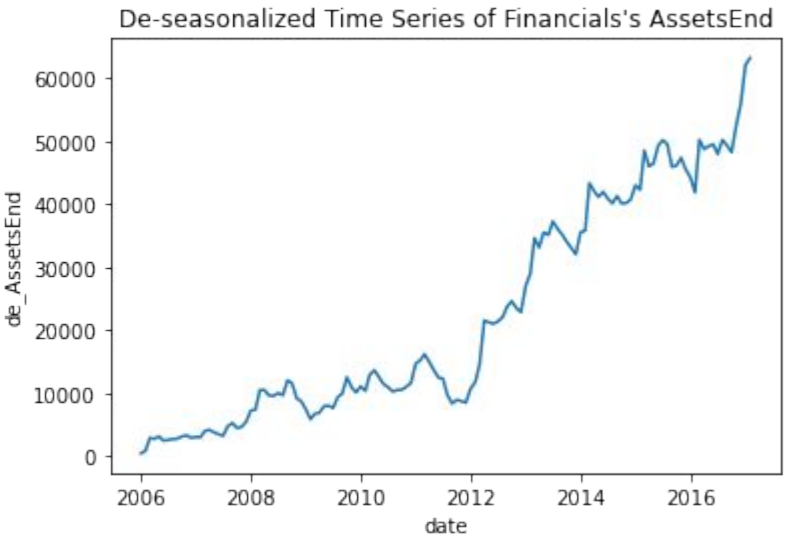
- AssetsEnd
- PortfolioChangePct

Function

Double Exponential Smoothing (DES) Forecast
——Function Building

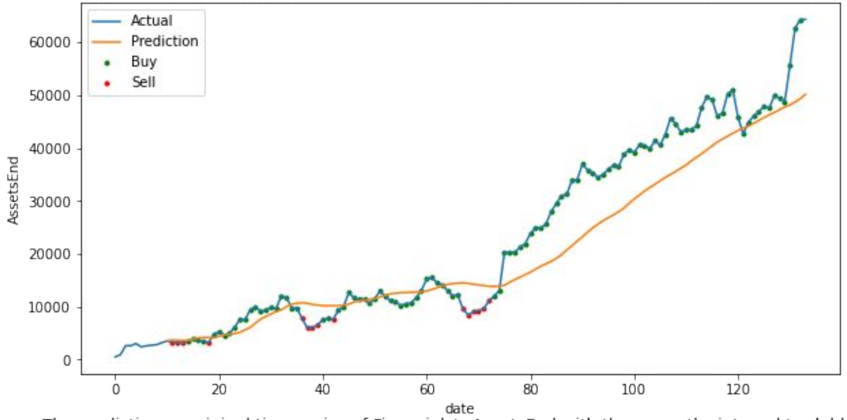
Function outputs



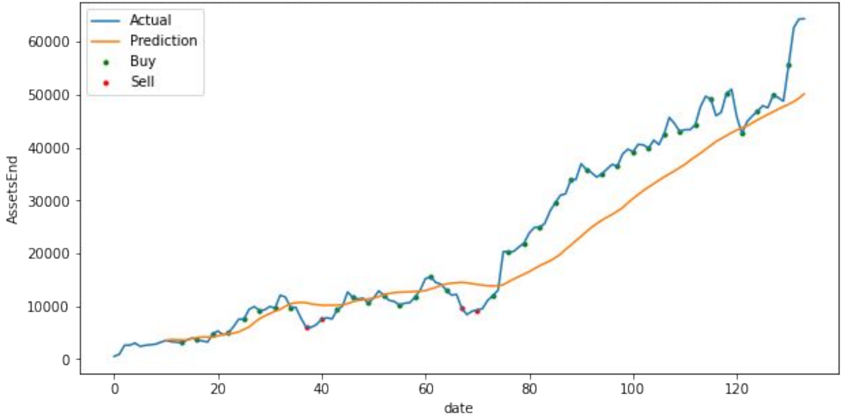


Double Exponential Smoothing (DES) Forecast
——Function Building

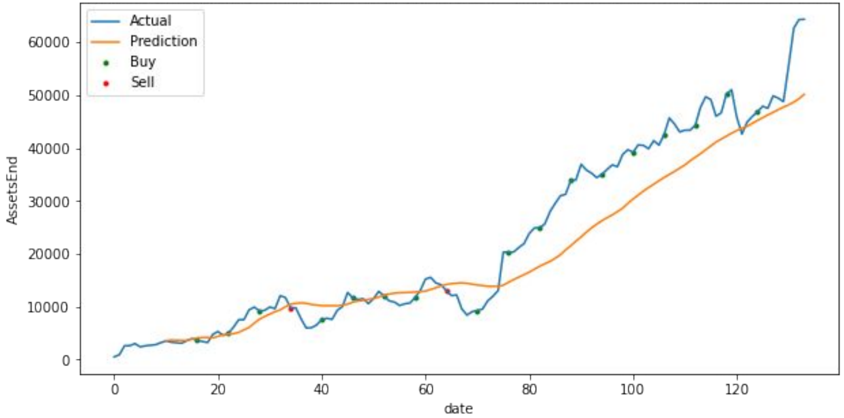
The prediction vs original time series of Financials's AssetsEnd with monthly tradable signals.



The prediction vs original time series of Financials's AssetsEnd with three months interval tradable signals.



The prediction vs original time series of Financials's AssetsEnd with six months interval tradable signals.



Double Exponential Smoothing (DES) Forecast ——Function Building

- MSE between prediction and original time series: 35348994.52
- One month interval:
  - Accuracy: 56.10%
  - AUC score: 0.47
- Three months interval:
  - Accuracy: 73.17%
  - AUC score: 0.51
- Six months interval:
  - Accuracy: 90.00%
  - AUC score: 0.75

One month interval	
True Positive: 65	False Negative: 11
False Positive: 43	True Negative: 4

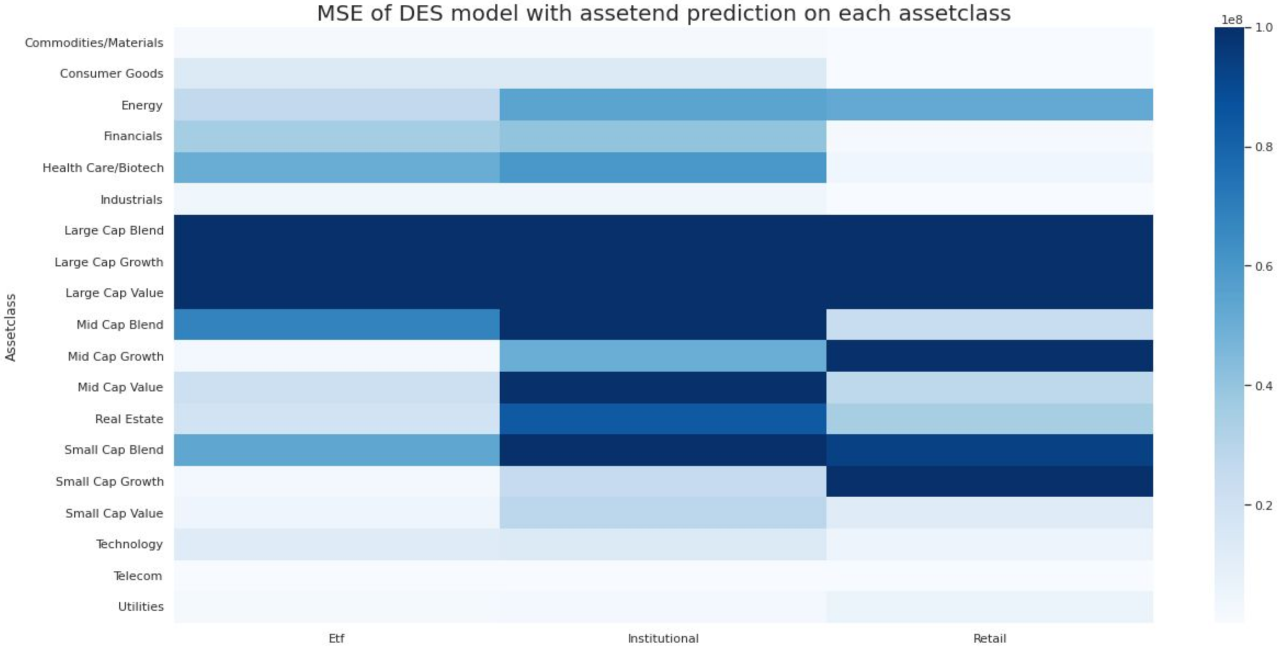
Three months interval	
True Positive: 29	False Negative: 3
False Positive: 8	True Negative: 1

Six months interval	
True Positive: 16	False Negative: 0
False Positive: 2	True Negative: 2

Double Exponential Smoothing (DES) Forecast
——Combined results

AssetsEnd MSE

Assetclass	Etf	Institutional	Retail
Commodities/Materials	1.470587e+06	1.546950e+06	9.847805e+03
Consumer Goods	1.371557e+07	1.371557e+07	3.421118e+03
Energy	2.647876e+07	5.443744e+07	5.210162e+07
Financials	3.534899e+07	4.053019e+07	1.578976e+06
Health Care/Biotech	5.076275e+07	6.012232e+07	3.936656e+06
Industrials	4.193728e+06	4.193728e+06	1.465923e+02
Large Cap Blend	3.538437e+09	1.198842e+10	9.281338e+08
Large Cap Growth	1.085933e+08	8.044643e+08	4.811277e+08
Large Cap Value	1.831616e+08	6.693287e+08	7.465348e+08
Mid Cap Blend	6.805939e+07	1.395607e+08	2.384740e+07
Mid Cap Growth	2.489287e+06	5.026462e+07	1.036738e+08
Mid Cap Value	2.143987e+07	1.307688e+08	2.771365e+07
Real Estate	1.884065e+07	8.394621e+07	3.482314e+07
Small Cap Blend	5.357041e+07	1.310897e+08	9.311448e+07
Small Cap Growth	2.096678e+06	2.530742e+07	1.093129e+08
Small Cap Value	4.848605e+06	2.820317e+07	1.206728e+07
Technology	1.232921e+07	1.350849e+07	5.554423e+06
Telecom	9.252694e+04	9.252694e+04	1.724722e+02
Utilities	1.758965e+06	2.211257e+06	5.931096e+06

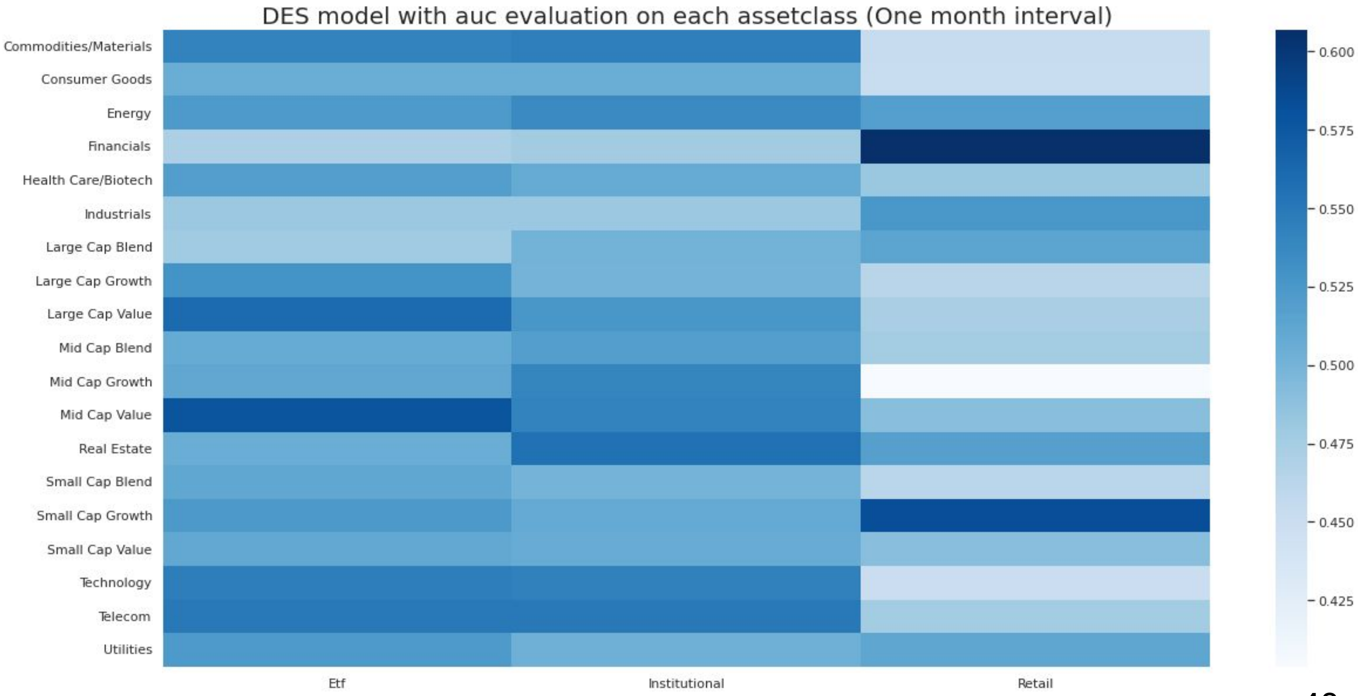


Double Exponential Smoothing (DES) Forecast

——Combined results

One month AUC

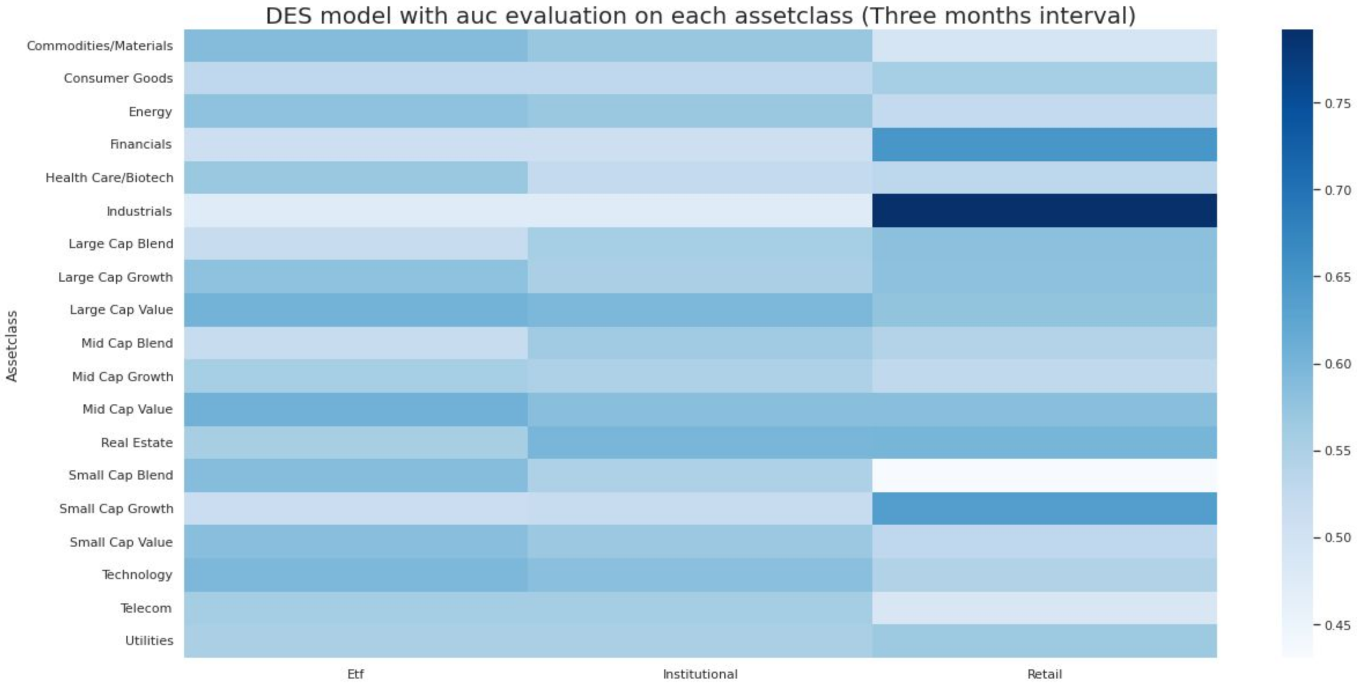
Assetclass	Etf	Institutional	Retail
Commodities/Materials	0.541948	0.545045	0.452233
Consumer Goods	0.505721	0.505721	0.451447
Energy	0.523455	0.535983	0.519380
Financials	0.471110	0.476891	0.607015
Health Care/Biotech	0.520396	0.508333	0.481548
Industrials	0.481419	0.481419	0.525862
Large Cap Blend	0.478221	0.501113	0.513988
Large Cap Growth	0.528730	0.500313	0.463636
Large Cap Value	0.560373	0.526391	0.472540
Mid Cap Blend	0.508333	0.520122	0.476399
Mid Cap Growth	0.510965	0.540778	0.403618
Mid Cap Value	0.577687	0.542560	0.490428
Real Estate	0.505446	0.555952	0.517429
Small Cap Blend	0.511773	0.500147	0.462500
Small Cap Growth	0.523830	0.508537	0.583333
Small Cap Value	0.510355	0.507317	0.490556
Technology	0.546189	0.544038	0.449657
Telecom	0.549906	0.549906	0.476653
Utilities	0.522807	0.502985	0.512698



Double Exponential Smoothing (DES) Forecast
——Combined results

Three months AUC

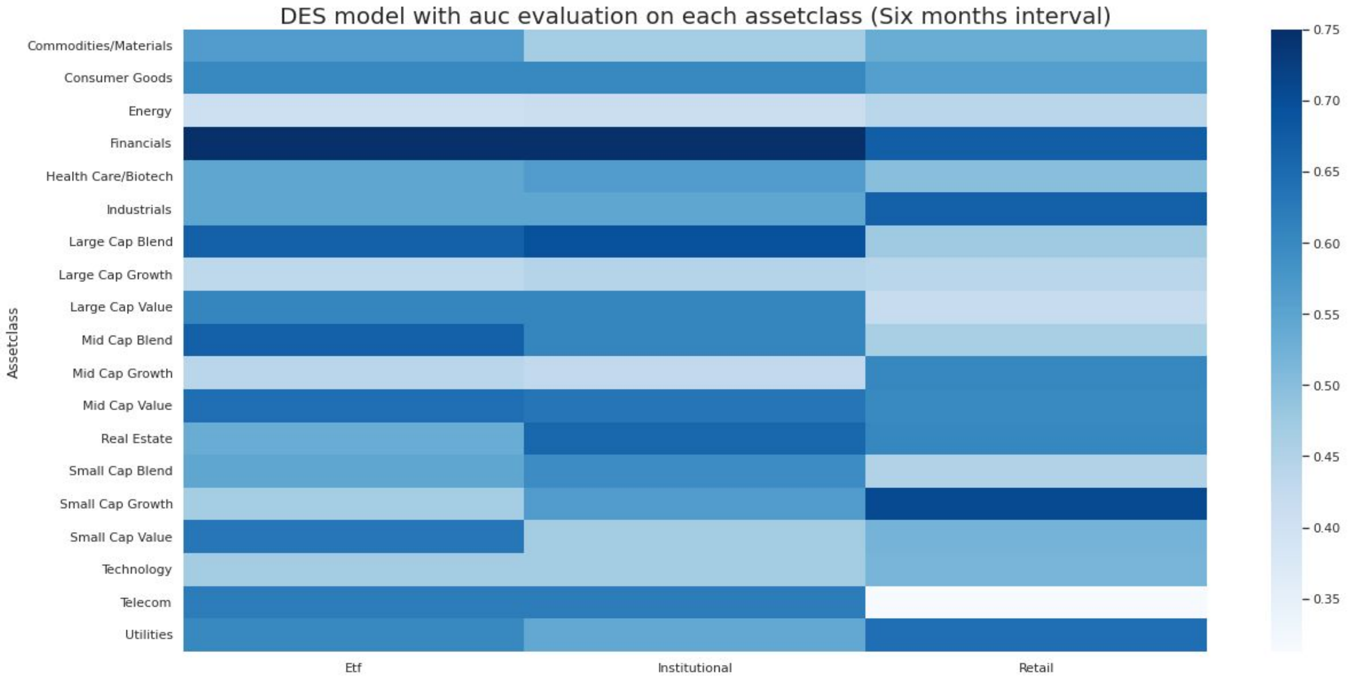
Assetclass	Etf	Institutional	Retail
Commodities/Materials	0.587500	0.570513	0.491667
Consumer Goods	0.528788	0.528788	0.556818
Energy	0.579670	0.570106	0.523352
Financials	0.508681	0.508681	0.648284
Health Care/Biotech	0.570106	0.523352	0.531746
Industrials	0.475641	0.475641	0.792208
Large Cap Blend	0.517742	0.557576	0.582418
Large Cap Growth	0.579861	0.551613	0.580460
Large Cap Value	0.603030	0.594697	0.576250
Mid Cap Blend	0.519355	0.564236	0.541209
Mid Cap Growth	0.557576	0.548851	0.526923
Mid Cap Value	0.605769	0.585484	0.584656
Real Estate	0.556034	0.597701	0.601449
Small Cap Blend	0.586364	0.548611	0.430622
Small Cap Growth	0.512121	0.517742	0.636905
Small Cap Value	0.585979	0.567742	0.528205
Technology	0.595486	0.583871	0.544686
Telecom	0.557971	0.557971	0.486715
Utilities	0.551932	0.551932	0.567460



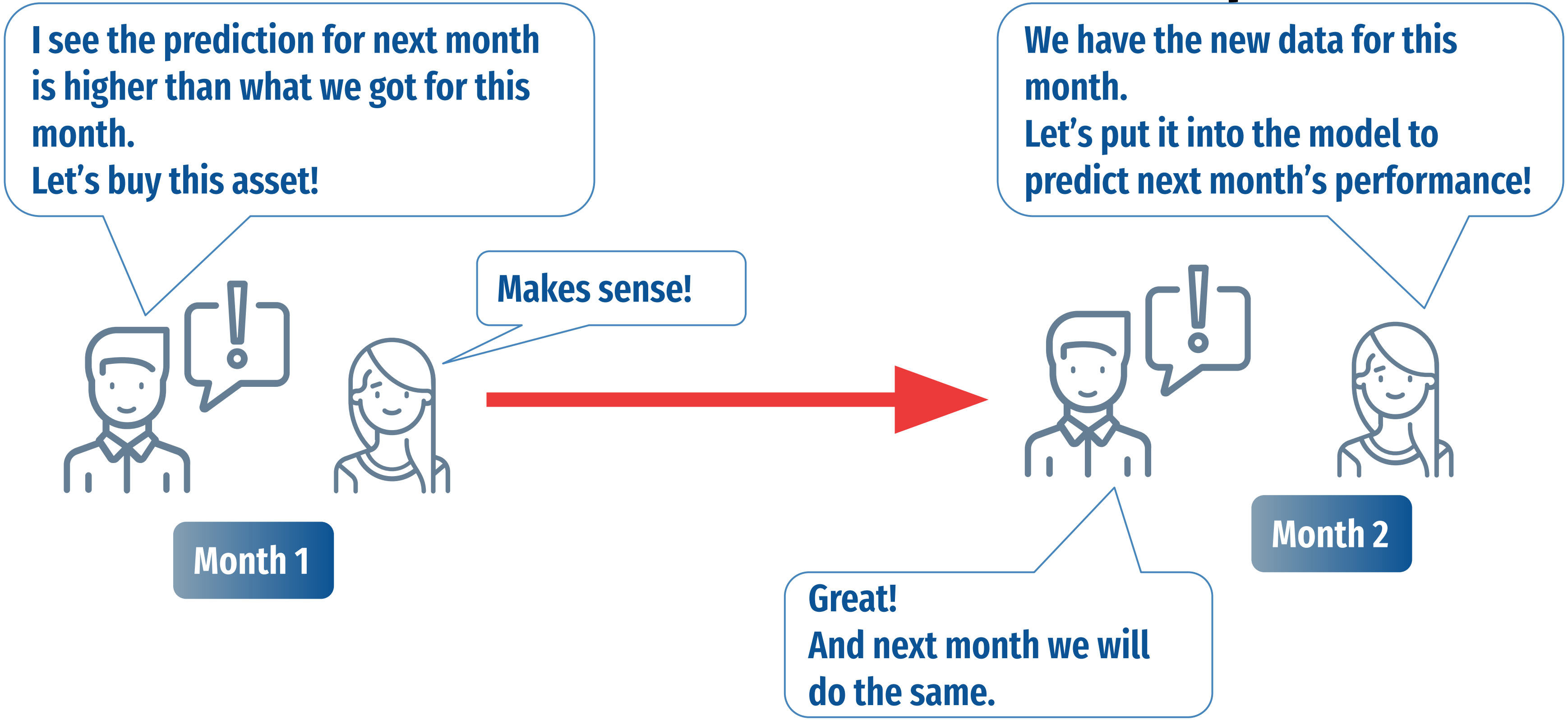
Double Exponential Smoothing (DES) Forecast
——Combined results

Six months AUC

Assetclass	Etf	Institutional	Retail
Commodities/Materials	0.566667	0.466667	0.533333
Consumer Goods	0.600000	0.600000	0.561111
Energy	0.406250	0.411765	0.437500
Financials	0.750000	0.750000	0.670330
Health Care/Biotech	0.547619	0.566667	0.500000
Industrials	0.547619	0.547619	0.666667
Large Cap Blend	0.666667	0.694444	0.476190
Large Cap Growth	0.433333	0.444444	0.440476
Large Cap Value	0.607843	0.607843	0.417582
Mid Cap Blend	0.666667	0.607843	0.461538
Mid Cap Growth	0.437500	0.428571	0.604167
Mid Cap Value	0.642857	0.633333	0.598901
Real Estate	0.533333	0.656250	0.604167
Small Cap Blend	0.547619	0.593750	0.450000
Small Cap Growth	0.466667	0.566667	0.707071
Small Cap Value	0.630952	0.468750	0.520833
Technology	0.468750	0.468750	0.516484
Telecom	0.621212	0.621212	0.312500
Utilities	0.600000	0.541667	0.642857



Double Exponential Smoothing (DES) Forecast ——Implementation



Double Exponential Smoothing (DES) Forecast
——Conclusion

- Double Exponential smoothing may not do well with prediction of a single month's value, but it's good at capturing the overall trend. So it could be used for identifying tradable signals of longer period.
- The accuracy or AUC score of this method varies between different asset classes, because they all have different trend during the 10 years between 2006 to 2017.
- Since most of the asset classes have an upward trend between 2006 to 2017, the model we have now do much better at predicting the "buy" signals than "sell" signals.
- When making actual investment decisions, one should consider whether the fluctuation is significant enough to action. In other words, whether the potential profit is worth the risk.

Conclusion

Findings | Interpretation | Limitations

Interpretations
Of Best Models

SARIMAX has performed well in various industries, and to a certain extent can be used as a guide for medium-term transactions

DES performed well for medium term transactions with three months interval, but overall is better for long-term trend prediction

Findings

- Out of VAR, ARIMA, logistic regression, Double Exponential Smoothing and SARIMAX:

  Double Exponential Smoothing provides the best consistent AUC overall

- However, since we are running models for 60 sets(3 datasets for ETF, institutional and retail × 20 asset classes for each), when comparing between Double Exponential Smooth and SARIMAX, SARIMAX provides a larger AUC on certain asset classes

- In the future, we suggest to use SARIMAX when investing on different asset classes, and use DES for investing overall.

Limitations and Improvements

Limitations

The amount of information in the data and the size of the dataset is small, so that the transaction information of the market itself is difficult to support long-term forecasts.

Improvements

- **Introduce external information such as interest rates**
- **Separate analysis of systematic risk and idiosyncratic risk using risk exposure analysis models**
  - **Suggestions for creating label to measure idiosyncratic risk: Percent change of future rolling average for future excess return OR T-test result for future excess return**

Thank you for listening

Any Questions?