Data Driven Methods in Finance Final Presentation

Group Yuan
Chengyu(William) Chen, Kexin(Kathy) Qian, Weiting(Eden) Wang, Yuan(Douglas) Yang

Introduction

1. Weighting Methods → **2. Manual Adjustment** → **3. Ranking Methods**

Historical Stock Prices from Yahoo Finance

Week 1-2

Week 3-5

Fundamental Factors (Financial Ratios)

Rolling Window

Week 6-7

Week 8-11

Machine Learning Optimization

From Riskfolio to OLS

Kick-Start from RF → **Import More Factors** → **Experiment with a Parametric Model**

Quantitative Strategic Asset Allocation
Inputs: The percentage of the return is all what it requires.
Parameter: The model for optimizing weights allocation; the risk measure standard; risk-free and risk aversion rate, etc.

```
model='FM' # Could be FM(Factor Model, BL (Black Litterman), or Classic
rm = 'MAD' # Risk measure, standard deviation / mean absolute deviation
obj = 'Sharpe' # Objective function, could be MinRisk, MaxRet, Utility or Sharpe
hist = True # Use historical scenarios for risk measures that depend on scenarios
rf = 0 # Risk free rate
l = 0 # Risk aversion factor

w = port.optimization(model=model, rm=rm, obj=obj, rf=rf, l=l, hist=hist)
```

Output:
1. The optimal weights of investments.
2. The mean expected return vector
3. The covariance matrix indicating the fluctuation

Past prices, trade volume, what else can we use?

	Current	6/30/2022	3/31/2022
Market Cap (intraday)	237.30B	270.79B	286.43B
Enterprise Value	300.27B	336.69B	353.28B
Trailing P/E	18.98	21.97	25.13
Forward P/E	11.51	10.93	11.48
PEG Ratio (5 yr expected)	N/A	N/A	N/A
Price/Sales (ttm)	4.16	4.80	5.13
Price/Book (mrq)	16.19	16.63	18.59
Enterprise Value/Revenue	5.24	23.09	26.10

A quick benchmark: sklearn_linear model

- Training_Set: Set the time horizon as 8 weeks, putting the stock price in each past week in a single column (x) and regard the stock price in the most recent week as the target variable.
- Additional inputs: Besides price, the weekly updates from companies' balance can also be properly find their places
- Prediction: Once we add the most recent week's stats into the model, we will be able to generate the predictions for this week!

Factor-Driven Idea

Date	Week8	Week7	Week6	Week5	Week4	Week3	Week2	Week1	Past_Target	Most_recent
Stock										
ABBV	143.509995	138.039993	142.600006	141.850006	136.350006	136.279999	141.419998	144.059998	143.059998	134.210007
ACN	306.260010	309.350006	320.440002	315.290009	298.130005	284.070007	290.549988	272.679993	259.980011	257.299988
AEP	97.791840	98.684830	103.779999	104.940002	101.820000	101.089996	104.709999	100.360001	97.739998	86.449997
AIZ	175.058945	154.652985	165.597916	171.145065	161.550003	158.899994	163.699997	154.830002	147.899994	145.270004
ALLE	105.244537	101.779533	105.951485	105.991302	97.239189	94.481117	99.200691	89.690002	92.040001	89.680000
...	...	...	...	...	...	...	...	...	...	...
SOL-USD	41.926998	40.561031	45.334759	35.890259	31.726362	31.227930	34.739765	32.208858	33.676636	33.212444
DOGE-USD	0.068909	0.069765	0.072345	0.067939	0.063572	0.061635	0.064018	0.060423	0.063336	0.061653
DOT-USD	8.185587	8.813465	9.479537	7.325466	6.923963	7.302052	7.770865	6.896843	6.438069	6.315190
SHIB-USD	120.000000	120.000000	130.000000	130.000000	120.000000	120.000000	130.000000	120.000000	110.000000	110.000000
AVAX-USD	24.522364	24.949528	29.343828	22.487879	20.598030	18.860632	20.400782	18.056599	17.959776	17.196198

Not surprisingly, this doesn't give out good predive results…

```
regression.score(predict_x,compare.current)
```

```
: -0.021663729206367055
```

What's Next?

- We need to refine our factors, rather than blindly putting the available inputs there, we probably should find some thesis with a mature methodology.
- Back test: Here we simply split them into training and test datasets and then simulate the scenario once, making the final prediction directly. Instead, it should be tested multiple times and we should let the model learn from the errors it made in the previous period

Factor-Driven Model

It all starts from following attributes:

- Open price
- Close price
- Volume
- High price
- Low price
- Adjusted Close Price

And those serve as the bricks we need to set up factors

Under the guidance from previous research paper, we constructed the factors as following:

```
'-1*ts_Decay((ts_Decay(Close,10)-ts_Decay(VWAP,10))/VWAP*(High-Low),40)',
'-1*ts_Decay(IdioRet*(Volume/ts_Delay(Volume,1)),40)',
 '-1*ts_Decay(IdioRet*(High-Low)/Close*(Volume/ts_Delay(Volume,1)),60)',
 '-1*ts_Decay(ts_Delta(Close*ts_Decay(TotalRet,5),5)*(High-Low)/Close*(Volume/ts_Delay(Volume,1)),60)',
 '-1*ts_Decay(VWAP/Close*(High-Low)/Close*(Volume/ts_Delay(Volume,1)),10)',
 '-1*ts_Decay(VWAP/Close*(High-Low)/Close*(Volume/ts_Delay(Volume,1)),40)',
 '-1*ts_DecayExp(ts_Mean(IdioRet*(High-Low)/Close*(Volume/ts_Delay(Volume,1)),20),60)',
 '-1*ts_DecayExp(ts_Mean(VWAP/Close*(High-Low)/Close,20),20)',
 '-1*ts_DecayExp(VWAP/Close*(High-Low)/Close,60)',
'-1*ts_Mean((ts_Decay(Close,10)-ts_Decay(VWAP,10))/VWAP*(High-Low)/Close,20)'
```

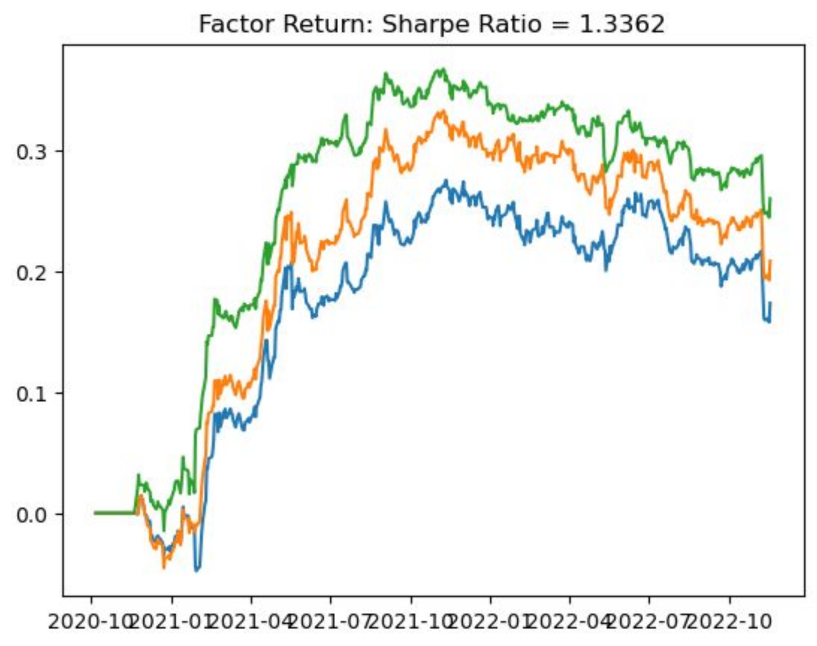
Each factor represents a distinguished dimension of stocks

Assigning weights to different factors:

1. Fixed Weights
2. Rolling Adjusted Weights based on back test

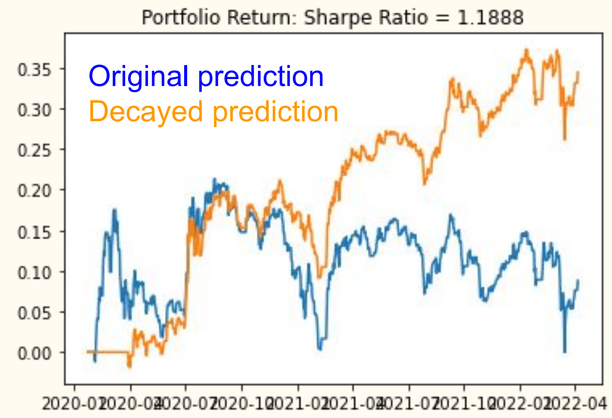
Rolling by Sliding Windows

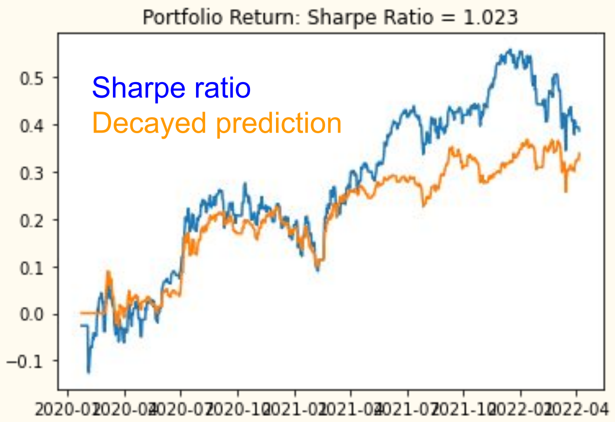
- Redefine the composite factors
  - Use a window for 30 days
  - Get the moving average returns by the sliding window
  - Get the **moving average sr/ret/markowitz** value by the sliding window
  - For **each day(>=30) we get a weight list** and composite a factor
  - Output the prediction list determined by the sliding window
- Use the prediction list for back testing
  - **Each day the weight changes** based on past 30 days' data
  - Better for continuous prediction and back testing, less efficient for single prediction



ML LightGBM Selection

- Train_set: composite factor value (days * stocks)
    - 2019_01_01 - 2021_10_01
- Validation_set
    - 2021_10_01 - 2022_06_30
- Test
    - 2022_06_30 - Now
- Params
    - Boosting type: gbdt
    - objective : regression
    - Metric: l2
    - Round: 100
    - Early stopping rounds: 10
- Use predicted value for backtesting
    - Try decay
    - Compare with sharpe ratio methods





Manual Adjustment

- Adjustment on S&P 500 stocks:
    - Market analyst reports
    - Events, e.g. earnings announcements
    - News
- Adjustment on ETFs:
    - Yahoo Finance Top ETFs
- Adjustment on Cryptos:
    - News
    - Constraint: < 5% weight in overall portfolio

name	Analyst Report	Weights			Events
AbbVie	Buy	0.0082822124	Total Weight	1.00000000000000	
Accenture	Buy	0.0096217002	Stocks Weight	0.468274253	
American Electric Powe	Strong Buy	0.0118196610	ETF Weight	0.531725747	
Assurant	Hold	0.0000000000	Crypto Weight	0	
Allegion	Buy	0.0104432419			
Applied Materials	Strong Buy	0.0412643161			17.48

Events Calendar for: Nov 13, 2022 - Nov 19, 2022

Find earnings for symbols

Prev | 13 Nov Sun 83 Earnings | 14 Nov Mon 705 Earnings | 15 Nov Tues 256 Earnings | 16 Nov Wed 153 Earnings | 17 Nov Thu 112 Earnings | 18 Nov Fri 50 Earnings | 19 Nov Sat | Next

Earnings on Thu, Nov 17 1-100 of 112 results Add to Portfolio

Symbol	Company	Event Name	Earnings Call Time	EPS Estimate	Reported EPS	Surprise(%)
BABA	Alibaba Group Holding Limited	-	TAS	11.61	12.92	+11.25
VJET	voxeljet AG	-	TAS	-0.44	-1.23	-176.4
NEXCF	NexTech AR Solutions Corp.	-	TAS	-0.06	-0.05	+16.67
PANW	Palo Alto Networks, Inc.	-	TAS	0.69	0.83	+20.64
KLIC	Kulicke and Soffa Industries Inc	Q4 2022 Earnings Call	TAS	0.98	1.19	+21.43

FINANCE

Credit Suisse sells most of its securitized products business to Apollo as it speeds up restructure

The Downfall of FTX: Could It Signal a Long Crypto Winter?

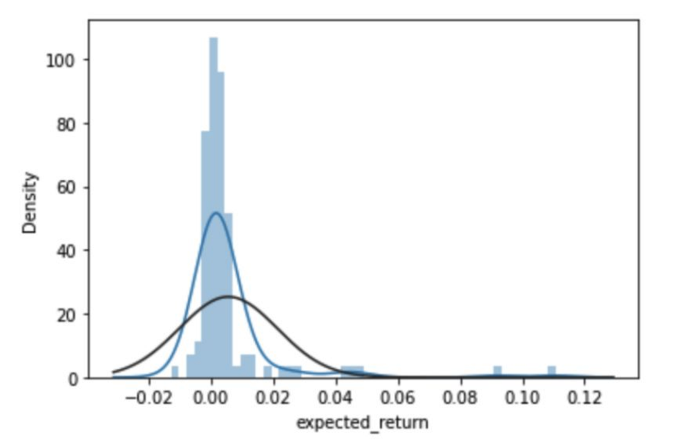
By Wall Street Journal

Rank Distribution - Fitting Stock Returns

- Check whether the data is skewed ➡ p-value of the Shapiro-Wilk test < 0.05
- ➡ data is indeed right-skewed
- The poor fitting result into normal distribution also supports the same conclusion.

```
from scipy.stats import shapiro
shapiro(expectations['expected_return'])[1]
```

```
3.750297517382212e-18
```

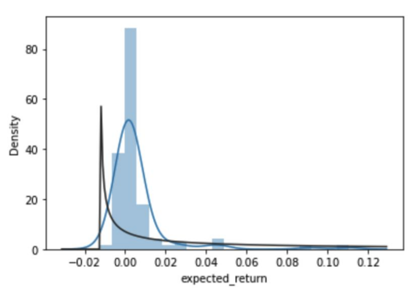


Rank Distribution - Fitting Tryouts

- To remove the skewness, we tried four ways:

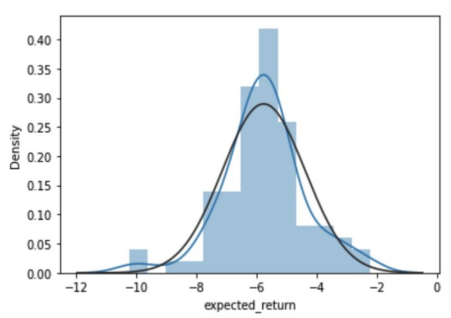
Gamma Distribution

We fitted the data into a gamma distribution, but the result was not ideal.



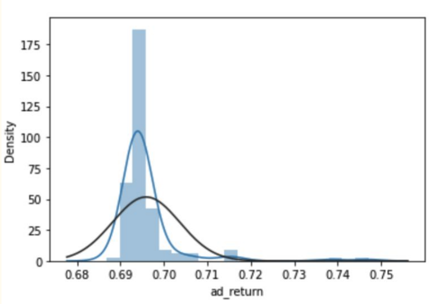
Log Normal Distribution

This approach performed better. However, simply taking the log of all expected returns will automatically drop all negative values, as the log cannot be applied to negative numbers.



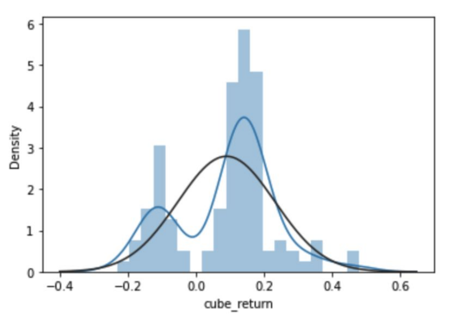
Adjusted Log Normal Distribution

We added 1 for each expected return and then performed the log-normal fitting model. This approach allows us to include the negative expected returns in the distribution.



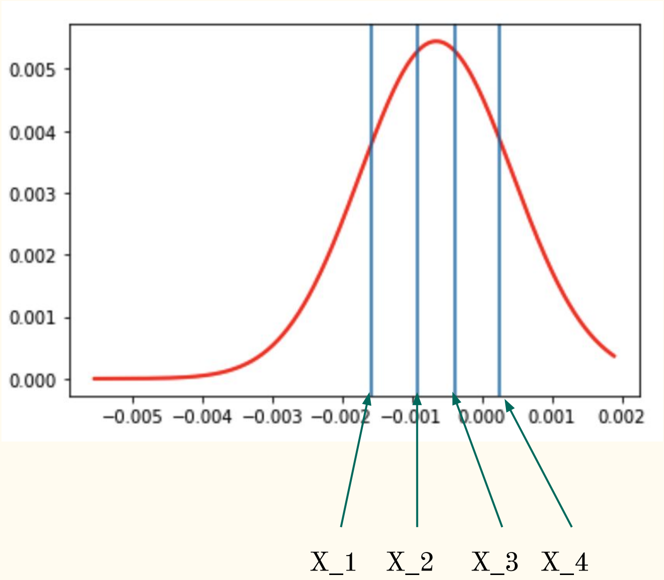
Cube Root Transformation

We first took each expected return's cube roots which can be applied to both negative and positive values. Then we fitted the cube-rooted results into a normal distribution. However, the result showed a bi-model shape which is not compliant with our data.



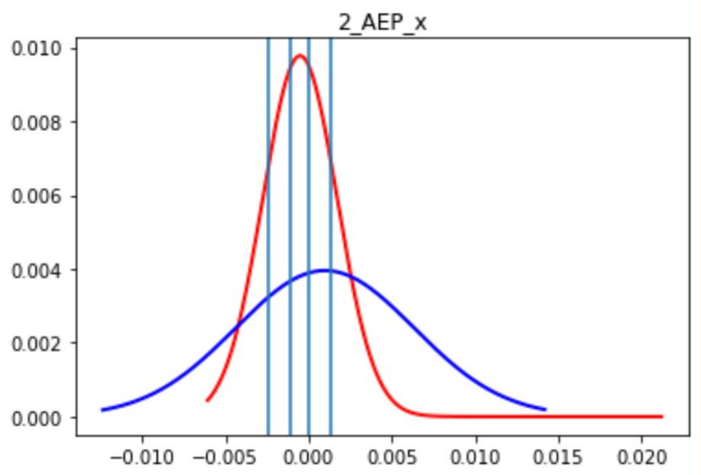
Rank Distribution - Finding Ranking Benchmarks

- After choosing the distribution of 110 stocks' expected returns, we decide on **benchmarks** to calculate ranking probabilities.
- Split the fitted distribution into 5 equal parts:
  - Derive 4 numbers: X_1, X_2, X_3, X_4
  - Segment the distribution into 5 parts
  - The area under PDF (Probability Density Function) in each part = 20%



Rank Distribution - Simulating Distribution

- Knowing each stock's expected return and variance ➡ simulate the normal distribution and generate **CDF** (Cumulative Density Function) for each stock
- Use the CDF to compute the probability for each benchmark:
  - P(X<X_1), P(X<X_2), P(X<X_3), P(X<X_4)
- Probability for this stock's return falling in rank 1 = **P(X<X_1)**
- Probability for this stock's return falling in rank 2 = **P(X<X_2)—P(X<X_1)** and similarly for rank 3, 4 and 5



Conclusion & Future Works

Conclusion

- Weighting Methods:
  - Riskfolio, Factor-Driven Model, ML Optimization, Manual Tuning
- Ranking Distribution:
  - Benchmarking, Simulating Distribution

Future Works

- How to include financial data from I/S, B/S, and SCF in models?
- More information to adjust ETFs and Cryptos?

Thank You!