AI in Finance Assignment 4

Yijia Zeng yzeng323@gatech.edu GTID: 903629003

Data Loading

Data Cleaning

- Handling date variable
- selectikng a sample of 10 firms per industry

```
This is the shape of my cleaned dataframe (28972, 22)

This is the shape of my training dataframe (8222, 22)
This is the shape of my validating dataframe (5296, 22)
This is the shape of my testing dataframe (6398, 22)
```

Neural Network Model Construction

- input features will include vwretd, ret, vol, spread, and industry
    - vwretd: This is the return of the value-weighted market index, which typically used as the market return when calculating beta. Beta is essentially the stock's sensitivity to the market return, so this variable is critical.
    - ret: The return of the stock is a crucial input when calculating beta, as beta represents the relationship between a stock's returns and the overall market returns.
    - vol: Volume is a key indicator of liquidity. Stocks with higher liquidity tend to have more stable prices and lower betas, while less liquid stocks can be more volatile and have higher betas.

- spread (ask - bid): A tight bid-ask spread indicates high liquidity and lower transaction costs, which generally corresponds to a lower beta. Wider spreads may suggest illiquidity and higher risk, leading to higher beta
- industry: different industries might have different general pattern for beta

NeuralBeta Class Construction & custom_loss function

Create Features

- create lookback features for variables except for industry, representing by adding lag_k columns

Standardize & Record Data

- standardize data used in neural network
- keep ret and vwretd to calculate loss

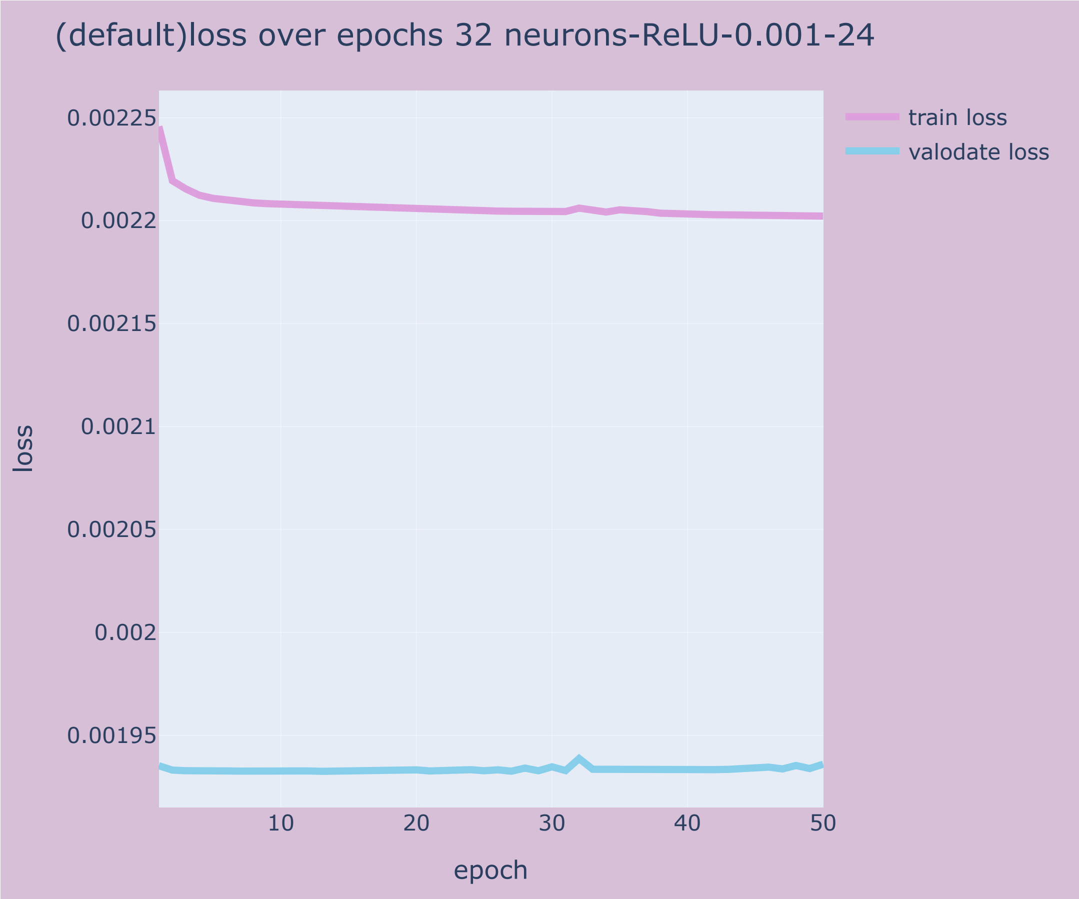
Model Training

Neural Network Model Implementation

- choose lookback window as 24
- use linear activation function to maintain the value for predicted beta
- batch size as 64, which offer a lot of information for distribution in each sampling, and also lead to better generalization performance
- optimizer as Adam
- learning rate as 0.001
    - candidates: 0.001, 0.0001, 0.00001

- epoch as 50
- number of neurons in hidden layer as 64
  - candidates: 32, 64, 128
- activation function as linear
  - candidates: linear, sigmoid, tanh, relu

```
Epoch 5/50, Train Loss: 0.00221084, Val Loss: 0.00193252
Epoch 10/50, Train Loss: 0.00220802, Val Loss: 0.00193269
Epoch 15/50, Train Loss: 0.00220642, Val Loss: 0.00193255
Epoch 20/50, Train Loss: 0.00220571, Val Loss: 0.00193318
Epoch 25/50, Train Loss: 0.00220487, Val Loss: 0.00193277
Epoch 30/50, Train Loss: 0.00220459, Val Loss: 0.00193467
Epoch 35/50, Train Loss: 0.00220527, Val Loss: 0.00193342
Epoch 40/50, Train Loss: 0.00220307, Val Loss: 0.00193303
Epoch 45/50, Train Loss: 0.00220246, Val Loss: 0.00193345
Epoch 50/50, Train Loss: 0.00220221, Val Loss: 0.00193592
```

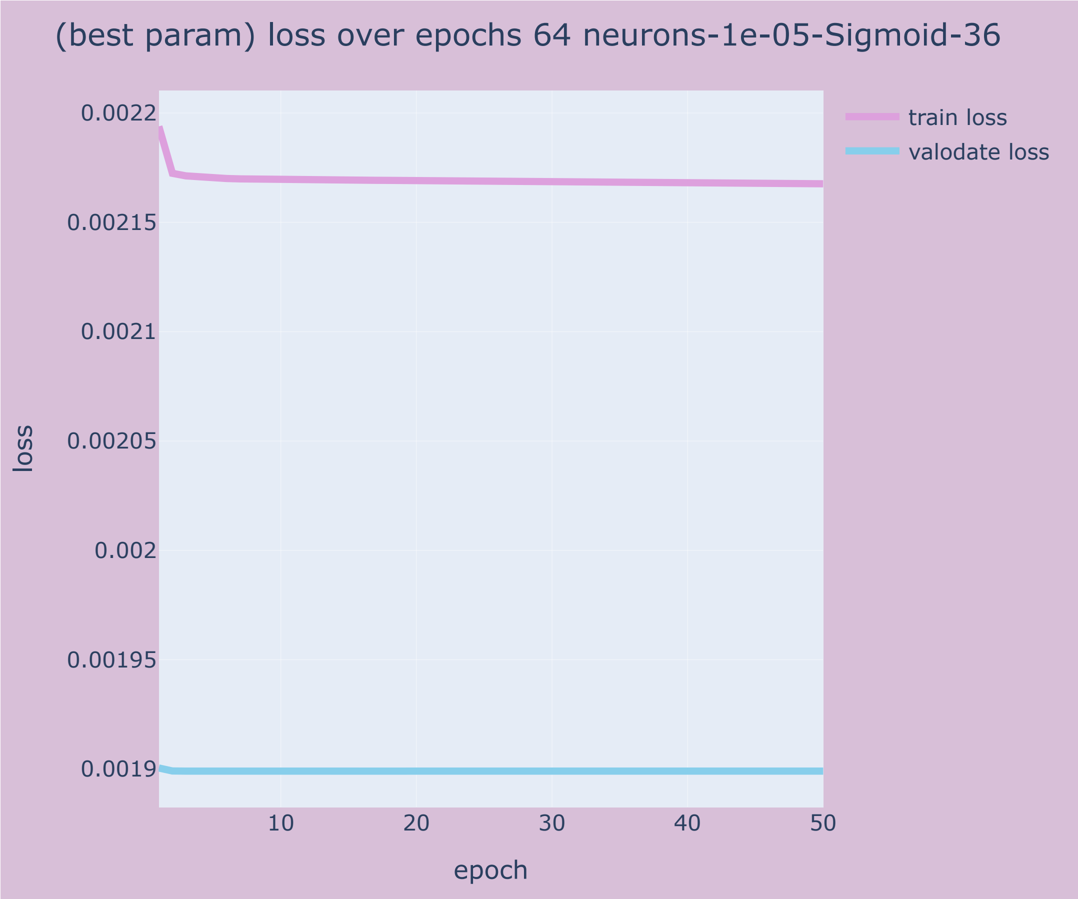


Hyper Parameter Tuning

```
current params: hidden_size = 32, activation = relu, lr = 0.001, lookback = 12
current params: hidden_size = 32, activation = relu, lr = 0.0001, lookback = 12
current params: hidden_size = 32, activation = relu, lr = 1e-05, lookback = 12
current params: hidden_size = 32, activation = tanh, lr = 0.001, lookback = 12
current params: hidden_size = 32, activation = tanh, lr = 0.0001, lookback = 12
current params: hidden_size = 32, activation = tanh, lr = 1e-05, lookback = 12
current params: hidden_size = 32, activation = sigmoid, lr = 0.001, lookback = 12
current params: hidden_size = 32, activation = sigmoid, lr = 0.0001, lookback = 12
current params: hidden_size = 32, activation = sigmoid, lr = 1e-05, lookback = 12
current params: hidden_size = 32, activation = linear, lr = 0.001, lookback = 12
current params: hidden_size = 32, activation = linear, lr = 0.0001, lookback = 12
current params: hidden_size = 32, activation = linear, lr = 1e-05, lookback = 12
current params: hidden_size = 32, activation = relu, lr = 0.001, lookback = 24
current params: hidden_size = 32, activation = relu, lr = 0.0001, lookback = 24
current params: hidden_size = 32, activation = relu, lr = 1e-05, lookback = 24
current params: hidden_size = 32, activation = tanh, lr = 0.001, lookback = 24
current params: hidden_size = 32, activation = tanh, lr = 0.0001, lookback = 24
current params: hidden_size = 32, activation = tanh, lr = 1e-05, lookback = 24
current params: hidden_size = 32, activation = sigmoid, lr = 0.001, lookback = 24
current params: hidden_size = 32, activation = sigmoid, lr = 0.0001, lookback = 24
current params: hidden_size = 32, activation = sigmoid, lr = 1e-05, lookback = 24
current params: hidden_size = 32, activation = linear, lr = 0.001, lookback = 24
current params: hidden_size = 32, activation = linear, lr = 0.0001, lookback = 24
current params: hidden_size = 32, activation = linear, lr = 1e-05, lookback = 24
current params: hidden_size = 32, activation = relu, lr = 0.001, lookback = 36
current params: hidden_size = 32, activation = relu, lr = 0.0001, lookback = 36
current params: hidden_size = 32, activation = relu, lr = 1e-05, lookback = 36
current params: hidden_size = 32, activation = tanh, lr = 0.001, lookback = 36
current params: hidden_size = 32, activation = tanh, lr = 0.0001, lookback = 36
current params: hidden_size = 32, activation = tanh, lr = 1e-05, lookback = 36
current params: hidden_size = 32, activation = sigmoid, lr = 0.001, lookback = 36
current params: hidden_size = 32, activation = sigmoid, lr = 0.0001, lookback = 36
current params: hidden_size = 32, activation = sigmoid, lr = 1e-05, lookback = 36
current params: hidden_size = 32, activation = linear, lr = 0.001, lookback = 36
current params: hidden_size = 32, activation = linear, lr = 0.0001, lookback = 36
current params: hidden_size = 32, activation = linear, lr = 1e-05, lookback = 36
current params: hidden_size = 64, activation = relu, lr = 0.001, lookback = 12
current params: hidden_size = 64, activation = relu, lr = 0.0001, lookback = 12
current params: hidden_size = 64, activation = relu, lr = 1e-05, lookback = 12
current params: hidden_size = 64, activation = tanh, lr = 0.001, lookback = 12
current params: hidden_size = 64, activation = tanh, lr = 0.0001, lookback = 12
current params: hidden_size = 64, activation = tanh, lr = 1e-05, lookback = 12
current params: hidden_size = 64, activation = sigmoid, lr = 0.001, lookback = 12
current params: hidden_size = 64, activation = sigmoid, lr = 0.0001, lookback = 12
current params: hidden_size = 64, activation = sigmoid, lr = 1e-05, lookback = 12
```

```
current params: hidden_size = 64, activation = linear, lr = 0.001, lookback = 12
current params: hidden_size = 64, activation = linear, lr = 0.0001, lookback = 12
current params: hidden_size = 64, activation = linear, lr = 1e-05, lookback = 12
current params: hidden_size = 64, activation = relu, lr = 0.001, lookback = 24
current params: hidden_size = 64, activation = relu, lr = 0.0001, lookback = 24
current params: hidden_size = 64, activation = relu, lr = 1e-05, lookback = 24
current params: hidden_size = 64, activation = tanh, lr = 0.001, lookback = 24
current params: hidden_size = 64, activation = tanh, lr = 0.0001, lookback = 24
current params: hidden_size = 64, activation = tanh, lr = 1e-05, lookback = 24
current params: hidden_size = 64, activation = sigmoid, lr = 0.001, lookback = 24
current params: hidden_size = 64, activation = sigmoid, lr = 0.0001, lookback = 24
current params: hidden_size = 64, activation = sigmoid, lr = 1e-05, lookback = 24
current params: hidden_size = 64, activation = linear, lr = 0.001, lookback = 24
current params: hidden_size = 64, activation = linear, lr = 0.0001, lookback = 24
current params: hidden_size = 64, activation = linear, lr = 1e-05, lookback = 24
current params: hidden_size = 64, activation = relu, lr = 0.001, lookback = 36
current params: hidden_size = 64, activation = relu, lr = 0.0001, lookback = 36
current params: hidden_size = 64, activation = relu, lr = 1e-05, lookback = 36
current params: hidden_size = 64, activation = tanh, lr = 0.001, lookback = 36
current params: hidden_size = 64, activation = tanh, lr = 0.0001, lookback = 36
current params: hidden_size = 64, activation = tanh, lr = 1e-05, lookback = 36
current params: hidden_size = 64, activation = sigmoid, lr = 0.001, lookback = 36
current params: hidden_size = 64, activation = sigmoid, lr = 0.0001, lookback = 36
current params: hidden_size = 64, activation = sigmoid, lr = 1e-05, lookback = 36
current params: hidden_size = 64, activation = linear, lr = 0.001, lookback = 36
current params: hidden_size = 64, activation = linear, lr = 0.0001, lookback = 36
current params: hidden_size = 64, activation = linear, lr = 1e-05, lookback = 36
current params: hidden_size = 128, activation = relu, lr = 0.001, lookback = 12
current params: hidden_size = 128, activation = relu, lr = 0.0001, lookback = 12
current params: hidden_size = 128, activation = relu, lr = 1e-05, lookback = 12
current params: hidden_size = 128, activation = tanh, lr = 0.001, lookback = 12
current params: hidden_size = 128, activation = tanh, lr = 0.0001, lookback = 12
current params: hidden_size = 128, activation = tanh, lr = 1e-05, lookback = 12
current params: hidden_size = 128, activation = sigmoid, lr = 0.001, lookback = 12
current params: hidden_size = 128, activation = sigmoid, lr = 0.0001, lookback = 12
current params: hidden_size = 128, activation = sigmoid, lr = 1e-05, lookback = 12
current params: hidden_size = 128, activation = linear, lr = 0.001, lookback = 12
current params: hidden_size = 128, activation = linear, lr = 0.0001, lookback = 12
current params: hidden_size = 128, activation = linear, lr = 1e-05, lookback = 12
current params: hidden_size = 128, activation = relu, lr = 0.001, lookback = 24
current params: hidden_size = 128, activation = relu, lr = 0.0001, lookback = 24
current params: hidden_size = 128, activation = relu, lr = 1e-05, lookback = 24
current params: hidden_size = 128, activation = tanh, lr = 0.001, lookback = 24
current params: hidden_size = 128, activation = tanh, lr = 0.0001, lookback = 24
current params: hidden_size = 128, activation = tanh, lr = 1e-05, lookback = 24
```

```
current params: hidden_size = 128, activation = sigmoid, lr = 0.001, lookback = 24
current params: hidden_size = 128, activation = sigmoid, lr = 0.0001, lookback = 24
current params: hidden_size = 128, activation = sigmoid, lr = 1e-05, lookback = 24
current params: hidden_size = 128, activation = linear, lr = 0.001, lookback = 24
current params: hidden_size = 128, activation = linear, lr = 0.0001, lookback = 24
current params: hidden_size = 128, activation = linear, lr = 1e-05, lookback = 24
current params: hidden_size = 128, activation = relu, lr = 0.001, lookback = 36
current params: hidden_size = 128, activation = relu, lr = 0.0001, lookback = 36
current params: hidden_size = 128, activation = relu, lr = 1e-05, lookback = 36
current params: hidden_size = 128, activation = tanh, lr = 0.001, lookback = 36
current params: hidden_size = 128, activation = tanh, lr = 0.0001, lookback = 36
current params: hidden_size = 128, activation = tanh, lr = 1e-05, lookback = 36
current params: hidden_size = 128, activation = sigmoid, lr = 0.001, lookback = 36
current params: hidden_size = 128, activation = sigmoid, lr = 0.0001, lookback = 36
current params: hidden_size = 128, activation = sigmoid, lr = 1e-05, lookback = 36
current params: hidden_size = 128, activation = linear, lr = 0.001, lookback = 36
current params: hidden_size = 128, activation = linear, lr = 0.0001, lookback = 36
current params: hidden_size = 128, activation = linear, lr = 1e-05, lookback = 36

 Best parameters: {'hidden_size': 32, 'lookback': 36, 'activation': 'sigmoid', 'lr': 0.001}


Epoch 5/50, Train Loss: 0.00217030, Val Loss: 0.00189897
Epoch 10/50, Train Loss: 0.00216955, Val Loss: 0.00189898
Epoch 15/50, Train Loss: 0.00216930, Val Loss: 0.00189902
Epoch 20/50, Train Loss: 0.00216912, Val Loss: 0.00189904
Epoch 25/50, Train Loss: 0.00216895, Val Loss: 0.00189905
Epoch 30/50, Train Loss: 0.00216876, Val Loss: 0.00189905
Epoch 35/50, Train Loss: 0.00216854, Val Loss: 0.00189905
Epoch 40/50, Train Loss: 0.00216828, Val Loss: 0.00189904
Epoch 45/50, Train Loss: 0.00216798, Val Loss: 0.00189904
Epoch 50/50, Train Loss: 0.00216763, Val Loss: 0.00189906
```



Discussion

- After using the best parameter for the neural network, the loss for both training and validation datasets decrease significantly

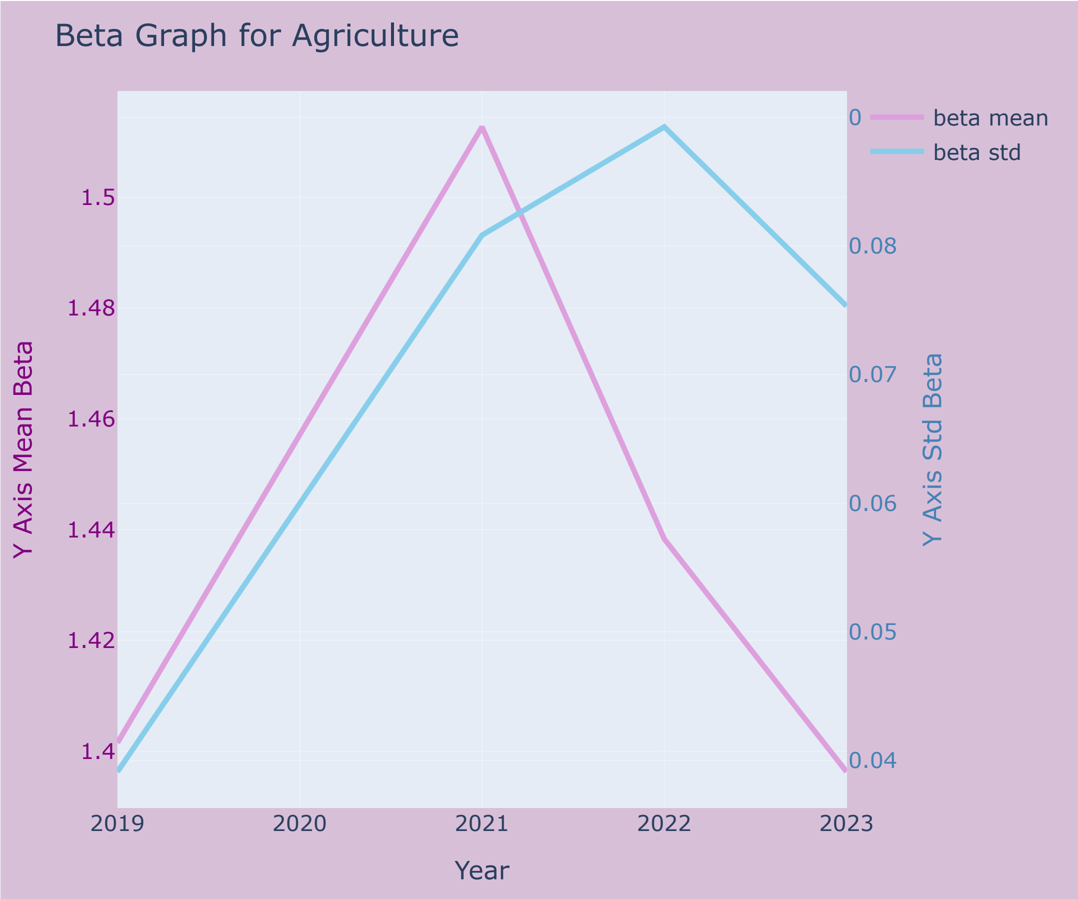
Beta in Test Dataset

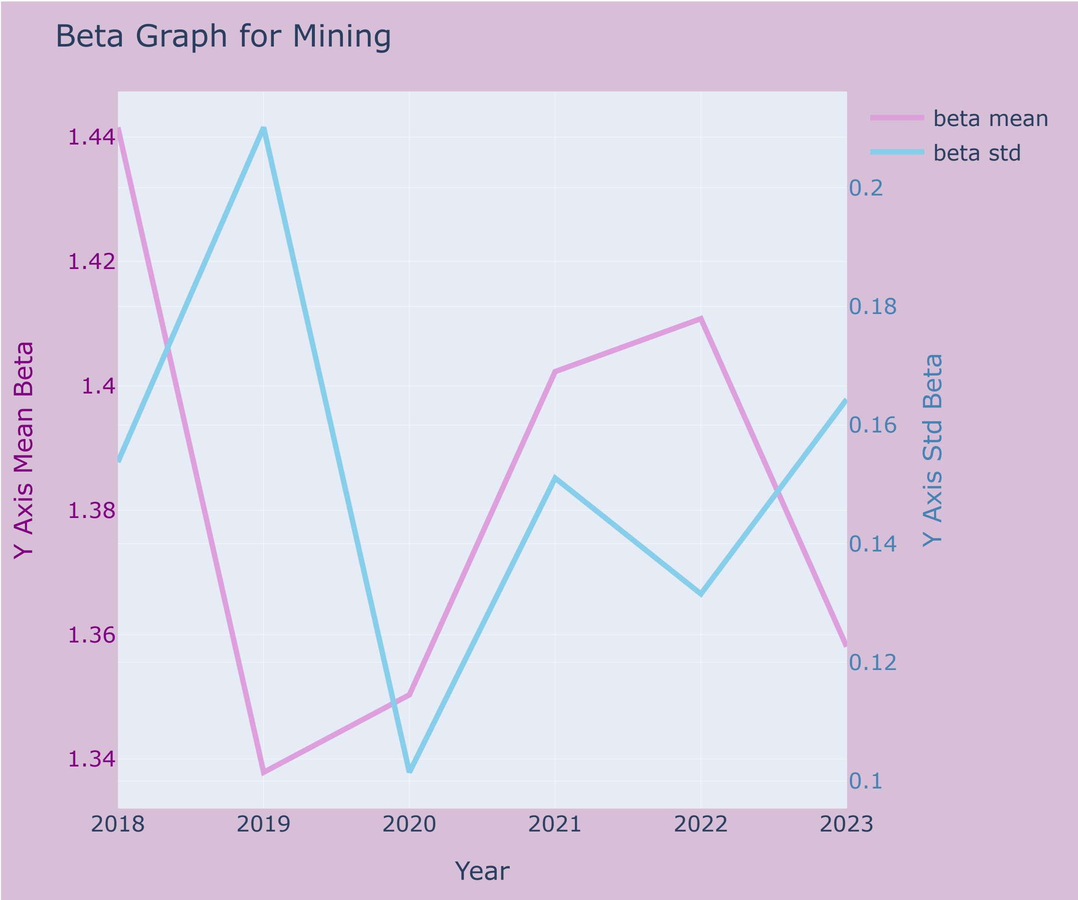
```
--------------------------------------------------

This is the Statistics of predicted beta for industry Agriculture
count       85.000000
mean         1.435956
std          0.086104
min          1.226304
max          1.614698
skew        -0.202027
kurtosis    -0.218002
1%           1.245660
5%           1.276113
25%          1.385251
50%          1.436983
75%          1.499366
95%          1.565781
99%          1.614060
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta for industry Mining
count       695.000000
mean          1.384359
std           0.160617
min           0.819927
max           2.085316
skew         -0.126340
kurtosis      2.706187
1%            0.879811
5%            1.106420
25%           1.304823
50%           1.396536
75%           1.473402
95%           1.600856
99%           1.786088
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta for industry Construction
count       649.000000
mean          1.339346
std           0.078162
min           1.057832
max           1.564469
skew         -0.255078
```

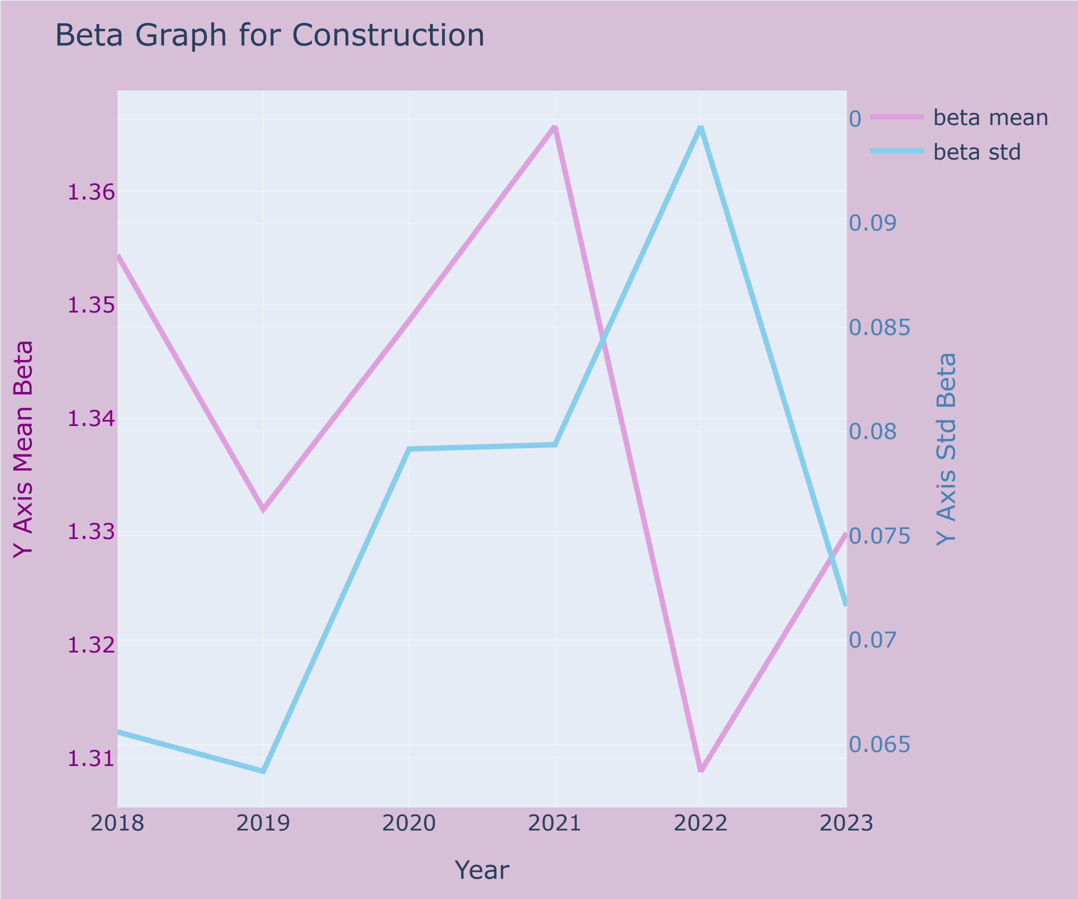
```
kurtosis      0.413654
1%            1.127184
5%            1.213595
25%           1.290241
50%           1.340973
75%           1.393395
95%           1.458426
99%           1.512999
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta for industry Manufacturing
count       961.000000
mean          1.284349
std           0.091588
min           1.012201
max           1.648317
skew          0.054476
kurtosis      0.351230
1%            1.071873
5%            1.130394
25%           1.222457
50%           1.285441
75%           1.344218
95%           1.433840
99%           1.516615
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta for industry Transportation
count       751.000000
mean          1.220181
std           0.087261
min           0.984575
max           1.495746
skew         -0.051122
kurtosis     -0.170963
1%            1.013013
5%            1.074966
25%           1.164294
50%           1.217533
75%           1.282349
95%           1.361231
99%           1.412134
Name: beta_pred, dtype: float64
```

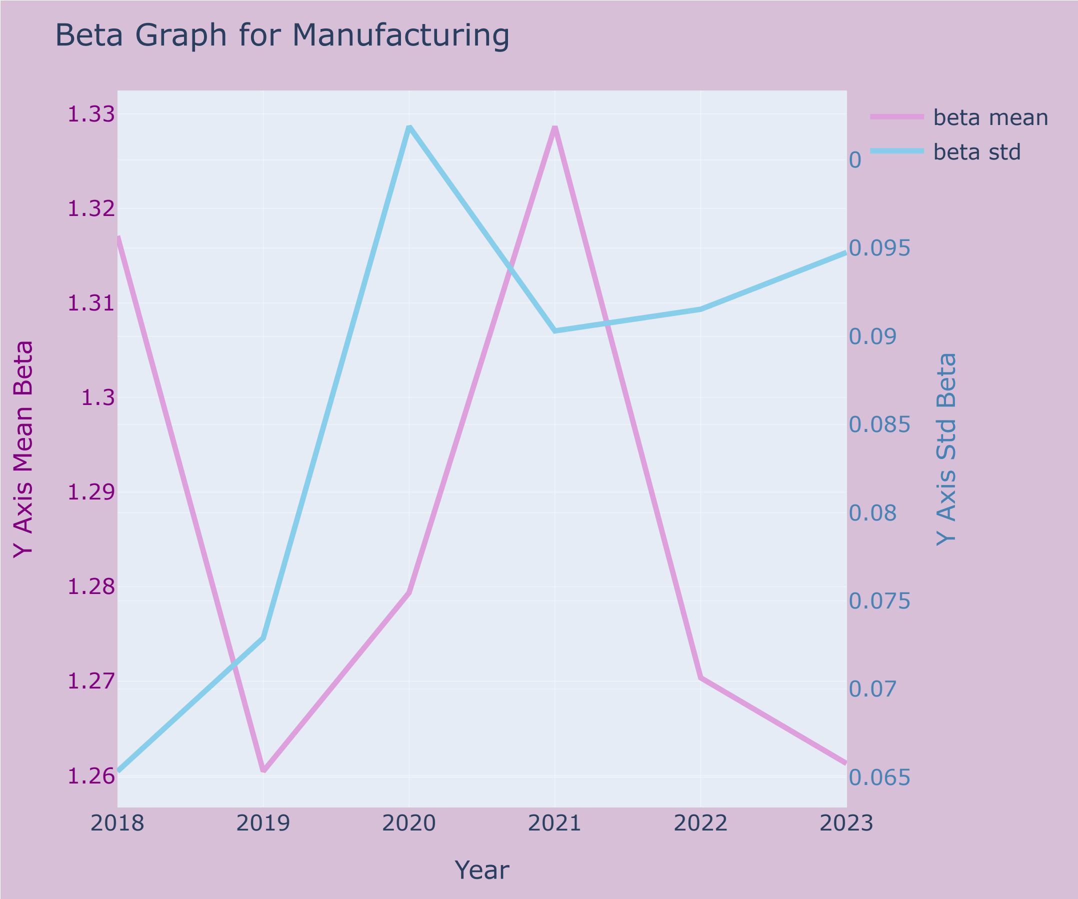
```
-------------------------------------------------

This is the Statistics of predicted beta for industry Wholesale
count       633.000000
mean          1.156721
std           0.085339
min           0.844869
max           1.365776
skew         -0.191017
kurtosis     -0.266311
1%            0.959099
5%            1.014612
25%           1.098715
50%           1.157650
75%           1.218058
95%           1.290654
99%           1.327476
Name: beta_pred, dtype: float64
-------------------------------------------------

This is the Statistics of predicted beta for industry Retail
count       649.000000
mean          1.119266
std           0.087249
min           0.868881
max           1.356848
skew         -0.181635
kurtosis     -0.238044
1%            0.914926
5%            0.970210
25%           1.061692
50%           1.123232
75%           1.179931
95%           1.253765
99%           1.299979
Name: beta_pred, dtype: float64
-------------------------------------------------

This is the Statistics of predicted beta for industry Finance
count       857.000000
mean          1.052207
std           0.117448
min           0.793067
max           1.734120
skew          1.911434
```

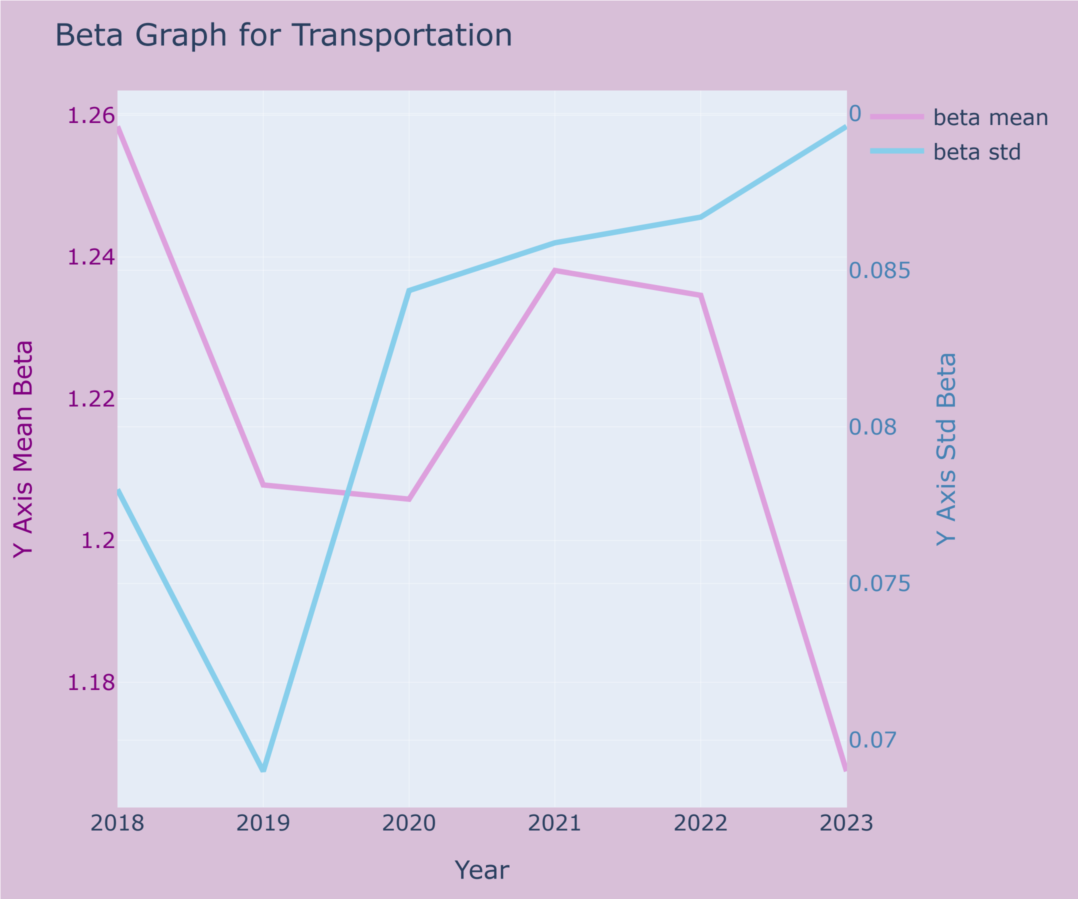
```
kurtosis      7.768735
1%            0.841435
5%            0.898133
25%           0.984260
50%           1.044477
75%           1.095576
95%           1.223929
99%           1.586350
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta for industry Service
count       698.000000
mean          1.006591
std           0.121798
min           0.743417
max           1.846172
skew          1.915878
kurtosis      8.025247
1%            0.808209
5%            0.854782
25%           0.928580
50%           0.996251
75%           1.060029
95%           1.192028
99%           1.451041
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta for industry Public
count       73.000000
mean         0.925875
std          0.082036
min          0.733613
max          1.106821
skew        -0.134973
kurtosis    -0.172284
1%           0.740399
5%           0.779228
25%          0.880695
50%          0.924155
75%          0.980826
95%          1.064403
99%          1.082728
Name: beta_pred, dtype: float64
```

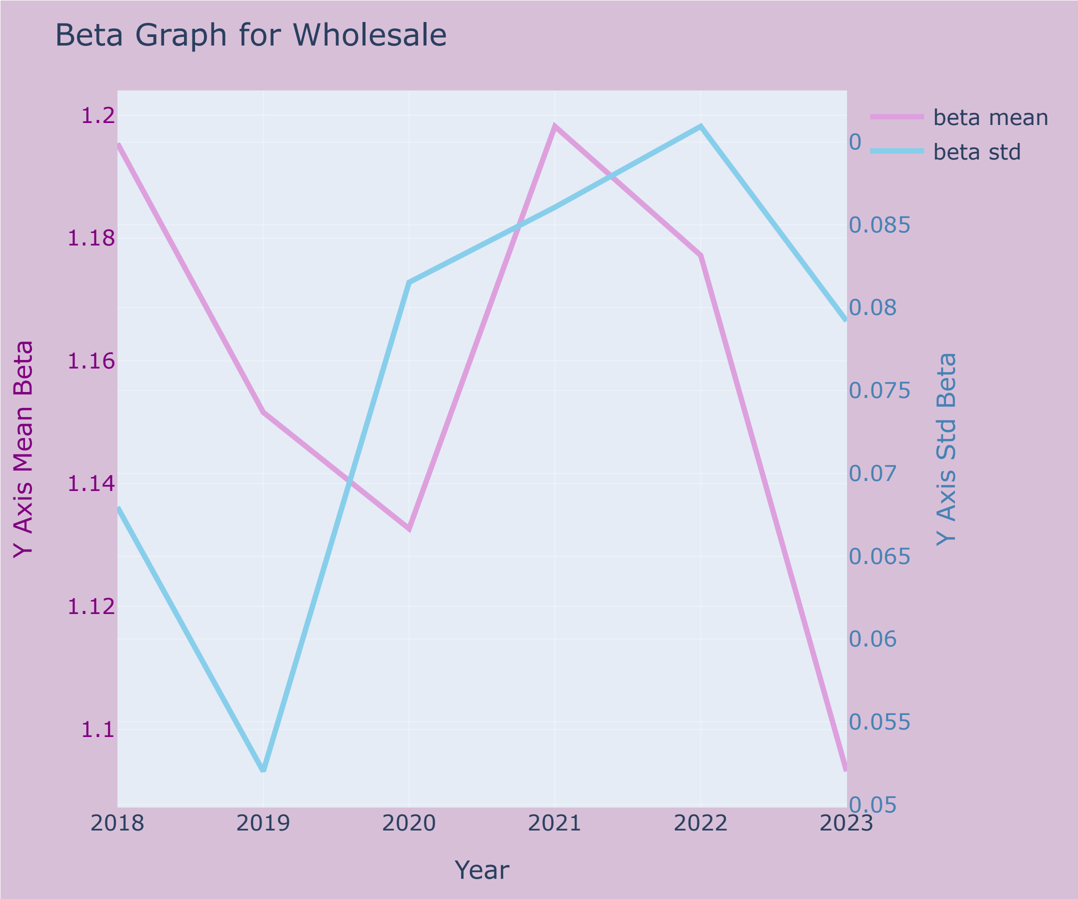

Beta Graph for Agriculture
beta mean
beta std
Y Axis Mean Beta
1.5
1.48
1.46
1.44
1.42
1.4
Y Axis Std Beta
0
0.08
0.07
0.06
0.05
0.04
2019
2020
2021
2022
2023
Year

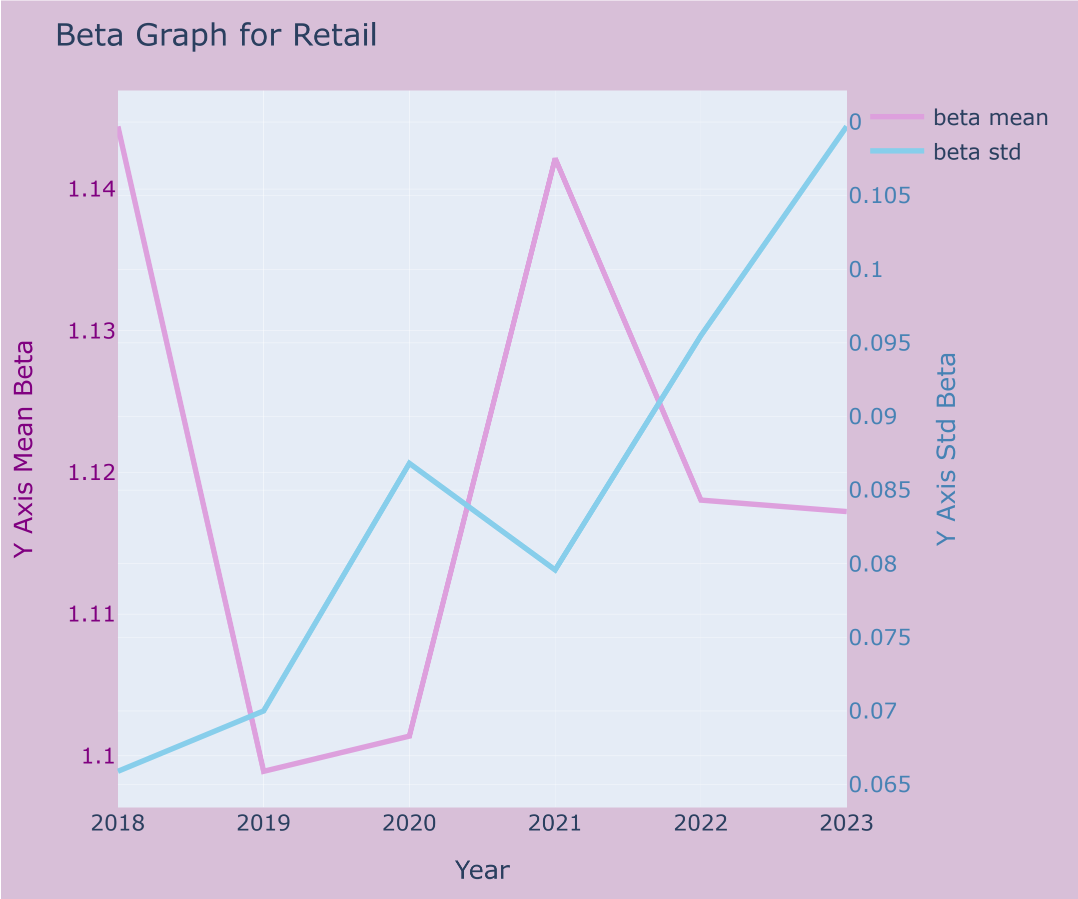

Beta Graph for Mining
beta mean
beta std
Y Axis Mean Beta
1.44
1.42
1.4
1.38
1.36
1.34
Y Axis Std Beta
0.2
0.18
0.16
0.14
0.12
0.1
2018
2019
2020
2021
2022
2023
Year

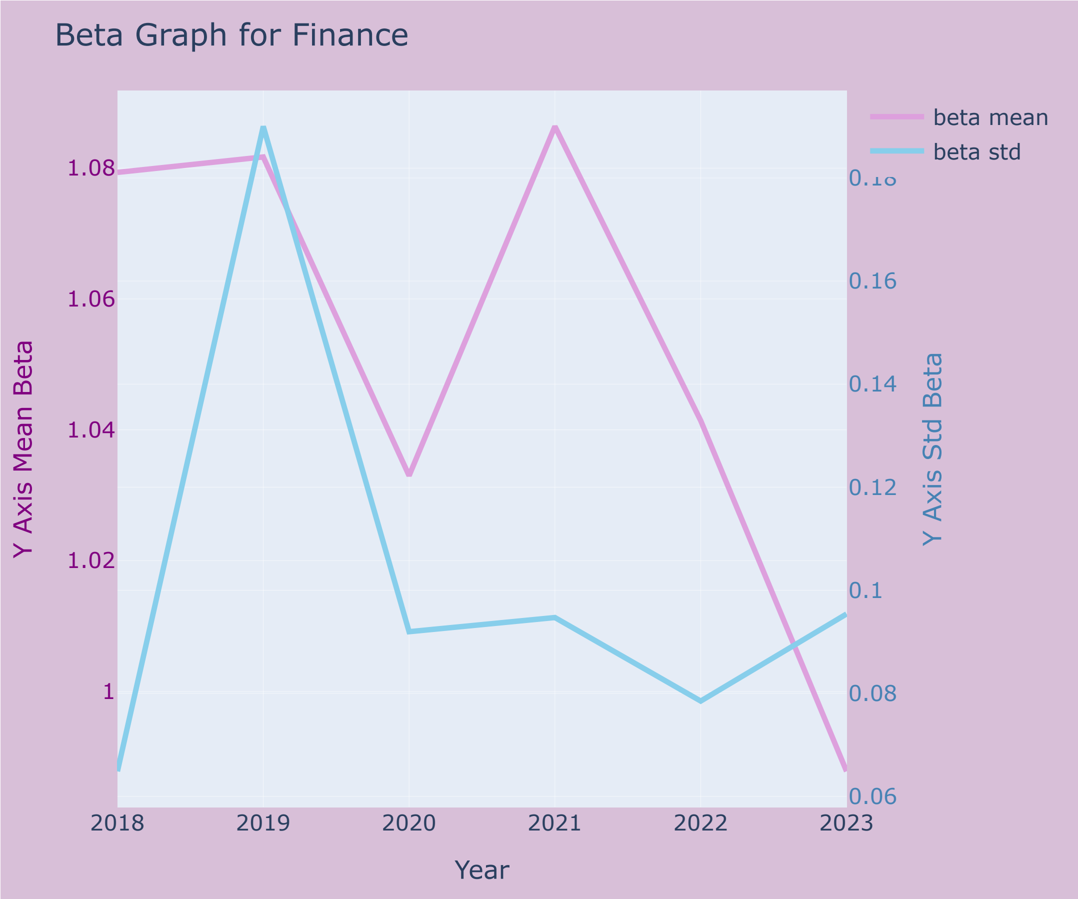

Beta Graph for Construction
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
1.36
1.35
1.34
1.33
1.32
1.31
0.09
0.085
0.08
0.075
0.07
0.065
2018
2019
2020
2021
2022
2023
Year

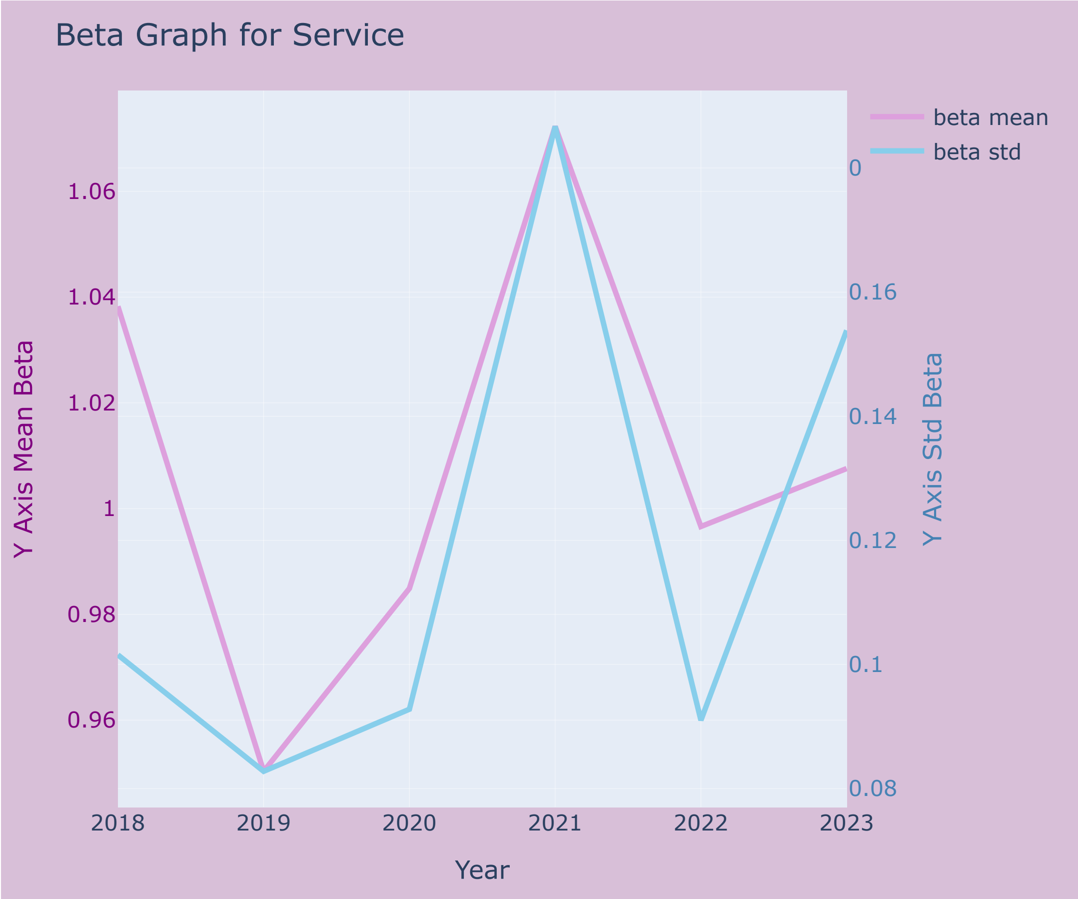

Beta Graph for Manufacturing
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
1.33
1.32
1.31
1.3
1.29
1.28
1.27
1.26
0
0.095
0.09
0.085
0.08
0.075
0.07
0.065
2018
2019
2020
2021
2022
2023
Year

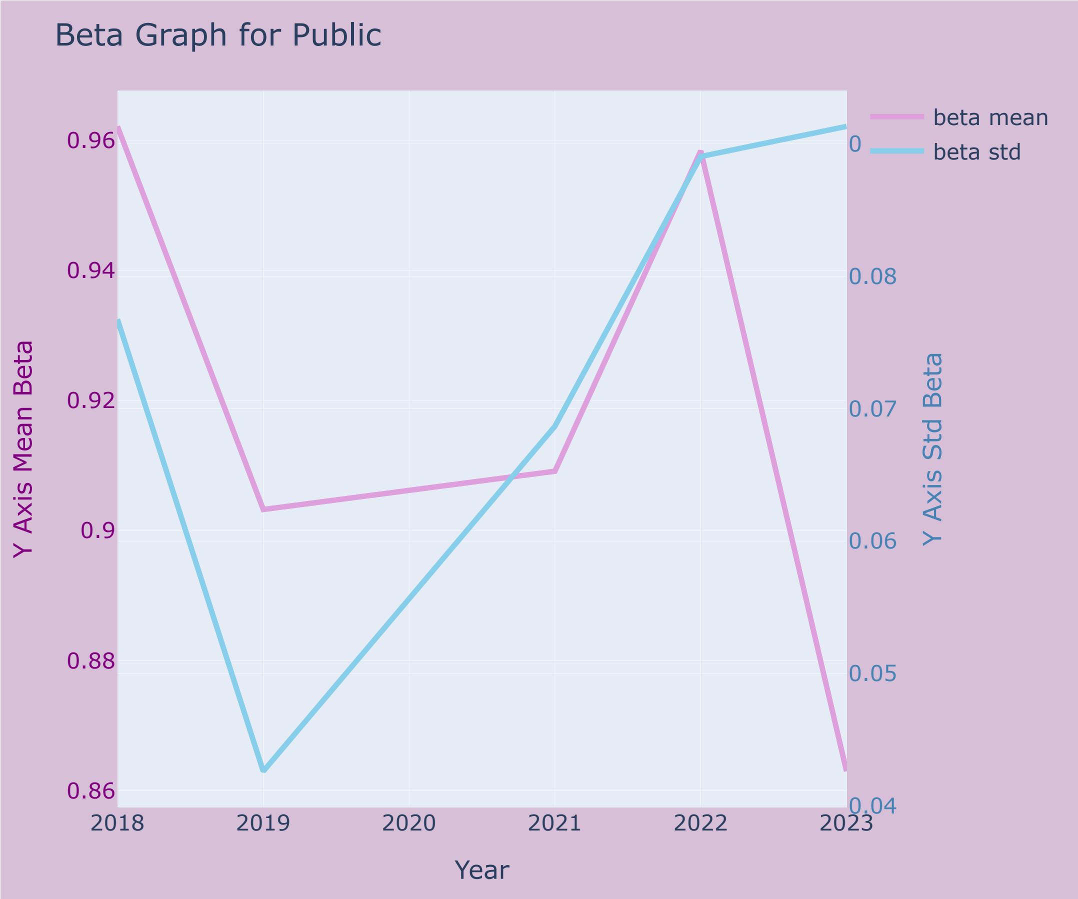

Beta Graph for Transportation
beta mean
beta std
Y Axis Mean Beta
1.26
1.24
1.22
1.2
1.18
Y Axis Std Beta
0
0.085
0.08
0.075
0.07
2018
2019
2020
2021
2022
2023
Year


Beta Graph for Wholesale
beta mean
beta std
Y Axis Mean Beta
1.2
1.18
1.16
1.14
1.12
1.1
Y Axis Std Beta
0.085
0.08
0.075
0.07
0.065
0.06
0.055
0.05
2018
2019
2020
2021
2022
2023
Year


Beta Graph for Retail
beta mean
beta std
Y Axis Mean Beta
1.14
1.13
1.12
1.11
1.1
Y Axis Std Beta
0
0.105
0.1
0.095
0.09
0.085
0.08
0.075
0.07
0.065
2018
2019
2020
2021
2022
2023
Year


Beta Graph for Finance
beta mean
beta std
Y Axis Mean Beta
1.08
1.06
1.04
1.02
1
Y Axis Std Beta
0.18
0.16
0.14
0.12
0.1
0.08
0.06
2018
2019
2020
2021
2022
2023
Year


Beta Graph for Service
beta mean
beta std
Y Axis Mean Beta
1.06
1.04
1.02
1
0.98
0.96
Y Axis Std Beta
0
0.16
0.14
0.12
0.1
0.08
2018
2019
2020
2021
2022
2023
Year



Observations for Beta Mean & Std

- Except for Agriculture, all other industries all have a relatively lower beta in around 2019 and 2020, and their beta recover in 2021 and 2022. In 2019 and 2020, global economic uncertainty, largely driven by the COVID-19 pandemic, led to market disruptions, changes in investor behavior, and a significant reduction in systematic risk across many industries. This could explain the drop in beta, as industries were less sensitive to overall market movements during the crisis.
- Betas for Public are mainly below 1, which is the lowest beta among all industries. A beta below 1 indicates that the Public sector stocks typically exhibit lower fluctuations in their returns relative to the broader market. This could be due to the nature of the Public sector, which often involves stable and essential services with predictable demand.

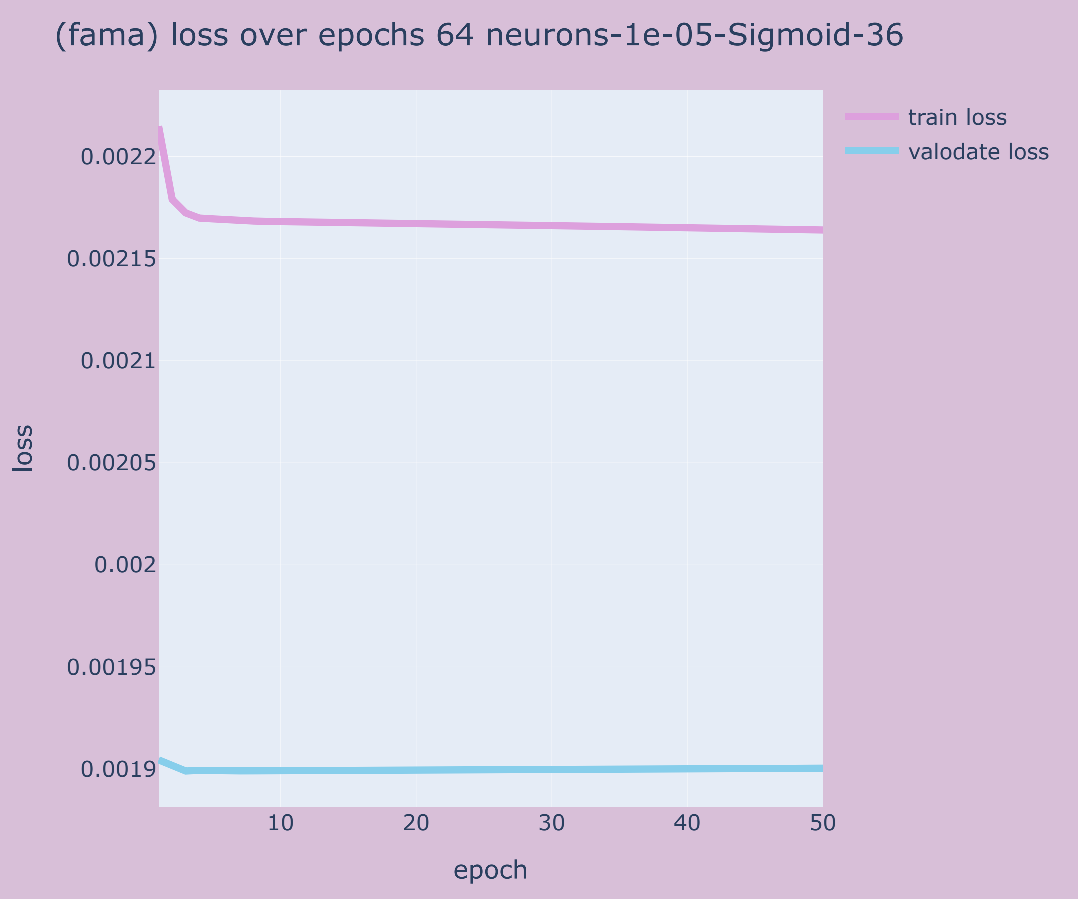
- Volatility in Beta is lower in Public, Retail, Wholesale, Transportation, Construction and Agriculture. These sectors are often tied to essential services or goods, infrastructure, and steady demand, which makes their performance more predictable and less reactive to short-term market fluctuations. As a result, their beta tends to remain more consistent, showing less fluctuation compared to industries that are more susceptible to economic cycles or external shocks.
- Beta for Agriculture, Mining, Construction, Manufacturing are mostly above 1.2.A beta above 1 indicates that these sectors tend to experience greater fluctuations in their returns compared to the broader market. This could be due to their exposure to external factors such as commodity prices (in the case of Agriculture and Mining), economic cycles (in Construction and Manufacturing), and global trade. Higher betas reflect the fact that these industries are more influenced by economic conditions and market volatility, leading to greater risk but also potential for higher returns in favorable market conditions.

Fama-French Factor

According to Commonly used Fama-Frech 5-factor model, which helps explain stock returns based on five different factors, these factors are often used to estimate the beta of a stock and provide a more nuanced understanding of risk by considering multiple dimensions beyond just market exposure.

- SMB - Size - The return difference between small-cap and large-cap stocks
- HML - Value - The return difference between high book-to-market and low book-to-market stocks
- RMW - Profitability - The return difference between stocks with robust profitability and weak profitability
- CMA - Investment - The return difference between companies that invest conservatively and those that invest aggressively
- RM-RF - Market risk premium - The excess return of the market over the risk-free rate

```
Epoch 5/50, Train Loss: 0.00217030, Val Loss: 0.00189897
Epoch 10/50, Train Loss: 0.00216955, Val Loss: 0.00189898
Epoch 15/50, Train Loss: 0.00216930, Val Loss: 0.00189902
Epoch 20/50, Train Loss: 0.00216912, Val Loss: 0.00189904
Epoch 25/50, Train Loss: 0.00216895, Val Loss: 0.00189905
Epoch 30/50, Train Loss: 0.00216876, Val Loss: 0.00189905
Epoch 35/50, Train Loss: 0.00216854, Val Loss: 0.00189905
Epoch 40/50, Train Loss: 0.00216828, Val Loss: 0.00189904
Epoch 45/50, Train Loss: 0.00216798, Val Loss: 0.00189904
Epoch 50/50, Train Loss: 0.00216763, Val Loss: 0.00189906
```

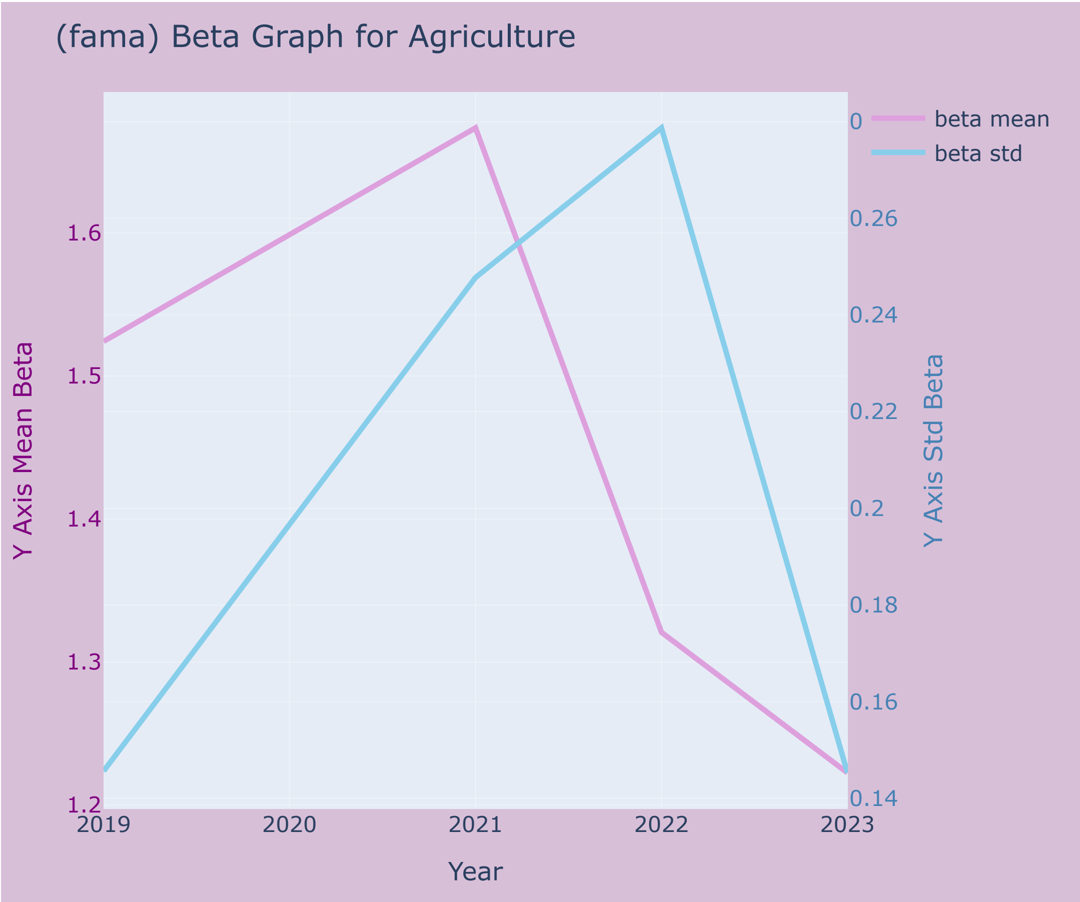

(fama) loss over epochs 64 neurons-1e-05-Sigmoid-36
train loss
valodate loss
loss
0.0022
0.00215
0.0021
0.00205
0.002
0.00195
0.0019
10
20
30
40
50
epoch

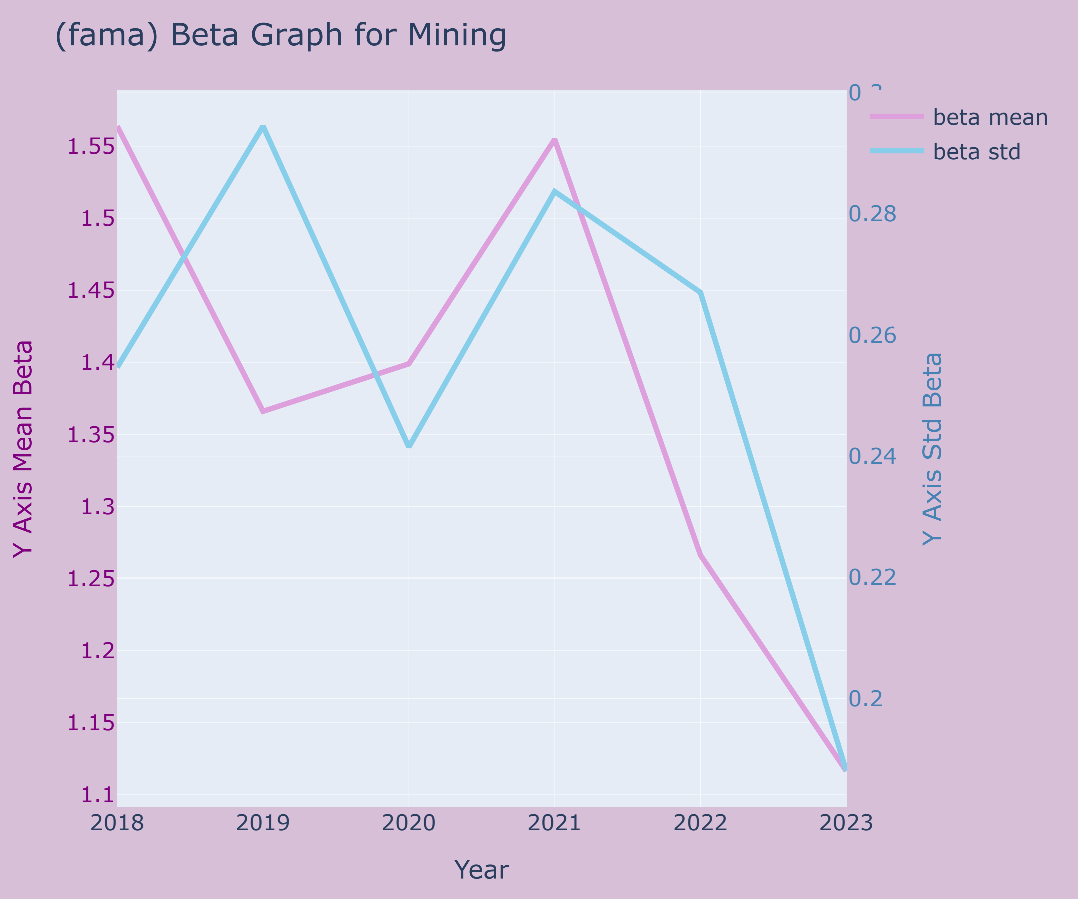
```
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Agriculture
count       85.000000
mean         1.386078
std          0.277596
min          0.879347
max          2.092015
skew         0.197421
kurtosis    -0.619461
1%           0.908810
5%           0.970113
25%          1.153826
50%          1.365737
75%          1.594365
95%          1.826639
99%          1.972119
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Mining
count       695.000000
mean          1.384857
std           0.302984
min           0.470261
max           2.455456
skew          0.221592
kurtosis      0.228026
1%            0.700425
5%            0.897584
25%           1.178943
50%           1.381884
75%           1.566089
95%           1.919885
99%           2.099978
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Construction
count       649.000000
mean          1.353834
std           0.261904
min           0.511526
max           2.043094
skew         -0.130389
```

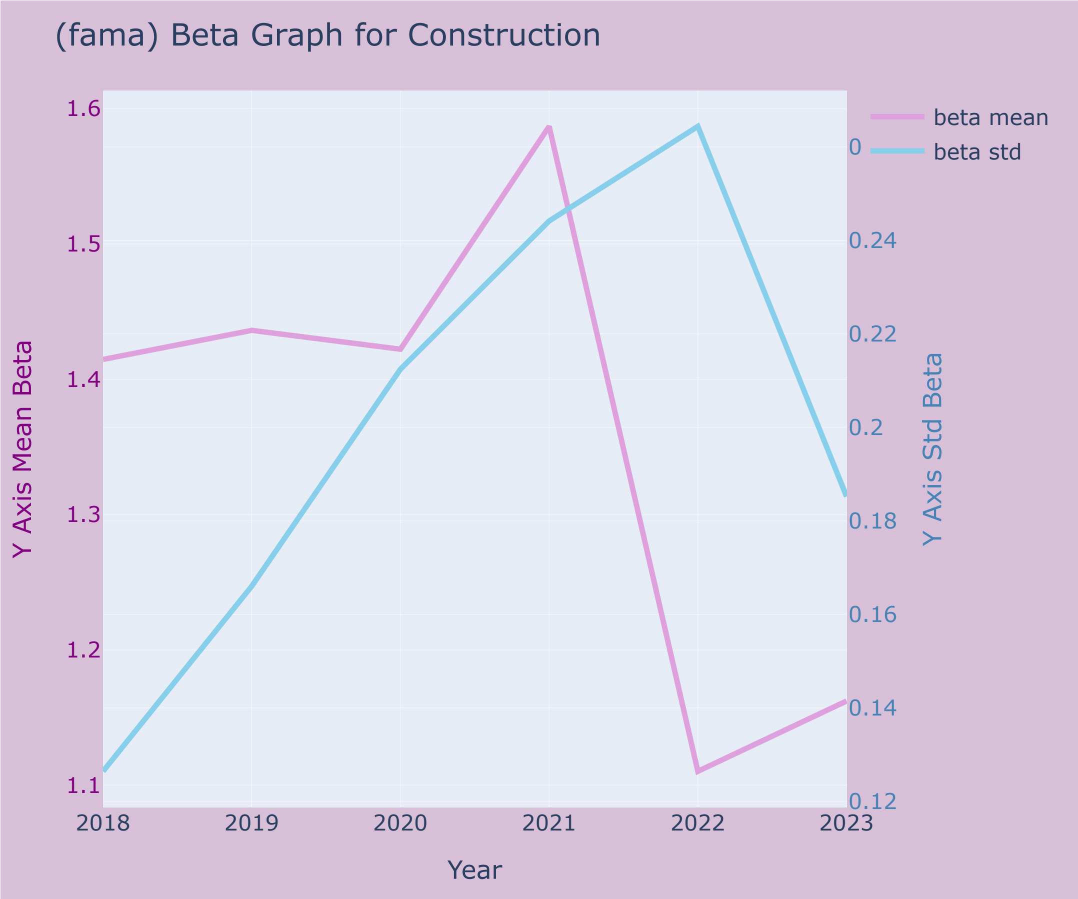
```
kurtosis      0.267870
1%            0.754922
5%            0.877659
25%           1.214203
50%           1.360045
75%           1.505029
95%           1.791960
99%           1.983273
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Manufacturing
count       961.000000
mean          1.291597
std           0.279145
min           0.654168
max           2.545329
skew          0.851960
kurtosis      2.604472
1%            0.748781
5%            0.838551
25%           1.125730
50%           1.277599
75%           1.441196
95%           1.741669
99%           2.347778
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Transportation
count       751.000000
mean          1.217331
std           0.251454
min           0.617366
max           1.872561
skew         -0.003260
kurtosis     -0.145285
1%            0.662248
5%            0.765107
25%           1.069525
50%           1.217834
75%           1.377279
95%           1.642257
99%           1.800071
Name: beta_pred, dtype: float64
```

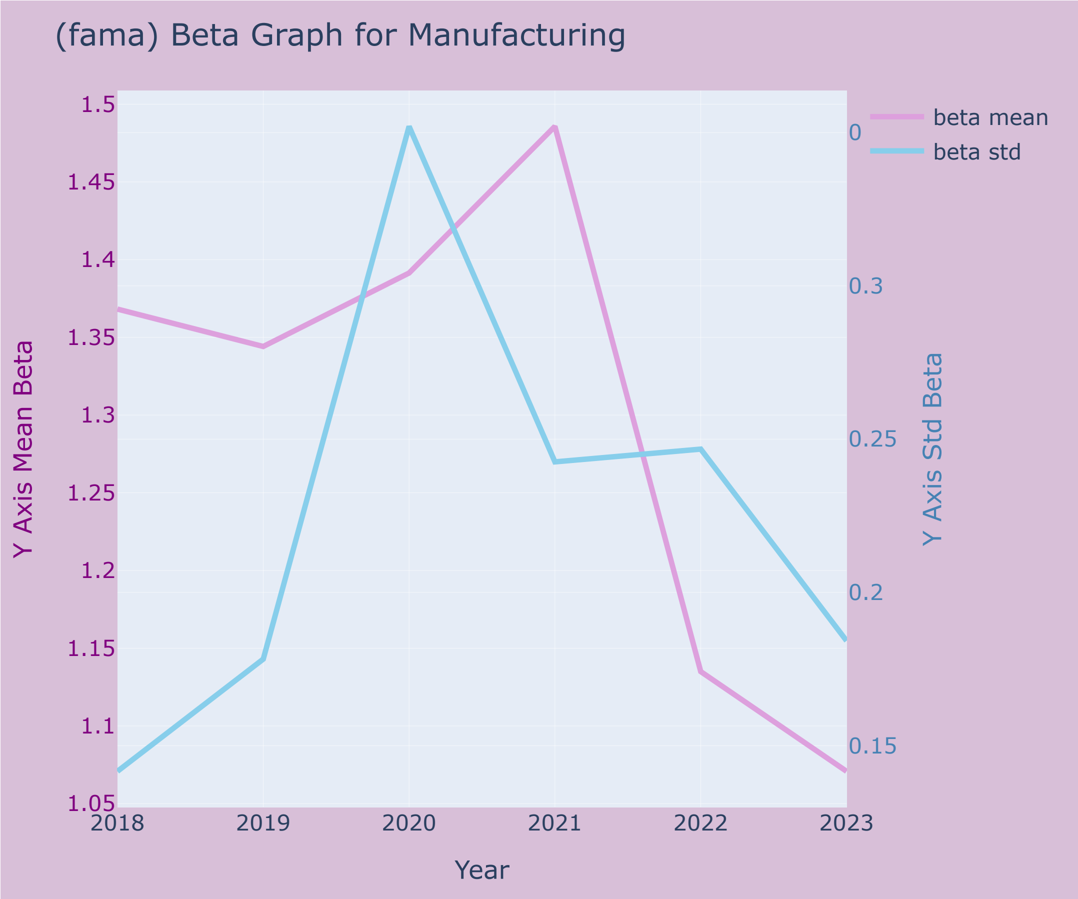
```
------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Wholesale
count       633.000000
mean          1.147328
std           0.267677
min           0.518902
max           1.909435
skew          0.187846
kurtosis     -0.098774
1%            0.595560
5%            0.690129
25%           0.984517
50%           1.136373
75%           1.309447
95%           1.630669
99%           1.805778
Name: beta_pred, dtype: float64
------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Retail
count       649.000000
mean          1.090775
std           0.249326
min           0.499394
max           1.803532
skew          0.226586
kurtosis     -0.050902
1%            0.583531
5%            0.666898
25%           0.925599
50%           1.082597
75%           1.232326
95%           1.545423
99%           1.721195
Name: beta_pred, dtype: float64
------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Finance
count       857.000000
mean          1.027503
std           0.281752
min           0.181520
max           1.875697
skew          0.245777
```

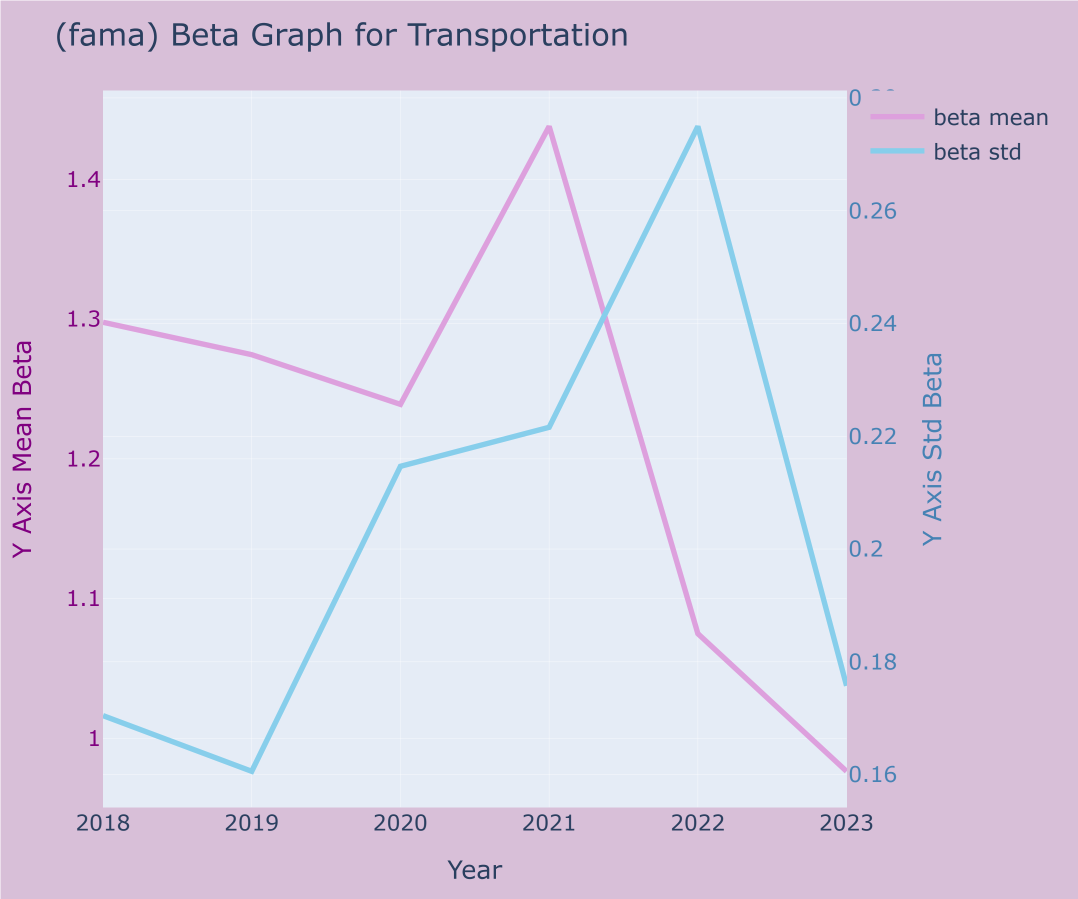
```
kurtosis      0.302034
1%            0.438468
5%            0.562948
25%           0.857065
50%           1.009103
75%           1.191074
95%           1.532519
99%           1.770036
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Service
count       698.000000
mean          0.944316
std           0.248592
min           0.328225
max           1.828201
skew          0.348858
kurtosis      0.232203
1%            0.449414
5%            0.529527
25%           0.778488
50%           0.926257
75%           1.097359
95%           1.392397
99%           1.606690
Name: beta_pred, dtype: float64
--------------------------------------------------

This is the Statistics of predicted beta (with fama factor) for industry Public
count       73.000000
mean         0.893677
std          0.226938
min          0.401152
max          1.454449
skew        -0.178129
kurtosis    -0.353178
1%           0.428114
5%           0.478379
25%          0.750958
50%          0.939204
75%          1.026312
95%          1.208259
99%          1.338064
Name: beta_pred, dtype: float64
```

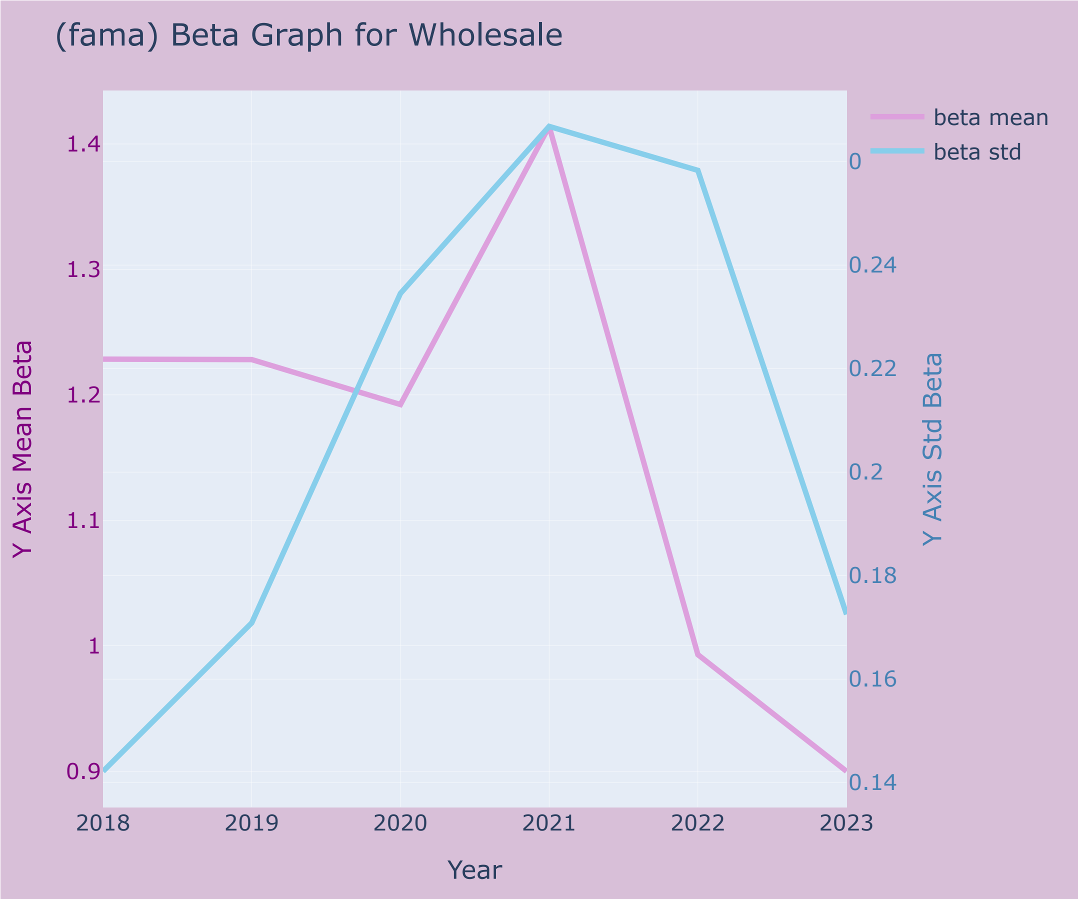

(fama) Beta Graph for Agriculture
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
Year
1.2
1.3
1.4
1.5
1.6
0
0.26
0.24
0.22
0.2
0.18
0.16
0.14
2019
2020
2021
2022
2023

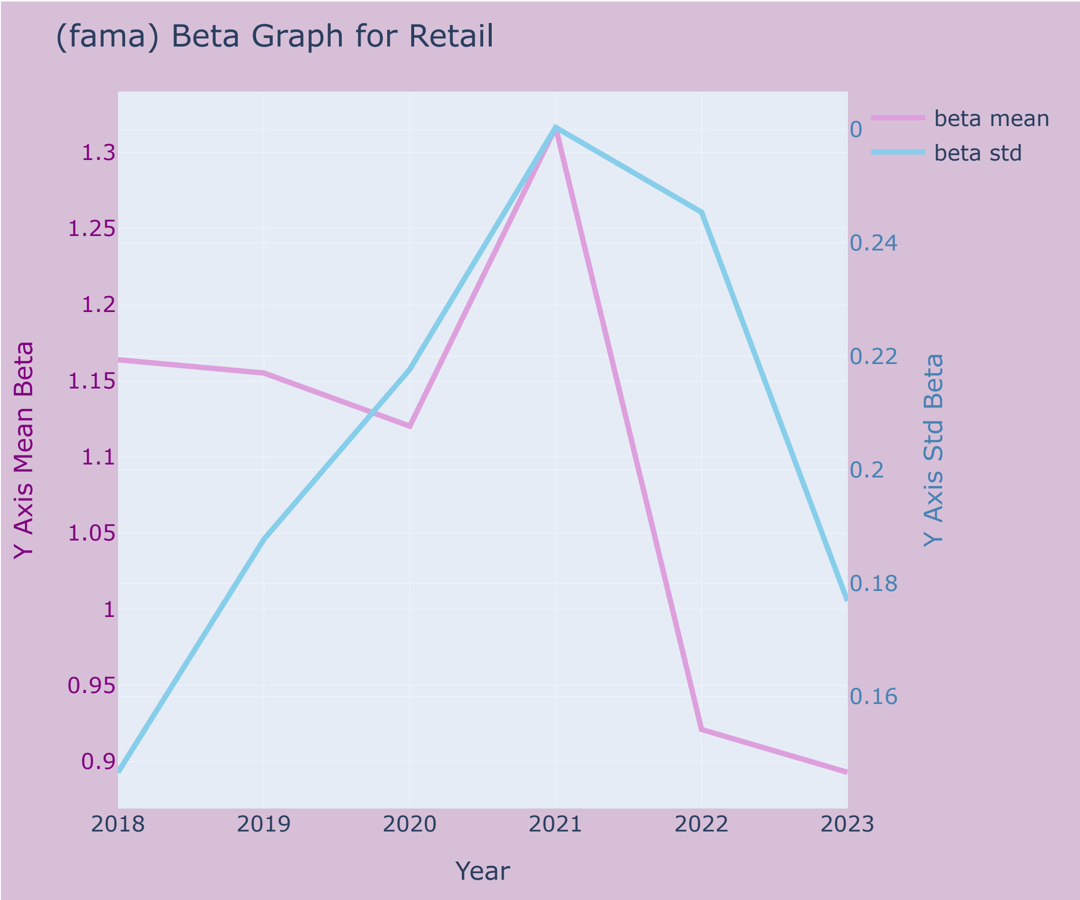

(fama) Beta Graph for Mining
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
Year
2018
2019
2020
2021
2022
2023
1.1
1.15
1.2
1.25
1.3
1.35
1.4
1.45
1.5
1.55
0.2
0.22
0.24
0.26
0.28

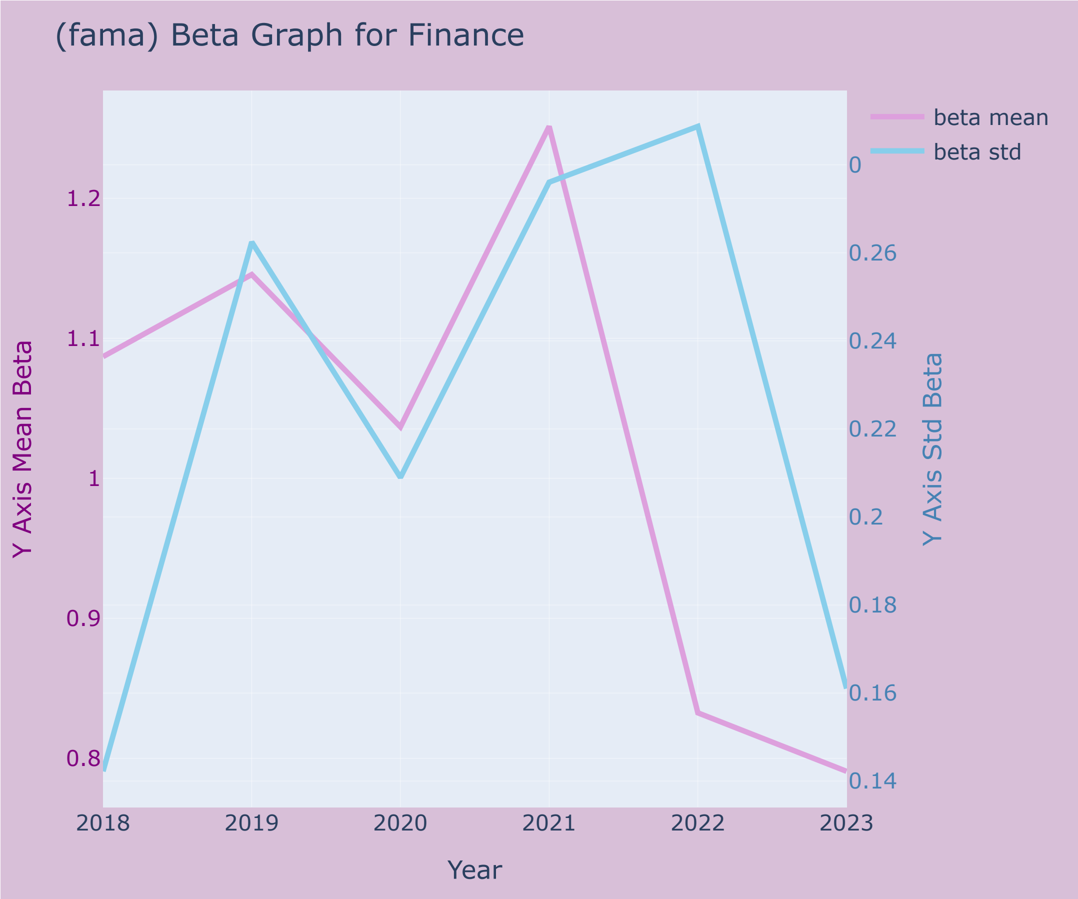

(fama) Beta Graph for Construction
beta mean
beta std
Y Axis Mean Beta
1.6
1.5
1.4
1.3
1.2
1.1
Y Axis Std Beta
0
0.24
0.22
0.2
0.18
0.16
0.14
0.12
2018
2019
2020
2021
2022
2023
Year

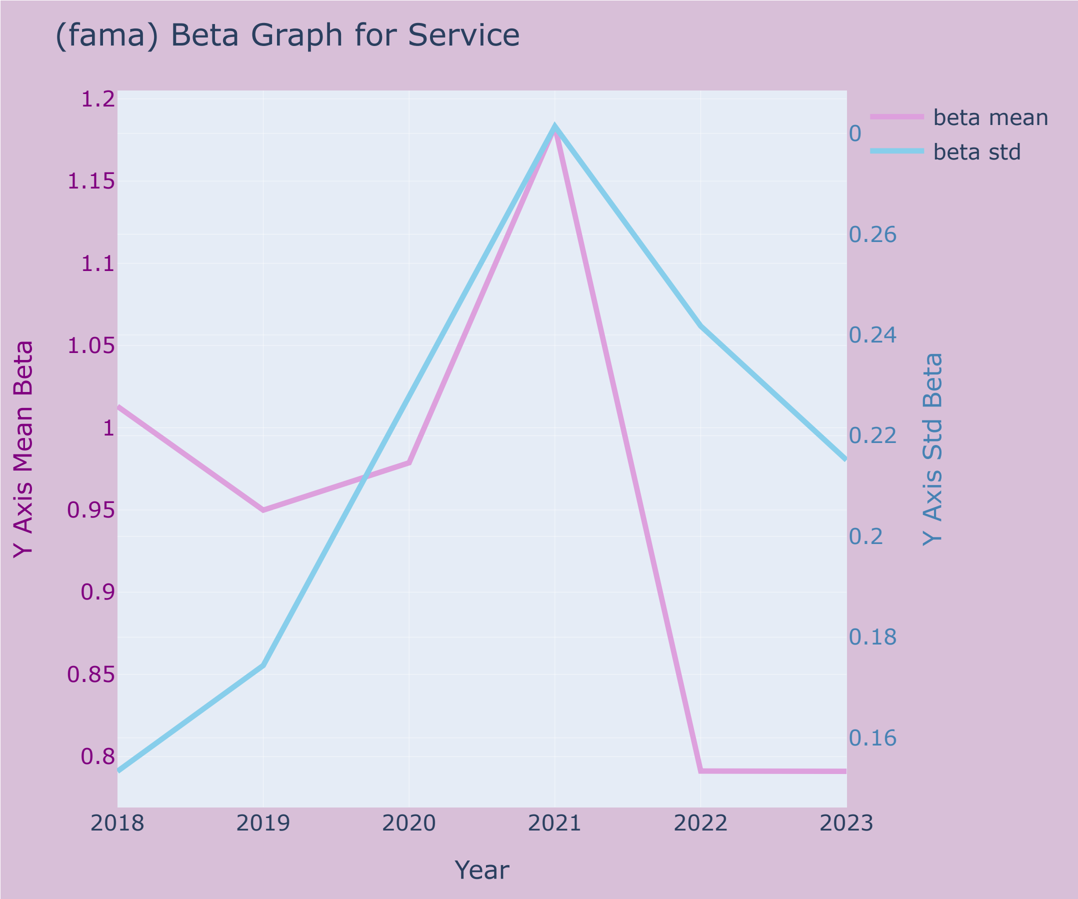

(fama) Beta Graph for Manufacturing
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
1.5
1.45
1.4
1.35
1.3
1.25
1.2
1.15
1.1
1.05
0
0.3
0.25
0.2
0.15
2018
2019
2020
2021
2022
2023
Year

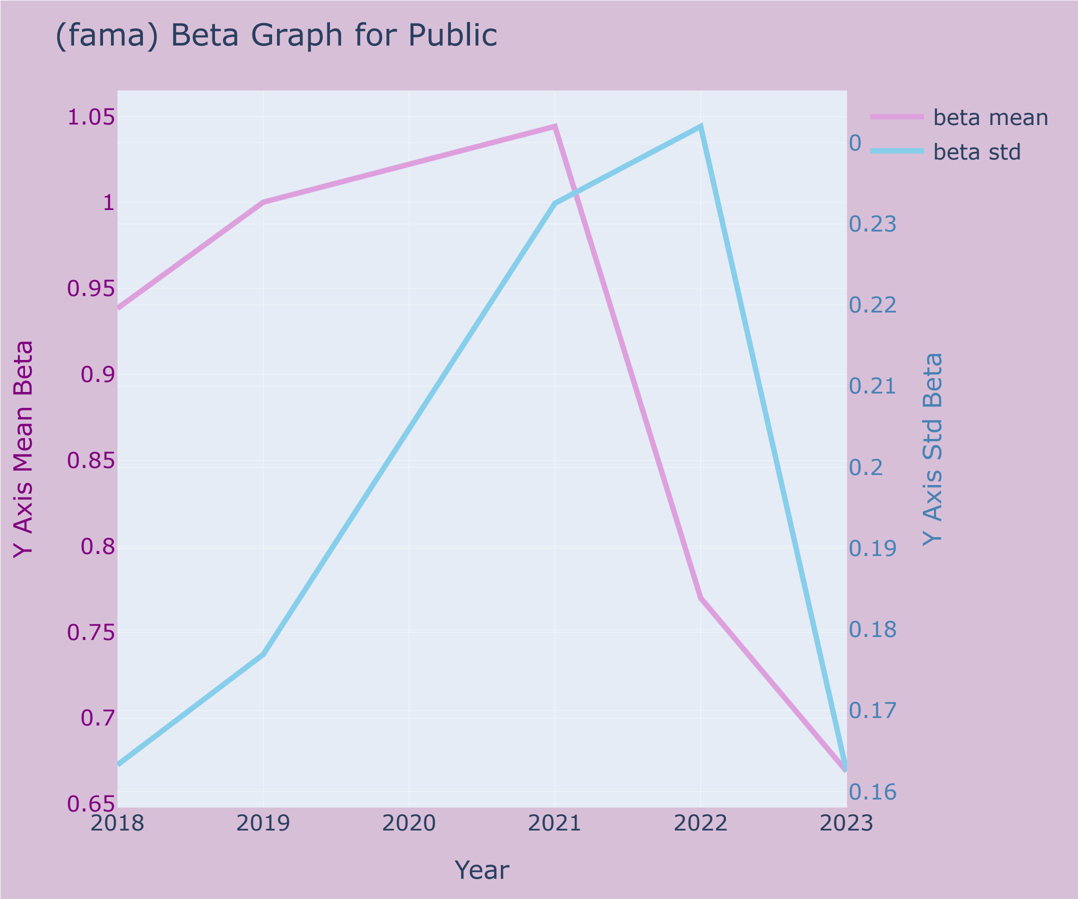

(fama) Beta Graph for Transportation
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
Year
1.4
1.3
1.2
1.1
1
0.26
0.24
0.22
0.2
0.18
0.16
2018
2019
2020
2021
2022
2023


(fama) Beta Graph for Wholesale
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
1.4
1.3
1.2
1.1
1
0.9
0
0.24
0.22
0.2
0.18
0.16
0.14
2018
2019
2020
2021
2022
2023
Year


(fama) Beta Graph for Retail
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
Year
2018
2019
2020
2021
2022
2023
1.3
1.25
1.2
1.15
1.1
1.05
1
0.95
0.9
0
0.24
0.22
0.2
0.18
0.16


(fama) Beta Graph for Finance
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
1.2
1.1
1
0.9
0.8
0
0.26
0.24
0.22
0.2
0.18
0.16
0.14
2018
2019
2020
2021
2022
2023
Year


(fama) Beta Graph for Service
beta mean
beta std
Y Axis Mean Beta
Y Axis Std Beta
Year
1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
0
0.26
0.24
0.22
0.2
0.18
0.16
2018
2019
2020
2021
2022
2023



Observation & Comparison

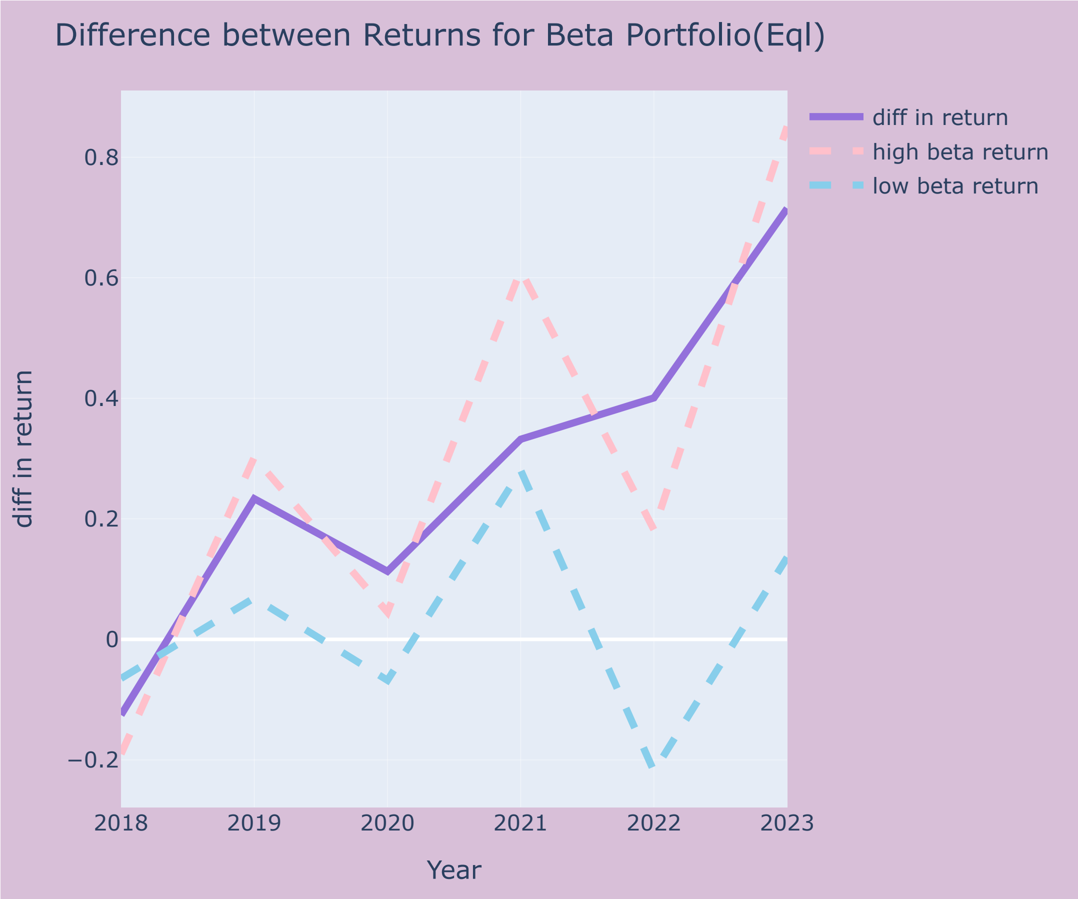
- The RMSE does not change alot after including fama-french factor
- Distributions for all betas are higher after include fama-french factor to predict them. The Fama-French factors in the beta prediction model captures additional sources of risk, leading to an overall increase in beta distributions and volatilities
- Volatilities for all betas also rise. The rise in beta volatilities indicates that these factors introduce more sensitivity to market movements across industries.
- most of patterns of the drop of beta in 2019 and 2020 are maintained with a slight less extreme trend. The fact that the general pattern of the beta drop in 2019 and 2020 is maintained, though slightly less extreme, implies that the Fama-French model smooths out some of the market anomalies but still reflects the underlying market dynamics during that period, such as the impact of the pandemic.

- Volatilities has an obvious pattern with a peak in 2020. This suggests that 2022 marked a period where industries experienced more unpredictable shifts in returns, with greater fluctuations in their betas. This aligns with broader market behavior during times of economic instability or transition.

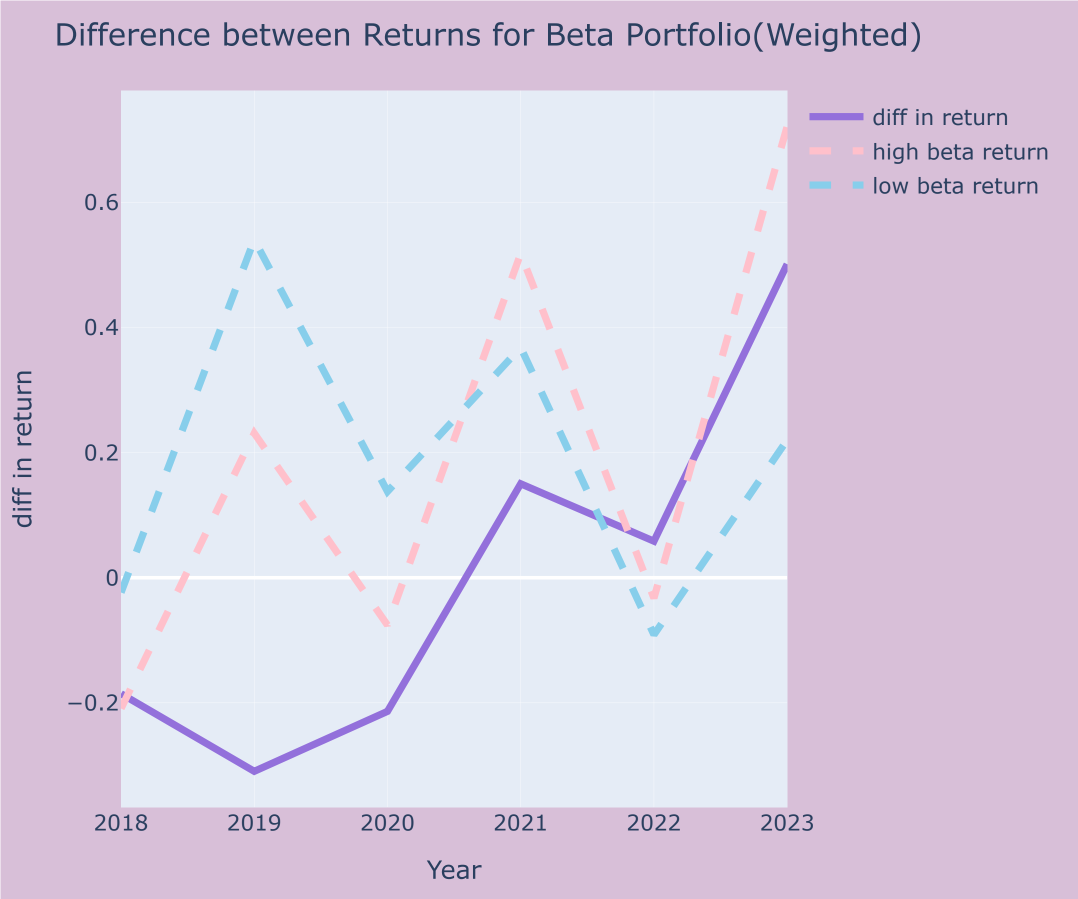
Beta and Stock Returns

Beta estimated w/o fama factor

```
Average return rate for high beta equal portfolio is 0.30044238482009095
Average return rate for low beta equal portfolio is 0.02231612256142222
```

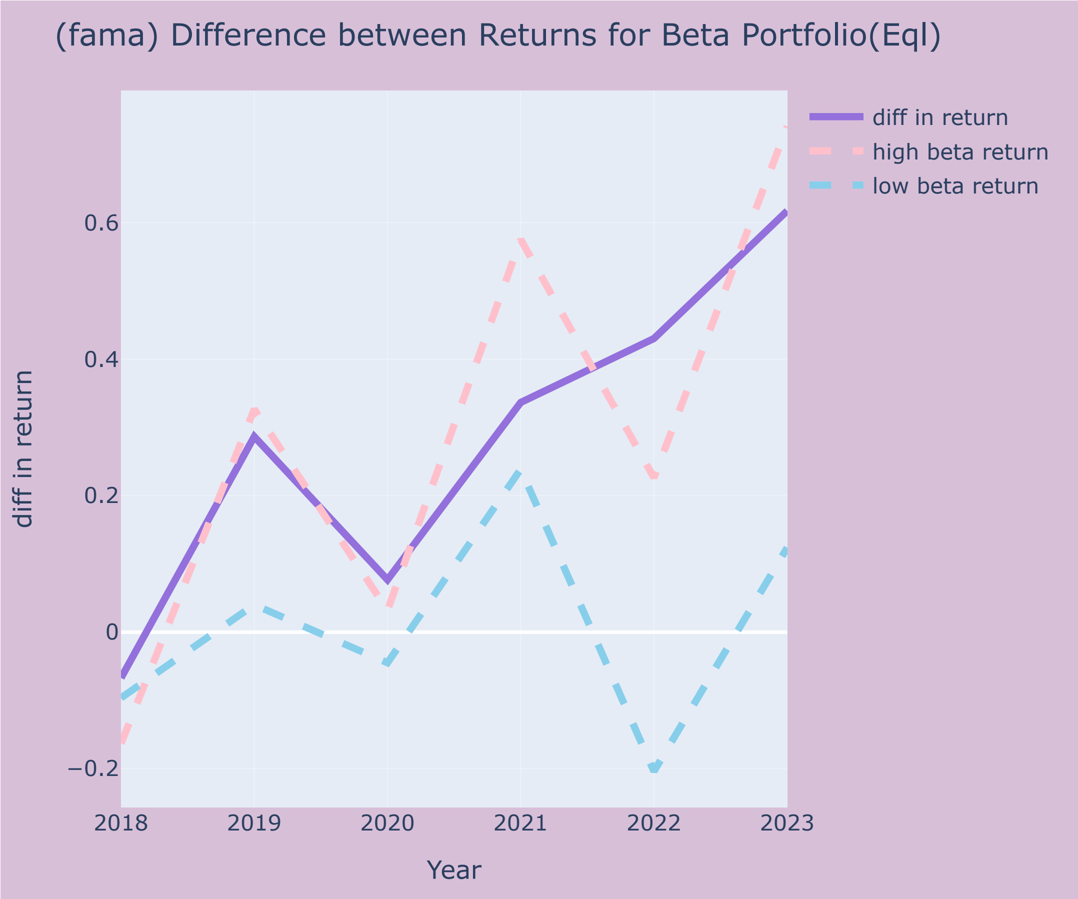



```
Return rate for high beta weighted portfolio  is 0.1923534679498584
Return rate for low beta weighted portfolio  is 0.1920980712952716
```

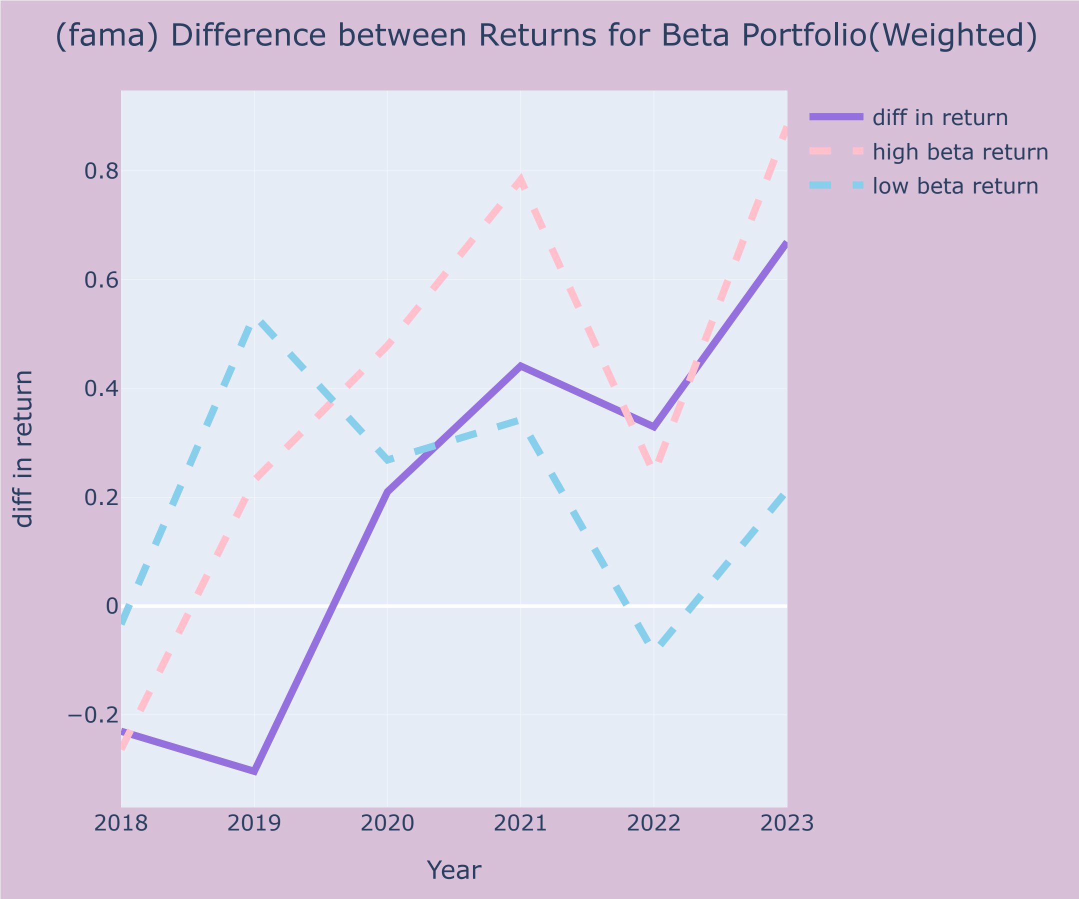


Beta estimated with fama factor

```
Average return rate for high beta equal portfolio (fama) is 0.2897096496528299
Average return rate for low beta equal portfolio (fama) is 0.00962667804433102
```




```
Return rate for high beta weighted portfolio (fama)  is 0.3926676021300841
Return rate for low beta weighted portfolio (fama)  is 0.20659285398547847
```



Discussion

- Equal weighted portfolio w/o fama factor
  - This portfolio performs about 25% better (0.30 portfolio return) than the portfolio using the beta predicted by linear regression (0.24 return). Moreover, the difference in return for high beta and low beta has most of the part lying above 0, showing a general better performance of the portfolio constructed by high beta. This might be due to the better estimation of beta generated by neural network. However, since we only predict beta from 2018 to 2023, more validations should be made to prove the effectiveness of such beta prediction
- Value weighted portfolio w/o fama factor

- The value weighted portfolio perform worse than the equal weighted portfolio, and the portfolio with high beta does not outperform the portfolio with low beta. The underperformance of the value-weighted portfolio compared to the equal-weighted portfolio could indicate that larger, more dominant stocks in the portfolio (which have more weight in a value-weighted approach) may not have performed as well during the observation period. Smaller or less dominant stocks, which get equal representation in an equal-weighted portfolio, might have experienced better performance, driving the better returns in the equal-weighted approach.

- Equal weighted portfolio with fama factor

  - This portfolio also has a good performance as the equal weighted portfolio w/o fama factor. It maintains a more obvious pattern that the high beta return is overall better than the low beta return. This suggests that including Fama-French factors in the equal-weighted portfolio enhances its performance, likely by better capturing the sources of risk and return linked to market, size, and value factors. The fact that the portfolio now shows a more obvious pattern where high-beta returns outperform low-beta returns indicates that the Fama-French model may help explain the risk premia associated with high-beta stocks more effectively than a model without these factors.

- Value weighted portfolio with fama factor

  - This portfolio has a better performance than the Value weighted portfolio w/o fama factor, but it still doesn't diversify the performance of high beta and low beta portfolio, having a siginificant portion of low beta portfolio with higher return. While the Fama-French factors improve the performance of the value-weighted portfolio, they don't fully capture or account for the expected relationship between beta and returns. The fact that a significant portion of the low-beta portfolio still outperforms the high-beta portfolio suggests that, even with the Fama-French model, other factors or market conditions may favor lower-risk stocks in the value-weighted approach. This could imply that larger, more stable stocks (which tend to dominate value-weighted portfolios) may benefit more from lower risk and steady returns, making them outperform despite their lower beta.

- Overall all the differences curves have increasing trends, showing that the difference in beta might diversify them better as time goes. This could indicate that as time progresses, the differences in beta are becoming a stronger factor in diversification, potentially allowing for better risk-adjusted returns as portfolios with varying betas respond differently to market conditions.