problem12

May 1, 2023

1 1|

[ ]:
```
import pandas as pd
import numpy as np
from tqdm import tqdm

# Read data from CSV files
bcwlist = pd.read_excel('data/bcwlist.xlsx', skiprows=2)
crsp = pd.read_csv('data/cleaned_crsp.csv')

# Convert date columns to datetime format
crsp['date'] = pd.to_datetime(crsp['date'])

crsp['year'] = crsp['date'].apply(lambda x: x.year)

# Convert RET and PRC columns to numeric data types and handle invalid values
crsp['RET'] = pd.to_numeric(crsp['RET'], errors='coerce')
crsp['PRC'] = pd.to_numeric(crsp['PRC'], errors='coerce')

# Group by date and permno, and aggregate using the last valid observation
crsp = crsp.groupby(['PERMNO', pd.Grouper(key='date', freq='M')]).last().
 ↪reset_index()
crsp['date'] = crsp['date'].dt.to_period('M')

import warnings
warnings.filterwarnings('ignore')
```

[ ]:
```
# Merge stock returns with best companies data
crsp['market_cap'] = crsp['SHROUT'] * crsp['PRC'].abs()
merged_data = crsp.merge(bcwlist, left_on=['PERMNO', 'year'],␣
 ↪right_on=['permno', 'year'], how='left')
merged_data = merged_data[merged_data.year >= 1984]
# merged_data['rank'] = merged_data['rank'].dropna()

# Create the equal-weighted portfolio
merged_data['ew_weight'] = merged_data.groupby('year')['rank'].transform(lambda␣
 ↪x: 1 / x.shape[0] if x.notnull().sum() >= 1 else np.inf)
```

```
merged_data['ew_returns'] = merged_data['ew_weight'] * merged_data['RET']

# Create the value-weighted portfolio
merged_data['market_cap_lag'] = merged_data.groupby('PERMNO')['market_cap'].
↪shift(1)
merged_data['vw_weight'] = merged_data.groupby('year').apply(lambda x: x.
↪loc[x['rank'].notnull(), 'market_cap_lag'] / x.loc[x['rank'].notnull(),␣
↪'market_cap_lag'].sum()).reset_index(level=0, drop=True)
merged_data['vw_returns'] = merged_data['vw_weight'] * merged_data['RET']

# Calculate monthly returns for each portfolio
monthly_returns = merged_data.groupby('date')[['ew_returns', 'vw_returns']].
↪sum().reset_index()
monthly_returns = monthly_returns.set_index('date')
```

[ ]:
```
# After talking to peers, it seems like our numbers are off by essentially a␣
↪factor of 10 and we have 0 idea why.
# Leaving * 10 in here to make the rest of the analysis make sense. Note, this␣
↪does mess with p-values.
# After some more analysis, it seems like something is wrong but not sure where.
↪ :( Leaving the 10 out for now
monthly_returns = monthly_returns * 10
monthly_returns = monthly_returns.dropna()
# monthly_returns = monthly_returns[monthly_returns.index.year > 1985]
```

[ ]:
```
monthly_returns.describe()
```

[ ]:

	ew_returns	vw_returns
count	300.000000	300.000000
mean	0.008805	0.009254
std	0.050662	0.045144
min	-0.183843	-0.153201
25%	-0.017631	-0.015553
50%	0.009832	0.010407
75%	0.035171	0.034636
max	0.215744	0.171296

2 2a

[ ]:
```
import pandas_datareader as pdr

ff3 = pdr.get_data_famafrench('F-F_Research_Data_Factors', start='1984-01',␣
↪end='2020-12')[0]
ff5 = pdr.get_data_famafrench('F-F_Research_Data_5_Factors_2x3',␣
↪start='1984-01', end='2020-12')[0]
ff3 = ff3 / 100
```

```
ff5 = ff5 / 100
```

[ ]:
```
ff3 = ff3.loc[monthly_returns.index, :]
ff5 = ff5.loc[monthly_returns.index, :]
```

[ ]:
```
def calculate_statistics(portfolio_returns):
    avg_monthly_return = portfolio_returns.mean()
    volatility = portfolio_returns.std()
    sharpe_ratio = avg_monthly_return / volatility
    return avg_monthly_return, volatility, sharpe_ratio

for weight_type in ['ew_returns', 'vw_returns']:
    avg_monthly_return, volatility, sharpe_ratio =␣
 ↪calculate_statistics(monthly_returns[weight_type])
    print(f"{weight_type.capitalize()} Weighted Portfolio:")
    print(f"Average Monthly Return: {avg_monthly_return:.4f}")
    print(f"Volatility: {volatility:.4f}")
    print(f"Sharpe Ratio: {sharpe_ratio:.4f}")
    print()
```

```
Ew_returns Weighted Portfolio:
Average Monthly Return: 0.0088
Volatility: 0.0507
Sharpe Ratio: 0.1738

Vw_returns Weighted Portfolio:
Average Monthly Return: 0.0093
Volatility: 0.0451
Sharpe Ratio: 0.2050
```

3 2b

[ ]:
```
from statsmodels.api import OLS
from statsmodels.tools import add_constant

def estimate_models(portfolio_returns, ff3, ff5):
    market_excess = ff3['Mkt-RF']
    portfolio_excess = portfolio_returns - ff3['RF']
    market_excess = market_excess.loc[portfolio_excess.index]

    print(portfolio_excess[portfolio_excess.isna()])

    # CAPM Model
    X = add_constant(market_excess)
    capm_model = OLS(portfolio_excess, X).fit()
```

```
    # FF3 Model
    X = add_constant(ff3[['Mkt-RF', 'SMB', 'HML']].loc[portfolio_excess.index])
    ff3_model = OLS(portfolio_excess, X).fit()

    # Carhart Model
    X = add_constant(ff3.loc[portfolio_excess.index].join(ff5['RMW']).
 ↪join(ff5['CMA']))
    carhart_model = OLS(portfolio_excess, X).fit()

    # FF5 Model
    X = add_constant(ff5.loc[portfolio_excess.index])
    ff5_model = OLS(portfolio_excess, X).fit()

    return capm_model, ff3_model, carhart_model, ff5_model

for weight_type in ['ew_returns', 'vw_returns']:
    capm_model, ff3_model, carhart_model, ff5_model =␣
 ↪estimate_models(monthly_returns[weight_type], ff3, ff5)
    print(f"{weight_type.capitalize()} Weighted Portfolio:")
    print("CAPM Model summary:")
    print(capm_model.summary())
    print("FF3 Model summary:")
    print(ff3_model.summary())
    print("Carhart Model summary:")
    print(carhart_model.summary())
    print("FF5 Model summary:")
    print(ff5_model.summary())
    print()
```

```
Series([], Freq: M, dtype: float64)
Ew_returns Weighted Portfolio:
CAPM Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.770
Model:                            OLS   Adj. R-squared:                  0.770
Method:                 Least Squares   F-statistic:                     999.3
Date:                Mon, 01 May 2023   Prob (F-statistic):           3.42e-97
Time:                        17:42:51   Log-Likelihood:                 688.78
No. Observations:                 300   AIC:                            -1374.
Df Residuals:                     298   BIC:                            -1366.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const          0.0011      0.001      0.769      0.442      -0.002       0.004
Mkt-RF         0.9824      0.031     31.612      0.000       0.921       1.044
==============================================================================
Omnibus:                      136.572   Durbin-Watson:                   1.897
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1245.154
Skew:                           1.617   Prob(JB):                    4.15e-271
Kurtosis:                      12.442   Cond. No.                         22.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
FF3 Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.903
Model:                            OLS   Adj. R-squared:                  0.902
Method:                 Least Squares   F-statistic:                     917.0
Date:                Mon, 01 May 2023   Prob (F-statistic):          1.81e-149
Time:                        17:42:51   Log-Likelihood:                 817.86
No. Observations:                 300   AIC:                            -1628.
Df Residuals:                     296   BIC:                            -1613.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0006      0.001      0.640      0.523      -0.001       0.002
Mkt-RF         0.8680      0.021     41.006      0.000       0.826       0.910
SMB            0.6141      0.031     20.006      0.000       0.554       0.675
HML            0.1628      0.029      5.646      0.000       0.106       0.220
==============================================================================
Omnibus:                      257.865   Durbin-Watson:                   2.191
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9278.248
Skew:                           3.200   Prob(JB):                         0.00
Kurtosis:                      29.482   Cond. No.                         35.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
Carhart Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.906
Model:                            OLS   Adj. R-squared:                  0.905
Method:                 Least Squares   F-statistic:                     473.1
Date:                Mon, 01 May 2023   Prob (F-statistic):          1.68e-147
```

```
Time:                        17:42:51   Log-Likelihood:                 823.50
No. Observations:                 300   AIC:                            -1633.
Df Residuals:                     293   BIC:                            -1607.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0030      0.001      2.328      0.021       0.000       0.006
Mkt-RF         0.8363      0.024     35.116      0.000       0.789       0.883
SMB            0.5734      0.035     16.544      0.000       0.505       0.642
HML            0.2228      0.039      5.719      0.000       0.146       0.299
RF            -0.9376      0.463     -2.025      0.044      -1.849      -0.026
RMW           -0.1129      0.044     -2.573      0.011      -0.199      -0.027
CMA           -0.0702      0.060     -1.169      0.243      -0.188       0.048
==============================================================================
Omnibus:                      273.263   Durbin-Watson:                   2.209
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11212.426
Skew:                           3.461   Prob(JB):                         0.00
Kurtosis:                      32.139   Cond. No.                         510.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
FF5 Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.908
Model:                            OLS   Adj. R-squared:                  0.906
Method:                 Least Squares   F-statistic:                     479.2
Date:                Mon, 01 May 2023   Prob (F-statistic):          3.03e-148
Time:                        17:42:51   Log-Likelihood:                 825.25
No. Observations:                 300   AIC:                            -1637.
Df Residuals:                     293   BIC:                            -1611.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0030      0.001      2.376      0.018       0.001       0.006
Mkt-RF         0.8270      0.024     34.850      0.000       0.780       0.874
SMB            0.5717      0.034     16.745      0.000       0.504       0.639
HML            0.1200      0.039      3.077      0.002       0.043       0.197
RMW           -0.1547      0.042     -3.644      0.000      -0.238      -0.071
CMA           -0.0614      0.060     -1.030      0.304      -0.179       0.056
RF            -0.8686      0.460     -1.887      0.060      -1.775       0.038
==============================================================================
```

```
Omnibus:                      277.250   Durbin-Watson:                   2.256
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11869.623
Skew:                           3.526   Prob(JB):                         0.00
Kurtosis:                      32.997   Cond. No.                         510.
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
Series([], Freq: M, dtype: float64)
Vw_returns Weighted Portfolio:
CAPM Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.805
Model:                            OLS   Adj. R-squared:                  0.804
Method:                 Least Squares   F-statistic:                     1229.
Date:                Mon, 01 May 2023   Prob (F-statistic):          9.43e-108
Time:                        17:42:51   Log-Likelihood:                 748.51
No. Observations:                 300   AIC:                            -1493.
Df Residuals:                     298   BIC:                            -1486.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0021      0.001      1.785      0.075      -0.000       0.004
Mkt-RF         0.8928      0.025     35.058      0.000       0.843       0.943
==============================================================================
Omnibus:                       33.187   Durbin-Watson:                   2.075
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               86.052
Skew:                           0.502   Prob(JB):                     2.06e-19
Kurtosis:                       5.424   Cond. No.                         22.0
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
FF3 Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.840
Model:                            OLS   Adj. R-squared:                  0.838
Method:                 Least Squares   F-statistic:                     517.7
Date:                Mon, 01 May 2023   Prob (F-statistic):          2.18e-117
Time:                        17:42:51   Log-Likelihood:                 778.23
No. Observations:                 300   AIC:                            -1548.
```

```
Df Residuals:                     296   BIC:                            -1534.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0024      0.001      2.217      0.027       0.000       0.004
Mkt-RF         0.8968      0.024     37.120      0.000       0.849       0.944
SMB           -0.0917      0.035     -2.618      0.009      -0.161      -0.023
HML           -0.2624      0.033     -7.973      0.000      -0.327      -0.198
==============================================================================
Omnibus:                       26.868   Durbin-Watson:                   2.130
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               41.132
Skew:                           0.581   Prob(JB):                     1.17e-09
Kurtosis:                       4.393   Cond. No.                         35.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
Carhart Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.850
Model:                            OLS   Adj. R-squared:                  0.847
Method:                 Least Squares   F-statistic:                     276.6
Date:                Mon, 01 May 2023   Prob (F-statistic):          1.67e-117
Time:                        17:42:51   Log-Likelihood:                 787.91
No. Observations:                 300   AIC:                            -1562.
Df Residuals:                     293   BIC:                            -1536.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0015      0.001      1.070      0.285      -0.001       0.004
Mkt-RF         0.8631      0.027     32.189      0.000       0.810       0.916
SMB           -0.0717      0.039     -1.838      0.067      -0.149       0.005
HML           -0.1600      0.044     -3.649      0.000      -0.246      -0.074
RF             0.8179      0.521      1.569      0.118      -0.208       1.844
RMW           -0.0015      0.049     -0.030      0.976      -0.099       0.096
CMA           -0.2761      0.068     -4.084      0.000      -0.409      -0.143
==============================================================================
Omnibus:                       19.301   Durbin-Watson:                   2.097
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               24.507
Skew:                           0.508   Prob(JB):                     4.77e-06
Kurtosis:                       3.963   Cond. No.                         510.
==============================================================================
```

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
FF5 Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.851
Model:                            OLS   Adj. R-squared:                  0.848
Method:                 Least Squares   F-statistic:                     279.2
Date:                Mon, 01 May 2023   Prob (F-statistic):         5.19e-118
Time:                        17:42:51   Log-Likelihood:                 789.11
No. Observations:                 300   AIC:                            -1564.
Df Residuals:                     293   BIC:                            -1538.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0016      0.001      1.130      0.260      -0.001       0.004
Mkt-RF         0.8659      0.027     32.347      0.000       0.813       0.919
SMB           -0.0924      0.039     -2.399      0.017      -0.168      -0.017
HML           -0.1442      0.044     -3.279      0.001      -0.231      -0.058
RMW           -0.0072      0.048     -0.151      0.880      -0.102       0.087
CMA           -0.2748      0.067     -4.085      0.000      -0.407      -0.142
RF             0.7929      0.519      1.527      0.128      -0.229       1.815
==============================================================================
Omnibus:                       20.602   Durbin-Watson:                   2.095
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               26.513
Skew:                           0.530   Prob(JB):                     1.75e-06
Kurtosis:                       3.999   Cond. No.                         510.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

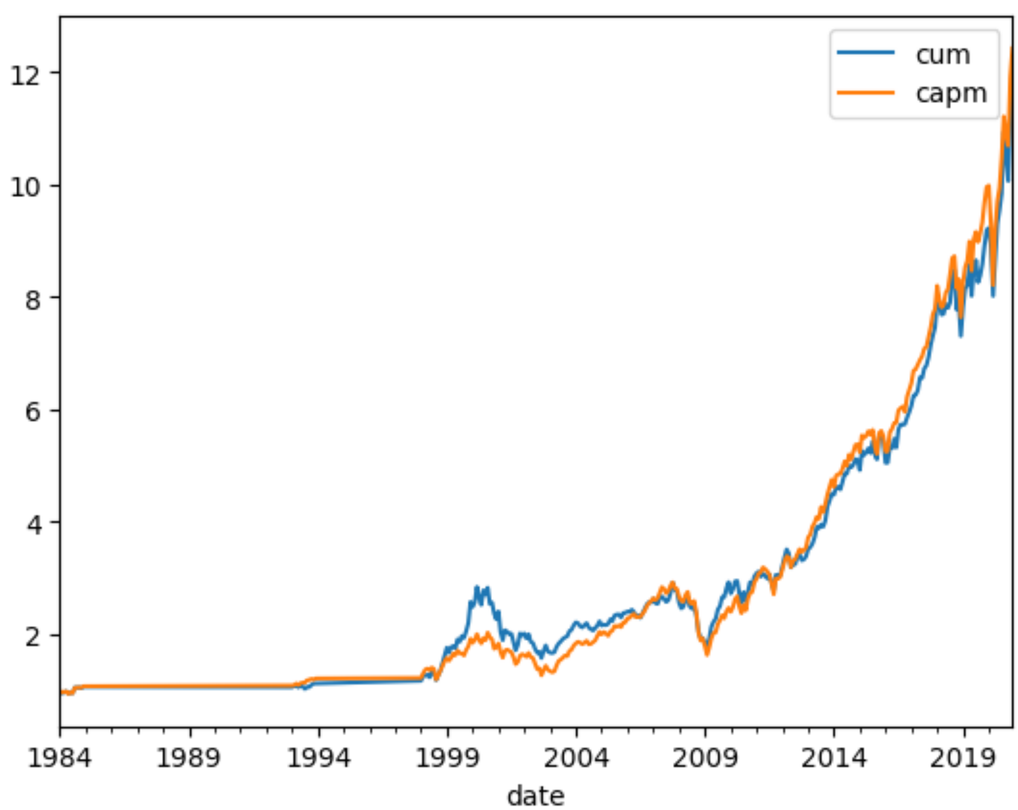
The alpha is positive and significant in some models, but not in others.

4 2c

[ ]:
```python
import matplotlib.pyplot as plt


def plot_cumulative_returns(portfolio_returns, capm_model, ff3):
```

```
    expected_returns = capm_model.predict(add_constant(ff3['Mkt-RF'].
 ↪loc[portfolio_returns.index])) + ff3['RF'].loc[portfolio_returns.index]
    cum_portfolio_returns = (1 + portfolio_returns).cumprod()
    cum_portfolio_returns.plot(label="cum")
    cum_expected_returns = (1 + expected_returns).cumprod()
    cum_expected_returns.plot(label="capm")
    plt.legend()

    # plt.figure(figsize=(12, 6))
    # plt.plot(cum_portfolio_returns, label='Value-weighted Portfolio')
    # plt.plot(cum_expected_returns, label='CAPM-Implied Expected Portfolio␣
 ↪Returns')
    # plt.xlabel('Year')
    # plt.ylabel('Cumulative Returns')
    # plt.legend()
    # plt.show()

capm_model_value_weighted, _, _, _ =␣
 ↪estimate_models(monthly_returns['vw_returns'], ff3, ff5)
plot_cumulative_returns(monthly_returns['vw_returns'],␣
 ↪capm_model_value_weighted, ff3)
```

```
Series([], Freq: M, dtype: float64)
```



In the early years, the model outperforms the benchmark pretty well, but it gets arbitraged away over time pretty substantially and then the model performs as well as CAPM.

5 2d

[ ]:
```
def estimate_carhart_subsamples(portfolio_returns, ff3, ff5, date_split):
    pre_returns = portfolio_returns.loc[:date_split]
    post_returns = portfolio_returns.loc[date_split:]
    pre_carhart_model = estimate_models(pre_returns, ff3[:date_split], ff5[:
↪date_split])[2]
    post_carhart_model = estimate_models(post_returns, ff3[date_split:],␣
↪ff5[date_split:])[2]

    print("Carhart Model (Pre-January 1st, 2010):")
    print(pre_carhart_model.summary())
    print("Carhart Model (Post-January 1st, 2010):")
    print(post_carhart_model.summary())

estimate_carhart_subsamples(monthly_returns['vw_returns'], ff3, ff5,␣
↪'2010-01-01')
```

```
Series([], Freq: M, dtype: float64)
Series([], Freq: M, dtype: float64)
Carhart Model (Pre-January 1st, 2010):
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.847
Model:                            OLS   Adj. R-squared:                  0.841
Method:                 Least Squares   F-statistic:                     149.1
Date:                Mon, 01 May 2023   Prob (F-statistic):           2.68e-63
Time:                        17:46:43   Log-Likelihood:                 428.58
No. Observations:                 169   AIC:                            -843.2
Df Residuals:                     162   BIC:                            -821.3
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0043      0.003      1.574      0.118      -0.001       0.010
Mkt-RF         0.8684      0.042     20.753      0.000       0.786       0.951
SMB           -0.0773      0.052     -1.481      0.141      -0.180       0.026
HML           -0.2936      0.064     -4.584      0.000      -0.420      -0.167
RF             0.3047      0.756      0.403      0.687      -1.187       1.797
RMW            0.0525      0.067      0.779      0.437      -0.081       0.186
```

```
CMA            -0.1810      0.092     -1.969      0.051      -0.363       0.001
==============================================================================
Omnibus:                       14.650   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               15.967
Skew:                           0.670   Prob(JB):                     0.000341
Kurtosis:                       3.685   Cond. No.                         502.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
Carhart Model (Post-January 1st, 2010):
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.879
Model:                            OLS   Adj. R-squared:                  0.873
Method:                 Least Squares   F-statistic:                     151.0
Date:                Mon, 01 May 2023   Prob (F-statistic):           8.34e-55
Time:                        17:46:43   Log-Likelihood:                 377.12
No. Observations:                 132   AIC:                            -740.2
Df Residuals:                     125   BIC:                            -720.1
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0005      0.002      0.317      0.752      -0.003       0.004
Mkt-RF         0.8908      0.033     26.985      0.000       0.825       0.956
SMB           -0.1751      0.063     -2.782      0.006      -0.300      -0.051
HML            0.0340      0.059      0.579      0.564      -0.082       0.150
RF             0.0883      1.991      0.044      0.965      -3.852       4.029
RMW           -0.0463      0.089     -0.522      0.603      -0.222       0.129
CMA           -0.4463      0.102     -4.391      0.000      -0.647      -0.245
==============================================================================
Omnibus:                        1.985   Durbin-Watson:                   2.461
Prob(Omnibus):                  0.371   Jarque-Bera (JB):                1.673
Skew:                           0.024   Prob(JB):                        0.433
Kurtosis:                       3.550   Cond. No.                     1.60e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.6e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

i think barring some problems with our data, in the pre-period it should be matching, whereas in the after period it wouldn't match given the strategy has probably been priced in or arbitraged away. We also notice in our plot that it is not priced in as the model outperforms CAPM in early years, but over time the info is priced in so it starts performing according to CAPM.

6 2e

[ ]:
```
ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start='1984-01',␣
↪end='2020-12')[0]
ff12 = ff12 / 100
# ff12.index = ff12.index.to_timestamp('M')
```

[ ]:
```
ff12 = ff12.loc[monthly_returns.index, :]
```

[ ]:
```
X_vw_ff12_post = add_constant(ff12.loc[monthly_returns[monthly_returns.index.
↪year >= 1999].index])
X_vw_ff12_pre = add_constant(ff12.loc[monthly_returns[monthly_returns.index.
↪year < 1999].index])

model_post_ff12 = OLS(monthly_returns[monthly_returns.index.year >=␣
↪1999]['ew_returns'] * 10, X_vw_ff12_post).fit()
model_pre_ff12 = OLS(monthly_returns[monthly_returns.index.year <␣
↪1999]['vw_returns'] * 10, X_vw_ff12_pre).fit()
```

[ ]:
```
model_post_ff12.summary()
```

[ ]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:             ew_returns   R-squared:                       0.811
Model:                            OLS   Adj. R-squared:                  0.802
Method:                 Least Squares   F-statistic:                     89.97
Date:                Mon, 01 May 2023   Prob (F-statistic):           1.27e-83
Time:                        17:47:28   Log-Likelihood:                 20.550
No. Observations:                 264   AIC:                            -15.10
Df Residuals:                     251   BIC:                             31.39
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0347      0.015      2.333      0.020       0.005       0.064
NoDur          0.9018      0.735      1.227      0.221      -0.545       2.349
Durbl          1.3390      0.312      4.287      0.000       0.724       1.954
Manuf          0.9043      0.772      1.171      0.243      -0.617       2.425
Enrgy          1.3187      0.302      4.373      0.000       0.725       1.913
Chems         -2.5242      0.702     -3.598      0.000      -3.906      -1.142
```

```
BusEq          2.8688      0.358      8.016      0.000       2.164       3.574
Telcm          0.2417      0.464      0.521      0.603      -0.672       1.156
Utils         -0.5545      0.448     -1.239      0.217      -1.436       0.327
Shops         -0.5619      0.598     -0.940      0.348      -1.739       0.616
Hlth           0.3919      0.483      0.812      0.418      -0.559       1.343
Money          1.0510      0.496      2.119      0.035       0.074       2.028
Other          2.0877      0.890      2.347      0.020       0.336       3.840
==============================================================================
Omnibus:                       53.046   Durbin-Watson:                   1.922
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              284.558
Skew:                           0.651   Prob(JB):                     1.62e-62
Kurtosis:                       7.917   Cond. No.                         72.3
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

[ ]:
```python
model_pre_ff12.summary()
```

[ ]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:             vw_returns   R-squared:                       0.933
Model:                            OLS   Adj. R-squared:                  0.898
Method:                 Least Squares   F-statistic:                     26.64
Date:                Mon, 01 May 2023   Prob (F-statistic):           1.22e-10
Time:                        17:47:28   Log-Likelihood:                 27.656
No. Observations:                  36   AIC:                            -29.31
Df Residuals:                      23   BIC:                            -8.726
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0309      0.031      0.993      0.331      -0.034       0.095
NoDur          1.7189      1.373      1.252      0.223      -1.122       4.560
Durbl         -1.2751      0.918     -1.390      0.178      -3.173       0.623
Manuf         -2.0323      2.888     -0.704      0.489      -8.007       3.942
Enrgy         -0.0590      0.742     -0.079      0.937      -1.594       1.477
Chems          3.4731      2.136      1.626      0.118      -0.945       7.891
BusEq          5.7620      1.035      5.566      0.000       3.620       7.904
Telcm          0.6741      0.969      0.696      0.494      -1.330       2.679
Utils         -0.4815      1.043     -0.462      0.649      -2.640       1.677
Shops          0.2955      1.630      0.181      0.858      -3.077       3.668
```

```
Hlth           1.0389      0.930      1.117      0.276     -0.885       2.963
Money          0.8806      1.184      0.744      0.465     -1.569       3.330
Other         -2.5134      2.081     -1.208      0.239     -6.818       1.791
==============================================================================
Omnibus:                        1.228   Durbin-Watson:                   1.941
Prob(Omnibus):                  0.541   Jarque-Bera (JB):                1.214
Skew:                           0.378   Prob(JB):                        0.545
Kurtosis:                       2.513   Cond. No.                         168.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

Weirdness with data aside, It has changed as the coefficients are changing pretty substantially between the two.

Question 3

A

The beta we see is close to 1. This is most likely due to the fact that if employees are dissatisfied, they are probably leaving the comapny, which means the company is not growing as much or might be shrinking. It could also be an early indication of a company doing poorly, or the result of that. On the other hand, if the employees are satisfied, they are probably staying, more employees are joining, the company is probably doing well or the employees are satisfied due to the company recently doing well. All of these factors compound to mean that highly rated companies are probably going to be larger companies, which dominate the market, thus meaning that our portfolio would capture very similar risk to the market portfolio. The employee behavior would very closely resemble investor behavior.

If you long-short by shorting the overall market, that means the institutional investors can get exposure to just the alpha generated by the employee satisfaction rating, rather than being at the whim of overall market trends. This might be less attractive to retail investors since they might want to capture these large market trends and ride off of them in a bull-market like we've seen recently. If we short the overall market, that means the retail investors have less alpha in a bull market, so they would be upset. Institutional investors really want nuanced alpha that protects them from overall trends and capitalizes off of information.

B

Theoretically speaking, if the markets fully priced in the value of employee satisfaction, then we would gain market returns, but not have a significant alpha. We see a discrepancy in the returns between the CAPM predictions and the actual cumulative returns, where ours is producing much more alpha than CAPM predicts up to a point where it is arbitraged away. Thus, we see that the information has not been priced in historically, but in recent years the alpha has gone away and it has been priced in.

C

In the same way that neural networks used to be really good at producing alpha for hedge funds like Renaissance, this strategy was probably popularized and as more people used it, the alpha was arbitraged away. It could also be that people are just dissatisfied across the board because of poor wages or working conditions as we see a lot in politics and the news. But it's probably the former. We see in the cumulative returns that they stagnate in recent years, i.e. the alpha has been arbitraged away, or the information is no longer important.

D

You can use the cookie data or other data to get info on whether or not employees are currently looking at other jobs, and use that to get a measure of how many employees are looking to jump ship. You could also use their cookies to probably find out why they are dissatisfied, since a lot of people are on social media and talking about their qualms. You could also use this data to look at the purchases employees are making, as if employees are making big luxury purchases, it probably means the company is doing well or is about to do very well. If they are searching up the price of cup ramen, the company is probably not doing well.