Problem5

April 23, 2023

[ ]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from datetime import datetime
from tqdm import tqdm
from tqdm.contrib.concurrent import process_map
from tqdm.contrib import tmap

# Enable tqdm for Pandas
tqdm.pandas()
```

```
/Users/esmirmesic/opt/anaconda3/envs/bem114/lib/python3.11/site-
packages/tqdm/auto.py:21: TqdmWarning: IProgress not found. Please update
jupyter and ipywidgets. See
https://ipywidgets.readthedocs.io/en/stable/user_install.html
  from .autonotebook import tqdm as notebook_tqdm
```

1 A

[ ]:
```
crsp_data = pd.read_csv("data/cleaned_crsp.csv")
crsp_data['date'] = pd.to_datetime(crsp_data['date'])
crsp_data['RET'] = crsp_data['RET'].str.replace('C', '')
crsp_data['RET'] = pd.to_numeric(crsp_data['RET'], errors='coerce')
crsp_data['date'] = pd.to_datetime(crsp_data['date'], format='%Y-%m-%d')
```

[ ]:
```
import pandas_datareader as pdr

start_date = '1926-01-01'
end_date = '2020-12-31'


ff5_factors = pdr.get_data_famafrench('F-F_Research_Data_5_Factors_2x3',␣
 ↪start=start_date, end=end_date)[0]
ff5_factors = ff5_factors / 100  # Convert to decimal
ff5_factors.index = ff5_factors.index.to_timestamp('M')  # Convert index to␣
 ↪monthly-end dates
```

```python
ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,␣
↪end=end_date)[0]
ff12 = ff12 / 100
ff12.index = ff12.index.to_timestamp('M')
```

```
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:7
: FutureWarning: The argument 'date_parser' is deprecated and will be removed in
a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff5_factors = pdr.get_data_famafrench('F-F_Research_Data_5_Factors_2x3',
start=start_date, end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:7
: FutureWarning: The argument 'date_parser' is deprecated and will be removed in
a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff5_factors = pdr.get_data_famafrench('F-F_Research_Data_5_Factors_2x3',
start=start_date, end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
```

```
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
/var/folders/sg/4dp480wd1cjd288xvby34rpr0000gn/T/ipykernel_12833/3641822038.py:1
1: FutureWarning: The argument 'date_parser' is deprecated and will be removed
in a future version. Please use 'date_format' instead, or read your data in as
'object' dtype and then call 'to_datetime'.
  ff12 = pdr.get_data_famafrench('12_industry_Portfolios', start=start_date,
end=end_date)[0]
```

[ ]:

```
import yfinance as yf

top10_holdings = pd.read_csv("data/top10_holdings_brk_arkk.csv")

# Get BRK-A and ARKK data from Yahoo Finance
brk = yf.download("BRK-A", start="1980-01-31", end="2020-12-31", interval="1mo")
brk.index = pd.to_datetime(brk.index)
arkk = yf.download("ARKK", start="2014-10-31", end="2020-12-31", interval="1mo")
arkk.index = pd.to_datetime(arkk.index)

# Make sure the index is datetime
brk.index = pd.to_datetime(brk.index)
arkk.index = pd.to_datetime(arkk.index)

brk.index = brk.index.to_period("M").to_timestamp("M")
arkk.index = arkk.index.to_period("M").to_timestamp("M")

# Calculate monthly stock returns
brk["Return"] = brk["Adj Close"].pct_change()
arkk["Return"] = arkk["Adj Close"].pct_change()

brk = brk.dropna()
arkk = arkk.dropna()

# Estimate the FF5 model for each strategy over their full histories and the␣
 ↪same sample period
# Merge data
brk_ff5 = pd.merge(brk, ff5_factors, left_index=True, right_index=True)
```

```
arkk_ff5 = pd.merge(arkk, ff5_factors, left_index=True, right_index=True)

# Find the common time period for both stocks
start_date = max(brk_ff5.index.min(), arkk_ff5.index.min())
end_date = min(brk_ff5.index.max(), arkk_ff5.index.max())

# Create the same sample period data
brk_ff5_same_period = brk_ff5.loc[start_date:end_date]
arkk_ff5_same_period = arkk_ff5.loc[start_date:end_date]

# Perform regressions for the same sample period
X_brk_same_period = sm.add_constant(brk_ff5_same_period[["Mkt-RF", "SMB",␣
 ↪"HML", "RMW", "CMA"]])
X_arkk_same_period = sm.add_constant(arkk_ff5_same_period[["Mkt-RF", "SMB",␣
 ↪"HML", "RMW", "CMA"]])

model_brk_same_period = sm.OLS(brk_ff5_same_period["Return"],␣
 ↪X_brk_same_period).fit()
model_arkk_same_period = sm.OLS(arkk_ff5_same_period["Return"],␣
 ↪X_arkk_same_period).fit()

# Perform regressions
X_brk = sm.add_constant(brk_ff5[["Mkt-RF", "SMB", "HML", "RMW", "CMA"]])
X_arkk = sm.add_constant(arkk_ff5[["Mkt-RF", "SMB", "HML", "RMW", "CMA"]])

model_brk = sm.OLS(brk_ff5["Return"], X_brk).fit()
model_arkk = sm.OLS(arkk_ff5["Return"], X_arkk).fit()

# Regress returns for each strategy on the Fama French 12 Industry Portfolios
X_brk_ff12 = sm.add_constant(ff12.loc[brk_ff5.index])
X_arkk_ff12 = sm.add_constant(ff12.loc[arkk_ff5.index])

model_brk_ff12 = sm.OLS(brk_ff5["Return"], X_brk_ff12).fit()
model_arkk_ff12 = sm.OLS(arkk_ff5["Return"], X_arkk_ff12).fit()
```

```
[**********************100%**********************]  1 of 1 completed
[**********************100%**********************]  1 of 1 completed
```

[ ]:
```
model_arkk.summary(), model_brk.summary()
```

[ ]:
```
(<class 'statsmodels.iolib.summary.Summary'>
 """
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Return   R-squared:                       0.814
Model:                            OLS   Adj. R-squared:                  0.800
Method:                 Least Squares   F-statistic:                     58.61
```

```
Date:                Sun, 23 Apr 2023   Prob (F-statistic):           3.92e-23
Time:                        20:34:29   Log-Likelihood:                136.38
No. Observations:                73   AIC:                             -260.8
Df Residuals:                    67   BIC:                             -247.0
Df Model:                         5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0062      0.005      1.266      0.210      -0.004       0.016
Mkt-RF         1.5209      0.120     12.672      0.000       1.281       1.760
SMB            0.5290      0.212      2.493      0.015       0.105       0.953
HML           -0.7020      0.194     -3.619      0.001      -1.089      -0.315
RMW           -0.1581      0.341     -0.464      0.644      -0.839       0.523
CMA           -0.7889      0.348     -2.265      0.027      -1.484      -0.094
==============================================================================
Omnibus:                       10.376   Durbin-Watson:                   2.246
Prob(Omnibus):                  0.006   Jarque-Bera (JB):               10.269
Skew:                           0.839   Prob(JB):                      0.00589
Kurtosis:                       3.750   Cond. No.                         80.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
""",
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Return   R-squared:                       0.341
Model:                            OLS   Adj. R-squared:                  0.333
Method:                 Least Squares   F-statistic:                     43.96
Date:                Sun, 23 Apr 2023   Prob (F-statistic):           1.60e-36
Time:                        20:34:29   Log-Likelihood:                670.86
No. Observations:                 431   AIC:                            -1330.
Df Residuals:                     425   BIC:                            -1305.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0074      0.003      2.824      0.005       0.002       0.013
Mkt-RF         0.8224      0.063     13.044      0.000       0.698       0.946
SMB           -0.3503      0.094     -3.745      0.000      -0.534      -0.166
HML            0.4275      0.114      3.745      0.000       0.203       0.652
RMW            0.3473      0.123      2.832      0.005       0.106       0.588
```

```
CMA           -0.0129      0.176     -0.073      0.942      -0.358       0.333
==============================================================================
Omnibus:                       94.435   Durbin-Watson:                   2.078
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              212.375
Skew:                           1.125   Prob(JB):                     7.65e-47
Kurtosis:                       5.601   Cond. No.                         79.6
==============================================================================

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

[ ]:
```
model_arkk_ff12.summary(), model_brk_ff12.summary()
```

[ ]:
```
(<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Return   R-squared:                       0.833
Model:                            OLS   Adj. R-squared:                  0.800
Method:                 Least Squares   F-statistic:                     25.02
Date:                Sun, 23 Apr 2023   Prob (F-statistic):           5.99e-19
Time:                        20:34:29   Log-Likelihood:                 140.43
No. Observations:                  73   AIC:                            -254.9
Df Residuals:                      60   BIC:                            -225.1
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0052      0.005      0.977      0.332      -0.005       0.016
NoDur          0.0154      0.262      0.059      0.953      -0.510       0.540
Durbl          0.3145      0.094      3.360      0.001       0.127       0.502
Manuf          0.4304      0.295      1.461      0.149      -0.159       1.020
Enrgy          0.0141      0.094      0.149      0.882      -0.174       0.203
Chems         -0.3705      0.272     -1.364      0.178      -0.914       0.173
BusEq          0.9235      0.194      4.750      0.000       0.535       1.312
Telcm         -0.2029      0.217     -0.936      0.353      -0.637       0.231
Utils         -0.0491      0.169     -0.290      0.773      -0.387       0.289
Shops          0.0344      0.231      0.149      0.882      -0.428       0.497
Hlth           0.5693      0.179      3.179      0.002       0.211       0.927
Money         -0.4316      0.208     -2.071      0.043      -0.849      -0.015
Other          0.0753      0.419      0.180      0.858      -0.763       0.914
==============================================================================
Omnibus:                        1.505   Durbin-Watson:                   2.602
Prob(Omnibus):                  0.471   Jarque-Bera (JB):                1.006
```

```
Skew:                          0.274   Prob(JB):                     0.605
Kurtosis:                      3.173   Cond. No.                      107.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
""",
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Return   R-squared:                       0.402
Model:                            OLS   Adj. R-squared:                  0.385
Method:                 Least Squares   F-statistic:                     23.45
Date:                Sun, 23 Apr 2023   Prob (F-statistic):           7.44e-40
Time:                        20:34:29   Log-Likelihood:                691.94
No. Observations:                 431   AIC:                           -1358.
Df Residuals:                     418   BIC:                           -1305.
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0066      0.003      2.602      0.010       0.002       0.011
NoDur          0.3477      0.123      2.833      0.005       0.106       0.589
Durbl         -0.0154      0.059     -0.260      0.795      -0.132       0.101
Manuf         -0.1576      0.143     -1.104      0.270      -0.438       0.123
Enrgy         -0.0013      0.055     -0.024      0.981      -0.109       0.107
Chems          0.0557      0.119      0.468      0.640      -0.178       0.290
BusEq         -0.2874      0.062     -4.634      0.000      -0.409      -0.165
Telcm          0.0550      0.076      0.722      0.470      -0.095       0.205
Utils          0.0322      0.080      0.404      0.686      -0.124       0.189
Shops          0.1805      0.103      1.755      0.080      -0.022       0.383
Hlth          -0.0022      0.083     -0.026      0.979      -0.165       0.161
Money          0.3294      0.087      3.803      0.000       0.159       0.500
Other          0.3441      0.149      2.303      0.022       0.050       0.638
==============================================================================
Omnibus:                       95.333   Durbin-Watson:                   2.075
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              245.090
Skew:                           1.080   Prob(JB):                     6.02e-54
Kurtosis:                       5.998   Cond. No.                         75.0
==============================================================================

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

```
"""")
```

[ ]:
```
model_arkk_same_period.summary(), model_brk_same_period.summary()
```

[ ]:
```
(<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Return   R-squared:                       0.814
Model:                            OLS   Adj. R-squared:                  0.800
Method:                 Least Squares   F-statistic:                     58.61
Date:                Sun, 23 Apr 2023   Prob (F-statistic):           3.92e-23
Time:                        20:34:26   Log-Likelihood:                136.38
No. Observations:                  73   AIC:                            -260.8
Df Residuals:                      67   BIC:                            -247.0
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0062      0.005      1.266      0.210      -0.004       0.016
Mkt-RF         1.5209      0.120     12.672      0.000       1.281       1.760
SMB            0.5290      0.212      2.493      0.015       0.105       0.953
HML           -0.7020      0.194     -3.619      0.001      -1.089      -0.315
RMW           -0.1581      0.341     -0.464      0.644      -0.839       0.523
CMA           -0.7889      0.348     -2.265      0.027      -1.484      -0.094
==============================================================================
Omnibus:                       10.376   Durbin-Watson:                   2.246
Prob(Omnibus):                  0.006   Jarque-Bera (JB):               10.269
Skew:                           0.839   Prob(JB):                      0.00589
Kurtosis:                       3.750   Cond. No.                         80.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
""",
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Return   R-squared:                       0.703
Model:                            OLS   Adj. R-squared:                  0.681
Method:                 Least Squares   F-statistic:                     31.68
Date:                Sun, 23 Apr 2023   Prob (F-statistic):           2.07e-16
Time:                        20:34:26   Log-Likelihood:                163.59
No. Observations:                  73   AIC:                            -315.2
```

```
Df Residuals:                       67   BIC:                            -301.4
Df Model:                            5
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          0.0010      0.003      0.293      0.770      -0.006       0.008
Mkt-RF         0.9092      0.083     10.997      0.000       0.744       1.074
SMB           -0.4787      0.146     -3.275      0.002      -0.771      -0.187
HML            0.3490      0.134      2.612      0.011       0.082       0.616
RMW            0.0164      0.235      0.070      0.945      -0.453       0.485
CMA            0.3688      0.240      1.537      0.129      -0.110       0.848
==============================================================================
Omnibus:                        0.446   Durbin-Watson:                   2.159
Prob(Omnibus):                  0.800   Jarque-Bera (JB):                0.532
Skew:                          -0.174   Prob(JB):                        0.766
Kurtosis:                       2.769   Cond. No.                         80.9
==============================================================================

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

2 B.

Cathie Wood and Warren Buffet do not appear to have similar investment strategies (during the period in which we have data from both). Based on the OLS regression over the FF12 data, the coefficients across their two models vary drastically, indicating that their models are different.

Warren Buffett is more like a value investor due to his positive and statistically significant HML coefficient (both historically and recently, although he has been acting less like a value investor in recent periods, as indicated by his decline in HML). In contrast, Cathie Wood has a negative and statistically significant HML coefficient, indicating that she is acting more like a growth investor than a value investor.

Warren Buffett's portfolio behaves closest to Consumer Nondurables (NoDur), Shops, Banking Sector (Money), and (Other). Cathie Wood's portfolio behaves closest to Consumer Durables (Durble), Manufacturing (Manuf), Business Equipment (BusEq), Health (Hlth).

Both Buffett and Wood focus on consumer goods (although different types), and Buffett focuses on Banking. The top 10 holdings focus on banking and consumer goods, so the portfolio behavior analysis tracks.