Master Thesis - Final Working Code

June 27, 2021

[10]:
```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
from IPython.display import Image
from scipy import linalg
from datetime import datetime, timedelta
import time
from numpy.linalg import inv
import seaborn as sns
from matplotlib.ticker import StrMethodFormatter
from statsmodels.formula.api import logit
from scipy.stats import chi2_contingency

from statsmodels.stats.diagnostic import het_breuschpagan
from statsmodels.stats.diagnostic import het_white

from statsmodels.formula.api import ols
%matplotlib inline
```

[11]:
```
totaldataframe = pd.read_excel('Total-Dataset.xlsx')
totaldataframe.head()
```

[11]:
```
   FUND ID  FIRM ID                   NAME     ASSET CLASS  Vintage year  \
0    18207    29928       Vendis Capital I  Private Equity          2009
1    56055    38771                  JET I  Private Equity          2015
2    28073    70067      Capidea Kapital I  Private Equity          2006
3    63397    25679  Via equity Fond III K/S  Private Equity          2016
4    13775    20804         Erhvervsinvest  Private Equity          2004

   Fund size  Net IRR  Net multiple   RVPI     DPI  Called  Quartile rank  \
0     145.16     26.7          3.74   59.0  315.00   98.20              1
1     142.83     11.0          1.47   52.6   94.50   93.60              2
2     128.47     10.0          1.50    0.0  150.00   85.00              3
3     145.57     16.1          1.36  136.2    0.00   91.50              2
4      45.55     21.8          2.52    0.0  251.71   93.51              2

  DATE REPORTED      STATUS    FUND MANAGER        COUNTRY  Fund number  \
```

```
0    2018-12-31      Closed  Vendis Capital         Belgium             1
1    2020-03-31      Closed  Jet Investment  Czech Republic             1
2    2021-03-31  Liquidated         Capidea         Denmark             1
3    2020-12-31      Closed      Via Equity         Denmark             1
4    2021-03-31  Liquidated  Erhvervsinvest         Denmark             1

  Region focus                      Core Industries Geographic focus
0       Europe                          Diversified           Europe
1       Europe                          Diversified               UK
2       Europe                          Diversified            Italy
3       Europe  Healthcare, Information Technology           Europe
4       Europe                          Diversified      West Europe
```

[12]:
```
workingdataframe = totaldataframe.drop(['FUND ID','FIRM ID', 'NAME', 'ASSET␣
 ↪CLASS', 'DATE REPORTED',
                         'STATUS', 'Geographic focus', 'FUND MANAGER',␣
 ↪'COUNTRY', 'Region focus',
                         'Core Industries'], axis = 1)

workingdataframe
```

[12]:
```
     Vintage year  Fund size  Net IRR  Net multiple   RVPI     DPI  Called  \
0            2009     145.16     26.7          3.74   59.0  315.00   98.20
1            2015     142.83     11.0          1.47   52.6   94.50   93.60
2            2006     128.47     10.0          1.50    0.0  150.00   85.00
3            2016     145.57     16.1          1.36  136.2    0.00   91.50
4            2004      45.55     21.8          2.52    0.0  251.71   93.51
..            …          …        …          …       …     …       …
412          2007   17708.40      8.0          1.50    0.0  150.00   99.00
413          2006     856.30     13.2          1.97    6.0  191.00   96.00
414          2012    7517.00     16.0          1.81   71.0  110.00  100.00
415          2015     504.00     13.8          1.41  126.0   15.00   59.00
416          2016    9481.00     24.8          1.50  147.0    3.00   92.00

     Quartile rank  Fund number
0                1            1
1                2            1
2                3            1
3                2            1
4                2            1
..             …          …
412              3           29
413              1           30
414              2           31
415              3           32
416              1           33
```

```
[417 rows x 9 columns]
```

[13]:
```
# Create correlation matrix
corrmatrix = workingdataframe.corr()
corrmatrix.style.background_gradient(cmap='coolwarm')
```

[13]:
```
<pandas.io.formats.style.Styler at 0x7feefec36880>
```

[17]:
```
dataframe = totaldataframe.drop(['FUND ID','FIRM ID', 'NAME', 'ASSET CLASS',␣
 ↪'DATE REPORTED',
                         'STATUS', 'Geographic focus', 'FUND MANAGER',␣
 ↪'COUNTRY'], axis = 1)

dataframe
```

[17]:
```
     Vintage year  Fund size  Net IRR  Net multiple   RVPI     DPI  Called  \
0            2009     145.16     26.7          3.74   59.0  315.00   98.20
1            2015     142.83     11.0          1.47   52.6   94.50   93.60
2            2006     128.47     10.0          1.50    0.0  150.00   85.00
3            2016     145.57     16.1          1.36  136.2    0.00   91.50
4            2004      45.55     21.8          2.52    0.0  251.71   93.51
..            …          …        …            …      …       …       …
412          2007   17708.40      8.0          1.50    0.0  150.00   99.00
413          2006     856.30     13.2          1.97    6.0  191.00   96.00
414          2012    7517.00     16.0          1.81   71.0  110.00  100.00
415          2015     504.00     13.8          1.41  126.0   15.00   59.00
416          2016    9481.00     24.8          1.50  147.0    3.00   92.00

     Quartile rank  Fund number          Region focus  \
0                1            1                Europe
1                2            1                Europe
2                3            1                Europe
3                2            1                Europe
4                2            1                Europe
..             …            …                   …
412              3           29                Europe
413              1           30         North America
414              2           31                Europe
415              3           32  Middle East & Israel
416              1           33                Europe

                      Core Industries
0                         Diversified
1                         Diversified
2                         Diversified
3    Healthcare, Information Technology
4                         Diversified
```

```
..                               …
412                    Diversified
413                    Diversified
414                    Diversified
415                    Diversified
416         Consumer Discretionary

[417 rows x 11 columns]
```

[18]:
```python
#Renaming variables
dataframe.rename(columns = {'Fund size':'Size','Net IRR':'IRR','Net multiple':
 ↪'TVPI',
                            'Called':'Capital called'},inplace=True)
dataframe.head()
```

[18]:
```
   Vintage year    Size   IRR  TVPI   RVPI     DPI  Capital called  \
0          2009  145.16  26.7  3.74   59.0  315.00           98.20
1          2015  142.83  11.0  1.47   52.6   94.50           93.60
2          2006  128.47  10.0  1.50    0.0  150.00           85.00
3          2016  145.57  16.1  1.36  136.2    0.00           91.50
4          2004   45.55  21.8  2.52    0.0  251.71           93.51

   Quartile rank  Fund number Region focus                          Core Industries
0              1            1       Europe                              Diversified
1              2            1       Europe                              Diversified
2              3            1       Europe                              Diversified
3              2            1       Europe  Healthcare, Information Technology
4              2            1       Europe                              Diversified
```

[19]:
```python
# All descriptive statistics
descriptive = dataframe.describe()
descriptive
```

[19]:
```
       Vintage year          Size         IRR        TVPI        RVPI  \
count    417.000000    417.000000  417.000000  417.000000  417.000000
mean    2007.724221   1187.904341   17.584149    1.815228   36.528801
std        5.181626   2051.500741   17.365884    0.670897   50.024370
min     1999.000000      8.250000  -14.400000    0.380000    0.000000
25%     2004.000000    193.140000    9.000000    1.410000    0.000000
50%     2007.000000    400.590000   15.080000    1.680000    9.000000
75%     2012.000000   1128.150000   23.000000    2.120000   67.000000
max     2016.000000  17708.400000  239.800000    5.820000  329.410000

              DPI  Capital called  Quartile rank  Fund number
count  417.000000      417.000000     417.000000   417.000000
mean   144.990983       93.897674       2.326139     4.928058
std     87.426320       12.913989       1.053434     6.889555
```

```
min       0.000000      24.800000      1.000000     1.000000
25%      87.000000      88.810000      1.000000     2.000000
50%     146.000000      96.000000      2.000000     3.000000
75%     193.290000     100.000000      3.000000     5.000000
max     580.000000     140.840000      4.000000    57.000000
```

[73]:
```
#Exporting table to Excel
file_name = 'Descriptive.xlsx'
descriptive.to_excel(file_name)
```

[20]:
```
#Figure X: Average by quartile rank
groupby_quartile = dataframe.groupby('Quartile rank')['IRR', 'TVPI', 'Size']
groupby_quartile.mean()
```

```
<ipython-input-20-73dba711cb18>:2: FutureWarning: Indexing with multiple keys
(implicitly converted to a tuple of keys) will be deprecated, use a list
instead.
  groupby_quartile = dataframe.groupby('Quartile rank')['IRR', 'TVPI', 'Size']
```

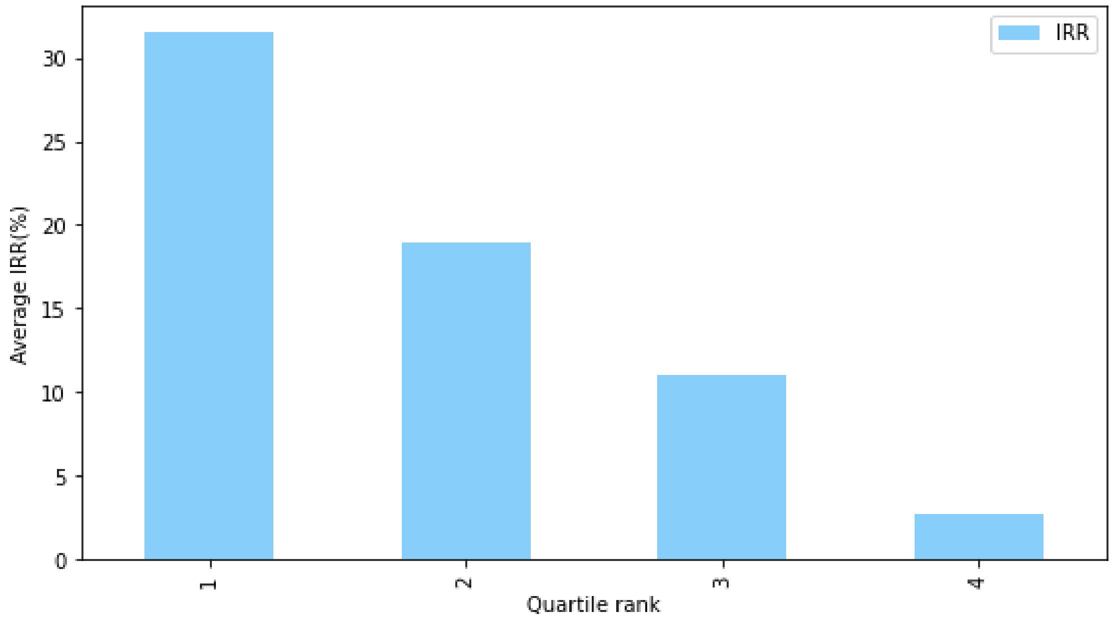
[20]:

	IRR	TVPI	Size
Quartile rank			
1	31.582143	2.497411	1025.612500
2	18.963846	1.864615	1338.392769
3	11.059804	1.488529	1361.554608
4	2.766986	1.137123	926.272329

[21]:
```
# Figure 7: Average Net IRR by quartile rank, total sample A

dataframe.groupby('Quartile rank').mean()['IRR'].plot.bar(legend=True,␣
 ↪figsize=(9,5), color='lightskyblue')
plt.xlabel('Quartile rank')
plt.ylabel('Average IRR(%)')
```

[21]:
```
Text(0, 0.5, 'Average IRR(%)')
```



[22]:
```
# Parts of Table 4: Vintage year statistics.
groupby_vintage = dataframe.groupby(['Vintage year'])[['IRR', 'TVPI']].
 ↪agg(['count','mean'])
groupby_vintage
```

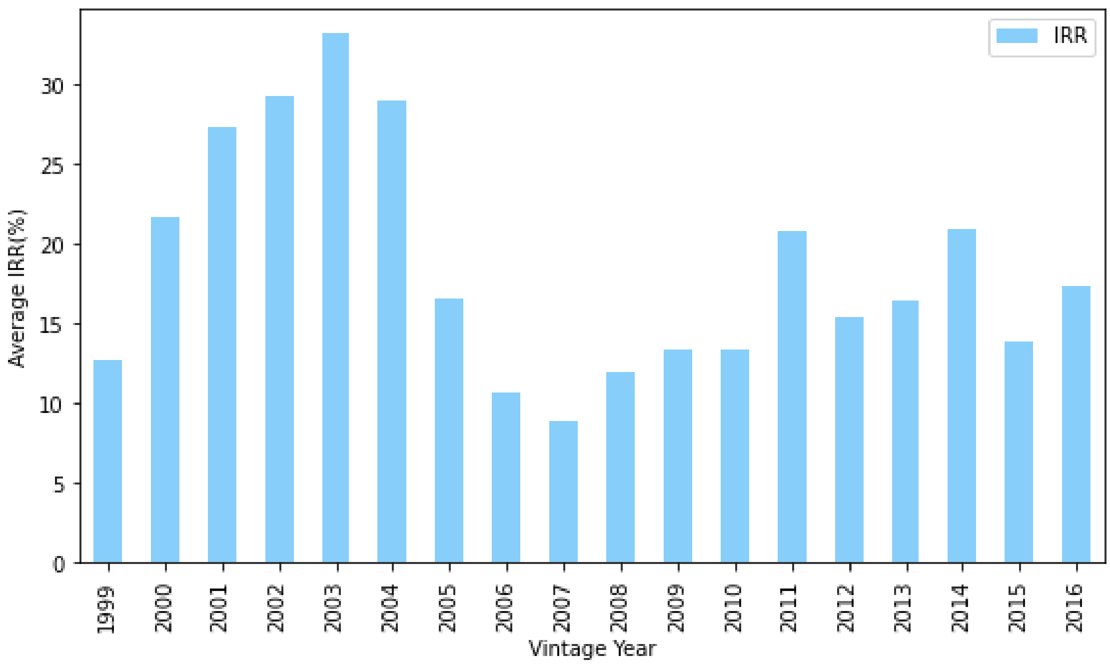
[22]:

	IRR		TVPI	
	count	mean	count	mean
Vintage year				
1999	21	12.773333	21	1.738571
2000	25	21.698000	25	2.029200
2001	17	27.319412	17	2.221176
2002	19	29.259474	19	2.317368
2003	22	33.173636	22	2.071818
2004	14	29.014286	14	2.240000
2005	27	16.550370	27	1.800000
2006	38	10.738684	38	1.743158
2007	34	8.890588	34	1.598235
2008	23	11.947391	23	1.759565
2009	21	13.394762	21	1.852857
2010	14	13.362143	14	1.750000
2011	22	20.786364	22	2.063636
2012	19	15.388421	19	1.877368
2013	19	16.480000	19	1.745263
2014	24	20.985417	24	1.696667
2015	30	13.816333	30	1.434333
2016	28	17.293571	28	1.426429

[23]:
```
# Figure 5. Average Net IRR by vintage year, total sample A

dataframe.groupby('Vintage year')['IRR'].mean().plot.bar(legend=True,
    figsize=(9,5), color='lightskyblue')
plt.xlabel('Vintage Year')
plt.ylabel('Average IRR(%)')
```

[23]:
```
Text(0, 0.5, 'Average IRR(%)')
```



[24]:
```
# Creating Dummy variables
dataframe['Industrial diversification'] = dataframe['Core Industries'].
    apply(lambda x: '1' if x == 'Diversified' else '0')
dataframe['Geographic focus'] = dataframe['Region focus'].apply(lambda x: '1'
    if x == 'Europe' else '0')
dataframe['Vintage year 06-08'] = dataframe['Vintage year'].apply(lambda x: 1
    if (x >=2006) & (x <= 2008) else 0)

dataframe
```

[24]:
```
   Vintage year    Size   IRR  TVPI   RVPI     DPI  Capital called  \
0          2009  145.16  26.7  3.74   59.0  315.00           98.20
1          2015  142.83  11.0  1.47   52.6   94.50           93.60
2          2006  128.47  10.0  1.50    0.0  150.00           85.00
3          2016  145.57  16.1  1.36  136.2    0.00           91.50
4          2004   45.55  21.8  2.52    0.0  251.71           93.51
```

```
..           …       …    …     …     …      …            …
412         2007  17708.40   8.0  1.50    0.0  150.00          99.00
413         2006    856.30  13.2  1.97    6.0  191.00          96.00
414         2012   7517.00  16.0  1.81   71.0  110.00         100.00
415         2015    504.00  13.8  1.41  126.0   15.00          59.00
416         2016   9481.00  24.8  1.50  147.0    3.00          92.00

     Quartile rank  Fund number          Region focus  \
0                1            1                Europe
1                2            1                Europe
2                3            1                Europe
3                2            1                Europe
4                2            1                Europe
..             …            …                     …
412              3           29                Europe
413              1           30         North America
414              2           31                Europe
415              3           32  Middle East & Israel
416              1           33                Europe

                        Core Industries Industrial diversification  \
0                           Diversified                          1
1                           Diversified                          1
2                           Diversified                          1
3    Healthcare, Information Technology                          0
4                           Diversified                          1
..                                  …                       …
412                         Diversified                          1
413                         Diversified                          1
414                         Diversified                          1
415                         Diversified                          1
416              Consumer Discretionary                          0

    Geographic focus  Vintage year 06-08
0                  1                   0
1                  1                   0
2                  1                   1
3                  1                   0
4                  1                   0
..               …                 …
412                1                   1
413                0                   1
414                1                   0
415                0                   0
416                1                   0

[417 rows x 14 columns]
```

[25]:
```
# Logging Fund variable
dataframe['Fund size'] = np.log(dataframe['Size'])
dataframe.head()
```

[25]:
```
   Vintage year    Size   IRR  TVPI   RVPI     DPI  Capital called  \
0          2009  145.16  26.7  3.74   59.0  315.00           98.20
1          2015  142.83  11.0  1.47   52.6   94.50           93.60
2          2006  128.47  10.0  1.50    0.0  150.00           85.00
3          2016  145.57  16.1  1.36  136.2    0.00           91.50
4          2004   45.55  21.8  2.52    0.0  251.71           93.51

   Quartile rank  Fund number Region focus  \
0              1            1       Europe
1              2            1       Europe
2              3            1       Europe
3              2            1       Europe
4              2            1       Europe

                     Core Industries Industrial diversification  \
0                        Diversified                          1
1                        Diversified                          1
2                        Diversified                          1
3  Healthcare, Information Technology                          0
4                        Diversified                          1

  Geographic focus  Vintage year 06-08  Fund size
0                1                   0   4.977837
1                1                   0   4.961655
2                1                   1   4.855695
3                1                   0   4.980657
4                1                   0   3.818811
```

[26]:
```
# OLS Regression using entire dataset

y1 = dataframe[['IRR']]
X1 = dataframe[['TVPI', 'Capital called', 'Fund size', 'Fund number',␣
 ↪'Industrial diversification', 'Geographic focus']]
X2 = X1.astype(float)

reg1 = sm.OLS(y1,sm.add_constant(X2), missing='drop')

results1 = reg1.fit()
results1.summary()
```

[26]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:                    IRR   R-squared:                       0.525
Model:                            OLS   Adj. R-squared:                  0.518
Method:                 Least Squares   F-statistic:                     75.60
Date:                Sun, 27 Jun 2021   Prob (F-statistic):           2.82e-63
Time:                        10:45:40   Log-Likelihood:                -1626.2
No. Observations:                 417   AIC:                             3266.
Df Residuals:                     410   BIC:                             3295.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
==============
                                 coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
const                         -0.2392      5.862     -0.041      0.967
-11.762      11.283
TVPI                          18.2634      0.898     20.327      0.000
16.497      20.030
Capital called                -0.1262      0.047     -2.694      0.007
-0.218      -0.034
Fund size                     -0.7041      0.470     -1.498      0.135
-1.628       0.220
Fund number                    0.0478      0.095      0.503      0.615
-0.139       0.235
Industrial diversification     0.3506      1.739      0.202      0.840
-3.067       3.768
Geographic focus               0.3199      2.788      0.115      0.909
-5.161       5.801
==============================================================================
Omnibus:                      668.364   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           293721.414
Skew:                           8.714   Prob(JB):                         0.00
Kurtosis:                     131.845   Cond. No.                         964.
==============================================================================

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

[27]:
```
# White test
white_test = het_white(results1.resid,  results1.model.exog)
white_test # p-values in 2nd and 4th row, both below 0.05, indicates␣
 ↪heteroskedasticity
```

[27]: (45.72338168208282,
 0.006908510925284029,
 1.9260940609393213,
 0.005281544092138827)

[28]:
```
# Breusch-Pagan test
Breusch = het_breuschpagan(results1.resid,  results1.model.exog)
Breusch # p-values in 2nd and 4th row, both below 0.05, indicates␣
 ↪heteroskedasticity
```

[28]: (17.31045150117482,
 0.008207519444725777,
 2.9594990837839124,
 0.007710036957622111)

[29]:
```
## STANDARD ROBUST ERROR MODEL, total data sample

y1 = dataframe[['IRR']]
X1 = dataframe[['TVPI', 'Capital called', 'Fund size', 'Fund number',
                'Industrial diversification', 'Geographic focus']]
X2 = X1.astype(float)

reg2 = sm.OLS(y1,sm.add_constant(X2), missing='drop')

results2= reg2.fit(cov_type='HC1')
results2.summary()
```

[29]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    IRR   R-squared:                       0.525
Model:                            OLS   Adj. R-squared:                  0.518
Method:                 Least Squares   F-statistic:                     24.76
Date:                Sun, 27 Jun 2021   Prob (F-statistic):           4.61e-25
Time:                        10:46:01   Log-Likelihood:                -1626.2
No. Observations:                 417   AIC:                             3266.
Df Residuals:                     410   BIC:                             3295.
Df Model:                           6
Covariance Type:                  HC1
================================================================================
==============
                                 coef    std err          z      P>|z|
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
const                         -0.2392      5.955     -0.040      0.968
```

```
-11.910      11.432
TVPI                            18.2634      2.227      8.200      0.000
13.898      22.629
Capital called                  -0.1262      0.057     -2.225      0.026
-0.237      -0.015
Fund size                       -0.7041      0.517     -1.361      0.173
-1.718       0.310
Fund number                      0.0478      0.051      0.929      0.353
-0.053       0.149
Industrial diversification       0.3506      1.253      0.280      0.780
-2.106       2.807
Geographic focus                 0.3199      2.442      0.131      0.896
-4.467       5.107
==============================================================================
Omnibus:                      668.364   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           293721.414
Skew:                           8.714   Prob(JB):                         0.00
Kurtosis:                     131.845   Cond. No.                         964.
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

[30]:
```python
## STANDARD ROBUST ERROR MODEL for CRISIS DUMMY


y1 = dataframe[['IRR']]
X1 = dataframe[['TVPI', 'Capital called', 'Fund size', 'Fund number',
                'Industrial diversification', 'Geographic focus', 'Vintage year␣
 ↪06-08']]
X2 = X1.astype(float)

reg1 = sm.OLS(y1,sm.add_constant(X2), missing='drop')

results1 = reg1.fit(cov_type='HC1')
results1.summary()
```

[30]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    IRR   R-squared:                       0.546
Model:                            OLS   Adj. R-squared:                  0.538
Method:                 Least Squares   F-statistic:                     29.19
Date:                Sun, 27 Jun 2021   Prob (F-statistic):           1.24e-32
Time:                        10:46:06   Log-Likelihood:                -1616.8
```

```
No. Observations:                 417   AIC:                             3250.
Df Residuals:                     409   BIC:                             3282.
Df Model:                           7
Covariance Type:                  HC1
================================================================================
==============
                                 coef    std err          z      P>|z|
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
const                         -0.5673      5.887     -0.096      0.923
-12.105      10.970
TVPI                          17.9818      2.199      8.176      0.000
13.671      22.292
Capital called                -0.1100      0.056     -1.955      0.051
-0.220       0.000
Fund size                     -0.4987      0.510     -0.978      0.328
-1.498       0.500
Fund number                    0.0290      0.048      0.605      0.545
-0.065       0.123
Industrial diversification     0.6371      1.207      0.528      0.598
-1.729       3.003
Geographic focus              -0.4298      2.354     -0.183      0.855
-5.043       4.183
Vintage year 06-08            -6.0755      0.814     -7.460      0.000
-7.672      -4.479
==============================================================================
Omnibus:                      689.296   Durbin-Watson:                   1.973
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           343934.234
Skew:                           9.237   Prob(JB):                         0.00
Kurtosis:                     142.476   Cond. No.                         964.
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

[86]:
```python
# SECTION B
dfPMEall = pd.read_excel('PME-Dataset-1.xlsx')
dfPMEall.head()
```

[86]:
```
   FUND ID  FIRM ID                              NAME  Vintage year  Fund size  \
0      634      233             Doughty Hanson & Co IV          2004    2030.70
1     4675       49                    Apax Europe VI          2005    5590.03
2     5696    94970  CVC European Equity Partners IV          2005    7258.81
3     7539        4                     3i Eurofund V          2006    6598.51
4     7234     2212                      Altor Fund II          2006    1395.98
```

```
   Net IRR  Net Multiple   RVPI     DPI  Called  Quartile rank DATE REPORTED  \
0     9.73          1.51   0.67  149.95  106.60              4    2020-06-30
1    12.70          2.02  14.00  188.00  100.00              2    2020-06-30
2    16.60          2.00   0.10  199.98   89.79              2    2020-06-30
3    12.40          2.43   2.00  241.00  102.23              1    2020-09-30
4    10.68          1.97  27.55  169.58  100.00              1    2020-09-30

   STATUS  S&P 500 PME  RUSSELL 2000 PME  MSCI PME  Fund number Region focus  \
0  Closed     1.130821          1.143126  1.201371            5       Europe
1  Closed     1.291170          1.299398  1.467259           28       Europe
2  Closed     1.593540          1.565293  1.921410            6       Europe
3  Closed     1.178697          1.237254  1.881750           10       Europe
4  Closed     1.180189          1.192204  1.699409            2       Europe

  Core Industries
0     Diversified
1     Diversified
2     Diversified
3     Diversified
4     Diversified
```

[87]:
```
dfPME = dfPMEall.drop(['FUND ID','FIRM ID', 'NAME','DATE REPORTED',
                        'STATUS'], axis = 1)

dfPME.head()
```

[87]:
```
   Vintage year  Fund size  Net IRR  Net Multiple   RVPI     DPI  Called  \
0          2004    2030.70     9.73          1.51   0.67  149.95  106.60
1          2005    5590.03    12.70          2.02  14.00  188.00  100.00
2          2005    7258.81    16.60          2.00   0.10  199.98   89.79
3          2006    6598.51    12.40          2.43   2.00  241.00  102.23
4          2006    1395.98    10.68          1.97  27.55  169.58  100.00

   Quartile rank  S&P 500 PME  RUSSELL 2000 PME  MSCI PME  Fund number  \
0              4     1.130821          1.143126  1.201371            5
1              2     1.291170          1.299398  1.467259           28
2              2     1.593540          1.565293  1.921410            6
3              1     1.178697          1.237254  1.881750           10
4              1     1.180189          1.192204  1.699409            2

  Region focus Core Industries
0       Europe     Diversified
1       Europe     Diversified
2       Europe     Diversified
3       Europe     Diversified
4       Europe     Diversified
```

[88]:
```
#Renaming variables
dfPME.rename(columns = {'Fund size':'Size','Net IRR':'IRR','Net Multiple':
 ↪'TVPI',
                        'Called':'Capital called'},inplace=True)
dfPME.head()
```

[88]:
```
   Vintage year     Size    IRR  TVPI   RVPI     DPI  Capital called  \
0          2004  2030.70   9.73  1.51   0.67  149.95          106.60
1          2005  5590.03  12.70  2.02  14.00  188.00          100.00
2          2005  7258.81  16.60  2.00   0.10  199.98           89.79
3          2006  6598.51  12.40  2.43   2.00  241.00          102.23
4          2006  1395.98  10.68  1.97  27.55  169.58          100.00

   Quartile rank  S&P 500 PME  RUSSELL 2000 PME  MSCI PME  Fund number  \
0              4     1.130821          1.143126  1.201371            5
1              2     1.291170          1.299398  1.467259           28
2              2     1.593540          1.565293  1.921410            6
3              1     1.178697          1.237254  1.881750           10
4              1     1.180189          1.192204  1.699409            2

  Region focus Core Industries
0       Europe     Diversified
1       Europe     Diversified
2       Europe     Diversified
3       Europe     Diversified
4       Europe     Diversified
```

[89]:
```
# Creating Dummy Variables
dfPME['Industrial diversification'] = dfPME['Core Industries'].apply(lambda x:␣
 ↪'1' if x == 'Diversified' else '0')
dfPME['Geographic focus'] = dfPME['Region focus'].apply(lambda x: '1' if x ==␣
 ↪'Europe' else '0')
dfPME['Vintage year 06-08'] = dfPME['Vintage year'].apply(lambda x: 1 if (x␣
 ↪>=2006) & (x <= 2008) else 0)

dfPME.head()
```

[89]:
```
   Vintage year     Size    IRR  TVPI   RVPI     DPI  Capital called  \
0          2004  2030.70   9.73  1.51   0.67  149.95          106.60
1          2005  5590.03  12.70  2.02  14.00  188.00          100.00
2          2005  7258.81  16.60  2.00   0.10  199.98           89.79
3          2006  6598.51  12.40  2.43   2.00  241.00          102.23
4          2006  1395.98  10.68  1.97  27.55  169.58          100.00

   Quartile rank  S&P 500 PME  RUSSELL 2000 PME  MSCI PME  Fund number  \
0              4     1.130821          1.143126  1.201371            5
1              2     1.291170          1.299398  1.467259           28
```

```
2               2      1.593540          1.565293  1.921410           6
3               1      1.178697          1.237254  1.881750          10
4               1      1.180189          1.192204  1.699409           2

  Region focus Core Industries Industrial diversification Geographic focus  \
0       Europe     Diversified                          1                1
1       Europe     Diversified                          1                1
2       Europe     Diversified                          1                1
3       Europe     Diversified                          1                1
4       Europe     Diversified                          1                1

   Vintage year 06-08
0                   0
1                   0
2                   0
3                   1
4                   1
```

[90]:
```python
# Logging Fund variable
dfPME['Fund size'] = np.log(dfPME['Size'])

dfPME.head()
```

[90]:
```
   Vintage year     Size    IRR  TVPI   RVPI     DPI  Capital called  \
0          2004  2030.70   9.73  1.51   0.67  149.95          106.60
1          2005  5590.03  12.70  2.02  14.00  188.00          100.00
2          2005  7258.81  16.60  2.00   0.10  199.98           89.79
3          2006  6598.51  12.40  2.43   2.00  241.00          102.23
4          2006  1395.98  10.68  1.97  27.55  169.58          100.00

   Quartile rank  S&P 500 PME  RUSSELL 2000 PME  MSCI PME  Fund number  \
0              4     1.130821          1.143126  1.201371            5
1              2     1.291170          1.299398  1.467259           28
2              2     1.593540          1.565293  1.921410            6
3              1     1.178697          1.237254  1.881750           10
4              1     1.180189          1.192204  1.699409            2

  Region focus Core Industries Industrial diversification Geographic focus  \
0       Europe     Diversified                          1                1
1       Europe     Diversified                          1                1
2       Europe     Diversified                          1                1
3       Europe     Diversified                          1                1
4       Europe     Diversified                          1                1

   Vintage year 06-08  Fund size
0                   0   7.616136
1                   0   8.628740
```

```
2                    0    8.889971
3                    1    8.794599
4                    1    7.241352
```

[91]:

```python
# OLS Regression S&P PME

y1 = dfPME[['S&P 500 PME']]
X1 = dfPME[['TVPI', 'IRR', 'Capital called', 'Fund size', 'Fund number',␣
 ↪'Industrial diversification', 'Geographic focus']]
X2 = X1.astype(float)

reg1 = sm.OLS(y1,sm.add_constant(X2), missing='drop')

results1 = reg1.fit(cov_type='HC1')
results1.summary()
```

[91]:

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:            S&P 500 PME   R-squared:                       0.850
Model:                            OLS   Adj. R-squared:                  0.825
Method:                 Least Squares   F-statistic:                     29.38
Date:                Sat, 26 Jun 2021   Prob (F-statistic):           2.05e-14
Time:                        13:35:14   Log-Likelihood:                 33.831
No. Observations:                  51   AIC:                            -51.66
Df Residuals:                      43   BIC:                            -36.21
Df Model:                           7
Covariance Type:                  HC1
================================================================================
==============
                                  coef    std err          z      P>|z|
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
const                           0.6609      0.163      4.051      0.000
0.341       0.981
TVPI                            0.3486      0.066      5.258      0.000
0.219       0.479
IRR                             0.0142      0.004      4.052      0.000
0.007       0.021
Capital called                 -0.0031      0.002     -1.838      0.066
-0.006        0.000
Fund size                       0.0031      0.017      0.182      0.856
-0.030        0.037
Fund number                     0.0014      0.004      0.352      0.725
-0.006        0.009
```

```
Industrial diversification     0.0004      0.088      0.004      0.997
-0.173       0.173
Geographic focus              -0.0891      0.051     -1.749      0.080
-0.189       0.011
==============================================================================
Omnibus:                        1.466   Durbin-Watson:                   1.454
Prob(Omnibus):                  0.481   Jarque-Bera (JB):                1.179
Skew:                           0.370   Prob(JB):                        0.555
Kurtosis:                       2.923   Cond. No.                         872.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

[92]:

```python
# Table 8
groupby_quartile = dfPME.groupby(['Quartile rank'])[['S&P 500 PME', 'RUSSELL␣
↪2000 PME', 'MSCI PME']]
groupby_quartile.mean()
```

[92]:

Quartile rank	S&P 500 PME	RUSSELL 2000 PME	MSCI PME
1	1.428617	1.613200	1.901465
2	1.147958	1.217936	1.427902
3	0.959974	1.082474	1.206026
4	0.870756	0.951636	1.038346

[93]:

```python
# Table 7 with added counts of funds (manually)
groupby_quartile = dfPME.groupby(['Vintage year'])[['S&P 500 PME', 'RUSSELL␣
↪2000 PME', 'MSCI PME']]
groupby_quartile.mean()
```

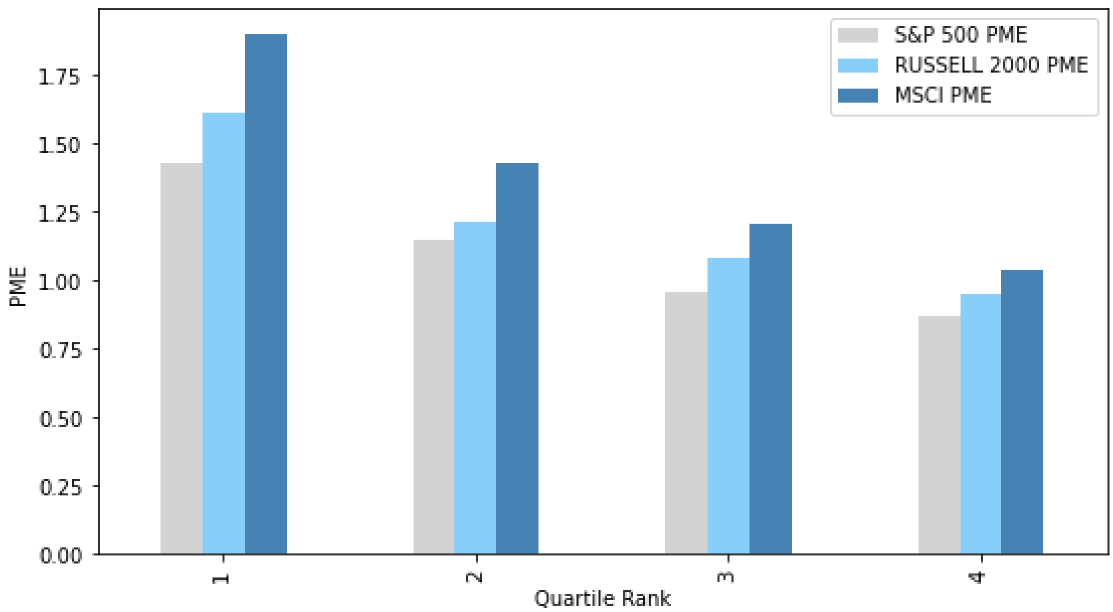
[93]:

Vintage year	S&P 500 PME	RUSSELL 2000 PME	MSCI PME
2004	1.130821	1.143126	1.201371
2005	1.442355	1.432346	1.694335
2006	1.134049	1.155781	1.652822
2007	0.859477	0.871468	1.120113
2008	1.111372	1.169606	1.529009
2009	0.959303	0.991482	1.245379
2010	1.183450	1.302273	1.563973
2011	1.012500	1.120783	1.340573
2012	1.346200	1.549890	1.798030
2013	1.289633	1.443942	1.612654
2014	1.156979	1.323494	1.448773
2015	1.030277	1.169761	1.220427
2016	1.051914	1.255534	1.227035

[94]:
```
# Figure 8 : Average PME by quartile rank

my_colors = ['lightgrey','lightskyblue','steelblue', 'lavender']
dfPME.groupby('Quartile rank').mean()[['S&P 500 PME',
        'RUSSELL 2000 PME', 'MSCI PME']].plot.bar(legend=True, figsize=(9,5),␣
 ↪color=my_colors)
plt.xlabel('Quartile Rank')
plt.ylabel('PME')
```

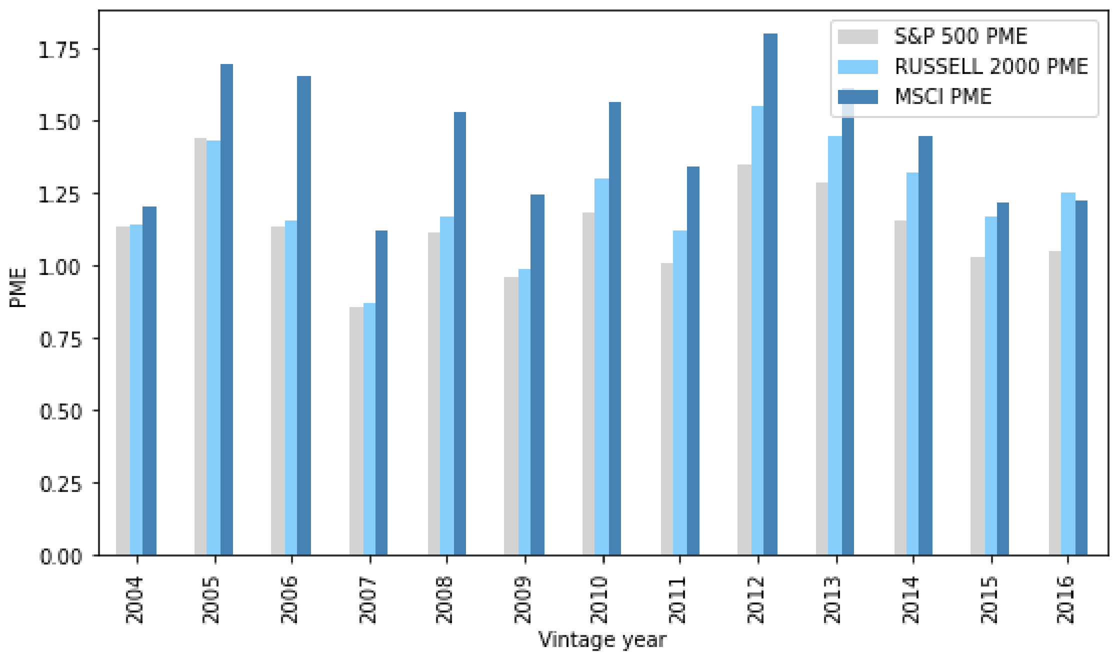
[94]: Text(0, 0.5, 'PME')



[95]:
```
# Figure 6 : Average PME by vintage year

my_colors = ['lightgrey','lightskyblue','steelblue']
dfPME.groupby('Vintage year').mean()[['S&P 500 PME', 'RUSSELL 2000 PME',
    'MSCI PME']].plot.bar(legend=True, figsize=(9,5), color=my_colors)
plt.xlabel('Vintage year')
plt.ylabel('PME')
```

[95]: Text(0, 0.5, 'PME')



[100]:
```python
## Chi-Square Test
df = pd.read_excel('Total-Dataset-CHI.xlsx')
df.head()
```

[100]:
```
   FUND ID  FIRM ID                                NAME     ASSET CLASS  \
0       16        4                     3i Eurofund III  Private Equity
1     2470      908         Alpha Private Equity Fund 3  Private Equity
2      742      708                           Astorg II  Private Equity
3     3717     2211                     Capvis Equity I  Private Equity
4      430      162  Charterhouse Capital Partners VI Aux  Private Equity

   Vintage year  Fund size  Net IRR  Net Multiple  RVPI     DPI  Called  \
0          1999    2507.00    19.30          2.04   0.0  204.00   90.50
1          1999     314.18    11.50          1.80   0.0  180.00  100.00
2          1999     190.74    23.70          3.06   0.0  306.40   89.06
3          1999     209.81    12.40          1.74   0.0  173.90   96.00
4          1999     567.77    32.32          2.80   0.0  280.42   75.90

  Quartile rank DATE REPORTED      STATUS Geographic focus  \
0           Two    2021-03-31  Liquidated           Europe
1           Two    2021-03-31  Liquidated      West Europe
2           One    2021-03-31  Liquidated           France
3         Three    2021-03-31  Liquidated      West Europe
4           One    2021-03-31  Liquidated           Europe
```

```
                   FUND MANAGER      COUNTRY  Fund number Region focus  \
0                            3i           UK            7       Europe
1                   Alpha Group   Luxembourg            3       Europe
2                        Astorg       France            1       Europe
3                     Capvis AG  Switzerland            1       Europe
4  Charterhouse Capital Partners           UK            3       Europe

                      Core Industries
0                         Diversified
1                         Diversified
2                         Diversified
3  Healthcare, Information Technology
4                         Diversified
```

[101]:

```
contigency = pd.crosstab(df['Core Industries'], df['Region focus'], margins =␣
↪False)
print(contigency)
```

```
Region focus                        Africa  Asia  Europe  \
Core Industries
Business Services                        0     0       3
Consumer Discretionary                   0     0      18
Diversified                              2     9     342
Energy & Utilities                       0     0       3
Financial & Insurance Services           0     0       4
Healthcare                               0     0       2
Healthcare, Information Technology       0     0       2
Industrials                              0     0      11
Information Technology                   0     0       8
Telecoms & Media                         0     0       3

Region focus                        Middle East & Israel  North America
Core Industries
Business Services                                      0              0
Consumer Discretionary                                 0              0
Diversified                                            1              7
Energy & Utilities                                     0              0
Financial & Insurance Services                         0              0
Healthcare                                             0              0
Healthcare, Information Technology                     0              0
Industrials                                            0              2
Information Technology                                 0              0
Telecoms & Media                                       0              0
```

[102]:

```
# Chi-square test of independence.
c, p, dof, expected = chi2_contingency(contigency)
# Print the p-value
```

```
print(p) #0.999 means do not reject null hypothesis
```

0.9997256770063442

[103]:
```
# Fama French five-factor model
# Loading IRR data
irrtest = pd.read_excel('Irrtest.xlsx')
irrtest.head()
```

[103]:

	Year	MeanIRR
0	1999	12.77
1	2000	21.70
2	2001	27.32
3	2002	29.26
4	2003	33.17

[104]:
```
# Loading Fama French data - downloaded from Kenneth R. French data library
ff5 = pd.read_excel('FF5factor.xlsx')
ff5.head()
```

[104]:

	Year	Mkt-RF	SMB	HML	RMW	CMA	RF
0	1999	20.57	8.71	-31.63	-27.20	-8.44	4.68
1	2000	-17.60	2.87	46.07	25.14	32.74	5.89
2	2001	-15.20	23.56	18.67	17.19	11.79	3.83
3	2002	-22.76	5.67	8.20	20.42	14.44	1.65
4	2003	30.75	24.21	4.06	-20.49	16.96	1.02

[105]:
```
# Note: All data is in percentage terms
# Merging the two datasets on "year"

merge = pd.merge(irrtest,ff5, on='Year')
merge
```

[105]:

	Year	MeanIRR	Mkt-RF	SMB	HML	RMW	CMA	RF
0	1999	12.77	20.57	8.71	-31.63	-27.20	-8.44	4.68
1	2000	21.70	-17.60	2.87	46.07	25.14	32.74	5.89
2	2001	27.32	-15.20	23.56	18.67	17.19	11.79	3.83
3	2002	29.26	-22.76	5.67	8.20	20.42	14.44	1.65
4	2003	33.17	30.75	24.21	4.06	-20.49	16.96	1.02
5	2004	29.01	10.72	7.12	7.47	8.41	-7.88	1.20
6	2005	16.55	3.09	-0.91	9.63	1.47	-5.01	2.98
7	2006	10.74	10.60	1.67	12.12	4.02	8.19	4.80
8	2007	8.89	1.04	-8.01	-17.02	4.78	-7.80	4.66
9	2008	11.95	-38.34	3.19	0.57	15.33	4.16	1.60
10	2009	13.39	28.26	8.11	-9.21	3.64	-2.69	0.10
11	2010	13.36	17.37	13.36	-5.33	-1.94	9.98	0.12
12	2011	20.79	0.44	-5.57	-8.40	13.15	-0.88	0.04
13	2012	15.39	16.27	-0.11	9.97	-5.53	9.41	0.06

```
14  2013     16.48    35.20    7.60    2.53  -3.67    1.33  0.02
15  2014     20.99    11.71   -8.00   -1.52   1.39   -1.77  0.02
16  2015     13.82     0.08   -5.82   -9.50   0.92   -8.52  0.02
17  2016     17.29    13.30    9.08   22.86   4.42    9.93  0.20
```

[106]:
```python
# Creating dependent variable (IRR minus risk-free return of the market)
merge['MeanIRR-RF']= merge.MeanIRR - merge.RF
merge.head()
```

[106]:

	Year	MeanIRR	Mkt-RF	SMB	HML	RMW	CMA	RF	MeanIRR-RF
0	1999	12.77	20.57	8.71	-31.63	-27.20	-8.44	4.68	8.09
1	2000	21.70	-17.60	2.87	46.07	25.14	32.74	5.89	15.81
2	2001	27.32	-15.20	23.56	18.67	17.19	11.79	3.83	23.49
3	2002	29.26	-22.76	5.67	8.20	20.42	14.44	1.65	27.61
4	2003	33.17	30.75	24.21	4.06	-20.49	16.96	1.02	32.15

[107]:
```python
# Preparing the model
y = merge['MeanIRR-RF']
X = merge[['Mkt-RF', 'SMB','HML', 'RMW', 'CMA']] # the five market risk factors

X_sm = sm.add_constant(X)
```

[108]:
```python
# Running the model, using OLS and constant variance
model = sm.OLS(y,X_sm)
results = model.fit()
results.summary()
```

```
/opt/anaconda3/lib/python3.8/site-packages/scipy/stats/stats.py:1603:
UserWarning: kurtosistest only valid for n>=20 … continuing anyway, n=18
  warnings.warn("kurtosistest only valid for n>=20 … continuing "
```

[108]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:             MeanIRR-RF   R-squared:                       0.279
Model:                            OLS   Adj. R-squared:                 -0.021
Method:                 Least Squares   F-statistic:                    0.9304
Date:                Sat, 26 Jun 2021   Prob (F-statistic):              0.495
Time:                        14:09:31   Log-Likelihood:                -58.679
No. Observations:                  18   AIC:                             129.4
Df Residuals:                      12   BIC:                             134.7
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         13.8777      2.739      5.067      0.000       7.910      19.845
```

```
Mkt-RF         0.0627      0.153      0.409      0.689     -0.271      0.396
SMB            0.3924      0.245      1.601      0.135     -0.142      0.926
HML            0.0800      0.214      0.373      0.715     -0.387      0.547
RMW            0.1508      0.275      0.548      0.594     -0.449      0.751
CMA           -0.0592      0.305     -0.194      0.850     -0.725      0.606
==============================================================================
Omnibus:                        2.427   Durbin-Watson:                   1.104
Prob(Omnibus):                  0.297   Jarque-Bera (JB):                1.489
Skew:                           0.450   Prob(JB):                        0.475
Kurtosis:                       1.916   Cond. No.                         36.4
==============================================================================

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

[109]:

```
# Running the model with robust standard errors
model1 = sm.OLS(y,X_sm)
results1 = model1.fit(cov_type='HC1')
results1.summary()
```

```
/opt/anaconda3/lib/python3.8/site-packages/scipy/stats/stats.py:1603:
UserWarning: kurtosistest only valid for n>=20 … continuing anyway, n=18
  warnings.warn("kurtosistest only valid for n>=20 … continuing "
```

[109]:

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:              MeanIRR-RF   R-squared:                       0.279
Model:                             OLS   Adj. R-squared:                 -0.021
Method:                  Least Squares   F-statistic:                     1.403
Date:                 Sat, 26 Jun 2021   Prob (F-statistic):              0.291
Time:                         14:09:39   Log-Likelihood:                -58.679
No. Observations:                   18   AIC:                             129.4
Df Residuals:                       12   BIC:                             134.7
Df Model:                            5
Covariance Type:                   HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         13.8777      2.078      6.680      0.000       9.806      17.950
Mkt-RF         0.0627      0.115      0.544      0.586      -0.163      0.288
SMB            0.3924      0.218      1.801      0.072      -0.035      0.819
HML            0.0800      0.202      0.397      0.692      -0.315      0.475
RMW            0.1508      0.260      0.581      0.561      -0.358      0.660
CMA           -0.0592      0.307     -0.193      0.847      -0.660      0.542
```

```
==============================================================================
Omnibus:                        2.427   Durbin-Watson:                   1.104
Prob(Omnibus):                  0.297   Jarque-Bera (JB):                1.489
Skew:                           0.450   Prob(JB):                        0.475
Kurtosis:                       1.916   Cond. No.                         36.4
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

```
[ ]:
```