

```
In [1]: import pandas as pd
import statsmodels.api as sm
from helper import ordinary_least_squares_regression

%matplotlib inline

#HSBC and ETF data was from Nasdaq and all other columns were from Ken French https://mba.tuck.dartmouth.edu
df = pd.read_excel("HSBC.ods", engine="odf")
```

```
In [2]: # cleaning data, data from French is already expressed in percent
for column in ["XLF-rf", "IXG-rf", "HSBC-rf"]:
    df[column] = pd.to_numeric(df[column], errors='coerce') * 100
df_cleaned = df.dropna()
df_cleaned.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 1202 entries, 53 to 1254
Data columns (total 28 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date                   1202 non-null   object
1   HSBC Close/Last        1202 non-null   object
2   Volume                 1202 non-null   int64
3   Open                   1202 non-null   object
4   High                   1202 non-null   object
5   Low                    1202 non-null   object
6   HSBC return            1202 non-null   object
7   XLF Close/Last         1202 non-null   float64
8   XLF return             1202 non-null   object
9   IXG Close/Last         1202 non-null   float64
10  IXG return             1202 non-null   object
11  Mkt-RF                 1202 non-null   float64
12  SMB                    1202 non-null   float64
13  HML                    1202 non-null   float64
14  RMW                    1202 non-null   float64
15  CMA                    1202 non-null   float64
16  RF                     1202 non-null   float64
17  XLF-rf                 1202 non-null   float64
18  IXG-rf                 1202 non-null   float64
19  Banks-rf               1202 non-null   float64
20  Insur-rf               1202 non-null   float64
21  RlEst-rf               1202 non-null   float64
22  Fin -rf                1202 non-null   float64
23  Banks                  1202 non-null   float64
24  Insur                  1202 non-null   float64
25  RlEst                  1202 non-null   float64
26  Fin                    1202 non-null   float64
27  HSBC-rf                1202 non-null   float64
dtypes: float64(19), int64(1), object(8)
memory usage: 272.3+ KB

```

correlation table

```
In [3]: df_cleaned[['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'XLF-rf', 'IXG-rf', 'Banks-rf', 'Insur-rf', 'RlEst-rf', 'Fin
```

Out[3]:

	Mkt-RF	SMB	HML	RMW	CMA	XLf-rf	IXG-rf	Banks-rf	Insur-rf	RIEst-rf	Fin -rf
Mkt-RF	1.000000	0.241713	-0.149459	-0.231911	-0.334568	0.667645	0.665665	0.772806	0.701076	0.787788	0.872955
SMB	0.241713	1.000000	0.341334	-0.300474	0.056323	0.336009	0.320983	0.445459	0.210226	0.518404	0.367765
HML	-0.149459	0.341334	1.000000	0.402650	0.602861	0.335870	0.288290	0.398252	0.247018	0.124394	0.148594
RMW	-0.231911	-0.300474	0.402650	1.000000	0.332726	-0.025180	-0.041786	-0.058386	0.041845	-0.203812	-0.188287
CMA	-0.334568	0.056323	0.602861	0.332726	1.000000	-0.016949	-0.020420	-0.070568	0.003107	-0.182833	-0.204975
XLf-rf	0.667645	0.336009	0.335870	-0.025180	-0.016949	1.000000	0.967325	0.814004	0.680716	0.653767	0.770289
IXG-rf	0.665665	0.320983	0.288290	-0.041786	-0.020420	0.967325	1.000000	0.766868	0.642052	0.646327	0.738759
Banks-rf	0.772806	0.445459	0.398252	-0.058386	-0.070568	0.814004	0.766868	1.000000	0.737845	0.774963	0.886440
Insur-rf	0.701076	0.210226	0.247018	0.041845	0.003107	0.680716	0.642052	0.737845	1.000000	0.613500	0.730809
RIEst-rf	0.787788	0.518404	0.124394	-0.203812	-0.182833	0.653767	0.646327	0.774963	0.613500	1.000000	0.800856
Fin -rf	0.872955	0.367765	0.148594	-0.188287	-0.204975	0.770289	0.738759	0.886440	0.730809	0.800856	1.000000
HSBC-rf	0.304714	0.201876	0.300325	0.055650	0.114450	0.618767	0.689784	0.451253	0.332055	0.345507	0.389772

With IXG only

```

In [4]: column_combos = [['Mkt-RF', 'SMB', 'HML',
                           'RMW', 'CMA', 'IXG-rf'], ['Mkt-RF', 'SMB', 'HML',
                           'RMW', 'CMA', 'IXG-rf', 'Banks-rf', 'Insur-rf', 'RIEst-rf', 'Fin -rf']]

y = df_cleaned["HSBC-rf"]

for column_combo in column_combos:
    print(column_combo)
    X = sm.add_constant(df_cleaned[column_combo])
    results, df_residual = ordinary_least_squares_regression(y=y, X=X)

```

```
['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'IXG-rf']
```

OLS Regression Results

=====						
Dep. Variable:	HSBC-rf		R-squared:	0.522		
Model:	OLS		Adj. R-squared:	0.519		
Method:	Least Squares		F-statistic:	217.1		
Date:	Wed, 26 Mar 2025		Prob (F-statistic):	2.55e-187		
Time:	21:36:18		Log-Likelihood:	-2114.0		
No. Observations:	1202		AIC:	4242.		
Df Residuals:	1195		BIC:	4278.		
Df Model:	6					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.0961	0.056	1.728	0.084	-0.013	0.205
Mkt-RF	-0.4298	0.053	-8.087	0.000	-0.534	-0.326
SMB	0.0039	0.065	0.060	0.952	-0.123	0.130
HML	-0.0668	0.060	-1.117	0.264	-0.184	0.051
RMW	0.0876	0.078	1.120	0.263	-0.066	0.241
CMA	0.2094	0.093	2.252	0.025	0.027	0.392
IXG-rf	1.0912	0.040	27.621	0.000	1.014	1.169
=====						
Omnibus:	133.949		Durbin-Watson:	2.159		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	939.013		
Skew:	-0.226		Prob(JB):	1.25e-204		
Kurtosis:	7.306		Cond. No.	5.28		
=====						

Notes:

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

```
['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'IXG-rf', 'Banks-rf', 'Insur-rf', 'RlEst-rf', 'Fin -rf']
```

OLS Regression Results

=====			
Dep. Variable:	HSBC-rf	R-squared:	0.527
Model:	OLS	Adj. R-squared:	0.523
Method:	Least Squares	F-statistic:	132.6
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	1.29e-185
Time:	21:36:18	Log-Likelihood:	-2107.4
No. Observations:	1202	AIC:	4237.
Df Residuals:	1191	BIC:	4293.
Df Model:	10		

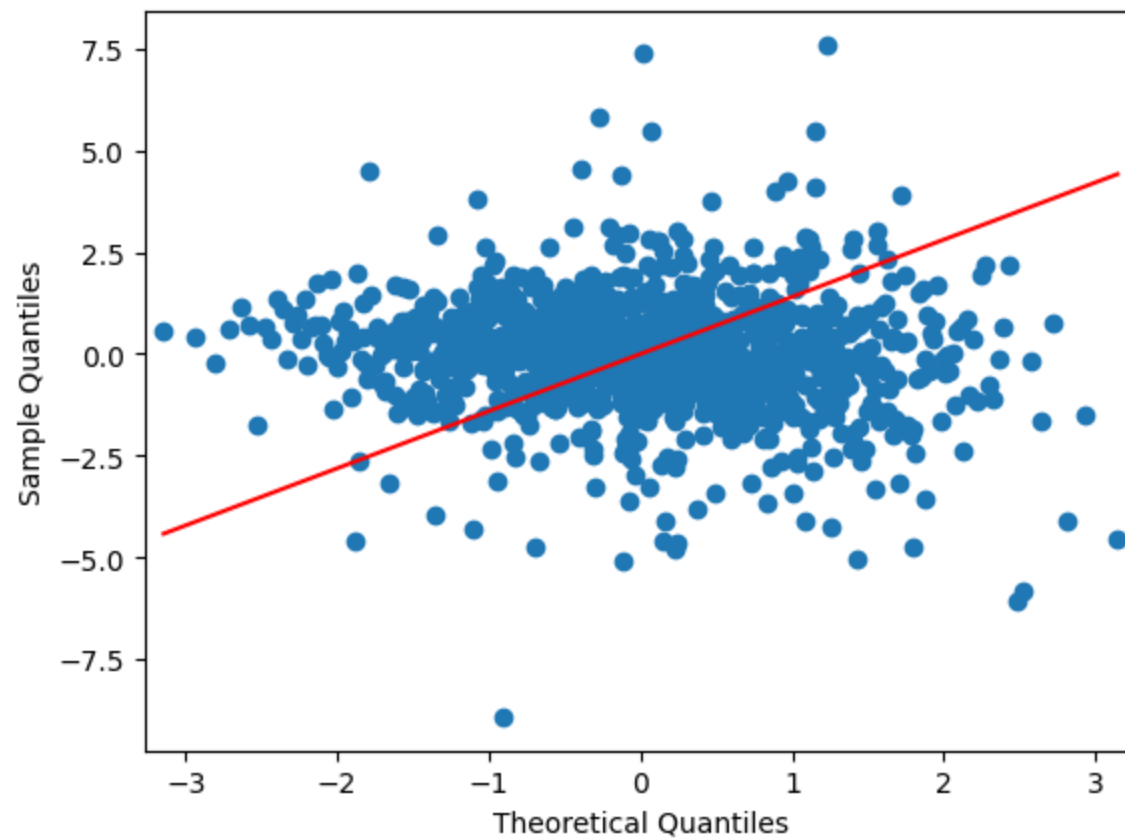
Covariance Type:		nonrobust				
	coef	std err	t	P> t	[0.025	0.975]
const	0.1135	0.056	2.018	0.044	0.003	0.224
Mkt-RF	-0.2798	0.113	-2.470	0.014	-0.502	-0.058
SMB	-0.0386	0.071	-0.542	0.588	-0.178	0.101
HML	-0.0151	0.089	-0.169	0.866	-0.191	0.160
RMW	0.0915	0.080	1.139	0.255	-0.066	0.249
CMA	0.2343	0.100	2.332	0.020	0.037	0.431
IXG-rf	1.1060	0.041	27.248	0.000	1.026	1.186
Banks-rf	0.0496	0.079	0.626	0.532	-0.106	0.205
Insur-rf	-0.1804	0.054	-3.359	0.001	-0.286	-0.075
RlEst-rf	0.0183	0.041	0.450	0.653	-0.061	0.098
Fin -rf	-0.0839	0.079	-1.059	0.290	-0.239	0.072
Omnibus:		134.352	Durbin-Watson:		2.172	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		962.234	
Skew:		-0.215	Prob(JB):		1.13e-209	
Kurtosis:		7.362	Cond. No.		13.1	

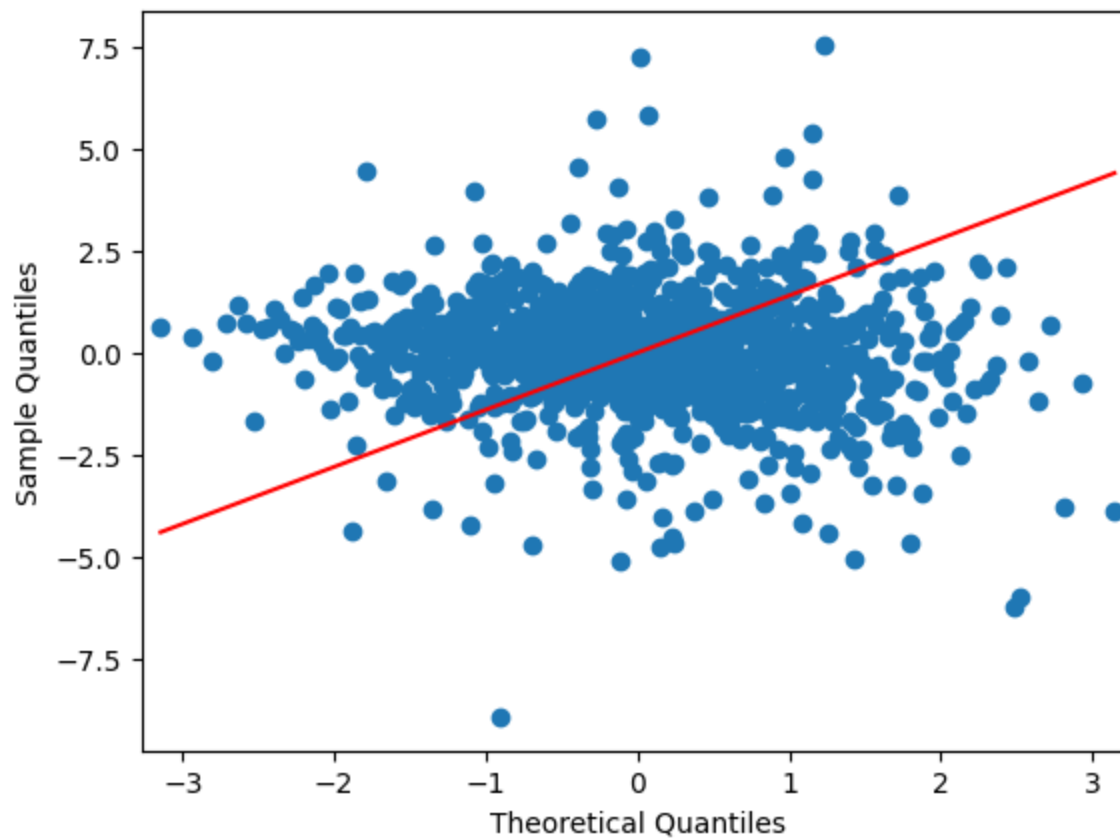
Notes:

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

/home/george/HSBC/helper.py:12: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
plot.show()

/home/george/HSBC/helper.py:12: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
plot.show()





With XLF only

```
In [5]: column_combos = [['Mkt-RF', 'SMB', 'HML',  
                        'RMW', 'CMA', 'XLF-rf'], ['Mkt-RF', 'SMB', 'HML',  
                        'RMW', 'CMA', 'XLF-rf', 'Banks-rf', 'Insur-rf', 'RlEst-rf', 'Fin -rf']]  
  
for column_combo in column_combos:  
    print(column_combo)  
    X = sm.add_constant(df_cleaned[column_combo])  
    results, df_residual = ordinary_least_squares_regression(y=y, X=X)
```

['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'XLF-rf']

OLS Regression Results

=====						
Dep. Variable:	HSBC-rf	R-squared:	0.409			
Model:	OLS	Adj. R-squared:	0.406			
Method:	Least Squares	F-statistic:	138.1			
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	6.61e-133			
Time:	21:36:18	Log-Likelihood:	-2240.5			
No. Observations:	1202	AIC:	4495.			
Df Residuals:	1195	BIC:	4531.			
Df Model:	6					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.1221	0.062	-1.983	0.048	-0.243	-0.001
Mkt-RF	-0.2693	0.061	-4.413	0.000	-0.389	-0.150
SMB	0.0199	0.072	0.278	0.781	-0.121	0.161
HML	-0.0644	0.069	-0.926	0.355	-0.201	0.072
RMW	0.0777	0.087	0.894	0.371	-0.093	0.248
CMA	0.3021	0.104	2.915	0.004	0.099	0.505
XLF-rf	0.8686	0.044	19.781	0.000	0.782	0.955
=====						
Omnibus:	108.979	Durbin-Watson:	2.134			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	620.948			
Skew:	-0.161	Prob(JB):	1.46e-135			
Kurtosis:	6.506	Cond. No.	5.60			
=====						

Notes:

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

['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'XLF-rf', 'Banks-rf', 'Insur-rf', 'RlEst-rf', 'Fin -rf']

OLS Regression Results

=====			
Dep. Variable:	HSBC-rf	R-squared:	0.418
Model:	OLS	Adj. R-squared:	0.414
Method:	Least Squares	F-statistic:	85.69
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	1.08e-132
Time:	21:36:18	Log-Likelihood:	-2231.3
No. Observations:	1202	AIC:	4485.
Df Residuals:	1191	BIC:	4541.
Df Model:	10		

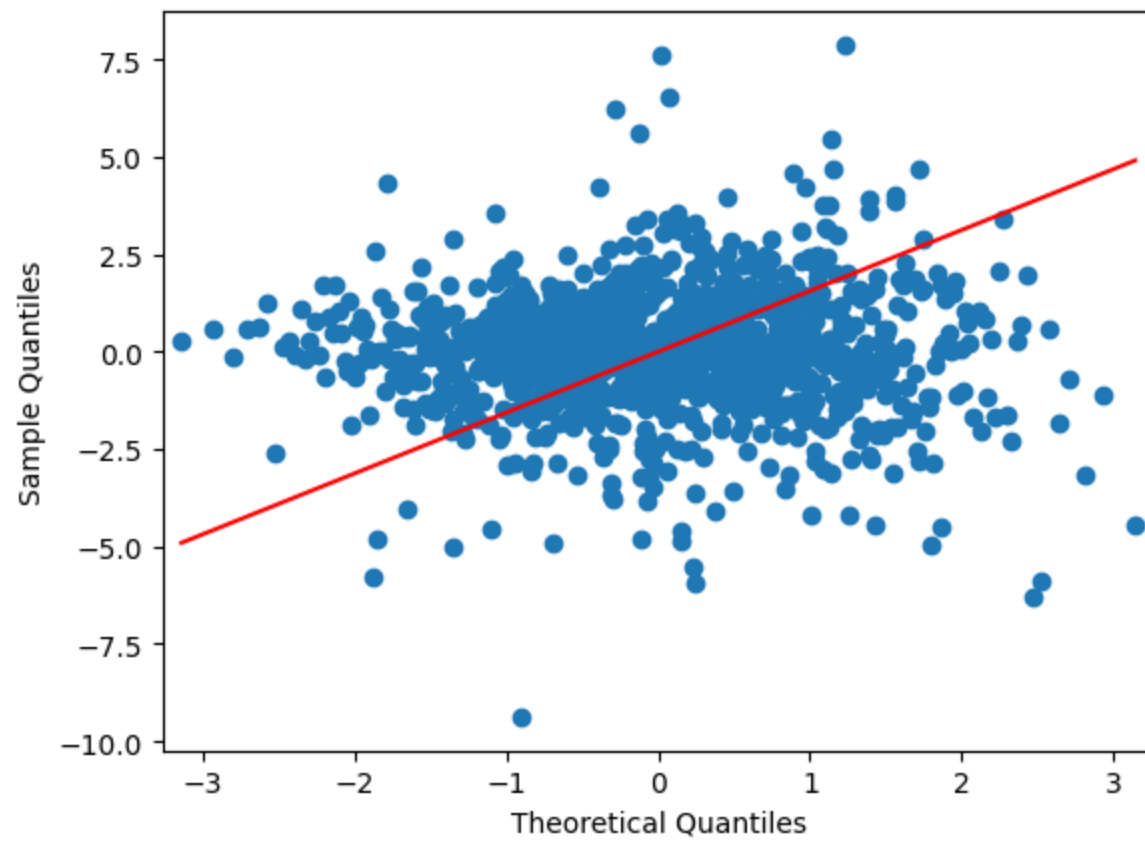
Covariance Type:		nonrobust				
	coef	std err	t	P> t	[0.025	0.975]
const	-0.0769	0.063	-1.215	0.225	-0.201	0.047
Mkt-RF	-0.0409	0.125	-0.327	0.744	-0.286	0.204
SMB	-0.0489	0.079	-0.619	0.536	-0.204	0.106
HML	0.0562	0.099	0.567	0.571	-0.138	0.251
RMW	0.0684	0.089	0.768	0.443	-0.106	0.243
CMA	0.3003	0.111	2.699	0.007	0.082	0.519
XLF-rf	0.9130	0.047	19.550	0.000	0.821	1.005
Banks-rf	-0.0279	0.089	-0.313	0.754	-0.203	0.147
Insur-rf	-0.2231	0.060	-3.728	0.000	-0.341	-0.106
REst-rf	0.0542	0.045	1.203	0.229	-0.034	0.143
Fin -rf	-0.1131	0.088	-1.280	0.201	-0.286	0.060
Omnibus:		109.455	Durbin-Watson:		2.148	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		632.992	
Skew:		-0.154	Prob(JB):		3.53e-138	
Kurtosis:		6.542	Cond. No.		13.3	

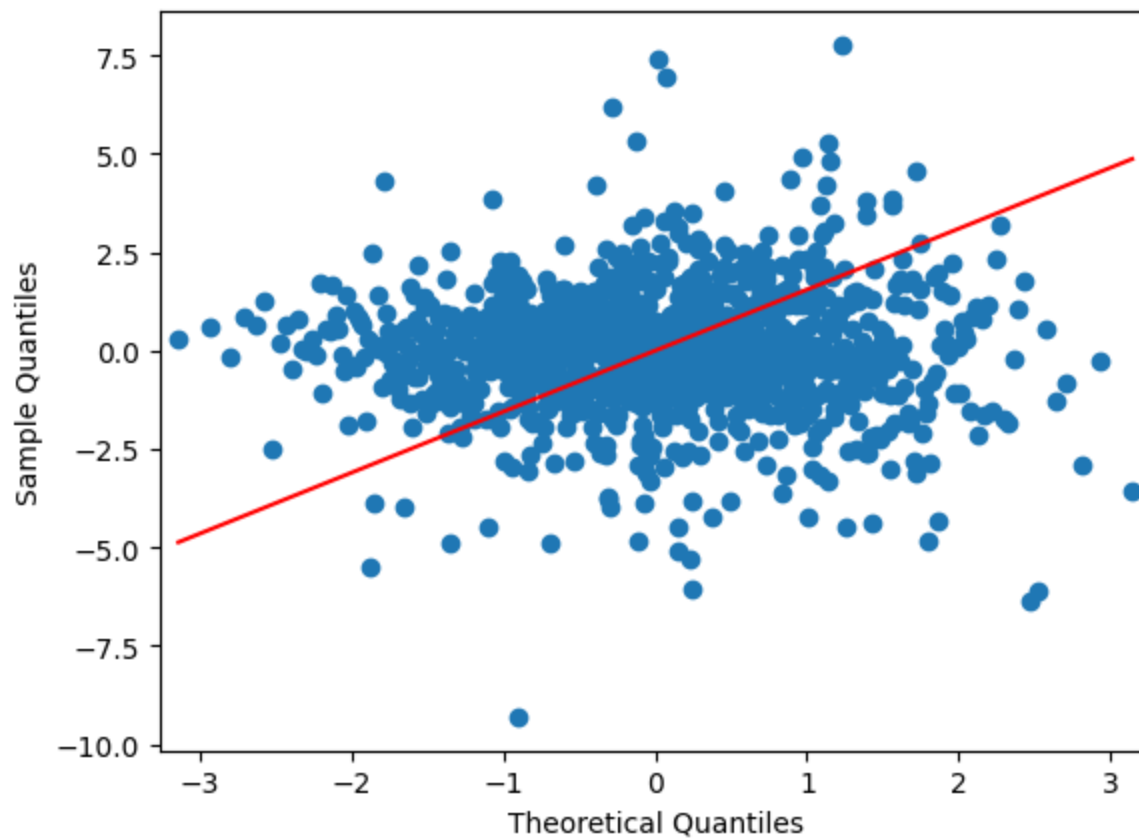
Notes:

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

/home/george/HSBC/helper.py:12: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
plot.show()

/home/george/HSBC/helper.py:12: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
plot.show()





Both XLF and IXG

```
In [6]: column_combos = [['Mkt-RF', 'SMB', 'HML',  
                        'RMW', 'CMA', 'XLF-rf', 'IXG-rf'], ['Mkt-RF', 'SMB', 'HML',  
                        'RMW', 'CMA', 'XLF-rf', 'IXG-rf', 'Banks-rf', 'Insur-rf', 'REst-rf', 'Fin -rf']]  
  
for column_combo in column_combos:  
    print(column_combo)  
    X = sm.add_constant(df_cleaned[column_combo])  
    results, df_residual = ordinary_least_squares_regression(y=y, X=X)
```

['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'XLF-rf', 'IXG-rf']

OLS Regression Results

Dep. Variable:	HSBC-rf	R-squared:	0.552			
Model:	OLS	Adj. R-squared:	0.549			
Method:	Least Squares	F-statistic:	210.2			
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	3.58e-203			
Time:	21:36:19	Log-Likelihood:	-2074.4			
No. Observations:	1202	AIC:	4165.			
Df Residuals:	1194	BIC:	4205.			
Df Model:	7					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.0339	0.054	0.625	0.532	-0.073	0.140
Mkt-RF	-0.3079	0.053	-5.789	0.000	-0.412	-0.204
SMB	-0.0193	0.062	-0.308	0.758	-0.142	0.103
HML	0.1139	0.061	1.859	0.063	-0.006	0.234
RMW	0.0646	0.076	0.853	0.394	-0.084	0.213
CMA	0.0928	0.091	1.020	0.308	-0.086	0.271
XLF-rf	-0.8795	0.097	-9.023	0.000	-1.071	-0.688
IXG-rf	1.8997	0.097	19.500	0.000	1.709	2.091
=====						
Omnibus:	135.755	Durbin-Watson:	2.222			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1078.176			
Skew:	-0.148	Prob(JB):	7.53e-235			
Kurtosis:	7.630	Cond. No.	9.83			
=====						

Notes:

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

['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'XLF-rf', 'IXG-rf', 'Banks-rf', 'Insur-rf', 'RlEst-rf', 'Fin -rf']

OLS Regression Results

=====			
Dep. Variable:	HSBC-rf	R-squared:	0.557
Model:	OLS	Adj. R-squared:	0.552
Method:	Least Squares	F-statistic:	135.8
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	3.38e-201
Time:	21:36:19	Log-Likelihood:	-2068.4
No. Observations:	1202	AIC:	4161.
Df Residuals:	1190	BIC:	4222.
=====			

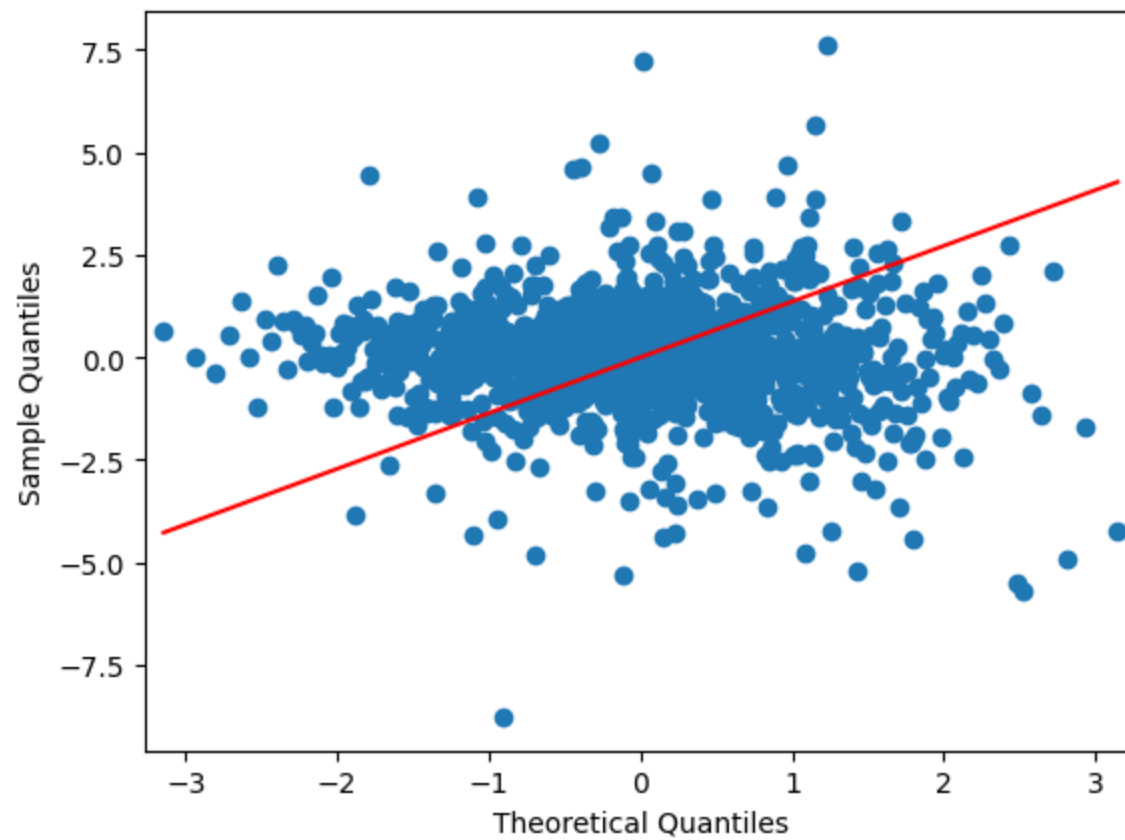
Df Model:	11					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.0186	0.056	0.335	0.738	-0.090	0.127
Mkt-RF	-0.4283	0.111	-3.860	0.000	-0.646	-0.211
SMB	-0.0449	0.069	-0.651	0.515	-0.180	0.090
HML	-0.0268	0.087	-0.310	0.757	-0.197	0.143
RMW	0.1078	0.078	1.385	0.166	-0.045	0.260
CMA	0.2031	0.097	2.086	0.037	0.012	0.394
XLF-rf	-0.9291	0.104	-8.930	0.000	-1.133	-0.725
IXG-rf	1.9295	0.100	19.248	0.000	1.733	2.126
Banks-rf	0.2157	0.079	2.732	0.006	0.061	0.371
Insur-rf	-0.1093	0.053	-2.078	0.038	-0.213	-0.006
REst-rf	-0.0059	0.039	-0.150	0.881	-0.083	0.072
Fin -rf	0.0025	0.077	0.033	0.974	-0.149	0.154
Omnibus:	136.681	Durbin-Watson:		2.228		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		1133.486		
Skew:	-0.113	Prob(JB):		7.35e-247		
Kurtosis:	7.752	Cond. No.		15.2		

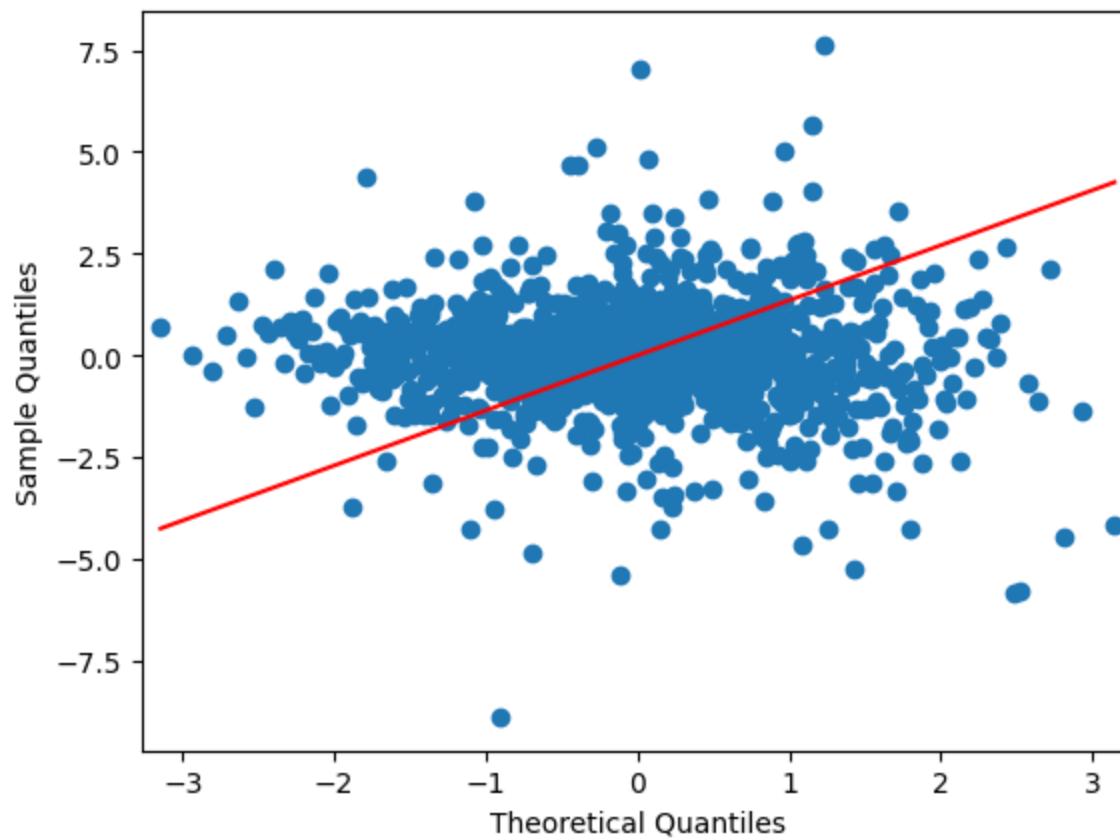
Notes:

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

/home/george/HSBC/helper.py:12: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
plot.show()

/home/george/HSBC/helper.py:12: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
plot.show()





```
In [7]: from matplotlib import pyplot as plt

fig = plt.figure()
ax = fig.add_subplot()
ax.set_xlabel("Actual HSBC-rf")
ax.set_ylabel("Residual")
ax.plot(df_cleaned["HSBC-rf"], df_residual["error"], 'o')
plt.show()
```

