Final Project

Ojus Sindwani

Aim of the Project

Undertake a financial/accounting analysis of a large set of U.S. companies during 5 key stock market crashes over the past few decades and explain determinants of stock returns.

1. **1987 Stock Market Crash** (Sept 1987 – Dec 1987)
2. **Dot Com Bubble and Crash** (Jan 2000 – Oct 2002)
3. **Great Recession & Financial Crisis** (Jan 2008 – Feb 2009)
4. **The Covid Shock** (Feb 2020 – March 2020)
5. **“The End of Easy Money”**(Jan 2022 – Sep 2022)

```
#Importing the necessary libraries and packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
```

```
#mounting the drive
from google.colab import drive
drive.mount('/gdrive')
```

```
Drive already mounted at /gdrive; to attempt to forcibly remount, call drive.mou
```

Uploading the G Sector Files

```
gsector_1987 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/GSECTOR-1986.csv')
gsector_2001 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/GSECTOR-2000.csv')
gsector_2008 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/GSECTOR-2007.csv')
gsector_2020 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/GSECTOR-2019.csv')
gsector_2022 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/GSECTOR-2021.csv')
```

An explanation of the G-Sectors is given as below:

`10.0` Energy

`15.0` Materials

20.0 Industrials

25.0 Consumer Discretionary

30.0 Consumer Staples

35.0 Health Care

40.0 Financials

45.0 Information Technology

50.0 Communication Services

55.0 Utilities

60.0 Real Estate

```
gsector_1987.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7219 entries, 0 to 7218
Data columns (total 4 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   gvkey    7219 non-null   int64  
 1   fyear    7219 non-null   int64  
 2   TICKER   7219 non-null   object 
 3   gsector  7210 non-null   float64
dtypes: float64(1), int64(2), object(1)
memory usage: 225.7+ KB
```

Uploading files containing financial ratios

These files have been derived from assigment 3

```
ratios_1987 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/financial_ratios_19
ratios_2001 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/financial_ratios_20
ratios_2008 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/financial_ratios_20
ratios_2020 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/financial_ratios_20
ratios_2022 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/financial_ratios_20
```

These files contain various financial ratios for each of the stock market crashes and the average returns for those crashes. The following analysis has already been performed on these datasets in Assignment 3 for the purpose of calculating regression.

1. **`Merging Data`** Merging financial data with average returns for the months leading up to the 5 stock market crashes.

2. **`Treating Missing Values`** Running logistic regression for Current Assets, Current Liabilites, Total Liabilities, Long-Term Debt, Cost of Goods Sold, and Retained Earnings.
3. **`Calculating financial ratios`** Calculating financial ratios like Net Profit Margin, Gross Profit Margin, Liquidity ratio, Leverage, Debt/Total Assets, Asset Turnover Ratio, and ROE.
4. **`Regression`** Regressing various financial ratios on avergae returns and interpreting the results.

```
ratios_1987.head()
```

consol	popsrc	datafmt	TICKER	curcd	act	...	RET	Net Profit Margin	Gross Profit Margin
C	D	STD	ANTQ	USD	13.090	...	-0.126467	0.021841	0.473422
C	D	STD	AIR	USD	168.950	...	-0.058069	0.051514	0.247015
C	D	STD	ACSE	USD	2.385	...	-0.069427	-0.145579	-0.006851
C	D	STD	AECE	USD	18.715	...	0.014061	0.015301	0.360856
C	D	STD	AELNA	USD	52.080	...	-0.099851	0.030642	0.263189

Uploading files containing the FF output for 5 years

These files have been derived from assignment 2

```
ff_1987 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/FF-Output-1983-1987.csv
ff_2001 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/FF-Output-1997-2001.csv
ff_2008 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/FF-Output-2004-2008.csv
ff_2020 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/FF-Output-2016-2020.csv
ff_2022 = pd.read_csv(r'/gdrive/MyDrive/Finance Final Project/FF-Output-2018-2022.csv
```

```
ff_1987.head()
```

	Unnamed: 0	TICKER	const	mktrf	smb	hml
0	1	CNK	0.008628	0.881037	0.581720	-0.000531
1	2	GENE	0.038865	1.157391	1.382452	-1.254665
2	3	GAMA	-0.005431	0.772968	-0.073854	0.189572
3	4	TCOMA	0.029071	0.926688	0.038847	-1.166874

1987 Stock Market Crash

The stock market crash of 1987, also known as Black Monday, occurred on October 19th, 1987. The crash was a sudden and severe drop in stock prices, with the Dow Jones Industrial Average falling by over 22% in one day. The crash was caused by a combination of factors, including overvalued stocks, a decline in international markets, and computerized trading programs.

Merging the three files together

```
# Renaming tic to TICKER in the gsector file
gsector_1987 = gsector_1987.rename(columns={'tic': 'TICKER'})
```

```
# Merge the DataFrames based on the 'TICKER' column
merged_df = pd.merge(gsector_1987, ratios_1987, on='TICKER')
merged_df = pd.merge(merged_df, ff_1987, on='TICKER')
```

```
merged_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 880 entries, 0 to 879
Data columns (total 38 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   gvkey_x                 880 non-null    int64
 1   fyear_x                 880 non-null    int64
 2   TICKER                  880 non-null    object
 3   gsector                 880 non-null    float64
 4   gvkey_y                 880 non-null    int64
 5   datadate                880 non-null    object
 6   fyear_y                 880 non-null    int64
 7   indfmt                  880 non-null    object
 8   consol                  880 non-null    object
 9   popsrc                  880 non-null    object
 10  datafmt                 880 non-null    object
 11  curcd                   880 non-null    object
 12  act                     880 non-null    float64
```

```
 13  at                    880 non-null    float64
 14  che                   880 non-null    float64
 15  cogs                  880 non-null    float64
 16  dltt                  880 non-null    float64
 17  lct                   880 non-null    float64
 18  lt                    880 non-null    float64
 19  ni                    880 non-null    float64
 20  re                    880 non-null    float64
 21  sale                  880 non-null    float64
 22  costat                880 non-null    object
 23  RET                   880 non-null    float64
 24  Net Profit Margin     880 non-null    float64
 25  Gross Profit Margin   880 non-null    float64
 26  Total Debt/Equity     880 non-null    float64
 27  Return on Equity      880 non-null    float64
 28  Liquidity             880 non-null    float64
 29  ROA                   880 non-null    float64
 30  Debt/Total Assets     880 non-null    float64
 31  Asset Turnover Ratio  880 non-null    float64
 32  at_log                880 non-null    float64
 33  Unnamed: 0            880 non-null    int64
 34  const                 880 non-null    float64
 35  mktrf                 880 non-null    float64
 36  smb                   880 non-null    float64
 37  hml                   880 non-null    float64
dtypes: float64(25), int64(5), object(8)
memory usage: 268.1+ KB
```

```
# Getting a sense of industry indicators
sector_counts = merged_df.groupby('gsector').size()
sector_counts
```

```
gsector
10.0     52
15.0     66
20.0    211
25.0    147
30.0     60
35.0     67
40.0     77
45.0    117
50.0     14
55.0     56
60.0     13
dtype: int64
```

Run OLS regression with the following three main categories

1. **Risk Exposures:** This included systematic risk (market risk), SML (Small Minus Big), which is the size risk, and HML (High minus Low), which is the Value vs Growth risk.

2. **Financial Ratios:** This includes various financial ratios, including but not limited to Liquidity, Leverage, Asset Turnover Ratio, Net and Gross Profit Margin, ROE and ROA.
3. **Industry Indicators:** This indicates which industries had the highest amount of affect on stock returns.

1987- Regression 1

Regression using all 3 Categories

```
# General regression code:
X1 = merged_df[['Gross Profit Margin', 'Debt/Total Assets', 'Asset Turnover Ratio', '
X2 = pd.get_dummies(merged_df['gsector'], prefix='sector')
X3 = merged_df[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2, X3], axis=1)

y = merged_df['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.242
Model:                            OLS   Adj. R-squared:                  0.227
Method:                 Least Squares   F-statistic:                     16.20
Date:                Sat, 06 May 2023   Prob (F-statistic):           1.39e-41
Time:                        12:27:41   Log-Likelihood:                 1397.9
No. Observations:                 880   AIC:                            -2760.
Df Residuals:                     862   BIC:                            -2674.
Df Model:                          17
Covariance Type:            nonrobust
================================================================================
                           coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                   -0.0087      0.008     -1.151      0.250      -0.024
Gross Profit Margin      0.0145      0.006      2.488      0.013       0.003
Debt/Total Assets       -0.0326      0.009     -3.500      0.000      -0.051
Asset Turnover Ratio     0.0066      0.003      2.425      0.016       0.001
at_log                   0.0019      0.001      1.858      0.064      -0.000
sector_10.0             -0.0190      0.007     -2.734      0.006      -0.033
sector_15.0              0.0180      0.006      2.921      0.004       0.006
sector_20.0              0.0046      0.004      1.138      0.255      -0.003
sector_25.0             -0.0103      0.005     -2.176      0.030      -0.020
```

```
sector_30.0            -0.0032      0.007     -0.438      0.661      -0.018
sector_35.0            -0.0048      0.006     -0.783      0.434      -0.017
sector_40.0            -0.0019      0.007     -0.288      0.773      -0.015
sector_45.0            -0.0142      0.005     -2.610      0.009      -0.025
sector_50.0            -0.0012      0.012     -0.096      0.924      -0.026
sector_55.0             0.0046      0.007      0.646      0.518      -0.009
sector_60.0             0.0186      0.013      1.443      0.149      -0.007
mktrf                  -0.0552      0.005    -11.793      0.000      -0.064
smb                    -0.0088      0.002     -3.987      0.000      -0.013
hml                     0.0031      0.003      1.203      0.229      -0.002
==============================================================================
Omnibus:                       39.819   Durbin-Watson:                   1.871
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              120.916
Skew:                          -0.040   Prob(JB):                     5.54e-27
Kurtosis:                       4.814   Cond. No.                     8.47e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The smallest eigenvalue is 4.49e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

The regression model shows the relationship between the dependent variable, RET(stock returns), and various independent variables. The R-squared value suggests that 24.2% of the variance in the dependent variable is explained by the model, with an adjusted R-Squared. The coefficients of the independent variables show their impact on the dependent variable, with Gross Profit Margin, Debt/Total Assets, Asset Turnover Ratio, sector_10.0, sector_15.0, sector_25.0, sector_45.0, mktrf, and smb having statistically significant impacts. The model has some potential issues with multicollinearity and singularity, therefore, further regressions are performed to reduce such issues.

1987- Regression 2

Regression using only financial ratios and industry indicators

```
# General regression code:
X1 = merged_df[['Gross Profit Margin', 'Debt/Total Assets', 'Asset Turnover Ratio', '
X2 = pd.get_dummies(merged_df['gsector'], prefix='sector')

X = pd.concat([X1, X2], axis=1)

y = merged_df['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()
```

```
# list regression output
print(model.summary())

                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.109
Model:                            OLS   Adj. R-squared:                  0.095
Method:                 Least Squares   F-statistic:                     7.574
Date:                Sat, 06 May 2023   Prob (F-statistic):           3.46e-15
Time:                        12:27:41   Log-Likelihood:                 1326.7
No. Observations:                 880   AIC:                            -2623.
Df Residuals:                     865   BIC:                            -2552.
Df Model:                          14
Covariance Type:            nonrobust
================================================================================
                           coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                   -0.0507      0.007     -7.376      0.000      -0.064
Gross Profit Margin      0.0193      0.006      3.086      0.002       0.007
Debt/Total Assets       -0.0501      0.010     -5.156      0.000      -0.069
Asset Turnover Ratio     0.0070      0.003      2.406      0.016       0.001
at_log                   0.0013      0.001      1.436      0.151      -0.000
sector_10.0             -0.0306      0.007     -4.205      0.000      -0.045
sector_15.0              0.0101      0.007      1.536      0.125      -0.003
sector_20.0             -0.0050      0.004     -1.172      0.242      -0.013
sector_25.0             -0.0218      0.005     -4.394      0.000      -0.032
sector_30.0             -0.0012      0.008     -0.152      0.879      -0.017
sector_35.0             -0.0139      0.007     -2.122      0.034      -0.027
sector_40.0             -0.0034      0.007     -0.478      0.632      -0.017
sector_45.0             -0.0298      0.005     -5.659      0.000      -0.040
sector_50.0             -0.0053      0.013     -0.396      0.692      -0.032
sector_55.0              0.0281      0.007      3.885      0.000       0.014
sector_60.0              0.0220      0.014      1.575      0.116      -0.005
==============================================================================
Omnibus:                       35.762   Durbin-Watson:                   1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               90.834
Skew:                          -0.140   Prob(JB):                     1.89e-20
Kurtosis:                       4.549   Cond. No.                     7.20e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The smallest eigenvalue is 6.03e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

The OLS regression model suggests that Gross Profit Margin, Debt/Total Assets, Asset Turnover Ratio, and some industry sectors have significant effects on the stock returns of the companies in the sample. However, the overall explanatory power of the model is weak, as indicated by the low R-squared value of 10.9%, and an adjusted R-Squared value 9.5%. Additionally, there may be issues of multicollinearity, and the design matrix may be singular, which could affect the reliability of the coefficient estimates. As risk exposures were excluded, the R2 dropped significantly indicating that 13.3% variance in the model is explained by risk exposures.

1987- Regression 3

Regression using only financial ratios and risk exposures

```
# General regression code:
X1 = merged_df[['Gross Profit Margin', 'Debt/Total Assets', 'Asset Turnover Ratio', '
X2 = merged_df[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2], axis=1)

y = merged_df['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.213
Model:                            OLS   Adj. R-squared:                  0.207
Method:                 Least Squares   F-statistic:                     33.81
Date:                Sat, 06 May 2023   Prob (F-statistic):           8.75e-42
Time:                        12:27:41   Log-Likelihood:                 1381.5
No. Observations:                 880   AIC:                            -2747.
Df Residuals:                     872   BIC:                            -2709.
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
                           coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                   -0.0114      0.008     -1.431      0.153      -0.027
Gross Profit Margin      0.0130      0.006      2.241      0.025       0.002
Debt/Total Assets       -0.0309      0.009     -3.527      0.000      -0.048
Asset Turnover Ratio     0.0069      0.002      3.051      0.002       0.002
at_log                   0.0025      0.001      2.531      0.012       0.001
mktrf                   -0.0586      0.004    -13.411      0.000      -0.067
smb                     -0.0097      0.002     -4.595      0.000      -0.014
hml                      0.0045      0.002      1.947      0.052   -3.57e-05
==============================================================================
Omnibus:                       36.429   Durbin-Watson:                   1.895
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              102.648
Skew:                          -0.059   Prob(JB):                     5.13e-23
Kurtosis:                       4.669   Cond. No.                         33.5
==============================================================================

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

This is a multiple linear regression with seven independent variables. The dependent variable is the RET (stock returns) and the independent variables are Gross Profit Margin, Debt/Total Assets, Asset Turnover Ratio, at_log, mktrf, smb, and hml. The R-squared value is 0.213, indicating that the model explains 21.3% of the variance in the dependent variable, with an adjusted R-Squareded of 20.7%. The t-tests and p-values indicate the statistical significance of the coefficients. The regression suggests that Gross Profit Margin, Debt/Total Assets, Asset Turnover Ratio, at_log, mktrf, and smb have significant impact on the return, while hml is borderline significant. This regression dropped the financial ratios, indicating that 2.9% of the variance was explained by the financial ratios.

1987- Regression 4

Regression using only risk exposure

```
#General regression code:
X = merged_df[['mktrf', 'smb', 'hml']]
y = merged_df['RET']
# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

#list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.191
Model:                            OLS   Adj. R-squared:                  0.188
Method:                 Least Squares   F-statistic:                     68.74
Date:                Sat, 06 May 2023   Prob (F-statistic):           6.36e-40
Time:                        12:27:41   Log-Likelihood:                 1368.9
No. Observations:                 880   AIC:                            -2730.
Df Residuals:                     876   BIC:                            -2711.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0006      0.005     -0.118      0.906      -0.010       0.009
mktrf         -0.0581      0.004    -13.287      0.000      -0.067      -0.050
smb           -0.0120      0.002     -6.493      0.000      -0.016      -0.008
hml            0.0015      0.002      0.649      0.516      -0.003       0.006
==============================================================================
Omnibus:                       45.581   Durbin-Watson:                   1.911
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              129.833
```

```
Skew:                          -0.190   Prob(JB):                     6.41e-29
Kurtosis:                       4.843   Cond. No.                         6.13
==============================================================================

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

This is a regression model with three independent variables, namely mktrf, smb, and hml for the risk exposures, and a dependent variable RET (stock returns). The model has an R-squared value of 0.191, indicating that the independent variables explain approximately 19.1% of the variance in the dependent variable, with an adjuted R-Squared of 18.8%. The coefficients of the independent variables suggest that mktrf and smb have negative relationships with RET, while hml has a positive relationship. Overall, the model can be used to predict the value of RET based on the values of the independent variables.

1987- Regression 5

Regression using only industry indicators

```
X_all_sectors = pd.get_dummies(merged_df['gsector'], prefix='sector')
y = merged_df['RET']
# Perform the regression using statsmodels
model_all_sectors = sm.OLS(y, X_all_sectors).fit()

# Identify the industry with the coefficient value closest to zero
reference_sector = model_all_sectors.params.abs().idxmin()

print(reference_sector)
```

```
sector_60.0
```

```
# Create dummy variables for the industry sectors
X_all_sectors = pd.get_dummies(merged_df['gsector'], prefix='sector')
y = merged_df['RET']

# Exclude the reference sector's dummy variable from the independent variables
X = X_all_sectors.drop(columns=[reference_sector])

# Add a constant to the independent variables
X = sm.add_constant(X)

# Perform the regression using statsmodels without the reference sector's dummy varia
model_5 = sm.OLS(y, X).fit()
```

```
# Print the regression output
print(model_5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.068
Model:                            OLS   Adj. R-squared:                  0.057
Method:                 Least Squares   F-statistic:                     6.311
Date:                Sat, 06 May 2023   Prob (F-statistic):           2.15e-09
Time:                        12:30:28   Log-Likelihood:                 1306.7
No. Observations:                 880   AIC:                            -2591.
Df Residuals:                     869   BIC:                            -2539.
Df Model:                          10
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const          -0.0346      0.015     -2.259      0.024      -0.065      -0.005
sector_10.0    -0.0533      0.017     -3.115      0.002      -0.087      -0.020
sector_15.0    -0.0079      0.017     -0.472      0.637      -0.041       0.025
sector_20.0    -0.0245      0.016     -1.554      0.121      -0.055       0.006
sector_25.0    -0.0406      0.016     -2.543      0.011      -0.072      -0.009
sector_30.0    -0.0124      0.017     -0.736      0.462      -0.046       0.021
sector_35.0    -0.0288      0.017     -1.720      0.086      -0.062       0.004
sector_40.0    -0.0406      0.017     -2.453      0.014      -0.073      -0.008
sector_45.0    -0.0443      0.016     -2.745      0.006      -0.076      -0.013
sector_50.0    -0.0296      0.021     -1.394      0.164      -0.071       0.012
sector_55.0    -0.0013      0.017     -0.074      0.941      -0.035       0.032
==============================================================================
Omnibus:                       46.203   Durbin-Watson:                   1.975
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              112.406
Skew:                          -0.268   Prob(JB):                     3.90e-25
Kurtosis:                       4.667   Cond. No.                         29.4
==============================================================================

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

The regression model examines the relationship between the stock returns (dependent variable) and industry sectors (independent variables) using Ordinary Least Squares (OLS) method.

The initial model includes dummy variables for all industry sectors, and the industry with the coefficient value closest to zero is identified as the reference sector. Then, a new model is created by excluding the reference sector's dummy variable from the independent variables.

The final model shows that the adjusted R-squared value is 0.057, indicating that the independent variables can explain 5.7% of the variation in the dependent variable.

Among the remaining industry sectors, the coefficients for sectors 10, 25, 40, and 45 are statistically significant at the 5% level, indicating that these sectors have a significant impact on stock returns. The coefficient for sector 15 is not statistically significant. The negative coefficients for all significant sectors suggest that stocks in these sectors have a negative impact on returns. The intercept coefficient is also negative and statistically significant, indicating that the expected return is negative even when all the independent variables are zero.

1987- Conclusion

```
Thus we note that the maximum variability in the data is explained by the risk exposure
```

This too in this case, by the systematic risk and the size risk(Small minus Big).

Dot Com Bubble and Crash

The dot com bubble crash of 2001 was a significant market downturn in the technology sector, fueled by the excessive speculation and overvaluation of internet-related stocks. This bubble began to burst in March 2000, leading to a sharp decline in stock prices of companies in the technology industry. By 2001, many dot com companies went bankrupt, resulting in massive job losses and a decline in overall economic growth. The crash was a painful reminder of the risks associated with speculative bubbles and market irrationality.

Merging the three files together

```
# Renaming tic to TICKER in the gsector file
gsector_2001 = gsector_2001.rename(columns={'tic': 'TICKER'})
```

```
# Merge the DataFrames based on the 'TICKER' column
merged_df_2001 = pd.merge(gsector_2001, ratios_2001, on='TICKER')
merged_df_2001 = pd.merge(merged_df_2001, ff_2001, on='TICKER')
```

```
merged_df_2001.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1973 entries, 0 to 1972
Data columns (total 38 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   gvkey_x               1973 non-null   int64
 1   fyear_x               1973 non-null   int64
 2   TICKER                1973 non-null   object
 3   gsector               1972 non-null   float64
```

```
 4   gvkey_y               1973 non-null   int64
 5   datadate              1973 non-null   object
 6   fyear_y               1973 non-null   int64
 7   indfmt                1973 non-null   object
 8   consol                1973 non-null   object
 9   popsrc                1973 non-null   object
 10  datafmt               1973 non-null   object
 11  curcd                 1973 non-null   object
 12  act                   1973 non-null   float64
 13  at                    1973 non-null   float64
 14  che                   1973 non-null   float64
 15  cogs                  1973 non-null   float64
 16  dltt                  1973 non-null   float64
 17  lct                   1973 non-null   float64
 18  lt                    1973 non-null   float64
 19  ni                    1973 non-null   float64
 20  re                    1973 non-null   float64
 21  sale                  1973 non-null   float64
 22  costat                1973 non-null   object
 23  RET                   1973 non-null   float64
 24  at_log                1973 non-null   float64
 25  Net Profit Margin     1973 non-null   float64
 26  Gross Profit Margin   1973 non-null   float64
 27  Total Debt/Equity     1973 non-null   float64
 28  Return on Equity      1973 non-null   float64
 29  Liquidity             1973 non-null   float64
 30  ROA                   1973 non-null   float64
 31  Debt/Total Assets     1973 non-null   float64
 32  Asset Turnover Ratio  1973 non-null   float64
 33  Unnamed: 0            1973 non-null   int64
 34  const                 1973 non-null   float64
 35  mktrf                 1973 non-null   float64
 36  smb                   1973 non-null   float64
 37  hml                   1973 non-null   float64
dtypes: float64(25), int64(5), object(8)
memory usage: 601.1+ KB
```

Run OLS regression with the following three main categories

1. **Risk Exposures:** This included systematic risk (market risk), SML (Small Minus Big), which is the size risk, and HML (High minus Low), which is the Value vs Growth risk.
2. **Financial Ratios:** This includes various financial ratios, including but not limited to Liquidity, Leverage, Asset Turnover Ratio, Net and Gross Profit Margin, ROE and ROA.
3. **Industry Indicators:** This indicates which industries had the highest amount of affect on stock returns.

2001- Regression 1

Regression using industry indicators, financial ratios and risk exposures.

```
# General regression code:
X1 = merged_df_2001[['Net Profit Margin', 'Gross Profit Margin', 'ROA', 'Debt/Total A
X2 = pd.get_dummies(merged_df_2001['gsector'], prefix='sector')
X3 = merged_df_2001[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2, X3], axis=1)

y = merged_df_2001['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.352
Model:                            OLS   Adj. R-squared:                  0.346
Method:                 Least Squares   F-statistic:                     55.87
Date:                Sat, 06 May 2023   Prob (F-statistic):          2.62e-168
Time:                        12:27:41   Log-Likelihood:                 3203.7
No. Observations:                1973   AIC:                            -6367.
Df Residuals:                    1953   BIC:                            -6256.
Df Model:                          19
Covariance Type:            nonrobust
===============================================================================
                           coef    std err          t      P>|t|      [0.025
-------------------------------------------------------------------------------
const                    0.0180      0.048      0.374      0.708      -0.076
Net Profit Margin       -0.0027      0.002     -1.423      0.155      -0.006
Gross Profit Margin      0.0057      0.003      2.244      0.025       0.001
ROA                     -0.0017      0.008     -0.228      0.819      -0.017
Debt/Total Assets        0.0071      0.005      1.316      0.188      -0.004
Asset Turnover Ratio     0.0039      0.002      2.297      0.022       0.001
sector_10.0             -0.0091      0.048     -0.188      0.851      -0.104
sector_15.0             -0.0208      0.048     -0.431      0.667      -0.115
sector_20.0             -0.0253      0.048     -0.526      0.599      -0.120
sector_25.0             -0.0111      0.048     -0.232      0.817      -0.106
sector_30.0             -0.0109      0.048     -0.225      0.822      -0.106
sector_35.0             -0.0069      0.048     -0.143      0.887      -0.101
sector_40.0             -0.0111      0.048     -0.231      0.817      -0.105
sector_45.0             -0.0235      0.048     -0.488      0.626      -0.118
sector_50.0             -0.0156      0.049     -0.320      0.749      -0.111
sector_55.0             -0.0162      0.048     -0.335      0.738      -0.111
sector_60.0             -0.0207      0.048     -0.428      0.669      -0.116
mktrf                   -0.0225      0.002    -10.815      0.000      -0.027
smb                     -0.0008      0.001     -0.584      0.559      -0.004
hml                      0.0350      0.002     22.916      0.000       0.032
==============================================================================
```

```
Omnibus:                      341.383   Durbin-Watson:                   2.051
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2043.791
Skew:                           0.671   Prob(JB):                         0.00
Kurtosis:                       7.802   Cond. No.                         312.
==============================================================================

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

The regression model examines the relationship between stock returns (RET) and various independent variables. The R-squared value of 0.352 indicates that 35.2% of the variance in the dependent variable is explained by the model, with an adjusted R-squared of 0.346. The coefficients of the independent variables indicate that Gross Profit Margin, Debt/Total Assets, Asset Turnover Ratio, sector_10.0, sector_15.0, sector_25.0, sector_45.0, mktrf, and hml have a statistically significant impact on the dependent variable. However, the model may have some issues with multicollinearity and singularity, and further regressions are required to address these issues.

2001- Regression 2

Regression using only industry indicators and risk exposure

```
# General regression code:
X1 = pd.get_dummies(merged_df_2001['gsector'], prefix='sector')
X2 = merged_df_2001[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2], axis=1)

y = merged_df_2001['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.348
Model:                            OLS   Adj. R-squared:                  0.343
Method:                 Least Squares   F-statistic:                     74.58
Date:                Sat, 06 May 2023   Prob (F-statistic):          4.29e-170
Time:                        12:27:41   Log-Likelihood:                 3197.1
No. Observations:                1973   AIC:                            -6364.
Df Residuals:                    1958   BIC:                            -6280.
Df Model:                          14
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0275      0.048      0.572      0.567      -0.067       0.122
sector_10.0   -0.0093      0.048     -0.193      0.847      -0.104       0.085
sector_15.0   -0.0210      0.048     -0.435      0.663      -0.116       0.074
sector_20.0   -0.0239      0.048     -0.496      0.620      -0.118       0.071
sector_25.0   -0.0088      0.048     -0.183      0.855      -0.103       0.086
sector_30.0   -0.0078      0.048     -0.162      0.871      -0.103       0.087
sector_35.0   -0.0084      0.048     -0.174      0.862      -0.103       0.086
sector_40.0   -0.0118      0.048     -0.246      0.806      -0.106       0.083
sector_45.0   -0.0217      0.048     -0.450      0.653      -0.116       0.073
sector_50.0   -0.0152      0.049     -0.312      0.755      -0.111       0.080
sector_55.0   -0.0173      0.048     -0.358      0.721      -0.112       0.078
sector_60.0   -0.0236      0.048     -0.487      0.627      -0.119       0.071
mktrf         -0.0229      0.002    -11.007      0.000      -0.027      -0.019
smb           -0.0016      0.001     -1.146      0.252      -0.004       0.001
hml            0.0358      0.001     24.337      0.000       0.033       0.039
==============================================================================
Omnibus:                      372.905   Durbin-Watson:                   2.049
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2388.177
Skew:                           0.728   Prob(JB):                         0.00
Kurtosis:                       8.190   Cond. No.                         255.
==============================================================================

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

This is an OLS regression summary, showing the results of a regression model that estimates the relationship between the dependent variable "RET" and the independent variables "const", "sector_10.0", "sector_15.0", "sector_20.0", "sector_25.0", "sector_30.0", "sector_35.0", "sector_40.0", "sector_45.0", "sector_50.0", "sector_55.0", "sector_60.0", "mktrf", "smb", and "hml".

The R-squared value of 0.348 indicates that 34.8% of the variance in the dependent variable is explained by the independent variables. The adjusted R-squared value of 0.343 is slightly lower, indicating that the model has not overfit the data.

Removing financial ratios from the regression reduced 0.4% explanatory power from the regression.

2001- Regression 3

Regression using only risk exposure and financial ratios

```
# General regression code:
X1 = merged_df_2001[['Net Profit Margin', 'Gross Profit Margin', 'ROA', 'Debt/Total A
X2 = merged_df_2001[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2], axis=1)
```

```
y = merged_df_2001['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.340
Model:                            OLS   Adj. R-squared:                  0.338
Method:                 Least Squares   F-statistic:                     126.7
Date:                Sat, 06 May 2023   Prob (F-statistic):          2.21e-171
Time:                        12:27:41   Log-Likelihood:                 3186.0
No. Observations:                1973   AIC:                            -6354.
Df Residuals:                    1964   BIC:                            -6304.
Df Model:                           8
Covariance Type:            nonrobust
================================================================================
                          coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                   0.0041      0.004      1.001      0.317      -0.004
Net Profit Margin      -0.0028      0.002     -1.511      0.131      -0.006
Gross Profit Margin     0.0051      0.003      2.026      0.043       0.000
ROA                    -0.0013      0.008     -0.166      0.868      -0.016
Debt/Total Assets       0.0086      0.005      1.810      0.070      -0.001
Asset Turnover Ratio    0.0027      0.001      1.977      0.048    2.12e-05
mktrf                  -0.0244      0.002    -12.750      0.000      -0.028
smb                    -0.0013      0.001     -0.956      0.339      -0.004
hml                     0.0363      0.001     27.187      0.000       0.034
==============================================================================
Omnibus:                      328.717   Durbin-Watson:                   2.050
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1877.112
Skew:                           0.654   Prob(JB):                         0.00
Kurtosis:                       7.596   Cond. No.                         14.9
==============================================================================

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

This is an OLS regression summary, showing the results of a regression model that estimates the relationship between the dependent variable "RET" and the independent variables "const", "Net Profit Margin", "Gross Profit Margin", "ROA", "Debt/Total Assets", "Asset Turnover Ratio", "mktrf", "smb", and "hml".

The R-squared value of 0.340 indicates that 34.0% of the variance in the dependent variable is explained by the independent variables. The adjusted R-squared value of 0.338 is slightly lower, indicating that the model has not overfit the data.

Removing gsector indicators reduced the explanatory power by only 0.8%.

2001- Regression 4

Regression using only risk exposure

```
#General regression code:
X = merged_df_2001[['mktrf', 'smb', 'hml']]
y = merged_df_2001['RET']
# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

#list regression output
print(model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.337
Model:                            OLS   Adj. R-squared:                  0.336
Method:                 Least Squares   F-statistic:                     333.4
Date:                Sat, 06 May 2023   Prob (F-statistic):          4.81e-175
Time:                        12:27:41   Log-Likelihood:                 3180.7
No. Observations:                1973   AIC:                            -6353.
Df Residuals:                    1969   BIC:                            -6331.
Df Model:                           3                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0134      0.002      6.440      0.000       0.009       0.017
mktrf         -0.0244      0.002    -12.906      0.000      -0.028      -0.021
smb           -0.0020      0.001     -1.518      0.129      -0.005       0.001
hml            0.0369      0.001     29.902      0.000       0.034       0.039
==============================================================================
Omnibus:                      355.132   Durbin-Watson:                   2.048
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2138.395
Skew:                           0.703   Prob(JB):                         0.00
Kurtosis:                       7.902   Cond. No.                         3.90
==============================================================================

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

This is a multiple linear regression model with three independent variables: mktrf, smb, and hml. The dependent variable is RET. The model has a significant overall fit with an R-squared of 0.337, meaning 33.7% of the variance in RET is explained by the three independent variables. The coefficients for mktrf and hml are significant at the 0.05 level, while the coefficient for smb is not significant at the 0.05 level. The model's intercept is also significant at the 0.05 level. The model assumes that the covariance matrix of the errors is correctly specified.

2001- Regression 5

This regression only takes into account the industry indicators.

```
X_all_sectors = pd.get_dummies(merged_df_2001['gsector'], prefix='sector')
y = merged_df_2001['RET']
# Perform the regression using statsmodels
model_all_sectors = sm.OLS(y, X_all_sectors).fit()

# Identify the industry with the coefficient value closest to zero
reference_sector = model_all_sectors.params.abs().idxmin()

print(reference_sector)
```

```
sector_35.0
```

```
# Create dummy variables for the industry sectors
X_all_sectors = pd.get_dummies(merged_df_2001['gsector'], prefix='sector')
y = merged_df_2001['RET']

# Exclude the reference sector's dummy variable from the independent variables
X = X_all_sectors.drop(columns=[reference_sector])

# Add a constant to the independent variables
X = sm.add_constant(X)

# Perform the regression using statsmodels without the reference sector's dummy varia
model_5 = sm.OLS(y, X).fit()

# Print the regression output
print(model_5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.148
Model:                            OLS   Adj. R-squared:                  0.144
Method:                 Least Squares   F-statistic:                     34.11
Date:                Sat, 06 May 2023   Prob (F-statistic):           9.67e-62
Time:                        12:35:50   Log-Likelihood:                 2933.6
No. Observations:                1973   AIC:                            -5845.
Df Residuals:                    1962   BIC:                            -5784.
Df Model:                          10
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
```

```
==============================================================================
const            0.0004      0.004      0.115      0.909      -0.007       0.008
sector_10.0      0.0219      0.007      3.115      0.002       0.008       0.036
sector_15.0      0.0144      0.006      2.214      0.027       0.002       0.027
sector_20.0      0.0037      0.005      0.758      0.448      -0.006       0.013
sector_25.0      0.0183      0.005      3.730      0.000       0.009       0.028
sector_30.0      0.0218      0.007      3.296      0.001       0.009       0.035
sector_40.0      0.0244      0.005      5.333      0.000       0.015       0.033
sector_45.0     -0.0412      0.005     -8.808      0.000      -0.050      -0.032
sector_50.0     -0.0125      0.010     -1.311      0.190      -0.031       0.006
sector_55.0      0.0204      0.008      2.493      0.013       0.004       0.036
sector_60.0      0.0110      0.008      1.348      0.178      -0.005       0.027
==============================================================================
Omnibus:                      169.744   Durbin-Watson:                   2.021
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              862.054
Skew:                           0.224   Prob(JB):                    6.42e-188
Kurtosis:                       6.207   Cond. No.                         11.5
==============================================================================

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

The regression output shows the results of a linear regression model that aims to predict the return (RET) of companies based on their industry sectors. The model uses dummy variables to represent the different industry sectors, and it excludes one reference sector to avoid multicollinearity.

The R-squared value of the model is 0.148, indicating that only 14.8% of the variation in the dependent variable is explained by the independent variables. The adjusted R-squared value is slightly lower, which suggests that the model might be overfitting the data.

Some of the coefficients have p-values higher than 0.05, indicating that they are not statistically significant. In particular, the coefficient for sector_50.0 has a p-value of 0.190, which suggests that it is not a good predictor of the dependent variable.

The coefficient for the reference sector (sector_20.0) is not shown in the output, as it was excluded from the model. The constant term is also included in the model, with a coefficient of 0.0004. Overall, the model suggests that the industry sectors of companies have a significant but relatively weak relationship with their returns.

2001- Conclusion

Therefore, the risk exposures explain the most variance in the stock returns with an explnantory power of 33.7%

Great Recession and Financial Crisis

The Great Recession of 2008 was a severe economic downturn that lasted from late 2007 to mid-2009, triggered by a collapse in the housing market and the subprime mortgage industry in the United States. It caused widespread job losses, foreclosures, and a significant decline in economic activity across the globe, leading to a financial crisis that affected the banking sector and financial markets worldwide. The recession had a lasting impact on many economies and societies, and policymakers implemented a range of measures to prevent another such event from occurring.

Merging the three files together

Run OLS regression with the following three main categories

1. **Risk Exposures:** This included systematic risk (market risk), SML (Small Minus Big), which is the size risk, and HML (High minus Low), which is the Value vs Growth risk.
2. **Financial Ratios:** This includes various financial ratios, including but not limited to Liquidity, Leverage, Asset Turnover Ratio, Net and Gross Profit Margin, ROE and ROA.
3. **Industry Indicators:** This indicates which industries had the highest amount of affect on stock returns.

```
# Renaming tic to TICKER in the gsector file
gsector_2008 = gsector_2008.rename(columns={'tic': 'TICKER'})


# Merge the DataFrames based on the 'TICKER' column
merged_df_2008 = pd.merge(gsector_2008, ratios_2008, on='TICKER')
merged_df_2008 = pd.merge(merged_df_2008, ff_2008, on='TICKER')
```

```
merged_df_2008.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2440 entries, 0 to 2439
Data columns (total 38 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   gvkey_x               2440 non-null   int64
 1   fyear_x               2440 non-null   int64
 2   TICKER                2440 non-null   object
 3   gsector               2440 non-null   float64
 4   gvkey_y               2440 non-null   int64
 5   datadate              2440 non-null   object
 6   fyear_y               2440 non-null   int64
 7   indfmt                2440 non-null   object
```

```
 8   consol                 2440 non-null   object
 9   popsrc                 2440 non-null   object
 10  datafmt                2440 non-null   object
 11  curcd                  2440 non-null   object
 12  act                    2440 non-null   float64
 13  at                     2440 non-null   float64
 14  che                    2440 non-null   float64
 15  cogs                   2440 non-null   float64
 16  dltt                   2440 non-null   float64
 17  lct                    2440 non-null   float64
 18  lt                     2440 non-null   float64
 19  ni                     2440 non-null   float64
 20  re                     2440 non-null   float64
 21  sale                   2440 non-null   float64
 22  costat                 2440 non-null   object
 23  RET                    2440 non-null   float64
 24  Net Profit Margin      2440 non-null   float64
 25  Gross Profit Margin    2440 non-null   float64
 26  Total Debt/Equity      2440 non-null   float64
 27  Return on Equity       2440 non-null   float64
 28  Liquidity              2440 non-null   float64
 29  ROA                    2440 non-null   float64
 30  Debt/Total Assets      2440 non-null   float64
 31  Asset Turnover Ratio   2440 non-null   float64
 32  at_log                 2440 non-null   float64
 33  Unnamed: 0             2440 non-null   int64
 34  const                  2440 non-null   float64
 35  mktrf                  2440 non-null   float64
 36  smb                    2440 non-null   float64
 37  hml                    2440 non-null   float64
dtypes: float64(25), int64(5), object(8)
memory usage: 743.4+ KB
```

2008- Regression 1

Regression using financial ratios, risk exposures and industry indicators

```
# General regression code:
X1 = merged_df_2008[['Total Debt/Equity', 'ROA',  'Asset Turnover Ratio']]
X2 = pd.get_dummies(merged_df_2008['gsector'], prefix='sector')
X3 = merged_df_2008[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2, X3], axis=1)

y = merged_df_2008['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()
```

```
# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.698
Model:                            OLS   Adj. R-squared:                  0.696
Method:                 Least Squares   F-statistic:                     350.8
Date:                Sat, 06 May 2023   Prob (F-statistic):               0.00
Time:                        12:27:42   Log-Likelihood:                 3446.3
No. Observations:                2440   AIC:                            -6859.
Df Residuals:                    2423   BIC:                            -6760.
Df Model:                          16
Covariance Type:            nonrobust
================================================================================
                           coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                    0.0156      0.003      4.883      0.000       0.009
Total Debt/Equity       -0.0001      0.000     -0.328      0.743      -0.001
ROA                      0.0512      0.009      5.783      0.000       0.034
Asset Turnover Ratio    -0.0017      0.002     -0.863      0.388      -0.006
sector_10.0             -0.0392      0.005     -7.901      0.000      -0.049
sector_15.0             -0.0044      0.005     -0.861      0.389      -0.014
sector_20.0              0.0019      0.004      0.538      0.591      -0.005
sector_25.0              0.0018      0.004      0.470      0.638      -0.006
sector_30.0              0.0062      0.006      1.074      0.283      -0.005
sector_35.0             -0.0014      0.004     -0.403      0.687      -0.008
sector_40.0              0.0189      0.004      4.744      0.000       0.011
sector_45.0              0.0012      0.003      0.373      0.709      -0.005
sector_50.0              0.0058      0.007      0.856      0.392      -0.007
sector_55.0              0.0155      0.006      2.433      0.015       0.003
sector_60.0              0.0094      0.006      1.531      0.126      -0.003
mktrf                   -0.1322      0.002    -65.514      0.000      -0.136
smb                     -0.0200      0.001    -15.081      0.000      -0.023
hml                     -0.0173      0.001    -13.158      0.000      -0.020
==============================================================================
Omnibus:                      146.830   Durbin-Watson:                   1.974
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              627.175
Skew:                           0.021   Prob(JB):                    6.47e-137
Kurtosis:                       5.483   Cond. No.                     2.27e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The smallest eigenvalue is 1.27e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

The OLS regression model examines the relationship between stock returns (RET) and various independent variables, including financial ratios and sector dummies. The model's R-squared value indicates that about 69.8% of the variation in the dependent variable is explained by the model, with an adjusted R-squared of 69.6%. The coefficients of the independent variables show that only ROA, sector_10.0, sector_40.0, sector_55.0, mktrf, smb, and hml have statistically significant impacts on stock returns. The model may suffer from multicollinearity and singularity issues, indicating a need for further analysis as done below.

2008- Regression 2

Regression using financial ratios and industry indicators

```
# General regression code:
X1 = merged_df_2008[['Total Debt/Equity', 'ROA',  'Asset Turnover Ratio']]
X2 = pd.get_dummies(merged_df_2008['gsector'], prefix='sector')

X = pd.concat([X1, X2], axis=1)

y = merged_df_2008['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.130
Model:                            OLS   Adj. R-squared:                  0.126
Method:                 Least Squares   F-statistic:                     27.95
Date:                Sat, 06 May 2023   Prob (F-statistic):           1.38e-64
Time:                        12:27:42   Log-Likelihood:                 2153.9
No. Observations:                2440   AIC:                            -4280.
Df Residuals:                    2426   BIC:                            -4199.
Df Model:                          13
Covariance Type:            nonrobust
================================================================================
                           coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                   -0.1292      0.004    -34.013      0.000      -0.137
Total Debt/Equity     8.641e-05      0.001      0.135      0.893      -0.001
ROA                      0.1095      0.015      7.379      0.000       0.080
Asset Turnover Ratio    -0.0016      0.003     -0.468      0.640      -0.008
sector_10.0             -0.0624      0.008     -7.503      0.000      -0.079
sector_15.0             -0.0381      0.009     -4.448      0.000      -0.055
sector_20.0             -0.0222      0.006     -3.794      0.000      -0.034
sector_25.0             -0.0351      0.006     -5.648      0.000      -0.047
sector_30.0              0.0321      0.010      3.310      0.001       0.013
sector_35.0             -0.0034      0.006     -0.558      0.577      -0.015
sector_40.0              0.0404      0.007      6.167      0.000       0.028
sector_45.0             -0.0387      0.005     -7.645      0.000      -0.049
sector_50.0             -0.0387      0.011     -3.380      0.001      -0.061
sector_55.0              0.0698      0.011      6.531      0.000       0.049
```

```
sector_60.0              -0.0329      0.010     -3.228      0.001      -0.053
==============================================================================
Omnibus:                      146.074   Durbin-Watson:                   1.955
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              602.311
Skew:                          -0.083   Prob(JB):                    1.62e-131
Kurtosis:                       5.428   Cond. No.                     1.86e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The smallest eigenvalue is 1.86e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

This is the output of a multiple linear regression analysis where the dependent variable is RET (return) and 13 independent variables have been used in the model.

The adjusted R-squared value for the model is 0.126, indicating that approximately 13% of the variation in the dependent variable can be explained by the independent variables included in the model. The adjusted R-squared value takes into account the number of independent variables in the model and adjusts the R-squared value accordingly. In this case, the adjusted R-squared value is slightly lower than the R-squared value of 0.130, which is the proportion of the variation in the dependent variable explained by the model.

Based on the output, it appears that Total Debt/Equity and Asset Turnover Ratio are not statistically significant as their p-values are greater than 0.05. The other 11 independent variables, including ROA (return on assets) and the sector variables, appear to be statistically significant in explaining the variation in the dependent variable.

The significantly low R-Squared suggests that most variability in the data was explained by the risk exposures.

2001- Regression 3

Regression using only financial ratios and risk exposure

```
# General regression code:
X1 = merged_df_2008[['Total Debt/Equity', 'ROA',  'Asset Turnover Ratio']]
X2 = merged_df_2008[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2], axis=1)

y = merged_df_2008['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()
```

```
# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.688
Model:                            OLS   Adj. R-squared:                  0.687
Method:                 Least Squares   F-statistic:                     893.5
Date:                Sat, 06 May 2023   Prob (F-statistic):               0.00
Time:                        12:27:42   Log-Likelihood:                 3404.0
No. Observations:                2440   AIC:                            -6794.
Df Residuals:                    2433   BIC:                            -6753.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
                           coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                    0.0208      0.003      6.271      0.000       0.014
Total Debt/Equity        0.0008      0.000      2.499      0.013       0.000
ROA                      0.0452      0.009      5.242      0.000       0.028
Asset Turnover Ratio    -0.0039      0.002     -2.220      0.026      -0.007
mktrf                   -0.1348      0.002    -69.398      0.000      -0.139
smb                     -0.0208      0.001    -15.912      0.000      -0.023
hml                     -0.0151      0.001    -12.209      0.000      -0.018
==============================================================================
Omnibus:                      161.737   Durbin-Watson:                   1.964
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              745.617
Skew:                          -0.067   Prob(JB):                    1.23e-162
Kurtosis:                       5.705   Cond. No.                         36.6
==============================================================================

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

This is the output of an Ordinary Least Squares (OLS) regression model with a dependent variable 'RET' (which stands for Return) and six independent variables: Total Debt/Equity, ROA, Asset Turnover Ratio, mktrf, smb, and hml. The model has been estimated using 2,440 observations. The model has an R-squared value of 0.688, indicating that the independent variables explain 68.8% of the variation in the dependent variable, with an adjusted R-Squared of 68.7%, indicating that there is no overfitting. All independent variables except Total Debt/Equity are statistically significant at a 5% level of significance.

2008- Regression 4

Regression using only risk exposure

```
#General regression code:
X = merged_df_2008[['mktrf', 'smb', 'hml']]
y = merged_df_2008['RET']
# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

#list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.683
Model:                            OLS   Adj. R-squared:                  0.683
Method:                 Least Squares   F-statistic:                     1749.
Date:                Sat, 06 May 2023   Prob (F-statistic):               0.00
Time:                        12:27:42   Log-Likelihood:                 3385.0
No. Observations:                2440   AIC:                            -6762.
Df Residuals:                    2436   BIC:                            -6739.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0239      0.003      9.171      0.000       0.019       0.029
mktrf         -0.1369      0.002    -71.477      0.000      -0.141      -0.133
smb           -0.0227      0.001    -17.874      0.000      -0.025      -0.020
hml           -0.0143      0.001    -11.768      0.000      -0.017      -0.012
==============================================================================
Omnibus:                      161.365   Durbin-Watson:                   1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              703.512
Skew:                          -0.127   Prob(JB):                    1.72e-153
Kurtosis:                       5.618   Cond. No.                         4.33
==============================================================================

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

In order to further understand the impact of risk exposures, the below above model is analyzed.

In this multiple linear regression model, the dependent variable is RET (stock returns) with six independent variables: Total Debt/Equity, ROA (return on assets), Asset Turnover Ratio, mktrf (systematic risk), smb (small minus big firms), and hml (high minus low book-to-market ratio firms).

Ths previous model showed that all six independent variables were statistically significant in predicting the dependent variable. The R-squared value was 0.688, which means that 68.8% of the variation in the dependent variable was explained by the independent variables.

This model shows a simplified version of the previous model, with only three independent variables: mktrf, smb, and hml. These three variables are also statistically significant in predicting the dependent variable. The R-squared value is 0.683, which means that 68.3% of the variation in the dependent variable is explained solely by the risk exposures.

2008- Regression 5

Regression with industry sectors

```
X_all_sectors = pd.get_dummies(merged_df_2008['gsector'], prefix='sector')
y = merged_df_2008['RET']
# Perform the regression using statsmodels
model_all_sectors = sm.OLS(y, X_all_sectors).fit()

# Identify the industry with the coefficient value closest to zero
reference_sector = model_all_sectors.params.abs().idxmin()

print(reference_sector)
```

```
sector_55.0
```

```
# Create dummy variables for the industry sectors
X_all_sectors = pd.get_dummies(merged_df_2008['gsector'], prefix='sector')
y = merged_df_2008['RET']

# Exclude the reference sector's dummy variable from the independent variables
X = X_all_sectors.drop(columns=[reference_sector])

# Add a constant to the independent variables
X = sm.add_constant(X)

# Perform the regression using statsmodels without the reference sector's dummy varia
model_5 = sm.OLS(y, X).fit()

# Print the regression output
print(model_5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.110
Model:                            OLS   Adj. R-squared:                  0.107
Method:                 Least Squares   F-statistic:                     30.15
Date:                Sat, 06 May 2023   Prob (F-statistic):           2.70e-55
Time:                        12:37:00   Log-Likelihood:                 2126.4
No. Observations:                2440   AIC:                            -4231.
Df Residuals:                    2429   BIC:                            -4167.
Df Model:                          10
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
```

```
const           -0.0565      0.011     -4.922      0.000      -0.079      -0.034
sector_10.0     -0.1251      0.014     -8.661      0.000      -0.153      -0.097
sector_15.0     -0.1049      0.015     -7.167      0.000      -0.134      -0.076
sector_20.0     -0.0903      0.013     -7.051      0.000      -0.115      -0.065
sector_25.0     -0.1067      0.013     -8.239      0.000      -0.132      -0.081
sector_30.0     -0.0359      0.015     -2.368      0.018      -0.066      -0.006
sector_35.0     -0.0846      0.013     -6.553      0.000      -0.110      -0.059
sector_40.0     -0.0298      0.012     -2.411      0.016      -0.054      -0.006
sector_45.0     -0.1128      0.012     -9.055      0.000      -0.137      -0.088
sector_50.0     -0.1077      0.017     -6.374      0.000      -0.141      -0.075
sector_60.0     -0.1014      0.016     -6.444      0.000      -0.132      -0.071
==============================================================================
Omnibus:                      140.013   Durbin-Watson:                   1.939
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              488.539
Skew:                          -0.182   Prob(JB):                    8.22e-107
Kurtosis:                       5.162   Cond. No.                         20.2
==============================================================================

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

The regression output shows the results of the OLS regression model that predicts the returns of companies based on their industry sectors. The R-squared value of the model is 0.110, which means that 11% of the variance in the dependent variable (RET) is explained by the independent variables (the dummy variables for industry sectors).

The coefficients for each industry sector are negative, which means that, on average, companies in these sectors have lower returns than the reference sector (which is not included in the model).

The reference sector (secotr 55) is not explicitly mentioned in the output, but it can be inferred from the fact that one dummy variable is excluded from the model. The reference sector is the industry sector that has the coefficient value closest to zero.

The coefficients of the different sectors indicate how much the return of a company is affected by the sector it belongs to, compared to the reference sector. For example, companies in sector 10.0, which is the energy sector, have returns that are 0.1251 lower, on average, than the reference sector, which is Utilities. Similarly, companies in Consumer Discretionary sector (25.0) have returns that are 0.1067 lower, on average, than the reference sector.

The intercept coefficient (const) is -0.0565, which means that, on average, a company in the reference sector has a return of -0.0565 when all other independent variables are held constant.

Overall, the results suggest that industry sector is a significant predictor of company returns, with different sectors having varying levels of impact on returns.

2001- Conclusion

For the Great Recession of 2008, the most explanatory power is held by the risk exposures.

The Covid Shock

The COVID-19 pandemic caused a global stock market crash in 2020, with major indices experiencing significant declines in March. Lockdowns and travel restrictions led to supply chain disruptions and reduced consumer spending, negatively impacting many industries.

Merging the three files together

```
# Renaming tic to TICKER in the gsector file
gsector_2020 = gsector_2020.rename(columns={'tic': 'TICKER'})


# Merge the DataFrames based on the 'TICKER' column
merged_df_2020 = pd.merge(gsector_2020, ratios_2020, on='TICKER')
merged_df_2020 = pd.merge(merged_df_2020, ff_2020, on='TICKER')
```

```
merged_df_2020.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2961 entries, 0 to 2960
Data columns (total 38 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   gvkey_x               2961 non-null   int64  
 1   fyear_x               2961 non-null   int64  
 2   TICKER                2961 non-null   object 
 3   gsector               2960 non-null   float64
 4   gvkey_y               2961 non-null   int64  
 5   datadate              2961 non-null   object 
 6   fyear_y               2961 non-null   int64  
 7   indfmt                2961 non-null   object 
 8   consol                2961 non-null   object 
 9   popsrc                2961 non-null   object 
 10  datafmt               2961 non-null   object 
 11  curcd                 2961 non-null   object 
 12  act                   2961 non-null   float64
 13  at                    2961 non-null   float64
 14  che                   2961 non-null   float64
 15  cogs                  2961 non-null   float64
 16  dltt                  2961 non-null   float64
 17  lct                   2961 non-null   float64
 18  lt                    2961 non-null   float64
 19  ni                    2961 non-null   float64
```

```
 20  re                     2961 non-null   float64
 21  sale                   2961 non-null   float64
 22  costat                 2961 non-null   object
 23  RET                    2961 non-null   float64
 24  Net Profit Margin      2961 non-null   float64
 25  Gross Profit Margin    2961 non-null   float64
 26  Total Debt/Equity      2961 non-null   float64
 27  Return on Equity       2961 non-null   float64
 28  Liquidity              2961 non-null   float64
 29  ROA                    2961 non-null   float64
 30  Debt/Total Assets      2961 non-null   float64
 31  Asset Turnover Ratio   2961 non-null   float64
 32  at_log                 2961 non-null   float64
 33  Unnamed: 0             2961 non-null   int64
 34  const                  2961 non-null   float64
 35  mktrf                  2961 non-null   float64
 36  smb                    2961 non-null   float64
 37  hml                    2961 non-null   float64
dtypes: float64(25), int64(5), object(8)
memory usage: 902.2+ KB
```

Run OLS regression with the following three main categories

1. **Risk Exposures:** This included systematic risk (market risk), SML (Small Minus Big), which is the size risk, and HML (High minus Low), which is the Value vs Growth risk.
2. **Financial Ratios:** This includes various financial ratios, including but not limited to Liquidity, Leverage, Asset Turnover Ratio, Net and Gross Profit Margin, ROE and ROA.
3. **Industry Indicators:** This indicates which industries had the highest amount of affect on stock returns.

2020- Regression 1

Regression with financial ratios, industry indicators and risk exposure

```
# General regression code:
X1 = merged_df_2020[['Gross Profit Margin', 'Liquidity', 'at_log']]
X2 = pd.get_dummies(merged_df_2020['gsector'], prefix='sector')
X3 = merged_df_2020[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2, X3], axis=1)

y = merged_df_2020['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
```

```
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.556
Model:                            OLS   Adj. R-squared:                  0.554
Method:                 Least Squares   F-statistic:                     217.2
Date:                Sat, 06 May 2023   Prob (F-statistic):               0.00
Time:                        12:27:42   Log-Likelihood:                 2648.2
No. Observations:                2961   AIC:                            -5260.
Df Residuals:                    2943   BIC:                            -5153.
Df Model:                          17
Covariance Type:            nonrobust
================================================================================
                          coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                  -0.1412      0.100     -1.416      0.157      -0.337
Gross Profit Margin    -0.0001      0.000     -0.362      0.717      -0.001
Liquidity              -0.0001      0.001     -0.125      0.901      -0.002
at_log                  0.0016      0.001      1.621      0.105      -0.000
sector_10.0             0.1460      0.100      1.468      0.142      -0.049
sector_15.0             0.1372      0.100      1.379      0.168      -0.058
sector_20.0             0.1429      0.099      1.438      0.150      -0.052
sector_25.0             0.1184      0.099      1.191      0.234      -0.077
sector_30.0             0.1424      0.100      1.429      0.153      -0.053
sector_35.0             0.1662      0.099      1.671      0.095      -0.029
sector_40.0             0.1178      0.099      1.186      0.236      -0.077
sector_45.0             0.1503      0.099      1.512      0.131      -0.045
sector_50.0             0.1294      0.100      1.298      0.194      -0.066
sector_55.0             0.0985      0.100      0.986      0.324      -0.097
sector_60.0             0.0960      0.100      0.964      0.335      -0.099
mktrf                  -0.1298      0.003    -45.836      0.000      -0.135
smb                    -0.0076      0.002     -3.651      0.000      -0.012
hml                    -0.0778      0.003    -27.502      0.000      -0.083
==============================================================================
Omnibus:                     4465.985   Durbin-Watson:                   1.964
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          5445507.010
Skew:                           8.722   Prob(JB):                         0.00
Kurtosis:                     212.365   Cond. No.                     1.56e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The condition number is large, 1.56e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The multiple linear regression model has an adjusted R-squared value of 0.554, indicating that the model explains around 55% of the variance in the dependent variable. Among the independent variables, only mktrf, smb, and hml have p-values less than 0.05, indicating that they are statistically significant predictors of the dependent variable at a 95% confidence level. The coefficients of these variables are negative, indicating that they have a negative effect on the stock returns. The other independent variables, including const, Gross Profit Margin, Liquidity, at_log, and sector variables, are not statistically significant at a 95% confidence level.

Here we also see that the as risk increased, including market risk, size risk and distress risk, this negatively impacted the stock returns.

However, the model also has some potential issues with multicollinearity and numerical problems.

2020- Regression 2

Regression with financial ratios and industry indicators

```
# General regression code:
X1 = merged_df_2020[['Gross Profit Margin', 'Liquidity', 'at_log']]
X2 = pd.get_dummies(merged_df_2020['gsector'], prefix='sector')

X = pd.concat([X1, X2], axis=1)

y = merged_df_2020['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.124
Model:                            OLS   Adj. R-squared:                  0.120
Method:                 Least Squares   F-statistic:                     29.83
Date:                Sat, 06 May 2023   Prob (F-statistic):           8.58e-75
Time:                        12:27:42   Log-Likelihood:                 1640.9
No. Observations:                2961   AIC:                            -3252.
Df Residuals:                    2946   BIC:                            -3162.
Df Model:                          14
Covariance Type:            nonrobust
================================================================================
                          coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                  -0.3420      0.140     -2.446      0.014      -0.616
Gross Profit Margin     0.0007      0.001      1.227      0.220      -0.000
Liquidity               0.0050      0.001      3.902      0.000       0.002
at_log                 -0.0020      0.001     -1.541      0.123      -0.005
sector_10.0             0.0764      0.140      0.547      0.585      -0.198
sector_15.0             0.1763      0.140      1.261      0.207      -0.098
sector_20.0             0.1779      0.140      1.275      0.202      -0.096
```

```
sector_25.0              0.1317      0.140      0.943      0.346      -0.142
sector_30.0              0.2656      0.140      1.897      0.058      -0.009
sector_35.0              0.2568      0.140      1.840      0.066      -0.017
sector_40.0              0.1799      0.140      1.289      0.197      -0.094
sector_45.0              0.2182      0.140      1.564      0.118      -0.055
sector_50.0              0.2084      0.140      1.489      0.137      -0.066
sector_55.0              0.2536      0.140      1.808      0.071      -0.021
sector_60.0              0.1734      0.140      1.240      0.215      -0.101
==============================================================================
Omnibus:                     4539.402   Durbin-Watson:                   1.929
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          5883822.268
Skew:                           9.007   Prob(JB):                         0.00
Kurtosis:                     220.637   Cond. No.                     1.55e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The condition number is large, 1.55e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

This is an OLS regression analysis of the dependent variable RET (stock returns) and 14 independent variables. The adjusted R-squared value is 0.120, which means that the model explains 12% of the variance in the dependent variable after adjusting for the number of independent variables. However, only the independent variables Liquidity and sector_30.0 are significant at a 95% confidence level (P < 0.05), whereas the other independent variables are not significant. There might be some numerical problems or strong multicollinearity in the data, indicated by the large condition number of 1.55e+03.

This significanly low explanatory power is a result of removing market, size and distress risk from the regression.

2020- Regression 3

Regression with financial ratios and risk exposure

```
# General regression code:
X1 = merged_df_2020[['Gross Profit Margin', 'Liquidity', 'at_log']]
X2 = merged_df_2020[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2], axis=1)

y = merged_df_2020['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()
```

```
# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.542
Model:                            OLS   Adj. R-squared:                  0.541
Method:                 Least Squares   F-statistic:                     583.2
Date:                Sat, 06 May 2023   Prob (F-statistic):               0.00
Time:                        12:27:43   Log-Likelihood:                 2601.5
No. Observations:                2961   AIC:                            -5189.
Df Residuals:                    2954   BIC:                            -5147.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
                          coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                  -0.0036      0.009     -0.386      0.700      -0.022
Gross Profit Margin    -0.0004      0.000     -1.003      0.316      -0.001
Liquidity               0.0013      0.001      1.393      0.164      -0.001
at_log                  0.0002      0.001      0.236      0.814      -0.002
mktrf                  -0.1251      0.003    -47.042      0.000      -0.130
smb                    -0.0063      0.002     -3.046      0.002      -0.010
hml                    -0.0842      0.003    -33.077      0.000      -0.089
==============================================================================
Omnibus:                     4454.761   Durbin-Watson:                   1.953
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          5375759.179
Skew:                           8.680   Prob(JB):                         0.00
Kurtosis:                     211.017   Cond. No.                         42.7
==============================================================================

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

This is a linear regression model with a dependent variable RET (stock returns) and six independent variables: Gross Profit Margin, Liquidity, at_log, mktrf, smb, and hml. The model has an R-squared of 0.542, indicating that 54.2% of the variability in the dependent variable is explained by the independent variables. Only the variables mktrf, smb, and hml have statistically significant coefficients at the 5% level. The constant term is not statistically significant.

As risk exposures were put back into the regression, the model's explanatory power and improved significantly.

2020- Regression 4

Regression with only risk exposure

```
#General regression code:
X = merged_df_2020[['mktrf', 'smb', 'hml']]
y = merged_df_2020['RET']
# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

#list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.542
Model:                            OLS   Adj. R-squared:                  0.541
Method:                 Least Squares   F-statistic:                     1165.
Date:                Sat, 06 May 2023   Prob (F-statistic):               0.00
Time:                        12:27:43   Log-Likelihood:                 2599.8
No. Observations:                2961   AIC:                            -5192.
Df Residuals:                    2957   BIC:                            -5168.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0012      0.004      0.318      0.751      -0.006       0.008
mktrf         -0.1252      0.003    -47.482      0.000      -0.130      -0.120
smb           -0.0059      0.002     -2.999      0.003      -0.010      -0.002
hml           -0.0852      0.002    -35.061      0.000      -0.090      -0.080
==============================================================================
Omnibus:                     4439.223   Durbin-Watson:                   1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          5268487.010
Skew:                           8.622   Prob(JB):                         0.00
Kurtosis:                     208.926   Cond. No.                         4.13
==============================================================================

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

The OLS regression model has an R-squared of 0.542, indicating that the model explains 54.2% of the variability in the dependent variable. The coefficients for Gross Profit Margin and Liquidity are not statistically significant at a 5% level. The final model includes the market factor (mktrf), size factor (smb), and value factor (hml), which are all statistically significant at a 5% level. The adjusted R-squared is 0.541, indicating that the model has a good fit. The model is also free from multicollinearity issues as evidenced by the low condition number.

It is also important to note the inverse relationship between risk exposures and stock returns.

2020- Regression 5

Regression with only industry indicators

```
X_all_sectors = pd.get_dummies(merged_df_2020['gsector'], prefix='sector')
y = merged_df_2020['RET']
# Perform the regression using statsmodels
model_all_sectors = sm.OLS(y, X_all_sectors).fit()

# Identify the industry with the coefficient value closest to zero
reference_sector = model_all_sectors.params.abs().idxmin()

print(reference_sector)
```

```
    sector_35.0
```

```
# Create dummy variables for the industry sectors
X_all_sectors = pd.get_dummies(merged_df_2020['gsector'], prefix='sector')
y = merged_df_2020['RET']

# Exclude the reference sector's dummy variable from the independent variables
X = X_all_sectors.drop(columns=[reference_sector])

# Add a constant to the independent variables
X = sm.add_constant(X)

# Perform the regression using statsmodels without the reference sector's dummy varia
model_5 = sm.OLS(y, X).fit()

# Print the regression output
print(model_5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.116
Model:                            OLS   Adj. R-squared:                  0.113
Method:                 Least Squares   F-statistic:                     38.64
Date:                Sat, 06 May 2023   Prob (F-statistic):           5.20e-72
Time:                        12:50:31   Log-Likelihood:                 1626.8
No. Observations:                2961   AIC:                            -3232.
Df Residuals:                    2950   BIC:                            -3166.
Df Model:                          10
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const          -0.0807      0.007    -11.584      0.000      -0.094      -0.067
sector_10.0    -0.1890      0.012    -15.490      0.000      -0.213      -0.165
sector_15.0    -0.0876      0.013     -6.819      0.000      -0.113      -0.062
sector_20.0    -0.0870      0.010     -9.002      0.000      -0.106      -0.068
sector_25.0    -0.1351      0.010    -12.949      0.000      -0.156      -0.115
sector_30.0    -0.0010      0.015     -0.069      0.945      -0.030       0.028
sector_40.0    -0.0899      0.009     -9.865      0.000      -0.108      -0.072
```

```
sector_45.0    -0.0424      0.010     -4.238      0.000      -0.062      -0.023
sector_50.0    -0.0590      0.015     -4.029      0.000      -0.088      -0.030
sector_55.0    -0.0206      0.017     -1.185      0.236      -0.055       0.013
sector_60.0    -0.0928      0.013     -7.280      0.000      -0.118      -0.068
==============================================================================
Omnibus:                     4500.352   Durbin-Watson:                   1.935
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          5598648.604
Skew:                           8.858   Prob(JB):                         0.00
Kurtosis:                     215.286   Cond. No.                         10.4
==============================================================================

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

This is a linear regression model with the dependent variable RET (return) and ten independent variables, which are dummy variables for industry sectors (excluding the reference sector with the coefficient closest to zero). The model has an R-squared value of 0.116, which means that 11.6% of the variation in the dependent variable can be explained by the independent variables.

Here, a one-unit increase in sector_10.0 (Energy Sector) is associated with a 0.1890 decrease in RET (stock return), holding all other independent variables constant.

The reference sector is sector_30.0, which is the Consumer Staples Sector, and its coefficient is not statistically significant with a p-value of 0.945, indicating that there is no significant difference in return between sector_30.0 and the reference sector.

The coefficients of the other independent variables are all statistically significant with p-values less than 0.05.

2020- Conclusion

`As in the models before for the previous 3 stock market crashes, risk exposures explain the most variability in returns for the months leading up to the Covid Crash in 2020.`

The end of Easy Money

The Easy Money stock market crash of 2022 occurred due to the Federal Reserve's decision to raise interest rates, leading to a significant drop in stock prices. The crash resulted in significant losses for investors and sparked concerns about the stability of the financial system.

Merging the three files together

```
# Renaming tic to TICKER in the gsector file
gsector_2022 = gsector_2022.rename(columns={'tic': 'TICKER'})


# Merge the DataFrames based on the 'TICKER' column
merged_df_2022 = pd.merge(gsector_2022, ratios_2022, on='TICKER')
merged_df_2022 = pd.merge(merged_df_2022, ff_2022, on='TICKER')


merged_df_2022.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3026 entries, 0 to 3025
Data columns (total 38 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   gvkey_x               3026 non-null   int64  
 1   fyear_x               3026 non-null   int64  
 2   TICKER                3026 non-null   object 
 3   gsector               3025 non-null   float64
 4   gvkey_y               3026 non-null   int64  
 5   datadate              3026 non-null   object 
 6   fyear_y               3026 non-null   int64  
 7   indfmt                3026 non-null   object 
 8   consol                3026 non-null   object 
 9   popsrc                3026 non-null   object 
 10  datafmt               3026 non-null   object 
 11  curcd                 3026 non-null   object 
 12  act                   3026 non-null   float64
 13  at                    3026 non-null   float64
 14  che                   3026 non-null   float64
 15  cogs                  3026 non-null   float64
 16  dltt                  3026 non-null   float64
 17  lct                   3026 non-null   float64
 18  lt                    3026 non-null   float64
 19  ni                    3026 non-null   float64
 20  re                    3026 non-null   float64
 21  sale                  3026 non-null   float64
 22  costat                3026 non-null   object 
 23  RET                   3026 non-null   float64
 24  Net Profit Margin     3026 non-null   float64
 25  Gross Profit Margin   3026 non-null   float64
 26  Total Debt/Equity     3026 non-null   float64
 27  Return on Equity      3026 non-null   float64
 28  Liquidity             3026 non-null   float64
 29  ROA                   3026 non-null   float64
 30  Debt/Total Assets     3026 non-null   float64
 31  Asset Turnover Ratio  3026 non-null   float64
 32  at_log                3026 non-null   float64
 33  Unnamed: 0            3026 non-null   int64  
 34  const                 3026 non-null   float64
```

```
 35  mktrf                  3026 non-null   float64
 36  smb                    3026 non-null   float64
 37  hml                    3026 non-null   float64
dtypes: float64(25), int64(5), object(8)
memory usage: 922.0+ KB
```

Run OLS regression with the following three main categories

1. **Risk Exposures:** This included systematic risk (market risk), SML (Small Minus Big), which is the size risk, and HML (High minus Low), which is the Value vs Growth risk.
2. **Financial Ratios:** This includes various financial ratios, including but not limited to Liquidity, Leverage, Asset Turnover Ratio, Net and Gross Profit Margin, ROE and ROA.
3. **Industry Indicators:** This indicates which industries had the highest amount of affect on stock returns.

2022- Regression 1

Regression with financial ratios, industry indicators and risk exposures

```
# General regression code:
X1 = merged_df_2022[['Gross Profit Margin', 'Liquidity', 'at_log']]
X2 = pd.get_dummies(merged_df_2022['gsector'], prefix='sector')
X3 = merged_df_2022[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2, X3], axis=1)

y = merged_df_2022['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.231
Model:                            OLS   Adj. R-squared:                  0.227
Method:                 Least Squares   F-statistic:                     53.20
Date:                Sat, 06 May 2023   Prob (F-statistic):          1.67e-157
Time:                        12:27:43   Log-Likelihood:                 3866.7
No. Observations:                3026   AIC:                            -7697.
Df Residuals:                    3008   BIC:                            -7589.
Df Model:                          17
```

```
Covariance Type:            nonrobust
=================================================================================
                          coef    std err          t      P>|t|      [0.025
---------------------------------------------------------------------------------
const                   0.0311      0.068      0.458      0.647      -0.102
Gross Profit Margin    -0.0004      0.000     -1.937      0.053      -0.001     4
Liquidity               0.0013      0.000      3.065      0.002       0.000
at_log                  0.0017      0.001      2.509      0.012       0.000
sector_10.0            -0.0252      0.068     -0.371      0.710      -0.158
sector_15.0            -0.0752      0.068     -1.109      0.267      -0.208
sector_20.0            -0.0455      0.068     -0.671      0.502      -0.178
sector_25.0            -0.0468      0.068     -0.690      0.490      -0.180
sector_30.0            -0.0391      0.068     -0.576      0.565      -0.172
sector_35.0            -0.0534      0.068     -0.788      0.431      -0.186
sector_40.0            -0.0575      0.068     -0.849      0.396      -0.190
sector_45.0            -0.0494      0.068     -0.729      0.466      -0.182
sector_50.0            -0.0655      0.068     -0.964      0.335      -0.199
sector_55.0            -0.0404      0.068     -0.594      0.553      -0.174
sector_60.0            -0.0705      0.068     -1.038      0.299      -0.204
mktrf                  -0.0508      0.002    -21.899      0.000      -0.055
smb                    -0.0139      0.001    -14.183      0.000      -0.016
hml                     0.0299      0.002     13.701      0.000       0.026
==============================================================================
Omnibus:                      505.374   Durbin-Watson:                   1.984
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             7643.782
Skew:                           0.299   Prob(JB):                         0.00
Kurtosis:                      10.763   Cond. No.                     1.62e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The condition number is large, 1.62e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The OLS regression model shows that the R-squared value is 0.231, indicating that the model explains only 23.1% of the variation in the dependent variable, with an adjusted R-Squared of 22.7%.

The regression coefficients for Gross Profit Margin, sector_10.0, sector_15.0, sector_20.0, sector_25.0, sector_30.0, sector_35.0, sector_40.0, sector_45.0, sector_50.0, sector_55.0, and sector_60.0 are not statistically significant at the 5% level. The regression coefficients for Liquidity, at_log, mktrf, smb, and hml are statistically significant at the 5% level.

The model's goodness of fit can be improved, as indicated by the Adjusted R-squared value of 0.227. The coefficients of smb and hml are negative and positive, respectively, suggesting that a higher SMB (size factor) and HML (value factor) are associated with lower and higher stock market returns, respectively.

The model has some potential issues with multicollinearity, as indicated by the large condition number.

2022- Regression 2

Regression with only financial ratios and industry indicators.

```
# General regression code:
X1 = merged_df_2022[['Gross Profit Margin', 'Liquidity', 'at_log']]
X2 = pd.get_dummies(merged_df_2022['gsector'], prefix='sector')

X = pd.concat([X1, X2], axis=1)

y = merged_df_2022['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.058
Model:                            OLS   Adj. R-squared:                  0.054
Method:                 Least Squares   F-statistic:                     13.34
Date:                Sat, 06 May 2023   Prob (F-statistic):           3.06e-31
Time:                        12:27:43   Log-Likelihood:                 3560.0
No. Observations:                3026   AIC:                            -7090.
Df Residuals:                    3011   BIC:                            -7000.
Df Model:                          14
Covariance Type:            nonrobust
================================================================================
                          coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                   0.0067      0.075      0.090      0.929      -0.140
Gross Profit Margin     0.0003      0.000      1.439      0.150      -0.000
Liquidity               0.0005      0.000      1.071      0.284      -0.000
at_log                  0.0028      0.001      3.910      0.000       0.001
sector_10.0            -0.0536      0.075     -0.715      0.475      -0.201
sector_15.0            -0.1057      0.075     -1.409      0.159      -0.253
sector_20.0            -0.0813      0.075     -1.086      0.278      -0.228
sector_25.0            -0.0940      0.075     -1.255      0.210      -0.241
sector_30.0            -0.0561      0.075     -0.747      0.455      -0.203
sector_35.0            -0.1003      0.075     -1.340      0.180      -0.247
sector_40.0            -0.0741      0.075     -0.990      0.322      -0.221
sector_45.0            -0.0995      0.075     -1.328      0.184      -0.246
sector_50.0            -0.1013      0.075     -1.349      0.177      -0.249
sector_55.0            -0.0490      0.075     -0.651      0.515      -0.197
sector_60.0            -0.0985      0.075     -1.313      0.189      -0.246
==============================================================================
Omnibus:                      427.615   Durbin-Watson:                   1.979
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5998.136
```

```
Skew:                           0.007   Prob(JB):                        0.00
Kurtosis:                       9.897   Cond. No.                    1.61e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The condition number is large, 1.61e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The R-squared value of the model is 0.058, indicating that only 5.8% of the variation in the dependent variable is explained by the independent variables. The Adjusted R-squared value is slightly lower at 0.054, suggesting that the model's goodness of fit could be improved.

The coefficients of the independent variables are mostly negative, indicating that all sectors and risk exposures, except for sml, lead to lower stock market returns. However, the p-values of most of these variables are greater than 0.05, suggesting that they are not statistically significant predictors of stock market returns.

The p-value of at_log is less than 0.05, indicating that it is a statistically significant predictor of stock market returns. This suggests that as a company's assets grow, its stock market returns also increase.

The p-values of the sector indicators are mixed, with some being statistically significant predictors of stock market returns (e.g., sector_15.0, which is Materials) and others not (e.g., sector_10.0, which is Energy).

The model's condition number is large, indicating the possible presence of strong multicollinearity or other numerical problems.

2022- Regression 3

Regression with financial ratios and risk exposures

```
# General regression code:
X1 = merged_df_2022[['Gross Profit Margin', 'Liquidity', 'at_log']]
X2 = merged_df_2022[['mktrf', 'smb', 'hml']]

X = pd.concat([X1, X2], axis=1)

y = merged_df_2022['RET']

# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()
```

```
# list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.211
Model:                            OLS   Adj. R-squared:                  0.209
Method:                 Least Squares   F-statistic:                     134.2
Date:                Sat, 06 May 2023   Prob (F-statistic):          4.77e-151
Time:                        12:27:43   Log-Likelihood:                 3826.7
No. Observations:                3026   AIC:                            -7639.
Df Residuals:                    3019   BIC:                            -7597.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
                         coef    std err          t      P>|t|      [0.025
--------------------------------------------------------------------------------
const                 -0.0188      0.006     -3.069      0.002      -0.031
Gross Profit Margin   -0.0004      0.000     -1.943      0.052      -0.001
Liquidity              0.0013      0.000      3.062      0.002       0.000
at_log                 0.0011      0.001      1.722      0.085      -0.000
mktrf                 -0.0495      0.002    -22.249      0.000      -0.054
smb                   -0.0135      0.001    -13.889      0.000      -0.015
hml                    0.0322      0.002     16.911      0.000       0.028
==============================================================================
Omnibus:                      488.542   Durbin-Watson:                   1.971
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6765.956
Skew:                           0.303   Prob(JB):                         0.00
Kurtosis:                      10.300   Cond. No.                         42.7
==============================================================================

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

The results indicate that the model has an R-squared value of 0.211, which implies that approximately 21.1% of the variance in the dependent variable can be explained by the independent variables. The F-statistic of 134.2 is statistically significant with a p-value of 4.77e-151. The results also indicate that some of the independent variables are statistically significant, while others are not.

Among the statistically significant variables, mktrf, smb, and hml are significant at a 95% level of significance. This implies that these variables have a strong impact on the dependent variable, RET. Specifically, mktrf has a negative coefficient of -0.0495, which means that an increase in the market risk premium will lead to a decrease in the firm's return on equity.

Similarly, smb has a negative coefficient of -0.0135, while hml has a positive coefficient of 0.0322, indicating that an increase in the size and value factors will lead to a decrease and increase, respectively, in the firm's returns.

2022- Regression 4

Regression with only the risk exposures

```
#General regression code:
X = merged_df_2022[['mktrf', 'smb', 'hml']]
y = merged_df_2022['RET']
# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()

#list regression output
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.207
Model:                            OLS   Adj. R-squared:                  0.206
Method:                 Least Squares   F-statistic:                     262.3
Date:                Sat, 06 May 2023   Prob (F-statistic):         2.75e-151
Time:                        12:27:44   Log-Likelihood:                 3819.1
No. Observations:                3026   AIC:                            -7630.
Df Residuals:                    3022   BIC:                            -7606.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0070      0.003     -2.625      0.009      -0.012      -0.002
mktrf         -0.0491      0.002    -22.090      0.000      -0.053      -0.045
smb           -0.0130      0.001    -14.784      0.000      -0.015      -0.011
hml            0.0308      0.002     17.179      0.000       0.027       0.034
==============================================================================
Omnibus:                      488.178   Durbin-Watson:                   1.981
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6859.814
Skew:                           0.294   Prob(JB):                         0.00
Kurtosis:                      10.353   Cond. No.                         5.21
==============================================================================

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

The R-squared value of the model is 0.207, which indicates that the independent variables explain approximately 21% of the variation in the dependent variable. The adjusted R-squared value is 0.206, which accounts for the degrees of freedom in the model.

The coefficients of the independent variables show the effect of each variable on the dependent variable, holding other variables constant. The constant is -0.0070, indicating that the expected value of the dependent variable when all independent variables are equal to zero is -0.0070.

The coefficients of "mktrf", "smb", and "hml" are -0.0491, -0.0130, and 0.0308, respectively. These coefficients suggest that "mktrf" and "smb" have a negative effect on the dependent variable, while "hml" has a positive effect. All three coefficients are statistically significant with p-values less than 0.05.

2022- Regression 5

Regression with only the industry indicators

```
X_all_sectors = pd.get_dummies(merged_df_2022['gsector'], prefix='sector')
y = merged_df_2022['RET']
# Perform the regression using statsmodels
model_all_sectors = sm.OLS(y, X_all_sectors).fit()

# Identify the industry with the coefficient value closest to zero
reference_sector = model_all_sectors.params.abs().idxmin()

print(reference_sector)
```

```
    sector_55.0
```

```
# Create dummy variables for the industry sectors
X_all_sectors = pd.get_dummies(merged_df_2022['gsector'], prefix='sector')
y = merged_df_2022['RET']

# Exclude the reference sector's dummy variable from the independent variables
X = X_all_sectors.drop(columns=[reference_sector])

# Add a constant to the independent variables
X = sm.add_constant(X)

# Perform the regression using statsmodels without the reference sector's dummy varia
model_5 = sm.OLS(y, X).fit()

# Print the regression output
print(model_5.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.053
Model:                            OLS   Adj. R-squared:                  0.049
Method:                 Least Squares   F-statistic:                     16.71
Date:                Sat, 06 May 2023   Prob (F-statistic):           8.51e-30
Time:                        12:49:18   Log-Likelihood:                 3550.5
No. Observations:                3026   AIC:                            -7079.
Df Residuals:                    3015   BIC:                            -7013.
Df Model:                          10
```

```
Covariance Type:            nonrobust
================================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          -0.0157      0.008     -1.908      0.057      -0.032       0.000
sector_10.0    -0.0095      0.010     -0.958      0.338      -0.029       0.010
sector_15.0    -0.0599      0.010     -5.957      0.000      -0.080      -0.040
sector_20.0    -0.0371      0.009     -4.128      0.000      -0.055      -0.019
sector_25.0    -0.0495      0.009     -5.392      0.000      -0.068      -0.032
sector_30.0    -0.0111      0.011     -1.058      0.290      -0.032       0.010
sector_35.0    -0.0589      0.009     -6.582      0.000      -0.076      -0.041
sector_40.0    -0.0261      0.009     -2.960      0.003      -0.043      -0.009
sector_45.0    -0.0563      0.009     -6.192      0.000      -0.074      -0.038
sector_50.0    -0.0557      0.011     -5.267      0.000      -0.077      -0.035
sector_60.0    -0.0528      0.010     -5.221      0.000      -0.073      -0.033
==============================================================================
Omnibus:                      422.698   Durbin-Watson:                   1.972
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5685.662
Skew:                          -0.075   Prob(JB):                         0.00
Kurtosis:                       9.714   Cond. No.                         21.5
==============================================================================

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

This is a regression analysis for the year 2022 using the Ordinary Least Squares (OLS) method. The dependent variable is the 'RET' column, and the independent variables are the dummy variables for the 'gsector' column of the 'merged_df_2022' dataset.

The regression output shows that the R-squared value is 0.053, which means that the model explains only 5.3% of the variance in the dependent variable.

The coefficient values for the different industry sectors indicate the impact of each sector on the dependent variable. The 'sector_15.0' industry has the most significant impact with a coefficient value of -0.0599, which means that a one-unit increase in this sector's dummy variable decreases the dependent variable by 0.0599 units, holding all other variables constant. The reference sector, 'sector_10.0', has a coefficient value of -0.0095, which is not statistically significant.

The model's overall fit can be improved as the adjusted R-squared value is only 0.049.

Overall, this regression analysis suggests that the industry sector has a statistically significant impact on the dependent variable, but the model's fit is not strong enough to make accurate predictions.

2022- Conclusion

For the End of Easy Money Stock Market Crash, the most variability in the stock returns are explained by the risk exposures

Links and References

1. https://en.wikipedia.org/wiki/ Global_Industry_Classification_Standard