HW 1 TOPICS IN ECONOMICS - GROUP 2 ¶

Aman Krishna

Tim Taylor

Yazmin Ramirez Delgado

1 ESG (40 points)

In [63]:

```
import warnings
import pandas as pd
import plotnine as p9
from plotnine import ggplot, aes, geom_line, scale_color_manual, scale_line
warnings.filterwarnings("ignore", category=UserWarning, module="plotnine")
```

1.

The file MSCI sample.csv contains MSCI ESG ratings for a random sample of firms. Following equations 1 and 2 of Pástor et al. (2022), compute the firm-level greenness measure, $g_{i,t}$. In order to do the value-weighting (this will also apply when you construct the GMB portfolio), use the fields PRC (price) and SHROUT (shares outstanding) to compute the market cap of each firm at each date. Then, $\bar{G}$ is the market cap weighted average. You do not need to report anything at this step, but will need to use this measure later.

In the paper the unadjusted "greenness score" is given by:

$$G_{i,t-1} = \frac{-(10 - E\_score_{i,t-1}) \times E\_weight_{i,t-1}}{100}$$

This measures the most recent score to month t, where t-1 is between 1 and 12 months before the current period.

The firm-level greenness score is given by:

$$g_{i,t} = G_{i,t} - \bar{G}_t$$

Which is the difference between the firm's level of greenness and the aggregate weighted average

In [64]:
```
# written by Tim Taylor
msci_sample_path =  r"data/MSCI_sample.csv"
msci_sample = pd.read_csv("MSCI_sample.csv")

# adding market cap column
msci_sample["MARKET_CAP"] = msci_sample.PRC * msci_sample.SHROUT

msci_sample.columns
msci_sample = msci_sample.dropna() # removing all NaN so the vector product
```

In [65]:
```
# computing Gi,t-1
msci_sample[r'G'] =(-(10 - msci_sample.ENVIRONMENTAL_PILLAR_SCORE) * msci_s
msci_sample.loc[msci_sample.ISSUER_NAME.str.contains('Exxon Mobil Corporati
```

Out[65]:

	ISSUER_NAME	ISSUERID	AS_OF_DATE	IVA_INDUSTRY	INDUSTRY_ADJUSTED_SCOI
4617	Exxon Mobil Corporation	IID000000002127471	2013-08-01	Integrated Oil & Gas	4
5026	Exxon Mobil Corporation	IID000000002127471	2013-09-01	Integrated Oil & Gas	4
5446	Exxon Mobil Corporation	IID000000002127471	2013-10-01	Integrated Oil & Gas	4

Taking a look at the numbers in columns for ENVIRONMENTAL_PILLAR_SCORE and ENVIRONMENTAL_PILLAR_WEIGHT, we can see that they don't match up with those in the paper, even though the calculation method is the same, this is due to working with a sample and not the original dataset.

In order to get G_bar, we would need to first get the weighted portfolio in each month. This is an issue, because all of the reporting dates take place in different reporting periods. For instance, Best Buy contains data until August 2013, but Exxon Mobil contains data until October 2013, so we may see a change in the "market portfolio" weights because Best Buy will not be included after August 2013.

In [66]:
```
# Getting the weighted average by each month's weight, since G_bar is speci
msci_sample["MONTHLY_TOTAL_MARKET_CAP"] = msci_sample.groupby("AS_OF_DATE")
msci_sample["WEIGHTED_MARKET_CAP"] = msci_sample.MARKET_CAP / msci_sample.M

# Creating G_bar
msci_sample["WEIGHTED_G"]  = (msci_sample.WEIGHTED_MARKET_CAP * msci_sample
msci_sample["G_bar"] = msci_sample.groupby("AS_OF_DATE")["WEIGHTED_G"].tran
```

In [67]:
```
msci_sample.loc[msci_sample.ISSUER_NAME.str.contains('Exxon Mobil Corporati
```

Out[67]:

	ISSUER_NAME	ISSUERID	AS_OF_DATE	IVA_INDUSTRY	INDUSTRY_ADJUSTED_SCOI
4617	Exxon Mobil Corporation	IID000000002127471	2013-08-01	Integrated Oil & Gas	
5026	Exxon Mobil Corporation	IID000000002127471	2013-09-01	Integrated Oil & Gas	
5446	Exxon Mobil Corporation	IID000000002127471	2013-10-01	Integrated Oil & Gas	

3 rows × 21 columns

In [68]:
```
# creating g_i,t
msci_sample["g"] = msci_sample.G - msci_sample.G_bar
```

Checking $w_t \cdot g_t = 0$ as per equation (3) of the paper.

In [69]:
```
abs(round(msci_sample.dropna().WEIGHTED_MARKET_CAP @ msci_sample.dropna().g
```

Out[69]:
```
2e-15
```

(a) (5 points)

In your own words, explain what the unadjusted greenness score, $G_{i,t}$, measures. Make sure to mention why we need to include E_{weight}.

\textcolor{#406A5F}{G_{i,t}} measures the weighted amount of points of greenness score is needed to give that firm a perfect score in the environmental category. The weight \textcolor{#406A5F}E_{weight} is essential to making sure the number is relative to the firm's overall exposure to Environmental risks. For instance, it may be easy for a company that relies on little fossil-fuel bulky capital to have a near-perfect G or "greenness points away from perfect" score, where 0 is perfect, so to measure the amount of improvement still needed, it makes sense to penalize firms with more exposure weight-wise if environmental exposure is a bigger part of their overall risk profile, e.g. a logistics company would be more exposed to environmental risk than an fund management firm, which may be more exposed to Social or Governance risks.

(b) (5 points)

Why does the paper focus on the adjusted greenness score?

The paper focuses on the adjusted greenness score because it is concerned with measuring the overall returns to greenness across different asset classes. This is the "E" component only of ESG. Additionally, an unadjusted score wouldn't fully capture the potential current impact of a firm's good or bad greenness score, for reasons explored above. To add to that discussion, it also enables for relative comparison of how green and dirty firms perform over time, which is more in line with how investors would make their investment decisions rather than some arbitrary number that doesn't fully capture a firm's overall risk exposure to ESG issues.

2. (10 points)

Use the description in Section 4 of the paper to replicate Figure 3.

From the paper, it constructs a portfolio of the top third of greenness score companies with the bottom third of greenness score companies and measure their returns over time.

These will be constracted by first allocating a -1 for bottom 3rd, 1 for top 3rd, and 0 otherwise and then using normalized weights to get aggregate firm performance.

In [70]:
```
# adding percentile indicators
percentiles = msci_sample.groupby("AS_OF_DATE")['g'].quantile([0.3, 0.7]).u
msci_sample['Q_INDICATOR'] = msci_sample.apply(lambda row: 1 if row['g'] >=
msci_sample.head(3)

# computing the percentage returns
msci_sample["MONTHLY_RETURNS"] = msci_sample.groupby("ISSUER_NAME")["PRC"].

# checking for Exxon
msci_sample.loc[msci_sample.ISSUER_NAME.str.contains('Exxon Mobil Corporati
```

Out[70]:

	ISSUER_NAME	AS_OF_DATE	MONTHLY_RETURNS	Q_INDICATOR
4617	Exxon Mobil Corporation	2013-08-01	0.03763	-1
5026	Exxon Mobil Corporation	2013-09-01	-0.07029	-1
5446	Exxon Mobil Corporation	2013-10-01	-0.01285	-1

In [71]:
```
def get_monthly_returns(data = msci_sample.set_index("AS_OF_DATE").dropna()
    """
        This computes the portfolio returns and cumulative returns based on
    """
    # removing the first dates of percent change
    weights = data[data.Q_INDICATOR == Q]["MARKET_CAP"] / data[data.Q_INDIC
    returns = data[data.Q_INDICATOR == Q]["MONTHLY_RETURNS"]
    df = pd.DataFrame({f"{name}": (weights * returns).groupby("AS_OF_DATE")
    df = df[(df.index > '2012-10-01') & (df.index < '2021-01-01')] # the sa
    df[f"{name}_Cumulative"] = (1 + df[f"{name}"]).cumprod() - 1
    return df

# Clean Returns
rets_clean = get_monthly_returns(data = msci_sample.set_index("AS_OF_DATE")

# Dirty Returns
rets_dirty = get_monthly_returns(data = msci_sample.set_index("AS_OF_DATE")

df_returns = pd.concat([rets_clean, rets_dirty], axis = 1)
df_returns
```

Out[71]:

AS_OF_DATE	Clean_Rets	Clean_Rets_Cumulative	Dirty_Rets	Dirty_Rets_Cumulative
2012-11-01	0.01792	0.01792	-0.00482	-0.00482
2012-12-01	0.00343	0.02141	-0.01303	-0.01778
2013-01-01	0.04390	0.06625	-0.00274	-0.02048
2013-02-01	0.08212	0.15381	0.04773	0.02628
2013-03-01	0.01397	0.16993	0.00291	0.02926
...	...	...	...	...
2020-08-01	0.10253	3.12552	0.03077	0.36210
2020-09-01	-0.04372	2.94517	0.01907	0.38807
2020-10-01	-0.06575	2.68578	-0.03311	0.34212
2020-11-01	-0.01662	2.62453	-0.00102	0.34075
2020-12-01	0.19936	3.34710	0.11868	0.49987

98 rows × 4 columns

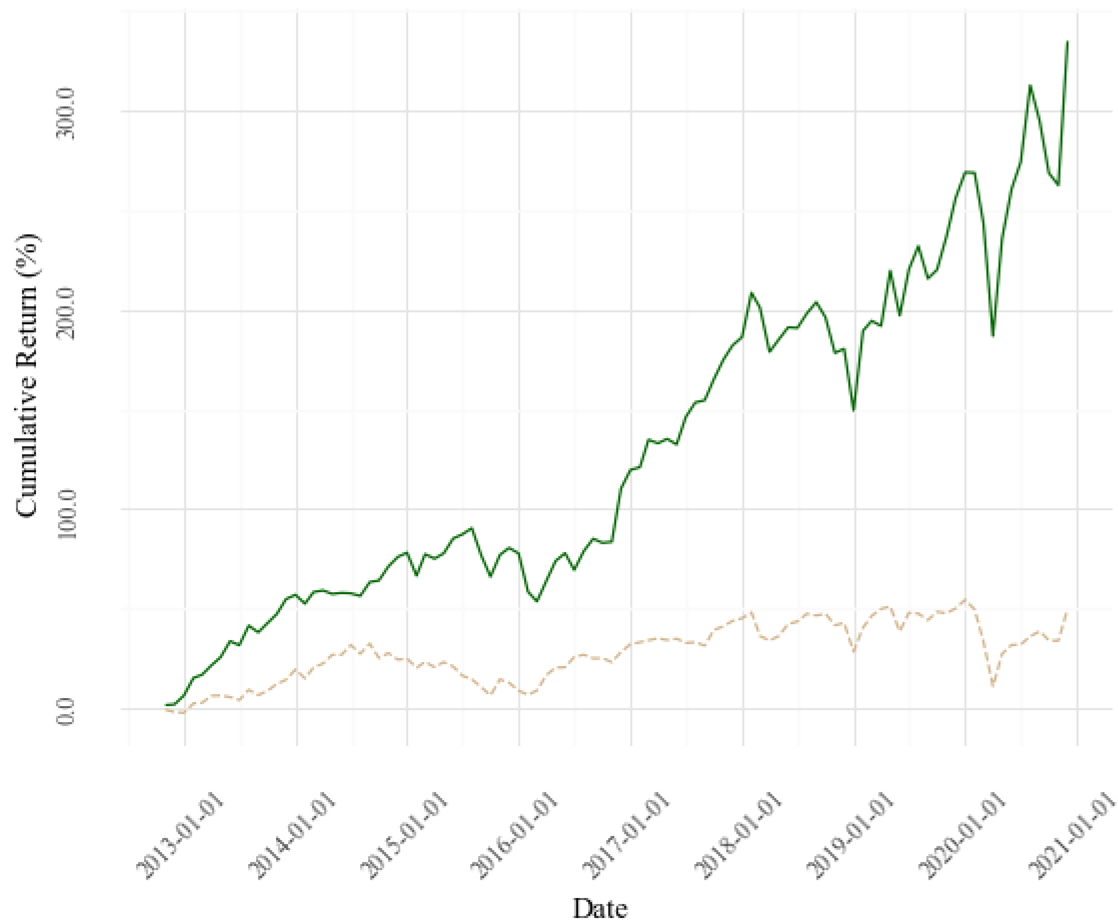
In [72]:

```
df_returns.index = pd.to_datetime(df_returns.index)
df = df_returns

line_styles = ['solid', 'dashed']
line_colors = ['#006400', '#D2B48C'] # green and brown

plot = (
    ggplot(df.reset_index(), aes(x='AS_OF_DATE'))
    + geom_line(aes(y='Clean_Rets_Cumulative'), color=line_colors[0], linet
    + geom_line(aes(y='Dirty_Rets_Cumulative'), color=line_colors[1], linet
    + scale_color_manual(values=line_colors)
    + scale_linetype_manual(values=line_styles)
    + labs(y='Cumulative Return (%)', x='Date')
    + theme_minimal()
    + scale_y_continuous(labels=lambda x: x * 100)
    + theme(axis_text_x=element_text(angle=45, family='Times New Roman'),
            axis_text_y=element_text(angle=90, family='Times New Roman'),
            legend_title=element_text(family='Times New Roman'))
    + theme(axis_title=element_text(family='Times New Roman'))
    # add a legend for the two lines
    + theme(legend_title="123", legend_position=(0.0, 0.0), legend_directic

)
plot
```



Out[72]: <ggplot: (702644962)>

3. (10 points)

What is the monthly return and Sharpe ratio of the green minus brown portfolio?

In [73]:
```
FF_path = r"data/F-F_Research_Data_Factors.csv"
FF_data = pd.read_csv("F-F_Research_Data_Factors.CSV",skiprows=2,nrows=1154
FF_data = FF_data.set_index("Unnamed: 0")
FF_data
### In the original dataset I removed the yearly data, headers, and footers
FF_data[(FF_data.index > 201210) & (FF_data.index < 202101)][["RF"]]
```

Out[73]:

	RF
Unnamed: 0	
201211	0.01
201212	0.01
201301	0.00
201302	0.00
201303	0.00
...	...
202008	0.01
202009	0.01
202010	0.01
202011	0.01
202012	0.01

98 rows × 1 columns

In [74]:
```
df_returns["Green_minus_Brown"] = df_returns.Clean_Rets - df_returns.Dirty_
display(df_returns[["Green_minus_Brown"]])

GMB_minus_RF = df_returns["Green_minus_Brown"].reset_index(drop=True) - (FF
# The sharpe ratio is calculated by taking the difference between the GMB a
print("Sharpe Ratio of GMB: ", round(float((GMB_minus_RF.mean()) / df_retur
```

	Green_minus_Brown
AS_OF_DATE	
2012-11-01	0.02274
2012-12-01	0.01646
2013-01-01	0.04664
2013-02-01	0.03439
2013-03-01	0.01106
...	...
2020-08-01	0.07176
2020-09-01	-0.06278
2020-10-01	-0.03264
2020-11-01	-0.01559
2020-12-01	0.08068

98 rows × 1 columns

```
Sharpe Ratio of GMB:  0.3024
```

The Sharpe ratio is a little bit lower than in the paper, this is likely due to a combination of the data included in the sample and the assumptions I made earlier regarding when to exclude missing data.

4. (10 points)

How does Pástor et al. (2022) explain that green stocks outperform bad stocks when the theoretical model from Pástor et al. (2021) suggests that brown stocks should outperform? Make sure to focus on the distinction between expected and realized returns.

In Pástor et al.'s 2022 paper, they find that the GMB portfolio returns have a significant positive return, however they note that when you control for the increased level of environmental issue awareness and concern level, a lot of these returns disappear. What they noted in their 2021 paper was that the green stocks should have a lower expected return as investors who buy them often make their decisions on measures of utility that include more than just the expected profitability. They reconcile this supposed flaw with a wider discussion about expected versus realized returns and note that even though we saw over the sample period that greener firms had a very high realized return, it does does disappear when you control for other factors, like environmental concern.

This is because even though realized returns are often used as a proxy for expected returns, they can often diverge as not only are there other proxies for expected returns, but there are other factors that should be controlled for that is missing in the usual assumptions for using realized returns as expected returns. Additionally, they note strange behaviour in price adjustment in large versus small firms, where small-cap firms tend to underreact to envronmental news, which suggests that green investing is not spread evenly. Overall, they are careful to note that their findings indicate this period of outperformance should not be interpreted to suggest that green firms should expect higher returns, and instead that these returns are driven by other factors.

2 Climate Risk (60 points)

In [75]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
import math as m
from arch import arch_model
from arch.univariate import GARCH, EWMAVariance
from sklearn import linear_model
import scipy.stats as stats
from statsmodels.regression.rolling import RollingOLS
import seaborn as sns
import warnings
from scipy.stats import norm
pd.set_option("display.precision", 5)
warnings.filterwarnings("ignore")
```

The first two parts of this problem are just to give you a chance to explore the industry returns data. The third part is specifically related to climate risk

1 (5 points)

The file 49_Industry_Portfolios.csv contains monthly returns of stocks in 49 different industries. Import this file into your program of choice. Restrict the sample to start in January 2004 and end in June 2018. Calculate the cumulative return of each industry over the sample period. Which were the three highest and three lowest performing industries in this sample?

In [76]:
```
# written by Aman Krishna and Yazmin Ramirez Delgado
# Reading in the data and dropping the first 11 rows and restricting the da
df = pd.read_csv("49_Industry_Portfolios.CSV", skiprows=11, index_col=0, nr

df = df.dropna()
```

Restricting Sample to provided dates

In [77]:
```
# Change the index to datetime of the form YYYYMM
df.index = pd.to_datetime(df.index, format='%Y%m')

#Change the Index column name to Date
df.index.name = 'Date'

#Restrict the sample to start between January 2004 and June 2018
df = df.loc['2004-01-01':'2018-06-01']

df
```

Out[77]:

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Boxes	Trans	Whls
Date														
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-1.09	-4.07	2.9
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	6.51	0.07	2.2
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.12	-0.54	2.1
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	-2.34	0.21	0.9
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	4.96	2.18	0.1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-3.76	-6.66	-5.4
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	-1.54	0.53	-1.0
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	-1.39	0.49	0.8
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	-3.09	4.85	1.5
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-2.51	-3.64	0.2

174 rows × 49 columns

Calculating Cumulative Return of each Insdustry over the sample period

In [78]:
```
# Calculate the cumulative monthly returns by summing the monthly returns
df_cumulative = df.cumsum()

df_cumulative_ranked = df_cumulative.iloc[-1]
```

In [79]:
```
df_cumulative
```

Out[79]:

Date	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Boxes	Tra
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-1.09	-4
2004-02-01	-0.72	7.09	11.36	-0.31	5.87	3.31	1.12	2.16	5.18	4.40	...	5.42	-4
2004-03-01	3.33	7.83	15.37	-0.52	1.84	-0.32	3.39	1.94	6.74	7.73	...	5.30	-4
2004-04-01	3.12	10.88	20.03	-0.14	3.75	-6.13	1.58	1.98	8.96	5.60	...	2.96	-4
2004-05-01	4.33	8.39	22.16	2.11	-8.86	-3.24	2.04	2.35	10.48	5.64	...	7.92	-2
...	...	...	...	...	...	...	...	...	...	...	...	...	
2018-02-01	165.46	139.59	184.28	149.37	233.75	91.10	212.32	52.55	107.81	190.58	...	197.16	170
2018-03-01	160.54	137.34	183.87	150.64	232.10	81.87	212.84	52.36	109.94	191.22	...	195.62	170
2018-04-01	168.53	133.83	183.08	146.35	217.94	88.29	215.93	51.90	104.59	194.38	...	194.23	171
2018-05-01	170.46	132.97	181.78	145.16	216.07	89.10	225.54	52.60	104.34	198.59	...	191.14	176
2018-06-01	167.39	139.01	186.23	149.58	219.14	94.78	230.45	54.93	107.93	205.57	...	188.63	172

174 rows × 49 columns

Three highest and three lowest performing industries in this sample

In [80]:
```
print("The three top performing industries are\n",df_cumulative_ranked.sort
print("----------------------------------------------")
print("The three lowest performing industries are\n",df_cumulative_ranked.s
```

```
The three top performing industries are
 Ships     282.98
Fun       230.45
Guns      228.34
Name: 2018-06-01 00:00:00, dtype: float64
----------------------------------------------
The three lowest performing industries are
 PerSv     80.27
Gold      70.70
Books     54.93
Name: 2018-06-01 00:00:00, dtype: float64
```

Top Performing Industries:

Ships: Shipping is an industry that may be exposed to climate change risk. The shipping industry may have benefited from increasing global trade during this period. As the world economy grew, the demand for transporting goods across borders also increased, which could have led to higher profitability for shipping companies.

Fun: This industry includes various entertainment and leisure businesses. During times of economic growth, people tend to spend more on entertainment and leisure, leading to increased revenues for companies in this sector. This might be one reason why we see it as a top performer.

Guns: The strong performance of this industry may not have a direct link to climate change concerns. Increased demand for firearms might have driven higher stock returns for companies in this industry.

Lowest Performing Industries:

PerSv: The poor performance of the Personal Services industry may not necessarily be driven by climate change concerns. With the automation and industrial development, many personal services have become redundant and moot.

Gold: The Gold industry typically reacts to various economic and geopolitical factors. It is typically considered a safe haven asset and behaves as somewhat of a hedge to the equity world. We do not see an immediate climate change connection.

Books : The decline in the "Books" industry's performance could be attributed to the rise of digital media and e-books, which disrupted traditional publishing. Also, books comes from trees.

2 (10 points)

For each industry, calculate the standard deviation of returns. Use this to calculate the Sharpe ratio. Which industries have the three highest and three lowest Sharpe ratios? How do these compare to those with the highest and lowest returns?

Calculating the standard deviation of returns for each industry

In [81]:
```
df_std = df.std()
df_std
```

Out[81]:
```
Agric     6.50389
Food      3.35111
Soda      5.65172
Beer      3.48047
Smoke     4.87271
Toys      6.15652
Fun       7.99779
Books     6.08752
Hshld     3.67395
Clths     5.70489
Hlth      5.12025
MedEq     4.57003
Drugs     3.86852
Chems     5.87956
Rubbr     5.89881
Txtls     9.10321
BldMt     6.77057
Cnstr     6.84579
Steel     8.51674
FabPr     8.02063
Mach      6.55334
ElcEq     6.06625
Autos     7.97213
Aero      5.26366
Ships     7.59158
Guns      5.09859
Gold     10.66172
Mines     9.22052
Coal     12.93151
Oil       5.80232
Util      3.57289
Telcm     4.24915
PerSv     5.90617
BusSv     4.46116
Hardw     5.91714
Softw     4.81853
Chips     5.79308
LabEq     5.15546
Paper     4.99164
Boxes     5.31394
Trans     5.04236
Whlsl     4.56167
Rtail     4.10749
Meals     4.03878
Banks     6.12945
Insur     5.30195
RlEst     8.75819
Fin       6.32057
Other     5.55620
dtype: float64
```

Calculating the Sharpe ratio for each industry

In [82]:
```
# Use the df_std and df_cumulative_ranked to calculate the Sharpe Ratio
df_sharpe = df_cumulative_ranked / df_std

df_sharpe
```

Out[82]:
```
Agric     25.73689
Food      41.48180
Soda      32.95104
Beer      42.97694
Smoke     44.97293
Toys      15.39507
Fun       28.81421
Books      9.02337
Hshld     29.37712
Clths     36.03402
Hlth      25.64716
MedEq     38.08070
Drugs     37.47943
Chems     32.64020
Rubbr     30.49087
Txtls     22.38550
BldMt     24.24906
Cnstr     16.12378
Steel     15.22883
FabPr     19.53960
Mach      28.52132
ElcEq     22.04328
Autos     12.67165
Aero      40.98861
Ships     37.27552
Guns      44.78491
Gold       6.63120
Mines     22.17770
Coal       7.11595
Oil       27.23739
Util      41.91003
Telcm     31.23211
PerSv     13.59088
BusSv     37.19882
Hardw     26.62264
Softw     37.90371
Chips     26.80958
LabEq     37.16057
Paper     25.44654
Boxes     35.49722
Trans     34.20024
Whlsl     32.39604
Rtail     40.65505
Meals     50.03986
Banks     14.63591
Insur     26.72790
RlEst     14.15589
Fin       21.50440
Other     16.14772
dtype: float64
```

Three highest and three lowest sharpe ratio industries in this sample

In [83]:
```
print("The three top Sharpe Ratio industries are\n",df_sharpe.sort_values(a
print("---------------------------------------------")
print("The three lowest Sharpe Ratio industries are\n",df_sharpe.sort_value
```

```
The three top Sharpe Ratio industries are
 Meals    50.03986
Smoke    44.97293
Guns     44.78491
dtype: float64
----------------------------------------------
The three lowest Sharpe Ratio industries are
 Books    9.02337
Coal     7.11595
Gold     6.63120
dtype: float64
```

The **Sharpe ratio** measures the risk-adjusted performance of an investment or portfolio. It considers not only the returns generated but also the level of risk taken to achieve those returns.

Top Sharpe Ratio Industries:

Meals: The returns from this industry were not only high but were achieved with relatively lower risk. Possible reasons could include stable demand for food-related products.

Smoke: This industry, which likely includes tobacco companies, also shows a high Sharpe ratio. This could be due to the relatively stable and consistent demand for tobacco products, which can lead to steady cash flows and returns.

Guns: The firearms industry has a high Sharpe ratio, indicating that it offered strong risk-adjusted returns. This may be linked to factors such as political events, concerns about personal safety, and changes in gun regulations, which could drive demand for firearms.

Bottom Sharpe Ratio Industries:

Books: This industry has the lowest Sharpe ratio among the listed industries. This suggests that the returns generated by this industry were relatively low compared to the level of risk taken. The decline of traditional print publishing due to the rise of digital media might have contributed to this lower risk-adjusted performance.

Coal: The coal industry's low Sharpe ratio reflects the challenges faced by this sector during the period. Factors like environmental concerns, the shift towards cleaner energy sources, and reduced demand for coal likely contributed to both lower returns and higher risk.

Gold: Gold, which is traditionally considered a safe-haven asset, had the second-lowest Sharpe ratio. This may be due to a combination of lower returns compared to equities and the relatively low risk associated with gold investments during this period.

In comparing the rankings based on cumulative monthly returns and Sharpe ratios, we can see that

3

In the next sub-questions, we are going to focus on days where attention to climate change is high. We will identify these days in two ways: using Google Trends searches and using the Ardia et al. (2020) climate concerns data.

(a) (20 points)

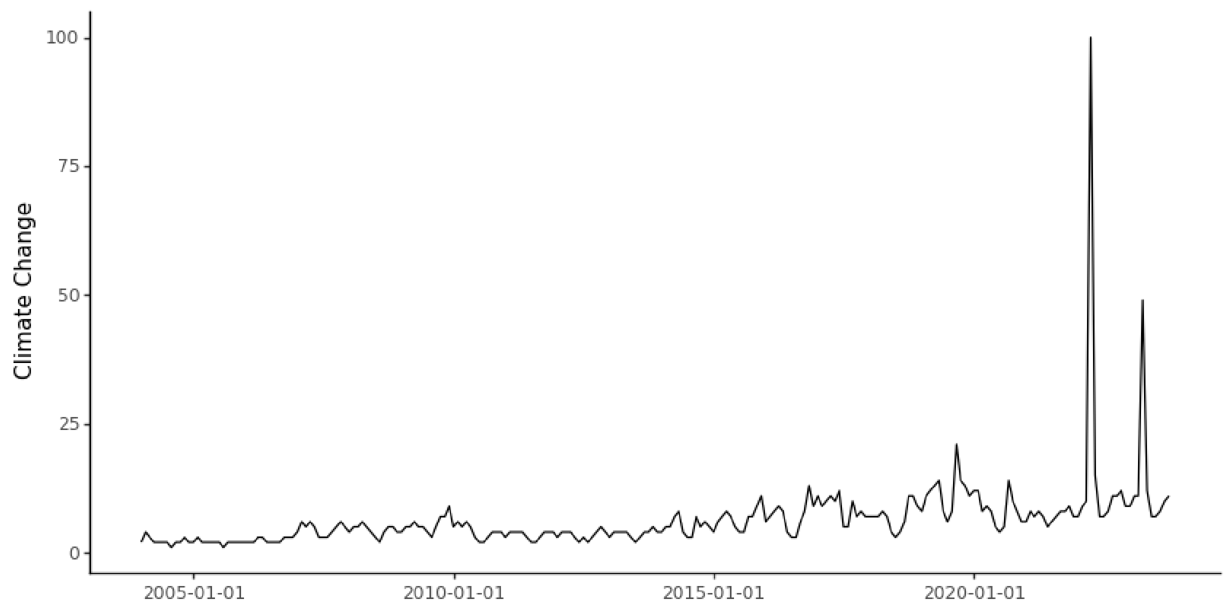
Go to [https://trends.google.com/trends/?geo=US (https://trends.google.com/trends/?geo=US)](https://trends.google.com/trends/?geo=US) and go to the page for the Climate Change topic.8 Change the timing to 2004-present and download the data.

Loading the Google Trends data. The range is 0-100.

In [84]:
```
df_google_trends = pd.read_csv("multiTimeline.csv",skiprows=2, index_col=0,

#Rename column to climate change
df_google_trends = df_google_trends.rename(columns={'climate change: (Unite

#PLot the google trends data using ggplot
p1 = (ggplot(df_google_trends, aes(x=df_google_trends.index, y="Climate Cha
p1
```



Out[84]: <ggplot: (703924218)>

i.

Pick three months where this measure was high. Can you find any news events from those months that might be leading Google searches for climate change to be particularly high? Discuss.

In [85]:
```
df_google_trends
```

Out[85]:

	Climate Change
Month	
2004-01-01	2
2004-02-01	4
2004-03-01	3
2004-04-01	2
2004-05-01	2
...	...
2023-06-01	7
2023-07-01	7
2023-08-01	8
2023-09-01	10
2023-10-01	11

238 rows × 1 columns

In [86]:
```
# Rank the google trends data to find the top 3 dates
df_google_trends.sort_values(by='Climate Change', ascending=False).head(3)
```

Out[86]:

	Climate Change
Month	
2022-04-01	100
2023-04-01	49
2019-09-01	21

Based on above graph and dates, we see the top three dates were April of 2022, 2023 and September of 2019.
We went ahead and tried to find any news events that might be leading Google searches for climate change to be particularly high. We found the following news articles that might be the reason for the spike in Google searches for climate change.
1. The April 2022 global surface temperature was 0.85°C (1.53°F) above the 20th century average and tied with 2010 as the fifth highest for April in the 143-year record. - https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202204#:~:text=The%20April%202022%20global%20surface,in%20the%20143%2Dy (https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202204#:~:text=The%20April%202022%20global%20surface,in%20the%20143%2Dy
2. The Southern Hemisphere had its warmest April and warmest month on record. As ocean temperatures set a record high for the month of April.- https://www.climate.gov/news-features/understanding-climate/global-climate-summary-april-2023#:~:text=Highlights,sea%20ice%20extent%20on% (https://www.climate.gov/news-features/understanding-climate/global-climate-summary-april-

2023#:~:text=Highlights,sea%20ice%20extent%20on%)
3.Globally, September 2019 was roughly 1.02 degrees Fahrenheit warmer than the average from 1981-2010, "making it the warmest September in our data record, although virtually on a par with 2016," the group said in a statement. -USA TODAY So it can be seen that the spike in Google searches for climate change is due to the increase in global temperature and global warming related news.

ii.

Merge the Google Trends data with the data from 49_Industry_Portfolios.csv. For each industry, regress returns on the Google Trends Climate Change topic score. Create a table with three columns: column 1 has the industry name, column 2 has the OLS regression coefficient and column 3 has the p-value for that coefficient. Sort the table from largest to smallest coefficient.

In [87]:
```
# Merge the google trends data with the industry data
df_merged = df.merge(df_google_trends, left_index=True, right_index=True)
df_merged
```

Out[87]:

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Trans	Whlsl	Rtail
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-4.07	2.95	-0.20
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	0.07	2.26	6.75
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.54	2.18	-0.13
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	0.21	0.91	-3.21
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	2.18	0.16	0.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-6.66	-5.41	-4.72
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	0.53	-1.04	-3.25
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	0.49	0.82	4.29
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	4.85	1.56	1.37
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-3.64	0.28	4.36

174 rows × 50 columns

In [88]:

```
def regression(df, regressor='Climate Change'):
    # Create a dataframe to store the results
    df_results = pd.DataFrame(columns=['Industry', 'Coefficient', 'P-Value'

    # Loop through each industry
    for industry in df.columns:
        # Create a dataframe with the industry and google trends data
        df_industry = df[[industry, regressor]]

        # Rename the columns
        df_industry.columns = ['Returns', 'Google Trends']

        # Drop any missing values
        df_industry = df_industry.dropna()

        # Run the regression
        model = sm.OLS(df_industry['Returns'], sm.add_constant(df_industry[
        results = model.fit()

        # Store the results in the dataframe
        df_results = df_results.append({'Industry': industry, 'Coefficient'

    return df_results

df_results = regression(df_merged, regressor='Climate Change')

# Sort the table from largest to smallest coefficient. Note: we are removin
df_results.sort_values(by='Coefficient', ascending=False)[1:]
```

Out[88]:

	Industry	Coefficient	P-Value
22	Autos	0.35786	0.16441
18	Steel	0.30430	0.26891
20	Mach	0.23978	0.25753
44	Banks	0.22550	0.25491
19	FabPr	0.22388	0.38807
40	Trans	0.21283	0.19123
28	Coal	0.21261	0.61145
36	Chips	0.20966	0.26276
13	Chems	0.20016	0.29226
37	LabEq	0.18126	0.27668
23	Aero	0.18028	0.28932
16	BldMt	0.17764	0.41725
11	MedEq	0.16881	0.25298
21	ElcEq	0.16859	0.39014
45	Insur	0.16776	0.32775
15	Txtls	0.16594	0.57324
47	Fin	0.16584	0.41724
38	Paper	0.16558	0.30482
27	Mines	0.15655	0.59983
33	BusSv	0.15498	0.28244
17	Cnstr	0.15220	0.49198
5	Toys	0.15109	0.44806
32	PerSv	0.14818	0.43797
7	Books	0.14508	0.46129
34	Hardw	0.13826	0.47012
6	Fun	0.13604	0.59916
0	Agric	0.12992	0.53702
35	Softw	0.12916	0.40722
46	RlEst	0.12177	0.66752
14	Rubbr	0.11595	0.54354
43	Meals	0.11330	0.38569
10	Hlth	0.10062	0.54365
41	Whlsl	0.09858	0.50418
42	Rtail	0.09146	0.49130
24	Ships	0.08064	0.74281

	Industry	Coefficient	P-Value
39	Boxes	0.06498	0.70558
12	Drugs	0.02809	0.82255
2	Soda	0.02445	0.89369
9	Clths	0.02145	0.90754
26	Gold	0.01713	0.96043
31	Telcm	0.01525	0.91173
3	Beer	-0.00576	0.95922
29	Oil	-0.00959	0.95930
48	Other	-0.01703	0.92458
25	Guns	-0.03852	0.81546
1	Food	-0.04846	0.65501
8	Hshld	-0.06546	0.58191
30	Util	-0.08726	0.45025
4	Smoke	-0.08926	0.57134

iii.

Comment on the ordering of the industries. Is it in line with what you would have expected?

Industries with Positive Coefficients:

Autos: This industry has the highest positive coefficient. This suggests that increased interest in climate change on Google Trends may have a positive impact on the automobile industry, possibly due to greater demand for environmentally friendly or electric vehicles.

Steel, Mach, Banks, FabPr: These industries also have positive coefficients, implying that they might benefit from higher climate change awareness. For example, the steel industry might see increased demand for materials used in renewable energy infrastructure.

Trans, Coal, Chips, Chems, LabEq, Aero: These industries also have positive coefficients but to a slightly lesser degree. They may experience some positive effects from heightened climate change concerns.

Industries with Negative Coefficients:

Smoke, Food, Hshld, Util, Smoke: These industries have negative coefficients, indicating that higher Google Trends Climate Change scores are associated with lower returns. This suggests that industries like tobacco, food, and utilities might face challenges or reduced demand in the context of increasing climate change awareness. For example, tobacco companies could face increased regulatory scrutiny.

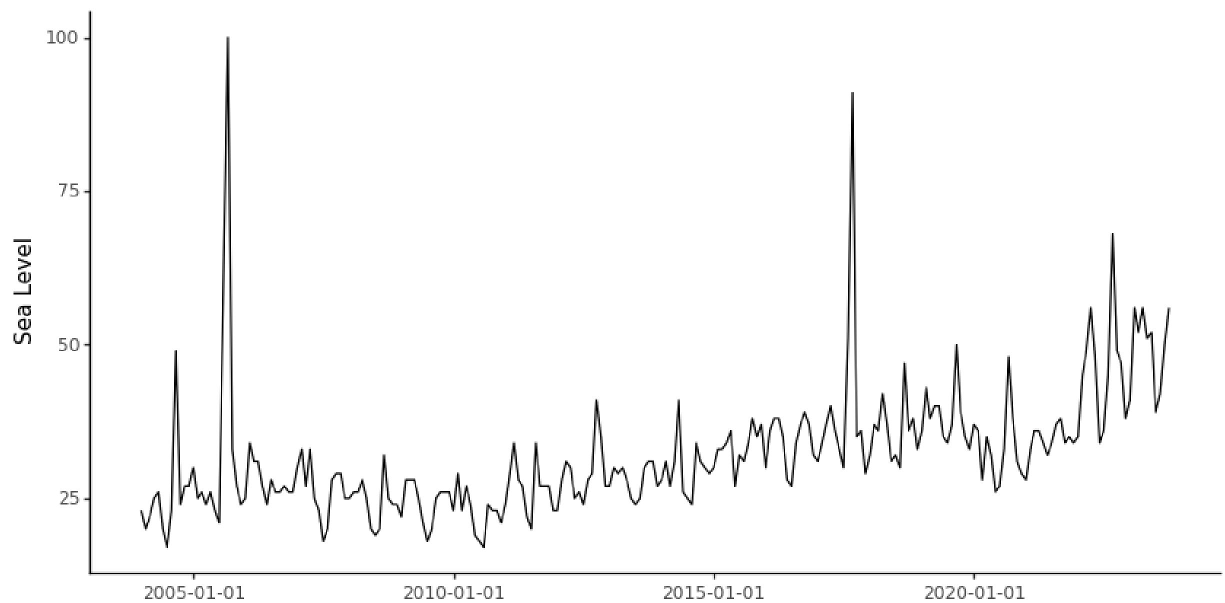
iv.

Repeat i.-iii. for two new climate change-related Google search terms of your choice. One of these terms should be related to physical risk (e.g. heat wave, hurricane, wildfire) and one should be related to transition risk (e.g. environmental regulation, carbon tax, cap and trade).

Physical Risk - "Sea Level"

In [89]:
```
df_google_trends1 = pd.read_csv("multiTimeline1.csv",skiprows=2, index_col=

#Rename column to climate change
df_google_trends1 = df_google_trends1.rename(columns={'Sea Level: (United S
#PLot the google trends data using ggplot
p2 = (ggplot(df_google_trends1, aes(x=df_google_trends1.index, y="Sea Level
p2
```



Out[89]: <ggplot: (703918728)>

In [90]:
```
# Rank the google trends data to find the top 3 dates
df_google_trends1.sort_values(by='Sea Level', ascending=False).head(3)
```

Out[90]:

Month	Sea Level
2005-09-01	100
2017-09-01	91
2022-09-01	68

Based on above dates, September 2005, 2017 and 2022-

1. Sep 2005: There are several news articles talking about severe Hurricanes in New Orleans. - https://edition.cnn.com/2005/WEATHER/09/01/orleans.levees/index.html (https://edition.cnn.com/2005/WEATHER/09/01/orleans.levees/index.html)
2. Sep 2017: There are several news articles talking about Hurricane Irma and Hurricane Maria. - https://www.cnn.com/2017/09/06/us/hurricane-irma-puerto-rico-florida/index.html (https://www.cnn.com/2017/09/06/us/hurricane-irma-puerto-rico-florida/index.html)
3. Sep 2022: There are several articles about Hurrican Ian and Hurrican Julia. - https://yaleclimateconnections.org/2022/09/ian-smashes-into-southwest-florida-with-historic-force/ (https://yaleclimateconnections.org/2022/09/ian-smashes-into-southwest-florida-with-historic-force/)

In [91]:
```
# Merge the google trends data with the industry data
df_merged1 = df.merge(df_google_trends1, left_index=True, right_index=True)
df_merged1
```

Out[91]:

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Trans	Whlsl	Rtail
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-4.07	2.95	-0.20
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	0.07	2.26	6.75
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.54	2.18	-0.13
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	0.21	0.91	-3.21
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	2.18	0.16	0.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-6.66	-5.41	-4.72
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	0.53	-1.04	-3.25
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	0.49	0.82	4.29
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	4.85	1.56	1.37
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-3.64	0.28	4.36

174 rows × 50 columns

In [92]:
```
df_results1 = regression(df_merged1, regressor='Sea Level')

# Sort the table from largest to smallest coefficient. Note: we are removin
df_results1.sort_values(by='Coefficient', ascending=False)[1:]
```

Out[92]:

	Industry	Coefficient	P-Value
47	Fin	0.07880	0.11929
26	Gold	0.06903	0.41973
29	Oil	0.06371	0.17040
16	BldMt	0.05738	0.29052
44	Banks	0.04833	0.32550
28	Coal	0.04683	0.65193
24	Ships	0.04525	0.45761
45	Insur	0.04345	0.30681
40	Trans	0.04189	0.30025
18	Steel	0.04081	0.55055
6	Fun	0.03595	0.57555
33	BusSv	0.03315	0.35430
36	Chips	0.03222	0.48837
21	ElcEq	0.02994	0.53864
17	Cnstr	0.02882	0.59999
27	Mines	0.02569	0.72866
20	Mach	0.02487	0.63642
25	Guns	0.02439	0.55128
37	LabEq	0.02395	0.56271
23	Aero	0.02093	0.62037
35	Softw	0.01860	0.63074
34	Hardw	0.01487	0.75436
38	Paper	0.01069	0.78964
5	Toys	0.00844	0.86448
22	Autos	0.00780	0.90301
14	Rubbr	0.00699	0.88270
41	Whlsl	0.00504	0.89063
48	Other	0.00500	0.91070
11	MedEq	0.00493	0.89320
0	Agric	0.00432	0.93408
13	Chems	0.00278	0.95308
30	Util	-0.00065	0.98191
46	RlEst	-0.00163	0.98148
42	Rtail	-0.00397	0.90418
7	Books	-0.00431	0.92969

	Industry	Coefficient	P-Value
4	Smoke	-0.00533	0.89174
19	FabPr	-0.00539	0.93326
43	Meals	-0.00749	0.81732
12	Drugs	-0.00890	0.77440
32	PerSv	-0.00903	0.84905
1	Food	-0.00935	0.72828
8	Hshld	-0.01264	0.66832
3	Beer	-0.01353	0.62826
15	Txtls	-0.01426	0.84538
10	Hlth	-0.01581	0.70050
39	Boxes	-0.02280	0.59298
9	Clths	-0.02703	0.55507
31	Telcm	-0.02895	0.39575
2	Soda	-0.05243	0.24716

Industries with Positive Coefficients:

Fin (Finance): The finance industry has the highest positive coefficient, indicating a positive association between the Google Trends "Sea Level" topic score and industry returns. This may suggest that the finance sector is positively influenced by increased interest or concerns related to sea level changes. This relationship could be due to investments in climate-resilient infrastructure or insurance related to sea level risks.

Gold, Oil, BldMt (Building Materials), Banks: These industries also have positive coefficients, suggesting that they might benefit from higher "Sea Level" topic scores. For instance, gold and oil industries might see increased demand due to sea level-related disruptions or investments in construction and banking sectors related to coastal development.

Ships, Insur, Trans, Steel: These industries show moderately positive coefficients, indicating some positive association with sea level-related trends. The shipping and insurance industries may be impacted by sea level changes, potentially leading to increased demand for their services.

Industries with Negative Coefficients:

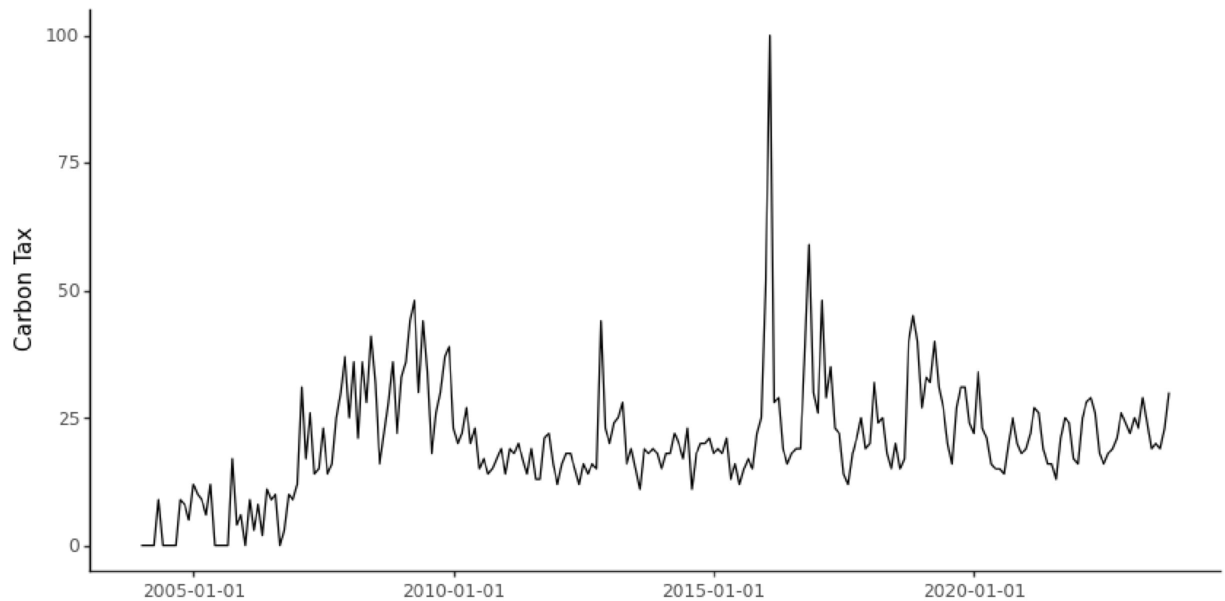
We do not see any straightforward explanation for the negative coefficients.

Transition Risk

In [93]:
```
df_google_trends2 = pd.read_csv("multiTimeline2.csv",skiprows=2, index_col=

#Rename column to climate change
df_google_trends2 = df_google_trends2.rename(columns={'carbon tax: (United

#PLot the google trends data using ggplot
p3 = (ggplot(df_google_trends2, aes(x=df_google_trends2.index, y="Carbon Ta
p3
```



Out[93]: <ggplot: (703923703)>

i.

Pick three months where this measure was high. Can you find any news events from those months that might be leading Google searches for climate change to be particularly high? Discuss.

In [94]:
```
df_google_trends2
```

Out[94]:

	Carbon Tax
Month	
2004-01-01	0
2004-02-01	0
2004-03-01	0
2004-04-01	0
2004-05-01	9
...	...
2023-06-01	19
2023-07-01	20
2023-08-01	19
2023-09-01	23
2023-10-01	30

238 rows × 1 columns

In [95]:
```
# Rank the google trends data to find the top 3 dates
df_google_trends2.sort_values(by='Carbon Tax', ascending=False).head(3)
```

Out[95]:

	Carbon Tax
Month	
2016-02-01	100
2016-11-01	59
2016-01-01	52

Based on above graph and dates, we see the top three dates were February and November and January respectively of 2016.

The Paris Agreement: In December 2015, the Paris Agreement was adopted, and in 2016, countries around the world were working on implementing their commitments to reduce greenhouse gas emissions. Carbon pricing mechanisms, including carbon taxes, were discussed as a means to achieve these emissions reduction targets. - United Nations

Canadian Carbon Pricing : In 2016, the Canadian government under Prime Minister Justin Trudeau announced its intention to implement a national carbon pricing system. This system would require provinces and territories to either implement their own carbon pricing mechanisms or adopt a federal carbon tax if they did not have an equivalent system in place. This initiative became known as the Pan-Canadian Framework on Clean Growth and Climate Change. - The New York Times

U.S. State-Level Initiatives: While there was no federal carbon tax in the United States in 2016, several states were taking action to implement their own carbon pricing mechanisms. Notable examples included California's cap-and-trade system and the Regional Greenhouse Gas Initiative (RGGI) in the northeastern U.S. - The New York Times

ii.

Merge the Google Trends data with the data from 49_Industry_Portfolios.csv. For each industry, regress returns on the Google Trends Climate Change topic score. Create a table with three columns: column 1 has the industry name, column 2 has the OLS regression coefficient and column 3 has the p-value for that coefficient. Sort the table from largest to smallest coefficient.

In [96]:
```python
# Merge the google trends data with the industry data
df_merged2 = df.merge(df_google_trends2, left_index=True, right_index=True)
df_merged2
```

Out[96]:

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Trans	Whlsl	Rtail
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-4.07	2.95	-0.20
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	0.07	2.26	6.75
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.54	2.18	-0.13
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	0.21	0.91	-3.21
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	2.18	0.16	0.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-6.66	-5.41	-4.72
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	0.53	-1.04	-3.25
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	0.49	0.82	4.29
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	4.85	1.56	1.37
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-3.64	0.28	4.36

174 rows × 50 columns

In [97]:
```
df_results2 = regression(df_merged2, regressor= 'Carbon Tax')

# Sort the table from largest to smallest coefficient. Note: we are removin
df_results2.sort_values(by='Coefficient', ascending=False)[1:]
```

Out[97]:

	Industry	Coefficient	P-Value
19	FabPr	0.08219	0.08979
26	Gold	0.07233	0.26246
22	Autos	0.03702	0.44354
5	Toys	0.03351	0.36898
28	Coal	0.03155	0.68743
13	Chems	0.03062	0.39010
34	Hardw	0.02593	0.46967
15	Txtls	0.02267	0.68131
38	Paper	0.02193	0.46843
20	Mach	0.02130	0.59192
18	Steel	0.02056	0.69056
40	Trans	0.01846	0.54583
27	Mines	0.01488	0.79010
36	Chips	0.01415	0.68718
21	ElcEq	0.00953	0.79562
0	Agric	0.00075	0.98489
16	BldMt	-0.00149	0.97113
14	Rubbr	-0.00281	0.93744
41	Whlsl	-0.00386	0.88896
10	Hlth	-0.00481	0.87701
37	LabEq	-0.00657	0.83351
33	BusSv	-0.00890	0.74217
43	Meals	-0.00891	0.71605
32	PerSv	-0.00926	0.79598
7	Books	-0.00935	0.80009
24	Ships	-0.00964	0.83411
3	Beer	-0.01034	0.62401
31	Telcm	-0.01164	0.65150
23	Aero	-0.01355	0.67108
42	Rtail	-0.01385	0.57818
11	MedEq	-0.01461	0.59794
1	Food	-0.01636	0.42039
35	Softw	-0.01639	0.57467
2	Soda	-0.01658	0.62858
39	Boxes	-0.01713	0.59502

	Industry	Coefficient	P-Value
30	Util	-0.01714	0.42864
9	Clths	-0.02027	0.55778
48	Other	-0.02064	0.54000
12	Drugs	-0.02104	0.36938
4	Smoke	-0.02134	0.46985
8	Hshld	-0.02216	0.31941
45	Insur	-0.02570	0.42372
25	Guns	-0.02593	0.40133
47	Fin	-0.02717	0.47819
46	RlEst	-0.02725	0.60780
6	Fun	-0.03168	0.51343
44	Banks	-0.03498	0.34618
17	Cnstr	-0.03531	0.39459
29	Oil	-0.05163	0.14114

iii.

Comment on the ordering of the industries. Is it in line with what you would have expected?

Positive Coefficients : Industries with positive coefficients are perceived to benefit or be neutral to discussions on carbon taxes. These could be sectors that are less carbon-intensive or those that might have solutions or products that would benefit from increased carbon taxes.

FabPr: This suggests that as discussions on carbon tax increase, that Fabricated Products provides alternative solutions that could benefit from a more stringent carbon tax regime or is relatively less affected by it.

Gold : This being second is interesting. One potential explanation could be that gold, as a non-productive asset, might be seen as a safe-haven during times of economic uncertainty, which discussions on carbon tax might bring about for certain sectors.

Autos: Given the push for electric vehicles and reducing carbon footprints, it's possible that positive discussions on carbon tax align with periods where electric or more efficient cars are in demand or being promoted.

Coal: Surprisingly, Coal has a positive coefficient. Given that coal is a significant carbon emitter, one might expect a negative relationship. However, this result could be due to various external factors not accounted for in this analysis or just statistical noise.

Negative Coefficients : Industries with negative coefficients are perceived to be adversely affected by discussions on carbon taxes. These could be sectors that are carbon-intensive and may face increased costs or regulations with higher carbon taxes.

Oil: Oil, being at the bottom, suggests it has the most negative relationship with the carbon tax topic. This is in line with expectations as the oil industry would be one of the most affected by stringent carbon tax policies.

Banks, Cnstr, RlEst: These sectors might have investments or dependencies on carbon-intensive industries, which could explain their negative coefficients.

(b) (25 points)

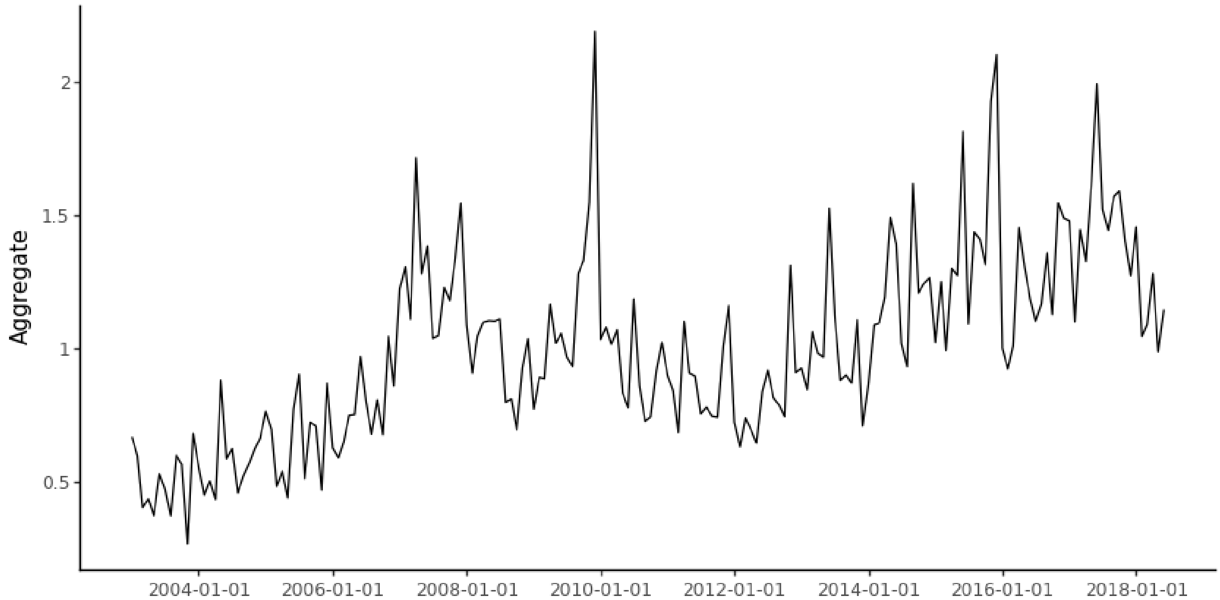
Download the Ardia et al. (2020) data from https://sentometrics-research.com/download/mccc/ (https://sentometrics-research.com/download/mccc/). Focus on the tab "SSRN 2022 version (monthly)." Repeat problem 3a using this data instead of Google Trends. Instead of Google searches for the Climate Change topic, use the column called "Aggregate." Instead of Google searches for physical and transition risk topics, choose one physical risk-related and one transition-risk related series from the Ardia et al. (2020) data. Try to choose the series that are most closely related to the Google searches you chose above. Comment on the similarities and differences in the industry rankings you find when using the Ardia et al. (2020) compared to the Google data.

Note: We did not get the SSRN 2022 monthly, so we used the MS 2022 monthly data.

Repeating 3a for Climate Change using the Aggregate column from the Ardia et al. (2020) data

In [100]:
```
df_ms_2022 = pd.read_excel("Sentometrics_US_Media_Climate_Change_Index.xlsx

#PLot the google trends data using ggplot
p4 = (ggplot(df_ms_2022, aes(x=df_ms_2022.index, y="Aggregate"))+ p9.geom_l
p4
```



Out[100]: <ggplot: (705509208)>

i.

Pick three months where this measure was high. Can you find any news events from those months that might be leading Google searches for **AGGREGATE** to be particularly high? Discuss.

In [103]:
```
# Rank the aggregate to find the top 3 dates
df_ms_2022.sort_values(by='Aggregate', ascending=False).head(3)[["Aggregate
```

Out[103]:

Date	Aggregate
2009-12-01	2.18944
2015-12-01	2.10294
2017-06-01	1.99219

Based on above dates-

1. Dec 2009: The Copenhagen Climate Change Conference was held in December 2009. This conference was a key step in the international climate change negotiations process. - United

Nations

2. Dec 2015: The Paris Agreement was adopted in December 2015. This agreement was a key milestone in the international climate change negotiations process. - United Nations

3. Jun 2017: President Donald Trump announced that the United States would withdraw from the Paris Agreement. This decision was widely criticized by the international community. - The New York Times

ii.

Merge the Google Trends data with the data from 49_Industry_Portfolios.csv. For each industry, regress returns on the Google Trends Climate Change topic score. Create a table with three columns: column 1 has the industry name, column 2 has the OLS regression coefficient and column 3 has the p-value for that coefficient. Sort the table from largest to smallest coefficient.

In [105]:
```
df_ms_2022_agg = df_ms_2022[["Aggregate"]]
# Merge the aggregate data with the industry data
df_merged3 = df.merge(df_ms_2022_agg, left_index=True, right_index=True)
df_merged3
```

Out[105]:

Date	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Trans	Whlsl	Rtail
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-4.07	2.95	-0.20
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	0.07	2.26	6.75
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.54	2.18	-0.13
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	0.21	0.91	-3.21
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	2.18	0.16	0.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-6.66	-5.41	-4.72
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	0.53	-1.04	-3.25
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	0.49	0.82	4.29
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	4.85	1.56	1.37
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-3.64	0.28	4.36

174 rows × 50 columns

In [108]:
```
df_results3 = regression(df_merged3, regressor='Aggregate')

df_results3.sort_values(by='Coefficient', ascending=False)
```

Out[108]:

	Industry	Coefficient	P-Value
22	Autos	2.83871	0.11663
15	Txtls	2.25385	0.27617
11	MedEq	2.04097	0.04861
36	Chips	2.03043	0.12252
0	Agric	1.99174	0.17760
18	Steel	1.98815	0.30463
7	Books	1.84482	0.18218
20	Mach	1.73651	0.24371
45	Insur	1.71237	0.15495
37	LabEq	1.60630	0.17013
35	Softw	1.53353	0.16110
23	Aero	1.52778	0.20145
44	Banks	1.51453	0.27714
38	Paper	1.44399	0.20296
13	Chems	1.42210	0.28744
21	ElcEq	1.41038	0.30660
47	Fin	1.38343	0.33589
33	BusSv	1.36693	0.17736
32	PerSv	1.29542	0.33488
39	Boxes	1.13508	0.34769
3	Beer	1.00533	0.20364
49	Aggregate	1.00000	0.00000
12	Drugs	0.99587	0.25746
40	Trans	0.95628	0.40454
46	RlEst	0.86208	0.66552
16	BldMt	0.84244	0.58468
34	Hardw	0.83986	0.53292
6	Fun	0.79344	0.66304
2	Soda	0.78862	0.53988
48	Other	0.78447	0.53508
42	Rtail	0.68888	0.46115
27	Mines	0.62955	0.76431
10	Hlth	0.61218	0.59947
17	Cnstr	0.56235	0.71829
14	Rubbr	0.55761	0.67803

	Industry	Coefficient	P-Value
25	Guns	0.53474	0.64505
41	Whlsl	0.50439	0.62721
31	Telcm	0.48374	0.61705
9	Clths	0.41925	0.74691
19	FabPr	0.32288	0.85970
8	Hshld	0.27457	0.74277
24	Ships	0.22944	0.89443
43	Meals	0.15163	0.86907
5	Toys	0.12120	0.93112
28	Coal	-0.02947	0.99202
1	Food	-0.04111	0.95704
30	Util	-0.38343	0.63739
26	Gold	-0.38989	0.87243
29	Oil	-0.40779	0.75760
4	Smoke	-0.78154	0.48097

iii.

Comment on the ordering of the industries. Is it in line with what you would have expected?

Industries with Positive Coefficients:

Industries with Positive Coefficients:

Auto:Automobiles are considered to be one of the major contributors to global Climate change. With advancements in EV it could be a positive aspect for nature.

Txtls (Textiles): The positive relationship might be influenced by various factors related to the textile industry's performance.

Chips (Semiconductor): The semiconductor industry also shows a positive coefficient. This may indicate that possible advancements or innovations in technology also affects the environmental climate.

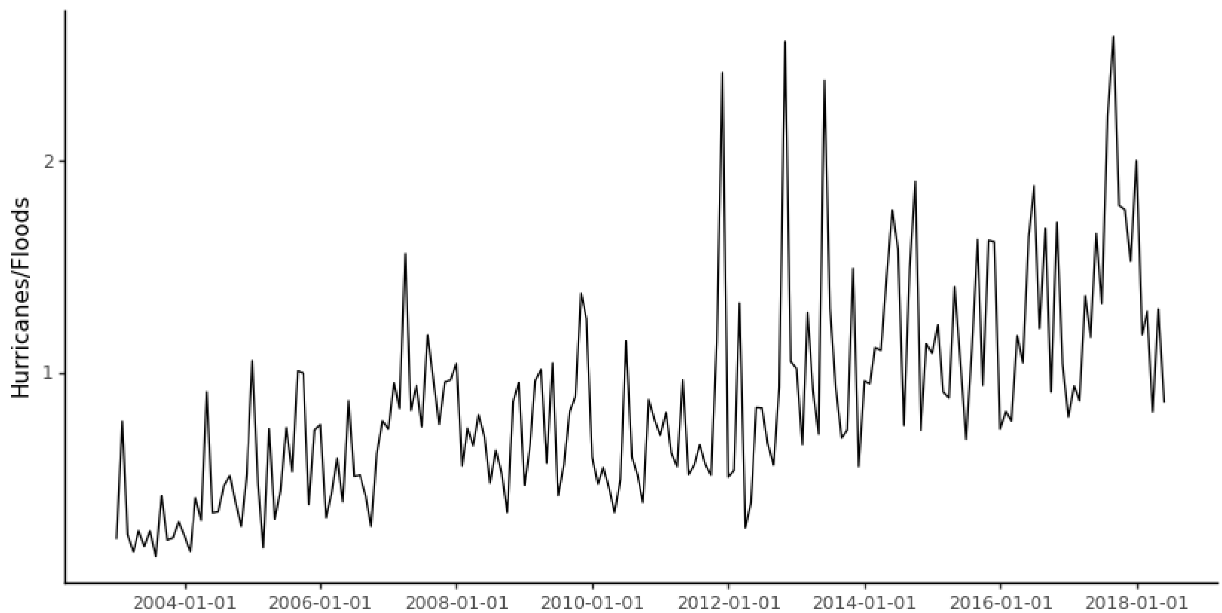
Industries with Negative Coefficients:

Smoke: The tobacco industry has a negative coefficient, indicating that higher Aggregate scores are associated with lower returns. This suggests that industries like tobacco could face increased regulatory scrutiny.

Oil: The oil industry also has a negative coefficient, suggesting that it might be negatively impacted by higher Aggregate scores. This could be due to factors like increased regulation or reduced demand for oil.

Repeating 3a for Physical Risk - "Hurricanes/Floods" Ardia et al. (2020) data

In [109]:
```
#PLot the google trends data using ggplot
p5 = (ggplot(df_ms_2022, aes(x=df_ms_2022.index, y="Hurricanes/Floods"))+ p
p5
```



Out[109]: <ggplot: (702813046)>

In [117]:
```
# Rank the hurricane to find the top 3 dates
df_ms_2022.sort_values(by='Hurricanes/Floods', ascending=False).head(3)[["H
```

Out[117]:

Date	Hurricanes/Floods
2017-09-01	2.58478
2012-11-01	2.56125
2011-12-01	2.41643

Based on above dates-

1. Sep 2017: There are several news articles talking about Hurricane Irma and Hurricane Maria. - [https://www.cnn.com/2017/09/06/us/hurricane-irma-puerto-rico-florida/index.html (https://www.cnn.com/2017/09/06/us/hurricane-irma-puerto-rico-florida/index.html)](https://www.cnn.com/2017/09/06/us/hurricane-irma-puerto-rico-florida/index.html)

2. Nov 2012: Hurricane Sandy was a devastating storm that hit the East Coast of the United States in November 2012. - The New York Times

3. Dec 2011: The 2011 Thailand floods were a major disaster that affected the country in December 2011. - The New York Times

In [111]:
```
df_ms_2022_hur = df_ms_2022[["Hurricanes/Floods"]]
# Merge the aggregate data with the industry data
df_merged4 = df.merge(df_ms_2022_hur, left_index=True, right_index=True)
df_merged4
```

Out[111]:

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Trans	Whlsl	Rtail
Date														
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-4.07	2.95	-0.20
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	0.07	2.26	6.75
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.54	2.18	-0.13
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	0.21	0.91	-3.21
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	2.18	0.16	0.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-6.66	-5.41	-4.72
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	0.53	-1.04	-3.25
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	0.49	0.82	4.29
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	4.85	1.56	1.37
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-3.64	0.28	4.36

174 rows × 50 columns

In [113]:
```
df_results4 = regression(df_merged4, regressor='Hurricanes/Floods')

df_results4.sort_values(by='Coefficient', ascending=False)
```

Out[113]:

	Industry	Coefficient	P-Value
36	Chips	1.84697	0.04599
44	Banks	1.82057	0.06323
18	Steel	1.68128	0.21828
47	Fin	1.65140	0.10265
22	Autos	1.63654	0.20039
11	MedEq	1.56307	0.03215
7	Books	1.53204	0.11603
45	Insur	1.43364	0.09111
25	Guns	1.41523	0.08277
33	BusSv	1.34391	0.05953
15	Txtls	1.31448	0.36834
38	Paper	1.22214	0.12633
12	Drugs	1.21677	0.04904
32	PerSv	1.18214	0.21192
23	Aero	1.06967	0.20498
40	Trans	1.06827	0.18628
21	ElcEq	1.01463	0.29721
49	Hurricanes/Floods	1.00000	0.00000
0	Agric	0.98784	0.34397
3	Beer	0.96143	0.08428
37	LabEq	0.92321	0.26428
20	Mach	0.90022	0.39216
16	BldMt	0.88284	0.41670
42	Rtail	0.77758	0.23785
46	RlEst	0.76414	0.58703
48	Other	0.69337	0.43703
35	Softw	0.68593	0.37518
17	Cnstr	0.66718	0.54400
6	Fun	0.52012	0.68565
10	Hlth	0.48685	0.55388
34	Hardw	0.43724	0.64555
39	Boxes	0.42934	0.61500
13	Chems	0.42534	0.65250
41	Whlsl	0.41741	0.56891
24	Ships	0.33816	0.78162

	Industry	Coefficient	P-Value
43	Meals	0.29725	0.64687
31	Telcm	0.25444	0.70938
14	Rubbr	0.17763	0.85136
8	Hshld	0.16787	0.77613
2	Soda	0.05451	0.95214
9	Clths	-0.15685	0.86415
30	Util	-0.16501	0.77380
19	FabPr	-0.17032	0.89487
29	Oil	-0.17713	0.84933
27	Mines	-0.24192	0.87031
1	Food	-0.32167	0.55010
4	Smoke	-0.71782	0.35872
5	Toys	-0.81745	0.40821
26	Gold	-1.07951	0.52843
28	Coal	-1.64721	0.42757

Industries with Positive Coefficients for regression against Hurricane/Floods:

Banks : Possibly the positive relationship could be due to the fact that banks might be involved in providing loans for rebuilding after natural disasters.

Steel: The positive relationship might be influenced by various factors related to the steel industry's performance, again indication the rebuilding part.

Fin: The finance industry also shows a positive coefficient, possibly being part of the rebuilding financing.

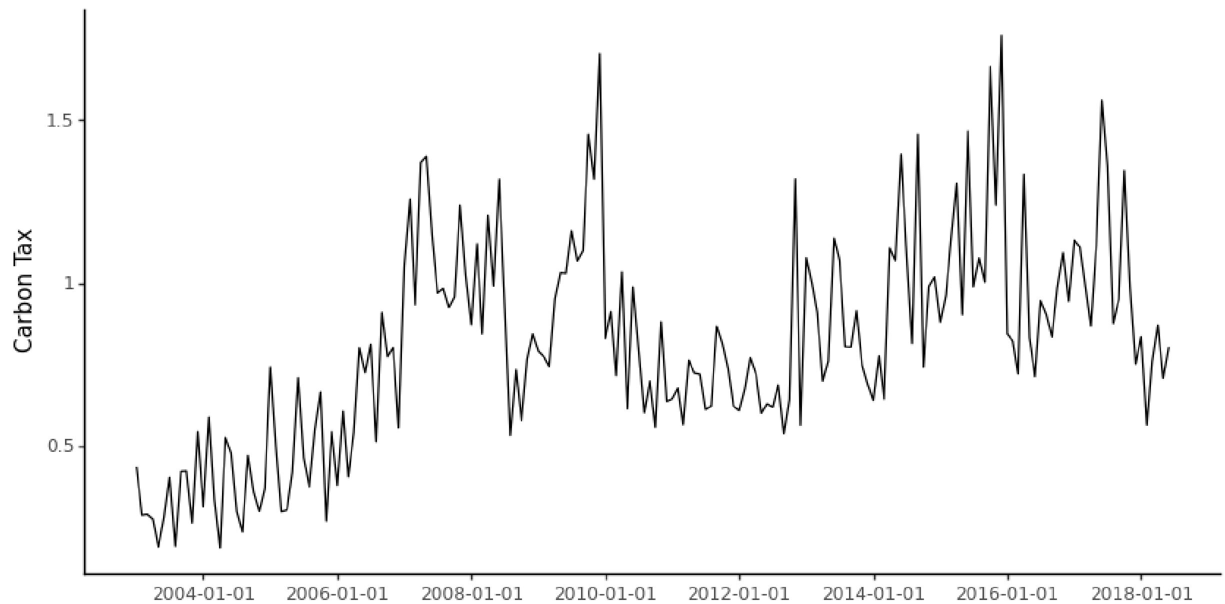
Industries with Negative Coefficients:

Coal: The coal industry has a negative coefficient, indicating that higher Hurricane/Floods scores are associated with lower returns. This suggests that industries like coal could face increased regulatory scrutiny.

Smoke: The tobacco industry also has a negative coefficient, suggesting that it might be negatively impacted by higher Hurricane/Floods scores. This could be due to factors like increased regulation or reduced demand for tobacco products.

Repeating 3a for Transition Risk - "Carbon Tax" Ardia et al. (2020) data

In [114]:
```
#PLot the Carbon Tax data using ggplot
p6 = (ggplot(df_ms_2022, aes(x=df_ms_2022.index, y="Carbon Tax"))+ p9.geom_
p6
```



Out[114]: <ggplot: (707364542)>

In [116]:
```
# Rank the hurricane to find the top 3 dates
df_ms_2022.sort_values(by='Carbon Tax', ascending=False).head(3)[["Carbon T
```

Out[116]:

	Carbon Tax
Date	
2015-12-01	1.75918
2009-12-01	1.70334
2015-10-01	1.66414

Based on above dates-

1. Dec 2015: The Paris Agreement was adopted in December 2015. This agreement was a key milestone in the international climate change negotiations process. - United Nations

2. Dec 2009: The Copenhagen Climate Change Conference was held in December 2009. This conference was a key step in the international climate change negotiations process. - United Nations

3. Oct 2015: The World Bank released a report on carbon pricing in October 2015. This report was a key milestone in the international climate change negotiations process. - World Bank

In [118]:
```
df_ms_2022_ct = df_ms_2022[["Carbon Tax"]]
# Merge the Carbon Tax data with the industry data
df_merged5 = df.merge(df_ms_2022_ct, left_index=True, right_index=True)
df_merged5
```

Out[118]:

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...	Trans	Whlsl	Rtail
Date														
2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	-1.34	...	-4.07	2.95	-0.20
2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	5.74	...	0.07	2.26	6.75
2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	3.33	...	-0.54	2.18	-0.13
2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	-2.13	...	0.21	0.91	-3.21
2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	0.04	...	2.18	0.16	0.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	-1.95	...	-6.66	-5.41	-4.72
2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	0.64	...	0.53	-1.04	-3.25
2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	3.16	...	0.49	0.82	4.29
2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	4.21	...	4.85	1.56	1.37
2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	6.98	...	-3.64	0.28	4.36

174 rows × 50 columns

In [119]:
```
df_results5 = regression(df_merged5, regressor='Carbon Tax')

df_results5.sort_values(by='Coefficient', ascending=False)
```

Out[119]:

	Industry	Coefficient	P-Value
15	Txtls	3.26036	0.15481
22	Autos	2.15959	0.28249
13	Chems	1.86968	0.20679
20	Mach	1.55493	0.34671
35	Softw	1.42572	0.24028
11	MedEq	1.42213	0.21673
38	Paper	1.42003	0.25896
21	ElcEq	1.40638	0.35790
3	Beer	1.32880	0.12920
37	LabEq	1.32223	0.30898
7	Books	1.29843	0.39773
27	Mines	1.24280	0.59329
18	Steel	1.22973	0.56721
5	Toys	1.19665	0.44099
36	Chips	1.18920	0.41573
12	Drugs	1.18913	0.22242
49	Carbon Tax	1.00000	0.00000
23	Aero	0.93448	0.48163
45	Insur	0.87648	0.51236
33	BusSv	0.84470	0.45291
0	Agric	0.78123	0.63414
47	Fin	0.77055	0.62908
26	Gold	0.74876	0.78087
34	Hardw	0.67293	0.65230
39	Boxes	0.60442	0.65225
19	FabPr	0.60069	0.76672
1	Food	0.55456	0.51192
42	Rtail	0.48677	0.63869
9	Clths	0.46759	0.74542
44	Banks	0.43273	0.77974
2	Soda	0.36698	0.79702
40	Trans	0.24880	0.84505
14	Rubbr	0.23653	0.87381
32	PerSv	0.21762	0.88396
8	Hshld	0.20956	0.82125

	Industry	Coefficient	P-Value
16	BldMt	0.20627	0.90395
41	Whlsl	0.20267	0.86031
31	Telcm	0.18748	0.86126
10	Hlth	0.17439	0.89269
46	RlEst	-0.06943	0.97495
48	Other	-0.13029	0.92600
6	Fun	-0.14578	0.94245
24	Ships	-0.20377	0.91533
43	Meals	-0.20910	0.83752
29	Oil	-0.25239	0.86321
25	Guns	-0.33607	0.79402
30	Util	-0.46006	0.60992
4	Smoke	-0.49219	0.68903
17	Cnstr	-0.97826	0.57122
28	Coal	-1.03333	0.75159

Industries with Positive Coefficients for regression against Carbon Tax:

Autos: The positive relationship to Carbon Tax would be due to the fact that the auto industry is one of the major contributors to global Climate change. With advancements in EV it could be a positive aspect for nature.

Chems (Chemicals) and Mach (Machinery): Chemicals and Machinery being positive could be due to the fact that these industries are involved in providing solutions to reduce carbon emissions.

Industries with Negative Coefficients:

Coal Oil Guns & Smoke: The coal, oil, guns and smoke industries have a negative coefficient, indicating that higher Carbon Tax scores are associated with lower returns. This suggests that industries like coal, oil, guns and smoke could face increased regulatory scrutiny.

3 Reading Response (15 points)

Optional for extra credit. Must be completed individually.

Read Lins et al. (2017) and answer the following questions.

1. (5 points)

Why are the authors particularly interested in studying the performance of socially responsible firms during crisis times (instead of studying performance of these firms more generally)?

2. (5 points)

Focusing on Section IV, describe the authors' findings related to the effect of CSR on operating performance and capital raising. Conceptually, explain why CSR would matter for these outcomes during crisis times?

3. (5 points)

What questions are you left with after reading this paper?

In [ ]:

```

```