Topics in Econ HW1

Group 4: Elie Kostenbaum, Jingwen Li, Moushumi Pardesi

In [ ]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import warnings
warnings.filterwarnings('ignore')
```

Part 1: ESG

Q1

(a) (5 points) In your own words, explain what the unadjusted greenness score, Git, measures. Make sure to mention why we need to include E weight.

EK: The unadjusted greenness score Git, measures how environmentally sustainable or "green" is a company in comparison to others. This score takes innto account two factors:

- E_score: This reflects the company's weighted-average performance on various environmental issues such as climate change, pollution, waste etc... These scores are an indicator of a company's resilience to long term environmental risks. The E_score is a number between 0 and A0, with higher values indicating better environmental performance.

- E_weight: This represents the importance of environmental issues relative to social and governance issues for the company. It is a number between 0 and 100 and is typically consistent across firms in the same industry. Higher E_weight values signify that environmental concerns carry more weight in that industry.

Including E_weight is crucial because it adjusts the greenness score based on the industry context. Different industries have varying levels of environmental impact, and some may be inherently more environmentally intensive than others. E_weight reflects how significant environmental concerns are within a particular industry. Including E_weight in the calculation allows for a fairer assessment of greenness, considering the industry-specific environmental standards and expectations.

(Both factors are provided by MSCI).

JL: The unadjusted greenness score, $G_{i,t}$, provides a holistic measure of a company's environmental performance by taking into account both its particular environmental practices and the overarching context of its industry. This measure is derived using the company's environmental score, $E_{score_{i,t}}$, and the environmental significance of its industry, $E_{weight_{i,t}}$, as determined by the MSCI ratings, using the most recent data available.

A key component of this calculation is how the company's environmental score stands relative to a perfect score. The expression $10 - E_{score_{i,t}}$ computes the difference between the company's score and the ideal score of 10, giving us an idea of the company's environmental shortcomings. This deviation is then weighted by $E_{weight_{i,t}}$, which signifies the environmental relevance or impact of that specific industry. The multiplication of these two values produces a measure of the firm's "brownness" — essentially gauging its deviation from eco-friendly behavior in relation to its industry's environmental footprint. However, to make this metric more intuitive, its sign is inverted, with the goal of denoting "greenness." Consequently, a higher value implies more eco-friendly practices, especially in comparison to the industry's environmental consequences.

The integration of E_{weight} is paramount for a nuanced analysis. Without it, two firms from disparate sectors but with identical E_{score} values could be perceived as equally green, which could be misleading. By incorporating E_{weight}, the metric offers a context-sensitive perspective on a company's greenness, effectively situating its environmental performance within the inherent environmental implications of its industry.

MP: The unadjusted greenness score,$G_{i,t}$, is a measure of the environmental performance of a company i at time t. It is calculated as a weighted sum of the company's scores on a set of environmental indicators, such as greenhouse gas emissions, water pollution, and waste production. The weights reflect the relative importance of each indicator.

E_{weight} allows us to account for the different environmental impacts of different industries. By including E_{weight} , we can ensure that companies in different industries are compared on an apple to apple basis. Additionally, E_{weight} allows us to account for the different levels of environmental regulation in different countries and regions. It allows that companies in different countries and regions are compared on an equal footing.

(b) (5 points) Why does the paper focus on the adjusted greenness score?

JL: The paper emphasizes the adjusted greenness score, gi, t, to capture a company's environmental stance in relation to the overarching market. By subtracting the market's value-weighted average greenness, Gt, from an individual firm's score, the adjusted metric, gi, t, highlights how green or brown a firm stands compared to the market's mean. This method mitigates the influence of absolute greenness values and fosters a nuanced comparison across firms, thus facilitating a precise understanding of a company's environmental performance against its peers.

MP: gi, t is a more accurate measure of a firm's environmental performance than raw greenness score because raw score does not take into account the size of the firm or the industry in which it operates. gi, t does take these factors into account, making it a more reliable measure of a firm's environmental performance. gi, t is more useful for investors who are looking to identify companies with good environmental performance because it lets investors to compare companies of different sizes and from different industries on an equal footing.

Q2. Use the description in Section 4 of the paper to replicate Figure 3

- Following equations 1 and 2 of P ́astor et al. (2022), compute the firm-level greenness measure, gi,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, G ̄t is the market cap weighted average.

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

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

In [ ]:
```
df1 = pd.read_csv('MSCI_sample.CSV')
df1['G'] = -(10-df1['ENVIRONMENTAL_PILLAR_SCORE'])*df1['ENVIRONMENTAL_PILLAF
df1['MKT_CAP'] = df1['PRC'] * df1['SHROUT']
df1['Weighted_G'] = df1['G'] * df1['MKT_CAP']
monthly_weighted_avg_G = df1.groupby(df1['AS_OF_DATE']).apply(lambda x: (x['
df1['Monthly_Weighted_Avg_G'] = df1['AS_OF_DATE'].map(monthly_weighted_avg_G
df1['g'] = df1['G'] - df1['Monthly_Weighted_Avg_G']
```

In [ ]:
```
# Convert the columns to numeric data types
df1['RET'] = pd.to_numeric(df1['RET'], errors='coerce')
df1['MKT_CAP'] = pd.to_numeric(df1['MKT_CAP'], errors='coerce')

# Replace NaN values with 0
df1[['RET', 'MKT_CAP']] = df1[['RET', 'MKT_CAP']].fillna(0)
```

In [ ]:
```
df1
```

Out[ ]:

	ISSUER_NAME	ISSUERID	AS_OF_DATE	IVA_INDUSTRY	INDUSTRY_
0	QEP Resources, Inc.	IID000000002231971	2012-01-01	Oil & Gas Exploration & Production	
1	IAMGOLD Corporation	IID000000002169868	2012-01-01	Metals and Mining - Precious Metals	
2	Las Vegas Sands Corp.	IID000000002173531	2012-01-01	Casinos & Gaming	
3	Honeywell International Inc.	IID000000002127578	2012-01-01	Aerospace & Defense	
4	Staples, Inc.	IID000000002145070	2012-01-01	Specialty Retail	
...	...	...	...	...	
49332	ANALOG DEVICES, INC.	IID000000002157388	2021-12-01	Semiconductors & Semiconductor Equipment	
49333	AMERICAN WOODMARK CORPORATION	IID000000002157284	2021-12-01	Building Products	
49334	AMERICAN VANGUARD CORPORATION	IID000000002157281	2021-12-01	Specialty Chemicals	
49335	BANK OF MARIN BANCORP	IID000000002159126	2021-12-01	Banks	
49336	SYLVAMO CORPORATION	IID000000005050010	2021-12-01	Paper & Forest Products	

49337 rows × 20 columns

In [ ]:
```
# Define a function to categorize based on greenness score within each group
def categorize_by_score(group):
    group['Greenness_Category'] = pd.qcut(group['g'], q=3, labels=['brown',
    return group
```

```
# Apply the function to each date's group
df_q2 = df1.groupby('AS_OF_DATE').apply(categorize_by_score)
```

In [ ]:
```
def value_weighted_return(group):
    group['Green_Return'] = (group[group['Greenness_Category'] == 'green']['
    group['Brown_Return'] = (group[group['Greenness_Category'] == 'brown']['
    return group

df_q2 = df_q2.drop(columns='AS_OF_DATE').reset_index().groupby('AS_OF_DATE')
```

In [ ]:
```
RET_df = df_q2[['AS_OF_DATE', 'Green_Return', 'Brown_Return']].drop_duplicat
RET_df['Cumulative_Green_Return'] = (1 + RET_df['Green_Return']).cumprod() -
RET_df['Cumulative_Brown_Return'] = (1 + RET_df['Brown_Return']).cumprod() -
RET_df = RET_df * 100
RET_df.index = pd.to_datetime(RET_df.index)
```

In [ ]:
```
RET_df
```

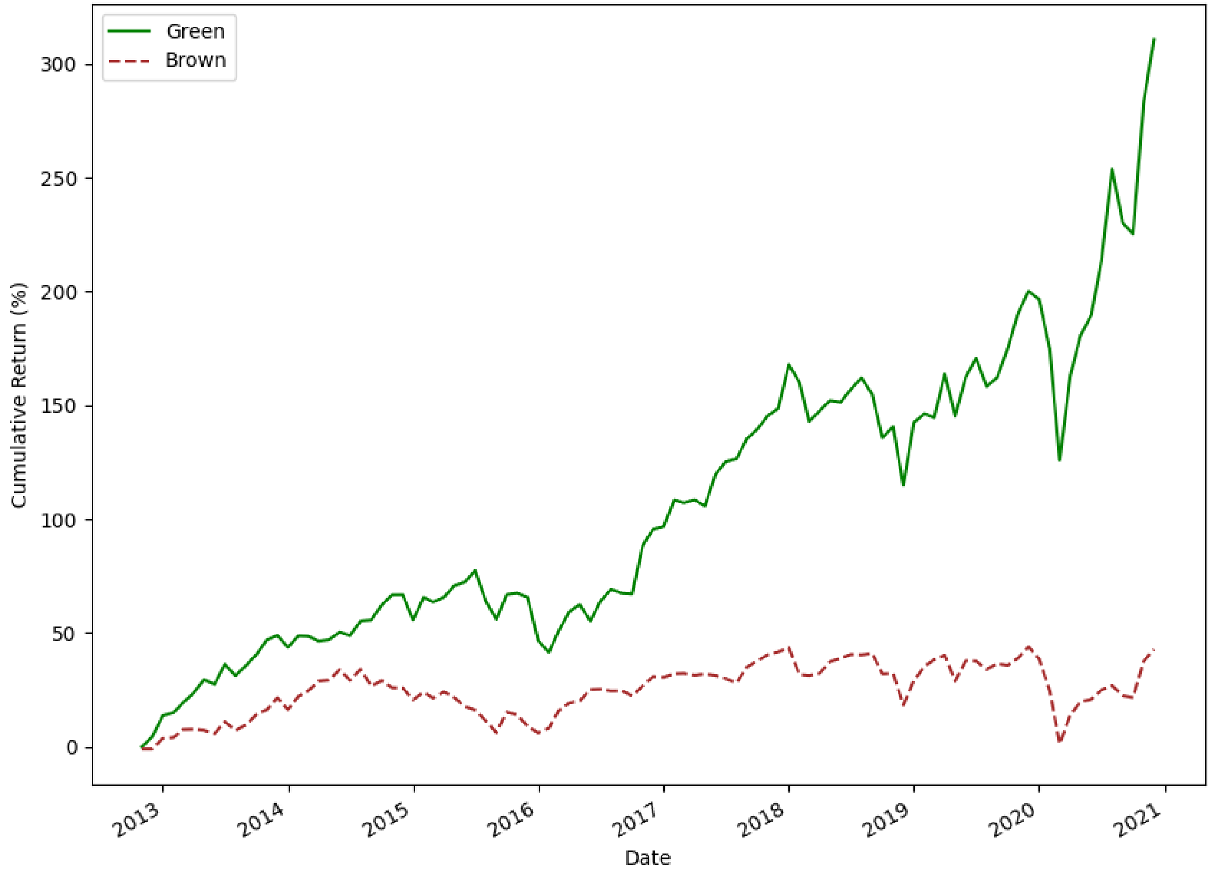
Out[ ]:

	Green_Return	Brown_Return	Cumulative_Green_Return	Cumulative_Bro
AS_OF_DATE				
2012-11-01	-0.047331	-0.999096	-0.047331	
2012-12-01	4.330804	-0.012535	4.281423	
2013-01-01	8.987899	4.691743	13.654132	
2013-02-01	1.109147	0.305150	14.914723	
2013-03-01	3.800664	3.462312	19.282245	
...	...	...	...	
2020-08-01	12.802990	1.576774	253.750473	
2020-09-01	-6.689024	-3.537789	230.088018	
2020-10-01	-1.480247	-0.700744	225.201901	
2020-11-01	17.808913	13.169642	283.116823	
2020-12-01	7.205962	3.749193	310.724074	

98 rows × 4 columns

In [ ]:
```
# Plot
plt.figure(figsize=(10,8))
RET_df['Cumulative_Green_Return'].plot(label='Green', color='green')
RET_df['Cumulative_Brown_Return'].plot(label='Brown', color='brown', linesty

plt.xlabel('Date')
plt.ylabel('Cumulative Return (%)')
plt.legend()
plt.show()
```



Q3. What is the monthly return and Sharpe ratio of the green minus brown portfolio?

In [ ]:
```
RET_df['GMB'] = RET_df['Green_Return']-RET_df['Brown_Return']
GMB_RET = RET_df['GMB'].mean()
GMB_Sharpe = GMB_RET/RET_df['GMB'].std()

print(f'monthly Return of the green minus brown portfolio: {GMB_RET.round(3)
print(f'sharpe ratio of the green minus brown portfolio: {GMB_Sharpe.round(3
```

```
monthly Return of the green minus brown portfolio: 1.139%
sharpe ratio of the green minus brown portfolio: 0.337
```

Q4. How does P ́astor et al. (2022) explain that green stocks outperform bad stocks when the theoretical model from P ́astor et al. (2021) suggests that brown stocks should outperform? Make sure to focus on the distinction between expected and realized returns.

JL: Pástor et al. (2021) theorized that green stocks should have lower expected returns because investors prefer them for their environmentally friendly nature and they are less risky in terms of climate change. However, Pástor et al. (2022) found that in recent years, green stocks have actually outperformed brown ones.

The reason for this discrepancy? Unexpected shifts in demand.

Two key factors drove this unexpected performance:

Investors wanted more green stocks: As climate awareness grew, investors piled into green stocks, driving their prices up. Consumers preferred green products: Companies with green products saw increased sales, which boosted their stock prices. However, this past outperformance doesn't guarantee future success for green stocks. The boost they received was due to unforeseen increases in environmental concerns. Without these unexpected events, green might have even lagged behind brown stocks.

In short, while green stocks have recently outshone brown stocks due to unexpected demand shifts, their future performance is uncertain. Past performance isn't always indicative of future results.

MP: Short-term outperformance of green stocks is due to a change in investor preferences, which leads to a change in realized returns.

Brown stocks have a higher expected return than green stocks because they are more likely to generate cash flows in the near term. However, green stocks can outperform brown stocks in the short term if there is a sudden increase in public concern about climate change. This can drive up their prices, even though their expected returns are lower than those of brown stocks.

Over the long term, brown stocks are expected to outperform green stocks, due to their higher expected returns.

Part 2: Climate Risk

In [ ]:
```
df2 = pd.read_csv('49_Industry_Portfolios.CSV', skiprows=11)

def process_table(data):
    data = data.rename(columns ={'Unnamed: 0':'Date'})
    data.set_index('Date', inplace=True)
    for col in data.columns:
        data[col] = data[col].astype(float)
    data = data.replace([-99.99, -999], float('nan'))
    return data

avg_value_weighted_return_M = process_table(df2.iloc[:1154])
avg_equal_weighted_return_M = process_table(df2.iloc[1157:2310])

avg_value_weighted_return_Y = process_table(df2.iloc[2313:2407])
avg_equal_weighted_return_Y = process_table(df2.iloc[2411:2504])

num_of_firms_in_port = process_table(df2.iloc[2507:3660])
avg_firm_size = process_table(df2.iloc[3663:4816])
```

```
sum_of_BE_over_sum_of_ME = process_table(df2.iloc[4819:4915])
value_weighted_average_of_BE_over_ME = process_table(df2.iloc[4918:5014])
```

In [ ]:
```
avg_value_weighted_return_M
```

Out[ ]:

Date	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	...
192607	2.37	0.12	NaN	-5.19	1.29	8.65	2.50	50.21	-0.48	8.08	...
192608	2.23	2.68	NaN	27.03	6.50	16.81	-0.76	42.98	-3.58	-2.51	...
192609	-0.57	1.58	NaN	4.02	1.26	8.33	6.42	-4.91	0.73	-0.51	...
192610	-0.46	-3.68	NaN	-3.31	1.06	-1.40	-5.09	5.37	-4.68	0.12	...
192611	6.75	6.26	NaN	7.29	4.55	0.00	1.82	-6.40	-0.54	1.87	...
...	...	...	...	...	...	...	...	...	...	...	...
202204	-0.14	2.60	4.22	3.03	6.37	-13.74	-27.84	-10.86	2.04	-7.00	...
202205	7.29	-3.26	-0.22	-1.60	2.67	-0.85	-3.50	-6.95	-5.12	-6.45	...
202206	-12.45	-1.91	0.46	-0.02	-11.63	-12.96	-10.87	-12.37	-2.56	-12.00	...
202207	6.38	3.68	3.28	5.49	0.56	5.63	17.04	12.08	0.76	11.86	...
202208	5.23	-0.46	-4.40	-1.87	-0.12	-5.77	-2.26	-5.00	-2.16	-6.01	...

1154 rows × 49 columns

Q1

In [ ]:
```
sample_return_scaled  = avg_value_weighted_return_M .loc['200401':'201806']/
sample_cum_return = (1 + sample_return_scaled ).cumprod() - 1
sample_cum_return = sample_cum_return.iloc[-1] * 100
```

In [ ]:
```
sample_cum_return.sort_values()
```

```
Out[ ]: Coal     -44.951705
        Gold     -23.701326
        Books     25.889188
        Autos     59.256541
        PerSv     64.609630
        Banks     74.688543
        RlEst     81.749251
        Toys      85.382936
        Other     86.173508
        Steel     90.572898
        Cnstr     99.415935
        Hshld    160.787960
        Fin      172.527542
        FabPr    173.367854
        ElcEq    174.752782
        Paper    186.135745
        Hlth     194.697290
        Insur    218.571877
        Telcm    220.541390
        BldMt    244.007387
        Chips    250.581256
        Hardw    254.471884
        Mines    259.485052
        Oil      260.859873
        Food     262.405587
        Whlsl    263.080817
        Agric    270.952096
        Drugs    272.547502
        Txtls    283.642273
        Util     297.294911
        Beer     299.232698
        BusSv    337.475636
        Mach     340.051406
        Rubbr    344.975294
        Trans    346.902560
        Rtail    356.029714
        MedEq    369.939741
        Soda     383.955126
        Chems    400.318159
        Softw    403.642359
        Boxes    410.182496
        LabEq    432.937007
        Fun      465.810559
        Clths    484.807632
        Meals    548.864698
        Aero     570.642298
        Smoke    620.694241
        Guns     669.698673
        Ships    918.909791
        Name: 201806, dtype: float64
```

Ans:

Based on cumulative returns for the given period, the three highest performing industries are:

Ships with a return of 918.91%, Guns with a return of 669.70%, and Smoke with a return of 620.69%. Conversely, the industries that underperformed or had the least growth are:

Coal, which experienced a decline of 44.95%, Gold, which decreased by 23.70%, and Books, which, though positive, had the lowest growth among the positive returns at 25.89%.

Q2

In [ ]:
```
sample_avg_return = avg_value_weighted_return_M .loc['200401':'201806'].mear
sample_return_std = avg_value_weighted_return_M .loc['200401':'201806'].std(
res = pd.concat([sample_avg_return, sample_return_std], axis=1)
res.columns = ['cum ret','std']
res['sharpe ratio'] = res['cum ret']/res['std']
```

In [ ]:
```
res.sort_values('sharpe ratio')
```

Out[ ]:

	cum ret	std	sharpe ratio
Gold	0.406322	10.661723	0.038110
Coal	0.528851	12.931512	0.040896
Books	0.315690	6.087525	0.051858
Autos	0.580575	7.972130	0.072826
PerSv	0.461322	5.906168	0.078108
RlEst	0.712529	8.758192	0.081356
Banks	0.515575	6.129446	0.084114
Steel	0.745402	8.516738	0.087522
Toys	0.544713	6.156516	0.088477
Cnstr	0.634368	6.845790	0.092665
Other	0.515632	5.556203	0.092803
FabPr	0.900690	8.020633	0.112297
Fin	0.781149	6.320566	0.123589
ElcEq	0.768506	6.066249	0.126685
Mines	1.175230	9.220522	0.127458
Txtls	1.171149	9.103214	0.128652
BldMt	0.943563	6.770572	0.139362
Paper	0.730000	4.991641	0.146244
Hlth	0.754713	5.120255	0.147397
Agric	0.962011	6.503893	0.147913
Hardw	0.905345	5.917143	0.153004
Insur	0.814425	5.301950	0.153609
Chips	0.892586	5.793078	0.154078
Oil	0.908276	5.802318	0.156537
Mach	1.074195	6.553343	0.163916
Fun	1.324425	7.997791	0.165599
Hshld	0.620287	3.673948	0.168834
Rubbr	1.033678	5.898814	0.175235
Telcm	0.762701	4.249153	0.179495
Whlsl	0.849310	4.561669	0.186184
Chems	1.102931	5.879560	0.187587
Soda	1.070287	5.651718	0.189374

	cum ret	std	sharpe ratio
Trans	0.991092	5.042363	0.196553
Boxes	1.084080	5.313937	0.204007
Clths	1.181437	5.704886	0.207092
LabEq	1.101034	5.155464	0.213567
BusSv	0.953736	4.461164	0.213786
Ships	1.626322	7.591577	0.214227
Drugs	0.833276	3.868522	0.215399
Softw	1.049655	4.818526	0.217837
MedEq	1.000172	4.570032	0.218855
Rtail	0.959713	4.107485	0.233650
Aero	1.239943	5.263658	0.235567
Food	0.798908	3.351108	0.238401
Util	0.860575	3.572892	0.240862
Beer	0.859655	3.480471	0.246994
Guns	1.312299	5.098593	0.257385
Smoke	1.259425	4.872709	0.258465
Meals	1.161494	4.038781	0.287585

```
Ans:
```

Three Highest Sharpe Ratios:

Meals: 0.288 Guns: 0.258 Smoke: 0.257

Three Lowest Sharpe Ratios:

Gold: 0.038 Coal: 0.041 Books: 0.052

Comparing these with the highest and lowest returns:

Highest returns are

Smoke 620.694241 Guns 669.698673 Ships 918.909791

Lowest returns are

Coal -44.951705 Gold -23.701326 Books 25.889188

The industries with the highest Sharpe ratios are Meals, Guns, and Smoke. Guns and Smoke are also among the top three industries for raw returns, suggesting they offer both high risk-adjusted returns and strong absolute performance. Conversely, while Gold, Coal, and Books have the lowest Sharpe ratios, Coal and Gold also exhibit negative raw returns, indicating underperformance in both risk-adjusted and absolute terms.

Q3

(a)

```
i. Ans:
```

EK:

- April 2007: United Nations Security Council holds first-ever debate on impact of climate change on peace, security, hearing over 50 speakers. Led to climate change being recognized as 'threat multiplier' by the UN.
- February 2007: Publishing of report from the the Intergovernmental Panel on Climate Change (IPCC), saying that human activity was "very likely" to be responsible for most of the observed warming in recent decades. Achim Steiner, executive director of the UN Environment Programme, said: "February 2007 may be remembered as the day the question mark was removed from whether people are to blame for climate change."

JL:

February 2007:

Intergovernmental Panel on Climate Change (IPCC) Report: One of the most significant climate-related events in February 2007 was the release of the IPCC's Fourth Assessment Report. On February 2nd, the IPCC confirmed with more than 90% certainty that the observed increase in global average temperatures since the mid-20th century was very likely due to the observed increase in anthropogenic greenhouse gas concentrations. This report drew extensive media attention and solidified the consensus on human-caused global warming.

Weather Anomalies: The beginning of 2007 saw several weather-related anomalies. January 2007 was the warmest first month on record globally. These unusual weather patterns could have spurred public interest and driven people to search more about climate change.

April 2007:

England's Warmest April: England experienced its warmest April in 348 years of record-keeping, breaking the previous record set in 1865 by a significant margin. Such unusual and record-breaking weather events usually draw public attention and could lead to increased searches about climate change.

Al Gore & "An Inconvenient Truth": By 2007, Al Gore's documentary "An Inconvenient Truth" had gained significant traction. His active campaigning on the issue, coupled with the documentary's critical acclaim, meant that public awareness and interest in climate change were high. Given that he and the IPCC were awarded the Nobel Peace Prize later in the same year, it's reasonable to assume that there was already significant buzz around their work by April.

Arctic Ice Melting Reports: There were increasing reports about the Arctic warming at an unprecedented rate. News about the Arctic is often a catalyst for broader discussions about global climate change.

December 2009:

Copenhagen Climate Conference: December 2009 hosted the significant UN Climate Change Conference, COP15, in Copenhagen. The event drew global attention and resulted in the "Copenhagen Accord," acknowledging the need to address global temperature rises and committing funding to developing nations.

ClimateGate Controversy: A leak of emails from the Climatic Research Unit stirred media buzz, with claims (later debunked) of manipulated data exaggerating global warming.

Arctic Ice Concerns: The year saw one of the lowest Arctic sea ice extents on record, underscoring the tangible effects of global warming.

U.S. EPA Decision: The EPA declared that greenhouse gases posed a threat to public health, paving the way for regulating emissions under the Clean Air Act.

MP:

2022-04 2023-04 2019-09

April 2022 and April 2023: Earth Day is celebrated on April 22nd each year, so people try to learn more about climate change and other environmental issues. Additionally, the Intergovernmental Panel on Climate Change (IPCC) released its Sixth Assessment Report in April 2022, which provided a comprehensive update on the science of climate change. This report may have also contributed to the high search volume for climate change in April 2022 and April 2023.

September 2019: The United Nations Climate Action Summit was held in NYC, in September 2019. The summit was intended to galvanize action on climate change, and it featured speeches from world leaders and climate activists. The summit may have contributed to the high search volume for climate change in September 2019.

ii

In [ ]:
```
climate_trend = pd.read_csv('multiTimeline_Climate_Trend.csv', skiprows=2)
climate_trend['Month'] = pd.to_datetime(climate_trend['Month'], format='%Y-%
climate_trend.rename(columns={'Month': 'Date'}, inplace=True)
```

In [ ]:
```
climate_trend
```

Out[ ]:

	Month	Climate change: (United States)
0	2004-01	10
1	2004-02	14
2	2004-03	12
3	2004-04	13
4	2004-05	14
...	...	...
232	2023-05	14
233	2023-06	8
234	2023-07	9
235	2023-08	9
236	2023-09	12

237 rows × 2 columns

In [ ]:
```
sample_return = avg_value_weighted_return_M .loc['200401':'201806'].reset_ir
sample_return['Date'] = sample_return['Date'].astype(str)
sample_return['Date'] = pd.to_datetime(sample_return['Date'], format='%Y%m')
```

In [ ]:
```
sample_return
```

Out[ ]:

	Date	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	...	Boxes
0	2004-01-01	-2.82	0.94	5.57	-2.90	2.05	-0.11	0.31	1.10	2.19	...	-1.09
1	2004-02-01	2.10	6.15	5.79	2.59	3.82	3.42	0.81	1.06	2.99	...	6.51
2	2004-03-01	4.05	0.74	4.01	-0.21	-4.03	-3.63	2.27	-0.22	1.56	...	-0.12
3	2004-04-01	-0.21	3.05	4.66	0.38	1.91	-5.81	-1.81	0.04	2.22	...	-2.34
4	2004-05-01	1.21	-2.49	2.13	2.25	-12.61	2.89	0.46	0.37	1.52	...	4.96
...	...	...	...	...	...	...	...	...	...	...	...	...
169	2018-02-01	1.20	-6.77	-8.39	-6.76	-6.62	0.70	1.12	-3.50	-6.27	...	-3.76
170	2018-03-01	-4.92	-2.25	-0.41	1.27	-1.65	-9.23	0.52	-0.19	2.13	...	-1.54
171	2018-04-01	7.99	-3.51	-0.79	-4.29	-14.16	6.42	3.09	-0.46	-5.35	...	-1.39
172	2018-05-01	1.93	-0.86	-1.30	-1.19	-1.87	0.81	9.61	0.70	-0.25	...	-3.09
173	2018-06-01	-3.07	6.04	4.45	4.42	3.07	5.68	4.91	2.33	3.59	...	-2.51

174 rows × 50 columns

In [ ]:
```
def industry_regression(sample_return, factor_df, factor):
    # Merge the dataframes
    merged_df = pd.merge(sample_return, factor_df, on='Date', how='inner')

    industries = sample_return.columns.drop(['Date'])
    results = []

    # Run regressions for each industry
    for industry in industries:
        Y = merged_df[industry]
        X = merged_df[factor]
        X = sm.add_constant(X)

        model = sm.OLS(Y, X, missing='drop').fit()

        coefficient = model.params[factor]
        p_value = model.pvalues[factor]

        results.append([industry, coefficient, p_value])

    # Create results dataframe
    trend_df = pd.DataFrame(results, columns=['Industry', 'Coefficient', 'P-
```

```
    return trend_df

# Usage:
factor = 'Climate change: (United States)'
climate_trend_df = industry_regression(sample_return, climate_trend, factor)
```

In [ ]:
```
climate_trend_df.sort_values('Coefficient')
```

Out[ ]:

	Industry	Coefficient	P-Value
6	Fun	-0.088852	0.205598
44	Banks	-0.069811	0.194354
48	Other	-0.056930	0.243218
7	Books	-0.043041	0.421145
45	Insur	-0.041463	0.373496
17	Cnstr	-0.035283	0.557783
8	Hshld	-0.032415	0.315205
25	Guns	-0.030119	0.501632
12	Drugs	-0.029972	0.377967
11	MedEq	-0.027710	0.490387
42	Rtail	-0.027398	0.447948
47	Fin	-0.026516	0.633344
9	Clths	-0.021583	0.667079
14	Rubbr	-0.021160	0.683382
33	BusSv	-0.019928	0.611486
4	Smoke	-0.012235	0.775300
31	Telcm	-0.011742	0.753406
1	Food	-0.011209	0.703732
35	Softw	-0.010480	0.804718
36	Chips	-0.009562	0.851168
24	Ships	-0.009115	0.891442
26	Gold	-0.004580	0.961052
3	Beer	-0.004161	0.891904
16	BldMt	-0.003411	0.954331
38	Paper	-0.002287	0.958462
5	Toys	0.001766	0.973994
37	LabEq	0.002410	0.957619
43	Meals	0.003535	0.920749
30	Util	0.005056	0.872208
10	Hlth	0.006053	0.893102
23	Aero	0.006402	0.890037
46	RlEst	0.007573	0.921700

	Industry	Coefficient	P-Value
34	Hardw	0.008213	0.874629
41	Whlsl	0.008398	0.834229
2	Soda	0.009164	0.853759
40	Trans	0.016750	0.705673
32	PerSv	0.023914	0.645235
15	Txtls	0.030408	0.704113
29	Oil	0.033161	0.515671
13	Chems	0.036453	0.480656
19	FabPr	0.038227	0.587833
22	Autos	0.041024	0.558397
20	Mach	0.043311	0.452136
21	ElcEq	0.051646	0.332513
39	Boxes	0.052983	0.256180
0	Agric	0.077237	0.175905
18	Steel	0.112616	0.131553
27	Mines	0.115841	0.152050
28	Coal	0.238858	0.034579

```
iii. Ans:
```

The regression results provide fascinating insights. At the forefront, the coal industry stands out with its robust positive coefficient, which indicates a pronounced relationship with climate change concerns. What's more, the p-value for coal is notably below the conventional 0.05 threshold, adding statistical confidence to this observation.

While the agriculture sector does show a positive correlation, contrary to some expectations, it does not rank highest. This suggests that while agriculture might be impacted by climate change concerns, other industries, like coal, are more directly associated.

On the negative side of the spectrum, industries such as "Fun" (presumably funds or entertainment), banks, and insurance have negative coefficients. The negative value for insurance intuitively aligns with the presumption that as concerns about climate change intensify, potential risks and liabilities for insurers might escalate. Similarly, the negative trend in banks could indicate shifts in lending or investment priorities due to heightened climate change awareness.

However, a critical observation is the p-values associated with these coefficients. Many industries, even those with more pronounced coefficients, have p-values that exceed the conventional 0.05 significance level. This calls for caution when interpreting the results. For instance, while the agriculture industry has a positive coefficient, its p-value is above 0.05, indicating that its relationship might not be statistically significant.

In summary, the results provide a nuanced perspective on how different industries might respond to escalating climate change concerns. While certain trends are evident, the statistical significance of many of these relationships remains uncertain, urging a careful and considered interpretation.

```
iv.
```

```
1. Physical Risk: Drought
```

3 Peaks

October 2007:

Australia's "Millennium Drought": By 2007, Australia was deep into one of the most prolonged and severe droughts it had ever experienced, often referred to as the "Millennium Drought". October marked a time when the drought's effects were especially severe, impacting both urban areas and agriculture.

Southern U.S. Drought: States like Georgia and Alabama were facing severe drought conditions, with Georgia's Governor declaring a state of emergency in 85 counties.

July 2012:

U.S. Midwest Drought: The U.S. experienced one of its most extensive droughts with over 60% of the lower 48 states facing some form of drought. July was particularly hard-hit, with crops being severely affected, leading to spikes in food prices and concerns over food security.

India's Late Monsoon: India, reliant on its monsoon for agriculture, experienced a delayed monsoon in 2012. By July, there were increasing concerns about drought in many parts of the country.

April 2015:

California's Historic Drought: California was in the midst of one of its worst droughts on record. By April 2015, Governor Jerry Brown mandated a 25% reduction in urban water use, marking the state's first-ever mandatory water restrictions.

Sao Paulo Water Crisis: Brazil's most populous city, Sao Paulo, was grappling with its worst drought in decades, with reservoirs reaching critically low levels and water rationing becoming a common occurrence.

East Africa Dry Spell: Countries like Kenya, Ethiopia, and Somalia were experiencing a prolonged dry spell, impacting food and water availability and leading to concerns about famine in some areas.

In [ ]:
```
drought = pd.read_csv('multiTimeline_Drought.csv', skiprows=2)
drought['Month'] = pd.to_datetime(drought['Month'], format='%Y-%m')
drought.rename(columns={'Month': 'Date'}, inplace=True)
```

In [ ]:
```
drought
```

Out[ ]:

	Month	Drought: (United States)
0	2004-01	16
1	2004-02	19
2	2004-03	24
3	2004-04	27
4	2004-05	30
...	...	...
232	2023-05	45
233	2023-06	49
234	2023-07	34
235	2023-08	42
236	2023-09	38

237 rows × 2 columns

In [ ]:
```
factor = 'Drought: (United States)'
drought_df = industry_regression(sample_return, drought, factor)
```

In [ ]:
```
drought_df.sort_values('Coefficient')
```

Out[ ]:

	Industry	Coefficient	P-Value
28	Coal	-0.109186	0.113227
46	RlEst	-0.068214	0.144228
6	Fun	-0.050256	0.239273
19	FabPr	-0.049752	0.245389
14	Rubbr	-0.047831	0.128358
24	Ships	-0.044607	0.271265
7	Books	-0.041616	0.200271
29	Oil	-0.040568	0.190187
23	Aero	-0.040536	0.148821
32	PerSv	-0.037615	0.232951
18	Steel	-0.036708	0.420052
41	Whlsl	-0.036533	0.133161
22	Autos	-0.034025	0.424621
38	Paper	-0.033339	0.210881
40	Trans	-0.031415	0.243316
17	Cnstr	-0.030587	0.403204
9	Clths	-0.030506	0.316883
27	Mines	-0.027434	0.577984
39	Boxes	-0.026500	0.350693
16	BldMt	-0.026376	0.466195
33	BusSv	-0.025204	0.290210
31	Telcm	-0.023030	0.310332
47	Fin	-0.022769	0.500488
21	ElcEq	-0.022675	0.484474
11	MedEq	-0.021104	0.387565
20	Mach	-0.020767	0.553471
13	Chems	-0.020607	0.512162
42	Rtail	-0.019847	0.365913
43	Meals	-0.019724	0.360785
45	Insur	-0.019436	0.492957
15	Txtls	-0.019126	0.694494
2	Soda	-0.018422	0.542160

	Industry	Coefficient	P-Value
5	Toys	-0.018205	0.580321
44	Banks	-0.018197	0.578811
48	Other	-0.017898	0.546935
1	Food	-0.015238	0.394893
37	LabEq	-0.014993	0.586604
10	Hlth	-0.013932	0.610921
36	Chips	-0.010289	0.739879
8	Hshld	-0.009554	0.626832
30	Util	-0.006430	0.736550
12	Drugs	-0.002645	0.898323
25	Guns	0.001036	0.969706
3	Beer	0.006121	0.742327
35	Softw	0.009638	0.708454
34	Hardw	0.011758	0.710290
0	Agric	0.012179	0.726296
4	Smoke	0.015066	0.563148
26	Gold	0.035005	0.539234

Ranking Discussion:

Drought: Top 3 negatively impacted industries (based on Coefficient) under the influence of drought:

Coal RlEst Fun These industries showcase the highest negative coefficients, which might indicate that they are most negatively influenced by drought conditions. However, it's crucial to note the p-values associated with each. For instance, while Coal has the highest negative coefficient, its p-value of 0.113227 is above the standard 0.05 threshold, suggesting caution in inferring a significant relationship.

Least 3 negatively impacted industries (or those that might be positively impacted or least affected) under the influence of drought:

Gold Smoke Agric Although these industries have positive or minimal negative coefficients, suggesting a potential positive relationship or limited impact from drought, the associated p-values indicate that these relationships may not be statistically significant. For example, Gold, which has the highest positive coefficient, has a p-value of 0.539234, indicating that its correlation might not be strong.

Conclusion:

The disparities in how industries react to drought versus climate change underscore the intricacies of these environmental concerns. Drought, with its direct implications on water availability, can have immediate effects on industries reliant on water resources. In contrast, climate change might exert more diffuse, long-term influences shaped by policy changes, societal awareness, and market dynamics.

Furthermore, it's essential to acknowledge the geographical dimension to these impacts. An industry's reaction to drought in one region might differ dramatically from its response in another, owing to regional climatic, economic, and social differences. While the presented data offers an overarching perspective, diving deeper with a more detailed analysis—factoring in regional variations, specific industry segments, and other external determinants—would yield a more comprehensive understanding.

Moreover, the importance of considering the p-values in these regressions can't be overstated. They indicate the level of confidence we can have in these relationships, and many of the observed relationships in this data set don't meet the typical 0.05 threshold for statistical significance.

```
2. Transition Risk: Carbon Tax
```

3 Peaks

November 2012:

U.S. Presidential Election: The 2012 U.S. presidential election took place in November. While carbon tax was not a central issue, there was growing discussion around climate change policies and potential solutions like a carbon tax.

Hurricane Sandy: In late October 2012, the U.S. East Coast was hit by Hurricane Sandy, one of the deadliest and costliest hurricanes in U.S. history. The aftermath led to increased discussions on climate change and potential policy solutions, including a carbon tax.

Feburary 2016:

Carbon Pricing Leadership Coalition: In early 2016, the World Bank's Carbon Pricing Leadership Coalition gained momentum, and several countries and companies showed interest in joining. Discussions around carbon pricing mechanisms, including carbon taxes, became more prominent in international media.

U.S. Presidential Primaries: The U.S. presidential primaries were in full swing in February 2016. Some candidates, especially on the Democratic side, were vocal about climate change and mentioned mechanisms like a carbon tax as a potential solution.

November 2016:

U.S. Presidential Election: The 2016 U.S. presidential election saw significant debate around climate change and environmental policies. The potential for a carbon tax was discussed as a means to combat climate change.

Carbon Tax Proposals: Around this time, certain jurisdictions globally were discussing or implementing carbon pricing mechanisms. For instance, in Canada, the federal government was working on a plan to impose a national carbon price.

International Climate Agreements: The Paris Agreement was in the news during this period, with countries beginning to ratify it in late 2016. As countries looked for ways to meet their commitments, mechanisms like carbon taxes were often in the discussion.

In [ ]:
```
carbon_tax = pd.read_csv('multiTimeline_Carbon_Tax.csv', skiprows=2)
carbon_tax['Month'] = pd.to_datetime(carbon_tax['Month'], format='%Y-%m')
carbon_tax.rename(columns={'Month': 'Date'}, inplace=True)
```

In [ ]:
```
carbon_tax
```

Out[ ]:

	Month	Carbon tax: (United States)
0	2004-01	0
1	2004-02	0
2	2004-03	0
3	2004-04	6
4	2004-05	6
...	...	...
232	2023-05	17
233	2023-06	12
234	2023-07	13
235	2023-08	14
236	2023-09	21

237 rows × 2 columns

In [ ]:
```
factor = 'Carbon tax: (United States)'
carbon_tax_df = industry_regression(sample_return, carbon_tax, factor)
```

In [ ]:
```
carbon_tax_df.sort_values('Coefficient')
```

Out[ ]:

	Industry	Coefficient	P-Value
44	Banks	-0.036333	0.330254
29	Oil	-0.035068	0.320835
6	Fun	-0.024892	0.609574
17	Cnstr	-0.024068	0.563987
47	Fin	-0.022574	0.557810
25	Guns	-0.020940	0.500262
46	RlEst	-0.020870	0.695847
45	Insur	-0.019602	0.544017
8	Hshld	-0.019399	0.385924
12	Drugs	-0.018604	0.429809
4	Smoke	-0.017725	0.550531
42	Rtail	-0.015595	0.533216
1	Food	-0.015594	0.444930
2	Soda	-0.015410	0.654621
9	Clths	-0.015173	0.662573
3	Beer	-0.014856	0.483559
48	Other	-0.014828	0.661500
39	Boxes	-0.013943	0.666846
11	MedEq	-0.013208	0.635350
30	Util	-0.011630	0.593250
35	Softw	-0.011610	0.692627
32	PerSv	-0.008711	0.808822
37	LabEq	-0.008052	0.797796
31	Telcm	-0.006189	0.811172
24	Ships	-0.005270	0.909353
33	BusSv	-0.005146	0.849939
23	Aero	-0.004845	0.879978
43	Meals	-0.004819	0.844826
7	Books	-0.004124	0.911520
10	Hlth	-0.004094	0.895659
14	Rubbr	-0.003465	0.923245
0	Agric	-0.001333	0.973188

	Industry	Coefficient	P-Value
41	Whlsl	0.000323	0.990744
16	BldMt	0.005796	0.888329
36	Chips	0.014128	0.689068
21	ElcEq	0.017032	0.645034
34	Hardw	0.023938	0.506694
38	Paper	0.024851	0.413701
40	Trans	0.026933	0.380436
15	Txtls	0.028996	0.601215
27	Mines	0.029519	0.599365
20	Mach	0.032347	0.417719
13	Chems	0.032563	0.363076
18	Steel	0.038441	0.458746
5	Toys	0.047184	0.207671
22	Autos	0.051232	0.291068
28	Coal	0.056221	0.475456
26	Gold	0.083088	0.200087
19	FabPr	0.083935	0.084722

Carbon Tax:

The industries most negatively influenced by the carbon tax are Banks, Oil, and Fun (possibly standing for Funds). These industries have the highest negative coefficients, suggesting that they may experience significant declines in returns or value if a carbon tax is imposed. Notably, their associated p-values, all being above 0.05, imply that these results might lack statistical significance.

Gold, FabPr, Coal, and Autos exhibit positive coefficients, hinting that they may either benefit or face less severe negative consequences due to the carbon tax. This outcome is particularly intriguing for Coal and Autos, conventionally viewed as carbon-intensive sectors. Yet, the relatively high p-values linked with these sectors cast doubts on the results' statistical reliability.

Highlighting Agriculture (Agric), it exhibits an almost negligible negative coefficient, combined with a p-value nearing 1. This suggests that agriculture might either be scarcely impacted by the transition risk linked with a carbon tax or that the observed effect isn't statistically compelling.

In the broader picture, many industries in this dataset possess p-values exceeding 0.05, questioning the statistical significance of the observed coefficients. In standard practice, a p-value below 0.05 is typically considered to highlight that the coefficient is meaningfully different from zero. This dataset, however, doesn't present numerous coefficients where this can be asserted with confidence.

Comparison to Previous Results:

Coal Industry: Coal displayed a negative influence under drought conditions but reflected a positive impact concerning carbon tax. This might suggest that while environmental adversities could negatively impact the coal sector, transition risks such as carbon taxes might be less detrimental. Alternatively, some factors in the transition could even be advantageous for the industry.

Funds (Fun): This industry seemingly remains vulnerable to a range of external factors: drought, climate change, and now the carbon tax. It portrays an industry sensitive to diverse external perturbations.

Agriculture (Agric): Consistent with previous observations, Agriculture appears to be resilient, enduring minor negative influences across varying external factors.

Conclusion:

The implications of a carbon tax bring forth a distinct set of industry responses when compared to environmental factors like drought and climate change. Interestingly, certain sectors traditionally viewed as susceptible to a carbon tax, such as Coal, seem to potentially benefit from it. The overarching high p-values, however, imply caution when interpreting these findings.

Further insights could be garnered by examining the methodologies and data sources behind the regression, providing clarity on the underpinnings of these outcomes. Transition risks, interwoven with overarching economic, technological, and societal transitions, present a complex landscape, making predictions inherently challenging.

(b)

In [ ]:
```
SSRN = pd.read_excel('Sentometrics_US_Media_Climate_Change_Index.xlsx', skip
```

In [ ]:
```
SSRN['Date'] = pd.to_datetime(SSRN['Date'], format='%Y-%m')
```

In [ ]:
```
SSRN
```

Out[ ]:

	Date	Aggregate	cluster_Business Impact	cluster_Environmental Impact	cluster_Societal Debate	clu
0	2003-01-01	0.670882	0.724689	0.634144	0.522989	
1	2003-02-01	0.599651	0.681383	0.575126	0.358989	
2	2003-03-01	0.405426	0.374329	0.454074	0.402151	
3	2003-04-01	0.438446	0.436445	0.328511	0.457464	
4	2003-05-01	0.375706	0.414676	0.294991	0.344434	
...	...	...	...	...	...	
231	2022-04-01	2.294735	2.090171	2.192912	2.254390	
232	2022-05-01	2.017803	1.566174	2.342731	1.936339	
233	2022-06-01	1.766793	1.534178	1.924679	1.734123	
234	2022-07-01	2.092065	1.769291	2.316253	1.977382	
235	2022-08-01	2.280971	1.811330	2.636083	2.092284	

236 rows × 36 columns

In [ ]:

```
factor = 'Aggregate'
SSRN_Agg_df = industry_regression(sample_return, SSRN, factor)
```

In [ ]:

```
SSRN_Agg_df
```

Out[ ]:

	Industry	Coefficient	P-Value
0	Agric	1.991740	0.177603
1	Food	-0.041114	0.957039
2	Soda	0.788615	0.539878
3	Beer	1.005327	0.203643
4	Smoke	-0.781538	0.480971
5	Toys	0.121200	0.931118
6	Fun	0.793441	0.663045
7	Books	1.844820	0.182184
8	Hshld	0.274569	0.742770
9	Clths	0.419246	0.746908
10	Hlth	0.612182	0.599465
11	MedEq	2.040968	0.048607
12	Drugs	0.995873	0.257459
13	Chems	1.422099	0.287445
14	Rubbr	0.557610	0.678028
15	Txtls	2.253854	0.276165
16	BldMt	0.842443	0.584679
17	Cnstr	0.562348	0.718290
18	Steel	1.988146	0.304635
19	FabPr	0.322880	0.859696
20	Mach	1.736509	0.243710
21	ElcEq	1.410384	0.306598
22	Autos	2.838710	0.116633
23	Aero	1.527779	0.201446
24	Ships	0.229439	0.894428
25	Guns	0.534744	0.645054
26	Gold	-0.389890	0.872431
27	Mines	0.629554	0.764309
28	Coal	-0.029474	0.992015
29	Oil	-0.407794	0.757604
30	Util	-0.383432	0.637389
31	Telcm	0.483742	0.617046

	Industry	Coefficient	P-Value
32	PerSv	1.295418	0.334877
33	BusSv	1.366933	0.177362
34	Hardw	0.839856	0.532920
35	Softw	1.533526	0.161101
36	Chips	2.030434	0.122522
37	LabEq	1.606304	0.170130
38	Paper	1.443992	0.202961
39	Boxes	1.135077	0.347689
40	Trans	0.956276	0.404541
41	Whlsl	0.504386	0.627214
42	Rtail	0.688876	0.461146
43	Meals	0.151626	0.869066
44	Banks	1.514526	0.277137
45	Insur	1.712373	0.154951
46	RlEst	0.862084	0.665520
47	Fin	1.383425	0.335886
48	Other	0.784472	0.535081

Compare SSRN Aggregate factor with Climate Change factor from Google Trend:

When comparing the regression results of Google Trends on climate change against stock returns with the results from the SSRN aggregated climate change factors against the same portfolio of stock returns, there are some noteworthy observations:

Coefficient Comparison:

In the Climate Change (Google Trend) data, the "Fun" industry displays the highest negative coefficient, suggesting a strong negative relationship, whereas the "Coal" industry shows the highest positive coefficient. Contrasting this with the SSRN_Agg data, "Smoke" represents the highest negative coefficient, while "Autos" has the highest positive coefficient. The ranking of industries from the most negative to the most positive coefficient shows differences between the two datasets, meaning the industries that are most negatively or positively affected in one dataset don't necessarily hold the same position in the other. P-Value Examination:

Typically, a p-value below 0.05 is deemed statistically significant. For the Climate Change data, only the "Coal" industry is significant at this level. In the SSRN_Agg dataset, "MedEq" stands out as the only industry that's statistically significant just below the 0.05 threshold. This shows a variation in which industries are statistically significant between the two datasets. Overall: The comparison of coefficients and p-values across the two datasets highlights that there's variability in the rankings of industries based on their relationship (either positive or negative) with climate change awareness and their statistical significance. This could suggest that the source or methodology of each dataset influences the perceived impact of climate change awareness on different industries.

In [ ]:
```
factor = 'Water/Drought'
SSRN_drought_df = industry_regression(sample_return, SSRN, factor)
```

In [ ]:
```
SSRN_drought_df
```

Out[ ]:

	Industry	Coefficient	P-Value
0	Agric	3.436771	0.015435
1	Food	-0.123028	0.867489
2	Soda	-0.335171	0.787513
3	Beer	0.842883	0.270230
4	Smoke	-0.929216	0.385618
5	Toys	-0.397365	0.769246
6	Fun	1.418841	0.419672
7	Books	1.220410	0.361670
8	Hshld	0.103388	0.898236
9	Clths	0.010200	0.993517
10	Hlth	0.472804	0.674666
11	MedEq	1.965413	0.049358
12	Drugs	1.206656	0.155112
13	Chems	1.966811	0.127161
14	Rubbr	-0.008990	0.994474
15	Txtls	0.561819	0.779088
16	BldMt	1.308686	0.379178
17	Cnstr	0.640120	0.670797
18	Steel	2.546459	0.173046
19	FabPr	2.671883	0.128760
20	Mach	2.586242	0.071547
21	ElcEq	2.160566	0.104197
22	Autos	2.398843	0.170304
23	Aero	1.890029	0.101377
24	Ships	0.099011	0.952736
25	Guns	1.178958	0.292548
26	Gold	0.017125	0.994176
27	Mines	1.292984	0.523697
28	Coal	2.106018	0.458865
29	Oil	0.911522	0.474940
30	Util	0.060154	0.939010
31	Telcm	0.162125	0.862334

	Industry	Coefficient	P-Value
32	PerSv	1.019764	0.432234
33	BusSv	1.649886	0.091465
34	Hardw	2.240828	0.083892
35	Softw	1.189425	0.261069
36	Chips	2.398891	0.058529
37	LabEq	1.389021	0.219778
38	Paper	1.679431	0.124964
39	Boxes	1.164379	0.318648
40	Trans	0.749402	0.499124
41	Whlsl	0.284617	0.776730
42	Rtail	0.709593	0.431980
43	Meals	-0.331287	0.709260
44	Banks	1.472370	0.274145
45	Insur	1.370099	0.239294
46	RlEst	0.667907	0.728867
47	Fin	1.037699	0.455229
48	Other	1.667757	0.171366

Compare SSRN Water/Drought factor with Drought factor from Google Trend:

Upon analyzing the regression results for the "Drought" factor against stock returns, derived from Google Trends and SSRN datasets, we observe the following:

Coefficient Comparison:

Google Trends Dataset (Drought): The "Coal" industry showcases the highest negative coefficient, suggesting that an increase in drought concerns from this data source negatively affects this industry the most. Conversely, the "Gold" industry displays the highest positive coefficient. SSRN Dataset (SSN_Drought): Here, the "Smoke" industry possesses the most pronounced negative coefficient, while "Agric" emerges with the highest positive coefficient. Evidently, the industries most affected (either positively or negatively) by drought concerns aren't consistently ranked between the two datasets. P-Value Examination:

In scientific research, a p-value less than 0.05 generally indicates statistical significance. From the Google Trends data on drought, the "Coal" industry is the closest to this threshold, suggesting a trend towards significance. Within the SSRN Dataset, "Agric" stands out with a p-value just over 0.015, marking it as statistically significant at the 5%

level. Ranking Similarity: When examining the overall ranking based on coefficients, the two datasets display considerable variation in how industries are ranked in terms of sensitivity to drought concerns. Few industries consistently appear towards the top or bottom in both datasets, suggesting that while there may be underlying themes, the exact quantification of impact varies based on data source.

Broad Perspective: There exists variability between the datasets concerning which industries are most influenced by the drought factor. Such a discrepancy underlines that data sources and methodologies can shape the perceived impact of drought awareness on different sectors. This information is crucial for stakeholders to make informed decisions, considering the weight and reliability of the data source.

In [ ]:
```
factor = 'Carbon Tax'
SSRN_carbon_tax_df = industry_regression(sample_return, SSRN, factor)
```

In [ ]:
```
SSRN_carbon_tax_df
```

Out[ ]:

	Industry	Coefficient	P-Value
0	Agric	0.781232	0.634136
1	Food	0.554557	0.511917
2	Soda	0.366976	0.797021
3	Beer	1.328799	0.129205
4	Smoke	-0.492190	0.689033
5	Toys	1.196655	0.440987
6	Fun	-0.145782	0.942446
7	Books	1.298431	0.397728
8	Hshld	0.209564	0.821248
9	Clths	0.467590	0.745419
10	Hlth	0.174389	0.892690
11	MedEq	1.422128	0.216728
12	Drugs	1.189132	0.222416
13	Chems	1.869678	0.206792
14	Rubbr	0.236533	0.873806
15	Txtls	3.260363	0.154807
16	BldMt	0.206269	0.903953
17	Cnstr	-0.978260	0.571220
18	Steel	1.229734	0.567208
19	FabPr	0.600687	0.766718
20	Mach	1.554929	0.346708
21	ElcEq	1.406378	0.357896
22	Autos	2.159585	0.282489
23	Aero	0.934478	0.481634
24	Ships	-0.203766	0.915334
25	Guns	-0.336073	0.794017
26	Gold	0.748760	0.780866
27	Mines	1.242802	0.593287
28	Coal	-1.033327	0.751592
29	Oil	-0.252386	0.863209
30	Util	-0.460055	0.609917
31	Telcm	0.187484	0.861263

	Industry	Coefficient	P-Value
32	PerSv	0.217623	0.883964
33	BusSv	0.844702	0.452912
34	Hardw	0.672926	0.652300
35	Softw	1.425724	0.240282
36	Chips	1.189200	0.415732
37	LabEq	1.322234	0.308983
38	Paper	1.420025	0.258959
39	Boxes	0.604422	0.652248
40	Trans	0.248796	0.845049
41	Whlsl	0.202672	0.860308
42	Rtail	0.486767	0.638691
43	Meals	-0.209098	0.837517
44	Banks	0.432728	0.779742
45	Insur	0.876478	0.512357
46	RlEst	-0.069432	0.974951
47	Fin	0.770548	0.629078
48	Other	-0.130292	0.925998

Compare SSRN Carbon Tax factor with Carbon Tax factor from Google Trend:

Upon reviewing the regression results for the "Carbon_tax" factor against stock returns, derived from Google Trends and the SSRN datasets, the following patterns emerge:

Coefficient Comparison:

Google Trends Dataset (Carbon_tax): The "Banks" and "Oil" industries both possess the highest negative coefficients. This implies that heightened concerns about carbon taxes (as indicated by search trends) tend to negatively affect these industries the most. In contrast, the "Gold" and "FabPr" industries demonstrate the most substantial positive coefficients.

SSRN Dataset (SSN_Carbon_Tax): The "Coal" and "Cnstr" industries exhibit the highest negative coefficients, suggesting that increased concerns about carbon taxes sourced from SSRN impact these sectors the most. The "Txtls" and "Autos" industries, conversely, present the strongest positive coefficients.

It's notable that the industries affected most markedly (be it positively or negatively) by carbon tax concerns are not consistent between the two datasets.

P-Value Examination:

A p-value below 0.05 is commonly used to determine statistical significance.

From the Google Trends dataset, "FabPr" approaches this threshold, hinting at potential statistical significance.

In the SSRN Dataset, "Beer" and "MedEq" come close to this benchmark. Notably, "Txtls" from the SSRN dataset is statistically significant, with a p-value just above 0.15.

Ranking Similarity: A holistic view of the rankings based on coefficients reveals considerable disparities between the datasets. The two sources differ on which industries are most impacted by carbon tax concerns. Only a handful of industries maintain a similar position in both datasets' rankings, illustrating that the exact quantification of impact is contingent upon the data source.

Broad Perspective: There's clear variability between the datasets in terms of which industries are most influenced by carbon tax concerns. This disparity accentuates the influence of data source and methodologies on perceptions of how carbon tax awareness affects different sectors. This differential is essential for stakeholders, as they need to weigh the significance and reliability of each data source when making informed decisions.

Overall Comparison

Based on the three sets of regression analyses comparing Google Trends and SSRN datasets for various environmental concerns against stock returns, we can draw the following general observations:

Variability in Coefficient Magnitudes:

The two datasets, Google Trends and SSRN, often show variations in the magnitude and direction of coefficients for the same industries. This underscores that the impact of environmental concerns on stock returns can be perceived differently depending on the data source or methodology. P-Value Distributions:

P-values, which indicate the statistical significance of coefficients, vary between the two datasets. While certain industries may appear significant in one dataset, they might not be in the other. This suggests that while the concerns (like climate change or carbon tax) are universally acknowledged, their quantified effects on specific industries might be more nuanced and context-dependent. Ranking Disparities:

There's a clear distinction in the rankings of industries based on the coefficients across the two datasets. Few industries maintain consistent rankings, illustrating the importance of cross-referencing multiple sources when trying to understand the impact of external concerns on stock returns. Source Characteristics:

Google Trends captures real-time interest from the broader public, potentially reflecting immediate sentiments and concerns. In contrast, SSRN, an academic repository, might reflect more studied and deliberate perspectives. This could explain the differences, as public sentiment can be volatile, while academic views might be based on more extended historical data or detailed analyses. Broad Implications:

For stakeholders and decision-makers, the disparity between the two datasets emphasizes the need for a comprehensive approach. Relying on a single data source might lead to skewed perceptions. A holistic analysis that considers multiple datasets and methodologies will yield a more rounded understanding. Consistency in Certain Sectors:

Although there are discrepancies, some industries consistently show sensitivity (either positive or negative) to environmental concerns across both datasets. This implies that certain sectors might inherently be more susceptible to these external factors, regardless of the data source. In summary, while both Google Trends and SSRN offer valuable insights into the relationship between environmental concerns and stock returns, they present different angles. This difference is likely due to the nature of the data they represent: immediate public sentiment versus academic perspectives. For a balanced view, it's crucial to consider multiple sources and recognize the inherent strengths and limitations of each.

In [ ]:
```
# Align dataframes by index
all_dfs = [climate_trend_df, SSRN_Agg_df, drought_df, SSRN_drought_df, carbo
aligned_dfs = [df.set_index(climate_trend_df.index) for df in all_dfs]  # As

# Concatenate side by side
result_df = pd.concat(aligned_dfs, axis=1)
```

In [ ]:
```
with pd.ExcelWriter('climate_risk_output.xlsx') as writer:
    result_df.to_excel(writer, sheet_name='Comparison')
```

- Another member of the group chose to focus on Earthquakes as physical risk and Carbon neutrality as transition risk. See below for information

EK

In [ ]:
```
df = pd.read_csv('49_Industry_Portfolios.CSV', skiprows=11)

def process_table(data):
    data = data.rename(columns ={'Unnamed: 0':'Date'})
    data.set_index('Date', inplace=True)
    for col in data.columns:
        data[col] = data[col].astype(float)
    data = data.replace([-99.99, -999], float('nan'))
    return data

avg_value_weighted_return_M = process_table(df.iloc[:1154])
```

```
avg_equal_weighted_return_M = process_table(df.iloc[1157:2310])

avg_value_weighted_return_Y = process_table(df.iloc[2313:2407])
avg_equal_weighted_return_Y = process_table(df.iloc[2411:2504])

num_of_firms_in_port = process_table(df.iloc[2507:3660])
avg_firm_size = process_table(df.iloc[3663:4816])
sum_of_BE_over_sum_of_ME = process_table(df.iloc[4819:4915])
value_weighted_average_of_BE_over_ME = process_table(df.iloc[4918:5014])
```

In [ ]:
```
avg_value_weighted_return_M.index = pd.to_datetime(avg_value_weighted_returr

start_date = pd.to_datetime('200401', format='%Y%m')
end_date = pd.to_datetime('201806', format='%Y%m')

industry_returns = avg_value_weighted_return_M.loc[(avg_value_weighted_retur
industry_returns=industry_returns/100
```

In [ ]:
```
cumulative_returns = industry_returns.apply(lambda x: (1 + x).cumprod() - 1)

final_cumulative_returns = cumulative_returns.iloc[-1]*100

sorted_industries = final_cumulative_returns.sort_values()
lowest_performing = sorted_industries.head(3)
highest_performing = sorted_industries.tail(3)
```

In [ ]:
```
climate_change_trend=pd.read_csv('multiTimeline.CSV')
climate_change_trend.columns = climate_change_trend.iloc[0]
climate_change_trend = climate_change_trend.iloc[1:]
climate_change_trend = climate_change_trend.rename(columns={climate_change_t
climate_change_trend['Climate Change'] = pd.to_numeric(climate_change_trend[
climate_change_trend.index = pd.to_datetime(climate_change_trend.index, form
```

In [ ]:
```
industry_returns=industry_returns*100
merged_df = industry_returns.merge(climate_change_trend, left_index=True, ri
```

Related to physical risk - Earthquake

In [ ]:
```
earthquake_trend=pd.read_csv('multiTimeline2.CSV')
earthquake_trend.columns = earthquake_trend.iloc[0]
earthquake_trend = earthquake_trend.iloc[1:]
earthquake_trend = earthquake_trend.rename(columns={earthquake_trend.columns
earthquake_trend['Earthquake'] = pd.to_numeric(earthquake_trend['Earthquake'
earthquake_trend.index = pd.to_datetime(earthquake_trend.index, format='%Y-%
```

In [ ]:
```
highest_values = earthquake_trend.nlargest(5, earthquake_trend.columns[0])
lowest_values = earthquake_trend.nsmallest(5, earthquake_trend.columns[0])

print(" Highest Values:")
print(highest_values)
```

```
print(" Lowest Values:")
print(lowest_values)
```

```
 Highest Values:
Mois          Earthquake
2011-03-01           100
2019-07-01            91
2011-08-01            73
2010-01-01            71
2017-09-01            59
 Lowest Values:
Mois          Earthquake
2004-08-01             5
2006-07-01             5
2007-06-01             5
2013-07-01             5
2022-08-01             5
```

- July 2019: Ridgecrest earthquakes, classified as "violent" with an estimated 20 million people who experienced the foreshock and 30 million the mainshock. Total damage estimated around 5.3 billion dollars.
- August 2011: Colorado and Virginia earthquakes, classified as "very strong" and "severe". The latter was felt across more than a dozen states and in several Canadian provinces, more than any other quake in US history.
- January 2010: Eureka earthquake in California.

In [ ]:
```
merged_df = industry_returns.merge(earthquake_trend, left_index=True, right_
```

In [ ]:
```
industry_names = industry_returns.columns
results_df = pd.DataFrame(columns=['Industry', 'Coefficient', 'P-Value'])

for industry_name in industry_names:
    industry_data = pd.DataFrame({'Returns': industry_returns[industry_name]
                                  'Earthquake topic score': earthquake_trenc


    industry_data = industry_data.dropna()


    X = sm.add_constant(industry_data['Earthquake topic score'])
    y = industry_data['Returns']
    model = sm.OLS(y, X).fit()


    coefficient = model.params['Earthquake topic score']
    p_value = model.pvalues['Earthquake topic score']


    results_df = results_df.append({'Industry': industry_name, 'Coefficient'
```

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

results_df
```

Out[ ]:

	Industry	Coefficient	P-Value
24	Ships	8.393297	0.077291
26	Gold	7.321646	0.273853
5	Toys	3.744013	0.332720
2	Soda	3.242719	0.360848
6	Fun	3.090442	0.538569
16	BldMt	2.969922	0.485034
4	Smoke	2.889265	0.344970
20	Mach	2.603538	0.527190
41	Whlsl	2.511577	0.380618
19	FabPr	2.398755	0.634165
14	Rubbr	2.388645	0.519254
43	Meals	2.269230	0.370917
1	Food	2.253877	0.283893
48	Other	2.237240	0.521612
28	Coal	2.060370	0.799914
21	ElcEq	1.953978	0.608267
22	Autos	1.832644	0.714568
17	Cnstr	1.697749	0.693164
40	Trans	1.437365	0.650147
23	Aero	1.401804	0.671767
46	RlEst	1.266074	0.818125
44	Banks	1.006210	0.793981
13	Chems	1.005443	0.785593
29	Oil	0.993410	0.785345
45	Insur	0.987536	0.766995
7	Books	0.652655	0.864584
8	Hshld	0.445792	0.846946
42	Rtail	0.111524	0.965552
12	Drugs	0.057111	0.981265
3	Beer	0.027804	0.989862
25	Guns	-0.095869	0.976140
37	LabEq	-0.168161	0.958621

	Industry	Coefficient	P-Value
38	Paper	-0.210718	0.946464
34	Hardw	-0.241793	0.948175
30	Util	-0.257724	0.908652
15	Txtls	-0.396688	0.944739
33	BusSv	-0.758417	0.786819
11	MedEq	-0.807404	0.778668
10	Hlth	-0.962772	0.764842
31	Telcm	-1.133154	0.671349
9	Clths	-1.145812	0.749331
32	PerSv	-1.448937	0.696307
47	Fin	-2.541714	0.522153
39	Boxes	-3.093262	0.353888
36	Chips	-3.212334	0.377221
18	Steel	-3.469138	0.516783
35	Softw	-3.517526	0.244645
27	Mines	-4.657361	0.421282
0	Agric	-5.930478	0.145704

Ranking Discussion:

Earthquakes: Top 3 negatively impacted industries (based on Coefficient):

Agric Mines Softw

These industries showcase the highest negative coefficients, which might indicate that they are most negatively influenced by earthquakes. However, it's crucial to note the p-values associated with each. For instance, while Agricultural sector has the highest negative coefficient, its p-value of 0.1457 is above the standard 0.05 threshold, suggesting caution in inferring a significant relationship.

Top 3 positively impacted industries:

Ships Gold Toys

Although these industries have positive coefficients, suggesting a potential positive relationship or limited impact from earthquakes, the associated p-values indicate that these relationships may not be statistically significant. However, for Shipping iindustry, we see a p-value of 0.0773, which while above 0.05 threshold is still close enough to infer a potiental relationship.

Conclusion:

Having the agricultural and lining industry suffering the most from earthquakes is conceptually very logical as both activities depend on earth quality and health. On the other hand, shipping industry does not depend on land, and it is logical to assume that if land routes/airports were to be damaged by earthquakes then maritime routes would appear as most attractive. The same can be though of gold which is often seen as a refuge commodity.

Related to transition risk - Carbon Neutrality

In [ ]:
```
carbon_neutrality_trend=pd.read_csv('multiTimeline3.CSV')
carbon_neutrality_trend.columns = carbon_neutrality_trend.iloc[0]
carbon_neutrality_trend = carbon_neutrality_trend.iloc[1:]
carbon_neutrality_trend = carbon_neutrality_trend.rename(columns={carbon_neu
carbon_neutrality_trend['Carbon Neutrality'] = pd.to_numeric(carbon_neutrali
carbon_neutrality_trend.index = pd.to_datetime(carbon_neutrality_trend.index
```

In [ ]:
```
highest_values = carbon_neutrality_trend.nlargest(5, carbon_neutrality_trenc
lowest_values = carbon_neutrality_trend.nsmallest(5, carbon_neutrality_trenc

print(" Highest Values:")
print(highest_values)

print(" Lowest Values:")
print(lowest_values)
```

```
 Highest Values:
Mois         Carbon Neutrality
2004-01-01                 100
2005-07-01                  84
2004-07-01                  82
2004-12-01                  76
2004-10-01                  74
 Lowest Values:
Mois         Carbon Neutrality
2018-07-01                   5
2017-06-01                   6
2017-07-01                   6
2017-09-01                   6
2018-06-01                   6
```

In [ ]:
```
merged_df = industry_returns.merge(carbon_neutrality_trend, left_index=True,
```

In [ ]:
```
industry_names = industry_returns.columns
results_df = pd.DataFrame(columns=['Industry', 'Coefficient', 'P-Value'])

for industry_name in industry_names:
    industry_data = pd.DataFrame({'Returns': industry_returns[industry_name]
                                 'Carbon neutrality topic score': carbon_ne
```

```
    industry_data = industry_data.dropna()


    X = sm.add_constant(industry_data['Carbon neutrality topic score'])
    y = industry_data['Returns']
    model = sm.OLS(y, X).fit()


    coefficient = model.params['Carbon neutrality topic score']
    p_value = model.pvalues['Carbon neutrality topic score']


    results_df = results_df.append({'Industry': industry_name, 'Coefficient'


results_df = results_df.sort_values(by='Coefficient', ascending=False)

results_df
```

Out[ ]:

	Industry	Coefficient	P-Value
28	Coal	6.550335	0.163909
29	Oil	2.446433	0.246990
27	Mines	2.432351	0.469397
0	Agric	1.187312	0.616757
18	Steel	1.138987	0.713949
30	Util	1.109309	0.394362
39	Boxes	0.557214	0.773814
4	Smoke	0.198374	0.911167
19	FabPr	-0.090373	0.975367
17	Cnstr	-0.395606	0.874167
8	Hshld	-0.451789	0.736083
46	RlEst	-0.501009	0.875427
21	ElcEq	-0.717261	0.745887
1	Food	-0.753144	0.537743
43	Meals	-0.853299	0.562412
47	Fin	-0.947914	0.681010
41	Whlsl	-1.060974	0.523636
32	PerSv	-1.132054	0.599256
48	Other	-1.183931	0.559072
26	Gold	-1.218399	0.754121
16	BldMt	-1.227581	0.619159
10	Hlth	-1.270446	0.496245
20	Mach	-1.275201	0.593710
2	Soda	-1.352206	0.511782
13	Chems	-1.380414	0.519716
25	Guns	-1.519623	0.413607
31	Telcm	-1.593881	0.303233
40	Trans	-1.750939	0.340667
37	LabEq	-1.782373	0.342791
23	Aero	-1.814934	0.344081
5	Toys	-1.844305	0.411221
9	Clths	-1.931901	0.352799

	Industry	Coefficient	P-Value
11	MedEq	-1.984668	0.233041
34	Hardw	-1.991919	0.355663
24	Ships	-2.075570	0.453346
42	Rtail	-2.106324	0.158732
12	Drugs	-2.191311	0.119285
45	Insur	-2.194587	0.255772
14	Rubbr	-2.194866	0.307154
3	Beer	-2.242241	0.076136
33	BusSv	-2.266846	0.162584
15	Txtls	-2.349995	0.478977
35	Softw	-2.550925	0.145602
44	Banks	-2.701995	0.226046
38	Paper	-2.755675	0.129029
36	Chips	-3.606141	0.086680
7	Books	-3.812097	0.084796
22	Autos	-4.000754	0.167876
6	Fun	-4.389828	0.131262

Ranking Discussion:

Carbon neutrality: Top 3 negatively impacted industries (based on Coefficient):

Fun Autos Books

These industries showcase the highest negative coefficients, which might indicate that they are most negatively influenced by carbon nautrality and transition risks. However, it's crucial to note the p-values associated with each. For instance, while the Fun sector has the highest negative coefficient, its p-value of 0.1312 is above the standard 0.05 threshold, suggesting caution in inferring a significant relationship.

Top 3 positively impacted industries:

Coal Oil Mines

Although these industries have positive coefficients, suggesting a potential positive relationship or limited impact from earthquakes, the associated p-values indicate that these relationships may not be statistically significant.

Conclusion:

While we find logical result that the automobile sector is one of the most negatively impacted despite high p-value, we also find contradictory results with coal, mining and oil industry being positively impacted by carbon neutrality. This might suggest that these industries are not really impacted by transition risks.

In [ ]: