Assignment 7: Predicting Operational Risk Using NLP

Name: Shiqi Hu GTID: 904061372

Analysis

Justification for Choosing Option 1 (Cybersecurity Risk Ratio)

• The cybersecurity risk ratio focuses on the relative importance of cybersecurity within each company's risk assessment. By using a ratio instead of a raw count, we can compare the emphasis on cybersecurity risk across firms, regardless of how much they write about risks overall. This makes it easier to compare cybersecurity risk across firms of different sizes or with varying levels of overall risk disclosure.

Observation on Cybersecurity Risk for All the Years Across Different Industry

- Finance, Insurance and Real Estate has the highest mean cybersecurity risk (0.020273), significantly higher than other sectors, with a high variability (standard deviation of 0.022927). This indicates that this sector faces the greatest exposure to cyber threats, likely due to its handling of sensitive financial data and strict regulatory requirements.
- Wholesale Trade demonstrates the second highest mean risk (0.013640), accompanied by relatively high variability (standard deviation of 0.016582). This suggests notable variation in cybersecurity risk exposure across companies in this sector.
- Agriculture, Forestry and Fishing has a relatively low mean risk (0.010314) and the lowest kurtosis (0.182918), suggesting a more uniform and generally lower
 perception of cybersecurity risks across companies in this sector.
- Manufacturing shows the lowest mean risk (0.007050) among all sectors but has a high kurtosis (3.741473), implying that while most companies in this sector report low cybersecurity risks, there are some significant outliers.
- Retail Trade exhibits interesting characteristics with a negative skewness (0.894671), suggesting a more uniform distribution of cybersecurity risks across
 companies in this sector, with fewer extreme outliers compared to others.
- Mining has the highest kurtosis (7.524308), reflecting extreme variation in cybersecurity risks, with certain companies reporting risks significantly higher than the sector average of 0.007442.
- Services shows a moderate mean risk (0.011581) and kurtosis (2.688571), indicating a balanced risk perception with some variability.
- Transportation and Other Utilities has a comparable mean risk (0.011583) and a standard deviation of 0.014042, reflecting moderate variability in cybersecurity risk across this industry.
- Construction reports a slightly higher mean risk (0.010290) than Agriculture, with a standard deviation of 0.011761 and a relatively high skewness (1.377157), indicating a heavier tail toward higher risks.
- Public Administration reports a mean risk of 0.012126 with moderate variability (standard deviation of 0.012076) and a balanced kurtosis (0.993258), reflecting relatively uniform risk perception within the industry.

Observation on Mean and Standard Deviation of Cybersecurity Risk for Each Industry for Each Year

- Finance, Insurance, and Real Estate demonstrates the most dramatic increase in cybersecurity risk, with the mean peaking at 0.053129 in 2020 and the standard deviation reaching its highest at 0.041241 in 2016. The sector has consistently shown elevated risk levels from 2016 to 2022, reflecting its exposure to cyber threats due to the sensitive nature of financial data and stringent regulatory compliance requirements.
- Services and Wholesale Trade exhibit significant risk growth after 2015. Services reached its highest mean of 0.038632 in 2022, with a notable increase in variability (standard deviation 0.016065) during its rise. Wholesale Trade peaked with a mean of 0.033379 in 2018, maintaining higher variability (standard deviation 0.021437) across companies, highlighting diverse risk exposures.
- Construction shows a steady upward trajectory, starting from near-zero levels in the 2005-2008 period and reaching its highest mean of 0.029413 in 2022. Its standard deviation remained relatively stable, reflecting a uniform increase in risk across the sector, possibly driven by the adoption of new technologies and infrastructure digitization.
- Agriculture, Forestry, and Fishing exhibits a more moderate but consistent increase in cybersecurity risk, with its mean rising from 0.000708 in 2006 to 0.018156 in 2021. Interestingly, this sector's standard deviation decreased over time (from 0.016709 in 2013 to 0.008591 in 2022), indicating more uniformity in perceived risk among companies.
- Mining shows volatile growth, with sharp jumps in mean risk post-2015, peaking at 0.027204 in 2022. The high variability (standard deviation frequently exceeding
 the mean) underscores the uneven distribution of risk across the sector, likely driven by varying levels of technological maturity and exposure to cyber threats.
- Public Administration demonstrates the most stable pattern among all sectors, with moderate risk growth over time. Its mean rose from 0.002591 in 2006 to 0.016899 in 2022, with standard deviations remaining relatively low and consistent, reflecting standardized cybersecurity practices and uniformity across organizations.
- Manufacturing shows slower growth in cybersecurity risk compared to other industries, with the mean increasing from 0.003038 in 2006 to 0.028401 in 2022. The
 sector has maintained a high standard deviation, indicating the presence of significant outliers, likely reflecting uneven adoption of cybersecurity measures.

Equal-Weighted Portfolio Comment and Discussion

- The returns for both high cybersecurity risk (Quintile 5) and low cybersecurity risk (Quintile 1) portfolios exhibit significant volatility, particularly during the early 2000s and 2003-2004. In 2003-2004, Quintile 5 displayed its highest return spike, exceeding 2.0, reflecting heightened sensitivity to market dynamics.
- Notable periods of divergence between high-risk and low-risk portfolios include:
 - Early 2000s, marked by a sharp reversal from high-risk underperformance to significant outperformance.
 - o 2003-2004, showing the largest return spread between high and low-risk portfolios.
 - o 2009-2010, with another phase of pronounced outperformance by high-risk portfolios, driven by recovery trends post-2008 financial crisis.
- The High Risk-Low Risk return spread (green line) remains volatile across the sample period, oscillating between positive and negative values. This highlights the instability of the relationship between cybersecurity risk and returns, with market conditions playing a significant role.
- Post-2015, the return differences between high and low-risk portfolios diminish noticeably, with the spread stabilizing around zero. Possible explanations include:
 - $\circ \;\;$ Improved efficiency in market pricing of cybersecurity risks.
 - O More robust risk management frameworks adopted across industries.

- o Increasing standardization of cybersecurity measures and disclosures.
- The recent period (2018-2022) exhibits relatively subdued fluctuations in both portfolio returns and their differences, suggesting maturation in how markets perceive and price cybersecurity risks, along with potentially reduced uncertainty regarding the economic impact of such risks.

Value-Weighted Portfolio Comment and Discussion

- The value-weighted portfolios show pronounced volatility, highlighted by two extreme events:
 - 2003-2004, with a sharp positive spike where both Quintile 5 and the High Risk-Low Risk return peaked near 2.0, indicating strong performance by high-risk, high-cap companies.
 - 2010, marked by a dramatic negative spike where the High Risk-Low Risk return fell to approximately -1.2, driven by significant outperformance of low-risk companies during economic stabilization.
- The relationship between high and low cybersecurity risk portfolios evolves through three distinct phases:
 - o 2000-2003, characterized by relatively stable fluctuations with no pronounced divergence.
 - o 2003-2010, a period of heightened volatility with extreme swings in return spreads, reflecting market turbulence and differential risk perceptions.
 - Post-2010, characterized by more subdued fluctuations and narrower return differentials, indicating convergence in risk management practices among larger firms.
- The most recent period (2015-2022) reflects the smallest return differentials, with the High Risk-Low Risk return line remaining close to zero. This indicates alignment in market capitalization-weighted returns for high and low cybersecurity risk companies, suggesting reduced differentiation in risk premiums.
- Unlike earlier periods, recent years demonstrate muted reactions to market stress events, indicating a more sophisticated and nuanced market approach to assessing cybersecurity risks among large-cap firms.

Descriptive Analysis

1. Compute the descriptive stats

	count	mean	std	min	max	skew	kurtosis	1%	5%	25%	50%	75%	95%	99%
industry														
Agriculture, Forestry and Fishing	144	0.010314	0.012193	0.0	0.044444	1.018134	0.182918	0.0	0.0	0.000000	0.005962	0.018966	0.034454	0.0444
Construction	172	0.010290	0.011761	0.0	0.058394	1.371757	1.989498	0.0	0.0	0.000000	0.006730	0.017033	0.031111	0.0505
Finance, Insurance and Real Estate	200	0.020273	0.022927	0.0	0.142857	1.779232	4.446305	0.0	0.0	0.000000	0.015244	0.028479	0.061169	0.0912
Manufacturing	269	0.007050	0.011229	0.0	0.060606	1.918140	3.741473	0.0	0.0	0.000000	0.000000	0.010989	0.031548	0.0414
Mining	186	0.007442	0.012125	0.0	0.067873	2.579830	7.524308	0.0	0.0	0.000000	0.002366	0.010194	0.035885	0.0569
Public Administration	155	0.012126	0.012076	0.0	0.055838	1.159158	0.993258	0.0	0.0	0.002444	0.008798	0.018512	0.035009	0.0474
Retail Trade	209	0.012711	0.013196	0.0	0.053571	0.894671	-0.010900	0.0	0.0	0.000000	0.010417	0.021127	0.038860	0.0474
Services	235	0.011581	0.014293	0.0	0.074689	1.537934	2.688571	0.0	0.0	0.000000	0.006757	0.020529	0.038795	0.0608
Transportation and other Utilities	191	0.011583	0.014042	0.0	0.068182	1.266751	1.056648	0.0	0.0	0.000000	0.006623	0.020036	0.039709	0.0503
Wholesale Trade	190	0.013640	0.016582	0.0	0.076923	1.541294	2.349952	0.0	0.0	0.000000	0.008753	0.021520	0.044679	0.0709

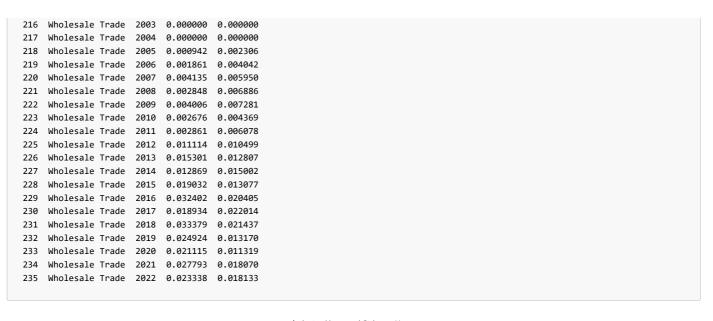
2. Compute the mean and standard deviation for each industry each year

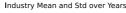
Industry: Agriculture, Forestry and Fishing industry year mean 0 Agriculture, Forestry and Fishing 2006 0.000708 0.002123 Agriculture, Forestry and Fishing 2007 0.001724 0.003220 Agriculture, Forestry and Fishing 2008 0.001227 0.002359 3 Agriculture, Forestry and Fishing 2009 0.001733 0.002327 4 Agriculture, Forestry and Fishing 2010 0.002784 0.004360 Agriculture, Forestry and Fishing 2011 0.003048 0.006541 Agriculture, Forestry and Fishing 2012 0.008742 0.015898 Agriculture, Forestry and Fishing 2013 0.010061 0.016709 8 Agriculture, Forestry and Fishing 2014 0.011957 0.015451 9 Agriculture, Forestry and Fishing 2015 0.016098 0.017545 10 Agriculture, Forestry and Fishing 2016 0.014394 0.015862 11 Agriculture, Forestry and Fishing 2017 0.019522 0.013353

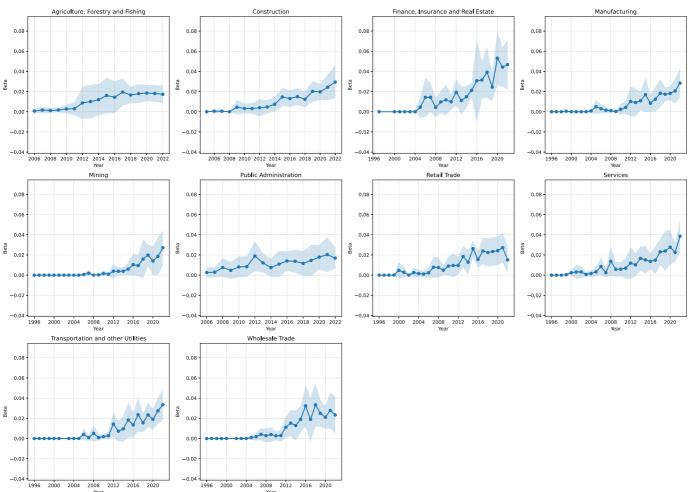
```
12 Agriculture, Forestry and Fishing 2018 0.016560 0.008096
13 Agriculture, Forestry and Fishing 2019 0.017963 0.009455
14 Agriculture, Forestry and Fishing 2020 0.018491 0.007950
15 Agriculture, Forestry and Fishing 2021 0.018156 0.008252
16 Agriculture, Forestry and Fishing 2022 0.017289 0.008591
Industry: Construction
       industry year
                          mean
                                    std
17 Construction 2005 0.000000 0.000000
18 Construction 2006 0.000552 0.001747
19 Construction 2007 0.000524 0.001656
20 Construction 2008 0.000000 0.000000
21 Construction 2009 0.004518 0.007190
22 Construction 2010 0.003256 0.004898
23 Construction 2011 0.003181 0.005250
24 Construction 2012 0.004010 0.008345
25 Construction 2013 0.004733 0.006433
26 Construction 2014 0.007467 0.007135
27 Construction 2015 0.014646 0.009085
28 Construction 2016 0.013081 0.008189
29 Construction 2017 0.014984 0.008342
30 Construction 2018 0.012340 0.007890
31 Construction 2019 0.020244 0.010143
32 Construction 2020 0.019679 0.007993
33 Construction 2021 0.024355 0.012600
34 Construction 2022 0.029413 0.016416
Industry: Finance, Insurance and Real Estate
                           industry year
                                              mean
                                                         std
35 Finance, Insurance and Real Estate 1997 0.000000
                                                         NaN
36 Finance, Insurance and Real Estate 2000 0.000000 0.000000
37 Finance, Insurance and Real Estate 2001 0.000000
38 Finance, Insurance and Real Estate 2002 0.000000
39
  Finance, Insurance and Real Estate 2003 0.000000 0.000000
40 Finance, Insurance and Real Estate 2004 0.000000 0.000000
41 Finance, Insurance and Real Estate 2005 0.004482 0.010391
42 Finance, Insurance and Real Estate 2006 0.014305 0.019463
43 Finance, Insurance and Real Estate 2007 0.014361 0.014544
44
   Finance, Insurance and Real Estate 2008 0.004375 0.009669
45 Finance, Insurance and Real Estate 2009 0.009719 0.010733
46 Finance, Insurance and Real Estate 2010 0.011774 0.015431
47 Finance, Insurance and Real Estate 2011 0.009881 0.007710
48
  Finance, Insurance and Real Estate 2012 0.019342 0.013179
49 Finance, Insurance and Real Estate 2013 0.011072 0.013920
50 Finance, Insurance and Real Estate 2014 0.014941 0.010711
51 Finance, Insurance and Real Estate 2015 0.021247 0.007220
52 Finance, Insurance and Real Estate 2016 0.030726 0.041241
53 Finance, Insurance and Real Estate 2017 0.031435 0.018436
54 Finance, Insurance and Real Estate 2018 0.039211 0.025305
55 Finance, Insurance and Real Estate 2019 0.024441 0.013881
56 Finance, Insurance and Real Estate 2020 0.053129 0.026964
57 Finance, Insurance and Real Estate 2021 0.044253 0.019310
58 Finance, Insurance and Real Estate 2022 0.046766 0.024764
Industry: Manufacturing
        industry year
                          mean
                                     std
59 Manufacturing 1996 0.000000 0.000000
60 Manufacturing 1997 0.000000 0.000000
61 Manufacturing 1998 0.000000 0.000000
62 Manufacturing 1999 0.000513 0.001622
63 Manufacturing 2000 0.000000 0.000000
64
   Manufacturing 2001 0.000000 0.000000
65 Manufacturing 2002 0.000000 0.000000
66 Manufacturing 2003 0.000000 0.000000
67 Manufacturing 2004 0.000769 0.002433
68 Manufacturing 2005 0.004938 0.006773
69
  Manufacturing 2006 0.003038 0.004628
70 Manufacturing 2007 0.001512 0.003194
71 Manufacturing 2008 0.001050 0.001727
72 Manufacturing 2009 0.000248 0.000783
73
   Manufacturing 2010 0.002408 0.005316
74 Manufacturing 2011 0.004300 0.006361
75 Manufacturing 2012 0.010261 0.014752
76 Manufacturing 2013 0.009141 0.013965
77 Manufacturing 2014 0.010788 0.009198
   Manufacturing 2015 0.016952 0.017275
79 Manufacturing 2016 0.008462 0.008801
```

```
80 Manufacturing 2017 0.012437 0.012240
81 Manufacturing 2018 0.018322 0.013526
82 Manufacturing 2019 0.017342 0.007058
83 Manufacturing 2020 0.018185 0.006853
84 Manufacturing 2021 0.020579 0.012352
85 Manufacturing 2022 0.028401 0.014373
Industry: Mining
   industry year
                     mean
                                std
86
   Mining 1996 0.000000 0.000000
87
     Mining 1997 0.000000
                               NaN
88
     Mining 1998 0.000000
                                NaN
    Mining 1999 0.000000
89
                               NaN
    Mining 2000 0.000000
    Mining 2001 0.000000
91
                               NaN
92
     Mining 2002
                 0.000000 0.000000
93
    Mining 2003 0.000000
                               NaN
94
    Mining 2004 0.000000
                                NaN
95
    Mining 2005 0.000000 0.000000
     Mining 2006 0.000606 0.001917
96
97
     Mining 2007
                 0.002001 0.002751
    Mining 2008 0.000000 0.000000
98
     Mining 2009 0.000265 0.000837
100 Mining 2010 0.001659 0.002808
101
     Mining 2011 0.000916 0.002079
102
    Mining 2012 0.003843 0.007017
103
    Mining 2013 0.003812 0.004157
104
    Mining 2014 0.003895 0.003867
     Mining 2015 0.005977 0.006241
105
106
     Mining 2016 0.010393 0.010741
    Mining 2017 0.009495 0.010510
107
108 Mining 2018 0.015909 0.019313
    Mining 2019 0.019857 0.010624
109
110
     Mining 2020 0.014025 0.014165
111
     Mining 2021 0.018564 0.018845
112 Mining 2022 0.027204 0.017123
Industry: Public Administration
               industry year
                                  mean
113 Public Administration 2006 0.002591 0.005181
114 Public Administration 2007 0.002933 0.005079
115 Public Administration 2008 0.007573 0.009339
116 Public Administration 2009 0.004755 0.008420
117 Public Administration 2010 0.008170 0.010261
118 Public Administration 2011 0.008436 0.010118
119 Public Administration 2012 0.018901 0.014773
120 Public Administration 2013 0.012210 0.009546
121 Public Administration 2014 0.007475 0.008782
122 Public Administration 2015 0.010956 0.013165
123 Public Administration 2016 0.014073 0.009592
124 Public Administration 2017 0.013734 0.011426
125 Public Administration 2018 0.011637 0.012193
126 Public Administration 2019 0.014549 0.015550
127 Public Administration 2020 0.017828 0.012161
128 Public Administration 2021 0.020353 0.017265
129 Public Administration 2022 0.016899 0.010421
Industry: Retail Trade
      industry year
                                    std
                         mean
130 Retail Trade 1996 0.000000
131 Retail Trade 1997 0.000000 0.000000
132 Retail Trade 1998 0.000000
133 Retail Trade 1999 0.000000 0.000000
134 Retail Trade 2000 0.004695 0.008132
135 Retail Trade 2001 0.002841 0.005682
136 Retail Trade 2002 0.000000 0.000000
137 Retail Trade 2003 0.002525 0.005051
138 Retail Trade 2004 0.001515 0.003711
139 Retail Trade 2005 0.001042 0.003294
140 Retail Trade 2006 0.002286 0.004828
141 Retail Trade 2007 0.007857 0.010454
142 Retail Trade 2008 0.007575 0.009942
143 Retail Trade 2009 0.004925 0.004528
144 Retail Trade 2010 0.008998 0.009404
145 Retail Trade 2011 0.009575 0.011835
    Retail Trade 2012 0.009664 0.008791
147 Retail Trade 2013 0.018412 0.010625
```

```
148 Retail Trade 2014 0.012798 0.012198
    Retail Trade 2015 0.026208 0.007977
150 Retail Trade 2016 0.015468 0.008258
151 Retail Trade 2017 0.023938 0.013057
152 Retail Trade 2018 0.022460 0.013950
153 Retail Trade 2019 0.023414 0.014184
154 Retail Trade 2020 0.024240 0.014097
155 Retail Trade 2021 0.027216 0.014891
156 Retail Trade 2022 0.015118 0.013594
Industry: Services
    industry year
                       mean
157 Services 1996 0.000000 0.000000
158 Services 1997 0.000000 0.000000
159 Services 1998 0.000000 0.000000
160 Services 1999 0.000412 0.001008
161 Services 2000 0.002331 0.002740
162 Services 2001 0.003126 0.006723
163 Services 2002 0.003208 0.006169
164 Services 2003 0.000751 0.002252
165 Services 2004 0.001563 0.003494
166 Services 2005 0.003217 0.004532
167 Services 2006 0.008513 0.008801
168 Services 2007 0.002406 0.004676
169 Services 2008 0.013659 0.014343
170 Services 2009 0.005792 0.007275
171 Services 2010 0.005872 0.007086
172 Services 2011 0.006901 0.010274
173 Services 2012 0.011882 0.015856
174 Services 2013 0.010207 0.011211
175 Services 2014 0.016559 0.011903
176 Services 2015 0.015091 0.013893
177 Services 2016 0.013639 0.007619
178 Services 2017 0.014825 0.013855
179 Services 2018 0.023081 0.013106
180 Services 2019 0.023930 0.019725
181 Services 2020 0.027661 0.017360
182 Services 2021 0.022566 0.008944
183 Services 2022 0.038632 0.016065
Industry: Transportation and other Utilities
                            industry year
                                               mean
                                                          std
184 Transportation and other Utilities 1996 0.000000
                                                          NaN
185 Transportation and other Utilities 1997 0.000000 0.000000
186 Transportation and other Utilities 1998 0.000000 0.000000
187 Transportation and other Utilities 1999 0.000000
188 Transportation and other Utilities 2000 0.000000
                                                         NaN
189 Transportation and other Utilities 2001 0.000000
                                                     0.000000
190 Transportation and other Utilities 2003 0.000000
191 Transportation and other Utilities 2004 0.000000 0.000000
192 Transportation and other Utilities 2005 0.000000 0.000000
193 Transportation and other Utilities 2006 0.004066 0.005970
194 Transportation and other Utilities 2007 0.000844 0.002669
195 Transportation and other Utilities 2008 0.005187 0.007860
196 Transportation and other Utilities 2009 0.000932 0.002946
197 Transportation and other Utilities 2010 0.001920 0.002561
198 Transportation and other Utilities 2011 0.002794 0.003670
199 Transportation and other Utilities 2012 0.014354 0.012041
200 Transportation and other Utilities 2013 0.007276 0.010210
201 Transportation and other Utilities 2014 0.009559 0.012613
202 Transportation and other Utilities 2015 0.018396 0.017476
    Transportation and other Utilities 2016 0.013374 0.012235
203
204 Transportation and other Utilities 2017 0.023583 0.016474
205 Transportation and other Utilities 2018 0.015569 0.009442
206 Transportation and other Utilities 2019 0.023456 0.012044
    Transportation and other Utilities 2020 0.019032 0.010139
    Transportation and other Utilities 2021 0.027407 0.012984
208
209 Transportation and other Utilities 2022 0.033481 0.015104
Industry: Wholesale Trade
           industry year
                             mean
210 Wholesale Trade 1996 0.000000 0.000000
211 Wholesale Trade 1997
                          0.000000 0.000000
212 Wholesale Trade 1998
                          0.000000
                                        NaN
213 Wholesale Trade 1999
                          0.000000 0.000000
214 Wholesale Trade 2000
                          0.000000
                                        NaN
215 Wholesale Trade 2002 0.000000
                                        NaN
```







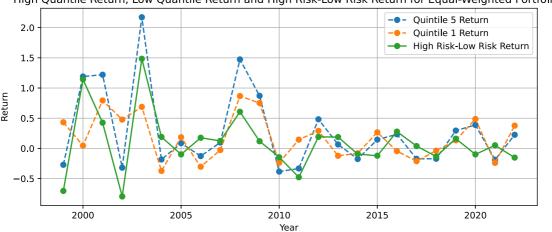
Step 5: Portfolio Analysis

- As all the companies in year 1996-1998 has 0 cybersecurity risk, it's not meaningful to cut portfolio into five quantile based on all 0 cybersecurity risk and make comparison in return for different quantile portfolio. Therefore, year 1996-1998 is dropped.
- For 1999-2010, given that over 80% of companies report zero cybersecurity risk, I use a binary classification: Quintile 1 for zero-risk companies and Quintile 5 for those with any non-zero risk.
- After 2010, as cybersecurity risk ratios become more diverse, I divide companies into five quintiles based on their cybersecurity risk levels and make comparison
 on equal-weighted and value-weighted return accordingly.

1. Form equal weighted portfolios

```
Average return for quintile 5 across all years: 0.2761
Average return for quintile 1 across all years: 0.1778
Equal-weighted return difference (High Risk-Low Risk) for each year:
Year 1999: -0.7040
Year 2000: 1.1437
Year 2001: 0.4266
Year 2002: -0.7960
Year 2003: 1.4844
Year 2004: 0.1896
Year 2005: -0.0986
Year 2006: 0.1765
Year 2007: 0.1244
Year 2008: 0.6064
Year 2009: 0.1190
Year 2010: -0.1459
Year 2011: -0.4776
Year 2012: 0.1915
Year 2013: 0.1883
Year 2014: -0.0941
Year 2015: -0.1211
Year 2016: 0.2786
Year 2017: 0.0388
Year 2018: -0.1336
Year 2019: 0.1612
Year 2020: -0.0989
Year 2021: 0.0513
Year 2022: -0.1518
```

High Quantile Return, Low Quantile Return and High Risk-Low Risk Return for Equal-Weighted Portfolios



2. Form value-weighted portfolio

```
Value-weighted return for quintile 5 across all years: 0.0323
Value-weighted return for quintile 1 across all years: 0.0205
Value-weighted return difference (High Risk-Low Risk) for each year:
Year 1999: 0.0203
Year 2000: 0.0755
Year 2001: 0.2246
Year 2002: -0.2862
Year 2003: 1.7670
Year 2004: 0.0097
Year 2005: 0.0777
Year 2006: 0.2337
Year 2007: 0.0702
Year 2008: -0.0433
Year 2009: -1.2910
Year 2010: -0.2469
Year 2011: 0.0043
Year 2012: -0.1513
```

```
Year 2013: 0.0805
Year 2014: -0.1539
Year 2015: -0.1190
Year 2016: 0.1235
Year 2017: 0.0772
Year 2018: -0.0133
Year 2019: -0.2872
Year 2020: 0.0101
Year 2021: 0.1621
Year 2022: -0.0615
```

