Asset Pricing Report – Replicate <Pollution Premium>

Group 2：Zhihao Wang, Qiaochu Liu, Yixin Cheng, Jiaao Liu, Xilong Liu

1. Executive Summary

We have replicated the first five tables of the Pollution Premium in this report. We chose the period from 1991 to 2022, as opposed to 1991 to 2016 in the original paper. We used the same TRI data, bridge dataset to connect TRI to Compustat dataset, and Fama-French five factor data. For all unpublished data including financial ratios and stock price information, we use WRDS. Due to data size constraints, we were not able to identify all the pollution premiums mentioned in the paper. However, our report supports the general idea that pollution premiums exist. With more market capitalization data, we will be able to have many more observations and get a better result.

2. Statistical Summary

B/M	book equity to market capitalization
I/K	capital expenditures divided by property, plant, and equipment
ROA	operating income after depreciation scaled by total assets
ROE	operating income after depreciation scaled by total assets
TANT	property, plant, and equipment divided by total assets
OL	summation of COGS and selling, general, and administrative expenses scaled by assets
Lev	summation of current liabilities and long-term debt scaled by total assets

In Table I Panel A (Appendix), we report the pooled mean, standard deviation (Std), 5th percentile (P5), 25th percentile (P25), median, 75th percentile (P75), and 95th percentile (P95). The sample period is 1991 to 2022 at an annual frequency.

The average emission is 1,914,023 pounds, and the standard deviation is 7,835,467 pounds, indicating that the company's emissions are different. The median number is 49,658 pounds, and most of the company's emissions are less than the average. For 'Market', the average value of the company's market is 4,227.37 million US dollars, the standard deviation is 21,105.52 million US dollars, and the median number is 634.37 million US dollars. As for 'Book market ratio', the average of B/M is 0.65, the standard deviation is 2.00, and the median is 0.51, indicating that the company generally has a high book value ratio. The average of ROA and ROE is 0.14 and 0.10, respectively. Most companies' ROA and ROE are relatively low. The standard difference between the ROE (0.77) indicates that the return on equity fluctuations fluctuate. TANT average is 3,661.05, the standard deviation is 14,740.79, and the company's investment differences in tangible assets are significantly different. Finally, the average of operating leverage (OL) is 2.41, the standard deviation is 1.53, and the average of the LEV is 0.57. The company has a large difference in operating leverage and financial leverage. As for Panel B, the number of positive relationships between the emissions and the market value is 0.21, and the company with a large market value is also large. Moreover, the positive relationship between ROA and ROE is 0.20, and the company's share capital return rate is also high. The correlation coefficient between TANT and the market value is 0.82, the market value of companies with higher tangible assets is also high. The correlation coefficient between OL and Lev is 0.17, companies with higher operating leverage often have higher financial leverage.

3. Data feature

We attribute the discrepancy between Table II and the original article primarily to the quality of the dataset. The authors, due to copyright constraints, have only provided a list of TRI facilities and mentioned that the ROE and I/A data were sourced from other researchers. Because of this, we collected most of the data by ourselves. Financial ratios, market capitalization and other firm data were from CRSP and other WRDS databases.

However, due to restrictions on database access, the dataset for market capitalization consists of approximately 15,000 monthly records, significantly smaller than the original dataset size of 150,000 records. This limitation has resulted in a sample size of only around 133 firms, with many facing listing or delisting issues during the 1998-2022 period. The number of firms within the five portfolios (from L to H) each year during this period is less than 20, with fewer than 10 firms at the beginning and end of the period. We believe this sample size causes the H-L portfolio cannot fully mitigate systematic risk, leading to the insignificant results seen in Table II and Table IV.

Another concern is the handling of outliers. From the portfolio construction perspective, outliers should be considered as they were part of the market. However, in a small sample size, outliers can significantly impact the sample mean, standard deviation, and consequently, the significance of the results. We believe this is one of the main reasons for insignificant result in Table II.

4. Univariate Portfolio Sorting

As the authors of the paper, we construct quintile portfolios sorted on firms' emissions scaled by total assets (AT) in Panel A, property, plant, and equipment (PPENT) in Panel B, sales (SALE) in Panel C, and outstanding shareholder's equity (EQ) in Panel D to report each portfolio's post-formation average stock return. The authors of Pollution Premium used market equity in Panel D while the data was not available. So, we chose outstanding shareholder's equity as the alternative. Besides, original authors used "industry peer' average" as the denominator in the scaling. As it is hard to pick an industry peer for all the companies in the industry, we choose to use the corresponding firm value in scaling. In this way, we can tag high-emission companies even if they may have relatively low absolute emissions. We then assign all firms with positive scaled emissions in the year t into quintile portfolios. The low (high) quintile portfolio contains firms with the lowest (highest) emissions.

To examine the emission-return relation, we follow the steps of the paper, forming an H-L portfolio that takes a long position in the high-emission portfolio and a short position in the low-emission portfolio and calculating the return on this portfolio. After forming the five portfolio sorts (from low to high), we calculate the market capitalization-weighted monthly returns on these portfolios over the next 12 months. In Panels A to D of Table II (Appendix), the top row presents the annualized average excess stock return in percentage, t-statistic, standard deviation, and Sharpe ratio.

Comparing our result to the outcome of the paper, we got the positive risk premium in group H-L while it is not significant. Besides, the excess return of 5 groups is not perfectly aligned with their emission situation. This may be due to the features we will discuss before.

Overall, our Table II conclusions align with the author's thesis in the paper. Consistent with the authors' approach, we will focus on emission intensity defined as annual emissions scaled by total assets and the associated portfolios.

5. Portfolios Firm Characteristics

Table III reports the average firm characteristics across quintile portfolios. In Table III, we find that firms in the high-emission group generated emissions of 8,445,205 on average, while firms in the low-emission group generated emissions of 1596.86. In addition, the emission intensity of the high (low) group is 16,001.83 (0.75). We further find that high-emission firms are smaller and have higher asset tangibility and higher operating leverage. However, there is little variation in B/M, I/K, ROA, financial constraints, and financial leverage across emission-sorted portfolios.

6. Asset Pricing Factor Tests

Table IV shows asset pricing factor tests for five portfolios sorted on emissions scaled by total assets relative to their industry peers, for which we use the Standard Industrial Classification (SIC) from WRDS and rebalance portfolios at the end of each fiscal year. The first 4 panel results reflect monthly data, while the last for HXZ model reflects annual data as ROE and I/A are reported yearly. The sample runs from July 1998 to December 2022 and excludes financial industries. To adjust for risk exposure, we perform time-series regressions of emission-sorted portfolios' excess returns on the market factor (MKT) as the CAPM model in Panel A, on the Fama and French (1996) three factors (MKT, the size factor-SMB, and the value factor-HML) in Panel B, on the Fama and French (1996) three factors plus Carhart (1997) factor (MKT, SMB, HML, and the momentum factor-UMD) in Panel C, on the Fama and French (2015) five factors (MKT, SMB, HML, the profitability factor-RMW, and the investment factor-CMA) in Panel D, and on the q-factors (MKT, SMB, the investment factor-I/A, and the profitability factor-ROE) in Panel E, respectively. Data on the Fama-French five factors and Carhart factor come from Kenneth French's website. Data on ROE factors are retrieved from WRDS while Investment to asset (I/A) is calculated based on company-month specific total asset (AT), with a formula of '$(\boldsymbol{AT_t} - \boldsymbol{AT_{t-1}})/\boldsymbol{AT_{t-1}}$'. t-Statistics are automatically generated by OLS regression function in Python statsmodels.api library, whose formula is '$\boldsymbol{E}(\boldsymbol{R})/(\boldsymbol{\sigma}/\sqrt{\boldsymbol{n}})$', with n for time range in this case.

In Table IV, we follow standard procedure and investigate the extent to which the variation in the average returns of the emission-sorted portfolios can be explained by existing risk factors. The table reports the alphas from the leading risk factor models, including the capital asset pricing model (CAPM), the Fama-French five-factor model (Fama and French (2015)), and the HXZ q-factor model. We find that the cross-sectional return spread across portfolio groups sorted on AT-scaled emission intensity (p_at) cannot be captured by these risk factors, as the alphas remain statistically significant in some groups. Though our 'H-L' alphas are not significant, the varying alpha significance still implies that our emission-based feature is an unattended factor in traditional pricing models. With more samples covering the full year and month range, constant stds are expected to scale down, making itself more statistically significant.

7. Fama-MacBeth Regressions

Table V reports Fama-MacBeth regressions of individual stock excess returns on their emission intensity in logarithm and other firm characteristics. We conduct cross-sectional regressions for each month of year 1991 to year 2022. In each month, monthly excess return of individual stocks are regressed on emission intensity in logarithm of year t-1, different sets of control variables know by the end of year t and industry fixed effects. Control variables include the natural logarithm of market capitalization (eq), the natural logarithm of book-to-market ratio(B/M), investment rate(I/K), return on equity(ROE), tangibility(TANT), book leverage(debt-to-asset) and industry dummies based on Fama and French 75-industry classification. All independent variables are normalized to zero mean and unit standard deviation to reduce the impact of outliers. T-statistics based on standard errors estimated are reported. The sample period is 1991 to 2022.

In Table V, we examine the emission-return relation by running Fama-MacBeth regressions to control for a variety of firm characteristics. The results of these regressions are not consistent with the results obtained when we sort portfolios on emission intensity, that emission intensity does not significantly positively predict future stock returns, which might be caused by lack of data in each year. For example, we only have 9 observations in 1991. In addition, the predictability of emission intensity is almost subsumed by known predictors of stock returns in the literature.

We also implement independent double sorts for emission intensity and size to alleviate the concern that the return predictability we document is driven by firm size. We find that high-emission firms continue to outperform low-emission firms in stock returns for both large-firm and small-firm groups.

Appendix

All data and code we used in this project:

https://drive.google.com/drive/folders/1VEO8n8twi4l96UeIF7qPzn0wunugWU7A?usp=drive_link

Table I Statistics and Correlations

Index	Emissions	EQ	B/M	ROA	ROE	TANT	OL	Lev
				Panel A: Summary Statistics				
Mean	1,914,023.45	4,227.37	0.65	0.14	0.10	3,661.05	2.41	0.57
Std	7,835,467.39	21,105.52	2.00	0.08	0.77	14,740.79	1.53	0.26
Min	0.00	-86,154.00	0.00	-0.86	-44.05	0.00	0.00	0.06
5%	0.00	22.76	0.15	0.03	-0.17	19.78	0.89	0.22
25%	3,336.38	182.31	0.33	0.10	0.05	109.53	1.57	0.44
50%	49,658.00	634.37	0.51	0.14	0.11	428.28	2.13	0.57
75%	477,186.00	2,002.12	0.77	0.18	0.19	1,713.75	2.86	0.69
95%	9,731,456.22	13,665.32	1.39	0.28	0.38	14,184.30	4.68	0.90
Max	134,540,257.09	506,199.00	170.93	0.90	15.89	259,651.00	23.84	18.54
Observations	14,605	14,598	14,325	14,584	14,348	14,602	13,148	14,605
				Panel B: Correlations				
Emissions	1	0.21	0.01	-0.07	0.00	0.27	0.05	0.06
EQ		1	0.00	-0.04	0.01	0.82	0.04	-0.02
B/M			1	-0.13	-0.05	0.01	0.01	0.10
ROA				1	0.20	-0.06	0.03	-0.11
ROE					1	0.01	0.00	-0.36
TANT						1	0.13	0.05
OL							1	0.17
Lev								1

Table II Univariate Portfolio Sorting

	L	2	3	4	H	H-L
			Panel A : AT			
E[r]-rf (%)	30.08	25.08	39.39	28.38	37.95	6
[t]	6.64	4.5	2.73	5.29	3.21	0.61
Std (%)	4.81	5.99	15.1	5.7	12.42	12.92
SR	6.25	4.19	2.61	4.98	3.06	0.46
			Panel B : PPENT			
E[r]-rf (%)	30.93	19.04	50.69	72.19	36.59	4.1
[t]	7.18	3.06	2.75	1.6	3.41	0.44
Std (%)	4.57	6.83	19.11	46.28	11.22	13.17
SR	6.77	2.79	2.65	1.56	3.26	0.31
			Panel C : SALE			
E[r]-rf (%)	29.75	23.48	57.02	33.16	39.18	7.56
[t]	6.79	4.34	2.2	4.07	3.25	0.71
Std (%)	4.66	5.84	26.8	8.6	12.62	13.35
SR	6.38	4.02	2.13	3.86	3.1	0.57
			Panel D : EQ			
E[r]-rf (%)	28.91	21.66	36.97	50.77	47.18	16.4
[t]	5.23	4.04	4.3	2.78	3.67	1.25
Std (%)	5.88	5.82	9.04	18.88	13.37	14.62
SR	4.92	3.72	4.09	2.69	3.53	1.12

Table III Firm Characteristics

	L	2	3	4	H
Raw Emissions	1596.86	29278.95	188262.89	898682.12	8445205.33
Emissions	0.75	14.86	97.60	487.11	16001.38
EQ	3211.97	4374.86	3381.01	4075.80	2413.29
B/M	0.53	0.57	0.63	0.60	0.70
I/K	0.20	0.19	0.18	0.16	0.14
ROA	0.15	0.15	0.15	0.15	0.13
TANT	0.69	0.75	0.90	1.19	2.61
OL	2.18	1.97	2.23	2.67	2.81
LEV	0.52	0.53	0.54	0.55	0.61
Num (average)	11	11	11	13	11

Table IV Asset Pricing Factor Tests

	L	2	3	4	H	H-L
			Panel A: CAPM			
αCAPM	6.68	5.75	6.54	2.31	1.33	-0.29
[t]	1.16	1.29	1.20	2.11**	2.56**	-0.16
MKT	2.48	1.76	3.45	1.20	1.22	0.01
[t]	2.05**	1.80*	2.82***	5.08***	10.74***	0.03

	L	2	3	4	H	H-L		L	2	3	4	H	H-L
			Panel B: FF3							Panel C: FF4			
αFF3	6.47	5.70	6.16	2.32	1.27	-0.24	*αFF4*	*5.96*	5.86	6.19	2.38	1.50	-0.65
[t]	1.12	1.28	1.13	2.12**	2.43**	-0.13	*[t]*	1.03	1.31	1.13	2.16**	2.85***	-0.36
MKT	1.89	1.57	3.25	1.06	1.17	-0.13	*MKT*	*2.3*	1.47	3.21	1.01	1	0.15
[t]	1.47	1.54	2.55**	4.26***	9.88***	-0.33	*[t]*	1.68*	1.34	2.33**	3.78***	7.90***	0.34
SMB	3.62	0.98	0.80	0.63	0.20	0.60	*SMB*	*3.6*	1.01	0.80	0.63	0.27	0.51
[t]	1.62	0.61	0.39	1.51	1.08	1.03	*[t]*	1.61	0.63	0.40	1.51	1.51	0.86
HML	-1.19	0.22	2.05	0.42	0.51	-0.54	*HML*	*-0.76*	0.1	2.03	0.38	0.35	-0.29
[t]	-0.67	0.16	1.22	1.25	3.14***	-1.02	*[t]*	-0.41	0.07	1.18	1.10	2.05**	-0.53
							MOM	*1.13*	-0.28	-0.07	-0.12	-0.43	0.67
							[t]	0.87	-0.27	-0.06	-0.46	-3.51***	1.74*

	L	2	3	4	H	H-L		L	2	3	4	H	H-L
			Panel B: FF5							Panel C: HXZ			
αFF3	*5.51*	6.30	7.27	2.26	1.23	-0.63	*αFF4*	11.66	42.45	17.78	25.85	16.40	6.05
[t]	*0.92*	1.36	1.28	1.99*	2.25**	-0.33	*[t]*	0.94	2.87***	1.86*	2.22**	2.22**	0.50
MKT	*1.91*	1.32	3.20	1.04	1.20	0.00	*MKT*	0.83	0.52	0.71	0.94	0.68	0.85
[t]	*1.38*	1.20	2.32**	3.91***	9.31***	0.01	*[t]*	2.53**	1.23	2.52**	2.48**	2.30***	2.13**
SMB	*4.90*	1.06	-0.99	0.80	0.14	0.69	*SMB*	1.52	2.16	1.12	1.76	1.01	-0.73
[t]	*2.03***	0.59	-0.44	1.78*	0.70	1.03	*[t]*	1.83*	2.12	1.64	2.04**	1.85*	-0.78
HML	*-0.90*	0.90	1.06	0.35	0.37	-1.11	*I/A*	0.15	0.06	0.26	0.27	0.04	0.26
[t]	*-0.38*	0.50	0.47	0.80	1.74*	-1.50	*[t]*	1.95*	0.39	0.50	3.75***	0.80	4.31***
RMW	*4.57*	0.08	-4.38	0.29	-0.09	0.15	*ROE*	0.46	-0.58	0.15	-0.33	0.49	-1.46
[t]	*1.58*	0.04	-1.66	0.56	-0.39	0.19	*[t]*	0.62	-0.64	0.28	-0.45	1.06	-1.92*
CMA	-4.71	-2.32	4.66	-0.34	0.32	0.97							
[t]	-1.25	-0.82	1.33	-0.50	0.98	0.89							

*Note: * for 10% significant level, ** for 5% significant level, and *** for 1% significant level*

Table V Fama-MacBeth Regressions

	Summary1	Summary2
Log Emissions	-0.150	-0.389
[t]	-0.522	-2.899
log eq	-0.089	-0.160
[t]	-0.502	-1.451
Log B/M	-0.023	-0.154
[t]	-0.090	-0.997
I/K	0.104	0.114
[t]	0.528	0.804
ROE	-0.153	-0.095
[t]	-0.802	-0.720
TANT		-0.793
[t]		-3.210
Lev		0.146
[t]		1.099
Observations	14,605	14,606
R^2	0.95	0.98
Industry FE	Yes	Yes

References

[1] Po-Hsuan Hsu, Kai Li, Chi-Yang Tsou, The Pollution Premium, 10.1111/jofi.13217, 78, 3, (1343-1392), (2023).

[2] Xiong, XI; Png, Ivan, 2019, "Location of U.S. Manufacturing, 1987-2014: A New Dataset", https://doi.org/10.7910/DVN/K4KBBR, Harvard Dataverse, V2.

[3] Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, Journal of Financial Economics 116, 1–22.

[4] Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, Review of Financial Studies 28, 650–705.

[5] Fama-French Factors https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[6] Risk Free Rate: https://pages.stern.nyu.edu/~adamodar/New_Home_Page/data.html

[7] TRI: https://www.epa.gov/toxics-release-inventory-tri-program/tri-data-and-tools