Question 1 Report

1. Introduction

To explore the intricate relationship between companies' environmental efforts and their financial performance. The study conducts a detailed regression analysis using environmental metrics from 1499 companies, offering a comprehensive insight into how environmental responsibility impacts financial success.

Key to this research is the use of Thomson Reuters ESG Scores from Eikon, chosen for their comprehensive and transparent assessment of corporate environmental performance. These scores provide a detailed view of a company's environmental management and are recognized for their objectivity and global applicability. Additionally, the report integrates the London Stock Exchange Group ESG Scores Methodology; the environmental indicators are selected from Resource use, Emissions, Innovation aspects.

The report is divided into four sections. Section 2 shows methodology and the data description; section 3 shows the study results and section 4 presents the conclusion.

2. Methodology and Variables

In the first trial, we used ESG data from MSCI STATS. According to the ISO Certification ISO 37301, 10 environmental indicators were selected from positive and negative aspects. They are Pollution & Waste (Env-B, Env-C, Env-N), Climate Change (Env-D, Env-K, Env-P), Environmental Opportunities (Env-A), and three control factors (Env-con-D, Env-con-F, Env-con-J). Besides, financial performance was evaluated with the firm’s earnings per share (EPS). However, the coefficients of the regression results were close to zero, indicating that there was no strong correlation between the selected independent variables and the dependent variable.

Therefore, we opt to use ESG score based on Refinitiv to look for potential linear relationships or interactions between variables.

The study sample comprises annual data for all firms during the period 2014 to 2023. This selection resulted in 15040 observations derived from 1499 listed firms (S&P 1500). All the data used were collected from the Refinitiv database via EIKON.

The next step is data cleaning. We filled the spaces with the median and do group by based on date, ensuring that the data is properly formatted and preparing for model.

```
[ ] # Create filling function
    def missing_to_median(column):
        median_value = column.median(skipna=True)
        return column.fillna(median_value)

    ### Interpolate annual median into NAN
    dates = df_new['date']

    gkx_filled = df_new.groupby('date').transform(missing_to_median)
    print(gkx_filled.columns)

    ### merge original dates column into gkx_filled
    gkx_filled['date'] = dates
    print('Original data:')

    df_new.head() # before filling

    print('Interpolation data:')
    print(gkx_filled)
```

Then rank normalization[1] is used to scales features in a uniform manner, help to perform grouped data analysis:

```
[ ]
    ## create rank_norm function

    def rank_norm(column):
        rank = column.rank(method='dense')
        return (rank-1)/(np.nanmax(rank)-1)

    # Apply the function to df_ranked and store the new columns with '_rank' suffix:
    df_ranked = df_ranked.groupby('date')[columns_to_rank].transform(rank_norm)
    df_ranked = df_ranked.add_suffix('_rank')

    df_ranked.head()
```

For dependent variable, this study uses Return on Assets (ROA) as a measure of profitability. It shows how successfully the firm uses its resources to generate profit.

The independent variable was measured using three disclosure indicators from LSEG

---

[1] We cross-sectionally rank all stock characteristics period-by-period and map these ranks into the [-1,1] interval following Kelly, Pruitt, and Su (2019) and Freyberger, Neuhierl, and Weber (2020)

ESG Scores Methodology. From the environmental pillar, scores of Resource use, Emissions, Innovation were selected. Three types of control variables were utilized in this study: close price, total assets and EBITDA, as Table 1.

Table 1. Description of the study variables

Variables	Label	Description/Formula
Dependent variable		
Return on Assets	ROA	Net Income/Total Assets
Independent variables		
Resource Use Score	RU	Refinitiv score
Emissions Score	ES	Refinitiv score
Environmental Innovation Score	EI	Refinitiv score
Control variables		
Price Close	PC	Stock price after the market close
Total assets	TA	The total assets of the company
EBITDA	EBITDA	A measure of corporate profitability

3. Study model

The study uses OLS linear modelling approach and the strategy focuses on minimizing the Mean Squared Error (MSE). This method allows us to find the best-fitting line by minimizing the sum of squared differences between observed and predicted values.

To better measure environmental factors impact on financial performance, the study estimates 2 linear regression models. One excludes the constant term, while the other includes it. The equations as following:

Model 1 without constant term:

$$ROA_{it} = \beta_1 ES_{it} + \beta_2 EI_{it} + \beta_3 RU_{it} + \beta_4 PC_{it} + \beta_5 TA_{it} + \beta_6 EBITDA_{it} + \varepsilon_{it}$$

Model 2 with constant term:

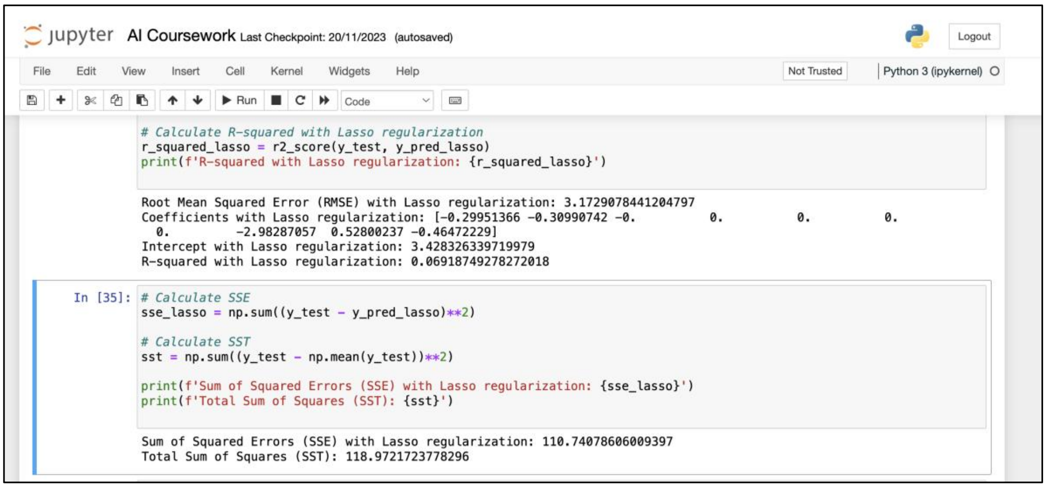
$$ROA_{it} = \beta_0 + \beta_1 ES_{it} + \beta_2 EI_{it} + \beta_3 RU_{it} + \beta_4 PC_{it} + \beta_5 TA_{it} + \beta_6 EBITDA_{it} + \varepsilon_{it}$$

where ROA_{it} are dependent variables, RU_{it} , ES_{it} , EI_{it} are independent variables, PC_{it}, TA_{it} , $EBITDA_{it}$ are control variables and ε_{it} is the error term for firm i in period t.

4. Empirical Results

4.1 Linear Regression Result with MSCI Dataset

Result from the linear regression using data from MSCI with EPS as a dependent variable can be seen below.



The R-squared value indicates that, with Lasso regularization, about 6.92% of the variance in the dependent variable is explained by the model. The RMSE is 3.172, indicating the average prediction error of the model.

The SSE is 110.740, representing the sum of the squared differences between the actual and predicted values. Considering the figure in conjunction with R-squared and RMSE, it indicates that the model may not be explaining the data well.

The Lasso regularisation has resulted in some coefficients being exactly zero, which means that certain features have been deemed less important and effectively excluded from the model. The RMSE is also comparable to the non-regularized model, suggesting that the regularization did not significantly impact the average prediction error. Therefore, we opted to do more experiments with the datasets from Refinitiv which is derived from EIKON.

4.2 Linear Regression Result with Refinitiv Dataset

This section includes Pearson correlation matrix, test results and regression results with Refinitiv Dataset.

- Correlation analysis

Table 2 provides the correlation matrix for each independent variable. As the table shows, there is no correlation between these scores and firm-level control variables.

Variables	ES	EI	RU	PC	TA	EBITDA
ES	1					
EI	0.472	1				
RU	0.826	0.486	1			
PC	0.209	0.154	0.205	1		
TA	0.476	0.297	0.434	0.249	1	
EBITDA	0.167	0.103	0.156	0.131	0.274	1

Table 2. Pearson correlation matrix

- Model evaluation and selection

After rank normalization and standardization, the dataset is methodically split into training and testing sets, following an 80/20 ratio.

To evaluate two linear regression models, R-square and the SSE are employed. A higher R-square value suggests a more substantial explanatory capacity of the model, indicating a closer fit to the observed data. SSE provides a quantitative measure of the model's accuracy, reflecting the aggregate squared deviation of the predicted values from the actual observed values.

From the table 3, the comparison can be found. Model 1 demonstrates superior performance in the training set with a high R-squared value of 0.782, suggesting it explains a significant portion of the variance. However, its performance drops in the testing set, indicating potential overfitting to the training data.

In contrast, Model 2 shows lower performance in training with an R-squared of 0.337 but maintains almost the same level of performance in the testing phase. This consistency suggests better generalization capabilities, despite its overall lower accuracy. Both models yield identical Sum of Squares Error (SSE) in the testing phase, indicating similar predictive error rates when applied to new data.

Model	**Data Set**	**R-squared**	**SSE**
Model 1	Training	0.782	533.276
	Testing	0.374	67.615
Model 2	Training	0.337	351.463
	Testing	0.373	67.615

Table 3. Model Fit Comparison

Table 4 and 5 presents the OLS regression results show that both models show significant positive relationships for EBITDA ranking and Price Close ranking, but significance and direction of environmental variables such as Emissions Score and Environmental Innovation Score are different from each other.

OLS Regression Results

Dep. Variable:	Return On Assets - Actual In the last 10 FY_rank	R-squared (uncentered):	0.782
Model:	OLS	Adj. R-squared (uncentered):	0.782
Method:	Least Squares	F-statistic:	7192.
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	0.00
Time:	18:36:19	Log-Likelihood:	1674.9
No. Observations:	12032	AIC:	-3338.
Df Residuals:	12026	BIC:	-3293.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
EBITDA (USD) In the last 10 FY_rank	0.4534	0.012	37.576	0.000	0.430	0.477
Emissions Score In the last 10 FY_rank	-0.0280	0.012	-2.379	0.017	-0.051	-0.005
Environmental Innovation Score In the last 10 FY_rank	-0.0190	0.008	-2.305	0.021	-0.035	-0.003
Price Close (USD) In the last 10 FY_rank	0.4393	0.007	66.585	0.000	0.426	0.452
Resource Use Score In the last 10 FY_rank	0.1269	0.012	10.912	0.000	0.104	0.150
Total Assets - Actual (USD) In the last 10 FY_rank	-0.2427	0.011	-22.299	0.000	-0.264	-0.221

Omnibus:	99.424	Durbin-Watson:	1.846
Prob(Omnibus):	0.000	Jarque-Bera (JB):	109.642
Skew:	0.184	Prob(JB):	1.55e-24
Kurtosis:	3.288	Cond. No.	9.17

Table 4. OLS Regression results of Model 1

OLS Regression Results

Dep. Variable:	Return On Assets - Actual In the last 10 FY_rank	R-squared:	0.337
Model:	OLS	Adj. R-squared:	0.337
Method:	Least Squares	F-statistic:	1272.
Date:	Tue, 12 Dec 2023	Prob (F-statistic):	0.00
Time:	15:47:09	Log-Likelihood:	6907.0
No. Observations:	15040	AIC:	-1.380e+04
Df Residuals:	15033	BIC:	-1.375e+04
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.3590	0.003	114.632	0.000	0.353	0.365
EBITDA (USD) In the last 10 FY_rank	0.4443	0.008	57.081	0.000	0.429	0.460
Emissions Score In the last 10 FY_rank	0.0015	0.008	0.196	0.844	-0.013	0.016
Environmental Innovation Score In the last 10 FY_rank	-0.0108	0.005	-2.022	0.043	-0.021	-0.000
Price Close (USD) In the last 10 FY_rank	0.1562	0.005	31.455	0.000	0.146	0.166
Resource Use Score In the last 10 FY_rank	0.0518	0.008	6.812	0.000	0.037	0.067
Total Assets - Actual (USD) In the last 10 FY_rank	-0.5312	0.007	-70.932	0.000	-0.546	-0.517

Omnibus:	1086.233	Durbin-Watson:	1.945
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1457.344
Skew:	0.636	Prob(JB):	0.00
Kurtosis:	3.842	Cond. No.	12.2

Table 5. OLS Regression results of Model 2

For better interpretation, we implement partial regression, which aims to research the relationship between one independent variable and the dependent variable independently, while control and isolate the effects of other independent variables.

The plots are shown below, the left figure is model 1 without constant term while the right one is model 2.

From Figure 1, both models show a negative coefficient for the Environmental Innovation Score, indicating a negative relationship with the dependent variable. Model 1 has a larger magnitude coefficient and a smaller p-value, suggesting a stronger and more significant negative impact compared to Model 2.

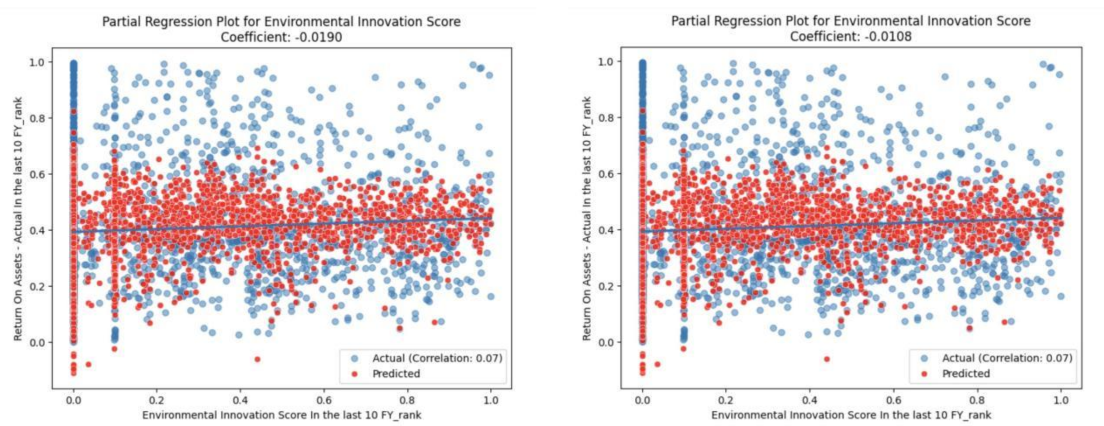


Figure 1. Partial regression results of Environmental Innovation Score (EI)

Figure 2 demonstrates the influence of variable ES. Model 1 visualizes a significant negative relationship, while in Model 2 this relationship is positive and not significant (p-value much higher than 0.05). The contrasting signs of the coefficients in the two models indicate a different interpretation of the Emission Score's impact on the ROA.

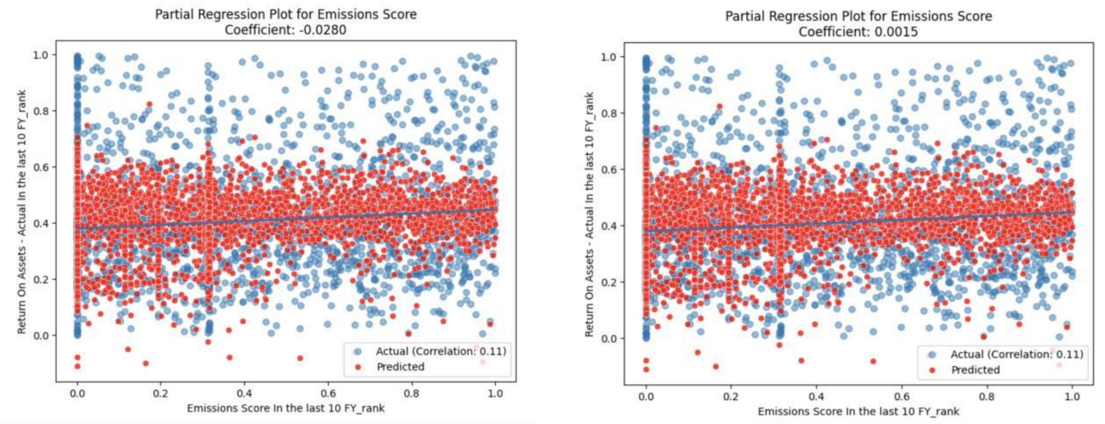


Figure 2. Partial regression results of Emission Score (ES)

The Figure 3 shows both models have a significant positive relationship between Resource Score and the ROA. However, the magnitude of the coefficient in Model 1 is larger, indicating a stronger positive impact on the financial performance.

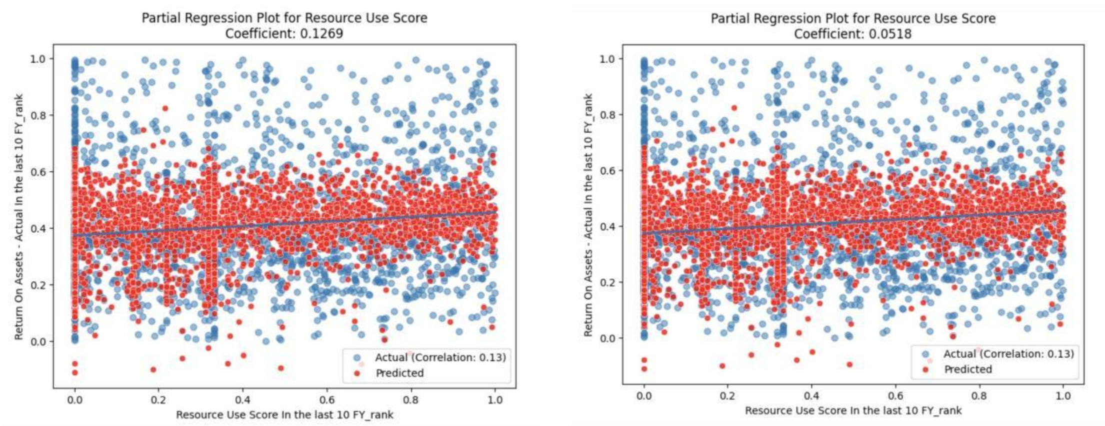


Figure 3. Partial regression results of Resource Score (RU)

5. Conclusion

This study investigates the relationship between companies' environmental efforts and performance. Analyzing pooled data from the Eikon database for the period 2014-2023, the study encompasses 1499 firms over ten years, yielding 15040 observations.

Based on the regression results from the two models, we can see that the addition of constant term has almost no effect on the interpretation. But after comparing the R-square and SSE indicators, we consider model 2 with less SSE is more convinced. Resource Use (RU) efficiency appears consistently beneficial across both models, indicating a clear positive link with profitability. However, Environmental Innovation (EI) and Emissions Score (ES) show more complex relationships. While they may have short-term costs, their long-term benefits impact on profitability might require further investigation.

In addition, the significant positive coefficient of EBITDA and Price Close (PC) enhanced its influence on profitability, while the negative coefficient of total assets indicated a potential inverse relationship between firm size and efficiency. The reason probably derives from decreasing marginal returns of large scale and high cost.

Finally, we stress the limitation of our study. As we only select 6 environmental factors, it produced an undesirable R-square value and limited insights on the research. It is suggested that further research could explore how other factors are affecting the financial performance.