



The Correlation Between Environmental, Social and Governance Disclosure Scores and Company's Financial Performance: A Case Study Based on Regression Model

Yizhe Jing^(✉)

School of Management, The University of Bristol, Bristol BS8 1QU, UK
yizhe_jing@163.com

Abstract. With the rapid development of machine learning and big data techniques, modelling and predictive analysis have been applied broadly in the financial area. From stock market prediction [11] to decision making support for financial institutions [9], researchers have shown that machine learning and big data techniques can be used in practical applications and make invaluable contributions. As sustainable development has become a worldwide consensus, the author is trying to justify whether the ESG disclosure score will affect a investment banking company's financial performance. The author used the data from S&P 1500, and selected the company's total assets, the Tobin's Q ratio and the return on common equity as the indicators of a company's financial performance. The author starts by building full regression models and using backward elimination to improve the model performance. The threshold of the F-test was set to be 0.05, and the results suggested that the ESG disclosure score has a significant non-linear correlation with the company's financial performance.

Keywords: EGS Disclosure Score · Regression Model · Financial Performance · Business Analysis · Investment Bank

1 Introduction

Can people predict what factors will make your business success, based on the rapid development of machine learning techniques and analytical tools in the past decade? There has already been an enormous amount of research exploring the correlation between environmental, social, and governance disclosure scores (ESG disclosure scores) with the financial performance of companies. Benlemlih's research has demonstrated that these disclosures scores negatively and significantly affect a company's total and idiosyncratic risk [2]. Enhancing corporate transparency and releasing appropriate and accurate environmental, social, and governance disclosures, according to Benlemlih, can strengthen a company's reputation and faith with the market and stockholders, as well as lower idiosyncratic and operational risk. A recent study [10] reported similar outcomes. Oncioiu's research looked at 143 organisations in the financial, energy, and

telecommunications services sectors from eight different nations. It demonstrates how environmental, social, and governance disclosure has become a vital tool for business governance and reputation management. Another argument raised in this study is that the focus of ESG disclosures can vary dramatically from one business to another. This viewpoint explains the phenomenon that happened in this work. When we create models for a specific sector, the variables that influence financial performance can be different from those in more general industry models. The author used Goldman Sachs, a well-known American investment bank and financial services firm, as an example in this paper to examine what factors influence an investment bank's financial performance.

2 Research Method

2.1 Data Cleaning

We are all in agreement that data cleaning is a crucial stage before beginning modelling. To maintain the quality of our training data, the row data must be cleansed, selected, and even sometimes altered. In that process, dealing with missing data is a frequent problem. Based on the cause of the missing value, Kang has categorised it into three types [7]. Value missing totally at random, value missing at random, and value missing not at random are the three types. Different approaches should be considered to address different kinds of missing values. Any missing value leads to bias and interruptions in analysis work, and numerous studies and experiments have been conducted to reduce the negative effects of missing values in various circumstances. The Bayesian network, mean value substitution, maximum likelihood estimation, and regression prediction substitution are all considered to be effective in different scenarios [1, 7]. The Bayesian network, for example, can deal with the missing value's dependence with other variables, at the same time, maintain the joint distribution of all variables. However, if the missing data is missed completely at random, the Bayesian method may not be accurate; or if the dataset is large, the method could require plenty of time to do the calculations.

Because the nature of missing data is uncertain in this example and the data set is considered to be large enough, the author will handle the missing value using the case deletion method [7]. As case deletion is not going to generate additional bias in the given data or complicate the research methodology, the author employs this straightforward methodology (Figs. 1 and 2).

2.2 Quick Correlation Check for Variables

The heat map given below shows that some variables appear to be related. The numeric values for the correlation coefficients between variables are provided by the correlation matrix at the same time. However, not all relationships are of interest to this research. This paper will primarily focus on the correlation between ESG score and the financial performance indicator (Fig. 3).

The author did not discover any evident linear correlations between the factors we care about. Only when the heat map is combined with the scatter plots matrix, however, we can observe that there are non-clear linear correlations between total assets and the ESG disclosure score. The author will next go into better detail on the factors that can affect a company's financial success (Fig. 4).

Factors Names	Count Not Null
NAME	4518
EQY_SIC_COD	4437
BS_TOT_ASSET	4452
ENVIRON_DISCLOSURE_SCORE	3179
SOCIAL_DISCLOSURE_SCORE	4168
GOVNCE_DISCLOSURE_SCORE	4371
TOBIN_Q_RATIO	4362
RETURN_COM_EQY	4232

Fig. 1. Variable selection and Null Value before Data Cleaning (photo credit: Original)

Factors Names	Count Not Null
NAME	3016
EQY_SIC_COD	3016
BS_TOT_ASSET	3016
ENVIRON_DISCLOSURE_SCORE	3016
SOCIAL_DISCLOSURE_SCORE	3016
GOVNCE_DISCLOSURE_SCORE	3016
TOBIN_Q_RATIO	3016
RETURN_COM_EQY	3016

Fig. 2. Variable Name and Null Value after Data cleaning (photo credit: Original)

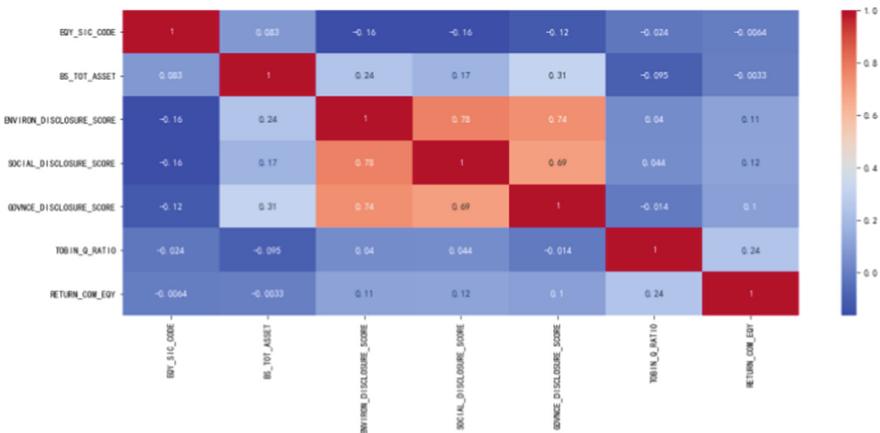


Fig. 3. The Variable Correlation Matrix Heat Map (photo credit: Original)

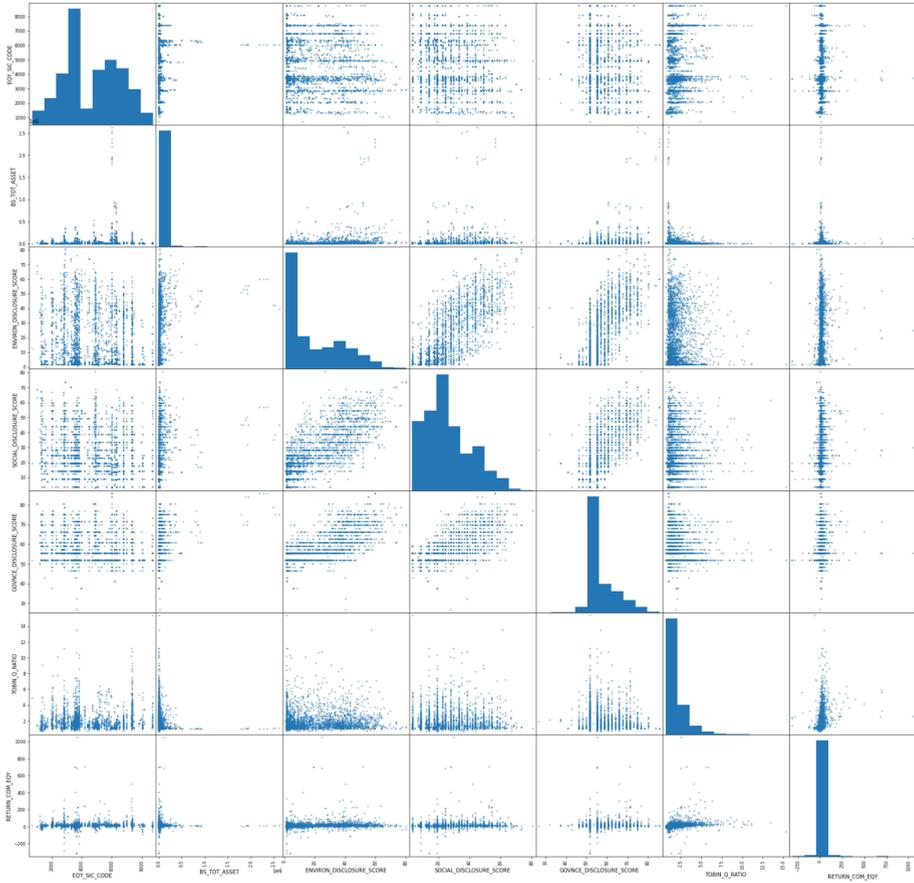


Fig. 4. The Variable Scatter Plots Matrix (photo credit: Original)

3 Factors Affecting a Company Financial Performance

The author used Goldman Sachs (GS), an American global investment bank and financial services enterprise, as an example in this study to examine the variables that impact a company's financial performance in that field. The author will utilise the SIC code to locate those organisation that is in the same sector as GS. Security Brokers, Dealers, and Flotation Companies are represented by the SIC code 6211 [13]. Security and Commodity Brokers, Dealers, Exchanges, and Services are classified as industry code 62 [14]. And the author will examine the situation from the most specific to the broader industry.

To begin, we review some of the GM's and the overall Security Brokers, Dealers, and Flotation industry's basic statistics (Fig. 5).

	GM	Security Brokers, Dealers, and Flotation Companies
Average total assets	902912.33	289282.24
Average environmental disclosure score	51.79	15.89
Average social disclosure score	20.0	15.21
Average governance disclosure score	60.71	55.28
Average Tobin's Q ratio	1.01	1.09
Average return on common equity	9.02	9.65

Fig. 5. Basic Statistics of the GM and the General Industry (photo credit: Original)

```

=====
                        OLS Regression Results
=====
Dep. Variable:          BS_TOT_ASSET      R-squared (uncentered):      0.890
Model:                 OLS              Adj. R-squared (uncentered):  0.885
Method:                Least Squares    F-statistic:                  385.4
Date:                  Mon, 28 Feb 2022   Prob (F-statistic):          1.71e-12
Time:                  10:02:57         Log-Likelihood:               -320.12
No. Observations:     24              AIC:                          642.2
Df Residuals:         23              BIC:                          643.4
Df Model:              1
Covariance Type:      nonrobust
=====
                                coef      std err          t      P>|t|      [0.025      0.975]
-----+-----
ENVIRON_DISCLOSURE_SCORE:GOVNCE_DISCLOSURE_SCORE  280.2929    20.586     13.616     0.000     237.707     322.879
=====
Omnibus:                16.783    Durbin-Watson:              1.555
Prob(Omnibus):          0.000    Jarque-Bera (JB):           31.541
Skew:                   1.105    Prob(JB):                    1.42e-07
Kurtosis:               8.163    Cond. No.                     1.00
=====
Notes:
[1] R2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Fig. 6. The Final Model of Total Assets (photo credit: Original)

3.1 Factors Affecting a Company’s Total Assets

We can observe that GS’s average total assets are significantly higher than the industry average by comparing its statistics to those of its industry. Furthermore, the average environmental disclosure score is also much higher than the industry norm. It’s reasonable to assume that the environmental disclosure score has a favourable positive impact on the company’s overall assets. To justify the assumption, the author began by creating the entire model and then optimised it using backward elimination (Fig. 6).

Our model supports the hypothesis that an environmental disclosure score has a favourable impact on a company’s total assets, as shown in the model summary. The effect, however, is not a simple linear relationship; it exists as an interacting term with the governance disclosure score together. When the two elements are assessed independently in the backward elimination process, the environmental or governance disclosure scores have no effect on the company’s total assets.

3.2 Factor Affecting a Company’s Tobin’s Q Ratio

In addition, the GS’s average Tobin’s Q ratio is slightly lower than the industry average, as we saw in the former table. That’s an unusual occurrence, given that Tobin’s Q ratio is an important indicator of a company’s profitability. Tobin’s Q ratio is defined as the market worth of a company divided by the assets’ replacement cost, and Nobel Laureate

OLS Regression Results						
Dep. Variable:	log_ratio	R-squared (uncentered):	0.637			
Model:	OLS	Adj. R-squared (uncentered):	0.604			
Method:	Least Squares	F-statistic:	19.29			
Date:	Mon, 28 Feb 2022	Prob (F-statistic):	1.45e-05			
Time:	10:03:50	Log-Likelihood:	-30.415			
No. Observations:	24	AIC:	-56.83			
Df Residuals:	22	BIC:	-54.47			
Df Model:	2					
Covariance Type:	nonrobust					

	coef	std err	t	P> t	[0.025	0.975]
ENVIRON_DISCLOSURE_SCORE:SOCIAL_DISCLOSURE_SCORE	-0.0003	5.92e-05	-4.369	0.000	-0.000	-0.000
SOCIAL_DISCLOSURE_SCORE:GOVNCCE_DISCLOSURE_SCORE	0.0002	3.31e-05	5.767	0.000	0.000	0.000

Omnibus:	0.865	Durbin-Watson:	1.395			
Prob(Omnibus):	0.649	Jarque-Bera (JB):	0.874			
Skew:	-0.366	Prob(JB):	0.646			
Kurtosis:	2.419	Cond. No.	5.72			

Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Fig. 7. The Final Model of Tobin's Q ratio (photo credit: Original)

in Economics James Tobin thought that this ratio should be close to one. Companies having a Tobin's Q ratio larger than one were deemed overvalued, and investors were encouraged to buy [6, 8]. According to Fu's research, a high Tobin's Q ratio indicates that a company's future performance will be better [5]. Because Tobin's Q ratio is such a key indication, the author believes it is highly appealing to think about: What makes the company overvalued? Or, to put it another way, what makes a company valuable in the eyes of the market?

We're applying the same technology we used in the prior sector, starting with the whole model and applying backward elimination to eliminate the extraneous variables. To make the value difference more significant, the author applied a log transformation to Tobin's Q ratio. Any factor with a P-value below 0.05 is thought to be related to Tobin's Q ratio (Fig. 7).

We can see from the summary of the final model that the three components have no effect on Tobin's Q ratio individually. The effects are only considered to be significant when they act interactively. When the environmental and social disclosure scores are combined together, Tobin's Q ratio is affected negatively. When the social and governance disclosure scores are taken together, however, the interactive term will show a positive correlation with Tobin's Q ratio. According to the author's final regression model, if companies in the relevant industry want to improve their Tobin's Q ratio, they should improve their governance disclosure score, attempting to reduce their environmental disclosure score while keeping an appropriate degree of social disclosure score. The author suggests that this phenomenon might be due to the fact that evidence shows a company in this industry is under proper governance will increase investors' confidence.

3.3 Factors Affecting a Company's Return on Common Equity

The return on common equity is a metric that gauges a company's profitability and is derived by dividing net income by shareholder equity [4]. All relevant variables failed to pass the F test by applying the F test, which means that the ESG disclosure score will not affect the return on common equity.

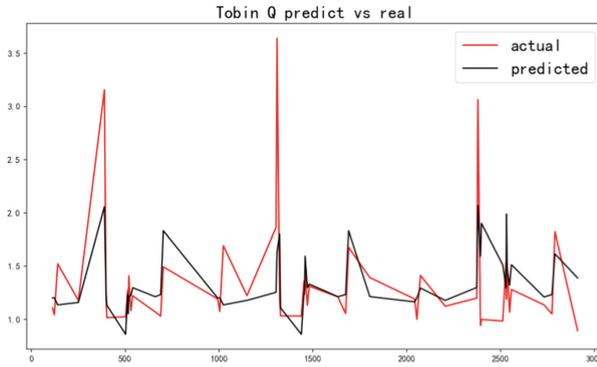


Fig. 8. The Predict Value versus Actual Value of Tobin’s Q Ratio (photo credit: Original)

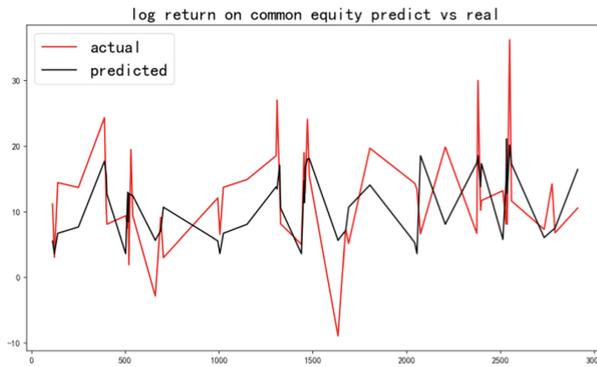


Fig. 9. The Predict Value versus Actual Value of Return on Common Equity (photo credit: Original)

4 The More General Industry

Furthermore, the elements that have a major impact on the return on common equity are assessed in this broader context. The company’s return on common equity is adversely connected with the environmental and social disclosure score interactive term, whereas the social, governance disclosure score interactive time has a positive association. The actual value versus projected value plot and the R square value both demonstrated that the model suited the data well (Figs. 8 and 9).

5 Conclusion

By selecting Goldman Sachs as an example, the author compared GS’S financial performance indicator and factors that may affect the financial performance with the Security Brokers, Dealers, and Flotation industry average. Although GS’s total asset is much larger than the industry average, Tobin’s Q ratio and return on common equity are marginally lower than the industry average. Initially, the author constructs the entire model with

every possible explanatory variable that might influence total asset, Tobin's Q ratio and return on common equity. The author then set the threshold to 0.05 and performed an f-test starting with the variable with the greatest p-value. A P-value greater than 0.05 indicates that the variable would be extremely unlikely to influence the model's conclusion. Eliminating the variables with the highest p-value one at a time, and repeating this process until all of the variables remaining have a significant effect on the model. According to the model, no single ESG disclosure score will have an impact on the company's total assets, Tobin's Q ratio, or return on common equity. As the interaction terms, however, significant impacts those indicators.

To summarise what we've discussed, companies in the relevant industry should focus on strengthening social and governance disclosures and limiting environmental disclosure in the future if they want to improve their financial performance.

In this study, there are two major limitations. The first is that to deal with the missing value, the study simply employed the case deletion procedure. The method is straightforward and efficient; however, if there is a probability of success identifying the reasons for the missing data, we would be able to utilise other ways to improve the data richness, such as the Bayesian network.

The study's second flaw is that it simply demonstrates correlation. Numerous studies [3, 12] show, on the other hand, that statistical correlation does not imply causality. The study can only show correlations between variables, but it cannot prove causality. We don't know if ESG disclosures scores affect financial performance or if there are other causal links waiting to be investigated.

References

1. Abidin, N.Z., Ismail, A.R., Emran, N.A., Performance analysis of machine learning algorithms for missing value imputation. *International Journal of Advanced Computer Science and Applications* 9, 442–447. 2018. DOI: <https://doi.org/10.14569/IJACSA.2018.090660>
2. Benlemlih, M., Qiu, Y., ... Court, S., Environmental and Social Disclosures and Firm Financial Risk: Evidence from UK Environmental and Social Disclosures and Firm Financial Risk : Evidence from UK. *Journal of Business Ethics* 0–41. 2016.
3. Cohen, P., Correlation and causality. *Evaluation and Program Planning* 3, 59–60. 1980. DOI: [https://doi.org/10.1016/0149-7189\(80\)90009-9](https://doi.org/10.1016/0149-7189(80)90009-9)
4. FERNANDO, J., How Return on Equity (ROE) Works. [online] Investopedia. 2021. Available at: <https://www.investopedia.com/terms/r/returnonequity.asp> [Accessed 12 January 2022].
5. Fu, L., Parkash, M., Singhal, R., Tobin's q Ratio and Firm Performance. *International Research Journal of Applied Finance*. 2017. DOI: <https://doi.org/10.0704/article-2>
6. HAYES, A., How the Q Ratio – Tobin's Q Works. [online] Investopedia. 2021. Available at: <https://www.investopedia.com/terms/q/qratio.asp> [Accessed 12 January 2022].
7. Kang, H., The prevention and handling of the missing data. *Korean Journal of Anesthesiology*. 2013. DOI: <https://doi.org/10.4097/kjae.2013.64.5.402>
8. Lindenberg, E.B., Ross, S.A., Tobin's q Ratio and Industrial Organization. *The Journal of Business* 54, 1. 1981. DOI: <https://doi.org/10.1086/296120>
9. Mezera, F., Křupka, J., Selected business intelligence methods for decision-making support in a finance institution. *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration* 24, 154–164. 2017.

10. Oncioiu, I., Popescu, D.M., ... Marin-Pantelescu, A., The role of environmental, social, and governance disclosure in financial transparency. Sustainability (Switzerland) 12. 2020. DOI: <https://doi.org/10.3390/SU12176757>
11. Ouahilal, M., Mohajir, M.El, ... Mohajir, B.E. El., A comparative study of predictive algorithms for business analytics and decision support systems: Finance as a case study, in: 2016 International Conference on Information Technology for Organizations Development, IT4OD 2016. Institute of Electrical and Electronics Engineers Inc., 2016. DOI: <https://doi.org/10.1109/IT4OD.2016.7479258>
12. Roy-García, I., Rivas-Ruiz, R., ... Palacios-Cruz, L., Correlation: Not all correlation entails causality. Revista Alergia Mexico 66, 354–360. 2019. DOI: <https://doi.org/10.29262/ram.v66i3.651>
13. SIC & NAICS Codes, Company Search, Business Lists - SICCODE.com. 2022. SIC Code 6211 Security Brokers, Dealers, and Flotation.... [online] Available at: <<https://siccode.com/sic-code/6211/security-brokers-dealers-flotation-companies#:~:text=Snapshot,SIC%20Code%206211%20%2D%20Security%20Brokers%2C%20Dealers%2C%20and%20Flotation%20Companies,estimated%20employment%20of%20410%2C552%20people>> [Accessed 18 February 2022].
14. SIC & NAICS Codes, Company Search, Business Lists - SICCODE.com. 2022. SIC Code 62 - Security and commodity brokers. [online] Available at: <<https://siccode.com/sic-code/62/security-commodity-brokers-dealers-exchanges-services#:~:text=Related%20Code%20Systems,SIC%20Code%2062%20%2D%20Security%20and%20Commodity,%2C%20Dealers%2C%20Exchanges%2C%20and%20Services>> [Accessed 19 February 2022].

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

