

ESG Risk and the Cost of Debt: Evidence from NIFTY50 Firms- Ankit Ram Samant

1. Introduction

Environmental, social, and governance (ESG) considerations have become central to how capital markets evaluate firms. A recurring claim in sustainable finance is that firms with weaker ESG performance face **higher financing costs**, including a higher **cost of debt**. In practice, however, the strength of this relationship can depend heavily on the market, sample, and measurement approach.

This project focuses on large Indian firms and asks:

Do NIFTY50 firms with higher ESG risk pay a higher cost of debt, after controlling for basic financial characteristics?

To answer this, I:

- Combine **NIFTY50 ESG risk scores (2024)** with detailed **NSE/BSE financial statement data**.
- Construct a **firm-level dataset** with ESG risk, cost of debt, profitability, leverage, and size.
- Estimate a **baseline OLS regression** to quantify the relationship.
- Benchmark several **predictive models** (Ridge, Lasso, Random Forest, XGBoost) to see whether more complex methods can better explain cost of debt.
- Build an **interactive Power BI dashboard** to visualize the patterns across ESG bands and sectors.

The final result is both a quantitative analysis and a reusable dataset + dashboard that can be extended in future work.

2. Data

2.1 ESG Data (NIFTY50)

The ESG data set covers companies in the **NIFTY50 index** and includes the following key variables:

- **nse** – NSE ticker symbol
- **company** – company name
- **Sector, Industry** – sector classification
- **esg_risk_score** – overall ESG risk score (higher values indicate worse ESG performance / higher risk)
- Additional context fields:
 - **esg_risk_exposure**
 - **esg_risk_management**
 - **esg_risk_level**
 - **controversy_score**

These ESG scores form the main independent variable of interest.

2.2 Financial Data (NSE/BSE)

The financials come from a comprehensive data pack for **4,456 NSE and BSE companies**, which includes multiple CSV files per company:

- ***_Basic_Info.csv**
 - Contains **NSE** ticker and **Company_name**.
- **Yearly_Balance_Sheet.csv**
 - Contains **Borrowings**, **Total Assets**, **Total Liabilities** across multiple years.
- **Yearly_Profit_Loss.csv**

- Contains `Interest`, `Net Profit` and other P&L items across multiple years.

For each **NIFTY50** firm:

1. I match the ESG ticker `nse` to the NSE ticker in `*_Basic_Info.csv`.
2. From `Yearly_Balance_Sheet.csv`, I select the **latest year** with non-missing `Total Assets`.
3. From `Yearly_Profit_Loss.csv`, I select the **latest non-TTM year** with non-missing `Net Profit`; if none, I fall back to `TTM`.

This yields a **single financial snapshot per firm**.

2.3 Final Dataset

After merging ESG and financials and dropping one firm with missing `cost_of_debt`, the final dataset contains:

- **49 NIFTY50 firms**
- One row per firm
- Core columns:
 - `nse`, `company`, `Sector`, `Industry`
 - `esg_risk_score`
 - `cost_of_debt`
 - `roa`
 - `leverage`
 - `size`

This cleaned dataset is saved as:

`nifty50_esg_cost_of_debt.csv` (in the `data_processed` folder of the repo)

and is the main input for both the regression and the Power BI dashboard.

3. Methodology

3.1 Variable Construction

Key variables are defined as follows:

- **Cost of debt**

$\text{cost_of_debt} = \frac{\text{Interest}}{\text{Borrowings}}$

This is interpreted as an approximate **interest rate on outstanding debt**. For example, a value of **0.08** corresponds to roughly an **8% cost of debt**.

- **Profitability (ROA)**

$\text{roa} = \frac{\text{Net Profit}}{\text{Total Assets}}$

- **Leverage**

$\text{leverage} = \frac{\text{Borrowings}}{\text{Total Assets}}$

- **Size**

$\text{size} = \log(\text{Total Assets})$

- **ESG Risk Band (for visualization)**

Based on `esg_risk_score`, firms are bucketed into three bands:

- Low risk

- Medium risk
- High risk

using fixed thresholds or quantiles. This banding improves interpretability in the dashboard.

3.2 Baseline OLS Model

To formally test the relationship between ESG risk and cost of debt, I estimate the following **linear regression**:

Regression model

$$\text{cost_of_debt}_i = \beta_0 + \beta_1 \text{esg_risk_score}_i + \beta_2 \text{roa}_i + \beta_3 \text{leverage}_i + \beta_4 \text{size}_i + \varepsilon_i$$

Where:

- cost_of_debt_i = cost of debt for firm i (interest expense \div borrowings)
- esg_risk_score_i = ESG risk score for firm i (higher = worse ESG)
- roa_i = return on assets for firm i (net profit \div total assets)
- leverage_i = leverage ratio for firm i (borrowings \div total assets)
- size_i = firm size for firm i (log of total assets)
- β_0 = intercept term
- $\beta_1, \beta_2, \beta_3, \beta_4$ = slope coefficients to be estimated
- ε_i = error term capturing unobserved factors for firm i

3.3 Predictive Models (Ridge, Lasso, Random Forest, XGBoost)

In addition to OLS, I benchmark several **predictive models** to see how well ESG + basic financials can predict cost of debt out-of-sample:

- **Ridge Regression** (L2-regularized linear model)
- **Lasso Regression** (L1-regularized, performs feature shrinkage)
- **Random Forest Regressor** (ensemble of decision trees)
- **XGBoost Regressor** (gradient-boosted trees)

Setup:

- Features: `esg_risk_score`, `roa`, `leverage`, `size`
- Target: `cost_of_debt`
- Train-test split: **80% train / 20% test**, `random_state = 42`
- Evaluation metrics:
 - **R²** on the test set
 - **RMSE** (root mean squared error) on the test set

The goal is not to build a production-grade forecasting model, but to see whether more flexible models uncover stronger predictive signals than OLS.

4. Results

4.1 Descriptive Statistics (High-Level)

For the final sample of 49 firms, the key variables have approximately:

- **Cost of debt**
 - Mean ≈ 0.105
 - Range ≈ 0.00 to 0.54
 - Distribution is **right-skewed** with a few extreme high values.
- **ESG risk score**
 - Mean ≈ 26
 - Range ≈ 11.4 to 52.2
 - Firms span low, medium, and high ESG risk levels.
- **ROA, leverage, size**

- Show substantial variation across firms and sectors.

These preliminary statistics already hint at a small sample with heterogeneous firms and a few notable outliers.

4.2 OLS Regression Results

The baseline OLS model yields:

- **R-squared ≈ 0.256**
- **Adjusted R-squared ≈ 0.188**
- **F-statistic p-value ≈ 0.0099**

This means:

- The model explains about **25.6%** of the variation in cost of debt across the 49 firms.
- As a group, the regressors (`esg_risk_score`, `roa`, `leverage`, `size`) are **jointly significant**.

Key coefficients (approximate):

- **ESG risk score**
 - Coefficient $\approx +0.0011$
 - p-value ≈ 0.49
 - Interpretation: directionally positive (higher ESG risk \rightarrow slightly higher cost of debt), but **statistically insignificant**. We cannot confidently claim a real effect in this sample.
- **ROA**
 - Coefficient $\approx +0.39$
 - p-value ≈ 0.093 (marginally significant at the 10% level)

- Interpretation: more profitable firms may have somewhat higher observed cost of debt, but this result is not very strong and may be influenced by sectoral composition or outliers.
- **Leverage**
 - Coefficient slightly negative; **not statistically significant**.
- **Size**
 - Coefficient slightly negative; **not statistically significant**.

Summary of OLS:

The model has moderate explanatory power overall, but the **ESG risk score is not a strong, statistically significant predictor of cost of debt** after controlling for basic financial metrics. The direction of the ESG effect is consistent with the “higher risk → higher cost” narrative, but the magnitude and significance are weak.

4.3 Predictive Model Performance

On an 80/20 train–test split, the predictive models achieve the following **test-set** performance:

- **Ridge Regression**
 - $R^2 \approx -0.22$
 - $RMSE \approx 0.0598$
- **Lasso Regression**
 - $R^2 \approx -0.25$
 - $RMSE \approx 0.0605$
 - Lasso shrinks some coefficients toward zero, indicating limited signal.
- **Random Forest**
 - $R^2 \approx -3.94$
 - $RMSE \approx 0.1202$
- **XGBoost**

- $R^2 \approx -4.90$
- $RMSE \approx 0.1313$

Interpretation:

- A model with $R^2 = 0$ would be as good as just predicting the **mean** cost of debt on the test set.
- **Negative R^2** means the model is **worse than predicting the mean**.

Therefore:

- Ridge and Lasso are the “**least bad**” models, but still yield **negative R^2** , meaning they do not reliably beat a simple baseline.
- Random Forest and XGBoost perform **much worse**, which is a classic sign of **overfitting** on such a small dataset (≈ 39 training points).

Predictive takeaway:

Even with more complex machine learning models, **ESG risk plus a handful of financial ratios does not provide strong predictive power for cost of debt** in this sample. The signal is weak enough that sophisticated models overfit and fail to generalize.

4.4 Dashboard and Visual Insights (Power BI)

The Power BI dashboard built for this project includes:

- **KPI cards**
 - Average cost of debt for the sample
 - Average ESG risk score
 - Number of firms
- **Firm-level table**
 - `nse, company, Sector, esg_risk_score, cost_of_debt, roa, leverage, size`

- **Scatter plot: ESG risk vs cost of debt**
 - Each point is a firm; a trendline is added.
 - The trendline is **slightly upward but very flat**, indicating only a weak relationship, visually consistent with the OLS coefficient.
- **Column chart: Average cost of debt by ESG risk band**
 - Low, Medium, and High ESG risk bands have **similar average cost of debt** with substantial overlap.
 - High ESG risk firms do not exhibit dramatically higher borrowing costs than low-risk firms.
- **Column chart: Average cost of debt by sector**
 - Cost of debt varies **more noticeably across sectors** than across ESG bands.
 - Some sectors have systematically higher borrowing costs, pointing to sectoral or structural drivers rather than ESG alone.
- **Slicers**
 - **ESG_Risk_Band** filter to view only low/medium/high-risk firms.
 - **Sector** filter to focus on specific industries.

The dashboard reinforces the statistical findings in a more intuitive, exploratory way.

5. Discussion

The combined evidence from:

- OLS regression
- Predictive models (Ridge, Lasso, Random Forest, XGBoost)
- Visualizations

leads to a consistent message:

1. **ESG risk is not a strong standalone driver of cost of debt in this NIFTY50 snapshot.**
 - The ESG coefficient in OLS is small and statistically insignificant.
 - Ridge and Lasso fail to improve predictive performance meaningfully.
 - Non-linear models (Random Forest, XGBoost) overfit and generalize poorly.
2. **Sectoral differences matter more.**
 - The dashboard's sector breakdown shows clearer variation in average cost of debt across sectors than across ESG risk bands.
 - This suggests that industry structure, regulation, or asset risk profiles may be more powerful determinants of borrowing costs than ESG scores alone, especially in a sample of large blue-chip firms.
3. **Data size and outliers are a big constraint.**
 - With only 49 firms, it is hard to estimate complex models reliably.
 - A few firms with extremely high cost of debt create heavy-tailed residuals and make inference noisier.

6. Limitations

Key limitations of this project include:

- **Sample size**
 - Only NIFTY50 firms are included (49 observations post-cleaning), which limits statistical power and makes overfitting easy for flexible models.
 - **Single-period cross-section**
 - The analysis is based on the latest available snapshot. There is no time-series component to study how changes in ESG risk affect changes in cost of debt.
 - **Approximate cost-of-debt measure**
 - Cost of debt is constructed from financial statements (interest expense / borrowings) and may differ from market-implied yields or bond spreads.
 - **Outliers and residual non-normality**
 - A few firms with unusually high cost of debt generate skewed, fat-tailed residuals, which can distort standard OLS inference.
 - **Limited feature set**
 - Only four explanatory variables are used (ESG risk score, ROA, leverage, size). Additional variables such as credit ratings, currency exposure, or market volatility could improve explanatory power.
-

7. Conclusion and Future Work

This project set out to test a simple but important hypothesis:

Do NIFTY50 firms with higher ESG risk face a higher cost of debt, controlling for basic financial characteristics?

Based on the analysis:

- The **OLS model** finds **no strong, statistically significant relationship** between ESG risk score and cost of debt once profitability, leverage, and size are included as controls.
- **Predictive models** (Ridge, Lasso, Random Forest, XGBoost) do **not outperform** a simple mean baseline on out-of-sample data, reinforcing the idea that cost of debt is difficult to predict from ESG + a small set of financial ratios in this sample.
- **Sector-level patterns** in cost of debt appear more pronounced than ESG-based differences, suggesting that industry structure and other risk factors may play a more dominant role.

In short:

Within this NIFTY50 snapshot, ESG risk does **not** emerge as a clear, standalone driver of firms' borrowing costs.

Future work could:

- Expand the dataset to include a wider universe of NSE/BSE firms.
- Use panel data over multiple years to examine how changes in ESG risk relate to changes in financing costs.
- Incorporate more granular measures such as bond yields, credit spreads, or external credit ratings.
- Decompose ESG into E, S, and G pillars to see if specific dimensions matter more than the aggregate score.