Eco-Risk Alpha: Measuring the Green Discount in Counterparty CVA

Summer school "Data science for sustainable finance and economics 2025"

Letizia Cazzaniga, Angelique Girod, Marco Piazza, Jacopo Raffaeli

Project provided by Deloitte, supervisor: Dr. Helio Quintanhilha Jr.

8 - 12 September 2025

HTW Berlin

Abstract

This report documents the analysis performed for the Eco-Risk Alpha project. The objective is to test whether ESG characteristics are priced in credit markets and to measure the economic magnitude of any resulting "green premium" in unilateral CVA for a benchmark interest-rate swap portfolio. The dataset comprises a corporate universe with CDS spreads, ESG metrics, financial controls, expected exposure profiles for a 5-year payer IRS, and EUR discount factors. Methods include data cleaning, baseline linear regressions, explainable machine learning, and CVA calculations based on model-implied hazard rates.

Contents

List of Tables

List of Figures

1 Introduction

The Eco-Risk Alpha project investigates whether sustainability indicators have measurable explanatory power for credit risk and if such relationships can be monetised through counterparty credit valuation adjustments (CVA). Credit markets increasingly incorporate environmental, social, and governance (ESG) information into investment and lending decisions, yet empirical evidence on a systematic "green discount" remains mixed. Some studies report that companies with stronger ESG profiles trade at tighter credit spreads, implying lower default risk, while others find no significant pricing differential once fundamental factors are controlled.

The case study focuses on a portfolio of plain vanilla interest rate swaps (IRS) where the dealer must compute unilateral CVA to price and hedge counterparty default risk. The core research questions are:

- Do ESG metrics help predict credit default swap (CDS) spreads when controlling for traditional firm-level risk factors?
- Can any ESG effect detected in CDS spreads be translated into a lower CVA charge, generating an economic "green premium" for counterparties with higher sustainability scores?

To address these questions, the project requires linking observed 5-year CDS mid-spreads for a set of roughly 50 corporates with their ESG scores and key financial fundamentals. Statistical and machine learning models are trained to explain the cross-sectional variation in CDS spreads. Real and Model-implied spreads are then converted to default intensities and used in the unilateral CVA formula to compute respectively real and predicted CVA. The difference between CVA computed from market CDS spreads and CVA predicted from ESG-based models provides a quantitative measure of the potential green premium.

The motivation for this analysis is twofold. From a risk management perspective, identifying ESG-related credit effects allows dealers to refine counterparty selection and capital allocation. From a sustainable finance perspective, robust evidence of a green premium would support the idea that markets reward sustainable behaviour through lower funding costs. The report presents the theoretical framework, modelling methodology, empirical testing, and interpretation of results.

2 Theory and Design

This section outlines the theoretical background and methodological framework used to investigate the presence of a "green discount" in Counterparty Credit Valuation Adjustment (CVA). Our aim is to evaluate whether counterparties with stronger environmental, social and governance (ESG) scores exhibit systematically lower credit spreads and, consequently, lower CVA charges.

2.1 Statistical and Financial Background

CVA represents the expected loss from counterparty default in an over-the-counter (OTC) derivatives transaction. It is computed as the discounted expectation of the exposure at default multiplied by the counterparty's default probability. For a benchmark interest rate swap (IRS), CVA can be expressed as:

$$\mathrm{CVA} = (1 - R) \int_0^T \mathrm{EE}(t) \cdot \mathrm{DF}(t) \cdot \lambda(t) \cdot S(t)\, dt \tag{1}$$

Where:

- R is the recovery rate and $R = 1 - \mathrm{LGD}$, where LGD is the Loss Given Default
- $\mathrm{EE}(t)$ is the Expected Exposure at time t
- $\mathrm{DF}(t)$ is the discount factor at time t
- $\lambda(t)$ is the default intensity (hazard rate)
- $S(t)$ is the survival probability at time t
- T is the final maturity

and the discrete approximation:

$$\mathrm{CVA} \approx \mathrm{LGD} \sum_{i=1}^{N} \mathrm{EE}_i \cdot \mathrm{DF}_i \cdot (S_{i-1} - S_i) \tag{2}$$

Where:

- $S_i = e^{-\lambda ti}$
- $\lambda = \frac{\mathrm{CDS}}{1-R}$

The central hypothesis is that ESG performance influences creditworthiness. If higher ESG scores lead to tighter credit spreads, the hazard rate extracted from these spreads will imply a lower default intensity and thus a smaller CVA. This relationship can be conceptualised as a "green discount" in counterparty risk.

2.2 Model Design and Assumptions

The analysis proceeds as follows:

1. **Data Exploration and Cleaning:** Visualize CDS spreads versus ESG scores, identify and handle outliers, manage missing values, and check for duplicates.
2. **Baseline Linear Model:** Regress CDS spreads on ESG scores and selected control variables to estimate the marginal effect of ESG on credit risk..
3. **Factor Selection Rationale:** Justify feature choices to isolate the ESG effect.

4. **CVA Computation:** Convert both observed and model-predicted CDS spreads into hazard rates, then apply the unilateral CVA formula using to produce CVA estimates.

5. **Nonlinear Machine Learning Model:** Train Machine Learning models (Random Forests, XGBoost) to capture potential nonlinearities and interactions. Feature importance and interpretability tools are used to assess the contribution of ESG and other factors. Repeat the feature selection justification and CVA computation.

6. **Insight Synthesis:** Compare actual versus model-predicted CVA, group firms by ESG tertile, and interpret the economic significance of the "green premium", specifically we define it as $GP = CVA_{real} - CVA_{predicted}$.

3 Empirical Study and Testing

3.1 Dataset and Sources

The analysis relies on a proprietary dataset of approximately 50 anonymised firms containing both market and firm-specific variables. All company identifiers are synthetic and provided solely for educational purposes. The dataset includes:

- **CDS Spreads**: 5-year CDS spreads, used as a proxy for counterparty credit risk, this is the target variable.
- **ESG Data**: Aggregate ESG score, as well as individual *Environmental (E)*, *Social (S)*, and *Governance (G)* pillar scores.
- **Carbon Intensity**: Emissions-related measure reflecting the environmental footprint of each firm, normalized per unit of output.
- **Fundamental Financial Metrics**:
  - EBITDA
  - Market Capitalization
  - Total Assets
  - Debt-to-Equity ratio
  - Leverage ratio
- **Categorical Variables**:
  - Sector (e.g., Energy, Consumer Goods, etc.)
  - Region (e.g., Europe, North America, etc.)

The raw data were supplemented with a benchmark Expected Exposure (EE) profile and risk-free zero rate curve, which are key inputs to the Credit Valuation Adjustment (CVA) calculations. This combination of market, ESG, and fundamental variables allows a comprehensive analysis of the potential impact of ESG factors on counterparty credit risk.

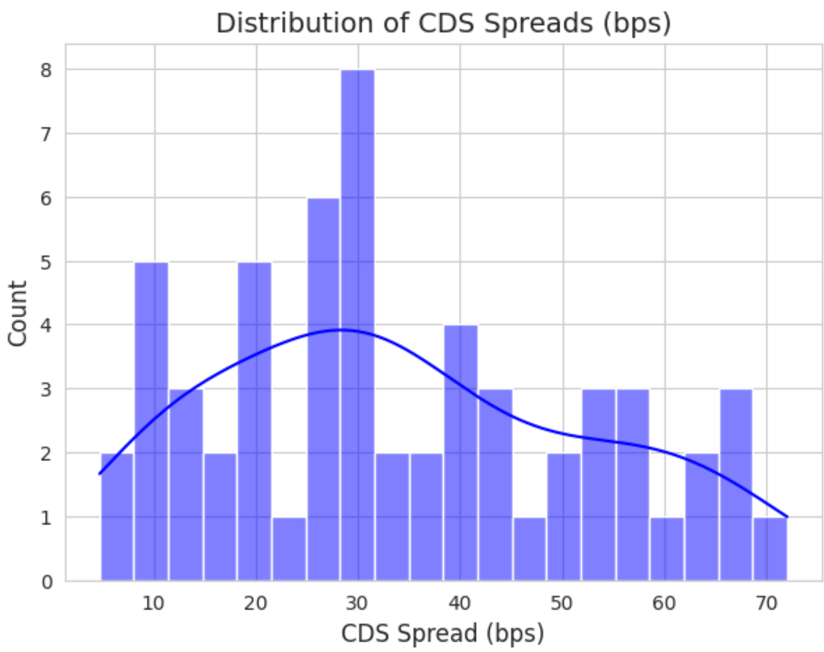

Figure 1: CDS spreads distribution.

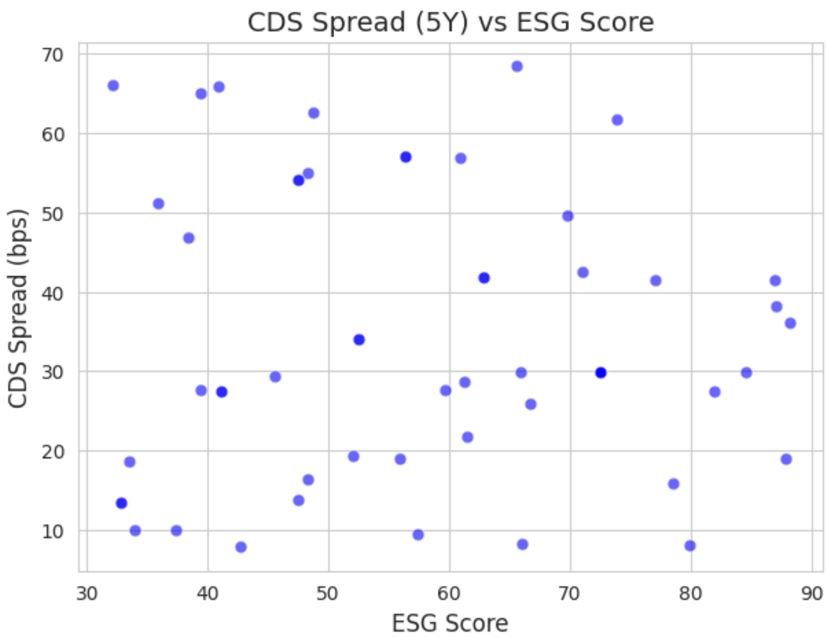

Figure 2: Scatter plot of Carbon Intensity vs CDS spreads.

3.2 Preprocessing and Data Cleaning

Data preparation included:

- **Missing values:** Sector-wise median imputation applied to Carbon Intensity.
- **Omitted features:** Market Capitalization was excluded due to many missing values and overlap with Total Assets, which is retained.
- **ESG score computation:** Missing ESG values were derived from pillar scores as $ESG = 0.5 \cdot E + 0.3 \cdot S + 0.2 \cdot G$.
- **Outlier detection:** Isolation Forest identified three potential outliers, which were retained to preserve real-world heterogeneity.

Exploratory analysis involved correlation matrices and bivariate plots to gauge initial relationships.

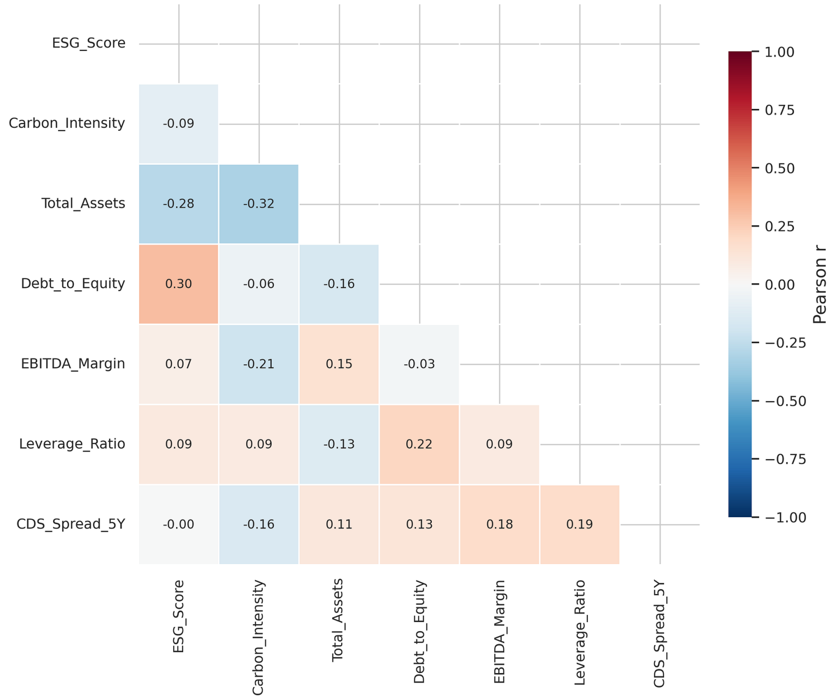

Figure 3: Correlation Matrix.

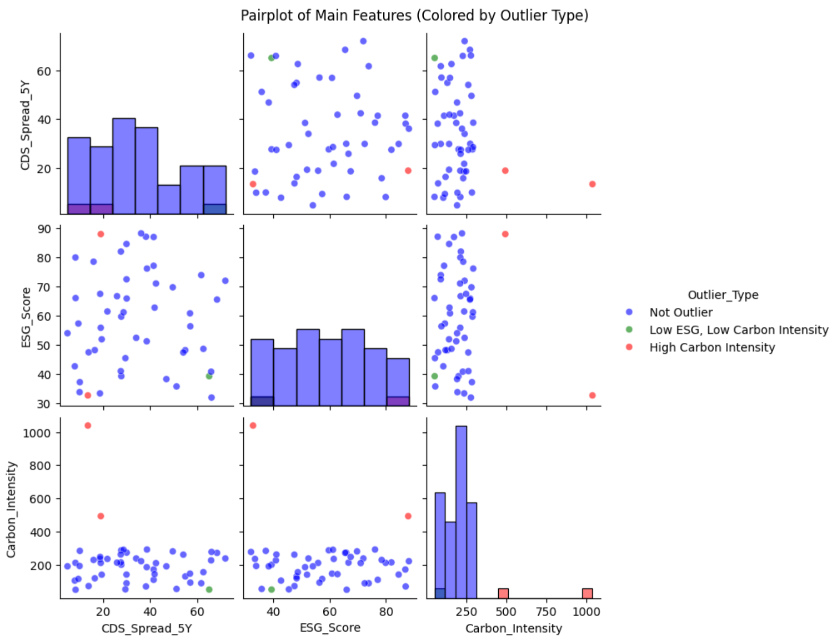

Figure 4: Outliers pairplot with main features.

3.3 Testing Strategy

The testing strategy is defined as follows:

1. **Baseline linear models**: Ordinary Least Squares (OLS) and Lasso regression were trained to explain CDS spreads using ESG and fundamental variables. Forward and backward Sequential Feature Selection (SFS) identified a parsimonious subset of predictors.

2. **Non-linear model**: A Random Forest (RF) regressor was implemented as a proof of concept to capture potential non-linear interactions between ESG factors and credit risk. Hyperparameters were tuned to mitigate overfitting.

3. **Validation**: Model performance was evaluated via Leave-One-Out cross-validation (LOOCV), reporting Root Mean Square Error (RMSE).

4. **CVA computation**: Both market CDS spreads and model-predicted spreads were converted to hazard rates and then to unilateral CVA using the provided EE profile and a fixed Loss-Given-Default (LGD) of 0.6.

5. **Green premium analysis**: Companies were bucketed into ESG score terciles. Actual CVA and model-implied CVA were compared across these groups to estimate a "green credit premium."

Model	**RMSE**
Linear Models	
Linear Regression (all features)	24.75
Linear Regression with Lasso	19.30
Stepwise Feature Selection (SFS)	18.74
Top-3 Correlated Features	19.39
Random Forest Models	
Random Forest (all features)	20.44
Random Forest (all features, tuned)	19.68
Random Forest SFS	18.03
Random Forest SFS (backward)	17.10
Random Forest SFS (forward, tuned)	17.48
Random Forest SFS (backward, tuned)	16.76

Table 1: Model performance comparison based on LOOCV RMSE.

3.4 Empirical Results

Key outcomes include:

- **Linear models**: Low explanatory power and RMSE around 18–25 bps, with Carbon Intensity, Leverage Ratio and the Energy sector dummy emerging as relevant features.
- **Random Forest**: Substantially higher in-sample fit and lower RMSE (down to ∼17 bps), especially after feature selection (EBITDA Margin, Leverage Ratio, Energy sector).
- **Green premium**: Median CVA charges for top-tercile ESG firms were modestly lower than for bottom-tercile firms, but differences remain within model error bands and should not be interpreted as definitive market pricing evidence.

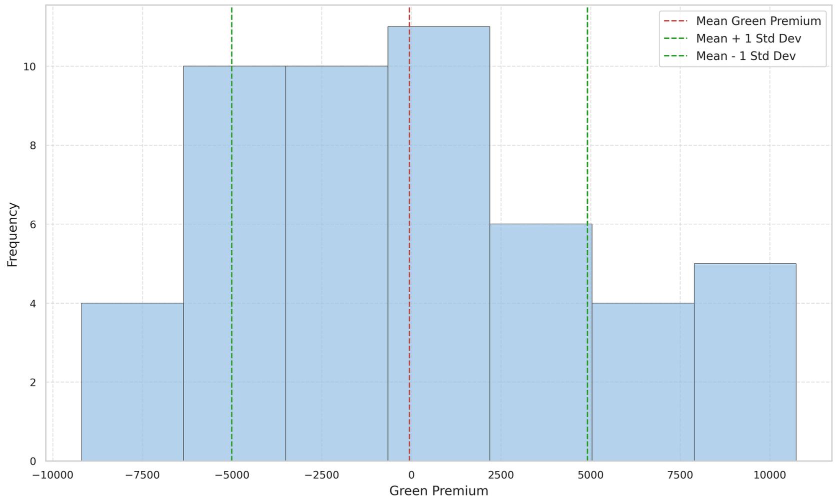

Figure 5: Green Premium distribution.

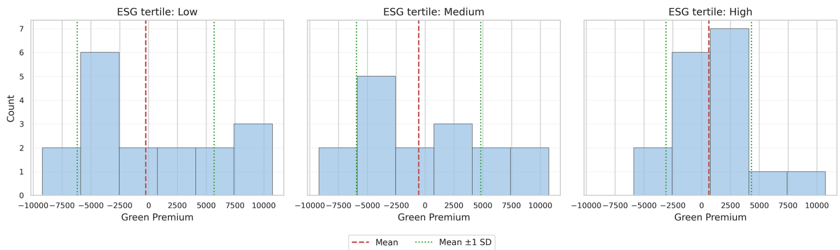

Figure 6: Green Premium distribution per tercile.

3.5 Stress Scenarios

Due to limited sample size, no formal stress testing of the zero rate curve was performed. This remains an important extension to assess CVA sensitivity under adverse market conditions.

4 Conclusion

The Eco-Risk Alpha project explored whether ESG factors have explanatory power for credit spreads and whether such relationships can generate a measurable "green credit premium" through CVA adjustments.

4.1 Take Home Messages

Our main findings are:

- Linear models indicate a modest correlation between ESG metrics and CDS spreads, with Carbon Intensity, Leverage Ratio, and Energy sector membership being the most relevant predictors.
- Random Forest models reveal stronger in-sample fit and lower prediction error, suggesting potential non-linear relationships between ESG factors and credit risk, though these results remain illustrative due to the small sample size.
- The computed green premium—differences between actual CVA and model-predicted CVA across ESG terciles—is relatively small and given from the model's prediction error, meaning it cannot be interpreted as definitive market evidence of pricing a "green discount."

4.2 Limitations and Possible Future Work

Challenges and limitations include:

- Limited dataset (~50 observations) restricts statistical power and precludes robust out-of-sample validation.
- The same dataset was used for feature selection, hyperparameter tuning, and model evaluation, which can introduce overfitting risk despite cross-validation.
- No stress tests of zero rate curves were performed, leaving sensitivity to market shocks unassessed.
- Synthetic data augmentation was considered but not applied, as small sample sizes can generate misleading patterns.

Future directions to strengthen the analysis:

- Expanding the dataset with additional counterparties or historical time series to improve model robustness.
- Performing nested cross-validation or maintaining a separate test set to better assess predictive performance.
- Applying stress scenarios to discount rates and exposures to evaluate CVA sensitivity under adverse market conditions.
- Investigating additional machine learning models (e.g., gradient boosting) with interpretability tools to confirm or refine ESG effects.

Overall, the project provides a structured framework for linking ESG scores to credit risk and quantifying potential impacts on CVA. While preliminary results hint at a green premium, the small sample and model limitations mean that further empirical validation is necessary before drawing definitive conclusions.

A Code repository and run instructions

Repository: Eco-Risk-Alpha

B AI usage declaration

Declaration of Authorship

"I hereby confirm that I have authored this report independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such. All verbatim or in-sentence copies and quotations, as well as all sections that were designed, written and/or edited with the help of AI-based tools, have been identified and verified. Information on the use of AI-based tools In the appendix of my work, I have listed all AI-based resources. These are identified with product names and formulated inputs (prompts) in an AI directory. I assure that I have not used any AI-based tools whose use has been explicitly excluded in writing by the examiner. I understand that the use of texts or other content and products generated by AI-based tools does not guarantee their quality. I am fully responsible for the adoption of any machine-generated passages used by me and bear the responsibility for any erroneous or distorted content, faulty references, violations of data protection and copyright law or plagiarism generated by the AI. I also affirm that my creative influence predominates in the present work."

References