Analysis of Probability of Default and Recovery Rate in CDS Pricing

Peter Emanuel and Will Line

CFEM

CORNELL FINANCIAL
ENGINEERING MANHATTAN

Table of Contents

I. Motivation

A Credit Default Swap (CDS) is an instrument that has become fundamental to the US financial markets ever since the global economic crisis of 2008. The owner of a CDS has protection against the default of a debtor, akin to owning a put option on the assets of a company or sovereign. The purpose of owning a CDS on a US company is self-evident, however, it is difficult to transact in US Corporate CDS because the market is, in many ways, inefficient. In contrast to the more liquid US Equities market, CDS are not exchange-traded, have far fewer participants, higher regulatory barriers to entry, and are relatively newer.

Moreover, CDS pricing models are far less developed. Two common metrics used to evaluate a CDS are the implied probability of default of the company over a given period of time, *P(0,t)* and the recovery rate a debtholder would expect to receive on their investment in the event of a default, *R*. Relating these quantities to CDS spread S_t yields the following relationship:

$$P(0,t) \;=\; 1 - exp(\frac{S_t}{1-R}) \text{ [1]},$$

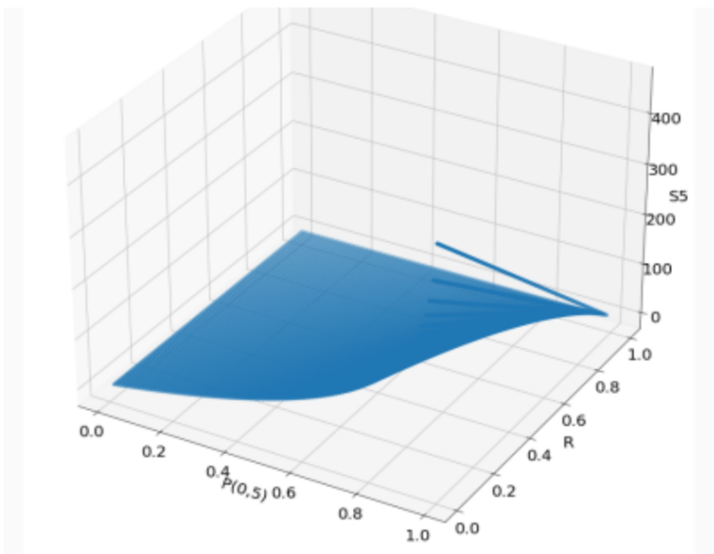

Figure 1: 5y CDS spread surface given 5y probability of default and recovery rate

where *P(0,t)* and *R* lie between 0 and 1 and S_t is expressed in percentage points. The market assumes a constant 40% *R*, restricting the pricing of CDS to one cross-section of the entire surface. The implied probability of default is often backed out of the equation, taking the market CDS spread as given.

We attempt to build a stronger pricing model for US High Yield Corporate 5-year CDS, S_5, by organically learning *P(0,5)* and by allowing *R* to vary across time and across companies. Having a stronger pricing model can help investors to better identify and act on mispricings. As more participants enter the market and adopt a stronger pricing methodology, the market will become more efficient. The positive feedback loop will benefit investors, market-makers, and risk managers alike.

II. Overview of Methodology

Our methodology is divided into three primary components:

1. Organic learning of *P(0,5)*

P(0,5) may vary significantly if *R* is allowed to vary from 40%. We believe a company's probability of default may vary linearly with the daily pricing data of their stock and bond, categoricals such as sector and rating, and macroeconomic indicators. After undergoing preprocessing and missing value imputation, we train a standard Ordinary Least Squares (OLS) regression on these features to predict a daily *P(0,5)* and introduce a lasso to derive more information about the features' explanatory power.

2. Classification and distributional analysis of *R*

We would like to form a more rigorous estimate of *R* by day and by company using the new learned *P(0,5)*. To start, we compute a new set of *R* with new *P(0,5)* and market S_5. We know that market S_5 may be inaccurate, so to reduce dependence on those observations, we instead categorize the predicted *R* as either over or under the 40% hurdle, and then train a classifier on each data point. We compare a Support Vector Classifier (SVC) with a Decision Tree (DT) on their ability to correctly classify *R* out-of-sample. We analyze average classification by month and by company to look for global trends in the new learned *R*. Finally, we briefly discuss the fairness and potential social impact of our classification.

3. Repricing of S_5, discussion of tradeability, and future work

Given a new learned *P(0,5)* and *R*, we solve one last time for S_5 and discuss how this model estimate relates to the market observation throughout the life of the dataset. We propose ideas for further work to improve the actionability of the output. In particular, we detail what needs to be done so that the model is consistently tradeable.

III. Description of Features

We select 153 High Yield companies to examine by taking every public company with at least one bond in the iShares iBoxx High Yield Corporate Bond ETF (HYG) holdings on October 28th, 2021 and labeling them by ticker symbol and sector. Daily Bloomberg S_5, stock price, 5-year US Treasury yield, and Market Volatility Index (VIX) data was taken from 9/1/2016 to 10/28/2021. 5-year bond yields used were the 5-year liquid benchmark bond yield at the time for each company, to account for new issuance. Bond Rating used was the Bloomberg Composite Rating, the mode of Moody's, S&P, and Fitch ratings, ranging from CC+ (lowest "junk" observation) to BBB- (highest "crossover" observation). Monthly Industrial Production and Consumer Price Index levels were taken from the FRED database[2] from 9/2016 to 10/2021.

In summary, the dataset used contains 153 tickers and 1,346 days of data. 2 features, stock price and 5-year bond yield, were unique by time and ticker. 2 features, rating and sector, were unique by ticker. The remaining 4 features were unique by time. Therefore, the feature space has 417,566 unique entries in total.

IV. Data Preprocessing

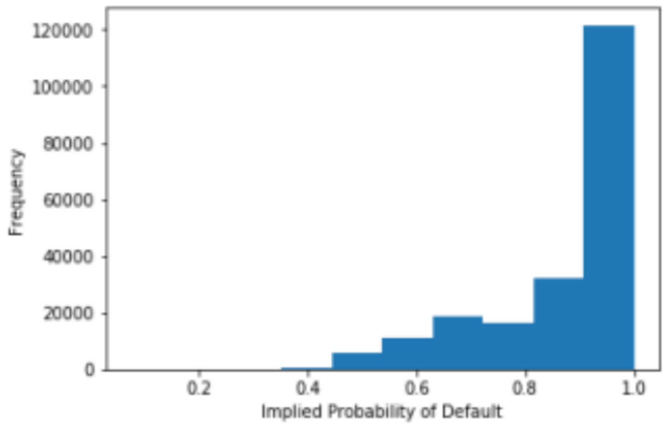

Figure 2: 1 - P(0,5) is roughly exponentially distributed

We standardize all numerical data to reduce potential autocorrelation and transform rating and sector with a manyhot encoding. Missing numerical values are front-filled; other than in a backtest, if a stock price is missing on day *d*, one would not have information about *d+1*, so the more reasonable estimate one could make for the stock price at *d* is the observation at *d-1*. Missing values at the beginning of the dataset are back-filled. We preprocess the labels, the *P(0,5)* as implied from S_5, with a Box-Cox transformation[3] because they do not resemble a normal distribution.

V. Learning of *P(0,5)*

1. Ordinary Least Squares (OLS) Regression

We train an OLS regression to predict *P(0,5)* from our preprocessed features. Each feature was individually added to the model and tested for statistical significance to determine if that feature should remain in the final model, yielding the following prediction of the daily Z-score of *P(0,5)*:

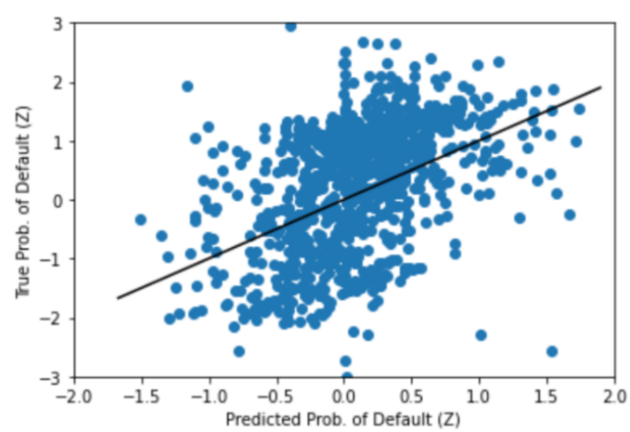

Figure 3: OLS Regression Pred. vs Actual; MSE: 0.9; R^2: 0.13

$$\hat{Y} = 0.20 * 5yearBondYield - 0.26 * StockPrice - 0.01 * 5yearTreasuryYield$$

We predict a higher *P(0,5)* as 5-year bond yield increases, but a lower *P(0,5)* as stock price or 5-year treasury yield increases. For example, for every unit increase in the standardized 5-year bond yield, we predict a 0.2 increase in the standardized *P(0,5)*, which is approximately equivalent to a 10% increase in *P(0,5)* itself, going from the $Z = 0$ to $Z = 0.2$ level. We calculate a Durbin-Watson statistic[4] of 1.8 for the residuals, implying no significant autocorrelation, verifying that a linear model is appropriate to model this relationship.

2. Lasso Regression

We implement a lasso regression to further explore the predictive power of the features. We use all of the features in the dataset to start, and allow the lasso to perform its selection, sending feature weights that add no predictive value to zero. The lasso regression had very similar mean-squared error (MSE) (0.86) and R^2 (15%) values to the OLS regression, returning weights as shown in the Results table below.

Results:	
Feature:	Weight:
Industrial Production	0.0012
Consumer Price Index	0.0007
Volatility Index	0.0024
Treasury Yield	-0.0076
Sector	0
Bond Rating	0
Bond	0.2027
Stock	-0.2631
Offset	0.0015

The lasso regression finds sector and rating weights to be zero; while sector and rating may be useful in the analysis of Investment Grade bonds, they more coarsely differentiate two High Yield companies. Sector and rating are also already highly correlated with 5-year bond yield, not in the numerical sense, but in the sense that different sectors have different fundamental properties that lead to certain ratings, which leads to sector-wide tighter or wider bond trading levels. The features with the largest magnitude weights are exactly the same as the features included in the final OLS regression: 5-year treasury yield, 5-year bond yield, and stock price. The remaining features have very low weights and do not add much predictive power. Industrial Production and Consumer Price Index levels are only reported monthly and therefore do not have a very significant impact on the daily predictions, given our model framework, and the VIX, being more closely tied to US Equity market volatility than probability of default, appears to not play a significant role in this analysis.

In summary, the OLS and lasso regression models have virtually the same MSE and R^2 values, producing roughly equivalent predictions for *P(0,5)* . The two models' equivalence increases our confidence in their predictions.

VI. Classification of *R*

Given the new learned *P(0,5)*, we now seek to develop a better understanding of how *R* should behave. To recalculate *R*, we convert the *P(0,5)* predictions back into probabilities by un-standardizing with ticker sample means and standard deviations. We then transform the new *R* into a series of +/- 1's to represent the data points that should recover more/less than 40% in the event of a default, which we will call over or under-recoverers. At this stage, we prefer to model *R* with classification, rather than regression. We know that market S_5 are not always accurate, but we now must use them as an input in the pricing formula to solve for *R*. As a result, we have less confidence in differentiating between two data points of similar *R*, but at the very least, we can say something about those two points in tandem if they are both well above or below 40%. We train a Support Vector Classifier (SVC) and a Decision Tree (DT) on the dataset, comparing the results on their out-of-sample misclassification rates and confusion matrices[5].

1. Support Vector Classifier (SVC)

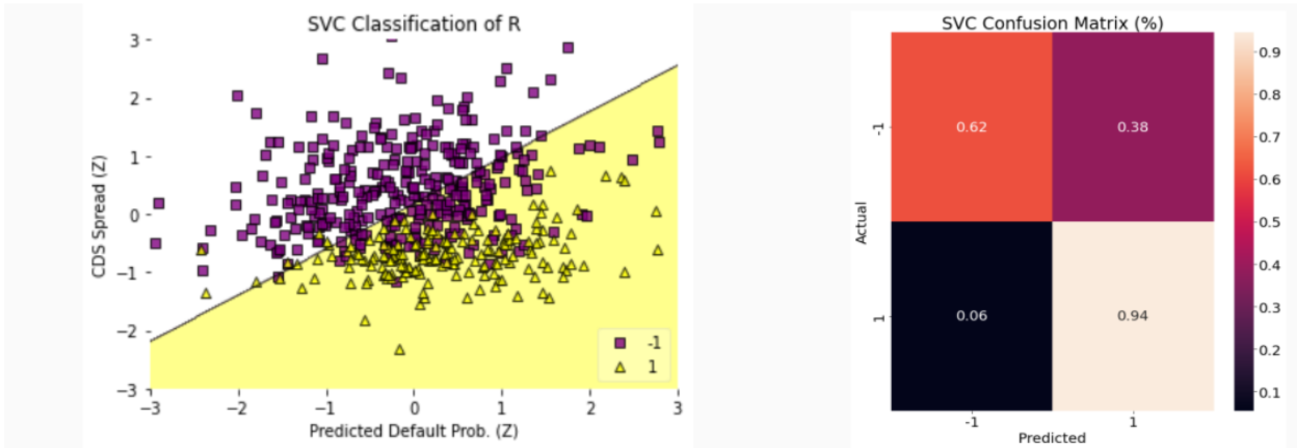

Figure 4: SVC Classifier; OOS Misclassification Error: 26%

2. Decision Tree (DT)

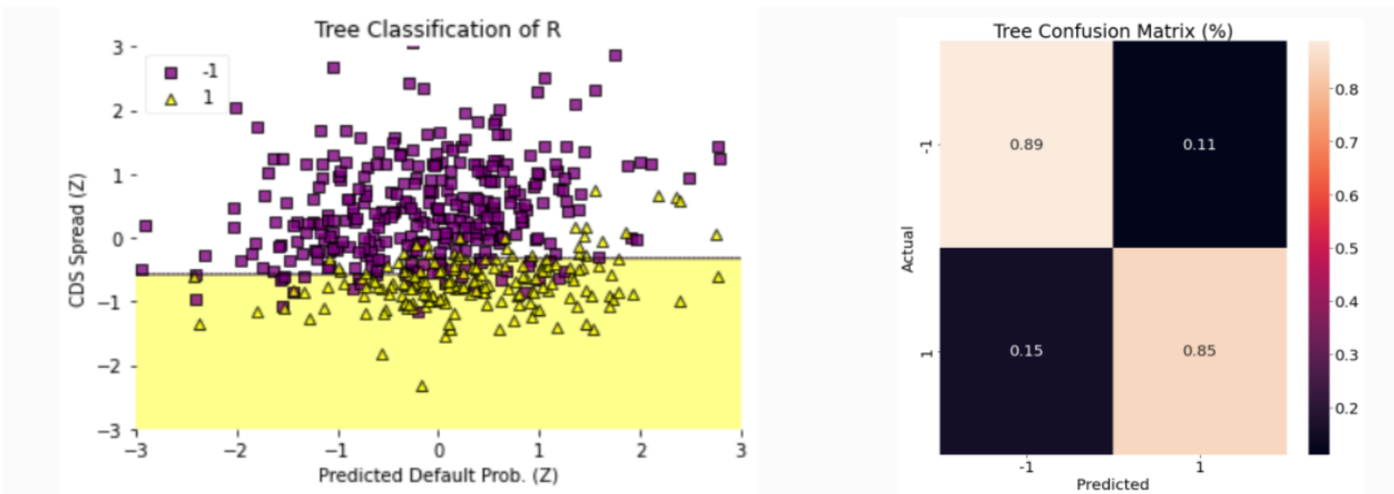

Figure 5: Decision Tree Classifier; OOS Misclassification Error: 13%

We choose the optimal l_2-regularization parameter for the SVC to be .01 from $\lambda \in (0,1)$ and the optimal number of splits for the DT to be 4 from $s \in [2, 3, \ldots, 10]$ in the cross-validation stage. The DT has half the out-of-sample misclassification error of the SVC, a low number of splits, and relatively equal entries along the diagonal of the confusion matrix. Moreover, the DT's nonlinearity more closely tracks the Theoretical Classification Boundary (TCB) that would have been observed, given the original market-implied $P(0,5)$, than the linear SVC.

We assert that the DT is a good model for classifying future over or under-recoverers, given that the only four pieces of information one needs to know to do so are standardized stock price, 5-year bond yield, 5-year treasury yield, and market S_5. The linear regression predicts $P(0,5)$ with the first three pieces, then the DT makes a directional statement about R with the new $P(0,5)$ and the fourth piece.

VII. Global Trends and Distributional Analysis of *R*

We revisit the motivating idea that the assumption of a flat 40% *R* is naïve. The results of the DT imply that *R* should instead vary by both time and ticker.

1. Ticker-variation of *R*

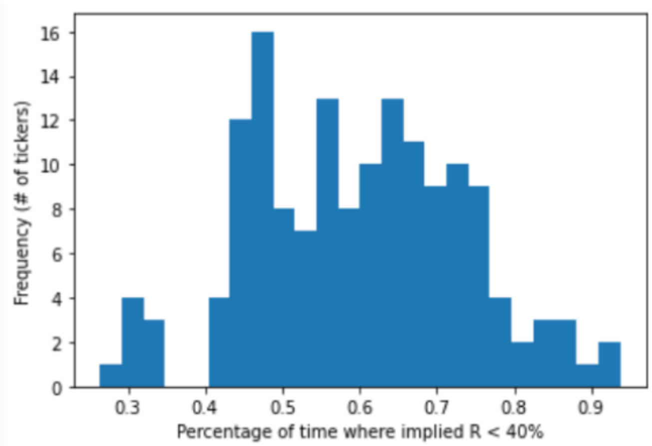

Figure 7: Average classification by ticker over full dataset

We assess the average classification (shifted from -1/+1 to 0/1) by ticker. If a 40% *R* were appropriate, the histogram in Figure 7 would be closely clustered at 0.5. However, there is instead a significant amount of weight in the tails of the distribution, implying that the market is systemically mispricing S_5 for those tickers. In the left tail, we find implied *R* higher than 40% around 70% of the time, while in the right tail, we find implied *R* lower than 40% around 90% of the time. Each ticker's *R* distribution is approximately normal (around varying means) and not significantly skewed. We conclude from this information that *R* should surely not be the same 40% value for each ticker.

2. Time-variation of *R*

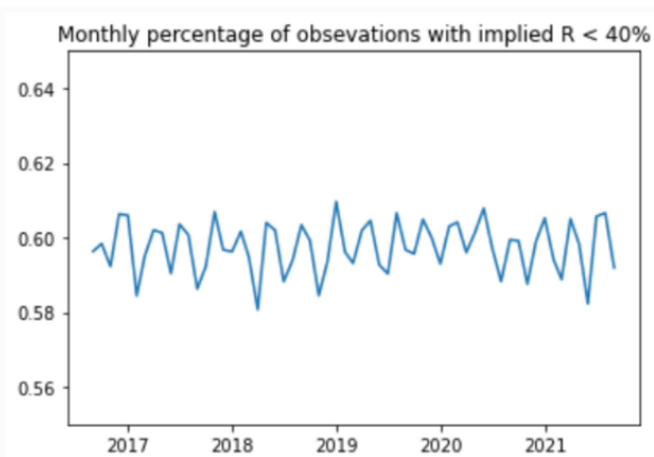

Figure 8: Average classification across all tickers by month

We assess the average monthly classification of all tickers. If a 40% *R* were appropriate, the plot in Figure 8 would be close to a horizontal line at 0.5. We instead observe a cyclical pattern centered around 0.6, suggesting that the companies in our dataset spend more time as under-recovers than over-recoverers. Indeed, the average *R*, by month in the dataset, is 30%, not 40%. We conclude from this information that *R* should surely not be the same 40% value at each point in time, even on average for all tickers.

3. Fairness and Social Impact Considerations

Though we acknowledge the importance of fairness in algorithms and machine learning models, we do not think the aforementioned analysis is sensitive to it. No model discussed is biased against a protected attribute. Perhaps further investigation of demographic parity as applied to the ESG spectrum could uncover differences in *R* classification between low and high-rated companies. However, ESG is already encapsulated in company ratings[6], which were not significant in predicting *P(0,5)* and so were not used in the classification of *R*. Moreover, we do not believe these models are Weapons of Math Destruction[7] because the results are easily measurable, have positive consequences for all market participants in the form of increased pricing accuracy and efficiency, and do not create negative feedback loops in terms of reduced ability to transact in CDS.

VIII. Implications of Varying *R* on S_5 Repricing

Given new estimates of *P(0,5)* and classifications and distribution-level information about *R*, we compute S_5 once more to produce a model S_5 and compare its evolution across the lifetime of the dataset to that of the market S_5. In the median case, we observe a general market underpricing of S_5, particularly in the first two years. This observation supports the claim that, if a new *R* were to be used for all tickers, it should surely be lower than 40%, as a lower *R* assumption implies a more valuable, expensive, and thus higher S_5. Investors may view these periods of time as outright buy opportunities (though the opportunity to earn a credit-risk-free bond-CDS basis[8] may go away), market-makers may skew their bid/offer upward to compensate for shorting, and risk managers may adjust their *R* assumptions downward. These actions would collectively move the market upward until the market matches the model, yielding a more efficient pricing of High Yield Corporate 5-year CDS.

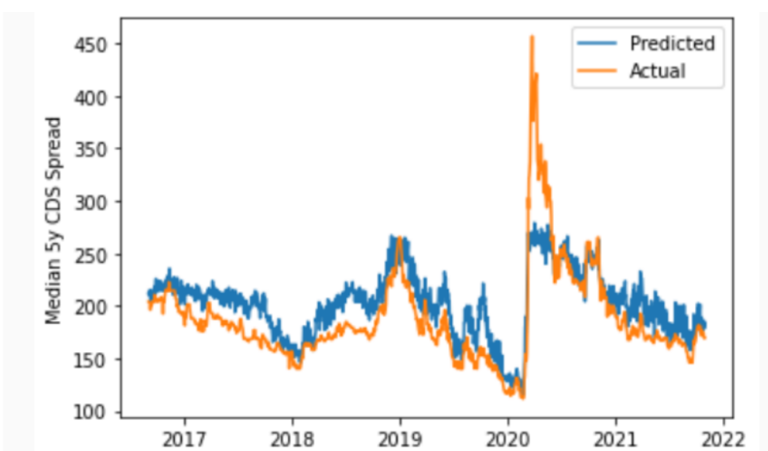

Figure 9: Periods of market underpricing compared to model, but model is noisy and fails to capture outlier events

IX. Future Work

Further analysis would seek to address shortcomings in the model S_5. We may first seek to get a more precise estimation of *R*. For example, we may use the TCB itself at the 30% *R* boundary, our mean dataset *R*, as a classifier, to attempt to outperform the DT. We may also choose to apply a nonlinear model, such as a neural network, to learn *R*.

We may also seek to address the final prediction of S_5. In particular, the model S_5 does not capture the spike in market S_5 that occurred in early 2020 due to the COVID-19 Pandemic. Removing the normality assumption of bond yields, for example, by tweaking preprocessing to account for their higher kurtosis, could address this problem. The pricing formula itself is also not able to compute a S_5 too high, else *R* will become negative. For example, the maximum S_5 given nonnegative *R*, at the 99.9% P(0,5) level, is 690 basis points (bps). As a result, our model cannot compute anything above 690 bps, given 3 significant figures of accuracy on P(0,5), and would always assume extreme market observations to be overpriced. Tweaking the pricing formula to capture extreme pricing cases would address this issue.

The model S_5 is also noisier than the market observation at almost all points in time. This phenomenon can be explained by the fact that the illiquid market S_5, in many cases, stays flat for multiple trading days or weeks before jumping when the next trade occurs. The model S_5, on the other hand, uses a new *R* and *P(0,5)* each day. The stock, corporate bond, and treasury markets are far more liquid; at least one of these features is expected to change each and every trading day. Additionally, the model S_5 is now a function of two variables, naturally creating more variation than the market, which holds *R* constant. Although the computed-daily, multivariate-input model S_5 may in fact be more robust than the market, to drastically change bid/offer, re-evaluate investment targets, or re-calibrate risk models daily may be counterproductive. Using a time series smoothing technique or ARIMA on the model S_5 would address this issue.

X. References

[1]Derived from Hull, J. C. (2014). Options, Futures, and Other Derivatives (9th ed.). Pearson.

[2]Federal Reserve Bank of St. Louis. (n.d.). *Federal Reserve Economic Data: Fred: St. louis fed*. FRED. Retrieved November 4, 2021, from https://fred.stlouisfed.org/?gclid=Cj0KCQiA47GNBhDrARIsAKfZ2rA9NvJu4tMyk-jDL_aQG9vutMtcq2pOCs5jA2dtq04OzF4xkrePL3oaAt2IEALw_wcB.

[3]Box, G. E. P., and D. R. Cox. "An Analysis of Transformations." *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, [Royal Statistical Society, Wiley], 1964.

[4]Durbin, J., and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression. II." *Biometrika*, vol. 38, no. 1/2, [Oxford University Press, Biometrika Trust], 1951

[5]Narkhede, Sarang. "Understanding Confusion Matrix." *Medium*, Towards Data Science, 15 June 2021, https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

[6]"Credit FAQ: How Does S&P Global Ratings Incorporate Environmental, Social, and Governance Risks into Its Ratings Analysis." www.spglobal.com/ratings/en/research/articles/171121-credit-faq-how-does-s-p-global-ratings-incorporate-environmental-social-and-governance-risks-into-its-ratings-10321964.

[7]O'neil, Cathy. *Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy*. London, Penguin Books, 2018.

[8]Bai, Jennie and Collin-Dufresne, Pierre, The CDS-Bond Basis (September 12, 2018). Georgetown McDonough School of Business Research Paper No. 2024531, Available at SSRN: https://ssrn.com/abstract=2024531 or http://dx.doi.org/10.2139/ssrn.2024531