

Random Forest を用いた ESG 情報による信用格付 予測モデル構築の提案と CDS 市場分析への応用可能性の検討

ニッセイアセットマネジメント

運用戦略部 ファイナンシャルテクノロジー運用室 シニア・ポートフォリオ・マネジャー

中央大学大学院 理工学研究科 確率解析・金融工学・保険数理研究室

東出 卓朗¹

1. はじめに

情報技術の進化に伴い、企業分析を行ううえで多種多様なデータとモデルが扱われるようになってきた。かかる中、Environment・Social・Governance(ESG)データは伝統的な財務関連データとは性質を異にし、経営などに関連した非財務データである。ESG を運用プロセスの中に取り込み、中長期的な目線で実践する持続可能な投資手法が国連グローバルコンパクト(UNGC)により推奨されて以来、一層の注目を集めている。実際、ESG ファクターが企業価値に影響を与えているといった報告もされている (Gutsche, Schulz, & Gratwohl, 2017; Lins & Servaes, 2017)。

ところで、研究者や実務家は ESG の概念を信用格付の判別問題に取り込む必要性を唱えてきた(Attig, El Ghouli, & Guedhami, 2013; Zeidan, Boechat, & Fleury, 2015)。Devalle, Fiandrino & Cantino (2017) は、ESG は信用格付に影響を与えると述べ、良好な ESG スコアを有する企業の信用格付は高格付けになる傾向があることを、Ordered Logistic 回帰を用いて実証した。信用格付の判別分析を行った最近の論文では、線形判別といった伝統的な手法(Altman, Marco & Varetto, 1994)だけでなく、機械学習を用いた手法も報告されており精度の高さが目を引く。例えば、Neural Network(Desai et al., 2000), Support Vector Machine (Zhong et al., 2014), Random Forest (RF) (Yeh, Lin & Hsu, 2012) を用いた分析がある。しかし、精度の低さとは裏腹に、依然として Logistic 回帰はモデルの解釈性の高さから、実務家より支持されている現状がある(Crook, Edelman & Thomas, 2007)。

本稿では、ESG データと RF (Breiman, 2001) を用いた信用格付の予測モデルを構築する手法を提案する。また、構築した RF モデルが伝統的な手法の一つである Multinomial Logistic 回

¹ ひがしで たくお

ニッセイアセットマネジメント株式会社 運用戦略部 ファイナンシャルテクノロジー運用室に
シニア・ポートフォリオ・マネジャーとして所属するとともに、中央大学大学院理工学研究科 確率解析・金融工学・
保険数理研究室に所属している。

本稿の中で示された内容はニッセイアセットマネジメント株式会社の公式見解を示すものではない。

〒100-8219 東京都千代田区丸の内 1-6-6 日本生命丸の内ビル(ニッセイアセットマネジメント)

〒112-8551 東京都文京区春日 1-13-27

E-mail: higashide_t590@nam.co.jp, takuo0823@gmail.com

謝辞：神戸大学大学院経済学研究科 羽森茂之教授、金京拓司教授、田中克幸研究員より有益なコメントを頂戴した。

帰(MLR)と比べてより高い予測力を有することを実験的に示すとともに、予測された信用格付がクレジット・デフォルト・スワップ(CDS)の価格を説明するうえで有効なダミー変数となり得るかについて検証を行う。RFは、モデルの過剰適合に対して頑健であり、外挿期間においてもエラー率が低いことが Breiman (2001)により理論的に示されている。また、解釈性にも秀でていることが知られている。以上の理由より RF を用いた。本稿は次の2点において、新規性がある。

- ・ モデルの枠組みが ESG データと RF の組み合わせであること。
- ・ 日本企業のデータを用いて実験を行っていること。

1点目については、RFを用いて信用格付を予測する先行研究では、説明変数に市場データを用いているものが多く、ESGと組み合わせたものは少ない。ここで、ESGデータは市場データとは性質が異なるものであることを改めて注意しておく。2点目については、日本は UNGC に加盟する企業数が多い国の一つであるにもかかわらず、ESGと信用格付けの関係について分析された先行研究が少ない。本稿における主な貢献は、日本企業を対象に ESG が信用格付けを予測するうえで十分な特徴量を兼ね備えたデータであることを実験的に示している点にある。

本稿の構成は以下の通りである。まず第2節で本研究に用いたデータとモデル構築の提案手法について説明する。次に第3節で実験的に信用格付予測の精度を測るとともに、CDS市場の分析への応用可能性について言及する。この際、MLRとRFによるモデルの精度についても述べる。最後に第4節で結論を纏める。

2. データと手法

データソース

日本企業のうち、社債を発行している先かつ ESG スコアが付与されている計 295 社を分析対象とした。ESG スコアは SUSTAINALYTICS (<https://www.sustainalytics.com/>) 社のデータを用いている。295 社の業種別内訳は 126 社が製造業、106 社が非製造業、63 社が金融業である。信用格付は、クレジット・プライシング・コーポレーション (<https://creditsurfer.jp/>) 社のデータを用いている。観測期間は 2013 年 1 月から 2017 年 7 月までの月次頻度のデータである。株価、CDS データは Bloomberg (<https://www.bloomberg.co.jp/>) より取得した。また CDS データは流動性の高い期間 5 年物を用いた。

ESG データ

図表 1 は ESG データの要約である。ESG データは 2 層からなり、上層は environment (E), social (S) そして governance (G) の 3 種類からなる。一方で下層は、上層の E・S・G 3 種類をそれぞれ 55, 56, 34 種類に分解したデータとなっている。本稿では、上層 3 種類のデータセットを用いる。

図表1 ESGデータの要約

ESG データセット			
層	E (Environment)	S (Social)	G (Governance)
上層	1	1	1
下層	55	56	34

(図表注1) 図表1は各層における変数の個数の要約である。

(出所) 筆者作成。

信用格付データ

図表2は信用格付データの要約である。訓練データはモデル構築を行ううえで用いるデータであるのに対して、試験データは外挿期間におけるモデル精度を計算するうえで用いる。ところで各格付け帯における不均衡データは、偏ったモデルを構築しかねない。不均衡の是正に対しては大きく2つの手法が存在する。一つは、モデル自体に改善を求めるか、或いはサンプリング技術を駆使するかである。前者については、注釈(1)をご覧いただきたい。一方で後者については、Lewis(1992)が、頑健で精度の高いモデルを構築するためには各格付け帯に1500個の標本数があれば十分であると述べている一方で、Mays(2001)は300個の標本数で十分であると述べている。また、Anderson(2007)はモデルの偏りを生じさせないような各集合における適切な標本数について明確な理論が存在していないことを指摘している。これらの議論を考慮して本稿で用いるデータは、AAAからBBBまでの適格銘柄に限定するとともに、各格付け帯におけるポジティブ / フラット / ネガティブの違いは考慮せず、業種別の格付け帯単位で分析を行う。

図表2 業種別の各格付け帯におけるサンプル数

業種	AAA	AA	A	BBB	BB	B
金融業 訓練データ	4	1329	976	108	0	31
金融業 試験データ	0	189	93	14	0	0
製造業 訓練データ	17	1884	3213	173	11	26
製造業 試験データ	0	291	474	10	1	3
非製造業 訓練データ	190	1898	2215	319	0	0
非製造業 試験データ	28	278	338	42	0	0

(図表注1) 図表2の数値は各格付け帯におけるサンプル数を示している。

(出所) 筆者作成。

以上のデータを用いて、次節よりRFとMLRの精度比較を行うが、モデルの頑健性を考慮して、訓練データに用いるデータは、図表2の格付け帯毎に10個の部分集合を作成し、それぞれの部分集合に対してRF分類器を構築する。その際、部分集合の各格付け帯が持つ標本数は100個とし、それに満たない格付け帯においては、全データを用いるものとした。

手法

モデルは RF²(Breiman, 2001) を用いる。RF は、バギングに基礎を置く異なる決定木のアンサンブル学習である。本稿では弱学習器は 100 個とした。また、決定木を構築する際の基準には Gini 係数を用いた。Gini 係数は各ノードを分割する際に最良の基準として知られており、値が小さければ小さいほど、純度が高いことを示している。

3. 実験結果

信用格付けの予測精度

まず、訓練期間のデータ(2013年1月～2016年12月)から生成される 10 個の部分データ集合を用いて RF と MLR の 2 つのモデルを構築する。次に、試験期間のデータ(2017年1月～2017年7月)のデータを構築したモデルで信用格付予測を行う。説明変数は ESG スコアとする一方で、被説明変数は信用格付であることを再度注意しておく。実験結果について述べる前に、記号と精度を測る基準について定義しておく。

定義

n : 部分集合数

m : 格付け帯数

i : 部分集合, $i \in 1, 2, \dots, 10$

j : 格付け帯, $j \in AAA, \dots$

k : 業種, $k \in$ 金融業, 製造業, 非製造業

$Acc_{i,j,k}$: 部分集合 i , 格付け帯 j , 業種 k におけるモデル精度

$$Ave_{j,k} \triangleq \frac{1}{n} \sum_{i=1}^n Acc_{i,j,k}$$

$$TotalAcc_k \triangleq \frac{1}{m} \sum_{j=1}^m Ave_{j,k}$$

図表 3 は格付け帯全体で見た $TotalAcc$ であり、実験結果による各モデルの信用格付け予測精度を示したものである。ご覧の通り、RF の予測力はいずれの業種においても訓練期間、試験期間とも MLR を上回っていることが分かる。特に金融業においては、試験期間において、RF モデルは 47.38% の正答率である一方、MLR モデルは 17.03% とその差は歴然である。また図表 4 は各モデルにおける説明変数の重要度の数値を示している。

² ページ数制約があるため、本稿で Random Forest の詳細については割愛するが、詳細は Brieman (2001) などを参照頂きたい。またサンプルサイズの問題に対して考慮した Balanced RandomForest や Weighted Random Forest (Chen et al., 2004) も参考になる。

図表3 各モデルにおける精度の要約

	訓練期間における精度						試験期間における精度					
	金融業		製造業		非製造業		金融業		製造業		非製造業	
	RF	MLR	RF	MLR	RF	MLR	RF	MLR	RF	MLR	RF	MLR
Total Acc	75.51	33.54	43.61	31.46	64.75	50.03	47.38	17.03	42.98	23.11	23.41	17.72

(出所) 筆者作成.

図表4 各モデルにおける説明変数の重要度

変数	金融業	製造業	非製造業
E	78.74	64.38	111.63
S	54.18	80.55	90.63
G	65.93	75.70	96.38

(出所) 筆者作成.

CDS 市場の分析への応用

次に前小節で予測した格付け(2017年1月~2017年7月)を用いて, CDS 市場の分析への応用可能性の検討を行うため, CDS(終値)を被説明変数とする以下の回帰式(1)³を考える.

$$(1) \quad CDS_i = \beta_0 + \beta_1 \cdot DD \text{ 確率}_i + \beta_2 \cdot \text{株価}_i + \beta_3 \cdot \text{Dummy}_i \\ + \beta_4 \cdot DD \text{ 確率}_i \cdot \text{Dummy}_i + \beta_5 \cdot \text{株価}_i \cdot \text{Dummy}_i + \epsilon_i, \quad \epsilon_i \sim i.i.d. (\mu, \sigma^2).$$

ただし, ダミー変数は格付け帯により構造変化が起きているところとし, 具体的には, AAA~AA までは0, A~BBBを1とするような変数である. この背景には求められる信用リスクプレミアムはある格付け帯から傾斜がより急勾配になる特徴が知られているためである. またその他の変数について説明しておく, Distance to Default (DD)確率は Ching et.al (2012)に基づき算出した1年以内の個別企業のデフォルト確率, 株価は個別企業の株価(終値)とする⁴. ダミー変数については, 定数項ダミーと係数ダミーともに用いた. また, ダミー変数の格付け帯閾値の決定には, 信用格付けが既知としたときに最も自由度修正済み決定係数の高かったものを採用した. 具体的には, 金融業は AA と A の間, 製造業は A と BBB の間, 非製造業は A と BBB の間となった. また, 回帰式(1)よりダミー変数項を取り除いた回帰式に対してダミー変数を既知の状態分散分析を行った結果, いずれも 1%有意水準で帰無仮説を棄却し, 各格付け帯の閾値が適切な構造変化点であったことが統計的に有意であることが裏付けられた. 具体的には回帰式(1)において,

³ 回帰式の背景については, Tanaka, K., Higashide, T., Kinkyu, T., & Hamori, S. (2017)を参照されたい.

⁴ CDS データが期間 5 年物に対してデフォルト確率を 1 年以内としている点は, CDS の期間構造が, 直近起り得るイベントを警戒して逆イールドとなる場合があるため整合性を欠く面もあるが, 本稿は手法の提案に焦点を当てているため, 簡便的に行っている点を明記しておく.

図表5 CDSに対する各変数の影響力

変数	MLR			RF		
	金融業	製造業	非製造業	金融業	製造業	非製造業
切片	0.13 (-1.06)	0.03 (0.53)	-0.03 (-0.57)	0.16 (-1.12)	-0.08 (-1.79)	-0.02 (-0.28)
DD確率	-0.14 (-1.07)	0.03 (0.59)	0.45*** (8.46)	-0.23 (-1.48)	0.05 (0.53)	0.65*** (7.27)
株価	-0.59*** (-5.57)	-0.08 (-1.66)	-0.12* (-2.33)	-0.55*** (-4.07)	-0.09* (-2.00)	-0.05 (-1.05)
ダミー変数	-0.78 (-1.27)	83.31* (2.58)	0.07 (0.7)	-0.12 (-0.55)	1.51*** (7.75)	0.10 (1.25)
DD確率×ダミー変数	-0.36 (-0.57)	633.64* (2.58)	0.02 (0.29)	0.43 (1.88)	-0.19 (-1.72)	-0.23* (-2.28)
株価×ダミー変数	0.30 (0.87)	-0.01 (-0.04)	0.40*** (4.26)	-0.07 (-0.36)	0.28 (1.00)	0.39*** (3.59)
自由度修正決定係数	0.27	0.01	0.23	0.28	0.12	0.24
サンプルサイズ	79	473	455	79	473	455

(図表注1) 各変数の2段階括弧内数値はt値.

(図表注2) ***, **, *, はそれぞれ有意水準 1% 5% 10% \leq 0と有意差があることを示す.

(出所) 筆者作成.

帰無仮説 $H_0: \beta_3 = \beta_4 = \beta_5 = 0$, 対立仮説 $H_1: \beta_3, \beta_4, \beta_5$ のうち少なくとも一つは0でない, としてF検定を行った結果, すべての業種において有意水準を1%より大きくする限り, 帰無仮説を棄却し, 対立仮説を採択する結果となった. 以上のデータを基準化したものを用いて回帰式(1)に対して最尤推定法で係数推定を行った結果が図表5である.

まず, いずれの業種においても自由度修正済み決定係数はRFモデルに基礎を置いた回帰モデルがMLRモデルのそれを上回っていることが分かる. 次に回帰係数の統計的に有意か否かに目を移す. 例えば, 金融業においてはDD確率の変数が統計的に有意でない一方で, 非製造業でのDD確率は有意であるなど結果は区々である. 概していえば, 係数の有意性はある程度, 業種に依存しているように見えるが, 業種分類の方法が主因の一つであると考えられる. 具体的には, 本稿ではメガバンクもノンバンクも同じ金融業という業種に属しているなど, 標本数の問題を考慮した粗い分類にしている点である. またRFの信用格付け予測に基礎を置く回帰モデルでは, ダミー変数がいずれも有意であることがわかり, 係数格差が存在していることを示している. また図表6はRFモデルの説明変数のVIF値であり, いずれの数値からも多重共線性がないことは明らかである.

図表6 VIF値

変数	金融業	製造業	非製造業
DD確率	2.52	4.65	4.83
株価	1.99	1.04	1.34
ダミー変数	1.34	1.21	1.04
DD確率×ダミー変数	2.18	5.17	4.77
株価×ダミー変数	1.85	1.46	1.32

(出所) 筆者作成.

4. 結論

本稿は, ESGデータとRFを用いた信用格付けの予測モデルを構築する手法の提案を行ったとともに, 予測した信用格付けをダミー変数として用いた回帰式を介し, CDS市場分析への応用可能性の検討を行った. 信用格付け予測に関しては, RFとESGの親和性は良いことがわかり, 伝統的なMLRなどの手法に比べて予測精度が高いことが実験的に示唆された. 一方で, CDS市場への応用は, 例えばRFモデルに基礎を置いた非製造業に関する回帰モデルから可能性を感じるものの, 業種によって区々であることが伺える. おそらく業種分類の仕方によっても当然結果が変わり得るであろうが, その際問題になり得るのが業種別/格付け帯別による標本数の偏りの問題であろう. この点は特に課題の一つであり, 注釈1で記載したような対応以外にも例えば

転移学習を導入することなどによる対応も考えられる。今後は ESG データの有する非財務諸表の影響力が UNGC に加盟している企業数によって異なるのかを調べたい。

参考文献

- Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of banking & finance*, 18(3), 505-529.
- Anderson, R. (2007). *The credit scoring toolkit: theory and practice for retail credit risk management and decision automation*. Oxford: Oxford University Press
- Attig, N., El Ghouli, S., & Guedhami, O. (2013). Corporate social responsibility and credit ratings. *Journal of Business Ethics*, 117(4), 679-694.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Chen, Chao, A. Liaw and Breiman (2004). Using random forest to learn imbalanced data, University of California, Berkeley 110: 1-12.
- Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447-1465.
- Desai, V. S., Conway, D. G., Crook, J. N., & Overstreet, G. A. (1997). Credit scoring models in the credit union environment using neural networks and genetic algorithms. *IMA Journal of Mathematics Applied in Business and Industry* 8, 323-346.
- Devalle, A., Fiandrino, S., & Cantino, V. (2017). The Linkage between ESG Performance and Credit Ratings: A Firm-Level Perspective Analysis. *International Journal of Business and Management*, 12(9), 53.
- Gutsche, R., Schulz, J. F., & Gratwohl, M. (2016). Firm-value effects of CSR disclosure and CSR performance. In *EFMA-Conference proceedings* (pp. 1-31). EFMA.
- Lewis, E.M. (1992). *An introduction to credit scoring*. San Rafael: Athena Press.
- Mays, E. (2001). *Handbook of credit scoring*. Chicago: Glenlake Pub. Co. Fitzroy.
- Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29(2), 449-470.
- Tanaka, K., Higashide, T., Kinkyō, T., & Hamori, S. (2017). Forecasting the Vulnerability of Industrial Economic Activities: Predicting the Bankruptcy of Companies. *Journal of Management Information and Decision Sciences*, 20, 1-24.
- Yeh, C. C., Lin, F., & Hsu, C. Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems*, 33, 166-172.

- Zhong, H., Miao, C., Shen, Z., & Feng, Y. (2014). Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. *Neurocomputing*, 128, 285-295.
- Zeidan, R., Boechat, C., & Fleury, A. (2015). Developing a sustainability credit score system. *Journal of Business Ethics*, 127(2), 283-296.

本資料に関する著作権は、株式会社大阪取引所にあります。
本資料の一部又は全部を無断で転用、複製することはできません。
本資料の内容は、株式会社大阪取引所の意見・見解を示すものではありません。
本資料は、デリバティブ商品の取引の勧誘を目的としたものではありません。