

# オルタナティブデータを使った運用実務について

## ストラテジストレポートの自然言語処理と環境指標

統計数理研究所 リスク戦略解析センター 客員教授  
 青山学院大学大学院 国際マネジメント研究科 客員教授  
 ニッセイアセットマネジメント 投資工学開発センター長

**吉野 貴晶**



### 1. はじめに

近年、運用業界ではオルタナティブデータが注目されている。オルタナティブデータには明確な定義はない。伝統的に投資判断に使われてきたデータを代替して実務面での利用が進んでいるデータを指す言葉である。決算発表に先行して企業の経済活動を捉えるPOS(Point of sale)、クレジットカードデータから、環境報告書などに掲載される温室効果ガス(GreenHouse Gas：GHG)の排出量まで、

様々な種類のデータがオルタナティブデータと呼ばれる。本稿はオルタナティブデータの整理と足元の運用実務におけるモデル開発に向けた一端を紹介して、今後の発展や留意点を考察する。

オルタナティブデータが注目される背景として、データのユーザー側と情報ベンダーなどの提供者側の双方にとっての大きなニーズがある。ユーザー側となる資産運用者にとっては、主に以下の2点が理由となる。

第1に、伝統的な投資指標などは、アクティブ運用の超過リターン<sup>(1)</sup>の源泉に乏しくなってきたからである。情報通信技術の発展に伴いインターネットを通じて、広く一般の人まで速やかに企業の情報等が取得できるようになっている。これまで情報収集面で優位であった機関投資家等は、その優位性が乏しくなっている。フェアディスクロージャールール<sup>(注1)</sup>の適用も優位性の低下につながっている。そこで新しい超過リターンの源泉としてオルタ

#### 〈目次〉

1. はじめに
2. オルタナティブデータの整理
3. ストラテジストレポートの自然言語処理の例
4. 環境関連のデータを使った分析例
5. おわりに

ナティブデータが注目されている。

第2の理由は、新しい銘柄評価軸の尺度としてのニーズが高まっていることである。ESGを例にあげると次のようである。ESGとは、環境(Environment)、社会(Society)とガバナンス(Governance)の頭文字をとったものである。世界的に環境破壊や人権問題が大きくなり、また企業不祥事が相次ぐなかで、これらの問題の解消への取り組みを強める企業に投資することがESG投資である。2006年に国連でPRI(責任投資原則)が提唱され、ESG推進を「投資家の取るべき行動」と定義されたことも大きな契機となっている。近年は独立社外取締役の人数などが銘柄の選別基準などにも用いられている<sup>(注2)</sup>。ESG指標に関しては伝統的な投資に使う財務等のデータと異なることから、本稿ではオルタナティブデータの範囲として捉える。

一方、データの提供者側に関しては次の2つの理由がある。第1に、情報通信技術の発展に伴い一般の人が取得可能な情報の範囲が広がり、既存のデータでは厳しいビジネス環境となっていることである。ただ情報技術の発展はデータベンダー側にもメリットがある。収集可能な情報が増えていることや、例えば手入力などで作ってきたデータベースを自動化して効率化が高まっている。第2の理由は、これまで一部の組織などで管理されていたデータも取引されるようになっており、別の組織でも分析や活用されるようになったことである。代表的なものがPOSなどのトラ

ンザクション(transaction：取引)データである。

そして、データのユーザー側、提供者側に共通する要因としては、情報技術発展とも関連するが、①クラウドなどを利用して大量データの保管が以前と比べて可能になってきたこと、②AI(人工知能)技術などの分析手法や分析ツールの発展などもある。AI分析ツールの発展の流れの1つがAIのコモディティ化である。例えば、データ入力と予測対象を指定することで、様々なモデルから最適な手法を提案するソフトウェアも公開されている。また学習済みモデルも提供されており機械学習をせず、高度な知識やITスキルがなくても高レベルのモデル化が可能になりつつある。今後もAIのコモディティ化が進むなかでオルタナティブデータの活用が進むだろう。

本稿の構成は次のようである。第2節はオルタナティブデータの整理を行う。第3節はオルタナティブデータの活用事例として自然言語処理をつかったストラテジストレポートの解析の具体例を紹介する。第4節はESGのうちの1項目である環境情報に関するリターンとの有効性の検証結果を示す。第5節はまとめを示す。

## ■ 2. オルタナティブデータの整理

本章では、オルタナティブデータの整理を

(図表1) 主なオルタナティブデータ

| カテゴリー          | 内容                                     | データタイプ | 予測の対象等                                                     |
|----------------|----------------------------------------|--------|------------------------------------------------------------|
| 情報の優位性         | POSデータ                                 | 数値情報   | ・個別企業の業績動向を決算前に捉える。<br>・マクロ経済の視点から購買の状況を捉える。               |
|                | クレジットカードデータ                            | 数値情報   | ・個別企業の業績動向を決算前に捉える。<br>・マクロ経済の視点から購買の状況を捉える。               |
|                | ポイントカードのデータ                            | 数値     | ・個別企業の業績動向を決算前に捉える。<br>・マクロ経済の視点から購買の状況を捉える(ナウキャストイング)。    |
|                | サプライチェーン                               | 数値     | ・サプライヤーやカスタマーに対する業績や株式リターンの影響の伝播を捉える。                      |
|                | ウェブサイトで公表される企業の事業活動に関する属性のデータ：空室率、利用数等 | 数値     | ・個別企業の業績動向を決算前に捉える。<br>・マクロ経済の視点から購買の状況を捉える(ナウキャストイング)。    |
|                | スマートフォンのアプリに関するデータ、アプリを通じて得られるデータ      | 文字・数値  | ・個人の購買に関するデータを通じて個別企業の業績予想を行う。<br>・利用度合いから関連する企業の業績の予想を行う。 |
|                | 企業のウェブサイトの閲覧に関するデータ                    | 文字・数値  | ・企業の人気の度合いなどを推計する。                                         |
|                | 企業アナリストレポートの文字情報                       | 文字     | ・アナリストの企業評価を捉える。                                           |
|                | 企業が発信する業績情報                            | 文字     | ・個別企業の業績動向を決算前に捉える。                                        |
|                | ニュース                                   | 文字     | ・個別企業の環境、信用リスクや企業間ネットワークを推計する。<br>・市場全体のセンチメントを推計する。       |
|                | SNS                                    | 文字     | ・個別企業の環境や企業間ネットワークを推計する。<br>・市場全体のセンチメントを推計する。             |
|                | アンケートなどの情報                             | 文字     | ・個別企業の環境や企業間ネットワークを推計する。<br>・市場全体のセンチメントを推計する。             |
|                | 政府等がウェブサイト等を通じて発信する文字情報                | 文字     | ・マクロ経済の動向(ナウキャストイング)や今後のイベントを予測。                           |
|                | 位置に関する情報や衛星、ドローンなどの画像情報                | 数値・画像  | ・個別企業の事業活動などを捉える。<br>・マクロ経済の視点から経済活動を捉える(ナウキャストイング)。       |
|                | 企業や政府の重要人物の発言                          | 文字・音声  | ・個別企業の事業活動などを捉える。<br>・マクロ経済の動向(ナウキャストイング)や今後のイベントを予測。      |
|                | 企業や政府の重要人物の表情                          | 画像     | ・個別企業の事業活動などを捉える。<br>・マクロ経済の動向(ナウキャストイング)や今後のイベントを予測。      |
| 新たな企業や市場などの評価軸 | 企業が公表するESGに関する定量情報                     | 数値     | ESG評価を行う。                                                  |
|                | 企業が公表する環境報告書等の文字情報                     | 文字     |                                                            |
|                | ESGに関するニュース                            | 文字     |                                                            |
|                | 企業の口コミに関するデータ                          | 文字     | 株式需給の予測を行う。                                                |
|                | 株式保有に関する情報                             | 数値・文字  |                                                            |
|                | 信用取引に関する情報                             | 数値     |                                                            |
|                | 企業が公表する情報                              | 数値・文字  |                                                            |
|                |                                        |        | 企業文化(organization culture)を把握する。                           |

(出所)筆者作成

行う。前述したように運用者側にとって、オルタナティブデータの利用には大きく次の2つの趣旨がある。①資本市場の効率性が高まるなか、これまで運用実務に使われていた伝統的な情報が開示される前に、POSデータなどを使って企業の活動を捉えるというもの、②ESGなどの新しい企業の評価軸の尺度として利用するというものである。資産運用業界では、これらの趣旨に応じてオルタナティブ

データの利用に向けた研究・開発が行われている。実際のオルタナティブデータに関して、データ自体の特徴で分類すると構造化データと呼ばれる数値情報と、非構造化データとされる文字、音声や画像情報があり、運用実務への応用に向けた研究が進んでいる。

これらの観点から図表1では主なオルタナティブデータをまとめている。

オルタナティブデータの利用は、研究の歴

史が浅いことや、AIなどの分析手法も発展するなかで、未だ確立されていない。

本稿では、自然言語処理を用いたストラテジストレポートの解析と、環境に関する指標の評価の2つの手法に関して取り上げる。実務への応用可能性が高い注目の2つの解析事例である。

### ■ 3. ストラテジストレポートの自然言語処理の例

#### (1) 分析の趣旨

本節では、ストラテジストレポートの自然言語処理によるスコアの構築と、その精度の検証を紹介する。手法に関しては高度に技術的なものは使わない。ストラテジストレポートをシンプルに自然言語処理した場合に、どの程度、効果的かを観察するためである。

運用実務で利用が進んでいるものの1つが自然言語処理と言われる、テキスト(文書)情報の評価である。代表的な手法の1つに、証券会社などの企業アナリストが作成する企業レポートの自然言語処理を行い、個別企業に得点を与えるものである。この得点はレポートを著すアナリストのセンチメントスコアとも呼ばれる。スコアが高い銘柄に投資する戦略が見られる。本節では、こうした企業アナリストレポートで行われる分析手法を、市場全体の行方を予測するストラテジストレポートに応用する事例を紹介する。

企業アナリストレポート解析の背後には次

の根拠がある。投資家はアナリストの将来の利益予想に注目しているため、業績の予想値が修正されると株価も反応する。しかし近年、情報の効率性が高まりアナリストが業績予想を修正すると株価が瞬時に反応するため、修正後の投資では超過リターン確保が難しい。そこでアナリストが業績予想を修正する前に、レポートの文脈から将来の業績予想の姿勢を捉えようとするのである。数値情報である業績予想の修正は、投資家へのインパクトも大きい。アナリストは確信を高めてから踏み切る傾向がある。企業レポートの自然言語処理による得点化は、将来のアナリストの利益予想修正の予兆を捉える目的がある。

一方、一般にストラテジストレポートを使った自然言語処理の評価でのモデル化は見られない。これは①ストラテジストが相場水準の見通しを修正しても株式市場に大きな影響を及ぼすとは考え難いことや、このため②レポートの文脈で将来の相場水準変更の予兆を捉えても有用性に乏しいと見られること、そもそも③将来の相場予想水準の変更が事前の文脈に表れているとは考え難いことなどがある。そして④ストラテジストは各々が独特の表現を持つため、学習サンプルの扱いが難しいこともある。

しかしストラテジストは市場の専門家であり、知見も豊富に持っており、市場全体のセンチメントを敏感に捉えている可能性がある。こうした傾向がレポートの文脈に表れることで

---

市場予想が可能となるかもしれない。そこでストラテジストレポートの解析により相場予測スコアを求めて、株式市場の予測可能性を検証する。

## (2) センチメント算出と検証方法

本節で使用するストラテジストレポートは、ある特定の証券会社の1社のみとする。ストラテジストには特有の文章の癖があるため、個別毎に評価することが妥当と考えるからである。学習データに関しては2012年9月～2018年9月の期間に発行されたレポート(注3)を対象にモデル化する。そのモデルを用いて2018年10月～2019年9月までの1年間のレポートの相場予測スコアを算出して、それぞれのレポートのスコアがその後の相場をどの程度予見できるかを検証する。

具体的な分析の流れは次のようである。スコア作成の準備として学習データである2012年9月～2018年9月の期間のレポートを対象に、相場見通しの要約が示されているレポート1ページ目のみの文書を取得して分ち書きする(注4)。そして、名詞、動詞と形容詞のみを抽出した後、各学習レポートをDoc2Vecにより数値ベクトルに変換する(注5)。

次に、各レポートの数値ベクトルについて、レポートの発行日の翌日から5営業日(暦日ベースで1週間程度)後までの配当込みTOPIXリターンが上昇したサンプルと下落したサンプルに対応するものとして2つに分

類する(注6)。そして上昇サンプル、下落サンプルの、2つのカテゴリーのそれぞれで文書の数値ベクトルを平均する。これらの平均ベクトルを、それぞれポジティブ基準ベクトル、ネガティブ基準ベクトルと呼ぶ。

一方、検証期間における1年間のレポートの相場予測スコアの算出は次のようである。各レポートの品詞ベクトルを学習データでつくったDoc2Vecのモデルにインプットして文書の数値ベクトルを求める。そして、それぞれのレポートに対応する数値ベクトルについて学習データでつくった、ポジティブ基準ベクトル、ネガティブ基準ベクトルとの類似の程度(注7)を計算する。こうした処理で各レポートに、ポジティブ、ネガティブ基準ベクトルに対する類似の程度が求められる。そして最終的には、ポジティブ基準ベクトルからの類似の程度から、ネガティブ基準ベクトルからの類似の程度を減じて相場予測スコアとする。

この相場予測スコアがどの程度、相場を予測できているかを検証するために、次のような検証を行う。

各レポートに付与される相場予測スコアを3つに分類する。値が0.003以上のケースを上昇予測、-0.003以下を下落予測とした。これらの間のケースを中立予測とする。そして5営業日後までの相場との関係を検証する。

(図表 2) 相場予測スコアの検証結果

| 事前予測  | 平均値    | 中央値    | 上昇割合  |
|-------|--------|--------|-------|
| 上昇    | 0.04%  | 0.48%  | 57.9% |
| 中立    | -0.13% | 0.39%  | 57.7% |
| 下落    | -0.41% | -0.46% | 46.4% |
| 全サンプル | -0.17% | 0.26%  | 54.2% |

(注)2018年10月～2019年9月までの1年間のレポートを対象とする検証。上昇割合はそれぞれの予測に対して相場が上昇した割合。検証に用いるリターンは配当込みTOPIXを用いる。  
(出所)筆者作成

### (3) 検証結果

図表 2 に示される結果から、ある程度の相場予測精度が確認できる。

レポートから求めた予測スコアが相場上昇を予測しているケースでは、実際に5営業日後に平均で0.04%相場が上昇した。また上昇予測していたなかで、実際に相場が上昇した割合を見る「上昇割合」も57.9%と高かった。下落を予測する場合も精度が高かった。5立会日後に平均で-0.41%と下落した。上昇割合も46.4%と下落の状況が大きかった。ストラテジストレポートを使った自然言語処理の手法である程度の相場予測が可能となることが示された。

しかし、この分析ではストラテジストが相場を予想できていたとは言えない。それは次の追加的な分析結果による。Doc2Vecを用いる文書ベクトルの分析は、単語の組み合わせで文書の傾向を捉えるものである。これに対してより予測スコアの作成手法として分かり易い方法が、極性辞書を用いるものがある。レポートに使われているワードに関して、ポジティブとネガティブワードを、その程度(極

性値)を使って集計する方法である。極性値は学習レポートを使って分析者が作ることも可能であるが本稿は、東京大学 和泉・坂地研究室が開発、公開されて誰でも取得可能な金融極性辞書<sup>(注8)</sup>を用いる。このような辞書が公開されていることで分析者が機械学習をしなくてもモデル化が可能となる良く知られた例である。例えば、同辞書で「増大」に関する極性値は0.49である、対義語となる「減少」は-0.71となっている。ポジティブがプラスとなり、ネガティブがマイナスである。この辞書により、アウトサンプル検証期間の2018年10月～2019年9月までの1年間のレポートを用いて、次の分析を行う。それぞれのレポートを、分かち書きした後、五島・高橋(2017)で用いられる計算と同様に、それぞれの極性値を各レポートにおける出現数で加重合計する。

ポジティブ(ネガティブ)の極性値はプラス(マイナス)値のため、加重合計値がプラス(マイナス)の場合には対象レポートは相場の強気(弱気)を示唆するものとする。

こうして求めた各レポートの極性値の加重

(図表3) 極性値で求めた予測スコアの結果

| 事前予測  | 5日後    |       |       | 10日後   |       |       |
|-------|--------|-------|-------|--------|-------|-------|
|       | 平均値    | 中央値   | 上昇割合  | 平均値    | 中央値   | 上昇割合  |
| 上昇    | -0.40% | 0.07% | 51.2% | -0.42% | 0.06% | 51.2% |
| 中立    | -0.18% | 0.06% | 53.1% | -0.80% | 0.24% | 51.0% |
| 下落    | 0.08%  | 0.32% | 58.3% | 0.03%  | 0.45% | 57.1% |
| 全サンプル | -0.17% | 0.26% | 54.2% | -0.42% | 0.38% | 53.0% |

(注)2018年10月～2019年9月までの1年間のレポートを対象とする検証。上昇割合はそれぞれの予測に対して相場が上昇した割合。検証に用いるリターンは配当込みTOPIXを用いる。

(出所)筆者作成

合計を相場予測スコアとする。レポートのスコアに関して、3つに分類する。値が0.007以上のケースを上昇予測、-0.003以下を下落予測とする。そして、これらの間のケースを中立予測とする。そして5営業日後の相場との関係を分析した結果が図表3である。

極性値を使った相場予測スコアは相場予測精度が高くない。上昇予測しているケースでは、5営業日後の平均で-0.40%と下落している。逆に、下落予測をしているケースでは、平均で0.08%と上昇してしまった。また上昇割合も下落予測の方が上昇予測よりも高い。参考に10日後の結果も同様に示したが、5営業日後と同様に予測精度が低かった。

こうした分析結果から言えることは、ストラテジストレポートを使った予測ではポジティブ、或いはネガティブのwordを単純に集計するだけでは将来の相場予測が難しい傾向があることだ。レポートに表れるストラテジストの特徴的なクセなどの特徴が相場予測に有効となるかもしれない。直観的に言うと例えば、Doc2Vecで把握されたストラテジスト

の強烈な強気の相場見通しの場合には、むしろ相場下落の可能性が高いという可能性もある。

今回はある1社のレポートのみを分析対象としたケースを紹介した。複数社のレポートのそれぞれの分析を行い、それらを融合する活用も相場予測に効果的となるかもしれない。

## 4. 環境関連のデータを使った分析例

### (1) ESG指標と市場評価について

本節ではオルタナティブデータの分析例として、ESG指標について取り上げる。ESG評価に関しては環境報告書の自然言語処理などの研究も進んでいる。ただ環境報告書に定量的な情報も掲載されており、この情報をESG評価に使う流れが見られる。しかしESG投資は企業の社会的責任(CSR)の実現が実際の経済的価値につながることを重視しているだけに、企業側が開示している数値情報をESG評

---

価として使うには企業業績や株式リターンとの関連を考える必要がある。このためどのように実際の指標として活用したらよいか難しいテーマとなっている。本節では、ESGのなかの環境に関する指標を幾つか取り上げて、実際のリターンとの関係を検証する。

ESGに関する指標は大きく2つある。1つがESGの開示を示す指標である。もう1つがESGの活動を示すものである。前者のESG開示指標に関しては、企業経営の透明性と関連が強い。投資家にとって企業の透明性が高いことは、知らないことから発生するリスクが低下するため市場評価(株価)が高く形成されると見られる。BloombergのESG開示スコアと株式リターンとの関係を検証した代表的な検証には湯山他 [2019] がある。同研究では2017年以降のコーポレートガバナンス・コード及びスチュワードシップ・コード改訂に向けた動きなどを受けて、ESG銘柄が注目されたため、2017年にはESG開示スコアが高い銘柄の株式リターンは有意にプラスとなった。しかし、2013年以降の分析期間を通じてみるとESG情報開示と株式リターンとの関係は、必ずしも有意にポジティブと言えなかったことを示した。ESGに関する市場での評価を捉えることは難しいことが分かる。

一方、後者のESG活動を評価する観点、そのESG活動自体が企業価値を高めるという考え方である。例えばガバナンス意識が高いということは株主価値最大化への努力をする企業であるため、業績も良くなり株式市場で

も高い評価を得られるというものである。ESG活動の指標と企業のパフォーマンスに関する先行研究としてFriede et al. [2015] がある。同論文では1970年以降の2,200以上の論文をレビューした結果、おおむね9割以上の研究でESGと企業の財務パフォーマンスの関係はある程度のポジティブであったことを示している。一方、我が国を対象とした実証研究では伊藤 [2018] などがあり、ESG活動が株式リターンと正の関係であることを指摘している。本節では先行研究でも広く使われるBloombergのデータを用いて、我が国における環境に関して注目の指標と株式リターンとの関係を検証する。

## (2) 分析指標とその趣旨

ESGに関しては情報ベンダーから数百にも及び様々な系列のデータが提供されている。しかし、どの指標をどのように処理して企業評価尺度として使うべきかは確立していない。本節では、①CO2排出量、②廃棄物量と③水使用量の主な3系列を対象として、その後の株式リターンとの関係を明らかにする。分析対象期間として、リターンの起点を2005年9月とする。環境関連の項目は以前は公表されていなかったものが多く、過去データの取得は難しい。本分析で3指標に絞った理由の1つは、東証1部企業を対象にクロスセクションの比較である程度のデータが取得できる項目との観点で抽出したためである。そして、これらの3項目を選んだ趣旨は次のよう

である。

### 【① CO2排出量】

地球温暖化の原因の多くは、人間が活動することにより排出される温室効果ガスの増加とされる。その温室効果ガスの4分の3をCO2(二酸化炭素)が占める。地球温暖化対策の国際的な取り組みとして、2015年にパリで開催されていたCOP21(国連気候変動枠組条約第21回締約国会議)で2020年以降の温暖化対策の国際枠組みであるパリ協定が採択された。これを受け、日本の削減目標は2030年度に2005年度25.4%の削減とする。また、2019年6月に「パリ協定に基づく成長戦略としての長期戦略」も閣議決定された。政府が温室効果ガスの削減の中での我が国の成長の方向を示している。政策面での取り組みが進むなか、企業にも温室効果ガスの削減に向けた努力が求められている。企業側には温室効果ガスの開示もあるが、開示銘柄数はCO2排出量の方が多い。同排出量は温室効果ガスの多くの部分を占めることから、本稿ではCO2排出量を用いる。

### 【② 廃棄物量】

ゴミを燃やすことで温室効果ガスが発生することや、リサイクル製品が作られなかった場合に同等の製品を作るために必要な燃料を燃やすことで温室効果ガスが発生するため、ゴミの削減は地球温暖化対策として重要である。こうしたなか世界的に注目されている流れがサーキュラー・エコノミー(循環型経済)である。回収・リサイクル、省資源製品の開

発で新たな天然資源の使用と廃棄物を減らす社会の仕組みを作ろうというものである。サーキュラー・エコノミーの実現でゼロエミッション(ゴミなど廃棄物をいっさい出さない)を目指すことが、政府や企業、そして個人レベルでも期待されている。積極的にゴミを減らした企業を割り出すゴミ削減の努力は日本全体での数値目標がないことから、むしろ企業側の自主的な姿勢が大きく反映するテーマとも考えられる。

### 【③ 水使用量】

水不足は食料不足と並んで、将来に向けた世界的な大きな懸念の1つである。仮想水という考え方がある。その国が輸入している農畜産物や工業製品を、国内で生産するとした場合に必要な水のことである。日本は食料の多くを輸入しているため仮想水の輸入が多い。世界的な水不足により農産物が十分に供給できなくなれば、日本の農産物の輸入ができなくなり食糧不足となる。我々も水資源を大切にすることが必要である。こうしたなか水使用の削減を目標にする企業がある。排水浄化、雨水利用などで効率的に水を利用する試みである。こうした企業は環境への配慮が高い企業として評価できる。

これらの3指標を使った分析には注意が必要である。例えば単純にCO2排出が少ないとなると、会社の経済活動が小さい企業が有利になってしまう。この修正のため、会社全体の活動量を示すものとして実績売上高でデフレートする。

(図表 4) 5 分位ポートフォリオの分析結果

| 指標       | リターンズブレッド<br>(Q5-Q1)の平均値 |          | Fama-French 3 ファクターモデルのアルファ |      |        |       |       |       |        |       |        |       |              |         |
|----------|--------------------------|----------|-----------------------------|------|--------|-------|-------|-------|--------|-------|--------|-------|--------------|---------|
|          | 値                        | t値       | Q1(低)                       |      | Q2     |       | Q3    |       | Q4     |       | Q5(高)  |       | スプレッド(Q5-Q1) |         |
|          |                          |          | 値                           | t値   | 値      | t値    | 値     | t値    | 値      | t値    | 値      | t値    | 値            | t値      |
| ① CO2排出量 | -3.94%                   | -1.35    | 0.08%                       | 0.82 | 0.12%  | 1.45  | 0.09% | 1.03  | -0.03% | -0.47 | -0.25% | -2.19 | -0.33%       | -1.78** |
| ② 廃棄物量   | -5.79%                   | -2.13*** | 0.20%                       | 2.19 | -0.07% | -0.81 | 0.00% | -0.06 | 0.01%  | 0.11  | -0.13% | -1.03 | -0.18%       | -1.09   |
| ③ 水使用量   | -3.79%                   | -1.57*   | 0.03%                       | 0.34 | 0.08%  | 0.97  | 0.00% | -0.02 | 0.04%  | 0.52  | -0.15% | -1.41 | -0.18%       | -1.09   |

(注) 各月において各変数を基準に、分析対象銘柄をクロスセクションに変数の昇順に並び替えて、5分位のそれぞれで銘柄数ベースの等ウェイトポートフォリオを構築する。そして、Q5のロングとQ1のショートにより実現するポートフォリオがリターンズブレッドである。またファクターアルファの値は、5分位それぞれの翌月のリターン及び、リターンズブレッドを被説明変数、FF3ファクターを説明変数とする時系列回帰分析の切片項である。値は年率換算。\*\*\*、\*\*、\*はそれぞれ両側有意水準、5%、10%、15%を示す。

(出所) Bloombergをもとに筆者作成

### (3) 検証方法

本節では5分位ポートフォリオの手法で検証を行う。分析サイクルは月次とする。毎年8月末に前年度の環境関連指標が取得できるとして、8月末以降のリターンとの検証に用いる。毎月末に3つの環境指標を前年度の売上高で除して東証1部企業のデータ取得可能な企業をランキングして、銘柄数が等しくなるように5分位に分ける。そして各指標÷前年度実績売上高が低い順にQ1からQ5とする。また、各指標÷前年度実績売上高が高い分位(Q5)の銘柄群をロングし、各指標÷前年度実績売上高が低い分位(Q1)の銘柄群をショートするロングショートポートフォリオ(Q5-Q1)も求める。分析に使うポートフォリオのリターンは等ウェイトとする。また将来のリターンの関係に関して、これらの、ファクターモデルでは説明できない部分の存在を明らかにする。実際には次の分析を行う。分析の対象はQ1からQ5までの各々のリターンと、ロングショートポートフォリオ(Q5

-Q1)のリターンとする。そしてFama and French [1993] で提案された3ファクターモデルを用いて、市場リターンに関するファクター、時価総額に関するファクター(SMB: small minus big)と自己資本の簿価と時価の比率に関するファクター(HML: high minus low)でコントロールできないリターンがあるかを観察する(注9)。

### (4) 検証結果と考察

指標と、将来のリターンとの関係の分析結果を図表4に示す。マイナスに大きい方が何れの情報も売上高と比べて小さいということから、環境面ではポジティブに評価される。まず、シンプルなりターンズブレッドの平均値を見ると何れの3指標ともに、マイナスとなっている。これは環境関連指標のリターンとの負の関係が見られる。②総廃棄量は3%水準でマイナスとなって、有意となった。

Fama-French 3ファクターモデルでは説明できない部分の存在を明らかにするアルフ

アでは、スプレッドを見ると①CO2排出量は-0.33となり10%水準で有意となっている。その他の2指標も統計的に有意とはならなかったがマイナスのスプレッドとなり環境面での努力が高い企業の将来のリターンが高いことが示された。

## 5. おわりに

本稿は、資産運用業界で注目のオルタナティブデータの整理を行った後に、実際の運用実務への適用に関する実際を紹介した。

現在、オルタナティブデータに関しては、新しいリターンの源泉となる期待が高まっている。その一方で、次のような留意点もあり、研究や運用モデルの開発のハードルにもなっている。

- ①時系列方向のサンプルが不足しており、検証を充分に行うことが難しい点。
- ②取得できるデータが全体の一部分であるため、全体を把握することが難しい点。
- ③クレンジングなどの前処理に時間が必要であること。その完成度をどの程度まで求める必要があるかが問題となる。
- ④ビッグデータでは、データ取得やデータ蓄積に関するコストが高い。

分析ツールに関しては、AIがコモディティ化するなかで、データをインプットすれば難なくモデル化が可能となっている。このた

めスモールデータのケースでは、扱い易さが加速しよう。

現在オルタナティブデータやAI技術の活用に関して運用実務への適用の多くは、試行錯誤の段階でもある。オルタナティブデータやAIを用いたモデル化では、単純にモデルの当てはまりを見るだけでなく、モデル構築の背後にあるファイナンスの理論などとの整合性を重視する必要があるだろう。

※本稿を作成するにあたり、ニッセイアセットマネジメント 投資工学開発室のメンバーより有益なコメントを頂いた。ここに記し、感謝申し上げます。本稿に掲載されている内容や意見は筆者個人に属し、筆者が所属する組織の公式見解ではない。また、あり得べき誤りは、全て筆者に帰属するものである。

### 〔参考文献〕

- ・伊藤桂一 [2018] 「ESG評価と株価」, 加藤康之編著「ESG投資の研究—理論と実践の最前線—」, 一灯舎.
- ・五島圭一・高橋大志 [2017] 「株式価格情報を用いた金融極性辞書の作成」『自然言語処理』, Vol.24, No.4, pp.547-577
- ・竹原均 [2019] 「マルチファクターモデルの実証的比較—自己資本コスト推定への応用上の諸問題」, 『証券アナリストジャーナル』, Vol.57, No.3, pp.8-16.
- ・湯山智教・白須洋子・森平爽一郎 [2019] 「ESG開示スコアとパフォーマンス」, 『証券アナリストジャーナル』, Vol.57, No.10, pp.72-83.
- ・Carhart, M.M. [1997] “On Persistence in Mutual Fund Performance”, *Journal of Finance*, Vol.52, No.1, pp.57-82
- ・Fama, E. F., and K. R. French [1993] “Common Risk Factors in the Returns of Stocks and Bonds”.

---

*Journal of Financial Economics*, Vol.33, No.1, pp.3 – 56

- ・ Fama, E. F., and K. R. French [2015] “A five – factor asset pricing model,” *Journal of Financial Economics*, Vol.116, No.1, pp.1 – 22.
- ・ Friede, G., T. Busch and A. Bassen [2015] “ESG and financial performance : aggregated evidence from more than 2000 empirical studies,” *Journal of Sustainable Finance & Investment*, Vol.52, No.4, pp.210 – 233.
- ・ Hou, K., C. Xue and L. Zhang [2015] “Digesting Anomalies : An Investment Approach,” *Review of Financial Studies*, Vol.28, No.3, pp.650 – 705.
- ・ Ito T., Sakaji H., Tsubouchi K., Izumi K., Yamashita T. [2018] “Text – Visualizing Neural Network Model : Understanding Online Financial Textual Data.” In : Phung D., Tseng V., Webb G., Ho B., Ganji M., Rashidi L.(eds), *Advances in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science*, Springer, vol. 10939, pp. 247 – 259.

(注1) 2018年4月に施行された改正金融商品取引法により導入されたもの。上場会社等が、未公表の重要な内部情報を証券会社のアナリストなど特定の第三者に対して選択的に開示することを禁じる規制。

(注2) 例えば、東証及び日本経済新聞社が共同算出するJPX 日経インデックス400では「独立した社外取締役の選任」が銘柄選定に用いられている。

(注3) 学習サンプル数は1,000を超えているものの充分とは言えない。しかし、今回はシンプルなモデル化での効果を観察する。

(注4) MeCabを使って。品詞分解する。

(注5) PythonライブラリのGensimに実装されている。詳細は省略するが品詞ベクトルをインプットして数値ベクトルに変換することで類似度などの

計算が可能となる。

(注6) 上昇サンプル数と下落サンプル数に偏りがあるが特に考慮しない。

(注7) コサイン類似度

(注8) <https://sites.google.com/socsim.org/izumi-lab/tools/>金融専門極性辞書

(注9) 近年、Carhart [1997] の4ファクターモデルやHou et al [2015] のqファクターモデル、更にFama and French [2015] の5ファクターモデルも注目されている。しかし竹原 [2019] で行われた我が国の検証では、モデル説明力と切片項の有意性からすれば、Fama – French 3ファクターモデルでも十分と指摘していることから、本稿ではFF 3モデルを検証に用いる。

