

## COVID19 と機関投資家行動： 機械学習手法による検証

山田 拓弥  
(東京工業大学大学院)

木村 遥介  
(東京科学大学)

中田 和秀  
(東京科学大学)

井上光太郎  
(東京科学大学)

### 要 旨

環境や社会面の新たな情報に対し、ES 配慮の機関投資家はその他の投資家と異なる投資行動を行うのか。この疑問に答えるため、機械学習手法を用いて機関投資家の行動分析を行い、パンデミックに対して脆弱なポートフォリオを保有した投資家、中でも短期集中保有投資家がパンデミックに耐性のある銘柄への逃避をいち早く進めたことを明らかにした。

キーワード：ESG 投資、機関投資家、機械学習、社会的距離、COVID19

### 1 はじめに

2019 年に発見された COVID-19 を引き起こす病原菌は、その後急速に全世界に感染拡大し、2020 年 3 月には WHO がパンデミック（感染爆発）と評価した。Covid-19 パンデミックは、従来の感染症の拡大と比較して株式市場に非常に大きなインパクトを与えたとされ、その理由として、感染拡大を防ぐためのロックダウンや外出自粛に伴って経済活動が制限されることが指摘されている (Baker et al., 2020)。

企業の活動に対するパンデミックの影響は非常に多様であった。ハイテク産業の企業はテレワークを活用することで社会的距離（ソーシャル・ディスタンス）を十分に確保することで事業活動を継続することができたが、飲食業・旅行業などのように顧客と従業員間の密接な接触が必要な業種では、社会的距離を確保することが事業の継続に大きく影響を与えた。したがって、Covid-19 パンデミックは従来意識されていなかった「社会的距離を確保できるかどうか」という要素が、企業活動に対して影響を与えるという事実を知らしめた。

社会的距離に対して耐性を持つことは、自然災害によってオフィス等が倒壊しても、企業が事業活動を継続できることを意味し、自然災害に対する耐性や従業員の安全性として、ESG 投資の中の環境・

社会 (ES) 項目と関連するテーマ性を持つと考えられる。社会的距離に対する耐性・脆弱性は、パンデミック発生以前には投資家や ESG 評価機関の意識の外にあり、従来の環境・社会評価尺度には組み込まれていなかったが、パンデミックを契機として株価形成に織り込まれた。Pagano et al. (2023) は、COVID19 パンデミック以降の株価のパフォーマンスに対し、社会的距離に対する脆弱性が有意な影響を及ぼしていた一方、ショック前に測定した環境スコアは一貫して有意な影響を及ぼさなかったことを報告している。

それでは COVID19 パンデミック以降、全ての投資家が社会的距離を考慮して売買を行うようになったのだろうか？ 株式市場には、アクティブ・パッシブのような投資スタイルや短期・長期といった投資ホライズンなど、多様な投資家が存在していることが知られている。本稿では、COVID19 パンデミックによって、どのような特徴を持つ投資家が、企業の社会的距離に対する耐性という新たな要素に基づいて売買行動を行ったかについて分析を行う。また、従来の企業の環境・社会に対する配慮という非財務情報に基づいて投資を行う投資家 (ES 投資家と呼ぶ) が、COVID19 パンデミックを契機に社会的距離に対する脆弱性を取り入れたかについても検証する。

本稿では、社会的距離に対する耐性・脆弱性を表す指標として、Koren and Petó (2020) が提案した指標 (KP スコア) を利用する。この指標は、他者とのコミュニケーションや物理的近接という観点に基づいて、社会的距離に関するルールを設定することで、産業レベルでどれだけの従業員が「影響を受けるか」を割合で表している。よって、KP スコアが高いほど社会的距離を確保すること (すなわちパンデミック) に対して脆弱であり、KP スコアが低いほど社会的距離を確保すること (パンデミック) に対して耐性を有することを意味する。

前述のように、本稿の目的は、COVID19 パンデミックをきっかけとして、社会的距離への脆弱性を売買の意思決定に取り入れた投資家の特徴を見出すことである。しかしながら、経済ショックの最中に株式を売買する際の意思決定には、投資先企業と投資家自体の異質性に起因する様々な売買要因が互いに影響を及ぼし合っている。例えば、Cella et al. (2013) は、2008 年に生じた金融危機において、長期投資家に比べて短期投資家が保有銘柄を大幅に売り越し、ショックを増幅させたことを示している。また、Kim et al. (2021) は、機関投資家を投資ホライズンとポートフォリオ集中度の 2 軸で分類し、リーマンショックの株価急落期においてショックを増幅させたのは短期投資家の中でも分散したポートフォリオを持つファンドだったことを示している。また、Au et al. (2023) は COVID19 ショックにおいて、COVID19 の感染者数が多く、パンデミックに対する悲観度の高い地域を拠点とするファンドが保有株式を有意に多く売却することを示している。このように、機関投資家の特徴と売買行動の間には複雑な関係が存在していることが予想される。そこで本研究では、多くの先行研究で一般的な、事前に線形関係を定式化した回帰分析の手法とは異なり、あらかじめモデル (関数形) を特定することなく様々な非線形関係を捉えることが可能である機械学習の手法を採用し、様々な売買要因を同時に考慮した包括的な分析を行う。言い換えれば、要因間の作用を考慮するため回帰分析で必要となる交差項の組み入れを特定せず、様々な要因間の作用を考慮可能な機械学習手法を用いて機関投資家の特徴と売買行動のメカニズムを明らかにすることを試みている。

機械学習による分析の枠組みは以下の通りである。まず準備として、類似した特徴を持つファンドを同一のクラスターへ集約する。その後、売買銘柄情報からその株式を売買したファンドの所属するクラスターを予測する分類モデルの訓練を行う。最後に、分類に対して入力した特徴量の貢献度を定量的に説明する手法としてもっとも広く使用されている、Lundberg and Lee (2017) が開発した解析ツールの SHAP を分類モデルに適用し、検証期間中に各クラスターが重視していた売買銘柄情報を明らかに

することで、機関投資家の売買パターンの要因を解析する。

本稿が扱う機械学習は教師あり学習であり、入力と出力が存在する場合に利用される。教師あり学習が適用される状況は二つある。一つは新たな入力に対応する未知の出力を予測することであり、もう一つは入力・出力の間に存在する関係を見出すことである。本研究における教師あり学習は後者を目的としている。

以下で機械学習手法について簡単に説明しよう。ファンドの売買の特徴が与えられたときに、ファンドの特徴によって定義されるグループの中から、もっともらしいグループにファンドを割り当てられる分類問題を考える。例えば、KP スコアを重視して取引を行うファンドを検知するために、ポートフォリオ集中度の高低によってファンドを2グループに分け、売買銘柄のKP スコアによって、ファンドがどちらのクラスターに分類される確率が高いかを訓練する。この訓練によって、売買銘柄のKP スコアという入力が、クラスターの分類という出力に対して、どの程度影響を与えるかを定量化することが可能であり、この影響度合いの大きさによって、売買行動がKP スコアに影響を受けているクラスターを識別することを試みている。

分析の結果、事前ポートフォリオに含まれる銘柄の社会的距離に対する脆弱性 (KP スコア) によって、COVID19 ショック中の売買における異質性が観察された。具体的には、社会的距離に対する脆弱性が大きいポートフォリオを構築していたファンドほど、COVID19 ショック中にKP スコアの低い銘柄 (社会的距離に対して耐性のある銘柄) を選好し、逆に社会的距離に対して脆弱性が小さいポートフォリオを構築していたファンドほど同期間にKP スコアの高い銘柄 (社会的距離に対して耐性の小さい銘柄) を購入した。この結果は、事前にCOVID19 パンデミックに対して脆弱なポートフォリオを持っていた機関投資家はショックへの反応としてリスク回避的に行動し、反対にCOVID19 パンデミックに対して耐性のあるポートフォリオを持っていた機関投資家は、COVID19 パンデミックに対して脆弱な銘柄を購入する行動を選択したと考えられる。いかなる機関投資家もCOVID19 パンデミックを事前に予測することは不可能だったと仮定すると、社会的距離に注目したポートフォリオのリバランスは、COVID19 パンデミック直前に保有していたポートフォリオの社会的距離に関する偏りという偶発的な要因によって引き起こされたことがわかった。また、ファンドの運用スタイルによってもその売買に違いがあり、短期・集中保有型ファンドはKP スコアの低い銘柄を選好し、長期・分散保有型のファンドはKP スコアの高い銘柄を購入する傾向にあった。

事前ポートフォリオの社会的距離への脆弱性は、ES 投資家に対しても影響を与えていた。パンデミックの前四半期である2019年第4四半期にES スコアを重視した売買を行っていたファンドの中で、パンデミック前にCOVID19 パンデミックに対して脆弱なポートフォリオを構築していたファンドは、パンデミック中にES 評価よりもKP スコアを重視し、逆にCOVID19 パンデミックに対して強靱なポートフォリオを構築していたファンドは、同期間にKP スコアよりもES 評価を重視して株式を購入していた。

本研究の貢献は主に3点ある。第一は、COVID19 ショックにおける機関投資家の売買行動の実態を調査する上で、機関投資家をその性質ごとに区別した詳細な分析を行った点にある。種々の機関投資家を一団として扱った場合には見えない個々のファンドの投資行動を可視化し、Baker et al. (2020) や Affinito and Santioni (2021) などの先行研究を補完した。第二は、COVID19 パンデミックによって新たに登場した社会的距離への耐性・脆弱性という売買要因が、市場価格に織り込まれるように取引を行った主体が短期・集中保有投資家であることを示した点にある。これは、環境や社会に関する側面の市場価格への反映が、長期投資家として位置づけられるES 投資家でなく、むしろ短期投資家が

主導するものであり、市場効率性に対する短期・集中保有投資家の重要性を確認すると共に、従来の社会的枠組みに基づくESスコアに依存したES投資の不完全性を浮き彫りにした。第三は、機関投資家の売買要因に関する新たな分析手法を提案した点にある。機械学習手法を活用することで、非常に多数の異質性を持つファンドの個々の上場企業に対する複雑な売買パターンを分析することが可能になり、従来の方法では観測することが難しかった非線形性を含む投資行動の可視化が可能になる。

本稿の構成は以下の通りである。第2節では先行研究を紹介し、第3節ではデータ及び機械学習モデルを解説、第4節では分析手法について説明する。第5節では評価測定及び分析の結果を示し、第6節で結論を述べる。

## 2 先行研究

本研究はCOVID19パンデミックにおける投資家行動や株価変動の研究に関連している。機関投資家がES投資を行う重要な動機として、CSR評価の高い銘柄が株価急落時における資金の逃避先としての機能を持つことが挙げられる。Lins et al. (2017) は2008年の金融危機において、CSR評価の高い銘柄はそうでない銘柄と比べて株価の実現リターンが有意に高かったことを示している。COVID19ショック時には、Albuquerque et al. (2020) や Garel and Petit-Romec (2021) はそれぞれES評価およびE評価の高い銘柄の株価下落率が有意に低かったことを示しており、ES評価に配慮するような投資家の投資期間構造が株価に影響を与えることを示唆している。他方、Glossner et al. (2022) は、COVID19ショックにおいて機関投資家がES評価を重視していないという見方と整合的な取引行動を行っていることを示した。同様に、Bae et al. (2021) は、パンデミック時にCSR評価が株式リターンに影響を与えた証拠はないと報告している。

また、COVID19ショックの発生に伴い、企業運営における社会的距離という新たな投資テーマが出現した。Affinito and Santioni (2021) は、COVID19ショックにおいて機関投資家が社会的距離を重視していたかどうかを検証した。40カ国以上の地域を拠点とする2万以上の大規模ファンドによる売買動向を調査し、2020年3月において機関投資家がCOVID19の蔓延に対する脆弱性の高い銘柄（ヘルスケアや小売業など）から低い銘柄（ITや金融業など）へポートフォリオをリバランスしていたことが確認された。さらに、Pagano et al. (2023) は、COVID19ショックの株価急落期において、社会的距離に対する脆弱性の低い銘柄は高い銘柄と比較して有意に高い実現リターンを記録したことを示しており、COVID19パンデミックをきっかけとして社会的距離に対する脆弱性という要素が株価に反映されたことを示唆している。

以上のように、先行研究では機関投資家がCOVID19パンデミックによる株価急落時に資金の逃避を行う上で、パンデミック発生前の枠組みに沿って算出されたES評価ではなく、社会的距離という新たな環境側面に注目していたことが示されている。しかし、上記の2つの文献を含む先行研究では機関投資家全体の傾向を検証するに留まり、どのような特徴を持つ機関投資家が社会的距離を重視したポートフォリオのリバランスを積極的に行ったのかについては明らかにされておらず、ここに機関投資家の異質性と投資行動の関係を明らかにする研究意義があると考えられる。

### 3 データ及び分類モデル

#### 3.1 機関投資家情報

本節では、分析に使用したデータ、及び機械学習の分類モデルの詳細について解説する。まず、機関投資家の株式保有状況および属性については、筆者の研究室が、株主調査を主たる業務とする IR Japan より提供を受けた機関投資家保有状況データベースを使用している。IR Japan が作成する機関投資家のファンド別保有データは、国内外各ファンドの日本株式の保有状況を、投資信託運用報告書、米国 SEC に提出されるファンド運用状況報告書といった公開情報、複数の情報ベンダーのデータから収集し、四半期ごとにまとめたものである。また、機関投資家の拠点・運用スタイル・運用機関などの情報は、各運用機関や各国の金融庁及びその関連機関の web ページを基に作成されている。

各ファンドの売買情報は、各ファンドの四半期末の株式保有状況から作成した。ポートフォリオに含まれる任意の銘柄に対して 2019 年第 4 四半期末時点と 2020 年第 1 四半期末時点での保有株式数を観測し、増加していれば 2020 年第 1 四半期にその銘柄を購入し、減少していればその銘柄を売却したと見做す。

ファンドに関する特徴量は、属性と保有銘柄情報の 2 種類に分類できる。ファンドの属性は、基本的な投資戦略を決定づける代表運用スタイル（アクティブまたはパッシブ）に加え、先行研究でショック中の売買に影響を与えることが確認されたターンオーバー<sup>1)</sup>、ポートフォリオ集中度<sup>2)</sup>、国別 COVID19 センチメントを用いた (Cella et al., 2013; Kim et al., 2021; Au et al., 2024)。国別 COVID19 センチメントの指標として、McKinsey & Company<sup>3)</sup> が計測した消費者センチメントを使用した。本指標は、各国における一般消費者の、COVID19 パンデミック収束後の経済状況に対する悲観度を表しており、機関投資家の所在地における国別のセンチメントを対応させた。補論の表 A1 に、ファンドに関する特徴量の定義を示す。

ファンドの保有銘柄情報に基づく特徴量として、本研究の主題である KP スコア及び ES スコアに加えて、Fama and French (2015) の 5 ファクターモデルを参考に、時価総額、PBR、ROE、純資産変化率のポートフォリオウェイトによる加重平均を使用する。保有銘柄情報は、各ファンドの四半期末時点でのポートフォリオに含まれる全銘柄を参照し、その保有額加重平均値を用いた。

#### 3.2 保有銘柄情報

機関投資家の売買銘柄情報については、過去 3 ヶ月リターン、過去 1 年リターン、対 TOPIX の  $\beta$  値、売上高変化率、純資産変化率、売上高営業利益率、ROE、ROA、自己資本比率、流動比率、現金比率、PBR、PER、業種別 KP スコア、ES スコアを用いる。ここで、本実験で使用した ES スコアは COVID19 パンデミック以前に定義された ES 評価指標であることから、パンデミック以前の社会的枠組みに基づく ES 評価の代理変数と解釈できる。本研究の主題となる KP スコア及び ES スコア以外の変数として、多くの先行研究で使用されている機関投資家の投資意思決定に影響する特徴量を網羅的に選択した。各保有銘柄の財務情報及び ES スコアは Refinitiv Eikon から取得した。本研究で注目する社会的距離の指標には、Pagano et al. (2023) や Fahlenbrach et al. (2021) などの先行研究に倣い、Koren and Petó (2020) が提唱した KP スコアを用いる。この KP スコアは産業ごとの社会的距離(す

なわち COVID19 の蔓延) に対する脆弱性の高さを表しており、社員のテレワーク可能性をはじめとした他者との物理的な近接性を基に算出されている<sup>4)</sup>。各業種の KP スコアは Koren and Petó (2020) で使用されている北米産業分類システム (NAICS) (2012) から日本標準産業分類 (JSIC) (2013) へ変換した。

### 3.3 ファンドの分類モデル情報

個々のファンドの分類モデルとして、Chen and Guestrin (2016) が考案した XGBoost<sup>5)</sup> を使用した。XGBoost は勾配ブースティング決定木というアルゴリズムを利用したフレームワークであり、アンサンブル学習によって決定木の誤りを次々に修正して高い予測精度を実現する。分類モデルとして XGBoost のような決定木アルゴリズムを用いた理由は、分類に使用された全ての特微量に対して、分類への寄与度の大きさを測定することができる SHAP という機械学習モデルのブラックボックス・テストツールとの親和性が高く、ファンドの売買要因分析に適していると判断したためである<sup>6)</sup>。モデルにおける各種ハイパーパラメータについては、全てデフォルトの設定を用いた。

次節で検証方法を説明する前に、多項ロジットモデルのような伝統的な手法ではなく、機械学習による分類モデルを採用する理由を説明する。まず、本稿の機械学習モデルは、機関投資家の売買銘柄情報と属性を入力と出力として扱い、それら間の非線形関係を捉えることが可能である。つまり、複雑なデータ構造に対しても適用可能であり、線形関係を仮定するモデルと比較して高い分類精度を期待することができる。一方、多項ロジットモデルのようにパラメータに関する線形性を仮定したモデルでは、回帰係数の意味が明確なため、どの特微量がどの程度結果に影響を与えるかを直接理解することが可能であり、統計的仮説検定を通じて推定値の統計的有意性を評価することができる。しかし、機械学習手法ではこれらの評価ができないことがデメリットである。以上のように、多項ロジットモデルに代表される伝統的な分析手法と機械学習手法は、問題設定や結果の解釈可能性に応じて選択されるべきであり、一般的に機械学習手法が優れていることを主張することはできない。本稿では、特微量の影響を直接解釈することよりも、より正確な分類結果を得ることを重視し、機械学習手法を選択した。

## 4 検証方法

### 4.1 分析の全体像

本節では、COVID19 ショックの初期にどのような性質を持つ機関投資家が社会的距離に対する脆弱性を重視して売買を行っていたのかについて検証するための方法を説明する。本研究の分析は二段階から構成される。第一段階では、COVID19 ショック中に KP スコアを重視して売買を行ったファンドの特徴を調査する。具体的には、ショックが生じた 2020 年第 1 四半期における売買要因と事前期間である 2019 年第 4 四半期における売買要因を比較し、その差分をショック期間特有の売買行動として注目する。COVID19 ショックの発生によって日経平均株価が 2020 年の 2 月の下旬から急落し、3 月の最終週で上昇し始めた。このことから、2020 年第 1 四半期における売買を観測することで機関投資家による COVID19 ショックへの対応の初動を捉えると想定した。

第二段階で、当ショックの以前から ES 評価への関心の高かったファンドの中で、当ショック中に

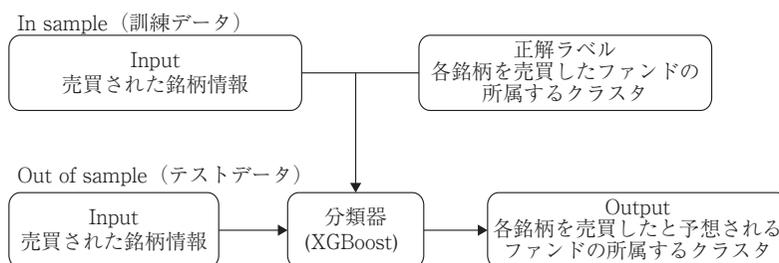
KP スコア重視の売買へ切り替えたファンドと、従来の社会的枠組みに沿って算出されたES スコアを重視し続けたファンドの性格の違いを検証する。はじめに、2019年第4四半期においてES スコアを重視して売買していたファンドを抽出し、ES 投資家と定義する。その後、第一段階と同様に2020年第1四半期における売買要因と事前期間における売買要因を比較することで、ショック中における売買の特徴を調査する。

本研究では、Affinito and Santioni (2021)をはじめとする先行研究に倣い、アクティブ投資家のみを分析対象とし、パッシブ投資家の売買に関する分析は行わない。

## 4.2 実験スキーム

本研究では、機関投資家の特徴がその売買に及ぼす複雑な関係を捉えるために、機械学習を用いた売買要因分析を行う。以下に、その具体的な手順を示すとともに、図1に本実験スキームの概要図を示す。

図1 本実験における分析フロー



出所：筆者作成

### 4.2.1 ファンドのクラスタリング

第一ステップとして、分析対象としている国内外の数千のファンドを全て個別に扱うと処理が煩雑になり、実験結果の一般化が困難になるため、類似した特徴を持つファンドを同一のクラスターへ集約し、ファンドの持つ情報を抽象化する。そのために、ファンドの性質を表す特徴量を用いて各ファンドを表すベクトルを作成する。次に、ポートフォリオ集中度や保有銘柄情報といった比率データを、中央値を閾値としてo(over)/u(under)の2値に変換する。代表運用スタイルやターンオーバーといったカテゴリデータといった二値データはそのまま使用する。連続変数だった特徴量を二値化してベクトルを置き換えてから、全く同じベクトル値を持つファンドを同一のクラスターへ分類する<sup>7)</sup>。同一のクラスターに属するファンドは、例えば、「集中保有・短期保有」や「分散保有・長期保有」のように同一の性質を持っていることを意味する。また、全ての特徴量が二値変数であるため、 $n$ 個の特徴量用いる場合、クラスター数は $2^n$ となる。

### 4.2.2 分類モデルの訓練

次のステップとして、銘柄情報を基に、その銘柄を売買したファンドが所属するクラスターを推定する分類モデルを作成する。まず、実現リターンや $\beta$ といった特徴量を基に、ファンドによって売買さ

れた銘柄を表すベクトルを作成する。その後、そのベクトルを入力、その銘柄を売買したファンドが所属するクラスタを正解ラベルとして分類モデルの訓練を行う。これにより、クラスタごとにファンドがどのような売買銘柄情報を重視していたのかをモデルに学習させることができる。

この分析では COVID19 ショックに特有の売買の異質性に注目するため、購入と売却の両側面について、ファンドによる売買行動が 2019 年第 4 四半期から 2020 年第 1 四半期にかけてどのように変化したのかについて注目する。それに伴い、実験では 2019 年第 4 四半期と 2020 年第 1 四半期、購入と売却のそれぞれのケースを調査するために合計 4 つの分類モデルを用意する。

### 4.2.3 SHAP を使用したモデルの解析

分類モデルの訓練及びテスト<sup>8)</sup>を終えた後、SHAP を用いてモデルを解析することで、各クラスタが売買においてどのような銘柄情報を重視していたのかを分析する。まず、COVID19 ショックにおいて KP スコアを重視して売買を行ったファンドをクラスタレベルで特定する。その後、それらのファンドに共通する特徴を抽出することで売買要因を解釈する。

以下に、SHAP を用いた解析の概要を説明する。この解析には、売買銘柄情報に含まれる各特徴量が、クラスタの予測に対してどの程度影響を与えたのかを表す SHAP 値を使用する。そのデータごとの SHAP 値を可視化した概念図が図 2 であり、『2020 年第 1 四半期における購入』のモデルにおける ES スコアに関する解析結果の例として 2 つのクラスタ(説明のため、クラスタ P とクラスタ N と呼ぶ)を掲載している。プロットされている各点は 1 つ 1 つがテストデータに対応しており、この図を参照することで、ES スコアの大小により、入力された銘柄がクラスタ P またはクラスタ N によって売買されたと予測される自信の度合いがどのように変化するかを直感的に理解することができる。この例では、黒い点が ES スコアの小さいデータ、白い点が ES スコアの大きいデータを表している。

図 2 『2020年第1四半期・購入』モデルのSHAPによるESに関する解析結果



注) 中央の点線を境目とし、右側はSHAP値が正の領域、左側はSHAP値が負の領域である。  
また、図にプロットされている点は1つ1つがテストデータに対応しており、白い点はESスコアが中央値以上のデータ、黒い点はESスコアが中央値未満のデータを表す。

まず、左側のクラスタ P の解析結果を見ると、SHAP 値が正の領域(右側)に ES スコアの高いデータ(白い点)が集中し、SHAP 値が負の領域(左側)に ES スコアの低いデータ(黒い点)が集中している。この結果は、分類モデルの内部において、テストデータに含まれる銘柄について、ES スコアが高いほどクラスタ P による購入だと推定される確率が他のクラスタと比較して相対的に上昇し、かつ ES スコアが低いほどその確率が減少することを表しており、より直感的に言えば、2020 年第一四半期の購入において、クラスタ P は ES スコアの高い銘柄(高 ES 銘柄)を选好し、ES スコアの低い銘柄(低 ES 銘柄)を回避したと解釈することができる。その一方で、右側のクラスタ N の解析結果では、白い点と黒い点がどちらも広域にプロットされており、それらの分布に偏りが見られない。この結果は売買した銘柄の ES スコアに応じた強い特徴が購入に現れていないことを表しており、このようなクラスタについては解釈が不可能となる。

このように、売買要因の解釈においてはプロットされたデータの特徴量の大きさとその分布が重要となる。この2つの項目について視覚的な判断ではなく、定量的な分析を行うため、以下のような手順でSHAPによる解釈を定量化した。まず、連続値をとっている各データのESスコアを、中央値を閾値としてダミー変数化した。次に、SHAP値の正負それぞれの領域ごとに、中央値以上の点と中央値未満の点を持つSHAP値の平均値をそれぞれ計算し、計4つの代表点を定める<sup>9)</sup>。ここで、領域ごとに処理を分ける理由は、0を境目として売買に対する態度が逆向きに変化するからである。それぞれの領域で、中央値以上と中央値未満のデータの集合を表す2つの代表点が離れているほど特徴量の大小による売買の差が強く現れているといえるため、この代表点間の距離を特徴の大きさの指標として用いた。

ここで、今回の実験ではSHAP値が負の領域に関する解釈は行わず、正の領域の解釈のみ行うこととした。その理由は、他のクラスタと比較して購入または売却を控えたというシグナルよりも逆の売買行動を積極的に行ったというシグナルの方が強力であることから、より信頼性の高い結果が期待できると判断したためである。最後に、全てのクラスタに対してSHAP値の正領域の代表点間の距離の大きさを測定し、その大きさが上位3分の1かつ内包するファンド数が第1四分位数である4以上のものを解析可能な強い特徴が現れたクラスタと定義した。ここで、内包するファンド数に下限を設けたのは、訓練データとテストデータの両方に十分な数の売買データが含まれているクラスタのみを解釈対象とすることで、解釈結果の信頼性を担保するためである。強い特徴が現れたクラスタのうち、中央値以上のデータのSHAP値の平均値が中央値未満のデータのそれよりも大きい(小さい)時、そのクラスタは注目している特徴量の大きさ(小ささ)を重視して売買していたと解釈することができる。

## 5 実験結果

### 5.1 ファンドに関する特徴量の選択

ファンドのクラスタ数は、使用するファンドに関する特徴量の数によって決定される。特徴量の数が多すぎると、ファンドをクラスタリングした際に1クラスタあたりに含まれるファンドが少なくなるため、各クラスタに所属するファンドの売買の特徴を捉えられなくなってしまう可能性がある。そこで、売買に与える影響の小さい特徴量を除外することで、実験で使用する特徴量の選択を行った。

売買に与える影響の大きさは、候補となる特徴量をそれぞれ単独で使用して本実験と同様の実験を行った際の分類精度によって測定する<sup>10)</sup>。さらに基準として、本論文で注目する特徴量であるKPスコアとESスコアの保有株平均による分類精度を上回る特徴量を本実験で使用する特徴量として選択する。ファンドに関する選択の方法の詳細は、補論A2に記載する。

結果として、KPスコアおよびESスコアの保有株平均に加え、代表運用スタイル(A/P)、ポートフォリオ集中度、ターンオーバー、保有株平均\_PBR、保有株平均\_純資産伸び率を特徴量として選択した。また、これらのファンドに関する基本統計量は表1の通りである。ここで、2019年第4四半期及び2020年第1四半期における解析では、前四半期末時点での銘柄情報及びファンドの属性情報を使用している。

表 1 記述統計

## パネルA ファンドの特徴量の比率データ

	N	Mean	Std.Dev.	Min	25%	50%	75%	Max
2019年3Q末								
ポートフォリオ集中度	2,033	0.19	0.28	0.00	0.02	0.06	0.22	1.00
保有株平均_純資産変化率	2,033	10.84	21.09	-7.11	4.71	7.92	10.51	235.88
保有株平均_PBR	2,033	2.36	1.29	0.44	1.46	2.23	2.83	16.07
保有株平均_KP	2,033	38.72	8.22	28.50	33.64	37.99	41.23	74.08
保有株平均_ES	2,033	54.99	11.86	9.58	49.73	56.01	61.05	84.54
代表運用スタイル	2,033	0.65	0.48	0	0	1	1	1
ターンオーバー	2,033	0.53	0.50	0	0	1	1	1
2019年4Q末								
ポートフォリオ集中度	2,012	0.19	0.28	0.00	0.02	0.05	0.22	1.00
保有株平均_純資産変化率	2,012	10.50	20.91	-7.11	4.59	7.24	10.05	235.88
保有株平均_PBR	2,012	2.78	1.82	0.48	1.60	2.53	3.28	21.40
保有株平均_KP	2,012	38.51	7.96	28.50	33.66	37.82	41.10	74.08
保有株平均_ES	2,012	57.56	12.02	13.42	52.40	58.78	63.86	88.86
代表運用スタイル	2,012	0.66	0.47	0	0	1	1	1
ターンオーバー	2,012	0.53	0.50	0	0	1	1	1

## パネルB 保有銘柄数・売買銘柄数の記述統計

	N	Mean	Std.Dev.	Min	25%	50%	75%	Max
保有銘柄数								
2019年3Q末	2,033	55.12	88.50	1.00	4.00	13.00	52.00	342.00
2019年4Q末	2,012	57.53	91.59	1.00	4.00	14.00	54.25	350.00
2020年1Q末	1,982	57.99	92.05	1.00	4.00	15.00	54.00	350.00
購入銘柄数								
2019年4Q末	1,170	29.24	56.92	1.00	2.00	6.00	22.00	336.00
2020年1Q末	1,087	32.25	64.83	1.00	2.00	7.00	25.00	348.00
売却銘柄数								
2019年4Q末	1,230	25.85	52.01	1.00	2.00	6.00	22.00	336.00
2020年1Q末	1,147	31.14	61.02	1.00	2.00	7.00	26.50	348.00

## 5.2 クラスタリング結果及びモデルの分類精度

前節で選択した特徴量に基づいたクラスタリングによって作成されたクラスタに関する基本統計量を表2に示す。ファンドに関する7つの特徴量は全て2値のみを取ることからクラスタ数は最大で $2^7=128$ 個であるが、2019年第4四半期では5つ、2020年第1四半期では3つのクラスタが欠損している。ここでの欠損は、これらのクラスタにファンドが分類されないことを意味している。欠損しているクラスタは全てパッシブファンド（クラスタ番号64~127）であり、本研究の分析対象はアクティブファンド（クラスタ番号0~63）のみであるため、この欠損が分析結果の妥当性に対して影響を及ぼすことはない。

表3には、それぞれの分析対象の四半期において、最低1銘柄の購入または売却を行ったファンド数、及び各モデルにおける分類精度を記載している。このうち、本稿の機械学習モデルを用いた場合の分類精度は、全てのケースで0.2から0.25の間の値を示した。クラスタを完全にランダムに予測する場合、全てのクラスタへ等しい確率で分類されるため、分類精度がクラスタ数の逆数(1/128)となること<sup>11)</sup>を考慮すれば、本分析の機械学習モデルの分類精度は、COVID19ショック中及びその事

表 2

	2019 年 第 4 四半期	2020 年 第 1 四半期
N	123	125
1 クラスタあたりのファンド数		
Mean	16.53	16.10
Std.Dev.	19.61	19.47
Min	1.00	1.00
25%	4.00	4.00
50%	9.00	8.00
75%	20.50	18.00
Max	103.00	102.00

表 3

分類モデル	売買した ファンドの総数	分類精度 (XGBoost)	分類精度 (MLR)
2019 年 4Q_ 購入	1,170	0.227	0.201
2019 年 4Q_ 売却	1,230	0.212	0.178
2020 年 1Q_ 購入	1,087	0.245	0.214
2020 年 1Q_ 売却	1,147	0.225	0.203

前期間において、機関投資家の売買パターンがその性質ごとに変化し、その特徴を本実験で構築した機械学習モデルがある程度正確に捉えられていることが分かる。また、同様の分類を多項ロジスティック回帰による結果と比較すると、機械学習モデルを用いた場合の方が平均して約 3% 分類精度が高い。したがって、本実験で使用した機械学習の分析スキームは、広く用いられている手法である回帰分析より正確に機関投資家の売買の特徴を捉えており、機械学習を機関投資家の売買要因分析に用いる合理性が明らかとなった。

### 5.3 SHAP の解析結果

本節では、4 つの分類モデルを SHAP で解析した結果を示す<sup>12)</sup>。最初に 5.3.1 節では COVID19 ショック中に KP スコアを重視して売買を行ったファンドの特徴を示し、次に 5.3.2 節では当ショックの以前から ES 評価への関心の高かったファンドの中で、当ショック中に KP スコア重視の売買へ切り替えたファンドや、従来の ES スコアを重視し続けたファンドの特徴を示す。なお、本研究では COVID19 ショック期間中の ES 投資に強い影響を与えたと考えられる ES スコアと KP スコアの重視度についての考察に限定し、その他の売買銘柄情報の重視度についての考察は割愛する。

#### 5.3.1 社会的距離を重視したファンドの解析

COVID19 ショック中の売買の特徴を捉えるために、その直前の四半期である 2019 年第 4 四半期における解析結果を示す。表 4 パネル A は、2019 年第 4 四半期において KP スコアを重視した売買を行っていたクラスタを示している。ここで、KP スコアを重視した売買を行ったクラスタは、KP スコアに関する SHAP 値の正領域の代表点間の距離の大きさを測定し、その大きさが全クラスタの中で上位 3 分の 1 かつ内包するファンド数が第 1 四分位数である 4 以上のクラスタとして定義した (4.2.3 節

参照)。

パネル A の「購入：高 KP 重視」の行は、高い KP スコアに注目して購入を行ったクラスタの特徴を表している。高 KP 重視のクラスタは全てで 10 個あり、そのうちターンオーバーが高であるクラスタが 8 個、ポートフォリオ集中度が高いクラスタが 5 個、保有株平均\_純資産成長率が高いクラスタが 5 個、保有株平均\_PBR が高いクラスタが 5 個、保有株平均\_KP が高いクラスタが 7 個、保有株平均\_ES が高いクラスタが 5 個であったことを表している。

KP スコアを重視した購入を行ったクラスタの全てが高 KP 銘柄に反応しており、低 KP 銘柄を選好したクラスタは存在しないという結果になった。ターンオーバー列を見ると、10 個のクラスタのうち 8 個が、ターンオーバーが高い短期投資家クラスタであることが分かる。また、保有株平均\_KP 列を見ると、事前に高 KP 銘柄を多く保有していたクラスタが 10 個中 7 個を占め、事前ポートフォリオに含まれる銘柄の特徴と類似した銘柄を購入するという一貫した投資戦略が結果として現れている。

売却側のモデルについても、同期間の購入側のモデルと同様に、KP スコアを重視した売却を行った全てのクラスタが高 KP 銘柄に反応していた。『保有株平均\_KP』列を見ると、事前に高 KP 銘柄を多く保有していたクラスタが 12 個中 11 個を占め、基本的に多くのファンドにおいて空売りは行われない中で、保有している銘柄しか売却できないという当然の事実を反映している。

以上のように、2019 年第 4 四半期には、高 KP 銘柄が類似した銘柄を多く保有するファンドによって特徴的に売買されていた一方で、KP スコアの低さを重視して売買していたクラスタは確認されなかった。このように、社会的距離が重要な投資テーマとして認識されるきっかけとなった COVID19 ショック以前に、KP スコアの大小による売買の偏りが生じている原因として、KP スコアと ES スコアが負の相関を持つことが挙げられる（全分析対象期間における全データを用いて測定された相関係数：-0.36）。5.3.2 で後述するように、2019 年第 4 四半期では高 ES 銘柄は事前ポートフォリオの状態に関わらず様々なファンドによって売買され、低 ES 銘柄は事前に低 ES 銘柄を多く保有するファンド間で循環しており、この ES スコアに対する選好が KP スコアに関する分析結果に現れたと解釈することができる。

上記の事前期間の分析結果を踏まえ、2020 年第 1 四半期における売買の特徴を解釈する。表 4 パネル B に、2020 年第 1 四半期において KP スコアを重視した売買を行っていたクラスタを示している。

購入について、「ポートフォリオ集中度」列を見ると、KP スコアの低さを重視して購入していたファンドは全て集中保有投資家であったことがわかる。さらに、「ターンオーバー」列を見ると短期保有投資家が 6 個中 5 個を占めることがわかる。また、「保有株平均\_KP」列から、これらのクラスタは全て事前に高 KP 銘柄を多く保有していたことがわかり、そのような特徴を持つファンドが KP スコアの高さを重視して購入していた前四半期とは明らかに異なる傾向が得られた。一方、KP スコアの高さを重視して購入していたファンドにおいては、「ポートフォリオ集中度」列を見ると、7 個中 4 個を分散保有投資家が占めている。また、「保有株平均\_KP」列を見ると、事前に KP スコアの高い銘柄を持っていたクラスタが 7 個中 2 個、すなわち、事前に KP スコアの低い銘柄を持っていたクラスタが高い KP スコアを重視して購入するという事前期間とは逆の傾向を示した。

売却に関する分析結果について、KP スコアの低さを重視して売却していたクラスタが全クラスタ中 1 個だけ存在している。しかし、1 個のクラスタでは特徴を見出すことが困難なため、この結果について解釈を行わない。一方、KP スコアの高さを重視して売却していたクラスタについては、『保有株平均\_KP』列を見ると、事前に高 KP 銘柄を多く保有していたクラスタが 13 個中 11 個を占めてい

表4：KPスコアを重視して売買したクラスタ

パネルA：2019年第4四半期においてKPスコアを重視して売買していたクラスタ

選好	クラスタ数	代表運用スタイル	ターンオーバー	ポートフォリオ集中度	保有株平均純資産成長率	保有株平均PBR	保有株平均KP	保有株平均ES
購入：低KP重視	0	-	-	-	-	-	-	-
購入：高KP重視	10	アクティブ	8 (0.8)	5 (0.5)	5 (0.5)	5 (0.5)	7 (0.7)	5 (0.5)
売却：低KP重視	0	-	-	-	-	-	-	-
売却：高KP重視	12	アクティブ	4 (0.33)	5 (0.42)	5 (0.42)	3 (0.25)	11 (0.92)	5 (0.42)

パネルB：2020年第1四半期においてKPスコアを重視して売買していたクラスタ

選好	クラスタ数	代表運用スタイル	ターンオーバー	ポートフォリオ集中度	保有株平均純資産成長率	保有株平均PBR	保有株平均KP	保有株平均ES
購入：低KP重視	6	アクティブ	5 (0.83)	6 (1.00)	2 (0.33)	3 (0.50)	6 (1.00)	2 (0.33)
購入：高KP重視	7	アクティブ	2 (0.29)	4 (0.57)	4 (0.57)	2 (0.29)	2 (0.29)	4 (0.57)
売却：低KP重視	1	アクティブ	0 (0.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	0 (0.00)
売却：高KP重視	13	アクティブ	7 (0.54)	7 (0.54)	4 (0.31)	4 (0.31)	11 (0.85)	7 (0.54)

注) 列名の変数が大きい値を示すクラスタの数を表し、括弧内の数値はすべてのクラスタ数に対する比率である。

る。この傾向は、事前期間を対象とした解析結果と一致しているため、COVID19 ショック中特有の行動であると断定することはできない。

以上の結果をまとめると、事前に高KP銘柄を多く保有していたファンドは低KP銘柄を選好して購入していたのに対し、事前に低KPスコアの銘柄を多く保有していたファンドは高KPスコアの銘柄を選好して購入していた。この結果から、事前にCOVID19パンデミックに対して脆弱な銘柄を多く保有していたファンドほどリスク回避的な行動を取った一方、事前にCOVID19パンデミックに対して強靱な銘柄を多く保有していたファンドは、パンデミックに対して脆弱な銘柄を買い支えたと解釈することができる。また、社会的距離に対する姿勢はファンドの投資戦略とも関連しており、上記のようなリスク回避の中心には短期・集中保有投資家が、リスク選好の中心には分散保有投資家、特にその中でも長期投資家の存在があったことが明らかになった。Kim et al. (2021) は、リーマンショックを分析対象とした実証分析において、短期集中保有投資家が持つ高い情報生産能力を指摘している。その情報優位性を活かし、社会的距離を組み込んだ新しいES投資の概念を構想した短期・集中保有投資家は、Affinito and Santioni (2021) において機関投資家全体の特徴として現れていた「パンデミックに対して強靱な銘柄への逃避」を先導し、市場の価格形成の役割を担っていたと考えられる。

また、いずれの分析結果においても『保有株平均ES』列に特徴的な傾向<sup>13)</sup>が見られなかったことから、事前ポートフォリオにおけるESスコアの大きさは、KPスコアを重視した売買に対して影響を及ぼさなかったと考えられる。この結果は、COVID19 ショック以前からESを配慮していたファンドが、その他のファンドに対してパンデミックに対する耐性が同程度であったことを表しており、従来の枠組みに沿って算出されたESスコアに依存したES投資はCOVID19パンデミックに対するリ

スクを低減できておらず、既存の社会的枠組みに基づいた ES 評価の限界が浮き彫りになっている。

### 5.3.2 ES 評価を重視したファンドの解析

本節では、前節と同様に 2019 年第 4 四半期において ES 評価を重視して売買していたファンドの特徴を調査し、その過程でショック以前から積極的に ES 投資を行っていたファンドを特定する。表 5 パネル A に、2019 年第 4 四半期において ES スコアを重視した売買を行っていたクラスタを示している。

購入側について、『ポートフォリオ集中度』列より、ES スコアの高さを重視して購入していたクラスタのうち 15 個中 13 個を集中保有投資家が占めている。さらに、『ターンオーバー』列を見ると、その中でも 13 個中 10 個が短期投資家であることがわかる。また、『保有株平均\_ES』列に注目すると、ES スコアの高さを重視して購入していたクラスタは、15 個中 10 個と若干事前に高 ES 銘柄を多く保有していたクラスタに偏っている一方、ES スコアの低さを重視して購入していたクラスタは、4 個中 3 個が事前に低 ES 銘柄を多く保有していたクラスタである。この結果は、高 ES 銘柄は事前ポートフォリオにおける ES スコアの大きさに限らず、比較的幅広いファンドに購入されているのに対して、あえて低 ES 銘柄を購入するのは事前に類似した銘柄を多く持つファンドが中心であることを示唆している。

売却側について、ES スコアの高さを重視して売却していたクラスタは 2 つのみであり、『ポートフォリオ集中度』列および『ターンオーバー』列を見ると、どちらも短期集中保有投資家であることがわかる。また、『保有株平均\_ES』列より、2 つの ES スコアの高さを重視して売却していたクラスタはどちらも事前に高 ES 銘柄を多く持つファンドであることが確認された。一方、ES スコアの低さを重視して売却していたクラスタについては、『ポートフォリオ集中度』列および『ターンオーバー』列に偏りが見られない。また、『保有株平均\_ES』列を見ると、10 個中 8 個が事前に低 ES 銘柄を多く持つファンドであることがわかり、ES スコアの高さを重視して売却していたクラスタについての解析結果と同様に、事前ポートフォリオと類似した銘柄を売却するという行動が確認された。

以上の解析結果をまとめると、以下の通りとなる。COVID19 ショック以前では、高 ES 銘柄は事前にそのような銘柄を多く保有するファンドによって集中的に売買されていた傾向にあるものの、そうでないファンドによる購入も一定程度確認された。ES 投資は、基本的に長期保有を前提とされているため、売却側では短期集中保有投資家の存在が目立っている。その一方で、低 ES 銘柄は高 ES 銘柄とは異なり、事前に類似した銘柄を多く持っていたファンド間で循環していたことを示唆している。この分析結果は、Nofsinger et al. (2019) が示した、一部の機関投資家が企業の ES 評価をネガティブスクリーニング材料として用いているという考察と整合的である。

上記の事前期間の分析結果を踏まえ、2020 年第 1 四半期における売買の特徴を解釈する。表 5 パネル B に、2020 年第 1 四半期において ES スコアを重視した売買を行っていたクラスタを示している。

購入側について、『ターンオーバー』列より、ES スコアの高さを重視して購入していたクラスタのうち 17 個中 11 個を短期保有投資家が占めている。それ以外については偏りが小さく、高 ES 銘柄は幅広いファンドに購入されていることが示唆される。ES スコアの低さを重視して購入していたクラスタは 5 個存在するが、『ポートフォリオ集中度』が高いクラスタが 4 個存在し、そのことから集中投資を行うファンドが ES スコアの低さに注目して購入していたことが確認された。

売却側について、ES スコアの高さを重視して売却していたクラスタは 11 個あり、『ポートフォリ

表5：ESスコアを重視して売買したクラスタ

パネルA：2019年第4四半期においてESスコアを重視して売買していたクラスタ

選好	クラスタ数	代表運用 スタイル	ターン オーバー	ポートフォ リオ集中度	保有株平均 純資産成長率	保有株平均 PBR	保有株平均 KP	保有株平均 ES
購入：低ES重視	4	アクティブ	2 (0.50)	2 (0.50)	1 (0.25)	1 (0.25)	3 (0.75)	1 (0.25)
購入：高ES重視	15	アクティブ	10 (0.67)	13 (0.87)	6 (0.40)	4 (0.27)	5 (0.33)	10 (0.67)
売却：低ES重視	10	アクティブ	4 (0.40)	3 (0.30)	6 (0.60)	4 (0.40)	5 (0.50)	2 (0.20)
売却：高ES重視	2	アクティブ	2 (1.00)	2 (1.00)	0 (0.00)	1 (0.50)	1 (0.50)	2 (1.00)

パネルB：2020年第1四半期においてESスコアを重視して購入していたクラスタ

選好	クラスタ数	代表運用 スタイル	ターン オーバー	ポートフォ リオ集中度	保有株平均 純資産成長率	保有株平均 PBR	保有株平均 KP	保有株平均 ES
購入：低ES重視	5	アクティブ	2 (0.40)	4 (0.80)	2 (0.40)	2 (0.40)	3 (0.60)	3 (0.60)
購入：高ES重視	17	アクティブ	11 (0.65)	7 (0.41)	8 (0.47)	9 (0.53)	8 (0.47)	8 (0.47)
売却：低ES重視	4	アクティブ	4 (1.00)	2 (0.50)	1 (0.25)	2 (0.50)	2 (0.50)	2 (0.50)
売却：高ES重視	11	アクティブ	6 (0.55)	10 (0.91)	2 (0.18)	4 (0.38)	4 (0.36)	8 (0.73)

注) 列名の変数が大きい値を示すクラスタの数を表し、括弧内の数値はすべてのクラスタ数に対する比率である。

オ集中度』列を見ると、集中保有投資家であることがわかる（11個中10個）。また、『保有株平均ES』列より、8個のクラスタが事前に高ES銘柄を多く持つファンドであることが確認された。一方、ESスコアの低さを重視して売却していたクラスタについては、『ターンオーバー』列より、全てのクラスタが短期保有投資家であることがわかる（4個中4個）。それ以外の項目については偏りが小さく、特徴を見出すことは難しい。

ESスコアに注目した以上の分析での重要な発見は、2019年第4四半期においては低ES重視の売却クラスタが多かったが、2020年第1四半期では高ES重視の売却クラスタが多くなったことである。COVID19ショック以前には、事前ポートフォリオのESスコアの特徴と一致した売買を行っているが、COVID19ショック以降では、事前ポートフォリオのESスコアが高いクラスタにおいて、高ESスコアの銘柄を売却する傾向が強く観測されるようになっている。

## 6 結 論

本研究では、環境や社会への人々の認識変化を伴う市場ショックが機関投資家のES投資戦略にどのような影響を与えたのかについて検証した。先行研究では、従来までES評価の優れた銘柄に備わっていた市場ショックにおける資金の逃避先としての機能の一部が、COVID19ショック以降では社会的距離の大きい銘柄、すなわちパンデミックに対して強靱な銘柄によって代替されていたことが示されている（Pagano et al., 2023）。一方、本稿の分析の結果、COVID19ショックの最中で社会的距離を重視した売買を行っていたのは、事前にCOVID19パンデミックに対して脆弱なポートフォリオを保

有していたファンドであることが判明した。また、その中でも、Affinito and Santioni (2021)において機関投資家全体の傾向として観測されていた「パンデミックに対して強靱な銘柄への逃避」を牽引し、積極的に価格形成に貢献していたのは短期集中保有投資家であることが示された。

企業のCSR活動を評価する上で、さまざまな機関が発表しているES評価のスコアリングを参照することは有効だが、COVID19パンデミックのような未曾有の事態により社会構造に変化が生じた場合、従来のES評価方法は市場の関心と乖離する。そして、COVID19ショックにおいてその変化への対応を左右していたのは、主に保有ポートフォリオのCOVID19パンデミックに対する脆弱度という偶発的な要因であり、短期投資家はその価格形成を牽引していた。すなわち、ESへ配慮している投資家でも、新たな種類の環境や社会面のリスクには対応できていないことが示唆される。この点で、投資家は市場でまだ認識されない環境、社会面のリスクを幅広く織り込むES評価指標を必要としている。

### 【補論】

論文の補論は、日本経営財務研究学会の公式ホームページに掲載している。

### 【付記】

本誌編集委員長の阿萬弘行先生（関西学院大学）と2名の匿名のレフェリーより有益なコメントを頂いた。また、株式会社IR Japanより機関投資家の株式保有状況データをご提供いただいた。ここに記して謝意を示したい。なお、本稿における誤りは全て筆者に帰属する。

### 【注】

- 1) ターンオーバーは、各ファンドの全観測期間における保有株の変動係数を正規化した数値を用いて定義されている。ファンドによって観測期間は異なるが、サンプルの開始時期は2004年、終了時期は2021年であり、複数期間のファンドの保有株数の変動を元に計算されている。この変動係数が高いほど頻繁に取引を行うファンド（短期投資家）、変動係数が低いほど取引が頻繁ではないファンド（長期投資家）として解釈することが可能である。
- 2) ポートフォリオ集中度は、ファンドのポートフォリオウェイトに基づくHerfindahl-Hirschman Index (HHI) で定義される。すなわちポートフォリオ集中度が高い場合、そのファンドのポートフォリオがいくつかの銘柄に集中していることを表し、ポートフォリオ集中度が低い場合、ポートフォリオが分散していることを表す。
- 3) <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/a-global-view-of-how-consumer-behavior-is-changing-amid-covid-19>, “Consumer sentiment and behavior continue to reflect the uncertainty of the COVID19 crisis”, October 26, 2020
- 4) KPスコアは産業レベルの変数であるため、企業間のばらつきに対する説明力を持たないことはこの変数の限界である。
- 5) [https://github.com/dmlc/xgboost/blob/62ed8b5fef01d960b5e180b6c3ab170b5f7a85d2/doc/python/python\\_api.rst](https://github.com/dmlc/xgboost/blob/62ed8b5fef01d960b5e180b6c3ab170b5f7a85d2/doc/python/python_api.rst)
- 6) XGBoostの他に、Ke et al. (2017) が考案したLight GBMや、Dorogush et al. (2018) が考案したCatBoostといったフレームワークにおいても勾配ブースティング決定木が利用されているが、今回の実験では、Jabeur et al. (2021) や、Pinelis and Ruppert (2022) をはじめとするSHAPによる要因分析の実用性を示した文献が豊富に存在するXGBoostを使用した。
- 7) クラスタ数を数種類に留めた方が結果の解釈を行いやすいものの、複雑な意思決定構造を持つファンドの売買の特徴をより詳細に捉えるため、ファンドの特徴量 $n$ 個を使用した2<sup>n</sup>個のクラスタを作成した。クラスタリング手法については、k-means法や階層型クラスタリング法では解釈が不可能な結果しか得られなかったため、各特徴量の中央値を閾値とした分類を採用した。
- 8) 本実験では、訓練データ数：テストデータ数が3：1になるよう、全売買データをランダムに選択した。
- 9) 領域ごとに代表点を取得すると、どちらかの領域にデータが偏った場合に少ない方の信頼性が損なわれてしまう

恐れがある。しかし、実際に全てのクラスタの解析結果における分布を確認したところ、6割を超えるデータがどちらかの領域に集中するクラスタは存在せず、また比較的偏りの大きい数個のクラスタを除いて解析を行っても結果には影響しなかった。

- 10) 精度指標は適合率と再現率の調和平均であるF値を採用する。F値が高いほど分類精度が高いことを意味する。
- 11) 多クラス分類問題において、N個のカテゴリに等確率でデータを割り当てるランダムクラス分類器の分類精度は以下のように計算される。各カテゴリにデータポイントを割り当てる確率は $1/N$ 、任意のデータポイントが正しいカテゴリに分類される確率も $1/N$ である。分類精度は全データポイントに対する正確なカテゴリ予測の平均確率として計算されるため、この分類器の分類精度は $1/N$ となる。
- 12) 補論において、すべてのクラスタに対するSHAPの解析結果を掲載している。
- 13) 例えば、KPスコアを重視して売買していたクラスタに関する分析結果を見ると、『保有株平均\_ES』の列では、KPスコアを重視する売買を行う機関投資家のクラスタの半数のESスコアが高く、半数のESスコアが低い結果を示しており、KPスコアを重視する売買を行うクラスタの中にESスコアに関する傾向がないことを意味している。

### 【引用文献】

- Affinito, M., & Santioni, R., 2021. When the panic broke out: COVID19 and investment funds' portfolio rebalancing around the world. Temi di discussione (Economic working papers), Bank of Italy, Economic Research and International Relations Area.
- Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C., 2020. Resiliency of environmental and social stocks: An analysis of the exogenous COVID19 market crash, *Review of Corporate Finance Studies* 9, 593-621.
- Au, S., Dong, M., & Zhou, X., 2024. Does social interaction spread fear among institutional investors? Evidence from Coronavirus Disease 2019. *Management Science* 70 (4), 2406-2426.
- Bae, K. H., El Ghouli, S., Gong, Z. J., & Guedhami, O., 2024. Does CSR matter in times of crisis? Evidence from the COVID-19 pandemic. *Journal of Corporate Finance*, 67, 101876.
- Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T., 2020. The unprecedented stock market impact of COVID-19. *Review of Asset Pricing Studies* 10 (4) 742-758.
- Cella, C., Ellul, A., & Giannetti, M., 2013. Investors' horizons and the amplification of market shocks. *The Review of Financial Studies* 26 (7), 1607-1648.
- Chen, T., & Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 785-794.
- Dorogush, A. V., Ershov, V., & Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.
- Fahlenbrach, R., Rageth, K., & Stulz, R. M., 2021. How Valuable Is Financial Flexibility When Revenue Stops? Evidence from the COVID-19 Crisis. *The Review of Financial Studies*, 34 (11), 5474-5521.
- Fama, E. F., & French, K. R., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116 (1), 1-22.
- Garel, A., & Petit-Romec, A. 2021. Investor rewards to environmental responsibility: Evidence from the COVID-19 crisis. *Journal of Corporate Finance* 68, 101948.
- Glossner, S., Matos, P., Ramelli, S., & Wagner, A. F., 2022. Do Institutional Investors Stabilize Equity Markets in Crisis Periods? Evidence from COVID-19 (September 3, 2022). Swiss Finance Institute Research Paper No. 20-56, European Corporate Governance Institute - Finance Working Paper No. 688/2020.
- Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L., 2021. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 1-21.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 3149-3157.
- Kim, D., Kim, H. D., Joe, D. Y., & Oh, J. Y. J., 2021. Institutional investor heterogeneity and market price dynamics: Evidence from

- investment horizon and portfolio concentration. *Journal of Financial Markets* 54, 100604.
- Koren, M., & Pető, R., 2020. Business disruptions from social distancing. *Plos one* 15 (9), e0239113.
- Lins, K. V., Servaes, H., & Tamayo, A., 2017. Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance* 72 (4), 1785-1824.
- Lundberg, S. M., & Lee, S. I., 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768-4777.
- Nofsinger, J. R., Sulaeman, J., & Varma, A., 2019. Institutional investors and corporate social responsibility. *Journal of Corporate Finance* 58, Pages 700-725.
- Pagano, M., Wagner, C., & Zechner, J., 2023. Disaster resilience and asset prices. *Journal of Financial Economics* 150 (2).
- Pinelis, M., & Ruppert, D. (2022). Machine learning portfolio allocation. *Journal of Finance and Data Science* 8, 35-54.
- Yoo, S., Keeley, A. R., & Managi, S., 2021. Does sustainability activities performance matter during financial crises? Investigating the case of COVID-19. *Energy Policy* 155, 112330.