

2023年7月19日 全8頁

ESG 評価の相違をいかに理解するか

「集約による混乱」の解消に向けた質的アプローチの重要性

金融調査部 研究員 中 澤

[要約]

- ESG 評価の重要性が高まる中、評価機関によって同一企業に対する評価が大きく異なる実態が明らかとなり、評価の妥当性や信頼性が疑問視されるようになった。一方、ESG は複雑で多様な概念であり、その評価が異なること自体は必ずしも問題とはいえない。本稿では、最近の実証研究の結果を検討することで、主に方法論の観点から ESG 評価に相違が生じる要因について理解を深めることを目的とする。
- Berg, Kölbel, and Rigobon (2022) “Aggregate Confusion: The Divergence of ESG Ratings” に基づくと、ESG 評価の相違は「範囲 (scope)」、「測定 (measurement)」、「ウェイト (weight)」に要因分解できる。同論文によると、6つの評価機関について一対の組み合わせごとに ESG 評価の相違を要因分解した結果を平均すると、相違の 56%が「測定」で説明される。残りの 38%が「範囲」、6%が「ウェイト」で説明され、「測定」による相違が大きいことが分かる。
- ESG 評価のプロセスとは情報を集約するプロセスであり、その方法論をめぐってはさまざまな選択がある。どのような選択をするかの違いが集約のプロセスの中で積み重なることによって、ESG 評価と総称されるものを全体として見た場合に、ESG 評価機関の個々の評価の間に相違が生じるものと理解できる。
- もっとも、ESG 評価の相違は評価機関の方法論のみに起因するものではなく、企業の情報開示の不足に起因する面もある。評価機関と利用者（主に機関投資家）の間のみならず、評価機関と企業との間でいかなるコミュニケーションが図られていくのかも、ESG 評価のこれからを考える際には無視できない点であろう。今後は、ESG 評価の妥当性や信頼性を確保するために、対話のような質的なアプローチを補完的に組み合わせることの重要性が増すのではないだろうか。

1. ESG 評価をめぐって

環境（「E」）、社会（「S」）、ガバナンス（「G」）要因を投資判断において考慮する ESG 投資の拡大に伴って、ESG 評価機関や ESG データ提供機関の役割が大きくなっている。企業に対する ESG 評価（ここでは便宜上、「AAA」や「BB+」といったアルファベットや記号で分類される「格付」

や、数値で表される「スコア」を特に区別せずに「ESG 評価」と呼ぶ)、およびそのために用いられる ESG データは、複雑で多様な ESG の諸要因を捉えるためのものとして、機関投資家の投資判断や企業の取組みのベンチマークとして活用されている。しかし、ESG 評価・データの重要性が高まる中、評価機関によって同一企業に対する評価が大きく異なっている実態が明らかとなり、評価の妥当性や信頼性が疑問視されるようになった。

このような見方が広まっていることは、ESG 評価・データ提供機関に対する規制強化の動きに見ることができる。証券監督者国際機構 (IOSCO) は 2021 年、規制当局、ESG 評価・データ提供機関、利用者、そして評価対象とされる企業に向けて 10 の提言を行っているが、それらのうち 7 つは ESG 評価・データ提供機関に向け、評価の方法論を中心に手続きの透明性や利益相反の管理・低減、情報開示等を求めるものである¹。日本では、早くも 2022 年に金融庁が「ESG 評価・データ提供機関に係る行動規範」を日本語と英語で公表し、(1) 品質の確保、(2) 人材の育成、(3) 独立性の確保・利益相反の管理、(4) 透明性の確保、(5) 守秘義務、(6) 企業とのコミュニケーションの 6 原則を定めた²。

2023 年 6 月、欧州委員会は、欧州連合 (EU) 域内において ESG 評価を提供する法人に対して欧州証券市場監督局 (ESMA) の認可を求める規則案を出しており、一般原則には評価機関の独立性、デュー・デリジェンスの実施、検証可能な方法論の継続的な使用など、評価機関の組織体制から評価の実務まで幅広い内容が盛り込まれている³。同規則案の付属書 III では方法論についての情報開示が詳細に規定されており、最低限の開示項目として、方法論の概要、データ処理の概要、方法論の科学的根拠の有無およびその根拠の用いられ方、評価の目的、範囲、ウェイト、データソースの限界など、11 項目が定められている⁴。これらの動向からは、ESG 評価の方法論とデータに課題があると認識されていることがうかがえる。

一方で、IOSCO の提言は利用者に対しても、評価の方法論やデータの精査を行い、用いようとする ESG 評価が「何を評価しているのか」を把握するよう求めている。主な利用者が債券投資家に限られる信用格付とは異なり、ESG 評価は特定のアセットクラスを想定したものではない。多面的な広がりを見せる ESG 要因のうち何を重視するかも機関投資家によってさまざまである⁵。各評価機関の評価が ESG 要因として何を評価しているのか、評価の相違が何によって生じているのか、利用者の側も理解することが重要である。

近年、このような問題意識は実務のみならず学術の世界においても共有されており、ESG 評価の相違を解き明かそうとする実証研究が大きく展開している。そこで本稿では、最近の実証研

¹ IOSCO, 2021, “Environmental, Social and Governance (ESG) Ratings and Data Products Providers Final Report”

² 金融庁, 2022, 「ESG 評価・データ提供機関に係る行動規範」および “The Code of Conduct for ESG Evaluation and Data Providers”

³ European Commission, 2023, “Proposal for a Regulation of the European Parliament and of the Council on the transparency and integrity of Environmental, Social and Governance (ESG) rating activities”

⁴ European Commission, 2023, “Annexes to the Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the transparency and integrity of Environmental, Social and Governance (ESG) rating activities”

⁵ 太田珠美, 2023, 「[ESG 格付けは今後選別されていくのか](#)」大和総研レポート (2023 年 4 月 17 日付)

究を取り上げ、その結果を検討することを通じて、主に方法論の観点から ESG 評価の相違に対する理解を深めたい。

2. ESG 投資の拡大と実証研究の進展

ESG 評価の問題が学術的に議論されるようになったのは、必ずしも最近のことではない。筆者が確認する限り、実証研究は遅くとも 2000 年代半ばから存在している。このような研究の展開を大きく進めたのが、マサチューセッツ工科大学のビジネススクールである MIT Sloan School of Management の Sustainability Initiative の下で行われている The Aggregate Confusion Project である。同プロジェクトは、より厳格で信頼できる ESG インテグレーション (ESG 要因を投資・運用プロセスに組み込むための手法) の開発を目的としている。マサチューセッツ州年金基金投資運用委員会 (Mass PRIM) をはじめ 5 つの機関が参画しており⁶、日本からはアセットマネジメント One が加わっている⁷。

ESG 評価の相違を対象とした研究はこれまでに 2 つ出版されている (1 本は *Review of Finance* 誌に掲載、もう 1 本は European Corporate Governance Institute のワーキングペーパーでプレプリントサーバにある) が、「どのような相違を問題とするか」をめぐって研究のアプローチは異なる。1 つは、同一企業に対する評価が、評価機関の間で異なるという意味の相違に注目する研究である (Berg, Kölbel, and Rigobon 2022)⁸。

もう 1 つは、同一時点の同一企業に対する評価が、データの利用時点で異なるという意味の相違を指摘するものである (Berg, Fabisik, and Sautner 2021)⁹。この意味での評価の相違とは、評価の「書き換え (rewriting)」である。これは、ある時点における評価の方法論の変更がその時点よりも過去の評価にも適用されてしまうことを指す。著者らによれば、このような変更 (データの書き換え) が、未発表で、繰り返し行われているという。いずれの「相違」も ESG 評価をめぐって重要な問題ではあるが、本稿では、同一企業に対する評価の違いに迫った前者の研究に注目し、その結果を検討する¹⁰。

⁶ Mass PRIM とアセットマネジメント One の他、MFS Investment Management、AQR Capital Management、Qontigo の 5 機関である。MIT Sloan School of Management の提供で PR Newswire が配信するプレスリリースで確認できる (URL: <https://www.prnewswire.com/news-releases/mit-sloans-aggregate-confusion-project-an-initiative-to-improve-esg-measurement-in-the-financial-sector-welcomes-four-new-investment-firms-as-founding-members-301408334.html>)

⁷ 同社の開示資料において、これに関する記載がある。アセットマネジメント One 「サステナビリティレポート 2022」を参照。

⁸ Florian Berg, Julian F. Kölbel and Roberto Rigobon. 2022. “Aggregate Confusion: The Divergence of ESG Ratings.” *Review of Finance* 26(6): 1315–1344, <https://doi.org/10.1093/rof/rfac033>.

⁹ Berg, Florian, Fabisik, Kornelia and Sautner, Zacharias. 2021. “Is History Repeating Itself? The (Un) Predictable Past of ESG Ratings.” *European Corporate Governance Institute Finance Working Paper* 708/2020. <http://dx.doi.org/10.2139/ssrn.3722087>.

¹⁰ 同一企業に対する ESG 評価が評価機関によって異なることは、(1) ESG 評価の主目的である評価をむしろ困難にする、(2) 企業にとって ESG パフォーマンスを向上させる誘因を低下させる、(3) ESG パフォーマンスが資産価格に与える効果を分散させる、(4) ESG パフォーマンスと役員報酬の連動を困難にさせる、(5) ESG 評価を用いた実証研究の結果を大きく左右してしまう、など、その影響が金融市場から実体経済まで広範囲に及ぶことで、特に問題が大きいと考えられる (Berg et al. 2022)。

3. 実証研究を読み解く

(1) ESG 評価の相違を生む「範囲」「測定」「ウェイト」

Berg et al. (2022) は、ESG 評価の相違をもたらす要因を分析するため、主に 2014 年のデータを用いて 6 つの評価機関 (KLD、Moody's ESG、MSCI、Refinitiv、S&P Global、Sustainalytics) に共通するサンプルとして 924 社の企業に対する ESG 評価を再現した (以下、「同論文」とはこの研究を指す)。その方法は次の通りである。

まず、6 つの評価機関から提供された 709 項目の指標に対し、共通の分類基準を用いて 64 の「カテゴリ」を付与し¹¹、再分類する。次に、評価機関ごとにカテゴリの平均値を求め、「カテゴリスコア」とする。そして、このカテゴリスコアを説明変数、ESG 評価を被説明変数とした回帰分析を行い、分析結果の回帰係数を求める。これらの結果を基に、ESG 評価の相違を、評価に用いられるカテゴリ（「範囲 (scope)」）による部分、カテゴリスコアの算出に用いられる指標（「測定 (measurement)」）による部分、そして前述のカテゴリスコアの回帰係数（「ウェイト (weight)」）による部分の 3 つに要因分解し、それぞれの寄与率を算出する。

重要なのは、ここでの要因分解の結果が、あくまでも再現であり、同論文の筆者らが独自に設定した分類基準、算出したカテゴリスコアの回帰係数に基づいているという点である。頑健性や再現性の確認は行われているものの¹²、同論文を読む上では、これらの分類基準の設定方法をめぐって異論や他の選択があり得ること、実証の方法が異なれば結果も異なったものになる可能性があることを念頭に置いておく必要がある。

図表 1 は、同論文を基に、ESG 評価がどのような段階を経て構成されるのか、どの段階において「範囲」「測定」「ウェイト」それぞれの相違が生じるのかを単純に図示したものである。最も下位（図表 1 の右）に位置するのが「指標」で、これが評価を構成する最小単位のデータである。まず、これらを上位（図表 1 の中央）の「カテゴリ」へと集約するが、このとき、同一の名称を付与されたカテゴリであっても用いられる指標が評価機関によって異なることで生じるのが「測定」による相違である。例えば、「労働慣行」というカテゴリを捉える指標として「離職率」を用いる評価機関もあれば、「労働関連の訴訟の数」を用いる評価機関もあり、これら下位指標が共通していないといった場合が考えられる。

次に、カテゴリをさらに集約して最終的な ESG 評価（図表 1 の左）を構成するが、このとき、評価機関によって用いられるカテゴリと用いられないカテゴリがあることで生じるのが、「範囲」による相違である。先述の「労働慣行」を例にとると、ESG 評価を構成するにあたって「労働慣行」を用いる評価機関と用いない評価機関があるということである。

そして、カテゴリから ESG 評価へとさらに集約するとき、どのカテゴリをどのくらい重要と考えるのかの違いが、「ウェイト」の違いである。例えば、ESG 評価に「労働慣行」を組み入れ

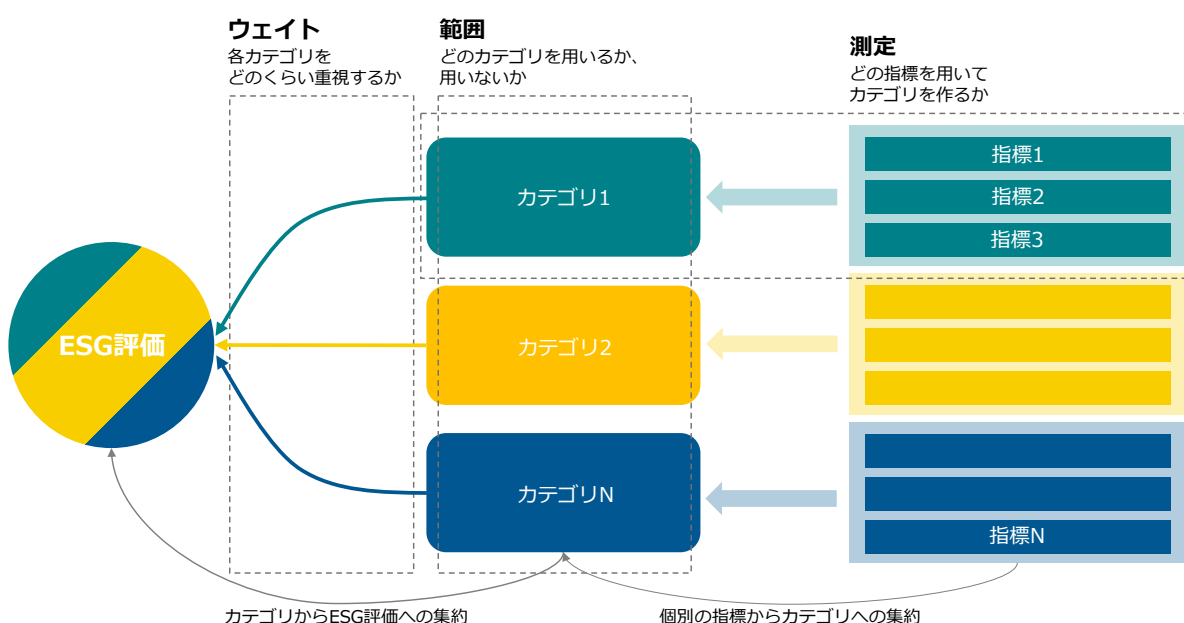
¹¹ 1 つの評価機関でしか用いられず再分類できないカテゴリは「未分類 (Unclassified)」とされている。

¹² この点について Berg et al. (2022) は、別の分類基準を用いる、異なる回帰分析の手法を用いるなどして頑健性を確認している。再現性の指標としては回帰分析の決定係数 (R^2) が用いられており、推計式は 79～99% (評価機関によって異なる) の説明力を持つとされている。

ることが2つの評価機関に共通していたとしても、一方の評価機関は全体の5%、もう一方の評価機関は全体の10%の比率で考慮しているといった場合がこれに該当する。

6つの評価機関について、一対の組み合わせごとに相違を要因分解した結果を平均すると、ESG評価の相違は、その56%が「測定」の相違によって説明される。残りの38%が「範囲」、6%が「ウェイト」によって説明され、「測定」による相違が大きいことが分かる。カテゴリ別に見ると、「測定」の相違は「気候リスクマネジメント」、「製品安全性」、「コーポレートガバナンス」、「汚職」、「環境マネジメントシステム」のカテゴリにおいて特に大きい。

図表1 ESG評価の構成と「範囲」・「測定」・「ウェイト」による相違のイメージ



出所：Berg et al. (2022) を基に大和総研作成

(2) ESG 評価機関ごとの方法論の特徴

Berg et al. (2022) は、ESG 評価の相違の要因だけでなく、それぞれの要因がどのくらい寄与するのかまで実証的に明らかにしたことに大きな意義がある。これに加えて、筆者はこの論文のもう1つの意義として、実証分析の過程で各評価機関の特徴が明らかになっている点が興味深いと考える¹³。例えば、「範囲」、「測定」、「ウェイト」の要因分解の結果も、評価機関ごとに見れば大きく異なっており、各評価機関の方法論の違いがよく表れている。

特徴が際立つのは MSCI である。同社と他の5社間の ESG 評価の相違をそれぞれ要因分解

¹³ 同論文では、評価機関の固定効果を推計し、それを「評価機関効果 (rater effect)」と呼んでその内実も議論しているが、解釈に留まる面が大きいようであるため本稿ではあえて取り上げなかった。「評価機関効果」とは、企業ごとのカテゴリスコアを規定する要因として、企業とカテゴリを統制してもなお残る、評価機関に固有の効果である。Berg et al. (2022) は「評価機関効果」を一種のバイアスであると考えており、あるカテゴリにおいて高い評価を得られた企業は、他のカテゴリにおいても高い評価を得られやすいのではないかと議論している。しかし、「評価機関効果」について、評価機関の何がそのような効果を生じさせるのか、具体的に議論するためには、各評価機関についてのより詳細な情報が必要と考えられる。

し、その結果を平均すると、MSCI と他社との相違は 68%が「範囲」の相違によって説明され（同様に分析すると、他の評価機関では 28～35%となる）、ESG 評価に用いるカテゴリが他と大きく異なっていることを示唆している（図表 2）。同論文では、その理由は MSCI が独自に用いる「エクスポージャー・スコア」（ESG 指標の評価ではなく、その ESG 指標がどのくらいその企業にとって重要かを評価する指標で、MSCI 以外の評価機関では用いられていない）にあるとしており、そこから他社との評価に大きな相違が生じていると考えられている。

図表 2 ESG 評価の相違に対する「範囲」・「測定」・「ウェイト」寄与率の評価機関ごとの平均値

評価機関	範囲	測定	ウェイト
KLD	34%	55%	10%
Moody's ESG	35%	60%	5%
MSCI	68%	36%	-4%
Refinitiv	29%	64%	7%
S&P Global	33%	64%	3%
Sustainalytics	28%	60%	12%

注：6つの評価機関について、一対の組み合わせごとに相違を要因分解した結果の、評価機関ごとの平均値。

出所：Berg et al. (2022) を基に大和総研作成

また、分析の過程を詳細に見ると、各評価機関がどのような ESG 指標を相対的に重視するのかが読み取ることができる。図表 3 は、分析の一部として行われた ESG 評価についての回帰分析の結果から、回帰係数の大きさを基に ESG 評価に影響を与える上位 3 カテゴリを評価機関ごとに取り出したものである。例えば、Sustainalytics では「サプライチェーン」、「環境マネジメントシステム」、「グリーン製品」の順で ESG 評価に影響を与える要因としての重要性が大きい。一方、S&P Global ではこれらが「従業員の能力開発」、「気候リスクマネジメント」、「資源効率」の順となっており、ESG 評価において重視するカテゴリが違ってくる。分析対象の 6 機関すべてに共通して上位 3 つに入るカテゴリは存在しない。

図表 3 ESG 評価に影響を与える上位 3 つのカテゴリ

評価機関	回帰係数の大きさ上位3カテゴリ
KLD	気候リスクマネジメント、報酬、製品安全性
Moody's ESG	環境ポリシー、ダイバーシティ、労働慣行
MSCI	製品安全性、従業員の能力開発、汚職
Refinitiv	取締役会、報酬、資源効率
S&P Global	従業員の能力開発、気候リスクマネジメント、資源効率
Sustainalytics	サプライチェーン、環境マネジメントシステム、グリーン製品

注：各評価機関の ESG 評価を被説明変数、Berg et al. (2022) において定義されたカテゴリスコアを説明変数として回帰分析を行った際の回帰係数に基づくもの。

出所：Berg et al. (2022) を基に大和総研作成

4. ESG 評価のこれから

本稿では、ESG 評価の相違をめぐる問題を考える上で、どのように相違が生じるのかを理解することが必要であるという観点から、最近の実証研究を取り上げて検討してきた。最後に、そこから得られた示唆と今後の方向性について議論し、結論に代えたい。

筆者が Berg et al. (2022) を読む中で、重要であると考えたポイントは主に 3 点ある。1 つ目は、ESG 評価のプロセスとは情報を集約するプロセスだということである。集約のプロセスにおいて、どのような指標からより上位の項目を構成するか、いずれの項目を最終的な評価に取り入れるかなどをめぐって、さまざまな選択肢があり得る。どのような選択をするかは評価機関に固有であり、それら選択は評価を作るプロセスのいくつかの段階で生じる。一連の選択の違いが積み重なることによって、「ESG 評価」と総称されるものを全体で見た場合に、それぞれの評価の間に相違が見出されるものと理解できる。同論文のタイトルである“Aggregate Confusion”（筆者訳「集約による混乱」）は、ESG 評価の背景に存在するこのようなプロセスを端的に表現している。

2 つ目は、同論文は ESG 評価のための方法論の「違い」を描き出したものではあっても、そのような違いを「問題」として指摘するものではないという点である。同論文は、ESG 評価の相違がどのような要因によって生じるのかを明らかにするものである。同一企業に対する評価の相違は、方法論の違いを反映しているに過ぎない。

3 つ目として、筆者は「集約による混乱」は「総称による混乱」と言い換えられる面もあるのではないかという印象を受けた。方法論の観点では、用いるデータや方法の違いから評価の相違を理解することになるが、そもそも「E」「S」「G」は異なるものである。例えば、ESG の「E」面に特化した企業評価を行う団体である英国の CDP のように ESG の特定の側面に焦点を当てた専門的な企業評価は既に存在する。このような取組みが「S」や「G」の面にも広がるのかは定かではない上、「E」の側面といっても気候変動リスク、森林、水、生物多様性とさまざまなテーマが含まれるが、少なくとも総合的な ESG 評価よりは焦点を絞った評価へと、専門化が進むことも考えられるのではないだろうか。

以上を踏まえると、今後、ESG 評価の相違について考える上で重要なことの 1 つとして、「E」「S」「G」それぞれをより詳細に見ていくことが挙げられるだろう。例えば「E」面では、温室効果ガス (GHG) のサプライチェーン排出量について、自社における直接排出 (Scope 1) や自社が購入したエネルギーに由来する間接排出 (Scope 2) のみならず、自社の事業活動に関連する他社の排出 (Scope 3) の算定へと関心が広がっているように、同一の指標であっても意味内容が変化しているものがある。このような変化は「範囲」、「測定」、「ウェイト」のいずれによる相違をもまた別の形で顕在化させると考えられ、継続的な調査・研究が必要である。その際は、ESG 評価を取り巻く規制環境や市場環境の変化を踏まえることも求められるだろう¹⁴。

¹⁴ 本稿で取り上げた The Aggregate Confusion Project のような最先端の研究プロジェクトには国内外の運用機関も参画している。このことから、一部の機関投資家は既に ESG 評価の方法論をめぐる議論に参加し、影響

もともと、ESG 評価の相違は、評価機関の方法論のみに起因するものではなく、企業の情報開示の不足に起因する面もある。本稿では主に前者の観点から検討してきたが、後者の点に関しても、近年では実証研究が進みつつあり、企業による ESG 情報開示の度合いと評価機関の間の評価の不一致の関連を検証するものがみられる。

例えば、浅野（2023）は、2015 年から 2021 年までの日本企業を対象として、Bloomberg の ESG 開示スコアでみた企業の ESG 情報開示の度合いと、MSCI と FTSE の 2 つの評価機関の間での ESG 評価の不一致度の関連を検証した¹⁵。その結果、時間の経過とともに日本企業の情報開示の度合いが高まる中、2 つの評価機関の間の ESG 評価の不一致も小さくなっており、この両者に関連があることが確認された。このような関連構造が分析期間を超えて観察されるのか、異なる評価機関の組み合わせでも成り立つのかは追加的な検証が必要と考えられるが、企業の情報開示の拡充が、評価機関の間の ESG 評価の相違を低減させる可能性を示唆する 1 つの結果であると解釈できる。

企業は ESG 評価の対象者であるが、自社に対する ESG 評価を踏まえて情報開示を見直す、経営の参考にするという意味では「利用者」ともなり得る。評価機関と機関投資家の間のみならず、評価機関と企業との間でいかなるコミュニケーションが図られていくのかも、ESG 評価のこれからを考える上では無視できない点であろう。

日本においては、年金積立金管理運用独立行政法人（GPIF）が、2017 年に国内株式を対象とした ESG 指数を選定して以来、ESG 評価機関との対話を行っている。対話のテーマには「ESG 評価機関と企業との対話促進」も含まれ、企業から寄せられた問い合わせや意見について評価機関からのフィードバックミーティングが実施されている¹⁶。このような対話は、評価機関と企業の相互理解を通じて、ESG 評価をめぐる混乱を収拾させる効果が期待される。今後は、ESG 評価の妥当性や信頼性を確保するために、対話のような質的なアプローチを補完的に組み合わせることの重要性が増すのではないだろうか。

を与える存在になっていると考えられる。これは、機関投資家の投資・運用プロセスにおいて外部の ESG 評価の重要性が相対的に低下していく可能性を示唆していると考えられる。

¹⁵ 浅野敬志，2023，「サステナビリティ情報開示の拡充と ESG 評価の不一致」『証券経済研究』122：63－81，日本証券経済研究所。

¹⁶ GPIF，2022，「2021 年度 ESG 活動報告」pp. 27－28。