【特集】

開かれた社会関連会計研究とは？

阪　　智香

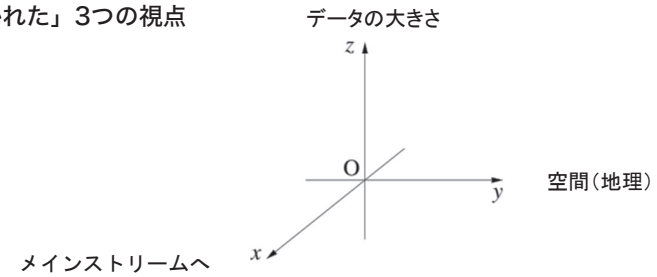
要旨

本研究では，「開かれた社会関連会計研究とは」をテーマとして，「開かれた」について，メインストリームへ，空間，データの大きさ，の3つの視点を示し，特に3つめの大規模データを扱うにあたって留意すべき点などを取り上げる。研究におけるデータ規模の増大は，機械的なデータ処理方法よりも，データから新しい知見を得る方法に変革をもたらし，探索的データ解析が有効となる。本研究では，財務データとESG Ratingsデータを用いた分析を例として，探索的データ解析を行うにあたり必要となるデータの前処理・ラングリング，欠測値の確認，データの分布・相関等の確認，そして，統計モデリングにあたってデータ分布と誤差構造に配慮することでより当てはまりの良い結果が得られること，最後に，再現可能研究について論じる。

1　「開かれた」3つの視点

本研究では，統一論題「開かれた社会関連会計研究を目指して：国際的，学際的，実践的な貢献の検討」をふまえ，「開かれた社会関連会計研究とは」をテーマとして，社会関連会計研究がどのような広がりをもつかとその論点を提供する。ここでは，「開かれた」を，メインストリームへ，空間，データの大きさ，の3つの視点から論じる（図表1）。

図表1　「開かれた」3つの視点



キーワード：探索的データ解析（Exploratory Data Analysis），データ前処理・ラングリング（Data Preprocessing and Wrangling），データ可視化（Data Visualization），欠測値（Missing Value），再現可能研究（Reproducible Research）

1.1　メインストリームへ

まず1つめは，図表1のx軸の「メインストリームへ」である。振り返って，1949年に公表された企業会計原則の序文には，「社会全体の利益のため」「我が国国民経済の民主的で健全な発達のため」という言葉がある。日本社会関連会計学会は，そのような目的に真正面から取り組んできたともいえる。一方，その後の1989年の国際会計基準審議会の概念フレームワークでは，「投資意思決定に有用な情報提供」に焦点が当てられ，財務報告の目的が変化した。このような会計制度は不平等を固定化し強化する傾向を有したことなどをふまえ（阪・國部・地道，2020），ビジネスのメインストリームで，ステークホルダー資本主義が注目されるようになった。アメリカの経済団体ビジネス・ラウンドテーブルは，2019年8月に発表した声明で，「顧客，従業員，サプライヤー，地域社会，株主といったすべてのステークホルダーの利益のために会社を導くことをコミットする」として，株主第一主義を脱し，ステークホルダーの利益を尊重する資本主義の実現を目指すとした。このステークホルダー資本主義は，2020年1月に開かれた世界経済フォーラム（World Economic Forum）年次総会（通称ダボス会議）でも重点テーマとなった。

日本社会関連会計学会は，創立の1988年から34年目，前身の企業生産性研究会から45年目を迎え，当初からステークホルダー資本主義を意識し，近年注目されている生産性や付加価値の研究，また，アカウンタビリティや環境会計，CSRディスクロージャー，ステークホルダーとの関係，最近ではESGやSDGsなどに関する多くの研究の蓄積がある。ようやくビジネスのメインストリームの意識がこの学会に近づいてきた。その時に，我々会員は，以前と同じ場所に戻るのではなく，新しいステークホルダー資本主義が，我々の研究の蓄積をもとに，社会がより高いレベルで議論ができるよう手助けをすることができるのではないか。これは社会関連会計研究がメインストリームの世界へ大きく開かれた，ということであり，社会の変革とその先へ，日本社会関連会計学会の過去の蓄積に基づきながら，我々が研究を通して貢献できると考える1つめの視点である。

1.2　空間（地理）

2つめの視点は，空間（地理）的な観点から「開かれた」ことある。会計研究がグローバル化し，世界数十か国のデータを使用した研究も増加した。多くの国を分析対象とする場合，特に新興国では，日本企業を対象にする場合と異なり，想像できないような事象が起きていることがある。例えば，トルコでは，2005年に100万トルコリラを1新トルコリラとするデノミネーションが行われた。それ以前の超インフレと6桁デノミの期間のデータをそのまま含めて分析している研究も存在するが，この期間のトルコ一国のデータが含まれていることで回帰分析の意味が失われるほどの影響を与えてしまうことになる。

そのため，まずは分析に使用するデータをよく確認する必要があり，その有効な方法は可視化である。国毎のデータを合計して，分析使用期間のデータの時系列可視化を実施すると，超インフレ・デノミ国の金額の発散と収束が確認でき，そのような国を識別することができる。この他

にも，グローバルな研究をする場合は，通常想定しないような財務諸表数値がみられることもあるため，分析前に，データを可視化してその特徴と実態を確認することが，これまで以上に重要となる。

1.3　データの大きさ

3つめは，図表1のz軸である「データの大きさ」の観点から，会計研究は大きく開かれてきた。データ規模の増大は，機械的なデータ処理方法よりも，データから新しい知見を得る方法に変革をもたらす。統計学者エフロンは，『大規模計算時代の統計推論（*Computer Age Statistical Inference*）』（Efron，2016）において，統計学の発展として，会計研究でも多く用いられているネイマン・ピアソンの仮説検定（1933年）は，その後，計算機の発展を機に，データアナリシス，データサイエンスに大きく方向転換してきたことを示している。

データがメモリ容量を遙かに超え，1台のコンピュータのストレージで収まらない程の量になると，一般にビッグデータとよばれる。データ規模が大きくなると，仮説に対する検証の精度が上がるため，統計学の使い方（会計研究への当てはめ方）も変わってくる。折しも，2016年には，米国統計協会が，統計的推論の利用に関する科学の実践を変えることを意図して，「統計的有意性とp値に関する声明」（警告）を公表し，p値は，データ取得やモデルの正しさ，影響の大きさには言及しない，ことを示した。アメリカ統計学会のジャーナル*The American Statistician*でも，2019年に，Statistical Influence in the 21st Century: Moving to a World Beyond “$p < 0.05$”という特集が組まれている。

大規模データを扱うにあたっては，データを可視化し，モデリングをし，フィッティングを確かめる一連のプロセスである探索的データ解析（Exploratory Data Analysis: EDA）（Tukey，1977）が有効となる（Saka and Jimichi, 2017）。次節では，大規模データを扱う場合に留意すべき点などを取り上げながら，企業の財務・株価情報として，Bureau van Dijk（BvD）のデータベースOsirisから連結決算主体で抽出した世界の全上場企業の財務データセット（以下，Osiris），および，ESG情報としてFTSE Russell 社のESG Ratings（以下，ESG レーティング）データを用いて，探索的データ解析によってESG情報と企業価値の関係を論じる（阪（2021b，2021c）を最新データで検証）。

2　データの前処理・ラングリング

一般に，粗データは，様々な理由から，直接データ解析環境に読み込むことができず，分析・解析することが難しい。そこで，まずは「読み込めるファイル形式」や「分析・解析できるオブジェクト形式」に変換する必要がある。この前処理の作業は，データを処理・分析・解析する全

工程の50％～90％を占めるとも言われる（地道，2018）。

Osirisは，154か国・30年間・96,377社のデータ，サイズが2GB（ギガバイト）超・約300万行である。このファイルをRで扱える形式に変換するためには，欠損値記号が存在しない欠損値の存在，レイアウトの不統一，金額に関するフォーマット（カンマ区切り）や特殊記号の存在（#など），データ規模が大きく通常のエディタやソフトでは読み込みや整形を行うことが困難といった問題に対処する必要がある。そこで，Unixコマンドやインタプリターを利用して整形し，データ解析環境Rを用いてデータファイルをRに読み込める形式（TSVファイル）に変換する，という手順で処理した。さらに，データファイルにおける文字列置換等の処理を行う際にGNU Parallelを利用してジョブを並列化することによって処理時間を短縮化するとともに，データをシリアライズするために，Apache Parquet形式を採用した。次に，このファイルをデータ解析できる形式に変換する工程であるデータラングリングを，Sparkを利用することによって行った。データ解析にRを利用することから，SparkRというRのパッケージを利用し，データファイルをSparkに読み込んだ。この読み込まれたオブジェクトを，SparkR，sparklyr，dplyrパッケージに付属する関数filter, select, collectを利用して，可視化・解析の用途に応じたRのデータ・フレーム・オブジェクトに変換した（地道，2018；地道・阪，2021）。

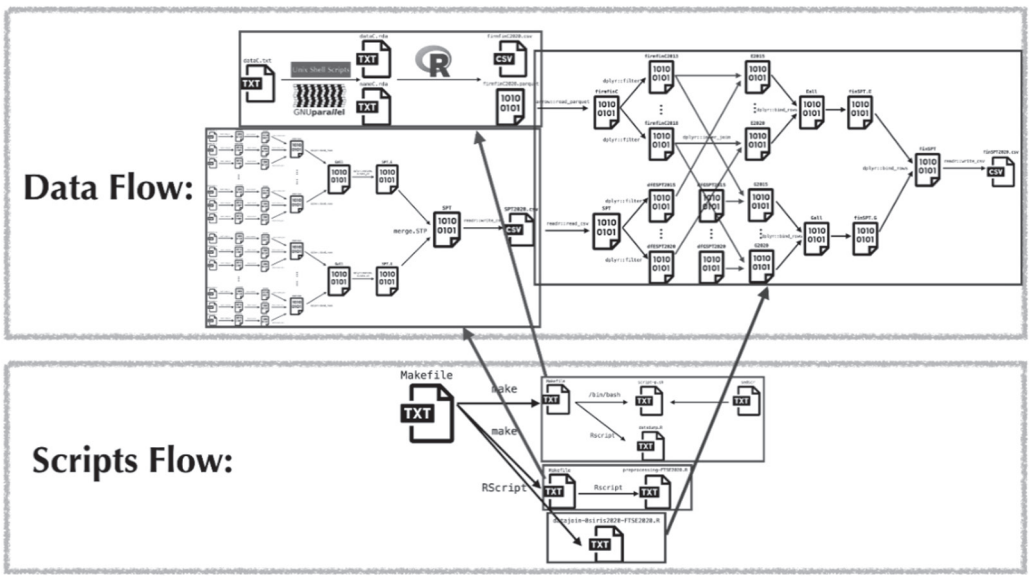
ESGレーティングは，FTSEが300以上の指標について，企業の有価証券報告書，CSR報告書，統合報告書，ウェブサイト等におけるESG開示情報を調査し，(1) 環境が5テーマ，社会が5テーマ，ガバナンスが4テーマの合計14テーマの評価，(2) 環境・社会・ガバナンスのそれぞれの総合評価（ピラー），(3) ESG全体の総合評価（サマリー）を算出し，評価にはエクスポージャーも考慮される[1)]。ESGレーティングの対象企業数は，2021年7月時点で，50カ国4,038社（先進国2,177社, 新興国1,861社）である。この4,038社の国別企業数は，中国の企業数が最も多く（1,012社），次いでアメリカ（609社），日本（507社），インド（172社），韓国（146社），イギリス（120社），オーストラリア（110社），台湾（108社），ブラジル（95社），香港（89社），フランス（82社），ドイツ（81社），南アフリカ（60社），スウェーデン（58社），カナダ（53社），タイ（53社），スイス（52社）という順となっている（50社以上の国のみ列挙）。このESGレーティングのファイルは，先進国と新興国が別になっており，さらに，上記のテーマ，ピラー，サマリー毎にファイルが分けられているため，これらのファイルを企業毎に結合するとともに，データ形式変換の処理を行う必要がある。そこで，先進国・新興国毎に，年情報を追加し，行結合し，並べ替えを行い，さらに，各列の型変換を行った。次に，先進国と新興国のデータを結合した。また，この前処理全体が自動的に行われるように設定した（地道・阪，2021）。

3　データの結合

本研究におけるOsirisとESGレーティングのように，複数のデータベースから抽出したデータを使用する研究では，データの結合（Join）が必要となる。その場合，2種類のデータの企業を一意に特定するためのコード（主キー）に何を使うべきか，それをどう入手するか，また，データベースによって企業の国名が異なるなど，実に様々な課題が生じる。1つ1つの課題については，データ元に確認しながら，必要に応じて追加の変数等を入手することで対応した。また，OsirisとESGレーティングとの統合にあたっては，どの年度のデータを対応させるかも重要となる。本研究では，ESGレーティングについては，企業の事業年度に係る情報がアニュアルレポート等で開示され，FTSEの初期調査，企業からのフィードバック，最終調査結果を経て，ESGレーティングが算出・公表されるまでに約2年を要することから，Osirisと同じ会計年度が評価対象となる2年後に公表されるESGレーティングを対応させて結合した。結合するためには，両方のデータに共通の変数（主キー）が必要であり，今回は証券識別コード（International Securities Identification Number: ISIN）を利用した。年毎に対応するOsirisデータとESGレーティングデータを結合し，先進国・新興国のそれぞれについて年に関するオブジェクトの結合を行った。

さらに，OsirisデータとESGレーティングデータの前処理から結合全体の全工程を，make コマンドを実行することによって自動化している。この自動実行の全工程のデータのフローとスクリプトのフローは，図表2に示すとおりである。OsirisとESGレーティングとの前処理から結合までの処理の詳細については，地道・阪（2021）を参照されたい。

図表2　OsirisデータとESGレーティングデータの前処理から結合までの全工程の自動実行（地道・阪，2021）



4　欠測状況の確認

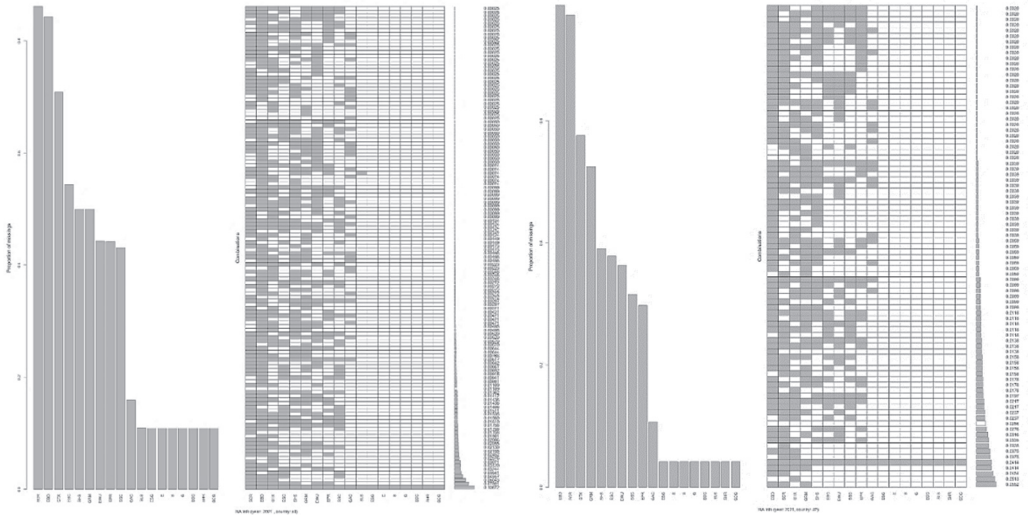
分析に入る前に，データの全体像を把握するために，データの欠測（欠損）状況を確認しておく必要がある。2021年7月時点のESGレーティングの対象全企業4,038社の欠測状況を可視化したものが図表3（左）である。例えば，テーマの顧客責任（Customer Responsibility），生物多様性（Biodiversity）では8割以上，税の透明性（Tax Transparency）では7割以上が欠測していることがわかる。これはFTSEの調査では，リスクエクスポージャーが低い業種は調査対象とならないという理由も影響している。一方，サマリーのESGや，ピラーの環境・社会・ガバナンスのレーティング，テーマの労働基準（Labour Standards），気候変動（Climate Change），人権と地域社会（Human Rights & Community），企業統治（Corporate Governance）は約9割のデータが入っていることもわかる。欠測パターンの図は，欠測値が出現する組み合わせを示している。例えば，最も出現している欠測パターンは全ての情報が欠測しているケース（全体の10.9%）であり，次は，顧客責任（Customer Responsibility），生物多様性（Biodiversity），税の透明性（Tax Transparency），環境サプライチェーン（Environmental Supply Chain），健康と安全（Health & Safety），水使用（Water Use），汚染と資源（Pollution & Resources）がセットで欠測しているケース（全体の7.8%）であり，この欠測パターンが全体に占める割合が右端のヒストグラムで示されている。

新興国と先進国では欠測しているテーマに違いがあり，新興国では全てのデータが欠測している企業が多い（図略）。国によっても欠測状況は異なっており，日本のみ図表3（右）に示す。日本企業で欠測の多いテーマは生物多様性（Biodiversity），顧客責任（Customer Responsibility: CR）であり，日本ではすべてのデータが欠測している企業は4.1%，欠測の最も多い組み合わせは，生物多様性（Biodiversity），顧客責任（Customer Responsibility），リスク管理（Risk Management）（5.5%），次いで税の透明性（Tax Transparency）（5.1%）である。中国では，すべてのデータが欠測している企業が2割を越えるが，これは2020年に新規対象の中国企業が増えた影響も大きいため（2019年の欠測企業は少ない），欠測データは時系列で確認しておくことも重要である。中国とインドは，税の透明性（Tax Transparency: TT）の欠測が100%である（図略）。

同様に，業種によっても欠測しているテーマに違いがある。ESGレーティングのIndustry Classification Benchmark（ICB）業種分類別企業数は，情報技術（Technology）368社，情報通信（Telecommunications）146社，ヘルスケア（Health Care）332社，金融（Financials）601社，不動産（Real Estate）258社，一般消費財（Consumer Discretionary）600社，生活必需品（Consumer Staples）319社，資本財（Industrials）701社，素材（Basic Materials）344社，エネルギー（Energy）161社，公益（Utilities）208社である。生活必需品では欠測データが少なく，情報技術，ヘルスケア，不動産では全てのデータが欠測している企業が比較的多い（図略）。

図表3　ESGレーティング（2021年7月）欠測状況の可視化

世界全体（4,038社）　　日本（507社）



5　国別・業種別のデータ確認

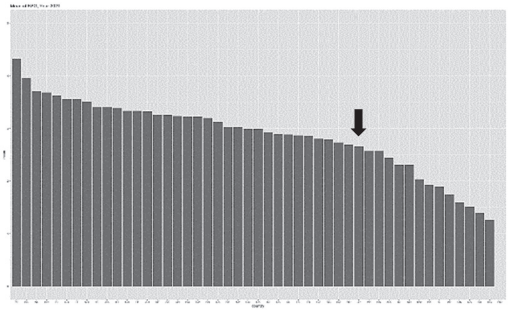
次に，分析を行うにあたり，国別・業種別のデータを確認し，その特徴を把握しておく。まず，国別の2021年7月のESGレーティングスコアの平均値を，サマリー，ピラーの順に取り上げたバーチャートが図表４である。ESG（サマリー）については，日本は上位から概ね中位に位置している（日本に矢印を付与。以下同様）。ピラーでは，環境（Environment）と社会（Social）で日本は上位から概ね2／3に位置している。ガバナンス（Governance）は全体に平均点が高いこともあり，日本は上位から概ね4／5に位置している。14のテーマの可視化は省略するが，日本のスコアが比較的高いテーマは，汚染と資源（Pollution & Resources），水利用（Water Use）であり，比較的低いテーマは，顧客責任（Customer Responsibility），人権と地域社会（Human Rights & Community）である。国別平均点は企業数が多くなると低下する傾向があり，日本企業の平均点が全体として中位以下であるのは，日本の対象企業数が国別で3位の507社もあることによることが大きいと考えられる。

業種毎では，サマリー（ESG）とピラー（E, S, G）については業種毎のスコアの差は大きくないが，生物多様性（Biodiversity），顧客責任（Customer Responsibility）では差が大きくなっている（図略）。これは，「生物多様性」，「顧客責任」のテーマについては，リスクエクスポージャーや活動国・地域によって適用外となる業種・企業が生じるため，スコアが付与されていない業種も存在することが理由として考えられる。全般に，情報通信，公益，エネルギー，金融の業種

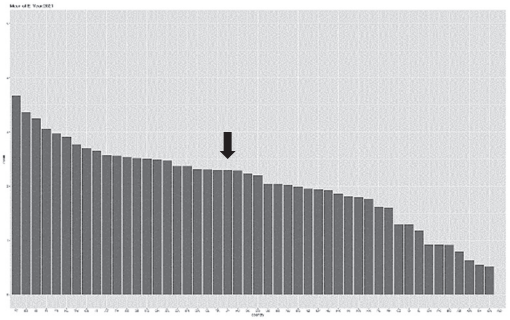
において平均値が高い傾向がみられる。

図表4　ESGレーティング（2021年7月）：国別平均値

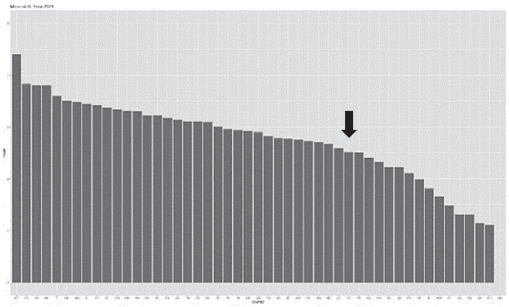
ESG（サマリー）



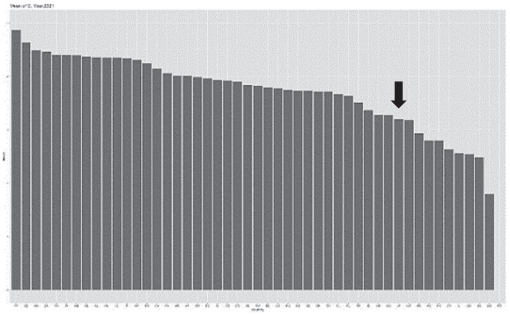
環境（ピラー）



社会（ピラー）



ガバナンス（ピラー）



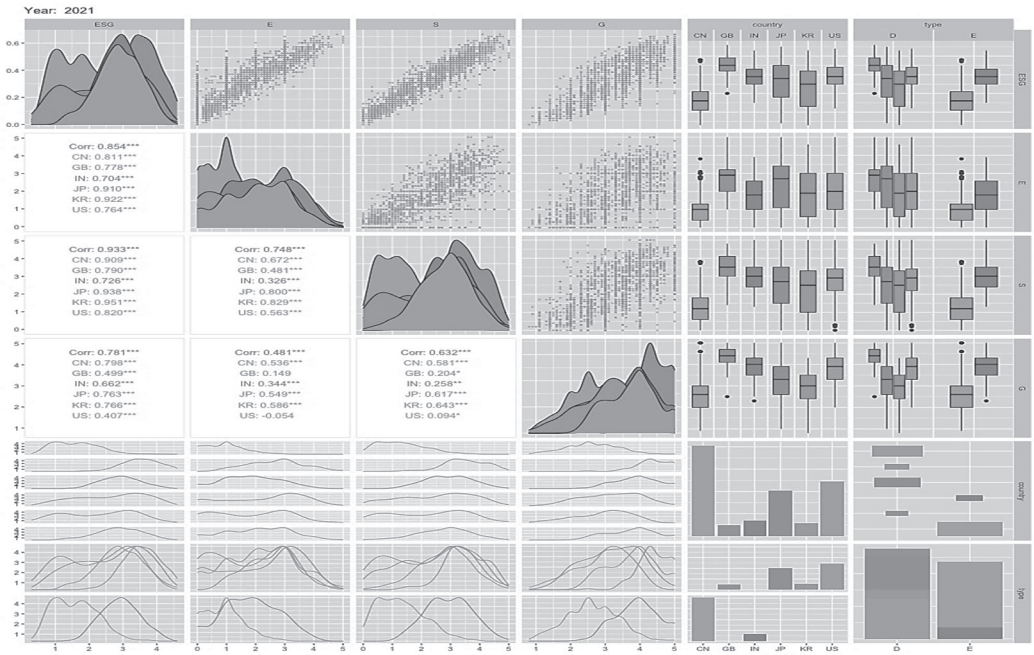
6　対散布図による分布と相関の確認

データを分析する際には，データを対散布図で可視化し，変数間に相関があるかどうかなどを事前に確認する必要がある。先進国の国別企業数上位4カ国であるアメリカ，日本，韓国，イギリス，新興国の国別企業数上位2カ国である中国，インドの計6カ国を取り上げ（これらは対象企業120社以上の国），ESG，環境，社会，ガバナンスの対散布図を描いたものが図表5である。右上段には散布図，対角線上には確率密度関数を推定したもの，左下段には相関係数が示されている。右2列は各変数の国毎のボックスプロット，下2行は国毎の各変数の確率密度関数の推定したものを示している。

図表5の確率密度関数やボックスプロットから，国毎のスコアに差があることが確認できる。また，ESGと環境，社会，ガバナンスがそれぞれ相関していることがわかる。また，右のボックスプロットと下の確率密度関数は，国別と先進国・新興国のタイプ別の2種類が示されている。

右の国毎のボックスプロットは右から順に，アメリカ，韓国，日本，インド，イギリス，中国となっており，この6カ国ではイギリスのスコアが最も高いこと，インドも比較的高く，中国は企業数が最も多いこともありスコアは低いが，経年比較（図略）からは6年間でスコアが上昇していることが確認できる。

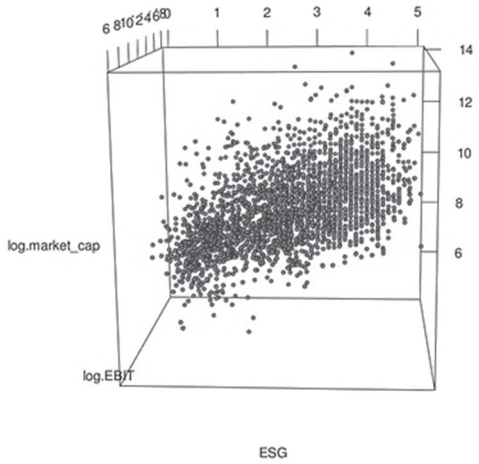
図表5　ESGレーティング（2021年7月）対散布図による可視化



7　統計モデリング

ESGレーティングと企業価値の統計モデリングを実施するにあたり，まずデータの関係を可視化する。4次元以上では可視化できないため，3つの変数を用いて，全企業のESGレーティング（x軸），利益（対数EBIT，y軸），企業価値（対数株式時価総額，z軸）を3次元散布図により可視化したものが図表6である。なお，Osirisデータは2020年入手のもの（2018会計年度）が最新であるため，本節では2020年7月のESGレーティングデータと2018会計年度の財務（Osiris）データを用いる。図6からは，ESGレーティングが高いほど株式時価総額が高いという相関があることがわかる。

図表6　ESGレーティング，利益，企業価値の3次元散布図



統計モデリングにあたっては，誤差の分布構造を考慮（Jimichi et al., 2018）することで，より当てはまりの良い結果が得られることにつながる。株価や財務データは通常極端に歪んでおり，対数変換してもなお歪みが残る。そこで，Ohlsonモデル（大鹿・阪・地道，2020）を念頭に，株式時価総額へ両対数モデルを当てはめるにあたり，誤差の対数に対して正規誤差，非対称正規誤差，非対称ティー誤差を仮定したモデルを考え，可視化とAICを用いて3つのモデルを比較検討することによってモデル選択を行った。全ての結果で，純資産，利益を所与としても，ESGの係数は有意に正であり，ESGスコアが高いほど企業価値が高いことが示されたとともに，図表7に示す非対称ティー誤差を仮定したモデルの当てはまり（フィッティング）が最もよいことがわかった（阪，2021c）。

図表7　ESGレーティングと企業価値の統計モデリング（非対称ティー誤差）

$\log(\text{株式時価総額i}) = \alpha_0 + \alpha_1 \log(\text{純資産i}) + \alpha_2 \log(\text{利益 i}) + \alpha_3 \text{ESG}_i + \alpha_4 \text{先進国ダミー} + \log(\varepsilon_i)$

	切片	log純資産	log利益	ESG	先進国ダミー	データ数
世界全体	-0.416	0.107	0.549	0.149	0.388	2,732
	(-2.391)*	(6.086)***	(31.733)***	(10.978)***	(13.277)***	

対数株式時価総額を被説明変数とする回帰分析の結果（非対称ティー誤差）。カッコ内はz比。
5%水準で有意な値に*、1%水準で有意な値に**、0.1%水準で有意な値に***を付している。
この結果は最尤法によって推定されたため修正済みR^2は掲載していない。

8　再現可能研究

本研究では，財務データとESGレーティングデータを用いた分析を例として，探索的データ解析を行うにあたり必要となるデータの前処理・ラングリング，欠測状況の可視化による確認，国別・業種別のバーチャートによるデータの特徴の把握，対散布図によるデータの分布・相関等の確認，そして，統計モデリングにあたってはデータの分布に配慮して対数変換するとともに，誤差構造の分布を考慮して誤差の対数に正規誤差，非対称正規誤差，非対称ティー誤差を仮定した3つのモデルを比較してモデル選択することで，より当てはまりのよい結果が得られることを示した。

さらに，通常研究論文を書くまでには，データを抽出し，前処理し，データ整形し，可視化・分析し，論文を執筆する工程があり，そのどこかの処理が再現できない際には研究の再現可能性が損なわれる。本研究では，データの処理をUnix系オペレーティングシステムのもとでシェルスクリプト，コマンド，データ解析環境Rを利用して実行しているが，この仕様は，前処理と結合の再現性を重視しているためである。全てのコードを残し，可視化した図などを常に元データから再現できるようにして，前処理・結合を含む探索的データ解析の一連の工程を再現可能研究（Reproducible Research）の観点から行った。また，工程の全てを対象とし，自動化することで，Peng（2011）の再現可能性のスペクトラムの最も高いレベルを確保している。

最後に，21世紀の科学のあるべき姿に関する会議の宣言では「社会における科学と社会のための科学」として，「科学研究の遂行と，その研究によって生じる知識の利用は，貧困の軽減などの人類の福祉を常に目的とし，人間の尊厳と諸権利，そして世界環境を尊重するものであり，しかも今日の世代と未来の世代に対する責任を十分に考慮するものでなければならない」と述べられている。日本学術会議「科学者の行動規範」（2013）でも，「科学者は，社会と科学者コミュニティとのより良い相互理解のために，市民との対話と交流に積極的に参加する」と述べられている。これらの問題意識は，日本社会関連会計学会の研究意識とも通じるところである。企業やステークホルダーと相互理解や対話・交流するにあたり，データは対話のツールとなり得る（阪，2021a）。データを正しく使い，データが示す証拠を，バイアスなく，わかりやすく可視化し，多くの人々に提示することは重要である。そして，そのような証拠や，社会関連会計研究の長い蓄積を活かすことによって，企業の社会関連情報開示やステークホルダーの行動に変化を促すことができるのではないかと考えている。

注

1）詳細は，FTSE RussellのESG Ratings https://www.ftserussell.com/ja/data/sustainability-and-esg-data/esg-ratings，日本取引所グループのESG評価機関の紹介
https://www.jpx.co.jp/corporate/sustainability/esgknowledgehub/esg-rating/02.html　参照

参考文献

Efron, B. and Hastie, T. (2016) *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science,* Cambridge University Press.

Jimichi, M., Miyamoto, D., Saka, C., and Nagata, S. (2018) “Visualization and statistical modeling of financial big data: double-log modeling with skew-symmetric error distributions”, *Japanese Journal of Statistics and Data Science*, Vol. 1, No. 2, pp. 347-371.

Peng, R. D. (2011) “Reproducible Research in Computational Science,” *Science,* Vol. 334, pp. 1226-1227.

Saka, C. and Jimichi, M. (2017) “Evidence of inequality from accounting data visualization,” *Taiwan Accounting Review*, Vol. 13, No. 2, pp. 193-234.

Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley Publishing Co.

大鹿智基・阪智香・地道正行 (2020) 「「社会にとってよい企業」への市場の評価とサステナ　ビリティ」『企業会計』第72巻第1号, 74-80.

阪智香（2021a）「データは対話のツール」『企業会計』第73巻第1号，42-43頁。

阪智香（2021b）「ESGと企業価値」『商学論究』第68巻第4号，149-170頁。

阪智香（2021c）「ESGと企業価値，将来に向けての視点」『會計』第199巻，12-24頁。

阪智香・國部克彦・地道正行（2020）「会計と不平等―付加価値分配率の探索的データ解析―」『国民経済雑誌』第221巻第4号，1-20頁。

地道正行（2018）「探索的財務ビッグデータ解析：前処理，データラングリング，再現可能性」『商学論究』第66巻第1号，1-31頁。

地道正行・阪智香（2021）「財務データとESGレーティングデータの前処理と結合」『商学論究』第68巻第3号，79-116頁。

日本学術会議（2013）「科学者の行動規範」

ユネスコ・世界科学者会議共催，21世紀の科学のあるべき姿に関する会議（1999）「科学と科学的知識の利用に関する世界宣言」

＜付記＞本研究は，JSPS科研費（19K02006，19K02026），2021年度学際大規模情報基盤共同利用・共同研究拠点（JHPCN 課題番号: jh211001-NWJ）等の助成を受けている。

阪　智香（関西学院大学）