

jh251004

# 財務ビッグデータの可視化と統計モデリング

地道正行（関西学院大学商学部）

## 概要

本研究では、財務ビッグデータを用いて、企業活動の利益率等の実態や、グローバル化もたらす負の側面（企業の租税回避、労働者と株主間の付加価値分配、企業の富の偏在等）に関する証拠と課題を提示した。この財務ビッグデータは、Moody's 社の 154 カ国・10 万社超の上場企業の（最長）30 年間・91 系列の財務（諸表）データ（Osiris, 300 万行, CSV 形式: 1.6GB）、および、非上場企業を含む 2,600 万社超の 10 年間・85 系列の財務（諸表）データ（Orbis, 2.9 億行, CSV 形式: 142GB 超）、さらに、Moody's DataHub システムから抽出された 2991 万社超の 47 年間・106 系列の財務（諸表）データ（DataHub, Parquet+DuckDB 形式: 44GB）、日本経済新聞社の日本の企業の延べ 17,000 社超の（最長）60 年間・2500 系列超の財務（諸表）データ（NEEDS, 295 万行, CSV 形式: 9.7GB）である。これらを、GPGPU 環境で PG-Strom、及び、CPU 環境で DuckDB と R を連動させて、探索的データ解析（Exploratory Data Analysis）にもとづき、時空間分析の観点からデータ可視化（Data Visualization）を行った。得られた知見に基づき、企業行動を高精度に予測する統計モデリング（Statistical Modeling）と実証分析を行うことでその有効性を検証した。これらの実証的エビデンスを社会に広く還元・周知し、社会・経済の持続可能な発展に向けて、企業行動の変容を促すことを目指すものである。

## 1 共同研究に関する情報

### 1.1 共同研究を実施した拠点名

- mdx I

### 1.2 課題分野

- データ科学・データ利活用課題分野

### 1.3 参加研究者の役割分担

地道正行（関西学院大学 商学部）:

- データ前処理, データラングリング
- 探索的データ解析にもとづく統計モデリング

宮本大輔（政策研究大学院大学 政策研究科）:

- データラングリング
- データ解析環境の構築

阪 智香（関西学院大学 商学部）:

- 財務データの経済・会計学的考察

永田修一（関西学院大学 商学部）:

- パネルデータ・時系列データ解析の理論構築

## 2 研究の目的と意義

本研究の目的は、次の 3 つである。

- (1) データの前処理・ラングリングとデータ抽出システムの構築
- (2) データ可視化とサステナビリティ基準開発と制度設計への活用
- (3) 統計モデリング

本研究で扱う財務諸表データ（補助資料参照）は、既存の研究において利用されてきたデータ量を凌駕する過去最大規模のものであるため、大容量のデータを高速に処理する計算機環境・ネッ

トワーク環境といった物理的な資源と、その環境を利用するための技術・知識・経験、さらにそれらを会計学・統計科学・情報科学の専門的観点から総合的に評価・分析できる多面的な知見をもつ人的資源が必要となる。これらの専門知識を有する研究メンバーにより、mdx I 環境 (以下、mdx と略すことがある) のもとで、財務ビッグデータを処理・可視化・モデリングするといった世界的にも希有の学際的研究を実践することが意義である。また、これらの研究成果と制度設計は、国内および国際的なサステナビリティ基準開発等に直接活用することができるということも、この研究の重要な意義である。

### 3 当拠点公募型研究として実施した意義

これまでの、研究において扱ってきた財務データセットは、数百 GB 超の複数のテキストファイルである。これらのファイルの前処理の高速化や可視化を含むデータ解析を実行するためには、それ相応の計算機環境が必要となり、このことが拠点公募型共同研究として実施してきた理由である。2022 年度に本格的に mdx 環境を利用し、これまでの研究を継続的に実行するために非常に優れた環境であることを確認している。また、Moody's 社のデータ抽出サービス DataHub を利用して抽出したデータファイルは、これまで扱ってきたデータ規模を超えるものとなり、このような規模のデータを前処理・可視化・モデリングするために mdx 環境が希求される。

また、本研究が扱う財務 (諸表) データは、会計学分野における研究におけるデータ量としては過去最大規模のものであり、これを扱うためには、高速な計算機環境・ネットワーク環境といった物理的な資源と、その環境を利用するための技術・知識・経験、さらにそれらを会計学・統計科学・情報科学の専門的観点から総合的に評価・分

析できる多面的な知見をもつ人的資源が必要となる。これらの専門知識を有する研究メンバーにより、学際的な共同研究を mdx 環境のもとで行うことが拠点公募型研究として実施した意義である。

### 4 前年度までに得られた研究成果の概要

2024 年 (度) に得られた各研究目的に対する研究成果は以下の通りである:

研究目的 (1) に対する成果:

- (a) 世界の全企業を対象として抽出されたデータセット DS-Orbis-C-2019, DS-Orbis-U-2019 の前処理を、これまでは mdx 上の 1 つの CPU ノードを利用することによって行ったきた (cf. 国内会議 (5)) が、3 つの CPU ノードをパラレルで利用することによって、スケールアウトを行い、従来よりも短時間 (20 分程度) で処理することができた。なお、この成果は、国内会議 (7) とワークショップ (1) で発表された。また、学術論文 (5) で公表され、国内会議 (16), (20), (26) でも発表された。
- (b) 日本における上場企業の財務データ (NEEDS) を抽出するシステム SKWAD <sup>\*1</sup> のバージョンアップ (2023 年版データの追加) を行った。このシステムの利用法については、マニュアル (3) にまとめられた。

研究目的 (2) に対する成果:

- (a) 現在、金融庁金融審議会でサステナビリティ情報の開示と保証のあり方が検討されている。論点の 1 つであるサステナビリティ基準の適用対象企業の検討において、我々の JHPCN の研究データを基にした分析結果が参考情報に掲

<sup>\*1</sup> SKWAD によるデータ抽出は、データのライセンスの関係上、弊学限定のサービスとなっている。

載され、議論を方向づけた。具体的には、サステナビリティ基準の適用対象企業を株式時価総額で区分する金融庁提案に対して、他の売上等の区分を確認すべきという複数の意見が委員から出されたことを受けて、JHPCN 研究データを利用して、株式時価総額、総資産合計、純資産、売上で区分した場合の対象企業のカバレッジを分析し、2018 年～2022 年の 5 年間の結果を示した。その結果から、時系列でも変数間でカバレッジ割合は大きくは異なることを示し、株式時価総額で区分することの妥当性を示すことができた（その他の業績 (1) 参照）。この分析結果を、金融審議会での検討状況に合わせてタイムリーに提供できたことは、JHPCN 研究で前処理・ラングリングしたデータを保持し、これまでも分析を重ねてきたことが大いに役立った。

- (b) DS-Osiris-C-2022 の財務データセットを用いて、株価純資産倍率 (Price Book-value Ratio:  $PBR = 1$  株当たり株式時価/1 株当たり純資産) の国別比較等の分析を行った。PBR は、株価が純資産を上回るかどうかをみる指標で、1 倍を下回ると純資産よりも安く買収されてしまうリスクがある。わが国の上場企業は PBR が低い (1 を下回る企業が多く存在する) ことが問題視されてきた。そこで、世界の上場企業の PBR の分布、実際にわが国企業の PBR が他国と比べて見劣りするかどうか、業種別の PBR 等について分析し、可視化した。その結果、先進国の中で日本は最も低い水準であることが確認できた (補助資料 図 1 参照)。また、上場企業の付加価値分配に関する各国の状況を、2021 年度までの 30 年間で可視化した。コロナ禍の 2020 年度は各国で利益が減少したが 2021 年度には回復したこと、一方、

1998 年前後のアジア危機はアジア諸国の企業の従業員へ人件費支払を激減させており、金融危機が与える影響の大きさが確認できた。

- (c) データセット DS Osiris-C-2020 と FTSE Russell ESG レーティングデータ DS-FTSE-2020 を結合したデータセット DS-Osiris-C-FTSE-2020 を用いて、株式時価総額と ESG データとの関係、財務データと ESG データの株式時価総額に対する説明力 (寄与率) などの探索的データ解析を継続して行い、この成果は、国内会議 (2), (3), (13) で発表された。さらに、このデータセットを用いて、回帰診断のプロット (補助資料 図 2 参照) を描くことによって、モデルの評価を行うとともに、当期純利益や純資産、ESG スコアが、国のタイプ (先進国、新興国) 別属性に依存することを条件付きプロット (conditioning plot) から、明らかにした (補助資料 図 3 参照)。この結果は、学術論文 (4) によって公表され、国内会議 (1), (12), (21) で発表された。
- (d) 日経 NEEDS 企業財務データを利用し、東京証券取引所プライム上場企業の売上高を高精度に予測するためのモデリングについて、2023 年度に得られた結果 (2022 会計年度のデータに対する非対称ティー誤差とダミー変数をもつ両対数モデルによる全要素生産性の考察を 1984 年 3 月期決算から 2023 年 3 月期決算の対数全要素生産性の推移を時系列プロット (補助資料 図 4 参照) をみることによって、業種毎の隆盛を考察し、国内会議 (4), (6) で発表された。

研究目的 (3) に対する成果:

- (a) データセット DS-Osiris-C-2020 と FTSE Russell ESG レーティングデータ DS-FTSE-2020 を結合したデータセット DS-Osiris-C-FTSE-2020 を用

いて、株式時価総額を財務データと非財務データで予測するための統計モデリングを行った。その際、先進国と新興国の国タイプの情報を考慮し、非対称分布族の有限混合回帰（混合比率が国タイプ属性に依存する場合）とダミー変数を持つ両対数モデルによるモデリングを行った。さらに、これらのモデルをデータに当てはめ、情報量規準によるモデル選択を行ったところ、非対称誤差とダミー変数を持つ両対数モデルが最も良いことがわかった。この理由をこの成果は、学術論文 (4), (6) で公表され、国内会議 (1), (12), (21) で発表された。

## 5 今年度の研究成果の詳細

各研究目的に対する今年度の成果は以下の通りである：

研究目的 (1) に対する成果：

- (a) 日本における上場企業の財務データ (NEEDS) を抽出するシステムとして、DuckDB を利用したものをローカル環境に構築した。このことにより、従来から利用しているシステム (SKWAD) より、高速にデータを抽出することが可能となった。この成果は、国内会議 (25), (28) で報告するとともに、学術論文 (9) で公表した。
- (b) 世界の全企業を対象として抽出されたデータセット DS-Orbis-C-2019, DS-Orbis-U-2019 の前処理を、これまでは mdx 上の 1 つの CPU ノードを利用することによって行ってきたが、3 つの CPU ノードをパラレルで利用することによって、スケールアウトを行い、従来よりも短時間 (20 分程度) で処理することができた。なお、この成果を、学術論文 (10) で公表した。
- (c) Moody's 社のデータ抽出システム

DataHub から世界の全企業を対象として 2024 年 11 月に抽出されたデータセットを、DukDB と R を協調するによって再データベース化し、探索的データ解析を実行するための環境を構築した。この成果を、国内会議 (38) で発表した。

- (d) 日本における上場企業の財務データ (NEEDS) を抽出するシステム SKWAD のバージョンアップ (2024 年版データの追加) を行った。このシステムの利用法については、マニュアル (4) にまとめられた。

研究目的 (2) に対する成果：

- (a) 内閣官房日本成長戦略本部の非財務情報可視化研究会の委員として、「人的資本可視化指針」の改訂の検討において、JHPCN 共同研究における研究成果 (Osiris による世界の上場・非上場企業の分析結果) を踏まえた意見を述べた。
- (b) 財務会計基準機構のサステナビリティ基準委員会 (SSBJ) 委員として、サステナビリティ開示基準開発の際に、JHPCN 共同研究における研究成果 (FTSE Russell ESG Rating と Osiris データを用いた ESG と利益 (EBIT), ESG と企業価値 (株式時価総額) との関係についての分析結果等) を踏まえて意見を述べている。
- (c) 金融庁金融審議会専門委員の「サステナビリティ情報の開示と保証のあり方に関するワーキング・グループ」、「サステナビリティ情報の保証に関する専門グループ」、および、日本公認会計士協会サステナビリティ能力開発協議会委員においても同様に、JHPCN 共同研究における研究成果 (世界の企業の財務特性と ESG 特性の分析) を踏まえた意見を述べている。この内容は、書籍 (1), 書籍 (2) と学術論文 (8) として執筆し、国際会議発表 (1), (2) を行った。また、

これらの研究成果は日本監査役協会の講演会にも含め、学術論文 (7) として執筆した。

- (d) 金融庁金融審議会専門委員の「ディスクロージャーワーキンググループ」において、Osiris による世界の上場・非上場企業の分析結果を基に、日本の資本市場の制度見直しの議論において意見を述べた。
- (e) 世界の会計士団体である国際会計士連盟 (International Federation of Accountants) における国際会計教育基準開発のためのパネル (the International Panel for Accounting Education) の委員として、国際会議発表 (3) を行った。また、日本公認会計士協会継続的専門研修制度協議会 IES 検討専門委員会専門委員としても JHPCN 共同研究における研究成果を還元している。さらに、日本公認会計士協会の研究大会で基調講演 (33) を行った。
- (f) 日本学術会議連携会員として参加したサステナビリティ経営の分科会において、JHPCN 共同研究における研究 (企業の付加価値分配の状況, ESG と企業価値との関係) を踏まえて議論し、分科会メンバーによる書籍『サステナブル経営—原理・潮流・実践』(2024) の第 5 章として執筆した。また、日本学術会議の国際会議「持続可能な社会のための科学と技術に関する国際会議 2025 (International Conference on Science and Technology for Sustainability)」において国際会議発表 (4) を行った。
- (g) JHPCN の研究成果を社会還元するために、産業界には講演 (24), (30), (34), (41), 環境省環境研究総合推進費関連のセミナーでの講演 (40) などを行った。
- (h) 以上の会議の検討において社会還元した JHPCN 共同研究における研究成果を、一般ビジネスパーソン向けの書籍と

して執筆した。最終校正を終え、2026 年 5 月に中央経済社から、阪智香・地道正行共著『会計ビッグデータで可視化する世界企業ギガトレンド』として出版される予定である。

- (i) 日経 NEEDS 企業財務データを利用し、東京証券取引所プライム上場企業の売上高を高精度に予測するためのモデリングを行い、1984 年 3 月期決算から 2024 年 3 月期決算の対数全要素生産性の推移を時系列プロット (補助資料 図 4 参照) をみることによって、業種毎の隆盛を考察した。この結果は、国内会議 (31) で発表された。なお、今年度に利用したデータは、今年度に得られた研究目的 (1) に対する成果 (a) である DuckDB による NEEDS 企業財務データのデータベースを利用した。
- (j) Apple Vision Pro による空間コンピューティングを利用し、3 次元空間内に多変量の財務データを可視化するために、生成 AI (クラウド, ローカル LLM) を利用し、Apple Vision Pro 用のアプリケーションを AI 駆動開発することを試みた。この結果は、国内会議 (29), (36), (43) で報告された。

研究目的 (3) に対する成果:

- (a) 日経 NEEDS 企業財務データを利用し、東京証券取引所プライム上場企業の売上高を高精度に予測するためのモデリングを行い、昨年度得られた結果が再現するかどうかを検証した。この結果は、国内会議 (31), (37) で発表された。なお、今年度に利用したデータは、今年度に得られた研究目的 (1) に対する成果 (a) である DuckDB による NEEDS 企業財務データのデータベースを利用した。
- (b) 2022 年 4 月から改編された東京証券取引所の新市場に対応する日経 NEEDS 財務データを利用し、業種と場部のカテ

ゴリ毎に分けられた企業の売上高、従業員数、資産合計の分布を統計モデリングし、パラメトリックブートストラップ法を用いることによって、合成疑似データを生成することができた。この結果は、論文 (11) として公表し、国内会議 (39) で発表した。

- (c) 日本の上場企業に関する財務データに基づき、企業価値を表す変数として株式時価総額を取り、キャッシュ・フロー変数が企業価値に対して持つ説明力を、可視化 (補助資料 図 5, 6 参照)、変数変換 (neg-log, log 変換)、誤差分布 (正規分布, 非対称正規分布, 非対称テーパー分布) の探索を行いながら検証した。この結果は、論文 (12) として公表し、国内会議 (44) で発表した。なお、利用したデータは、今年度で得られた研究目的 (1) に対する成果 (a) である DuckDB による NEEDS 企業財務データのデータベースを利用した。

なお、これらの研究成果に加えて、セミナー (2) (招待講演) と国内会議 (42) では、これまで得られた研究成果を総合報告するとともに、国内会議 (27) では、昨年度、参加したデータ解析コンペティション<sup>\*2</sup> において提供された規模の大きなデータ (CSV ファイル, 55 GB) を解析することによって得られた結果 (地理的加重回帰による食品 (特に惣菜) の売上高の統計モデリング) を報告するとともに、この取り組みを含む商学部におけるデータサイエンス教育についての総合報告 (32), (45) (共に招待講演) を行った。

## 6 進捗状況の自己評価と今後の展望

2025 年度は、これまで行ってきた非上場企業を含めて抽出したデータセットの前処理とラ

ングリング、データベース構築を mdx 環境で最新技術 (GNU Parallel, PG-Strom, Apache Arrow) を利用して実施してきたことを発展させ、CPU ノードをフルに活用した前処理を行うと共に、新たな技術として DuckDB を研究環境へ導入することと、PG-Strom が Parquet 形式に対応したことから、その利用に関する調査・検証を行った。このことにより、これまで扱うことが難しかった非上場企業を含めた富の蓄積と偏在、租税回避行動、付加価値分配に関する分析、可視化、モデリングを実行するための環境が整備され、新たなデータ可視化と統計モデリングを行うことができた。

2025 年度の具体的な研究成果として、学術論文 6 編、書籍 1 冊、学会・セミナー・ワークショップ等の報告 27 回 (招待講演 13 回含む)、統計数理研究所共同研究集会開催、テキストの発行、マニュアル・テクニカルレポート公表、プレス発表等に照らして、概ね順調に結果を得ることができているものと思われる。また、2017 年度から JHPCN の課題として採択していただき、2025 年度で 9 年目を迎え、これまで行ってきた研究の成果を総括するような機会を得ることができた (cf. セミナー (2), 国内会議 (42))。

これまでの進捗を勘案し、各研究目的に対する今後の課題と展望を与える：

研究目的 (1) について

- (a) 2017 年度に JHPCN 拠点として採択された時点からの懸案事項であった世界の全企業のデータ (データセット DS-Orbis-C-2019, DS-Orbis-U-2019) の前処理、データラングリング、可視化、統計モデリングを行うためには、高速にデータベースからデータを抽出する必要がある。このための方策として、mdx の GPGPU 環境のもとでの PG-Strom を利用してきたが、2023 年度から開発者である宮原徹氏、海外浩平氏らとの

<sup>\*2</sup> 経営科学系研究部会連合協議会 令和 6 年度データ解析コンペティション <https://jasmac-j.jimdofree.com>

共同研究を継続している。なお、2025 年末に PG-Strom が Parquet 形式のデータファイルへの対応が検証されており、2026 年度では Orbis データと DataHub データをこれらの環境で本格的に利用可能となる予定である。また、mdx 及びローカル環境でデータベース管理システム DuckDB でもデータベース化して、より高速にデータを抽出する環境を構築する予定である。

- (b) 2022 年 4 月から東京証券取引所（東証）の株式市場（第一部、第二部）の市場区分が「プライム」、「スタンダード」、「グロース」へ変更されたため、2023 年度からは、データベースを再整備する必要が生じたため、2024 年度もこの整備をおこなった 2026 年度は、この市場に対応した 2025 年版 NEEDS 財務データを DuckDB でもデータベース化して、より高速にデータを抽出する環境を構築する<sup>\*3</sup>。また、2022 年度から、日経 NEEDS 財務データにもとづいて、東京証券取引所（東証）の株式市場（第一部、第二部）に上場している企業の財務データの匿名化に取り組んできたが、2025 年度は、新市場区分に対応した財務データの匿名化に成功したので、2026 年度は引き続き匿名化の問題に取り組む予定である。
- (c) これまで主に利用してきたデータは、Moody's 社の Osiris, Orbis というデータベースから抽出したものであったが、これまでのサービスが 2022 年から DataHub へ順次移行したことに伴い、2025 年度は、このシステムからデータを抽出し、DuckDB を利用してデータベース化し、上場企業に関する探索的データ解析を行った。2026 年度は、最新のデータ（2025 年 11 月抽出分）を DuckDB を利用してデータベース化す

るとともに、mdx 環境において PostgreSQL と PG-Strom を用いてデータベース化し、研究に利用する予定である。

#### 研究目的 (2) について

- (a) これまで主に利用してきたデータは、日本経済新聞社のデータベース NEEDS や、Moody's 社の Osiris, Orbis, そして、FTSE Russell 社の ESG レーティングデータであったが、これに加えて、2026 年度は、DataHub のデータも利用し、様々な観点から可視化し、考察を行い、富の蓄積と偏在、租税回避行動、付加価値分配に関する分析を本格的に行う。これにより、富の蓄積と偏在、租税回避行動、付加価値分配に関する課題に取り組む予定である。また、国内および国際的なサステナビリティ基準開発と制度設計に向けたさらなる分析とデータ可視化の実施する。
- (b) 2025 年度は、Apple Vision Pro による空間コンピューティングを利用し、3次元空間内に多変量の財務データを可視化するためのアプリケーションを（生成）AI 駆動開発した。これにより、プロトタイプであるが、アプリケーションを開発することができ、隠蔽の問題をある程度解決することができた。2026 年度は、この研究を加速させ、従来からは「視る」ことができなかったデータの 3 次元構造を把握することによって、精緻な統計モデリングを行うためにつながるものとなることが期待される。なお、mdx にローカル LLM を導入することによって、開発環境を構築する予定である。

#### 研究目的 (3) について

- (a) 2023 年度に得られた成果（学術論文 (1)）では、日経業種分類の中分類をダミー変数と非対称ティー誤差をもつ両対数モデルが、2022 年 3 月期決算の東京証券取引所プライム上場企業の財務データへ良い当てはまりを与えること

<sup>\*3</sup> 2026 年 4 月の時点で 2025 年版のデータにもとづいて LAPP 環境と DuckDB 環境でデータベース化が完了している。

がわかった。さらに、この結果から得られた経済学的なインプリケーションとしては、コロナ禍における航空機産業と鉄道産業の売上高に対する大きな打撃があったこともわかった。2026 年度も最新の NEEDS データを DuckDB でデータベース化したものから高速にデータを抽出したものを利用することによって、このテーマに取り組む予定である。

- (b) 日経 NEEDS 企業財務データを利用し、東京証券取引所の新市場に準じた上場企業の財務指標に対する合成疑似データを生成するための統計モデリングに取り組む予定である。
- (c) DataHub から抽出したデータを DuckDB でデータベース化し、本格的に mdx 環境で利用して可視化や統計モデリングを実施する予定である。

2021 年度から 2023 年度まで実施してきたワークショップ (統計数理研究所・統計思考院・人材育成事業) を、研究色を強める形式で、2024 度は共同研究集会 (統計数理研究所) として実施した。2025 年度は研究種別を重点型研究\*<sup>4</sup> に移し、その公開研究報告会を開催した。2026 年度も引き続きこの研究を継続する予定である。

---

\*<sup>4</sup> 2025 年度統計数理研究所公募型共同利用 【研究種別】重点型研究「社会科学におけるデータモデリングの新展開」、企画立案責任者：川崎 能典 (統計数理研究所 学際統計数理研究系 教授) 地道 正行 (関西学院大学 商学部 教授)、佐藤 忠彦 (筑波大学大学院 ビジネスサイエンス系 教授)

補助資料

付録 A データ可視化

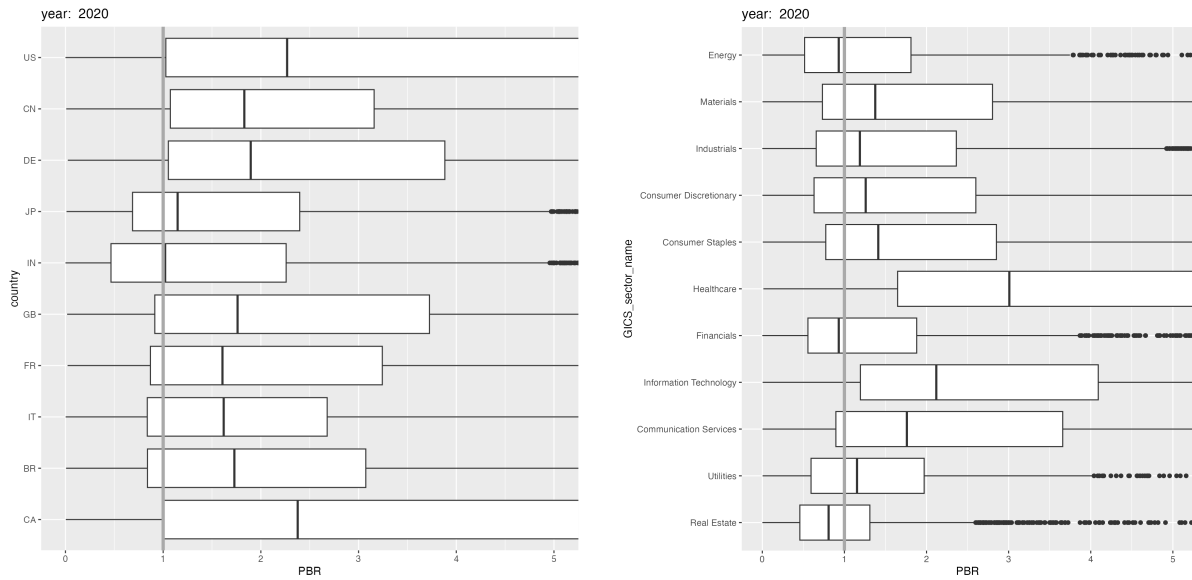


図1 国別・業種別の株価純資産倍率 ( $0 < \text{PBR} < 5$ ) (2020 年度): 左図: GDP トップ 10 カ国別, 右図: 業種別

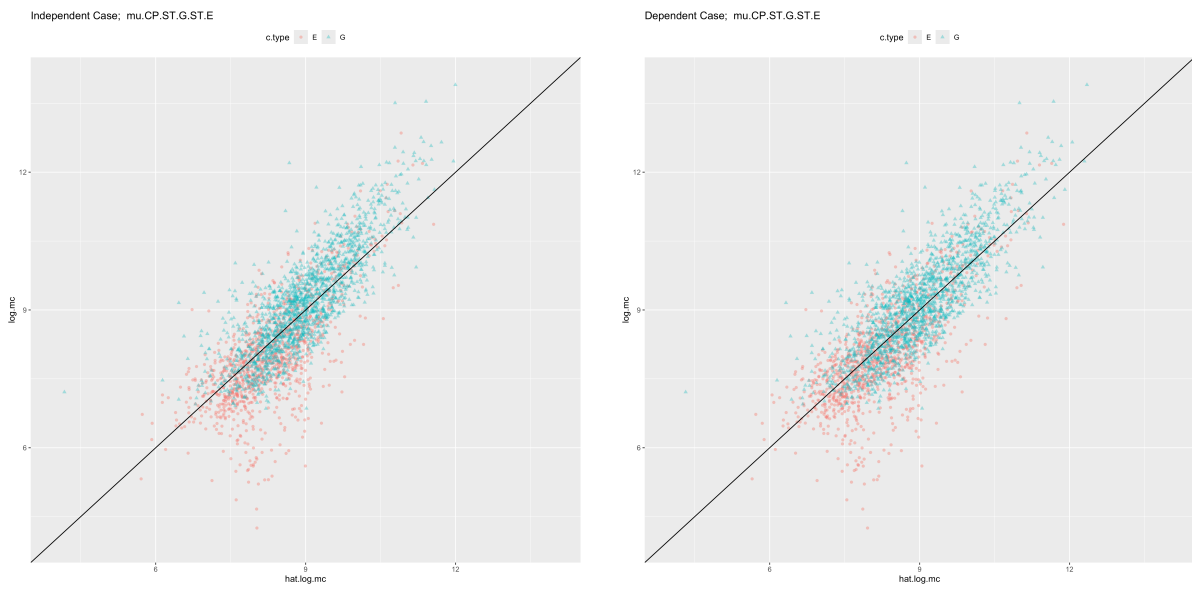


図2 対数株式時価総額の値 (縦軸) とその当てはめ値 (横軸) の散布図: 左図: 独立の場合, 右図: 従属の場合

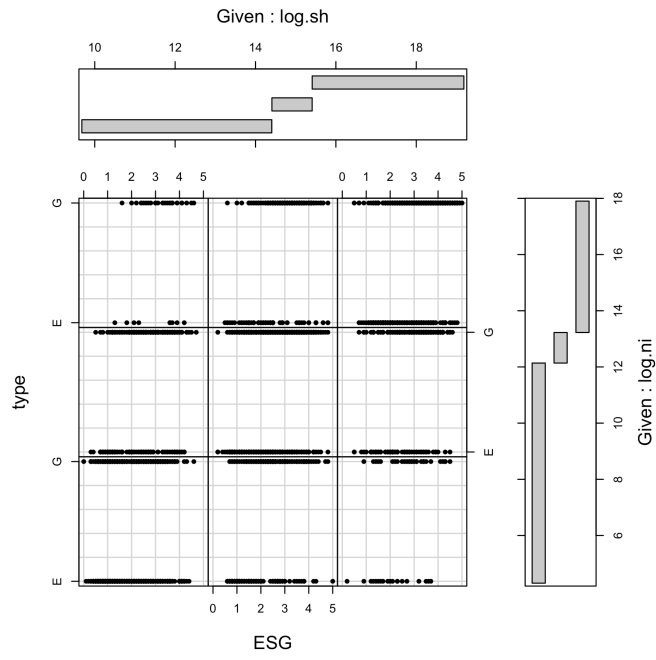


図3 条件付きプロット

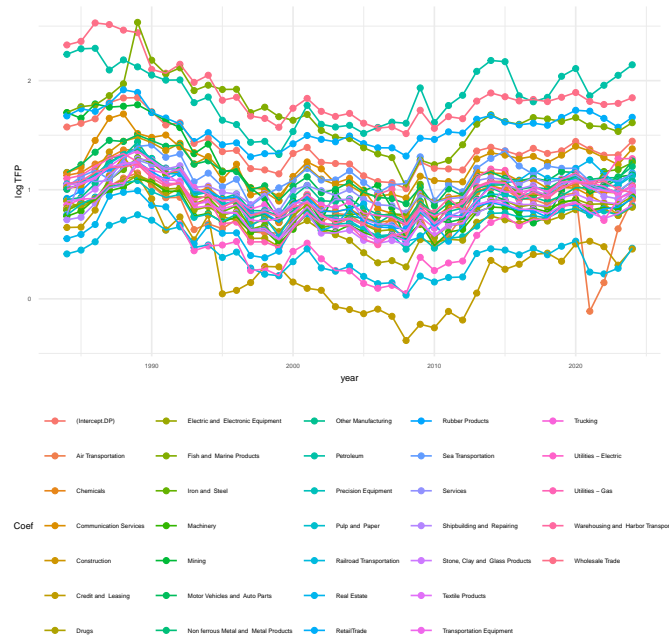


図4 対数全要素生産性の時系列プロット

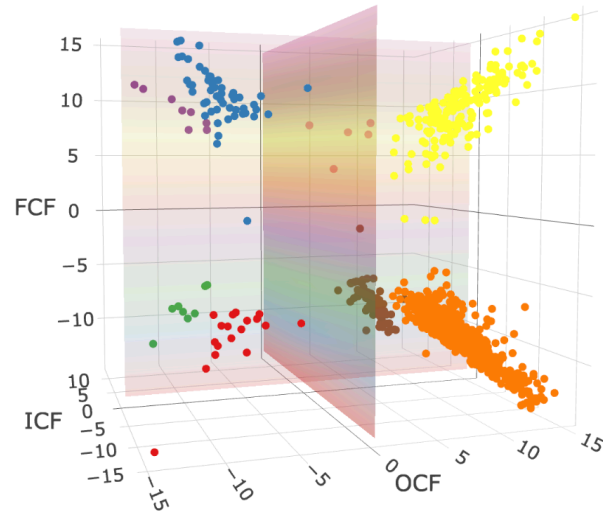


図5 neg-log 変換後の営業・投資・財務活動におけるキャッシュ・フローの3Dプロット(2024年)

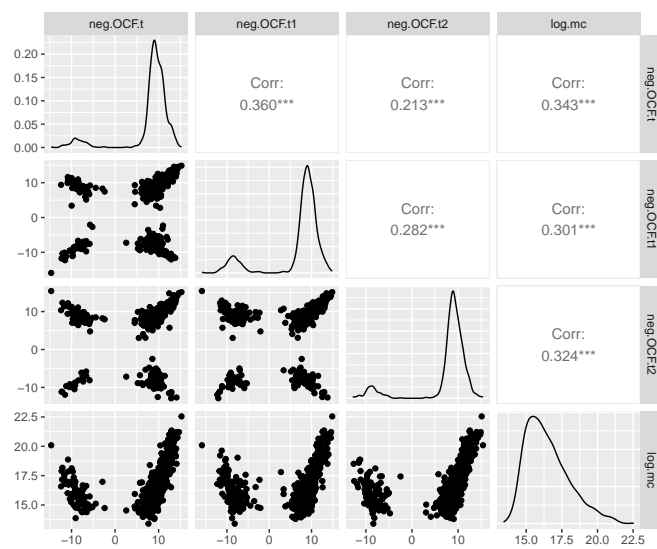


図6 neg-log 変換後の営業キャッシュフロー, 対数時価総額の対散布図(2024年)

## 付録 B データセット

表 1 データセット一覧

データセット名	年月版	データベース	上場・非上場	連結・非連結
DS-Osiris-C-2020	2020/3	Osiris	上場	連結
DS-Osiris-U-2020	2020/3	Osiris	上場	非連結
DS-Osiris-C-2021	2021/3	Osiris	上場	連結
DS-Osiris-U-2021	2021/3	Osiris	上場	非連結
DS-Osiris-C-2022	2022/3	Osiris	上場	連結
DS-Orbis-C-2019	2019/12	Osiris	上場・非上場	連結
DS-Orbis-U-2019	2019/12	Osiris	上場・非上場	非連結
DS-NEEDS-C-2024	2024/9	NEEDS	上場 (日本, 非上場含む)	連結
DS-NEEDS-U-2024	2024/9	NEEDS	上場 (日本, 非上場含む)	非連結
DS-DataHub-2023	2023/11	DataHub	上場・非上場	連結・非連結
DS-DataHub-2024	2024/12	DataHub	上場・非上場	連結・非連結

表 2 データセット: 規模

データセット名	企業数	規模
DS-Osiris-C-2020	96,377 社	約 300 万行, 1.6GB 超
DS-Osiris-U-2020	96,377 社	約 300 万行, 1.6GB 超
DS-Osiris-C-2021	100,542 社	約 300 万行, 1.7GB 超
DS-Osiris-U-2021	100,542 社	約 300 万行, 1.7GB 超
DS-Osiris-C-2022	105,276 社	約 300 万行, 1.6GB 超
DS-Orbis-C-2019	2,635 万社超	約 2.9 億行, 142GB 超
DS-Orbis-U-2019	2,635 万社超	約 2.9 億行, 142GB 超
DS-NEEDS-C, U-2024	17,424 社	約 295 万行, 9.7GB 超
DS-DataHub-2023	2,908 万社超	約 35GB (Parquet 形式)
DS-DataHub-2024	2,991 万社超	約 43GB (Parquet 形式)

## 付録 C 論文・講演等（投稿中・発表予定含む）

### 学術論文（査読あり）

- (1) Masayuki Jimichi, Yoshinori Kawasaki, Daisuke Miyamoto, Chika Saka, Shuichi Nagata, “Statistical Modeling of Financial Data with Skew-Symmetric Error Distributions”, *Symmetry*, Vol. 15, No. 9, 1772, September 2023, <https://doi.org/10.3390/sym15091772>.

### 学術論文（査読無し）

- (1) 地道正行『探索的財務ビッグデータ解析と再現可能研究: DataHub データのラングリングと Osiris データとの比較』, 商学論究, 第 71 巻, 第 2 号, pp. 1-50, 2024 年 3 月
- (2) 阪 智香『探索的データ解析の意義: ESG と企業価値研究の事例』, 商学論究, 第 71 巻, 第 3 号 (小菅正伸博士記念号), pp. 171-190, 2024 年 3 月.
- (3) 阪 智香『IFRS サステナビリティ開示基準 S1 号及び S2 号の概要』, 国際会計研究学会年報, 2023 年度第 2 号 (通号 54 号), pp. 15-32, 2024 年 7 月.
- (4) 地道正行『株式時価総額の統計モデリング: 独立性と従属性を考慮した有限混合回帰の適用』, 商学論究, 第 72 巻, 第 1 号, pp. 1-45, 関西学院大学商学研究会, 2024 年 10 月.
- (5) 地道正行『大規模データの前処理・ラングリング・可視化: 探索的財務データ解析と再現可能研究』, オペレーションズリサーチ, 2025 年 1 月号, pp. 32-43, 2025 年 1 月 1 日 (招待論文).
- (6) 地道正行『株式時価総額の統計モデリング: ダミー変数と非対称誤差をもつ両対数モデルの適用』, 商学論究, 第 72 巻, 第 3 号, pp. 1-43, 関西学院大学商学研究会, 2025 年 3 月.
- (7) 阪 智香『サステナビリティ開示基準を踏まえた企業経営と監査役等の役割』, 月刊 監査役, 2025 年 8 月 25 日号, 日本監査役協会, 778 号, 2025 年 8 月.
- (8) 阪 智香・地道正行『可視化最前線: ESG と財務実績』, 産業経理, 2025 年第 3 号, 産業経理協会, 2025 年 10 月.
- (9) 地道正行『DuckDB による探索的財務データ解析と再現可能研究: NEEDS 企業財務データの利用』, 商学論究, 第 73 巻, 第 1 号, pp. 1-54, 2025 年 11 月 <http://hdl.handle.net/10236/0002001604>.
- (10) 地道正行『探索的財務ビッグデータ解析と再現可能研究: mdx 環境による Orbis データの前処理再考』, 商学論究, 第 73 巻, 第 2 号, pp. 1-22, 2025 年 11 月 <http://hdl.handle.net/10236/0002001611>
- (11) 地道正行『財務データの匿名化: 日経 NEEDS 企業財務データ (東京証券取引所新市場区分対応版) を利用した合成疑似データの生成』, 商学論究, 第 73 巻, 第 4 号, pp. 1-71, 2026 年 3 月.
- (12) Miyu Fukui, Masayuki Jimichi, Chika Saka, Yoshinori Kawasaki, Statistical modeling of firm value based on cash flow, *The Institute of Statistical Mathematics Research Memorandum*, No.1229, 2026/03/31. <https://www.ism.ac.jp/editsec/resmemo-e.html>

### 国際会議発表

- (1) Chika Saka “Big Data & Exploratory Data Analysis (EDA) for Sustainability Issue”, The Valuism Conference 2025, Doshisha University, Kyoto, 28 August 2025.
- (2) Frendy, Tomoki Oshika, and Chika Saka, “Value Relevance and Determinants of Japanese Firms Pro-Forma Conformance to the IFRS S2 Climate-Related Disclosures”, 国際会計研究学会 IASB セッション, 神戸学院大学 (KOBE Co CREATION CENTER), 2025 年 8 月 29 日 (金).
- (3) Chika Saka, “Reaching the Next Generation of Talent”, 国際会計士連盟 (IFAC) Accountancy Education Directors Forum *Supporting the Full Professional Journey: Education, Development 2026 (London/Hybrid)*, 2026 年 3 月 5 日 (木), 6 日 (金). (招待講演)
- (4) Chika Saka, “Towards the Promoting International Talent Mobility and Circulation”, *International Conference on Science and Technology for Sustainability*, 2025, Science Council of Japan, 12. Feb. 2026. (招待講演)

### 国内会議発表

- (1) 地道正行\*, 川崎能典, 阪 智香, 宮本大輔, 永田修一『有限混合回帰による対数株式時価総額の統計モデリング』, 日本計算機統計学会, 第 38 回大会, ハイブリッド開催, 2024 年 5 月 25 日 (土).
- (2) 阪 智香『SSBJ の公開草案の概要と国際動向、そして今後の見通しについて』, サステナビリティ・コミュニケーション・ネットワーク (NSC) 定例勉強会, 2024 年 5 月 28 日 (火) (招待講演).
- (3) 阪 智香『「SSBJ 基準」公開草案の内容と解説』, 日刊工業新聞 産業研究所グリーンフォーラム主催, 「グリーンフォーラム特別シンポジウム 変わるサステナ情報開示 SSBJ 基準 起業への影響と対応」, 日刊工業新聞社 西日本支社 セミナールーム A (10F), 2024 年 5 月 30 日 (木) (招待講演).
- (4) 地道正行\*, 川崎能典, 宮本大輔, 阪 智香, 永田修一『探索的財務データ解析と再現可能研究: 東証プライム上場企業財務データの利用』, 日本経営数学会第 46 回 (通算 66 回) 全国研究大会, ハイブリッド開催 (専修大学神田キャンパス), 2024 年 6 月 15 日 (土).
- (5) 地道正行\*, 阪 智香\*, 宮本大輔, 永田修一『財務ビッグデータの可視化と統計モデリング』, 学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) 第 16 回 シンポジウム, ハイブリッド開催, 口頭発表, 2024 年 7 月 11 日 (木).
- (6) 地道正行\*, 川崎能典, 宮本大輔, 阪 智香, 永田修一『探索的財務データ解析と再現可能研究: 東証プライム上場企業財務データの利用』, 国際数理科学協会, 2024 年度年会「統計的推測と統計ファイナンス」分科会研究集会, 大阪公立大学中百舌鳥キャンパス, 現地開催, 2024 年 8 月 18 日 (日).
- (7) 地道正行『探索的財務ビッグデータ解析と再現可能研究: Orbis データの前処理とラングリング再考』, 国際数理科学協会, 2024 年度年会「統計的推測と統計ファイナンス」分科会研究集会, 大阪公立大学中百舌鳥キャンパス, 現地開催, 2024 年 8 月 18 日 (日).
- (8) 阪 智香『企業価値評価』, TKC・関西学院大学新月プログラム 2024, 関西学院会館, 2024 年 8 月 23 日 (金). (招待講演)

- (9) 阪 智香『サステナビリティ開示』, TKC・関西学院大学新月プログラム 2024, 関西学院会館, 2024 年 8 月 23 日 (金). (招待講演)
- (10) 阪 智香『制度化・基準化における実務家と研究者の役割分担』, アカウンティング・ウィーク 2024, 国際会計研究学会研究グループ報告「非財務情報の開示と企業価値」, 主査: 大鹿智基 (早稲田大学教授), 早稲田大学, 2024 年 8 月 26 日 (月).
- (11) 阪 智香『サステナビリティ開示の進展』, 日本会計研究学会第 83 回大会, 統一論題 (財務会計) 「社会変革と次世代の財務会計」報告, 早稲田大学, 2024 年 8 月 27 日 (火).
- (12) 地道正行\*, 川崎能典, 阪 智香, 宮本大輔, 永田修一『株式時価総額の統計モデリング: 有限混合回帰と非対称誤差・ダミー変数をもつ両対数モデルの利用』, 2024 年度統計関連学会連合大会, 東京理科大学神楽坂キャンパス, 現地開催, 2024 年 9 月 4 日 (木).
- (13) 阪 智香『SDGs13 気候変動対策とサステナビリティ開示』, 熊本学園大学大学院会計専門職研究科開設 15 周年記念シンポジウム「SDGs と会計・税務」, 熊本学園大学, 高橋守雄ホール, 2024 年 9 月 7 日 (土). (招待講演)
- (14) 地道正行\*, 川崎能典, 宮本大輔, 阪 智香, 永田修一『財務データの匿名化: NEEDS 財務データを利用した合成疑似データの生成』, 日本計算機統計学会第 38 回シンポジウム, 能楽堂ホール tenjin9/岡山市立オリエント美術館, ハイブリッド開催, 2024 年 10 月 26 日 (土).
- (15) 地道 正行, 阪 智香『探索的財務ビッグデータ解析』, データ活用社会創成シンポジウム 2024 \*<sup>5</sup>, オンライン開催, 2024 年 12 月 6 日. (招待講演)
- (16) 地道正行\*, 宮本大輔, 阪 智香, 永田修一『R による財務ビッグデータの前処理再考』, 統計数理研究所共同研究集会, 2024 年度「データ解析環境 R の整備と利用」, 統計数理研究所, ハイブリッド開催, 2024 年 12 月 8 日 (日).
- (17) 阪 智香『サステナビリティ開示で変わる経営と監査役役割』, 日本監査役協会, 情報交換会, ホテルグランヴィア広島, 2024 年 12 月 10 日 (火). (招待講演)
- (18) 阪 智香『サステナビリティ開示で変わる経営と監査役役割』, 日本監査役協会, 情報交換会, JR ホテルクレメント高松, 2024 年 12 月 11 日 (水). (招待講演)
- (19) 宮本大輔『ビッグデータ解析を取り巻く環境について』, 2024 年度統計数理研究所公募型共同研究集会 (研究課題名: 「ビッグデータ解析と再現可能研究」), 2025 年 1 月 25 日 (土).
- (20) 地道正行\*, 阪 智香\*『探索的財務ビッグデータ解析からみえる世界の企業行動の実態』, 2024 年度統計数理研究所公募型共同研究集会 (研究課題名: 「ビッグデータ解析と再現可能研究」), 2025 年 1 月 25 日 (土).
- (21) 地道正行\*, 川崎能典, 阪 智香, 宮本大輔, 永田修一『株式時価総額の統計モデリング』, 科学研究費シンポジウム「データサイエンスにおける統計的理論の展開研究」, 静岡がんセンターしおさいホール, 2025 年 1 月 30 日 (木).
- (22) 福井美優\*, 地道正行『飲料 POS データの探索的データ解析』, 科学研究費シンポジウム「データサイエンスにおける統計的理論の展開研究」, 静岡がんセンターしおさいホール, 2025 年 1 月

---

\*<sup>5</sup> <https://sites.google.com/g.ecc.u-tokyo.ac.jp/dp-sympo2024>

31 日 (金).

- (23) 阪 智香 『AI 活用時代における経営教育の変革：パネル討論 1 2030 年に向けた企業経営・会計監査の人材育成に資する教育変革』, 学術フォーラム, 日本学術会議主催, 日本学術会議講堂, 2025 年 3 月 1 日 (土). (招待講演)
- (24) 阪 智香 『サステナビリティ開示基準の解説と経営への影響』, ゼロカーボン講演会, 本州四国連絡高速道路本社, 2025 年 4 月 14 日 (水). (招待講演)
- (25) 地道正行 \*, 川崎能典, 阪 智香, 宮本大輔, 永田修一 『DuckDB によるデータベース構築: NEEDS 企業財務データの利用』, 日本計算機統計学会, 第 39 回大会, ハイブリッド開催, 2025 年 6 月 14 日 (土).
- (26) 地道正行 \*, 阪 智香 \*, 宮本大輔, 永田修一 『財務ビッグデータの可視化と統計モデリング』, 学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) 第 17 回 シンポジウム, ハイブリッド開催, 口頭発表, 2025 年 7 月 10 日 (木).
- (27) 市橋咲奈, 河本陽菜 \*, 福井美優, 伏見泰祐, 正井彩葉, 宮崎綾平, 地道正行 \* 『探索的レシートデータ解析: 主成分分析と地理的加重回帰による地域別消費傾向の探索』, 国際数理科学協会, 2025 年度年会「統計的推測と統計ファイナンス」分科会研究集会, 関西学院大学梅田キャンパス, 現地開催, 2025 年 8 月 24 日 (日).
- (28) 地道正行 \*, 川崎能典, 宮本大輔, 阪 智香, 永田修一 『DuckDB によるデータベース構築: NEEDS 企業財務データの利用』, 国際数理科学協会, 2025 年度年会「統計的推測と統計ファイナンス」分科会研究集会, 関西学院大学梅田キャンパス, 現地開催, 2025 年 8 月 24 日 (日).
- (29) 地道正行 『Apple Vision Pro のためのデータ可視化アプリケーションの AI 駆動開発』, 国際数理科学協会, 2025 年度年会「統計的推測と統計ファイナンス」分科会研究集会, 関西学院大学梅田キャンパス, 現地開催, 2025 年 8 月 24 日 (日).
- (30) 阪 智香 『サステナビリティ開示基準と企業の対応』, 九州生産性本部 第 21 期経理部長クラブ, 天神ビル (福岡), 2025 年 9 月 5 日 (金). (招待講演)
- (31) 地道正行 \*, 川崎能典, 阪 智香, 宮本大輔, 永田修一 『探索的財務データ解析と再現可能研究: NEEDS 企業財務データの利用』, 2025 年度統計関連学会連合大会, 関西大学 千里山キャンパス, ハイブリッド開催, 2025 年 9 月 8 日 (月).
- (32) 地道正行 『文系学部におけるデータサイエンス教育』, 2025 年度統計関連学会連合大会, 関西大学 千里山キャンパス, ハイブリッド開催, 2025 年 9 月 9 日 (火). (招待講演)
- (33) 阪 智香 『いのち輝く未来社会におけるアカウンティングの在り方—会計 (Accounting) の本質とサステナビリティ情報開示』, 第 46 回 日本公認会計士協会研究大会 2025, 大阪国際会議場, 2025 年 9 月 11 日 (木). (招待講演)
- (34) 阪 智香 『サステナビリティ開示について考える』, あずさ監査法人/KPMG, 東京大手町フィナンシャルシティサウスタワー, 2025 年 10 月 6 日 (月). (招待講演)
- (35) 阪 智香 「気候変動情報開示」サステナビリティ経営研究会, 関西大学梅田キャンパス, 2025 年 11 月 6 日 (木). (招待講演)
- (36) 地道正行 \*, 宮本大輔, 阪 智香, 永田修一 『Apple Vision Pro のためのデータ可視化アプリケー

- ションの AI 駆動開発』, 日本計算機統計学会第 39 回シンポジウム, 京都女子大学東山キャンパス, ハイブリッド開催, 2025 年 11 月 8 日 (土), 9 日 (日).
- (37) 地道正行\*, 川崎能典, 阪 智香, 宮本大輔, 永田修一『DuckDB による探索的財務データ解析と再現可能研究: NEEDS 企業財務データの利用』, 日本経営数学会第 46 回 (通算 66 回) 研究大会 2025 年秋季研究会, 専修大学神田キャンパス, 2025 年 11 月 22 日 (土).
- (38) 地道正行\*, 宮本大輔, 阪 智香, 永田修一『DuckDB+R による DataHub データの探索的財務ビッグデータ解析』, 統計数理研究所共同研究集会, 2025 年度「データ解析環境 R の整備と利用」, 統計数理研究所, ハイブリッド開催, 2025 年 11 月 23 日 (日).
- (39) 地道正行\*, 川崎能典, 阪 智香, 宮本大輔, 永田修一『財務データの匿名化: NEEDS 財務データ (新市場対応版) を利用した合成疑似データの生成』, 科学研究費・基盤研究 (B) 課題番号 23K28043 「高次元統計解析に有効な関数推定法の深化・展開研究」研究代表者: 内藤 貫太 (東北大学), 鹿児島シンポジウム, 鹿児島大学 郡元キャンパス学習交流プラザ 2 階 学習交流ホール, 2025 年 11 月 27 日 (木).
- (40) 阪 智香『自然関連財務情報開示の国際動向』, 東京大学森林科学セミナー, 東京大学農学部キャンパス (弥生キャンパス), 2025 年 12 月 23 日 (火). (招待講演)
- (41) 阪 智香『サステナビリティ開示基準』, 関西電力監査委員会室研修会, 関西電力本社, 2025 年 12 月 25 日 (木). (招待講演)
- (42) 地道正行\*, 川崎能典, 宮本大輔, 阪 智香, 永田修一『探索的財務ビッグデータ解析と再現可能研究』, 2025 年度 統計数理研究所 公募型共同利用 重点テーマ 3 公開研究集会 「社会科学におけるデータモデリングの新展開」, 統計数理研究所, 2026 年 2 月 21 日 (土).
- (43) 地道正行\*, 川崎能典, 宮本大輔, 阪 智香, 永田修一『データ可視化アプリケーションの AI 駆動開発 -Apple Vision Pro による 3 次元データ可視化-』, 科学研究費補助事業 (基盤研究 B) 「企業の税制適応行動とその経済的帰結に関する研究」 研究会, 神戸大学六甲台第 1 キャンパス 2026 年 3 月 1 日 (日). (招待講演)
- (44) 福井美優\*, 阪 智香, 地道正行, 川崎能典『キャッシュ・フローに基づく企業価値モデリング』, 第 20 回日本統計学会春季集会, 同志社大学今出川キャンパス, 2026 年 3 月 7 日 (土).
- (45) 地道正行『文系学部におけるデータサイエンス教育』, 第 23 回 統計・データサイエンス教育の方法論ワークショップ 「生成 AI 時代の統計・データサイエンス教育の再設計 -発達段階に応じた AI 支援と高大接続・指導要領の視点から-」, 統計数理研究所 (大講義室) ハイブリッド開催, 2026 年 3 月 28 日 (土). (招待講演)

#### ワークショップ・セミナー

- (1) 地道正行『探索的財務ビッグデータ解析と再現可能研究』, 第 115 回ワークショップ「企業の税制適応行動と戦略策定・政策立案 -税務行動研究の EBPM を通じた社会実装に向けて-」, NPO 法人 現代経営学研究所主催, 神戸大学大学院経営学研究科本館 208 教室, 2024 年 10 月 20 日. (招待講演)
- (2) 地道正行『探索的財務ビッグデータ解析と再現可能研究』, 神戸大学大学院人間発達環境研究科

数理交流セミナー（統計）, 2025 年 7 月 2 日（水）.（招待講演）

### 書籍

- (1) 阪 智香『サステナブル経営を実現する会計』, 西尾チヅル・上林憲雄編著『サステナブル経営』第 5 章 (pp.127-154), 同文館出版, 2025 年 1 月.
- (2) 阪 智香・水口剛『サステナビリティ基準がわかる』, 日本経済新聞出版, 2025 年 4 月.

### マニュアル・テクニカルレポート・テキスト

- (1) 地道正行『財務データ抽出システム SKWAD』, pp. 1-193, 関西学院大学リポジトリ, 2022 年 5 月 20 日, ISBN: 9784990553012\*<sup>6</sup>.
- (2) 地道正行『SKWAD ユーザマニュアル: NEEDS 企業財務データの抽出, Version 2.0』, pp. 1-100, 関西学院大学リポジトリ, 2023 年 4 月 12 日. \*<sup>7</sup>.
- (3) 地道正行『SKWAD ユーザマニュアル: NEEDS 企業財務データの抽出 Version2.1』, pp. 1-100, 関西学院大学リポジトリ, 2024 年 4 月 11 日, \*<sup>8</sup>.
- (4) 地道正行『SKWAD 2024-2025 ユーザマニュアル: NEEDS 企業財務データの抽出 Version2.2』, pp. 1-100, 関西学院大学リポジトリ, 2025 年 4 月 1 日, \*<sup>9</sup>.
- (5) 阪 智香『損益計算書』（第 4 章, pp.73-102）大阪商工会議所編『ビジネス会計検定試験公式テキスト 2 級 第 6 版』, 中央経済社, 2024 年 4 月.

### プレスリリース

- (1) 日刊工業新聞 朝刊「グリーンフォーラム特別シンポジウム 変わるサステナ情報開示 SSBJ 基準 起業への影響と対応」, 『「SSBJ 基準」公開草案の内容と解説 経営の財務軸と ESG 軸を近づける』に掲載, 2024 年 7 月 26 日（金）

### その他の業績

- (1) 金融庁, 第 2 回金融審議会「サステナビリティ情報の開示と保証のあり方に関するワーキング・グループ」, 「参考資料」に研究結果が掲載された. pp. 1-2, 2024 年 5 月 14 日\*<sup>10</sup>.

\*<sup>6</sup> <http://hdl.handle.net/10236/00030225>

\*<sup>7</sup> <http://hdl.handle.net/10236/00030672>

\*<sup>8</sup> <http://hdl.handle.net/10236/0002000348>

\*<sup>9</sup> <http://hdl.handle.net/10236/0002001099>

\*<sup>10</sup> [https://www.fsa.go.jp/singi/singi\\_kinyu/sustainability\\_disclose\\_wg/shiryou/20240514/02.pdf](https://www.fsa.go.jp/singi/singi_kinyu/sustainability_disclose_wg/shiryou/20240514/02.pdf)