ISSN 2436-9462

SBI金融経済研究所 所報
vol. 8
2025.8

SBI Research Review

巻頭言
政井 貴子｜SBI金融経済研究所 理事長

解題
副島 豊｜SBI金融経済研究所 研究主幹

次世代・デジタル金融の社会デザインを考える

SBI 金融経済研究所は、先端テクノロジーを活用した次世代・デジタル金融およびその市場のあり方を検討し、戦略的な提言を発信してまいります。
提言を通じて、日本社会全体のより良い発展に貢献することを目指します。

金融分野における因果推論の展開― 統計的手法・因果AI・自然言語処理の三潮流とその展望 ―
和泉 潔｜東京大学大学院工学系研究科 教授

視点の違いに着目した金融因果推論の高度化
市瀬 龍太郎｜東京科学大学 工学院 教授
許 子微｜東京科学大学 工学院 助教
陳 穎｜東京科学大学 工学院 博士課程在籍
井上 光太郎｜東京科学大学 工学院 教授

大規模言語モデルを活用した統計的因果探索の新展開
― 統計的因果プロンプティングによる専門知識とデータの融合 ―
三内 顕義｜京都大学理学部特定准教授
国立情報学研究所 大規模言語モデル研究開発センター 客員准教授
高山 正行｜文部科学省科学技術・学術政策研究所 第 1 調査研究グループ 研究官
滋賀大学 データサイエンス・AI イノベーション研究推進センター 客員研究員
清水 昌平｜大阪大学 産業科学研究所 教授
滋賀大学 データサイエンス学系 教授

オフ方策評価の概要とそのリスクリターンに基づく有用性評価
齋藤 優太｜コーネル大学コンピュータサイエンス学部博士課程在籍

暗号資産に投資するのはどんな人？― 因果推論手法の適用事例 ―
増島 稔｜ SBI 金融経済研究所 研究主幹
難波 了一｜SBI金融経済研究所 主任研究員

次世代金融インフラのあるべき姿
（2025 年 3 月 31 日）
次世代金融インフラの構築を考える研究会


SBI
Financial and Economic Research Institute

SBI Research Review

SBI金融経済研究所 所報 vol.8 2025.8

CONTENTS

巻頭言

政井 貴子｜SBI 金融経済研究所 理事長

政井 貴子

複数の外資系金融機関で約 20 年にわたり金融市場関連業務の拡大に貢献。その後新生銀行（現 SBI 新生銀行）に移り、2013 年、同行初の女性執行役員に就任し調査部門を率いる。2016 年、政府から日本銀行審議委員に任命され、唯一の女性委員として金融政策運営に参画。2021 年退任後、先端テクノロジーを活用した次世代・デジタル金融と金融市場のあり方等を調査研究する SBI 金融経済研究所を初代理事長として率いている。

進化する金融の未来へ

SBI 金融経済研究所は、本年で設立から 4 年を迎えました。この間、デジタル技術や金融制度、国際経済の構造が目まぐるしく変化する中で、私たちの研究活動も、次世代の金融インフラや 2040 年を見据えた経済・金融構造の探求といった新たな展開を迎えています。

本研究所では昨年度より、二つの新たな研究会を立ち上げました。一つは「次世代の金融インフラの構築を考える研究会」、もう一つは「2040 年の経済社会研究会」です。後者は、来るべき経済社会の全体像を探求し、あるべき姿、目指すべき姿を模索することで、前者の金融インフラ構想の前提となる議論を深めることを目的としています。両研究会は、急速に進化するデジタル技術や、人口動態・地域構造の変化といった日本経済の中長期的課題を見据え、社会全体に必要な制度設計や市場基盤のあり方を、理論と実証の両面から掘り下げていく取り組みを進めています。

前号、前々号の所報では、こうした研究会の主張提言を中心に取り上げ、データと金融経済制度、そして実体経済との接点に立脚した議論の蓄積をご紹介しました。本研究所が大切にしているのは、個別の論点に閉じるのではなく、将来を見通す複眼的な視座と、それに基づいた構想力です。次世代の金融インフラを構想することは、単なる技術導入や一部の制度改革にとどまらず、「どのような経済社会を目指すのか」という目指すべき社会像の共有が求められます。研究会では、そうした視点からの議論が、民間、行政、アカデミアといった多様な分野の有識者のご参加も得て、重ねられてきました。

本研究所の二つの研究会は、国際的な動向と呼応しながら、日本固有の構造課題に応答しうるものとなるよう議論を重ねています。国際動向の一部を紹介すると、たとえば BIS（国際決済銀行）は、中央銀行マネー、商業銀行預金、国債、証券といったマネーや金融資産のトークン化と相互連携を可能にする「統合型台帳（unified ledger）」の構想を打ち出しています。米国では、Saule Omarova 氏による公共デジタル通貨基盤「The People's Ledger」の提案が賛否両論を巻き起こしているほか、ケンブリッジ大学 CCAF が進める分散型金融に関する研究も注目を集めています。また、2040 年の経済社会構造に関しては、米国家情報会議（NIC）の『Global Trends 2040』や、欧州 ESPAS による地域将来像の報告書などが、長期的視野での制度設計に資する材料として各国の政策形成に活用されています。

本研究所のもう一つの柱である独自のアンケート調査も、着実に成果を積み重ねつつあります。本調査は、従来の株式・債券などの伝統的金融資産に加えて、暗号資産やステーブルコインといった新しい資産クラスも対象とし、リスク認識、情報取得、金融リテラシーとの関連性などを多角的に分析可能な設計となっております。今回で３回目となる調査でも、この設計に基づき実施いたしました。この調査は、単なるデータ収集にとどまらず、国内外でも類例の少ない独自の構成を追求しており、理論と実証を結びつける上で重要な基盤となりつつあります。すでに海外の研究機関にもデータ提供が始まるなど、着実に研究基盤としての厚みを増していると考えております。調査の設計から分析・発信に至るまで、研究所内のメンバーが主体的に関与している点も、本研究所の一つの特長と捉えております。

本号では、そうした研究所の活動の展開、特に研究会でのデジタル化社会を巡る議論や独自のアンケート調査から、政策やビジネスの判断において「何が結果をもたらしたのか」「ある介入や制度変更は、どのような影響をもたらすのか」といった因果関係をより深く把握することの重要性が改めて認識されたことを踏まえ、「因果推論の最前線」を特集として取り上げ、この分野の知見を一度整理しておくことといたしました。因果関係を把握することの重要性は、これまで以上に高まっています。ビッグデータの活用や AI 技術の進展と相まって、因果推論の手法や応用領域はここ数年で飛躍的に広がっています。本号では、統計的因果推論に関する最新の理論的展開に加え、自然言語処理や機械学習との統合、そして当研究所が実施したアンケート調査への応用まで、多面的な論考を掲載しています。

こうした流れを俯瞰すると、研究所の活動は、「研究会の社会提言」「データの蓄積」「理論と実証の架橋」という三つの軸がはっきりと見えてまいりました。これらの軸を今後さらに有機的に結びつけることが、私たちの研究の深化に肝要であると考えています。社会課題に対する構造的な視座と、現場に根差した実践的な問いをつなぐ知的な連携の仕組みを、研究所として徐々に形づくってまいりたいと思います。

本号が、研究所の現在地と今後の可能性を感じ取っていただく一助となり、読者の皆様それぞれの問いや関心と接点を持つことができれば幸いです。本号の刊行にあたり、ご協力いただいた諸先生方に深く感謝申し上げます。引き続き、実務と理論、制度とテクノロジーを架橋する知的挑戦を重ね、社会への貢献を目指してまいります。各界の皆様におかれましては、変わらぬご支援とご鞭撻を賜りますよう、心よりお願い申し上げます。

解題

副島 豊｜SBI 金融経済研究所 研究主幹

因果推論の最前線

今号の特集は「因果推論の最前線」である。因果推論とは、「あることが起きた原因は何か」「ある行動や政策をとった結果、何が起こるのか」といった原因と結果の関係を明らかにするための考え方や分析手法である。

たとえば、「ワクチンを打った、給付金支給や減税を行った、最低賃金を引き上げた、医療費負担率を引き上げた、非正規雇用にかかる労働契約法を変更した、株式投資促進策を採った、関税を引き上げた」場合、実現した結果が政策や介入の効果なのか、様々な他の要因に左右されたに過ぎないのかを判断するのは容易ではない。同様にビジネスにおいても、「価格を下げた、商品の内容量を減らした、クーポンを配布した、スポット広告を打った、店内レイアウトを変更した、AI 活用研修を行った、アプリに新機能を追加した、社外取締役を増やした、上場 /MBO/M&A を実施した」といった施策の効果を、他要因を排除して正確に計測するのは非常に困難である。

因果推論の原理的な困難さは、ある経済主体や個体について、時々の環境のもとで、「介入した場合」と「しなかった場合」の両方の結果を観測することができない点にある。これは、因果推論の根本問題（Fundamental Problem of Causal Inference）と呼ばれている。他の要因を完全に同一とした並行宇宙を用意し、世界Aと世界Bの比較を行ってみるわけにはいかないのである。この根本問題に対して、どのような手法で因果性を検証することが可能になるのか、長年にわたり研究が蓄積されてきた。

発展の歴史と主要な手法

観測されなかったもう一つの世界、すなわち、政策やビジネス戦略を行っていれば、もしくは行っていなければ、実現していたであろう世界のことを反実仮想（counterfactual outcome）とよぶ。これは、エコノミストにとって馴染みが深い概念であり、実証分析モデルを用いて反実仮想を推計する手法が知られている。消費税引き上げがなかりせば、量的緩和がなかりせば、資産バブルの早い段階で引き締め政策が採られていれば、不良債権問題への公的資金投入が早期に決定されていれば、等々である（前二者は政策が採られていなければ、後二者は採られていればという反実仮想である）。このほか、計量経済学

でよく知られた操作変数法も因果推論に応用されてきた。しかし、前者の経済構造モデル分析は「モデルが世界の近似として適切か」、後者は「適切な操作変数が利用可能か」という難しい課題を抱えている。

因果推論でもっとも著名な手法として、実験によって根本問題を乗り越えるというランダム化比較試験（RCT：Randomized Controlled Trial）がある。1930 年代に開発され、1940 年代には医療での適用が行われたが、経済分野では技術的・倫理的・予算的な制約により実験が難しいと考えられていた。しかし、1990 年代以降、社会政策や教育、経済学での応用が広がり、特に開発経済学の進歩において重要な役割を果たした。2019 年には開発経済学での RCT がノーベル経済学賞を受賞している。2010 年代になると、労働経済学や教育経済学、行動経済学、公共経済学でも RCT が活発に導入されるようになった。

因果推論の手法としては RCT や操作変数法のほかにも、傾向スコアマッチングや、差の差分法、回帰不連続デザイン、逆確率重み付けほか様々な手法がある。これらは、反実仮想が観測できないという制約を統計的仮定と設計上の工夫によって集団レベルで推定可能にするアプローチであり、潜在アウトカムモデルと総称される。こうしたアプローチとは別に、構造的因果モデルと呼ばれる手法も発展してきた。原因と結果の因果の方向性をグラフで表現し、これを数理モデル化することで個体レベルでの反実仮想が検証可能となっている。モデル化の方法論は異なるが、構造モデルを推計するという点では、前述の計量経済学の構造モデルと類似している。

政策やビジネスへの応用

以上のような因果推論の発展は、EBPM（Evidence-Based Policy Making）やデータマーケティングの観点からますます重視されるようになっている。最低賃金を引き上げると失業が増加するというトレードオフが必ずしも成立しないことを発見した米国の 90 年代初の研究や、コロナ給付金の相当な割合が消費に回らなかったことを銀行口座データから突き止めた研究は、いずれも因果推論の手法を用いている。排出権取引など環境規制が企業行動や大気汚染レベルに効果的であったか、少人数制学級が学力向上に寄与するのか、特定健診・保健指導が医療費の抑制効果を有していたか、警察官の配置増員や防犯カメラ設置が犯罪の発生を抑制したかなど、様々な分野で EBPM が進展し、因果推論は検証手法として重要な役割を担っている。

ビジネスにおいても応用事例は多い。ビデオ配信サービスにおいて推薦アルゴリズムが視聴時間を延ばすか・購買停止に抑止力を持つか、クーポン配布がタクシー乗車回数や収益を増やしているか（必要以上に配りすぎて収益機会を失っていないか）、金融商品キャンペーンやアプリの UI/UX 改良が商品購買や収益に貢献しているか、CRM（Customer Relationship Management）ツールや営業手法、トレーニングプログラムの導入が成約率向上や顧客単価、Life Time Value の引き上げに貢献しているかなど様々な検証が進められて

いる。このように、データサイエンスに基づくビジネスの高度化が競争力に直結する時代となって久しい。データ駆動型ビジネスの先端を行く GAFAM の株価時価総額の伸びがそれを象徴している。

その一方で、経験則や直感に基づく政策実施やビジネス戦略の決定も少なくないであろう。あるいは、データサイエンスの普及状況や、分析に必要な IT 基盤・データの未整備という現状から推測すると、誤った分析結果に基づく方策決定が政策立案やビジネス推進の現場で行われていることも十分考えられる。

因果推論の活用推進の動きも明確に観察されている。因果推論への注目度が高まるにつれ、専門書や啓蒙書、実践手法のテキストが出版されるようになった。特に、2020 年前後から翻訳書や日本の研究者による優れた書籍が多数執筆されている。Box 記事（p.8）にその代表的なものを示した。因果推論の考え方や実践方法をこれらの書籍で学ぶことは、政策やビジネス推進に大きく貢献する。

特集の狙い：新しい手法の登場、言語という分析対象への拡張

本号特集の狙いは、因果推論の最新動向の一部を紹介することである。近年の発展のドライビングフォースとなっているのは、自然言語処理の技術の発展、とりわけ大規模言語モデル（LLM：Large Language Model）の爆発的な進化により、因果性検証のための新データ、すなわち言葉で書かれた文章が利用可能になった点である（LLM については所報 6 号［2024 年］の副島論文を参照）。ここまで解説してきた統計的因果推論に加えて、新しい手法やテキストという膨大なデータソースにより、因果推論の潜在力が一段と高まる兆しが窺われる。こうした発展に伴い、研究領域の拡大が生じている。ものごとの因果性を検証するだけでなく、因果性を人間がどう認知しているか、その認知に基づきどう行動しているか、という人間の情報処理の仕方が研究対象となっている。本号掲載の 2 つの論文が、こうした視点からの研究であり、その手法について実証事例を含めて紹介している。

また、機械学習や AI の手法が発展し、これを因果推論に活用する研究やビジネスが増えている。推薦アルゴリズムの選択問題では、介入手段が数万の音楽や映像、書籍が対象になり、1 つの政策採用による介入に比べて解くべき問題が遥かに高次元化している。こうしたケースでは RCT や A/B テストの実施が困難であり、「ある方策や環境のもとで観察されたデータから意思決定を行う」ことを強いられる。また、医療のように試験的介入を行えない領域が広い分野も存在している。本号掲載の論文では、オフ方策評価と呼ばれる上記の意思決定問題を取り上げ、最近の技法発展を紹介している。

Box 記事に示した書籍の出版社名をみてもわかるように、因果推論を研究対象とする学問分野はもともと間口が広かった。AI や機械学習という新手法と、テキストという新データの登場により、情報処理、行動経済学、政治学、医学、法曹 / 法制度、経済史、比較制度分析、コミュニケーションサイエンス、マーケティング、計算社会科学（Computational Social Science）など、

因果推論を研究対象とする学問分野が拡大している。筆者は、5月に大阪で開催された人工知能学会の全国大会に参加したが、SNSデータを対象とした計算社会科学や、マーケティングにおける自然言語処理活用などのセッションが多数設定され、学界・実務界からの報告者も多分野・多産業にわたっている点が興味深かった。今号の収録論文も、いわゆるエコノミストによるものは1本のみであり、情報処理や人工知能、機械学習等の工学系の研究者のほか、教育行政、統計、自然言語、認知科学など、様々なバックグラウンドを持つ研究者の方々にご寄稿いただいている。

因果推論の可能性

こうした学際的な展開は、人間の意思決定における因果的思考の役割とその社会への影響を深く理解することに繋がる。インターネットは、情報の生産と拡散のされ方にグーテンベルクの印刷技術発明級の大きな変革をもたらした。政治や政策、企業のオフィシャルな情報発信やそれを伝える従来のメディア媒体と同じかそれ以上に、SNSでの情報発信や伝達のされ方が、世論形成や人々の行動変容、政治・経済政策、民主主義のあり方などに大きな影響を及ぼすようになった。悪意を持った情報操作、ポピュリズムの横行、プライバシーやセキュリティ情報の簒奪と悪用は、ネット社会の負の側面である。同時に、デジタル技術を活用した新しいデモクラシーへの期待や、ボトムアップ型の社会運営の可能性も広がっている。

以上のような社会環境において、因果推論の技法は一段と重要性を高めている。社会で起こる出来事の背景にどのような因果があるのかを正しく理解し、効果的な政策や制度設計を実施していくためには、原因と結果の関係を適切に検証することだけでなく、これを社会がどう認識し、行動に反映させていくかが、健全な社会運営を進めていくための鍵となる。地球温暖化や社会保障、税を巡る議論など個別論点の事例を考えるだけでも、その重要性が確認できよう。

ビジネスにおいても、従来の数字中心の構造化データでは捉えきれなかった因果関係を明らかにするという大きな潜在力を有している。多くのビジネス活動では、顧客の声や顧客対応、営業メモ、商品レビュー、チャット履歴、SNS投稿、FAQなど、非構造データであるテキスト情報が意思決定の基盤になっている。こうした情報は主に感情分析や話題抽出にとどまっていたが、ある文脈・表現・メッセージがどのような行動や評価に因果性をもって影響を与えているかを明らかにするために、因果推論の活用が始まっている。

広告や顧客対応などテキストや音声、映像を介した処置（介入）の効果、LLMによる大量のテキストデータの処理や構造化による因果推論の精度向上、テキスト生成による最適化介入設計など、応用範囲は更に拡がっていくであろう。

以下では、本号に収録された論文を実務への活用という視点を中心に紹介する。技術内容の詳細については、各論文やベースとなっている研究論文に当たられたい。

最近出版された因果推論の日本語書籍

1. 実務実践系

効果検証入門：正しい比較のための因果推論 / 計量経済学の基礎、安井翔太、2020、技術評論社
施策デザインのための機械学習入門：データ分析技術のビジネス活用における正しい考え方、安井翔太・齋藤優太、2021、技術評論社
反実仮想機械学習：機械学習と因果推論の融合技術の理論と実践、齋藤優太 (本号寄稿者)、2024、技術評論社
つくりながら学ぶ！Python による因果分析：因果推論・因果探索の実践入門、小川雄太郎、2020、マイナビ出版
マーケティングのための因果推論：偶然と相関の先へ進む因果思考、漆畑充・五百井亮、2025、ソシム

2. 教科書系（一部コード付き）

因果推論の計量経済学、川口康平・澤田真行、2024、日本評論社
開発経済学：実証経済学へのいざない、高野久紀、2025、日本評論社
はじめての統計的因果推論、林岳彦、2024、岩波書店
統計的因果推論の理論と実装 (Wonderful R)、高橋将宜、2022、共立出版
因果推論：基礎から機械学習・時系列解析・因果探索を用いた意思決定のアプローチ、金本拓、2024、オーム社
因果推論入門〜ミックステープ：基礎から現代的アプローチまで、Scott Cunningham、2023、技術評論社
政治学と因果推論：比較から見える政治と社会、松林哲也、2021、岩波書店
医学研究のための因果推論レクチャー、井上浩輔他、2024、医学書院

3. 世界の大御所系

インベンス・ルービン 統計的因果推論（上・下）、Guido Imbens・Donald Rubin、2023、朝倉書店
入門 統計的因果推論、Judea Pearl 他、2019、朝倉書店
ローゼンバウム 統計的因果推論入門：観察研究とランダム化実験、Paul Rosenbaum、2021、共立出版
政策評価のための因果関係の見つけ方：ランダム化比較試験入門、Esther Duflo 他、2019、日本評論社

4. 啓蒙書系

「原因と結果」の経済学：データから真実を見抜く思考法、中室牧子・津川友介、2017、ダイヤモンド社
データ分析の力：因果関係に迫る思考法、伊藤公一朗、2017、光文社新書
RCT 大全、Andrew Leigh、2020、みすず書房
因果推論の科学：「なぜ？」の問いにどう答えるか、Judea Pearl 他、2022、文藝春秋

和泉論文

最初は、金融分野における因果推論の応用を展望した和泉論文である。計量経済学の手法を含む統計的因果推論は、金融政策の効果や金融市場の変動分析、企業行動やガバナンスの研究に早くから活用されてきた。これを第一の潮流と位置付けたうえで、金融経済研究では注目される機会がまだ少ない他の二つのトレンドを紹介している。

一つは、機械学習との融合によって登場した「因果 AI」と呼ばれる新潮流である。深層学習による AI の性能向上や機械学習の技法発展は、高次元（多変量）で複雑な挙動を示すデータに潜む因果構造を発見し、因果効果の推定

や、反実仮想の生成を可能にした。市場変動の予兆や与信判断、政治経済情勢の先行き予測などにおいては、ベテランの経験則や直感が有効であり、そこには実務を通じて蓄積されてきたパターン認識や因果性認知が隠されている。こうした職人芸を AI や機械学習に取り込もうとする動きは過去からあった。森羅万象のデジタルデータ化が進展しつつあり、AI や機械学習の手法を適用することで、膨大なデータを網羅的に探索し、強力な因果発見力を活用することが可能になった。これらは既に、人間には到達不可能な領域に到達している。分野は異なるが、将棋や囲碁において AI が人間を超越したことがその一例であり、レントゲン画像の読み取り、創薬、物流最適化、異常検知などでも人間以上の高い精度やパフォーマンスを示している。

もう一つの潮流は、自然言語処理に基づく因果推論である。金融レポート、決算短信、ニュース、SNS 投稿といったテキストを対象に、因果関係の抽出と分析を試みる領域であり、言葉として人間に直接理解可能な因果構造を扱う分野である。本アプローチは統計的因果推論や因果 AI にはない特徴と重要性を有している。この点を本論文では、「経済現象のように人間の行動が深く関わる領域では、統計的因果だけでは事象を説明することが困難である。その主な理由は、人間が出来事をどう認知し、それにどう反応するかという人間の認知と行動ルール自体が因果構造の中核をなすためである」と端的に指摘している。また、具体例を示してテキストから因果を抽出する手法を紹介している。上場企業の決算短信は、企業を巡る外生環境や企業の戦略選択が決算という結果にどのような因果性をもって影響したかを示している。この因果連鎖の探索を行うシステムが実装・公開されており、その利用事例が示されている。

本稿は単に手法の紹介にとどまらず、因果推論を「統計的因果」と「経験的因果」という二つの概念枠に整理し、それらの補完性にも焦点を当てている。統計的因果は反実仮想モデルや構造方程式に基づく形式的手法であり、再現性と数量性を強みとする。一方、経験的因果は人間の認知や文脈的理解を含む因果の捉え方であり、稀少事象や未知の状況において直感的判断を支える。この両者の統合的視座が、現代の複雑化・非定常化する金融環境においてますます重要になることを指摘している。

取り上げられている技術分野の幅広さも本稿の特徴である。AI、機械学習、自然言語処理、統計学、計量経済学、マルチエージェント・シミュレーション、オルタナティブデータ、金融のドメイン知識といった異分野の知見を繋ぐ知的格闘技が因果推論のフロンティアを拡げていくことを期待させる。本特集号の編集企画にも、そうした視点が反映されている。

市瀬他論文

展望論文で紹介された第三の潮流に属する研究であり、筆者らが開発した最新技法を紹介する論文である。取り上げられている具体例から入ろう。有価証券報告書には MD & A（Management Discussion and Analysis、経営者による財政状態および経営成績の説明）が含まれている。経営者の視点からの

業績や財務状況の説明である。企業経営者を、株主重視、取引先や従業員などステークホルダー重視に二分したとすると、二つのグループで因果性の捉え方に相違はあるのだろうか。例えば、コーポレートガバナンスの改善が収益力の改善をもたらしたのか、収益面での余力が生じたのでコーポレートガバナンスの改善に取り組む機会をもてたのか、逆方向の因果性がそれぞれ存在しうる。株主重視の企業経営は前者であり、ステークホルダーを優先する企業経営は後者であるかもしれない。はたして MD & A のテキスト情報から、経営者の頭の中にあるこうした因果性を抽出することはできるのだろうか。「できる」というのが本論文のエッセンスであり、筆者らが開発したツールが紹介されている。

もう一つ紹介されている事例をあげよう。自信過剰な経営者は、自身の意思決定の正確さを過大評価し、確率的事象における不確実性を過小評価する傾向があるという指摘が先行研究によって行われている。こうした経営者は、良好な業績を自身の能力に起因させ、不振の結果を外部要因に帰する傾向もあると推測され、自己奉仕的帰属バイアスという名前が付けられている。上述のような経営者の認知バイアスや過剰な自信は、企業の戦略決定や資本構成、投資行動、業績や株価の予測などに強い影響を及ぼしうる。MD & A のテキスト情報から、どの企業の経営者にそうした傾向が強いかを炙り出すことができる。筆者らの研究によれば、自己奉仕的帰属バイアスと自信過剰・楽観予測は正相関している。また、同バイアスは、CEO の年齢が高いほど、在任期間が長いほど、理工系教育を受けているほど低下することを発見している。同バイアスが高い企業は、配当減少、自社株買い増加、高財務レバレッジ、企業価値棄損に繋がる M&A の実施という特徴を持つことも検証されている。

金融業で企業評価にかかわる者であれば、筆者らが開発した手法に強い関心を持つであろう。人間にしかできなかった、かつ労働集約的であり、サイエンスとして実行し難かった経営者評価を、定型化し自動処理化することが可能となり、人間による情報処理との比較検証も可能となる。論文では、テキストから因果関係を抽出する方法、これを整理して因果知識グラフにする方法、因果推論における視点の相違を抽出する方法が解説されている。また、機械学習を活用し、経済専門家の視点から因果知識グラフを自動生成するフレームワーク FinCaKG（筆者らによる開発）も紹介されている。計量経済学の視点からは、因果性検証に用いられる操作変数をテキスト解析によって膨大な候補の中から自動探索する技法（ETE-FinCa）も興味深い。

本手法のより一般的な貢献は、因果推論は用いるデータやモデル以上に「観測者の視点」に依存するという指摘である。「金融商品を推奨された / 金融教育を受けた個人は、その推奨 / 教育の情報に基づいてどう意思決定を変えたのか」といった問題設定では、個人の認知や理解プロセスそのものが因果推論の対象になる。そこでは、人間がどのように世界を理解しているか（因果認知モデル）を踏まえたモデル構築が必要となる。このようなアプローチは、行動経済学や認知科学の成果とも接続し、因果推論をより人間中心的な視点から再設計することを可能にする。

三内論文

LLM と統計的因果探索の統合が本論文のテーマである。統計データのみに依存する因果推論では、データを機械的に処理すると直感的にありえないような因果性を提示することがある。アイスクリームの売上が増えると溺死事故が増加するという因果は、気温と水辺のレジャー機会というチャンネルを無視したことによって導き出されうる。分析対象となる統計データに、これら４者が適切に含まれている保証はない。

実際の分析でこうした事態が生じにくいのは、分析者が自分の事前知識に基づき、注目すべき因果性の仮説設定を意識的・無意識的に行っているからである。このため、分析者の持つ情報次第ではバイアスが発生する可能性がある。因果のストーリーの誕生起点がどこにあるのかという問題である。アイスクリームの事例は誰もが持つ常識で解決できる（仮説として取り上げられることすらない）。では、財政支出による経済成長刺激が不足していたから長期にわたる低成長がもたらされたという因果仮説はどうであろうか。考察に入れるべき事象が非常に多く、1 枚のグラフで検証できるような仮説ではない。しかし、そこに因果性を導き出したいという強いバイアスを持つ者が少なからず存在する。なぜそうしたバイアスが人の心の中に生じるのかも、社会学に関する広義の因果推論の研究対象として重要であろう。

本論に戻ると、本論文の元となっている研究論文の筆者らは、LLM が膨大な分野の情報を大量に学習し、多様な専門家の知識やビジネスドメインの常識を備えている点に注目し、因果推論において統計的手法と LLM を融合させる方法を開発した。その着目点を順に整理してみる。

1）最初に統計データに基づく因果探索を行い、因果知識グラフを作成しておく。各種の統計的分析手法も実施しておく。

2）次に、検出された因果性、例えば A と B の事象について LLM が有している知識を引き出す。具体的には、「A が B に影響を与えるメカニズムについて、専門知識に基づいて説明してください」と問いかける。その際、各種の統計的手法の分析結果も与え、LLM の推論能力を引き出す CoT（Chain of Thought）の技法を用いる。すなわち、段階的な推論過程を明示させるようなプロンプト（LLM への指示文）を与える。

3）ここまでの分析で得られた因果推論を LLM に与え、これが正しいかどうかを Yes/No で LLM に回答させる。いわば、LLM の知識や推論能力を多重に活用する。確率的言語生成モデルである LLM の出力には揺らぎがあるため、同じ質問を繰り返して Yes/No の平均確率を採用する。

4）LLM から得た情報を統計的手法に活用し、全事象（変数）間の因果性を再検証する。以上の分析過程により、統計データに基づく因果探索や統計的手法の情報処理と、LLM が有する情報や推論能力を相互参照的に統合させる。

論文では具体例として、因果探索の研究で広く利用されているデータセット（自動車の燃費に関連した変数セット、気候地形関連の変数セット）や、LLM がデータセット情報を有していない（LLM の学習に利用されていない）非公

開の健康診断データセットを用いた検証を行っている。上記の手法を適用することで因果探索の精度がどれだけ改善するかを計測し、LLM の持つ情報と推論能力が改善に寄与することを確認している。

ちなみに、最新の LLM は専門家が有する医療情報も学習済みであり、健康診断データセットそのものに関する情報はなくとも、医療専門知識や推論能力が貢献していると考えられる。解題筆者は、自らの実証分析研究で大学院クラスの計量経済学の知識やその数式展開、プログラミング実装（コードの書き方やライブラリの存在・用法を含む）を LLM から引き出しているが、専門教科書と突き合わせてもハルシネーションが生じることが殆どなくなってきている。LLM の劇的な進化を因果推論に統合した本手法は、LLM のさらなる進化によってパフォーマンス改善が期待できる。

齋藤論文

統計的因果推論では、RCT のような実験介入や、他の環境が同一という仮定のもとで介入前後の変化を比較する手法（差の差分法）、介入処理群と非処理群で介入以外の要素が類似したペアを見つけて比較する手法（マッチング法）、偶然の環境差に依存する方法（回帰不連続デザイン）、操作変数法などがよく利用される。しかし、検証したい内容によってはこうした手法が利用できないケースも少なくない。例えば、EC や音楽映像配信の推薦システムなどパーソナライズドサービスでは提示候補が膨大にあり、単一の施策介入を前提とした分析手法では全く対応できない。ほかにも、「他の環境や要素が類似している」という仮定が成立していなかったり、介入が生じる前後の状況にしか適用できない、適切な操作変数が存在しないといった限界を各手法が抱えている。RCT や A/B テストがユーザー体験を損ねる場合もあるし、新薬テストや治療法選択のように倫理的に実行が難しいケースもある。

こうしたケースに対しては、過去のある環境・ある介入状況のもとで採集されたデータのみに基づいて、複数の介入方策のどれが適切なのか選択を迫られることがある。これを実現する技術はオフ方策評価（Off-Policy Evaluation）と呼ばれており、推薦システムを用いている企業などで実践されている。

本論文は、オフ方策評価の基本概念、主要な推定量、近年の研究動向に焦点を当てて解説したものである。概念解説だけでなく、数式を用いて具体的にどのような計算がなされているのかが平易に説明されている。前述の「過去のある環境」とは、例えばユーザーの視聴・購買履歴、過去の株価推移などであり、機械学習分野の用語法で特徴量と呼ばれる。それが観察されたもとで、ある介入（方策の選択）が行われている。例えば特定の推薦アルゴリズムの実行、株式ポートフォリオの選択などである。これを行動と呼ぶ。所与の特徴量と行動のもとで、報酬（購買、視聴時間、投資損益など）が決定され、これは機械学習などの関数で推計される。期待報酬の最大化を目指すのが、意思決定方策の役割である。

オフ方策評価の目標は、ある特徴量に対して、期待報酬を最大化するような方策が採られ、その結果、報酬が観察されたという「過去の状況に依存した」ログデータだけに基づき、いまだ試したことがない方策の性能（期待報酬）をより正確に推定する方法（推定量）を構築することである。これは難しい作業であるが、過去データのみから算出可能で、試験コストを要せず、迅速な実行が可能というメリットを有している。それゆえ、実務で利用されており、推定量改善に向けた努力もなされている。

論文では、三つの代表的な推定量が紹介され、その長所短所を実際の分析事例を用いて定量的に解説している。各推定量の精度は、期待報酬を推計する関数の精度や、観測値に含まれるノイズ、利用可能な観測値の数（多いほど精度が安定する）、検証すべき方策の数（多いほど精度にマイナスに効く）などに左右される。実務ではこうした要因を考慮しながら、各問題に適切な推定量が選択されている。

この推定量には、期待値としての正確さと、分散（誤差の振れ幅）の小ささという二つの評価軸がある。資産ポートフォリオ選択問題における平均分散アプローチと同様な性質である。様々な推定量が開発されるにつれ、対処する問題の性質にあわせて、どの推定量を選択すればよいかという実務的な課題が生じている。一つの評価法は、平均分散のトレードオフに関する自らの選好を明らかにしたうえで、選択を行うものである。例えば、医療手法の選択のように大きなリスクをとれない場合、分散の最小化に重きが置かれる。こうした発想は、資産ポートフォリオの運用方針の選択と類似している。本稿のもう一つの貢献は、この方針選択に関する新しい指標の提示である。シャープレシオを参考にリスク調整後の性能評価指標を明示的に算出し、その利用方法を解説している。

増島・難波論文

最後の論文は、当研究所が実施しているアンケート調査を用いたRCTと操作変数による因果推論の事例である。これらは統計的因果推論の典型的な活用法であり、当研究所でも因果推論の王道的な手法の実践が行われていることを紹介したものである。本論文の詳細は、研究所のワーキングペーパーシリーズで公表されている。

アンケート調査は、株式などの金融資産に加え、暗号資産やステーブルコインなど新しい金融資産を対象としており、米国や中国、ドイツでも同じ調査を実施しているため国際比較が可能となっている。暗号資産などについては日本の企業や組織が実施しているものの中では最大規模のものとなっている。過去3回のアンケート調査の集計結果も公表されている。

本論文では、暗号資産を保有している人の特徴分析に焦点が当てられている。一般的な傾向として、男性、若年層、高所得層、低学歴層、リスク回避度が低い層ほど暗号資産に投資する傾向が強いことを明らかにしている。また、インフレ期待が高いほど、インフレ期待と成長期待の不確実性が高いほど、投

資傾向が強いことも判明している。

本論文の白眉は、金融リテラシーが暗号資産投資に与える影響について、逆方向の因果性や欠落変数バイアスから生じる内生性を考慮した分析がなされている点である。本研究に先立つ研究（他の研究所スタッフによるもので、研究所ワーキングペーパーとして公表済み）では、金融リテラシーと暗号資産投資には非線形関係があることが指摘されていた。具体的には、金融リテラシースコアが中程度のグループにおいて暗号資産の投資経験割合が高く、高・低グループでは同割合が低いという関係である。本論文の分析は操作変数法を用いることにより、こうした関係が見せかけのものであり、金融リテラシーが高いほど暗号資産の投資経験割合が高まり、金融資産ポートフォリオに占める暗号資産のシェアが高まる可能性を示唆している。

これは、表面的な関係性や通常の回帰分析からは特定できない因果関係を識別したものであり、操作変数法の有効性を示した実証分析となっている。なお、計量分析に必要な条件を満たした適切な操作変数が利用できたのは、アンケートの設計段階からリサーチプランが検討されていたためであり、適切な操作変数が見つからないという同手法の弱点をリサーチデザインで克服している点も本研究の特徴である。

もう一つの特徴は RCT の活用であり、こちらもアンケート設計段階からリサーチデザインが実施されている。具体的には、情報提供が暗号資産の購入行動に与える影響を検証している。調査対象者を無作為に二つのグループに分け、何の情報も提供しない対照群とビットコインの収益率に関する情報、「ビットコインの価格は過去 5 年で 6 倍以上、過去 10 年で 100 倍以上になった」を提供した処置群について、1 年後の望ましいポートフォリオを尋ねている。その結果、情報提供が暗号資産の購入意欲や保有割合を高めることを示唆する結果を得ている。この分析は、金融資産に関する情報提供が投資行動を促進する可能性を示唆するものである。

研究会報告書

本号の最後には、当研究所が主催している「次世代金融インフラの構築を考える研究会」の第二次提言「例示」を掲載した。この研究会の問題意識は、本号のテーマと関連している。提言書の冒頭にある問題意識の提示では、「金融 API やブロックチェーン技術、ビッグデータの活用に代表される情報技術の革新によるデジタル化社会の進展に伴い、新しい決済・送金手段、暗号資産などのデジタル金融資産、分散型金融サービス (DeFi) の登場など、金融サービスの提供主体・手段等に変化が生じており、今やデータが収益の源泉となり、情報生産機能の高度化が金融機関の経営課題となっている」と指摘している。

提言書の中でも、以下のような言及がなされている。

- 「金融・非金融領域を跨ぐデータの活用による利用者ニーズの可視化」
- そのための「データ基盤等の整備や IT システムへの実装」
- 「非金融分野における利用者動向のデータ（会計情報・購買履歴・生産活

動情報等）の利活用を通じた金融サービスへの展開」

- 「業務サービスの一環として派生した顧客データを他の用途に活用することで、需要予測やダイナミックプライシング、推薦システムなどのパーソナライゼーションサービスを進化させて高収益化に繋げるデータ・マネタイゼーション」
- 「データサイエンス（AI/ 機械学習 / 因果推論等）と新しい IT システムの構築手法が活用され、ビジネス実装と運用結果のフィードバックに基づく連続的改善という高速サイクル」
- これにより「短期間で効率的にサービスの高度化や新サービスを発見」
- 「利用者データの権利（CDR ＝ Consumer Data Rights）の法制化」

これらを一言で要約すると、金融サービスは非金融サービスとのデータ連携によって統合され、「情報の循環回転ビジネスモデルやデータ駆動型システムに発展していく」というメッセージとなる。

ここまで解題を読まれた読者は、提言書の内容と因果推論の関連性に気付かれたであろう。因果推論は次世代金融インフラの重要なピースである。また、因果推論のポテンシャルを引き出すためには、1）データの観測・蓄積・活用に必要な IT インフラやこれを支える人材、2）データや IT インフラの標準化・相互互換性の推進と情報のセキュアな流通、3）分析推進のためのデータサイエンティストとビジネスドメイン知識を持った人材の連携、4）全体推進役としての経営者や政策当局者の理解が重要となる。

今号の特集、「因果推論の最前線」がその一助となれば幸いである。

金融分野における因果推論の展開

― 統計的手法・因果 AI・自然言語処理の三潮流とその展望 ―

和泉 潔｜東京大学大学院工学系研究科 教授

和泉 潔

東京大学大学院工学系研究科教授。1993 年東京大学教養学部基礎科学科第二卒業。1998 年同大学院総合文化研究科広域科学専攻博士課程修了、博士（学術）。同年、電子技術総合研究所（現・産業技術総合研究所）に入所し、2010 年まで勤務。2010 年より東京大学大学院工学系研究科システム創成学専攻准教授、2015 年より現職。専門は金融情報学、市場シミュレーション、金融データマイニング。IEEE、人工知能学会、情報処理学会、電子情報通信学会会員。

要約

本稿は、金融分野における因果推論の主要な三つの潮流を整理し、それぞれの背景・手法・応用事例を概観する。第一は統計的因果推論であり、政策効果測定や企業ガバナンスの影響分析といった実証研究に広く用いられている。第二は機械学習技術を活用した因果 AI であり、複雑・高次元データによる金融予測、非定常時系列における因果構造の発見、さらには説明可能な AI（XAI）への応用が進展している。第三は自然言語処理を用いた因果推論であり、稀な経済現象の要因解明、企業への経済変動の影響分析、株価に対するニュースインパクトやリード・ラグ効果の分析など、経験的因果の知識を活用する領域に応用されている。今後は、大規模言語モデル（LLM）と統計的因果推論の融合や、反実仮想シナリオを用いたエージェントシミュレーションの導入により、因果推論の精度が向上し、応用範囲が拡大することが期待される。

1．経済・金融は因果で動く

経済ニュース記事や相場解説では、株価の動き、商品の売り上げ、雇用、貿易といった経済事象における因果関係が頻繁に登場する。たとえば、「将来の少子高齢化が○○を引き起こし、それに関連した○○への需要が増加する」、あるいは「現在の株価下落は○○による市場参加者のリスク警戒心理を反映している」といった形で、経済事象の波及効果やその原因が説明される。このように、因果関係は経済や金融の動きを理解し説明する際に中心的な役割を果たしている。さらに、人々はこうした因果関係の理解に基づいて経済行動を選択し、それらの行動の集積が新たな経済現象を生み出していく。経済は、因果関係の理解と行動の連鎖によって日々動いているのである。

この因果への注目は、日常的な経済理解にとどまらない。アダム・スミスの『国富論』（Smith, 1776）の原題に「原因（causes）」という語が含まれているように、因果関係は経済学の基本的概念である。この伝統はアリストテレスにまで遡り、経済学は長らく物理学をモデルとしながら、経済現象における因果法則の解明を目指してきた（Hoover, 2018）。哲学者であり経済学者でも

あったデイヴィッド・ヒュームは、実践的な経済学を本質的に因果科学として捉え、「原理を知ることは公共政策の実行において有用である」と主張している（Hume, 1785）。一方で彼は、因果関係の本質を完全に理解することには懐疑的であり、この因果関係の認識論的な限界と政策的応用との間の緊張関係は、ヒューム以降の経済分析においても一貫して問題となり続けている。

1.1 背後の因果関係を知らなければ大規模データは価値を持たない

近年の金融市場分析では、従来の経済統計や財務諸表に加え、ソーシャルメディア、衛星画像、購買履歴など、これまで活用されてこなかった多種多様な大規模データの利用が急速に進展している。これら「オルタナティブデータ」は高度な機械学習技術と組み合わされ、金融機関は自社あるいは外部の情報技術者に最先端の機械学習手法を適用させ、より優位な分析手法の開発を推進している。しかし、データの背景にある構造的関係、特に因果関係を無視したまま、膨大な情報を機械的に処理して投資戦略に反映させるアプローチは、長期的に深刻なリスクを伴う可能性がある。

実際に、Dessaint et al.（2024）は、オルタナティブデータの利用が短期予測の精度を高める一方で、長期的な市場予測にはむしろ負の影響を及ぼすことを示している。Institutional Brokers' Estimate System（I/B/E/S）のデータを用いた分析によれば、1980 年代以降、アナリストによる 1 日～ 1 年先の業績予測精度は向上しているが、2 年以上先の長期予測精度は一貫して低下している。特にソーシャルメディア（Stocktwits など）へのアクセス頻度の高いアナリストほど、短期予測の精度は 0.54 ポイント改善する一方で、長期予測は 1.51 ポイント低下しており、長期予測能力の相対的な劣化が観察される。

過去データへの過信が構造変化に対応できないというデータ駆動型アプローチの限界は、以下の事例に顕著である。1998 年には、ロングターム・キャピタル・マネジメント（LTCM）がロシア危機という想定外の事態に対処できず破綻した。2013 年のツイッター・クラッシュでは、ホワイトハウスで爆発という誤報にアルゴリズム取引が過剰反応し、市場に混乱をもたらした。さらに、2020 年の COVID-19 感染拡大では、ビッグデータに基づく定量的ファンドが未曽有の事態に対応できず、大きな損失を被った。これらの事例は、金融市場における大規模データの活用が分析の精緻化と即時性をもたらす一方で、データの背後にある経済事象の波及効果や要因の理解という長期的かつ広範な視点を欠けば、投資判断の誤謬やシステミックリスクを引き起こす可能性があることを示している。今後の市場分析においては、単なる統計的相関に依拠するのではなく、人間行動や制度的背景を含む経済事象の構造的理解に基づいた因果関係の把握が不可欠である。

2. 因果推論：因果関係の解明に挑む科学的アプローチ

因果推論（causal inference）とは、観察データから、単なる相関関係で

はなく、原因と結果の関係を推定する手法である。Pearl and MacKenzie（2018）は、因果関係を理解するためには、人間が観察、介入、反事実という三つのレベルの認知能力を備えている必要があると述べている（図表 1）。これらは、本稿で紹介する因果推論の全ての手法の基礎となっている。

図表1.

1. **関連付け（Association）**：見る能力 観察に基づく理解。「X を見たら Y が起きることが多い」など、データ間のパターンを読み取る。 例：株価が下がると金価格が上がることが多い。
2. **介入（Intervention）**：行動する能力実際に行動・介入した場合に何が起こるかを予測し、介入を行う（行わない）を判断する能力。 例：中央銀行が金利を下げたら、住宅市場にどのような影響が出るか?
3. **反事実（Counterfactual）**：想像する能力 「もしあのとき別の選択をしていたら?」という仮想的な過去に立ち返る思考。物事の原因を深く理解しようとする能力。 例：もし破綻した金融機関 Z が救済されていたら、金融危機は防げたのか?

出所）Pearl and MacKenzie（2018）より筆者作成

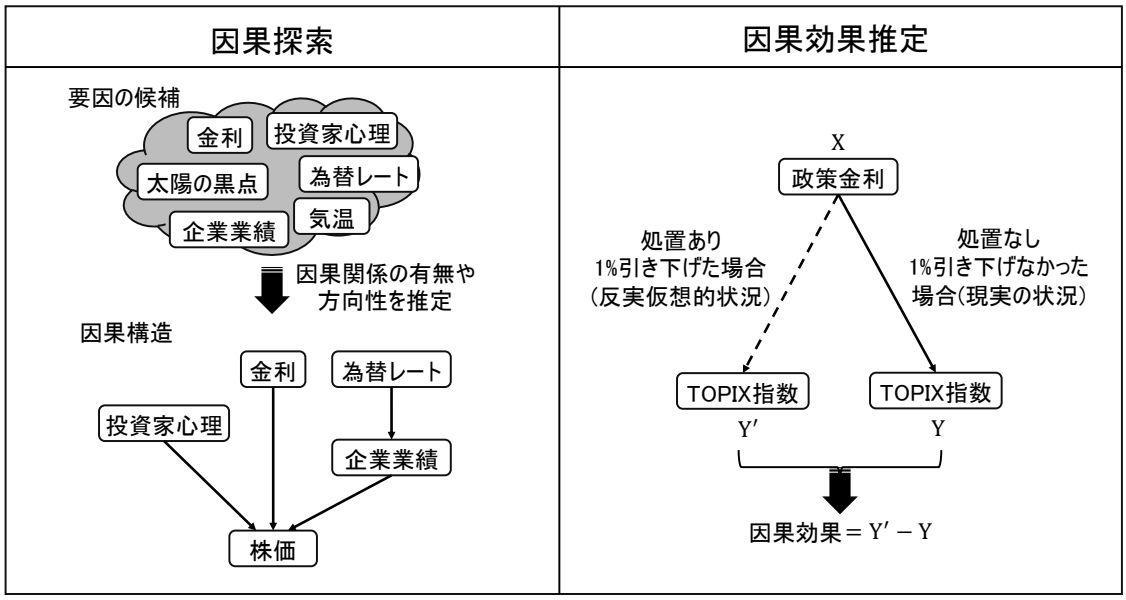
2.1 因果探索と因果効果推定

因果推論は、その分析目的に応じて主に二つのアプローチに大別される。第一に、因果探索（causal discovery）は、変数間の因果的依存関係が未知の場合に、その構造を体系的に表現し自動的に探索するアプローチである。第二に、因果効果推定（causal effect estimation）は、既知の因果関係に基づき、特定の介入や処置が結果変数に及ぼす定量的影響を推定するアプローチである。両者は補完的な関係にあり、特に金融市場のような高次元かつ動的な環境では、両者の適切な使い分けが重要となる。

たとえば、株式市場における複数の要因（金利、為替レート、企業業績、投資家心理指標など）と株価変動との関係を調査する場合を考える（図表 2）。因果探索の段階では、過去の観測データに基づき、「金利が株価に影響を与えるのか」「為替変動は業績を介して株価に影響するのか」といった因果関係の有無や方向性を探索する。この段階では因果構造は既知ではなく、複数の変数間のネットワーク構造の探索が焦点となる。一方、因果効果推定では、因果構造が既に定まっている前提のもと、特定の介入の効果を推定する。たとえば「政策金利を 1% 引き下げた場合、TOPIX 指数は平均してどの程度上昇するか」を推定する場面である。この場合、実際に介入が行われたかのような反実仮想的状況を模擬し、その因果効果を定量的に評価する。

このように、因果探索は因果構造に関する仮説形成のために用い、因果効果推定はその仮説に基づき具体的な政策や意思決定の効果を評価するものである。金融分野における因果推論の応用では、まず因果探索によりモデル構造を同定し、その上で因果効果推定により具体的な行動の有効性やリスクを定量化するという、段階的なアプローチが有効である。

図表2



出所）筆者作成

2.2 統計的因果推論、因果AI、自然言語処理に基づく因果推論

近年、因果推論の研究領域においては、統計学的手法を基盤とする「統計的因果推論（Statistical Causal Inference）」と、機械学習技術を応用した「因果 AI（Causal AI）」という二つの主要な潮流が形成されている。さらに、近年急速に発展している新たな潮流として、自然言語処理（Natural Language Processing, NLP）に基づく因果推論（NLP-based Causal Inference）が注目されている。これら三者の比較を図表 3 に示す。

まず、統計的因果推論とは、観察データあるいは実験データに基づいて、ある介入が結果に与える因果的影響を推定・検証するための統計学的手法の総称である。Rubin の因果モデル（Rubin Causal Model）および Pearl の構造的因果モデル（Structural Causal Model, SCM）に理論的基盤を置き、無作為割り当てや共変量による交絡制御などの明示的な仮定のもとで因果関係を識別・推定する。代表的な手法としては、差の差分析（Difference-in-Differences, DID）、傾向スコア法、操作変数法（Instrumental Variable Method, IV 法）、回帰不連続デザイン（Regression Discontinuity Design, RDD）、および構造推定に基づく因果探索手法（例：LiNGAM）などが挙げられる。

次に、因果 AI とは、機械学習や深層学習のアルゴリズムを活用して、因果構造の発見、因果効果の推定、反実仮想の生成といった因果推論的タスクを遂行する AI 技術群を指す。従来の予測 AI（Predictive AI）は相関に基づく予測性能の向上を主目的としていたのに対し、因果 AI は介入や反実仮想の推論を可能とする点で本質的に異なる。具体的には、データ駆動的な因果構造学習（例：NOTEARS）、ニューラルネットワークを用いた介入効果の推定（例：DragonNet）、さらには深層学習との融合による意思決定最適化など、多様な技術的展開が進められている。

最後に、自然言語処理（NLP）に基づく因果推論は、論文、ニュース記事、SNS 投稿といったテキストデータから因果関係を抽出・推定・構造化することを目的とした技術である。典型的には、因果的な言語表現（例：“X causes Y”、“Y is due to X”）の検出、言語モデルを用いた因果関係の識別、因果知識グラフの構築、さらには生成 AI による反実仮想文の生成などが含まれる。手法的には、ルールベースの情報抽出から、BERT や GPT などの言語モデルを用いた深層学習型アプローチに至るまで、多様な技術が併存している。

図表3

	統計的因果推論	因果 AI	NLP ベース因果推論
目的	数量データにおける因果探索・因果効果推定	複雑・高次元データへの因果推論の拡張	テキストから因果関係の発見・理解
入力データ	数値指標などの構造化データ	構造化データと非構造化データ	テキスト（文書、対話、SNS など）
モデルの解釈性	明示的・可解なモデル（回帰、確率モデル等）	非線形・ブラックボックス的モデル	自然言語による明示的・可解な因果関係提示
理論基盤	反実仮想モデル、構造的因果モデル	深層学習、最適化理論との融合	自然言語処理技術の応用

出所）筆者作成

3．金融分野における統計的因果推論の応用

統計的因果推論は、金融市場や制度の変化が経済主体に及ぼす影響を定量的に評価する有効な手段として、金融分野で広く活用されている。特に、観察データの特性やその発生構造の探求に重きを置く実証金融研究において、単なる相関ではなく因果関係を特定するための重要な方法論として確立されている。以下では、金融分野における統計的因果推論の具体的な応用事例のごく一部を紹介する。

3.1 金融政策の効果測定：操作変数法（IV法）

中央銀行による政策金利の変更が実体経済や金融市場に与える影響は、因果推論の代表的な研究対象である。しかし、政策金利には内生性が伴うため、単純な回帰分析ではバイアスが生じてしまう。この課題を解決するため、操作変数法（IV 法）が活用されている。Kuttner（2001）は、FOMC 声明発表前の連邦ファンド先物価格から算出した「予想外の政策金利変化（monetary policy shocks）」を操作変数として用い、金利変更が株価に与える因果効果を推定した。

3.2 ガバナンス制度改革と企業業績：差の差分析（DID）

制度変更が企業行動に与える影響を測定する手法として、差の差分析（DID）が効果的である。Gompers et al.（2003）は、企業統治の強化が株主リターンに与える影響を分析した研究において、州ごとのガバナンス規制変更を外生的ショックとして扱い、DID を用いて制度導入前後の変化を比較検証した。

3.3 金融資産間の因果構造探索：LiNGAM

リスク管理とポートフォリオ構築においても、因果構造を考慮したアプローチが進展している。金融資産間の構造的因果を特定することで、リスク要因の源泉を明らかにし、ヘッジ戦略の最適化やストレステストの設計が可能となる。最新の研究では、LiNGAM（Shimizu et al., 2006）などの因果探索手法を活用して、株式やマクロ指標間の因果構造を抽出する取り組みが報告されている（Yamamoto, 2020）。

4．金融分野における因果AIの応用

因果 AI は、統計的因果推論と機械学習を融合することで、意思決定に不可欠な課題に対する新たな分析枠組みを提供している（Hurwitz & Thompson, 2023）。従来の統計的因果推論が構造化された数値データを主な対象としていたのに対し、因果 AI は機械学習との融合により、テキスト・画像・音声などの非構造化データを含む、高次元で複雑なマルチモーダルデータまで扱えるようになった。この技術的進展を背景に、因果 AI の金融分野への応用も急速に広がっている。以下では、代表的な研究領域とその成果を概観する。

4.1 マルチモーダル因果学習による金融予測

Zhang et al.（2025）は、マクロ経済イベントに対する市場反応を予測する因果 AI フレームワーク「CAMEF」を提案した。この手法は、政策声明などのテキストと時系列金融データを統合し、因果学習モジュールと大規模言語モデル（LLM）を用いて反実仮想的シナリオを生成・評価する。米国の 5 種類の資産（株式、債券など）を対象とした実証分析では、従来の深層学習 Transformer ベースの手法を上回る予測性能を示した。

4.2 非定常時系列における因果構造発見

Sadeghi et al.（2023）は、金融時系列の非定常性と時間遅延構造を考慮した因果探索フレームワーク「CD-NOTS」を提案した。GDP・株価・金利など複数のマクロ・市場変数から時変的かつ非定常な因果構造を抽出し、要因ベースの投資判断やポートフォリオ構築への応用可能性を示した。特に、因果グラフに基づく意思決定支援の基盤構築に貢献している。

4.3 因果特徴選択によるリターン予測モデル

Oliveira et al.（2024）は、DynoTEARS や改良型 VAR-LiNGAM など

の因果探索アルゴリズムを用いて、株式・ETF のリターン予測モデルにおける説明要因となる特徴量の選択を最適化した。金融危機期などの高不確実性環境下で、因果的特徴選択に基づくモデルは非因果的アプローチより高い安定性と汎化性能を示し、過剰適合も抑制できることが明らかになった。

4.4 株価間の時間的因果ネットワーク

Li et al.（2024）の「CausalStock」フレームワークは、ニュース記事から株価間の時間的因果ネットワークを構築し、株価予測モデルに統合した。ラグ依存構造を考慮したニュースベースの因果探索と Functional Causal Model による統合的処理により、米国・中国・日本・英国を含む 6 つの主要市場で多銘柄同時予測の精度を向上させた。

4.5 説明可能なAI（XAI）の金融分野への応用

機械学習技術の急速な発展に伴い、その判断プロセスの「ブラックボックス化」による不透明性や説明責任の欠如が課題となっている。この課題に対し、機械学習モデルの判断根拠を人間が理解できる形で提示する説明可能な AI（Explainable AI, XAI）が注目を集めている。一方、因果 AI は介入や反実仮想といった因果推論の機能を AI に組み込み、因果理論に基づく説明性の構築を目指している。XAI がモデルの挙動の可視化と理解可能性を追求するのに対し、因果 AI はより構造的で実践的な意思決定支援のための説明性を目指している。

両アプローチは「説明性の向上」という目的を共有しており、近年では因果 AI を XAI の有力な手法の一つとする見方も増えている。XAI には因果推論とは異なる説明性実現手法もあり、代表的なものが特徴量帰属法（Feature Attribution）である。これは入力特徴量のモデル出力への影響度を定量化し、予測結果への各要因の寄与度を説明する。特に、ゲーム理論に基づく特徴量帰属法 SHAP（SHapley Additive exPlanations）は、金融分野で以下のような実践的応用が進んでいる。

- 原油価格の変動要因分析（金田他 , 2022）
- 金融システムにおけるリスク情報要因の特定（Bluwstein et al., 2023）
- 企業の格付け分類モデルにおける要因分析（橋本他 , 2023）

5. 金融分野における自然言語処理に基づく因果推論の応用

5.1 なぜ100年に一度の金融危機の要因がわかるのか

統計的な分析に必要な十分なデータが存在しない極めて稀な現象―たとえば「100 年に一度の金融危機」においても、専門家や一般の人々はその要因をある程度推定することができる。人間は、観察、介入、反事実という三層の認知能力（図表 1）を駆使することで、データが不十分な場合でも、複雑な現象の因果関係を構築することが可能である。

統計的因果推論や因果 AI が対象とするのは、比較的安定した再現性のある

因果関係であり、十分なデータがあれば統計的手法によって検証できる。これを「統計的因果」と呼ぶ。たとえば「物体に力を加えると運動する」といった自然科学の第一原理に基づく因果関係が典型例である。しかし、経済現象のように人間の行動が深く関わる領域では、統計的因果だけでは事象を説明することが困難である。その主な理由は、人間が出来事をどう認知し、それにどう反応するかという人間の認知と行動ルール自体が因果構造の中核をなすためである。こうした因果関係は状況や文脈によって変化しやすく、自然科学のような客観的・普遍的な因果系列を統計的に導き出すことは難しい。人間は自らの観察や経験に基づいて現象を理解しようとし、その際には状況に応じた因果メカニズムの把握が重視され、実践的・具体的な文脈に即した因果関係が追求される。このような因果の理解を「経験的因果」と呼ぶ。

統計的因果と経験的因果は対立するものではなく、むしろ相補的な関係にある。統計的因果推論は経験的因果の知見を数理的に定式化しようとする試みであり、経験的因果は統計的推論の前提や解釈の妥当性を支える基盤となる。両者を統合することで、因果推論の信頼性と説明力が高まり、より説得力のある理解が可能となる。この関係性は西洋医学と東洋医学の対比に似ている。西洋医学は実験や統計に基づく定量的・科学的アプローチを重視し、再現性のある知識体系を構築してきた。これは統計的因果に相当する。一方、東洋医学は患者個々の症状や体質に基づく全体論的・経験的アプローチをとり、経験的因果に近い。いずれか一方だけでは捉えきれない現象に対して、両者の知見を融合することでより包括的な理解と治療が可能となる。

同様の関係は、経済学における計量経済学とナラティブ経済学の間にも見られる。計量経済学は統計モデルや数理的手法により因果関係を定量的に分析する。一方、ナラティブ経済学は物語や歴史的文脈に着目し、経済主体の行動や意思決定の背後にある意味や動機を質的に解明しようとする。両者を組み合わせることで、単なるデータの相関を超えて、現象の背後にある人間的・社会的意味までを含めた深い理解が実現される。金融分野では、株価の変動要因や投資リターンの予測において統計的因果推論が中心的な役割を果たしているが、それだけでは市場参加者の複雑な行動を捉えきれない。実際には、トレーダーや投資家の行動、意思決定プロセスを質的に観察し、取引パターンや市場心理を経験的に理解することが求められている。こうした統計的因果と経験的因果の両輪によって、経済や金融といった複雑系の因果構造に対するより深い洞察が可能となるのである。

5.2 自然言語処理による経済的因果連鎖の構築と検索

経験的因果の知識は、人々が日常的に使用する自然言語によるテキストデータの中に記述されている。近年、自然言語処理（NLP）技術を用いて、大量のテキストデータから因果情報を機械的に抽出・活用する取り組みが進展している。鳥澤（2003）や乾他（2004）の研究をはじめとして、因果関係情報を扱う技術の研究開発も活発に行われている。

本節では、経済・金融分野に特化して開発した因果連鎖の検索システムの概要とその応用例を紹介する。和泉他（2021）は、経済関連のテキストデータ

から因果関係を抽出・データベース化し、特定のフレーズに関連する因果連鎖を検索できるシステムを開発した。このシステムにより、ユーザーは入力した語句に対する因果連鎖を確認・編集でき、経済的な波及効果とその要因を把握することができる。

5.2.1 経済的因果関係の抽出とデータベース構築

和泉他 (2021) では、上場企業の決算短信に記述された情報から、手がかり表現を用いて因果関係を抽出した。

- 使用したテキストデータ：2012 年 10 月から 2024 年 10 月に発行された決算短信テキスト
- 抽出された因果関係数：約 100 万件

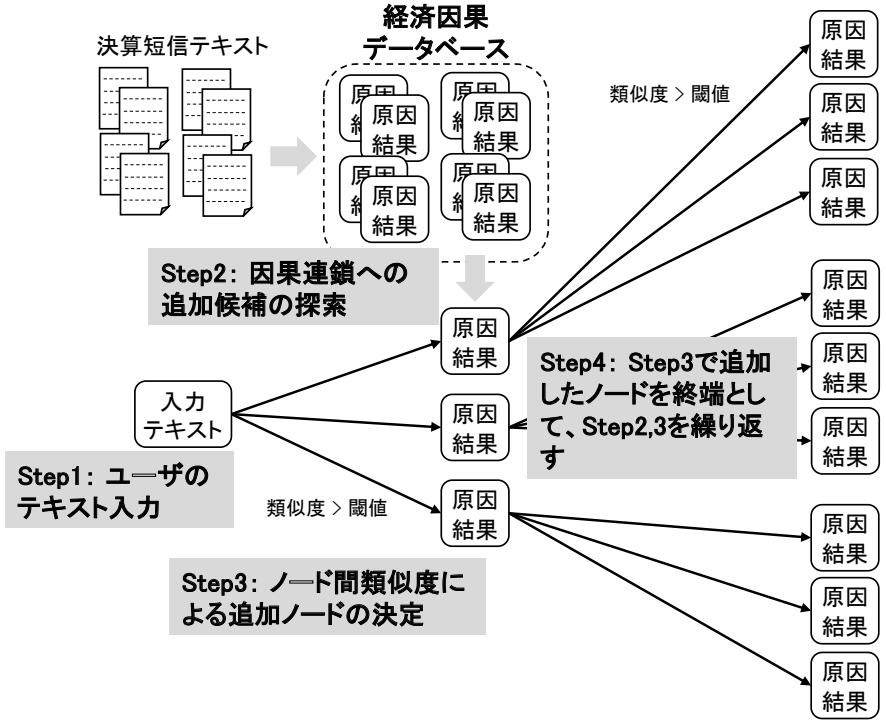
これらの因果関係は、発行日や銘柄情報とともにデータベースに保存され、検索・分析に活用される。

5.2.2 因果連鎖の構築アルゴリズム

因果連鎖の構築は、以下の四つのステップで実行する（図表 4）。

1. ユーザーによるフレーズの入力（最初の終端ノードの設定）
2. 入力された終端ノードと類似する因果関係の選択
3. ノード間の類似度を計算し、閾値αに基づいてネットワークを拡張
4. 新たに追加されたノードを次の終端ノードとして、ステップ 2・3 を繰り返す

図表4



出所）和泉他(2021)より筆者作成

因果連鎖構築アルゴリズムに基づき、ユーザーが因果関係を確認・編集できる検索システムを実装した（図表 5）。このシステムでは、ユーザーが検索の起点となるテキストを入力し、「波及効果」または「要因検索」を選択して検索を実行すると、関連性の高い因果関係が表示される。

図表5

出所）因果連鎖探索の試作システム https://socsim.t.u-tokyo.ac.jp の「公開ツール」-「因果チェーン検索」-「因果チェーン検索サービス」

5.2.3 経済的因果連鎖検索の応用例

和泉他 (2021) で提案されたような因果連鎖検索の手法を、経済金融分野に応用した事例を紹介する。

経済変動の企業への影響分析

令和４年度中小企業実態調査（デロイトトーマツコンサルティング合同会社，2023）では、感染症の流行、自然災害、資源価格の高騰などの経済社会上の変動が中小企業に及ぼす波及効果を分析した。これらの影響は、特定の企業だけでなく、サプライチェーンの断絶、風評被害、自粛の連鎖を通じて広範に波及する。このような複雑な影響構造を把握するには、単一の事象だけでなく、その因果的な派生を追跡する必要がある。本調査では、因果連鎖検索の手法として、決算短信のテキストと日経新聞記事を階層的に組み合わせるアプローチを採用した。分析の結果、「感染症の拡大」から「外出機会の減少」といったネガティブな影響から、「家飲み需要の増加」といったポジティブな派生効果まで、さらには製造業における「Mixed Reality（MR）技術の活用」といった新たな経済トレンドも特定された。この手法により、実務上重要な因果の連鎖を発見し、支援対象の特定や政策立案に活用できることが実証された。

株価へのニュースインパクト分析

因果連鎖検索は、ニュースが金融市場に与える影響を分析する上で有効な手法であり、ニュースの解釈と市場反応の関係を定量的に捉える基盤となる。

Izumi et al.（2020）では、小麦価格に関する二つのニュース事例を用いて、関連企業の株価への影響を実証的に検証した。2018 年 7 月 24 日の「フランスで小麦減産見込み」や 2018 年 9 月 6 日「ロシアが小麦の輸出制限をしない」というニュースで、市場はニュースの背景にある遠因に注目し、因果連鎖上の企業の株価に影響が及んだことを明らかにした。この手法は、ニュースに基づく株価変動の因果構造の理解に有効であり、関連企業の選定や市場反応の予測精度向上に貢献する。

株価におけるリード・ラグ効果分析

Nakagawa et al.（2019）では、因果連鎖を用いて、株価におけるリード・ラグ効果（ある先行銘柄のリターンが時間差を伴って別の遅行銘柄のリターンに影響を与える関係）を推定した。具体的には、因果連鎖における「原因」企業群に顕著な株価変動が生じた際に、その影響が「結果」企業群へとリード・ラグ的に波及するかを実証的に分析した。その結果、この手法に基づく投資戦略が有意な収益性を示し、日本株式市場において因果的に結びついた企業間での株価予測可能性が存在することが確認された。

6．経済金融分野における因果推論の将来

6.1 大規模言語モデルと統計的因果推論の融合

近年、大規模言語モデル（Large Language Models, LLM）の急速な進展により、従来の課題を克服し、因果推論タスクへの応用可能性が注目されている。統計的因果推論は観察データから因果関係を明らかにする強力な枠組みだが、現実世界の複雑な現象に対して意味のある一貫した因果モデルを構築するには、ドメイン専門家の知識を制約として導入することが不可欠である。しかし、専門知識を体系的かつ網羅的に取得することは困難であり、この点が実用上の主要な制約となっている。

このような課題に対し、LLM と統計的因果推論の統合による新たなアプローチが期待されている（Takayama et al., 2025 および三内他, 2025）。ただし、実用化には、LLM の出力の信頼性向上および統計的因果構造との相互作用に関する理論的・実証的理解が不可欠である。LLM は本質的に確率的パターン認識に基づいており、相関と因果の混同、専門用語の多い領域や非英語環境での誤推論の増加、言語表現やドメインの変化に対する脆弱性といった構造的限界が指摘されている（Takayanagi et al., 2024）。LLM の出力は因果構造の本質的理解というよりも、因果情報を含む文章を表面的に再現するような生成にとどまる傾向が強い。

より精度の高い因果推論を実現するには、LLM と統計的因果推論を統合する新たなフレームワークの構築が必要である。そのためには、構造化された因果モデル学習との連携を通じて、自然言語処理と因果推論の相互補完的活用を可能にする設計が求められる。

6.2 反実仮想としてのエージェントシミュレーション

社会経済分野において、反実仮想的状況を統計的手法や自然言語処理によって観察データから推定するアプローチには限界がある。特に、創発的で予測困難な現象や制度変更に伴う非線形的影響の把握には、既存のモデル構造に依存する従来手法では十分に対応できない。このような背景から、複数の自律的意思決定主体（エージェント）で構成されるマルチエージェントシミュレーション（Multi-Agent Simulation: MAS）が注目されている。MAS は実世界の社会経済構造を模倣し、制度変更や介入策に対する反実仮想的な影響を直接的に再現・評価できる枠組みとして期待される。

水田（2024）は、人工市場と呼ばれるエージェントベース金融市場シミュレーションに着目し、呼値刻み変更、レバレッジ規制、空売り規制といった実市場における制度導入の影響を評価した研究などを紹介している。反実仮想的な「ありえた副作用」の抽出および政策設計への応用可能性を示し、制度介入前に市場構造の変容や予期せぬ副作用を検証する手段としての有効性を主張している。

森下ら（2025）は、反実仮想的シナリオを仮想空間上で再現する手法として、LLM エージェントを活用した経済シミュレーション環境「EconGrowthAgent」を提案した。本環境は、マクロ経済理論に基づき、家計や企業などの経済主体の意思決定過程を大規模言語モデルで模倣し、それらの相互作用を通じて生産活動の動学を再現する。100 体のエージェントによる 20 年間のシミュレーションでは、経済成長とその周辺現象が再現可能であることが示され、モデルの妥当性が検証された。さらに、「超小さな政府への政権交代」や「地球滅亡級の隕石接近」といった現実では観察不可能な極端なシナリオを仮想的に設定し、その経済的影響を分析することで、反実仮想を直接検証するプラットフォームとしての有効性が確認された。

これらの研究はいずれも、「反実仮想を直接検証・予測するツール」として機能する点で共通している。今後、マルチエージェントシミュレーションは、人間集団の意識や行動の変容を仮想的に再現し、政策導入や社会状況変化に対する世論や行動反応を事前に推定するプラットフォームとして、因果関係をより精緻に検証する可能性を示している。

7. まとめ

本稿では、金融分野における因果推論の方法論を、統計的因果と経験的因果の二つに分類し、それぞれの理論的背景と応用可能性について概説した。また、統計的因果推論、因果 AI、自然言語処理を用いた因果推論の手法とその適用例を取り上げ、今後の展望についても論じた。

今後の応用が期待される領域としては、高頻度取引市場における価格変動の因果連鎖の解明や、金融政策の波及効果における統計的因果と市場心理との相互作用の分析が挙げられる。また、ESG（環境・社会・ガバナンス）投資では、企業の ESG 対応が中長期的に企業価値へ与える影響を統計的手法により

評価する一方で、個別企業の事例研究を通じてその意思決定プロセスやステークホルダーとの関係性を理解する定性的分析の重要性も増している。

統計的因果と経験的因果は独立したアプローチだが、相互に補完し合うことで、金融経済の複雑な因果構造を多面的かつ実践的に理解するための有力な手段となる。たとえば、計量分析によって検出されたマーケット・アノマリーの背後にある投資家行動や意思決定の原理を、エージェントベース・シミュレーションや自然言語処理などの経験的手法を通じて解明することが可能となる。両者を統合することで、因果推論はより実践的かつ説得力のある分析枠組みへと発展しうる。今後の因果推論研究においては、この統合的視座のもとで、多様な方法論を融合できる柔軟な手法の構築と金融分野での実践的な応用が求められる。

参考文献

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics,* 145.

Dessaint, O., et al. (2024). Does alternative data improve financial forecasting? The horizon effect. *The Journal of Finance, 79*(3), 2237–2287.

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate governance and equity prices. *Quarterly Journal of Economics, 118*(1), 107–156.

Hoover, K. D. (2018). Causality in economics and econometrics. In The New Palgrave Dictionary of Economics (pp. 1446–1457). London: Palgrave Macmillan UK.

Hume, D. (1785). *Essays, moral, political, and literary* (E. F. Miller, Ed., 1985). Indianapolis: Liberty Press.

Hurwitz, J. S., & Thompson, J. K. (2023). *Causal artificial intelligence: The next step in effective business AI*. Wiley.

Izumi, K., Suda, S., & Sakaji, H. (2020). Economic news impact analysis using causal-chain search from textural data. In *The AAAI-20 Workshop on Knowledge Discovery from Unstructured Data in Financial Services*, New York, USA.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. *Journal of Monetary Economics, 47*(3), 523–544.

Li, S., Sun, Y., Lin, Y., Gao, X., Shang, S., & Yan, R. (2024). CausalStock: Deep end-to-end causal discovery for news-driven multi-stock movement prediction. In *NeurIPS 2024*.

Nakagawa, K., Sashida, S., Sakaji, H., & Izumi, K. (2019). Economic causal chain and predictable stock returns. In *7th International Conference on Smart Computing and Artificial Intelligence (SCAI 2019)*, 655–660.

Oliveira, D., Lu, Y., Lin, X., Cucuringu, M., & Fujita, A. (2024). Causality-inspired models for financial time series forecasting. *SSRN*. https://ssrn.com/abstract=4971119

Pearl, J. & Mackenzie, D. (2018). *The book of why: The new science of cause and effect. Basic Book*.（栗田昭平訳,『因果推論の科学－「なぜ？」の問いにどう答えるか』, 草思社, 2019 年）

Sadeghi, A., Gopal, A., & Fesanghary, M. (2023). Causal discovery in financial

markets: A framework for nonstationary time-series data. *SSRN*. https://ssrn.com/abstract=4678484

Shimizu, S., Hoyer, P. O., Hyvärinen, A., & Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research, 7*, 2003–2030.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London: W. Strahan and T. Cadell.（アダム・スミス『国富論：国の豊かさの本質と原因についての研究』山岡洋一訳 日本経済新聞出版社 2007 年）

Takayama, M., Okuda, T., Pham, T., Ikenoue, T., Fukuma, S., Shimizu, S., & Sannai, A. (2025). Integrating large language models in causal discovery: A statistical causal approach. *Transactions on Machine Learning Research*, *5*.

Takayanagi, T., Suzuki, M., Kobayashi, R., Sakaji, H., & Izumi, K. (2024). Is ChatGPT the future of causal text mining? A comprehensive evaluation and analysis. In *2024 IEEE International Conference on Big Data (BigData)*, 6651–6660.

Yamamoto, R. (2020). Causal analysis of macro-financial linkages using LiNGAM. *Journal of Financial Econometrics, 18*(2), 348–374.

Zhang, Y., Yang, W., Wang, J., Ma, Q., & Xiong, J. (2025). CAMEF: Causal-augmented multi-modality event-driven financial forecasting by integrating time series patterns and salient macroeconomic announcements. In *KDD 2025* (to appear).

和泉潔・坂地泰紀・松島裕康（2021）『金融・経済分析のためのテキストマイニング』 岩波書店.

乾孝司・乾健太郎・松本裕治（2004）「接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得」『情報処理学会論文誌』45(3), pp.919–933.

金田規靖・木全友則・平木一浩・松栄共紘（2022）「SHAP を用いた機械学習モデルの解釈―原油価格の変動要因分析を例に―」日本銀行金融研究所 https://www.imes.boj.or.jp/jp/conference/finance/2022_slides/1111finws_slide2.pdf

三内顕義・高山正行・清水昌平（2025）「大規模言語モデルを活用した統計的因果探索の新展開－統計的因果プロンプティングよる専門知識とデータの融合－」『SBI 金融経済研究所 所報』vol.8 pp.46-61.

デロイト トーマツ コンサルティング合同会社（2023）. 令和４年度中小企業実態調査事業「中小企業の実態把握等のためのデータ利活用に関する委託調査」https://www.meti.go.jp/meti_lib/report/2022FY/000326.pdf

鳥澤健太郎（2003）「「常識的」推論規則のコーパスからの自動抽出」『言語処理学会 第９回年次大会 発表論文集』pp.318–321.

橋本龍一郎・三浦翔・吉崎康則（2023）「格付け分類モデルにおける機械学習の応用：機械学習の説明可能性を高める手法」日本銀行ワーキングペーパーシリーズ

水田孝信（2024）「人工市場：金融市場のコンピュータ・シミュレーション」『SBI 金融経済研究所 所報』vol.5 pp.54-66.

森下皓文・角掛正弥・山口篤季・永塚光一・友成光・森尾学・今一修・十河泰弘 (2025)「EconGrowthAgent: LLM エージェントと経済成長理論に基づくマクロ経済シミュレーション」『人工知能学会全国大会論文集 第 39 回』

視点の違いに着目した金融因果推論の高度化

市瀬 龍太郎｜東京科学大学 工学院 教授
許 子微｜東京科学大学 工学院 助教
陳 穎｜東京科学大学 工学院 博士課程在籍
井上 光太郎｜東京科学大学 工学院 教授

市瀬 龍太郎

1995 年東京工業大学卒業。2000 年東京工業大学博士課程修了。博士（工学）。国立情報学研究所助手、助教授、准教授、スタンフォード大学言語情報研究所客員研究員、文部科学省学術調査官、産業技術総合研究所招聘研究員、東京工業大学教授などを経て、現在、東京科学大学 工学院 教授。専門は人工知能。

要約

因果関係を理解することは、説明可能な AI の実現、特に経済分野における意思決定にとって極めて重要である。しかし、実世界においては、因果関係の同定は容易ではなく、多様な事象データの観測が困難であるという「可観測性の問題」によって、因果関係の把握は困難な問題となっている。本稿では、この課題に対処するために、膨大に存在するテキストから経済専門家の視点を反映した因果関係を取り出し、それを整理した因果知識グラフを自動構築する機械学習フレームワークについて述べる。本フレームワークには、特定の視点に基づき表現された因果推論を抽出するためのデータマイニングアルゴリズムが組み込まれている。実際の市場データを用いて評価した結果、統計的に有意かつ高品質な因果推論パターンが明らかとなり、その頑健性が裏付けられた。これは、視点を考慮した因果知識グラフが、因果推論、経済モデリング、政策分析の高度化に貢献し、多様な専門家の視点を通して経済現象へのより深い洞察を提供し得ることを示している。さらに、経営者の過信に起因する視点バイアスの違いを明らかにする新たなデータセット CC-BizEnv を生成するフレームワークについても述べる。それを用いて、上場企業の経営者に内在する認知バイアスを定量化することができる。

キーワード：因果知識グラフ、経済因果性、視点表現、機械学習、データマイニング、自信過剰

1．はじめに

因果関係を示す際に用いられる事例の多くは、理論的な仮定に強く依存しているため、非現実的で過度に単純化され、実際の状況を十分に反映していないことが多い。この問題は、観察された情報からだけでは因果関係を特定・観測することが困難であることに起因しており、一般に「可観測性の問題（observability problem）」として知られている（Lane, 2020）。因果関係の可観測性を高めるための有効な手段として、膨大に存在するテキストを利用する方法がある。テキストから因果関係を抽出し、整理したものとして、因果

知識グラフがある。因果知識グラフは、構造化された情報を格納し、複雑な関係を可視化する手法として注目されている。例えば、Wikidata（Vrandecic and Krötzsch, 2014）は汎用的な常識的知識を提供し、ICEWS（O' Brien, 2010）は国家や政治家間の政治的イベントや相互作用を記録している。また、FinCaKG（Xu et al., 2025）は、上場企業の年次報告書に基づき、経済領域に特化した因果関係を記録している。

経済分野においては、同一の状況下であっても、それぞれのエージェントが企業パフォーマンスに対して異なる評価を下すことがある。「視点（standpoint）」とは、エージェントの評価的態度を構造的に表現したものであり、そのエージェントが妥当または許容可能とみなす語彙や言語的解釈の集合を捉えたものである（Bennett, 2011）。視点を表現することは、異なるエージェントが、経済現象にどのように反応するかをモデル化する上で極めて重要であり、政策立案、市場予測、リスク評価に資する。さらに、視点が明確に定義されることで、異なる条件下における経済行動のシミュレーション精度が高まり、より豊かな洞察と頑健な結論を導くことが可能となる。

本稿では、異なる視点を持つエージェントが、異なる因果推論を行うことに着目し、それをどのように明らかにするのかについて議論を行う。そのために、テキスト情報から因果関係を抽出し、そこからどのように視点の違いに応じた因果関係を発見するかについて説明をする。第 2 節で、テキストから因果関係を抽出する方法、因果関係を整理して知識グラフにする方法、因果推論において視点の違いを抽出する既存研究について概観する。次に、第 3 節で、機械学習技術を利用することにより、経済専門家の視点を表現する因果知識グラフを自動生成するフレームワーク FinCaKG を説明する。続く第 4 節では、この因果知識グラフを利用し、専門家の視点に応じた因果推論を抽出するためのデータマイニングアルゴリズム ETE-FinCa を紹介する。さらに、実際の市場データを用いてアルゴリズムを評価した結果についても紹介する。第 5 節では、経営者の過信に起因する視点バイアスの違いを検証するための因果エンティティデータセット CC-BizEnv の生成フレームワークを提示する。そして、それを用いた上場企業の経営者における認知バイアスの定量化を行う手法、および、実験結果に基づき、経営者の過信傾向を捉える指標としての有効性についてについて議論する。最後の第 6 節で、本稿の結論および今後の展望について論ずる。

2. 関連研究

2.1 テキストからの因果関係の抽出

因果関係を整理して利用するために、本稿では、テキストから因果関係を抽出し、それを整理して知識グラフの構築を行う。テキストからの因果関係の抽出する方法は、大きく分けて、パターンマイニングアプローチ、機械学習アプローチ、生成的アプローチの 3 通りがある。以下、これらの手法を順に説明する。

2.1.1 パターンマイニングアプローチ

テキストに現れる因果関係は微妙かつ多様であるため、その特定は容易ではない。この課題に対処するため、因果関係を表す動詞、特定の構文構造、語彙的な手がかりなどの言語的パターンを用いて、因果関係を抽出するルールベースの手法が開発されてきた（Luo et al., 2016; Girju, 2003; Radinsky et al., 2012; Ali et al., 2021）。これらの手法では、特定のパターンに当てはまった場合に因果関係があるとして、関係の抽出を行う。さらに、これらの従来型のパターンに加えて、原因と結果のペアの出現確率に基づく統計的アプローチも利用されており、教師なしで因果関係を予測する手法も試みられている（Chang and Choi, 2004）。

2.1.2 機械学習アプローチ

暗黙的な因果関係を発見するためには、表層的なパターンだけでは十分ではない。機械学習手法では、表層的なパターンを超えた、深い意味の把握を試みる。このアプローチでは、因果関係の手がかりを特徴ベクトルとして表現し、分類、回帰、クラスタリングなどのアルゴリズムを用いてその意味を解釈する（Ali et al., 2021）。パターンベースの手法は明示的な特徴の抽出には有効であるが、機械学習は、これらの特徴から一般化を行うため、より微妙な因果関係の抽出に適している（Sorgente et al., 2013; Rink et al., 2010）。近年では、自動的に特徴を抽出できるという利点から、ニューラルネットワークを用いた深層学習モデルが注目を集めている。当初は、再帰型ニューラルネットワークや畳み込みニューラルネットワークが用いられたが、これらは因果構造の抽出において、効果が限定的であり（Elman, 1990; LeCun et al., 1998）、より特化したアーキテクチャの研究が進められてきた（Yin et al., 2017）。近年では、汎用的な関係抽出モデルが因果関係の発見のために応用されている。そのようなものとして、REBEL（Huguet Cabot and Navigli., 2021）、UniRel（Tang et al., 2022）、ReLiK（Orlando et al., 2024）が挙げられる。REBEL は自己回帰型のデコーディング手法により、テキスト情報から、二つのエンティティとその関係を生成することができる。UniRel は、統一的な表現と自己注意機構を導入することで、エンティティと関係を同時に抽出することができる。ReLiK は、特定の戦略を用いることで、エンティティの関連付けと関係抽出を一度の処理で実現するようにし、効率性を向上させている。

2.1.3 生成的アプローチ

生成モデルは、従来の教師あり深層学習手法に代わり、柔軟な選択肢を提供する。特にアノテーションに多大なコストと時間を要する従来手法の問題を克服する手段として注目を集めている（Ye et al., 2022）。これらのモデルは、多様なタスクに容易に適応できるため、知識グラフ構築における構造的情報の抽出に有用である。生成アプローチには、主に、2 つの戦略が用いられている。第 1 の戦略は、タスク特化型のパイプライン戦略であり、生成モデルを順番に適用して、テキストからの因果関係の抽出に必要な、固有表現抽出、関係抽出、エンティティリンキングを行っていく。例えば、LSTM-CRF モデル（Lample et al., 2016; Straková et al., 2019）は、入力トークンに対する注意機構を用い

て、高精度な固有表現抽出を行う。また、KnowGL（Rossiello et al., 2023）は、エンティティのペアを検出し、ラベルや型、優先度付き関係を含む構造化ファクトを生成することで、生成モデルにより、情報抽出の複数のサブタスクを処理する。GENRE（De Cao et al., 2021）や、mGENRE（De Cao et al., 2022）は、文脈を考慮しながら、エンティティ名を自己回帰的に生成することで、エンティティリンキングを行う。第 2 の戦略はエンドツーエンド型の戦略を用い、テキスト情報から構造化されたエンティティ同士の関係を直接抽出する。DeepStruct（Wang et al., 2022）や UIE（Lu et al., 2022）は、事前学習と統一的なフレームワークを活用し、テキストをエンティティ同士の関係を含む構造的表現へと変換する。これらのアプローチは、新たなドメインへの適応が容易であり、多様な情報抽出タスクにおいて高い性能を示している。

2.2 因果関係知識グラフの構築

次に、抽出された関係を用いて、知識グラフの構築を行う。因果関係知識グラフを構築するためには、テキスト中で表現された原因と結果の関係を特定し、体系的に整理する作業を行う必要がある。初期の研究では、主に単語レベルの因果関係のペア抽出に焦点を当てていた。例えば、「bacteria（細菌）」が「disease（病気）」を引き起こすといった関係が扱われていた。Cause Effect Graph（Li et al., 2020）や、CauseNet（Heindorf et al., 2020）では、大規模な語彙レベルの因果ペアが収集され、テキスト中での出現頻度に基づいて重み付けが施されている。しかし、その後の研究では、因果関係が単語単位ではなく、より長いテキストのまとまり（スパン）として表現されている。例えば、「間違えた投資のために、破産した」という例では、「間違えた投資」という原因が、「破産した」という結果をもたらしており、文脈を含んだスパンベースの因果関係表現が必要となる。Frattini et al.（2023）は、スパンレベルの因果関係は、自然言語においてより頻繁に現れるだけでなく、情報量も豊富であることを報告している。近年の研究では、このようなスパンのペアを活用した知識グラフ構築へと焦点が移っている。FinCaKG（Xu et al., 2025）は、その代表例であり、因果スパンを構造化されたグラフとして接続することで、複雑な経済領域の因果ダイナミクスを捉えている。また、CausalBank（Li et al., 2020）のような他のモデルでも、文脈的な手がかりを取り入れた構造的表現の構築が試みられている。このように、単純で表層的なパターンマッチングから、スパンレベルへの移行は、より深い意味理解へと研究の方向性が深化していることを示している。

2.3 因果推論におけるバイアス選択

実際の因果関係の発見において、どの変数を因果の連鎖に含めるかの選択は極めて重要である。特に、特定の方向を持つバイアス（視点）に注目する場合には、この選択が重要となる。従来の研究の多くは、バイアスを最小化することに注目してきたが、ある研究では、特定の操作変数（Instrumental Variables）が、処置効果の推定において、逆にバイアスを増幅させることがあることを示している（Pearl, 2010）。伝統的な手法では、こうした影響を排除することを目指してきた。例えば、傾向スコアマッチング（Propensity Score Matching）（Rosenbaum

and Rubin, 1983）は、処置群と対照群の共変量をバランスさせることで、選択バイアスを減らそうとした。また、PC アルゴリズム（Spirtes et al., 2000）では、潜在的交絡因子が存在しないという仮定のもとで、条件付き独立性検定を行い、因果構造を取り出している。これらのアプローチとは対照的に、本稿ではバイアスの役割を再考する。すなわち、操作変数を排除するのではなく、あえてそれらを因果連鎖の中で明示的に特定することにより、専門家の視点に応じたバイアスを可視化し、その特徴を捉えることを目指した。このアプローチは、因果推論におけるバイアスの新たな表現と解釈の枠組みを提供することができる。

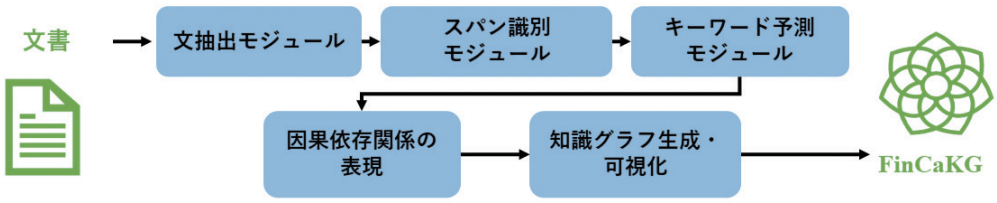
3．因果知識グラフ構築フレームワーク

3.1 概要

次に、因果知識グラフ FinCaKG（Xu et al., 2025）の構築手法について説明する。FinCaKG フレームワークの目的は、専門家レベルの知見と分野固有の信頼性を保ちながら、経済テキスト情報から因果関係の知識グラフを自動的に構築することである。この目標を達成するために、本研究では、各構成要素が明確かつ追跡可能な役割を担うモジュール型のフレームワークを設計した。このモジュール構造により、各処理段階での性能を個別に評価することが可能となり、最終的に生成される知識グラフの信頼性を担保することができる。

FinCaKG を作成する手法の概要を図表 1 に示す。図に示したように、FinCaKG は、以下の 5 つのモジュールから構成されている。

図表1　FinCaKG 構築のワークフロー



文抽出モジュール：因果関係を表現する文を特定して抽出する。
スパン識別モジュール：原因と結果に対応するテキストのスパンを検出する。
キーワード予測モジュール：検出されたスパンから主要な経済用語を抽出する。
因果依存関係の表現：原因と結果のキーワードペアを関連付けし、構造化された因果連鎖を形成する。
知識グラフ生成・可視化：最終的な知識グラフを構築して可視化する。

3.2 情報抽出手法

最初の 3 つの文抽出モジュール、スパン識別モジュール、キーワード予測モジュールでは、事前学習済み言語モデルを活用して、系列分類やトークン単位のラベル付けタスクを実行した。文の分類には BERT（Devlin et al.,

2019）を採用し、因果関係を含む文と含まない文を効果的に判別する。また、特定されたスパンからキーワードを予測するタスクには、XLM-Roberta（Conneau et al., 2020）を用いた。

これらのタスクの中でも、スパン識別が最も困難な課題である。特に、経済テキストにおける因果関係は微妙かつ複雑であるため、従来手法では十分な性能を発揮できない。例えば、次の文を考える。

> 中本氏は、1 日で 9,000 BTC を失い、12 億円の借金を抱えることになった（2014 年 2 月 10 日）

この文では、「1 日で 9,000 BTC を失った」が、原因 (cause) として特定され、「12 億円の借金を抱えることになった」が結果 (effect）として識別される。このような複雑な文に対処するため、私たちは、iLab-FinCau モデル（Xu et al., 2022）をスパン識別の中核として採用した。このモデルは、因果関係グラフの特徴を BERT 埋め込み層に統合することで、スパン識別能力を強化しており、グラフに基づく事前知識を活用することで、性能を向上させることができる。

3.3 因果依存関係の表現

主要なスパンとキーワードを特定した後に、これらのテキスト上の関係を構造化された因果知識グラフへと変換する。FinCaKG フレームワークでは、原因と結果のキーワードを、方向付けしたうえで、接続していく新たな関係表現の仕組みを導入した。この処理により、これまでのモジュールで得られた出力が統合され、FinCaKG 上で、意味的解釈が可能かつ、相互接続された因果関係が形成される。このようにして、FinCaKG フレームワークでは、自然言語処理とグラフベースの表現学習を統合することで、経済文書から、分野に特化した専門家の知識に基づく因果知識グラフを体系的に構築することができる。

4．視点に応じた因果関係の抽出

操作変数（IV: Instrumental Variable）は、反実仮想的なバイアスや交絡バイアスといった課題に対処するために、結果変数とは条件付きで独立でありながら、処置に影響を与える変数として機能する。経済分野における操作変数の構築は、通常、あらかじめ設計された合成操作変数に依存しており、その有効性は特定のアルゴリズムによって評価される。しかし、この従来の枠組みでは、より多様かつ具体的な操作変数が必要となる広範な課題には対応できない。本稿では、因果知識グラフに基づいた手法 ETE-FinCa（Chen et al., 2024）を説明する。この手法は、専門家の視点を最適化することにより、操作変数の情報源を構築し、専門的概念に基づいて操作変数を具体化するとともに、実際の経済データと統合して因果関係を解釈することができる。さらに、定義された高品質な操作変数は、二段階最小二乗法回帰（Two-Stage Least Squares Regression）において、処置変数と結果変数の間の因果関係を統計的に有意な形で同定することが可能である。

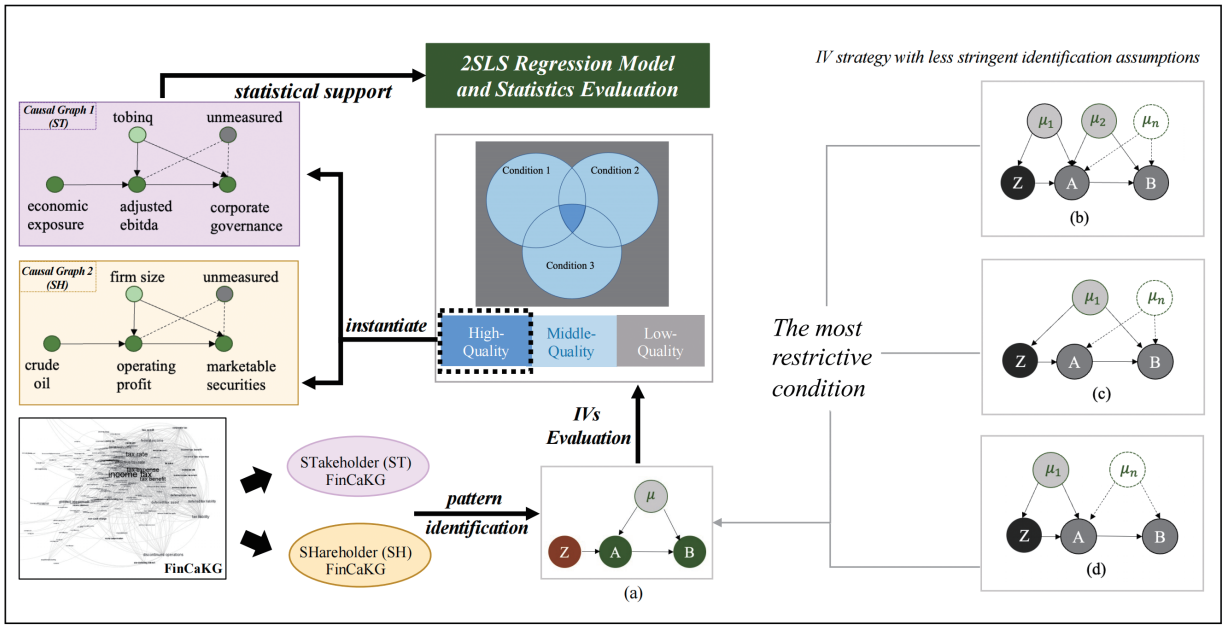
4.1 視点の抽出

「企業は誰のために存在するのか―株主のためか、それともステークホルダーのためか？」

この根本的な問いは、株式会社の登場以降、経済分野において、長年にわたり議論の的になってきた。Chen et al.（2024）は、経済テキストから構築した因果知識グラフを利用することで、辞書を利用した手法や数値指標を利用した手法といった伝統的なアプローチよりも、株主志向とステークホルダー志向という 2 つの視点を、より正確に判別できることを示した。因果知識グラフを利用することで、因果関係の論理的再構成や、認知連鎖に基づく推論が可能となり、企業の文脈に対するより深い知識を利用できることで、2つの立場の分類精度を向上させることができる。

異なるガバナンス視点を比較するために、因果知識グラフ C_{FinCaKG} 全体を、ステークホルダー指向の視点に基づく因果知識グラフ $C_{\text{ST-FinCaKG}}$ と、株主指向の視点に基づく因果知識グラフ $C_{\text{SH-FinCaKG}}$ の 2 つに分割する。そのために、「すべてのステークホルダーの利益のために事業を行う」と明言している企業を特定し、それらの企業が、過去 20 年間の年次報告書において開示したMD&A（経営者による財務状況および経営成績の分析）セクションから因果関係を抽出し、$C_{\text{ST-FinCaKG}}$ を構築した。同様に、「株主価値の最大化を主要な目標とする」企業に対して、$C_{\text{SH-FinCaKG}}$ を構築した。さらに、これら 2 つのサブグラフに含まれるすべての因果関係を統合し、全体グラフ C_{FinCaKG} を生成した。このグラフは、両視点の共通点および相違点を分析する際の基準として利用した。

図表2　ETE-FinCa における、因果変数の識別および解釈のためのワークフロー



出所）Chen et al.(2024)中の図を許諾を得て改変。

4.2 内生性要因の選択

図表 2 は、因果変数の同定と解釈を行うための ETE-FinCa のワークフローを示している。図の右側（b ～ d）は、識別仮定を緩和した操作変数を提示している。これらのうち、我々は最も厳格な条件（a）を採用した。これは、変数 Z が変数 B と直接的に関連づけられる確率を最小化している。先行研究によれば、因果連鎖は 3 ホップを超えると、因果的意味を失う傾向がある（Xu et al., 2025）とされているため、3 ホップ以内の関係を「有効な論理的接続」と定義した。したがって、本タスクは以下のように定式化される：

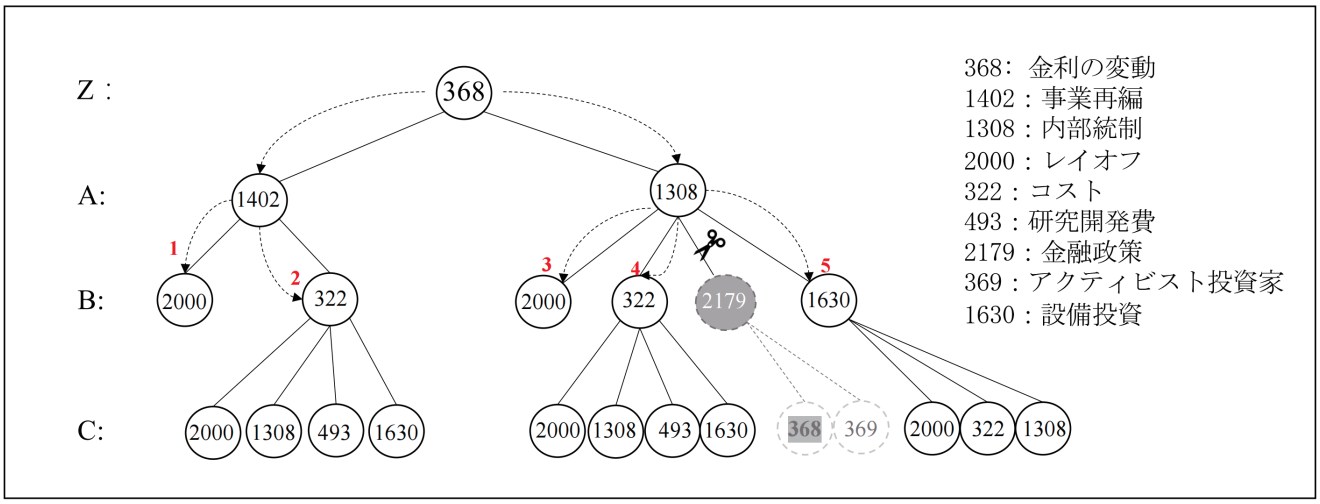
> 「変数 A に関連し、かつ変数 B に 3 ホップ以内で関連しない操作変数 Z を因果連鎖の中から探索する」

このために、FinCaKG 上で「$Z \perp B \mid A$」の条件を満たすパターンを探索する深さ優先探索アルゴリズムを構築した。

図表 3 は、深さ優先探索アルゴリズムにおける因果変数識別の具体例を示している。破線の矢印は、因果連鎖内での遷移を表している。灰色でハイライトされたノードは、グラフから無効と判断されて削除されたノードであることを示す。例えば、金利の変動（ノード 368）を Z としておいた時に、金融政策（ノード 2179）は、金利の変動（ノード 368）との関連が認められたため、因果連鎖から削除される。もし、「事業再編（ノード 1402）が必ずレイオフ（ノード 2000）につながるか否か」という因果関係を特定する場合、金利の変動（ノード 368）が有効な操作変数として機能する可能性が高いことがここから分かる。

操作変数の評価に関して、ETE-FinCa から抽出された操作変数を、以下のスコアリング条件に基づき、「高品質（*3* 点）」、「中品質（*1* ～ *2* 点）」、「低品質（*0* 点）」の 3 段階に分類した。評価指標は以下の通りである。

図表3　DFS アルゴリズムにおける因果変数識別の一例



出所）Chen et al.（2024）中の図を許諾を得て改変。

図表4　操作変数の品質およびエッジノードに関する結果

データセット	因果連鎖パターンの数			
	操作変数はエッジノード	低品質	中品質	高品質
全 FinCaKG	446	9,973	9,618	87
SH-FinCaKG	294	3,260	5,577	95
ST-FinCaKG	345	7,811	5,701	21

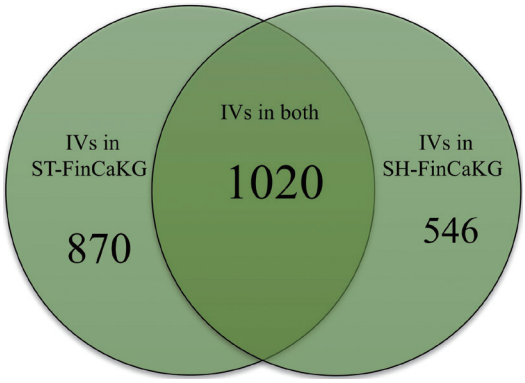
出所）Chen et al.(2024)中の図を許諾を得て改変。

- 変数 Z が、グラフの端のノード（edge node）である場合（＋ 1 点）
- Z と A の間の重み $w_{z,a}$ が 5.0 以上である場合（＋ 1 点）
- A と B の間の重み $w_{a,b}$ が 5.0 以上である場合（＋ 1 点）

図表 4 は、FinCaKG およびそのサブグラフにおける「操作変数―処置―結果変数」型の因果関係について、低品質、中品質、高品質に分類した件数を示している。SH-FinCaKG は、95 件という多くの高品質の操作変数と、それに対応する説明可能な因果関係を一意に識別している。しかし、ST-FinCaKG では、高品質操作変数を用いると、実際に識別可能な因果関係は 21 件にとどまる。一方で、ST-FinCaKG は、345 個のエッジノードが含まれており、より多くの専門知識を含んでいるが、高品質の因果関係は少ないため、その多くの因果関係は交絡している可能性があると言える。SH-FinCaKG は、ST-FinCaKG と比較して、より多くの高品質な因果連鎖を抽出できているため、「株主価値の最大化」という視点に対して、より明確で説得力のある説明を提供していると言える。この違いが生まれた原因として、ST-FinCaKG の方が、より多くのエッジノード（345 個）を含んでいるにもかかわらず、SH-FinCaKG と比べて変数間の結び付きが弱い（重みが小さい）ことが考えられる。これにより、SH-FinCaKG における専門知識は、より明確に「視点（standpoint）」を表している可能性があると言える。

また、異なるサブグラフから識別された操作変数のパターンの比較も行った。図表 5 に示すように、ST-FinCaKG に固有の操作変数は 870 個、SH-FinCaKG に固有の操作変数は 546 個、両方に共通して含まれる操作変数は 1020 個であった。この結果は、FinCaKG における専門知識の大部分が、この分野に共通するものであることを示唆している。ここで重要な問いが生ずる。特定のサブグラフでのみ観測される因果関係は、専門家の視点バイアスに起因するのか、それとも特定の企業タイプに固有の論理的パターンを反映しているのか。この問題については、次節において実際の経済データを用いてさらに検討する。

図表5　株主指向とステークホルダー指向の重なりを示すベン図



図表6　視点に基づく因果関係選択の実証分析結果

連鎖	ノード選択	二段階最小二乗法（ステージ）	二段階最小二乗法回帰係数	t 値	統計量（Anderson/CD Wald F）
ST-FinCaKG	Z：経済的エクスポージャー A：EBITDA B：コーポレートガバナンス	第 1 段階：Z → A 第 2 段階：A → B	-0.645*** 26.100***	-4.92 3.84	24.96***/12.49
SH-FinCaKG	Z：原油価格 A：営業利益 B：短期有価証券	第 1 段階：Z → A 第 2 段階：A → B	-1.102*** 0.340***	-3.06 2.55	11.43***/11.41

注）*** p<0.01, ** p<0.05, * p<0.1

出所）Chen et al.（2024）中の図を許諾を得て改変。

4.3 実験と結果

ここで、特定のサブグラフにのみ存在する高品質な操作変数を識別し、二段階最小二乗法回帰を実行することで、「処置―結果変数」の方向性効果および統計的有意性を解釈し、最終的に専門知識の一般化を図った。そして、因果関係の妥当性を検証するために、二段階最小二乗法回帰の統計的結果を算出し、因果推論の正当性を評価した。

全体の中で、高品質な操作変数に着目し、ST-FinCaKG における因果関係（経済的エクスポージャー → EBITDA → コーポレートガバナンス）、および SHFinCaKG における因果関係（原油価格 → 営業利益 → 短期有価証券）の 2 つの視点に基づく因果連鎖を検証した。図表 6 は、産業固定効果および年固定効果を考慮した 二段階最小二乗法回帰の結果を示している。ここでは、「経済的エクスポージャー」および「原油価格」を、それぞれ「為替エクスポージャー」および「原油価格エクスポージャー」で代替し、これらは産業ごとの為替および原油価格変動に対する感応度に基づいて算出した。また、「コーポレートガバナンス」の代理変数としては、社外取締役の比率を用いた。

為替エクスポージャーを用いた分析により、EBITDA がコーポレートガバナンスに与える正の効果が明らかとなり、その回帰係数は 26.1（$p<0.01$）であった。通常は、コーポレートガバナンスが良いと EBITDA を高める効果があると考えられるが、ここでは逆の関係となってる。今回は、ステークフォルダー指向の企業に関して見られた因果関係であり、こうした会社は EBITDA が高い時に、ガバナンスを整備するものであり、EBITDA を高めるために、

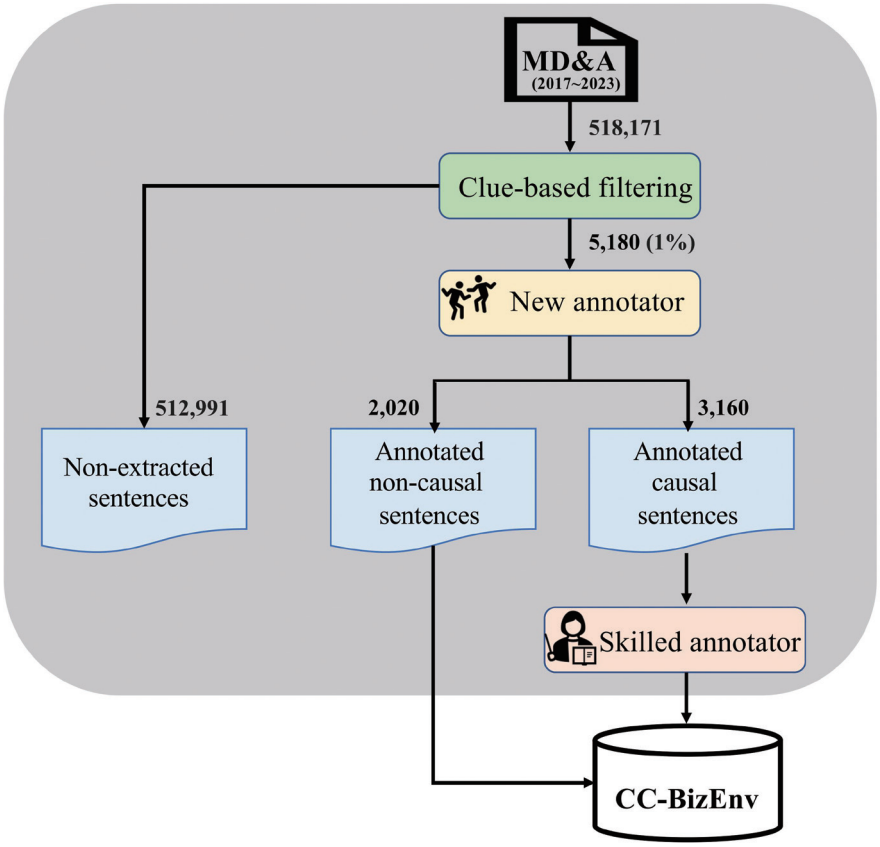
ガバナンスを強化するという発想が薄いことが示唆される。その一方で、株主指向の企業には、そうした因果関係は確認されていない。また、原油価格エクスポージャーに基づく分析では、営業利益が短期有価証券に与える正の効果が特定され、回帰係数は 0.34（$p<0.01$）であった。Anderson Canon. Statistics は有意（$p<0.01$）であり、また Craig Donald Wald F も 10 を超えていることから、操作変数 Z が関連性の仮定（$Cov(Z,A) \neq 0$ ）、および外生性の仮定（$Cov(Z, \varepsilon_2) = 0$）を満たしていることが示唆される。株主指向の企業では、環境の変化に対する利益の影響が保有する現金の決定係数になっており、現金保有が外部環境の変化に対して企業価値を守るための安全弁になっていると解釈できる。一方、ステークフォルダー指向の企業は、もしものために現金を蓄積する傾向があり、このような因果関係が見られない。これらの結果は、特定の因果関係が、全体においても統計的に有意であることを示しており、サブグラフにおいてのみ観測された因果関係が、現実世界における視点ベースのバイアスを反映している可能性が高いことを意味していると言える。

5．認知バイアスの特定

5.1 自己奉仕的帰属バイアス

経営者の視点の違いは、情報の解釈や対応の仕方に影響を与える認知バイアスに起因している可能性がある。経営者の認知バイアスは、行動ファイナンスの実証研究において注目されてきた。特に、自信過剰な経営者は、自身の意思決定の正確性を過大評価し、確率的事象における不確実性を過小評価する傾向がある。この理論に基づき、実験経済学や社会心理学では、経営者の認知バイアスや過剰な自信が、企業の政策決定、資本構成、投資行動、株価予測などの財務的意思決定に与える影響について、定義・実証分析が行われてきた。先行研究では、経営者の過信の測定は、主に心理実験や、CEO によるストックオプション行使行動などに関連する行動指標を用いた間接的手法によって測定されている（Malmendier and Tate, 2008）。しかし、日本では、一般に株価連動型のストックオプションが行使されることが少なく、この測定手法を日本企業に適用することは困難である。さらに、実証研究では、しばしば事例選択バイアスや事例サイズの不足といった課題がある。自信過剰な経営者は、良好な業績を自身の能力に起因させ、不振の結果を外部要因のせいにする傾向となる自己奉仕的帰属バイアス（Self-serving Attribution Bias）があることが知られている。本稿では、陳他（2024）の研究に基づき、データ・セット CC-BizEnv を用いてモデルの事前学習を行い、その後、財務文書によりファインチューニングを行うことにより、上場企業の経営者が毎年開示するテキストから、認知バイアスを定量化できる可能性があることを示す。

図表7　データセットの構築方法



出所）陳他（2024）

5.2 因果関係の抽出と分類

因果関係の抽出を行うために、データ・セット CC-BizEnv を構築した。CC-BizEnv データ・セットの構築（図表 7 参照）に当たり、経済の知識を有する 3 名のアノテータを採用した。そのうち 2 名は、経験のない新規アノテータであり、残りの 1 名は、3 万件以上のアノテーション経験を持つ熟練アノテータである。データ・セットの構築には、企業の年次報告書に含まれる MD&A（経営者による財務状況および経営成績の分析）セクションを用い、300 トークン以上が含まれる 518,171 個の文を抽出した。その中から、因果関係を示唆する文をさらに抽出し、その後、全体の 1% に相当する 5,180 文をランダムに選択した。これらをあらかじめ定義した 4 種類の因果パターン（例：内部要因、外部要因）に分類した。新規アノテータには以下の 2 つのタスクを割り当てた：

1. 文中に因果関係が含まれているかの判定
2. 「手がかり（clue）」、「内部要因」、「外部要因」、「肯定的結果」、「否定的結果」に該当するスパンに対して、タグ付けを行う作業

その後、熟練アノテータが全件をレビューし、必要に応じて修正を加えた。このプロセスにより、最終的に 3,160 文の高品質なラベル付きデータが得ら

れ、CC-BizEnv データセットが構築された。

経営者の因果的推論を抽出するために、LSTM、Bi-LSTM-CRF、ELECTRA、BERT を含む複数のモデルを比較した。評価指標は、スパンとの部分一致に基づき、適合率、再現率、F1 スコアを用いた。アノテータの専門性の違いを調べたところ、熟練アノテータによってアノテーションを精緻化した場合、学習結果が有意に改善されていた。同一の高品質データセットを用いた場合、BERT は他のモデルを上回る性能を示し、F1 スコアで 87.4% を達成した。これは、ELECTRA(53.4%)、LSTM(63.5%)、Bi-LSTM-CRF(67.8%)と比較すると顕著な差がある。他のモデルは、「肯定的結果」および「否定的結果」の識別において、一定の効果しか示さなかった。最後に、より高度な大規模言語モデル (LLM) を用いたゼロショット評価も実施した。CC-BizEnv データセットでファインチューニングされた BERT モデルは、ゼロショット設定下の Claude-3.5-Sonnet（67.9%)、GPT-3.5-Turbo（27.3%)、GPT-4o-2024-05-13（31.7%）を大きく上回り、全体で最高の性能（87.4%）を示した。

5.3 自信過剰の測定と経済的影響

ここでは、経営者が肯定的な成果を内部要因（例：経営能力）に、否定的な成果を外部要因（例：COVID-19 など）に帰属させる文を抽出する手法を説明する。

SAB 指標（Self-serving Attribution Bias Index）は、年次報告書内のすべての因果的帰属文のうち、自己奉仕的な帰属文の割合を測定するものであり、経営者の帰属バイアスを定量的に評価する指標である。これまでの実証研究（Chen et al., 2025）によれば、SAB 指標が高い企業は、誤った予測や過度の楽観主義といった既存の経営者過信指標と有意に相関していることが示されている。一方で、SAB 指標は、CEO の年齢、在任期間、および理工系学歴とは負の相関を示している。経済的観点から見ると、SAB 指標が高い企業は、配当支払いの減少、自社株買いの増加、高い財務レバレッジ、価値を毀損する M&A の実施頻度の増加といった、株主価値に対して負の影響を及ぼす行動をとる傾向がある。以上の知見は、SAB 指標が経営者の認知バイアスを評価し、それが企業の財務政策やパフォーマンスに与える潜在的影響を把握する上で、有用なツールとなることを示唆している。

6. おわりに

本稿では、テキストから因果関係を抽出、利用することで、経済的因果関係研究における長年の課題である「可視化問題」に取り組めることを実証した。テキストから因果知識グラフを構築するフレームワーク FinCaKG を活用することで、経済専門家の繊細な視点を取り込んだ因果知識グラフを構築することができる。そして、提案したデータマイニングアルゴリズム ETE-FinCa は、高品質な因果連鎖を効率的に抽出できることを示しており、実際の市場

データを用いた実証分析において、これらの因果連鎖の統計的有意性と頑健性が、厳密な計量経済分析によって確認された。これらの成果は、視点に基づく因果知識グラフが、因果推論の精緻化、より正確な経済モデリングの支援、政策分析の高度化に貢献し得ることを示唆している。すなわち、経済現象を構造的かつ解釈可能、かつデータ駆動型に表現する有力な手段として、本アプローチの有用性が示されたと言える。また、自信過剰といった経営者の視点の違いを、効果的に捉える CC-BizEnv データセットと合わせると、主観的な視点が経済的ナラティブや、意思決定にどのような影響を及ぼすのかを、包括的に理解するための道を切り拓くことができるであろう。因果テキスト分析と定量的モデリングを統合することで、企業コミュニケーションにおける行動的側面の検証に新たな視点を提供することができ、このような手法を利用することで、経済領域における因果ダイナミクスの解釈可能性を一層深化させることができる。

今後の展望として、因果知識グラフは、ニュース記事、ソーシャルメディア、専門家による解説といった追加的なデータソースを取り込むことで、より多様かつきめ細かな視点の抽出へと発展させることができる。さらに、高度な生成型大規模言語モデル技術や因果発見アルゴリズムとの統合により、因果関係抽出の精度およびスケーラビリティの向上が期待される。テキスト分析と因果推論を組みあわせることで、より透明性が高く、説明可能な AI システムの基盤技術として、ステークホルダーが的確な意思決定を行うために広く利用されていくことが期待される。

参考文献

Ali, W., Zuo, W., Ali, R., Zuo, X., and Rahman, G. (2021). Causality mining in natural languages using machine and deep learning techniques: A survey. *Applied Sciences,* 11:10064.

Bennett, B. (2011). Possible worlds and possible meanings: A semantics for the interpretation of vague languages. In *Proceedings of AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, pp. 23-29.

Chang, D.-S. and Choi, K.-S. (2004). Causal relation extraction using cue phrase and lexical pair probabilities. In *Proceedings of International Conference on Natural Language Processing,* pp. 61-70. Springer.

Chen, Y., Kimura, Y., and Inoue, K. (2025). Measuring self-serving attribution bias using text mining. *Available at SSRN 5130571*.

Chen, Y., Xu, Z., Inoue, K., and Ichise, R. (2024). Causal inference in finance: An expertise-driven model for instrument variables identification and interpretation. In *Proceedings of the 23rd International Conference on Machine Learning and Applications,* pp. 1089-1094.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, É., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8440-8451.

De Cao, N., Izacard, G., Riedel, S., and Petroni, F. (2021). Autoregressive entity

retrieval. *In Proceedings of the 9th International Conference on Learning Representations*.

De Cao, N., Wu, L., Popat, K., Artetxe, M., Goyal, N., Plekhanov, M., Zettle-moyer, L., Cancedda, N., Riedel, S., and Petroni, F. (2022). Multilingual autoregressive entity linking. *Transactions of the Association for Computational Linguistics,* 10:274-290.

Devlin, J., Chang, M.- W., Lee, K., and Toutanova, K. (2019). BERT: Pretraining of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* pp. 4171-4186.

Elman, J. L. (1990). Finding structure in time. *Cognitive science,* 14(2):179-211.

Frattini, J., Fischbach, J., Mendez, D., Unterkalmsteiner, M., Vogelsang, A., and Wnuk, K. (2023). Causality in requirements artifacts: prevalence, detection, and impact. *Requirements Engineering*, 28:49-74.

Girju, R. (2003). Automatic detection of causal relations for question answering. In *Proceedings of the ACL workshop on Multilingual summarization and question answering,* pp. 76-83.

Heindorf, S., Scholten, Y., Wachsmuth, H., Ngonga Ngomo, A.- C., and Potthast, M. (2020). Causenet: Towards a causality graph extracted from the web. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management,* pp. 3023-3030.

Huguet Cabot, P.- L. and Navigli, R. (2021). REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021,* pp. 2370-2381.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* pp. 260-270.

Lane, N. (2020). The new empirics of industrial policy. *Journal of Industry, Competition and Trade,* 20(2):209-234.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). *Gradient-based learning applied to document recognition. Proceedings of the IEEE,* 86(11):2278-2324.

Li, Z., Ding, X., Liu, T., Hu, J. E., and Van Durme, B. (2020). Guided generation of cause and effect. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence,* pp. 3629-3636.

Lu, Y., Liu, Q., Dai, D., Xiao, X., Lin, H., Han, X., Sun, L., and Wu, H. (2022). Unified structure generation for universal information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics,* pp. 5755-5772.

Luo, Z., Sha, Y., Zhu, K., Hwang, S.- W., and Wang, Z. (2016). Commonsense causal reasoning between short texts. In *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning,* pp. 421-430.

Malmendier, U. and Tate, G. (2008). Who makes acquisitions? ceo overconfidence and the market's reaction. *Journal of financial Economics,* 89(1):20-43.

O'Brien, S. P. (2010). Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review,* 12(1):87-104.

Orlando, R., Huguet Cabot, P.- L., Barba, E., and Navigli, R. (2024). ReLiK: Retrieve and LinK, fast and accurate entity linking and relation extraction on an academic

budget. In *Findings of the Association for Computational Linguistics ACL 2024,* pp. 14114-14132.

Pearl, J. (2010). On a class of bias-amplifying variables that endanger effect estimates. *In Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence,* pp. 417-424.

Radinsky, K., Davidovich, S., and Markovitch, S. (2012). Learning causality for news events prediction. In *Proceedings of the 21st International Conference on World Wide Web,* pp. 909-918.

Rink, B., Bejan, C. A., and Harabagiu, S. (2010). Learning textual graph patterns to detect causal event relations. In *Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference*.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41-55.

Rossiello, G., Chowdhury, M. F. M., Mihindukulasooriya, N., Cornec, O., and Gliozzo, A. M. (2023). Knowgl: knowledge generation and linking from text. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, pp. 16476-16478.

Sorgente, A., Vettigli, G., and Mele, F. (2013). Automatic extraction of causeeffect relations in natural language text. In *Proceedings of the 7th International Workshop on Information Filtering and Retrieval, co-located with the 13th International Conference of the Italian Association for Artificial Intelligence,* volume 1109, pp. 37-48.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. 2nd edition.

Straková, J., Straka, M., and Hajic, J. (2019). Neural architectures for nested NER through linearization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* pp. 5326-5331.

Tang, W., Xu, B., Zhao, Y., Mao, Z., Liu, Y., Liao, Y., and Xie, H. (2022). UniRel: Unified representation and interaction for joint relational triple extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* pp. 7087-7099.

Vrandecic, D. and Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM,* 57(10):78-85.

Wang, C., Liu, X., Chen, Z., Hong, H., Tang, J., and Song, D. (2022). DeepStruct: Pretraining of language models for structure prediction. In *Findings of the Association for Computational Linguistics: ACL 2022,* pp. 803-823.

Xu, Z., Nararatwong, R., Kertkeidkachorn, N., and Ichise, R. (2022). iLab at FinCausal 2022: Enhancing causality detection with an external cause-effect knowledge graph. In *Proceedings of the 4th Financial Narrative Processing Workshop,* pp. 124-127.

Xu, Z., Takamura, H., and Ichise, R. (2025). Fincakg: A framework to construct financial causality knowledge graph from text. *International Journal of Semantic Computing*, 19:02, pp. 321-340.

Ye, H., Zhang, N., Chen, H., and Chen, H. (2022). Generative knowledge graph construction: A review. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* pp. 1-17.

Yin, W., Kann, K., Yu, M., and Schütze, H. (2017). Comparative study of cnn and rnn for natural language processing. *arXiv preprint ar Xiv:1702.01923*.

陳穎・井上光太郎・市瀬龍太郎.（2024）. 有価証券報告書からの経営者による経営環境と企業業績に関する因果認知の抽出. 人工知能学会第二種研究会資料, Fin-033:163-168.

大規模言語モデルを活用した統計的因果探索の新展開

― 統計的因果プロンプティングによる専門知識とデータの融合 ―

三内 顕義｜京都大学理学部特定准教授
国立情報学研究所 大規模言語モデル研究開発センター 客員准教授
高山 正行｜文部科学省科学技術・学術政策研究所 第 1 調査研究グループ 研究官
滋賀大学 データサイエンス・AI イノベーション研究推進センター 客員研究員
清水 昌平｜大阪大学 産業科学研究所 教授
滋賀大学 データサイエンス学系 教授

三内 顕義

京都大学理学部特定准教授。東京大学博士（数理科学）。京都大学数理解析研究所特定助教、理化学研究所革新知能統合研究センター研究員、同上級研究員などを経て2023 年より現職。国立情報学研究所大規模言語モデル研究開発センター客員准教授、滋賀大学データサイエンス学部特定准教授、科学技術・学術政策研究所客員研究官などを兼任。
2025 年 6 月に、数学分野での AI エージェントで世界最高性能（SOTA）を達成。

要旨

本稿は、我々が提案した「統計的因果プロンプティング（SCP）」を解説する。データから因果関係を発見する統計的因果探索（SCD）は多分野で活用されているが、専門知識なしでは正確性に限界があった。この課題に対し、大規模言語モデル（LLM）と SCD を相互参照的に統合する新手法が開発された。SCP は、初期 SCD 実行、LLM による知識生成と確率的評価、事前知識を用いた SCD 再実行の 4 段階で構成される。実験では、ベンチマークデータで構造的ハミング距離が縮まり、精度も向上することが確認された。LLM の学習データに含まれない健康診断データでも有効性が確認された。本手法は、統計的証拠と専門知識を融合させる点で画期的だが、計算量が変数の数の 2 乗に比例する制約がある。また、医療・金融等の重要領域で応用する上では、専門家による検証が不可欠である。今後、LLM 技術の発展とともに、より大規模で複雑な因果関係の解明への貢献が期待される。これらの結果は Takayama, et al.(2025) に基づいたものである。

1．はじめに

1.1 因果推論の重要性と現代的課題

私たちは日々、様々な現象の原因と結果を理解しようとしている。なぜ特定の治療法が病気に効くのか、どのような経済政策が景気回復につながるのか、環境汚染がどのように健康に影響するのか。これらの問いに答えるには、単なる相関関係ではなく、真の因果関係を明らかにする必要がある。

近年、ビッグデータの時代を迎え、大量のデータから因果関係を発見する統計的因果探索（Statistical Causal Discovery: SCD）への期待が高まっている。SCD は、観察データから変数間の因果構造を推定する手法であり、ラン

ダム化比較試験が困難な状況でも因果関係の解明を可能にする。

しかし、SCD には重要な課題がある。純粋にデータ駆動的なアプローチでは、統計的には妥当でも現実世界の知識と矛盾する因果関係を導出することがある。例えば、アイスクリームの売上と溺死事故の間に強い相関があっても、両者の間に直接的な因果関係があるわけではない。このような誤った推論を避けるには、分野固有の専門知識、すなわちドメイン知識の活用が不可欠である。

従来、ドメイン知識の組み込みは専門家による手作業に依存していた。しかし、変数が増えるにつれて、すべての変数間の因果関係について専門家が判断することは現実的ではなくなる。さらに、専門家の知識にも偏りや限界があり、最新の科学的知見を常に反映させることは困難である。

1.2 大規模言語モデルの登場と新たな可能性

このような状況に大きな変革をもたらしたのが、大規模言語モデル（Large Language Model: LLM）の登場である。GPT-4 をはじめとする最新の LLM は、膨大な文献から学習した知識を活用し、専門的な質問にも高い精度で回答できるようになった。

LLM の特筆すべき点は、単に情報を記憶しているだけでなく、文脈に応じて推論し、知識を統合できることである。この能力は、因果推論においても大きな可能性を秘めている。例えば、医学的な因果関係について問われれば、解剖学的構造、生理学的メカニズム、疫学的証拠などを総合的に考慮した回答を生成できる。

しかし、LLM を因果推論に活用する試みはまだ始まったばかりである。LLM の出力には不確実性があり、時として事実と異なる情報を生成する「ハルシネーション」と呼ばれる現象も知られている。また、LLM が持つ知識と、特定のデータセットから得られる統計的証拠をどのように統合すべきかも明確ではなかった。

本稿で紹介する我々の研究は、まさにこの課題に正面から取り組んだものである。我々が提案した「統計的因果プロンプティング（Statistical Causal Prompting: SCP）」は、LLM の持つドメイン知識と SCD による統計的証拠を相互参照的に統合する画期的な手法である。この手法により、専門知識を考慮しつつ、データに基づいた信頼性の高い因果推論が可能になる。

以下では、この革新的なアプローチの詳細を解説し、その可能性と課題、そして今後の展望について論じていく。

2. 原論文 Takayama, et al.(2025)の概要

2.1 研究の背景と目的

我々の研究は、統計的因果探索が抱える根本的な課題に着目している。従来の SCD 手法は、データの統計的傾向のみに基づいて因果関係を推定するため、しばしば現実と乖離した結果を生む。例えば、病院の数と死亡率の間に正の相

関があったとしても、「病院が増えると死亡率が上がる」という因果関係は明らかに誤りである。このような問題を解決するには、「病気の人が多い地域に病院が建設される」という背景知識が必要となる。

研究の目的は、LLM が持つ広範な知識を活用して、このようなドメイン知識を体系的に SCD に組み込む方法を開発することである。特に重要なのは、LLM の知識とデータから得られる統計的証拠を単に組み合わせるのではなく、両者が相互に情報を提供し合う枠組みを構築した点である。

2.2 提案手法の概要

SCP の核心は、4 段階の反復的プロセスにある。

第 1 段階では、通常の SCD アルゴリズムを用いて、事前知識なしに因果グラフを推定する。この初期結果は、データに内在する統計的傾向を反映している。

第 2 段階では、LLM に対して変数間の因果関係について問いかける。ここで重要なのは、単に「A は B の原因か」と尋ねるのではなく、第 1 段階で得られた統計的情報も提示することである。例えば、「統計解析では A から B への因果関係が示唆されていますが、専門知識の観点からこれは妥当でしょうか」というかたちで問いかける。

第 3 段階では、LLM が生成した詳細な説明を基に、各因果関係の存在確率を定量的に評価する。因果関係の存在を LLM に「Yes」か「No」で回答させるかたちで問い、その出力確率を測定することで、因果関係の確信度を数値化する。

第 4 段階では、LLM から得られた確率的評価を事前知識行列に変換し、これを制約条件として用いて、再度 SCD を実行する。この際、統計的証拠と専門知識の両方を考慮した、より信頼性の高い因果グラフが得られる。

3. 主要な理論・手法の解説

3.1 統計的因果探索の基礎

統計的因果プロンプティング（SCP）を理解するには、まず基盤となる統計的因果探索（SCD）の主要手法を理解する必要がある。我々は、代表的な三つのアルゴリズムを採用している。

3.1.1 PCアルゴリズム

PC アルゴリズムは、ピーター・スパーティスとクラーク・グリマーによって開発された、制約ベースの因果探索手法である。その名前は開発者の頭文字に由来する。

このアルゴリズムの基本原理は、条件付き独立性の検定である。二つの変数が与えられたとき、他の変数を条件として独立かどうかを統計的に検定する。例えば、気温とアイスクリームの売上は相関するが、季節を条件とすると独立になる可能性がある。この場合、気温と売上の間に直接的な因果関係はなく、

両者は季節という共通原因を持つと推定できる。

PC アルゴリズムは、まず完全グラフから始めて、条件付き独立性が成立するペアのエッジを削除していく。その後、向きのない辺で構成されるグラフ構造（スケルトン）が得られたら、有向非巡回グラフ（DAG）の仮定等に基づく向きづけルール（オリエンテーションルール）を適用し、可能な範囲で因果の向きを特定していく。このプロセスにより、最終的に、データと統計的に整合する因果グラフが得られる。ただし、観察データのみからは、オリエンテーションルールで因果の方向を完全に特定できない場合もあり、その場合は無向辺を残した、部分的有向非巡回グラフ（CPDAG）として表現される。

3.1.2 Exact Searchアルゴリズム

Exact Search は、スコアベースの手法であり、可能なすべての DAG の中から、データに最も適合するものを探索する。我々は、A* アルゴリズムを用いた、ヒューリスティックなアプローチを導入している比較的効率的な実装を採用している。

この手法では、候補となる各因果グラフに対してベイズ情報量基準（BIC）などのスコアを計算し、そのスコアが最適となる因果グラフを選択する。BIC は、モデルの適合度とモデルの複雑さのバランスを考慮した指標であり、過学習を防ぐ効果がある。理論的には、サンプルサイズが十分大きければ、真の因果グラフを一致推定できることが知られている。

スコアベースの手法の中でも、Exact Search の利点は、グローバルに最適な解を保証できることである。一方、変数の数が増えると計算量が指数的に増加するため、大規模なネットワークには適用が困難という課題もある。

3.1.3 DirectLiNGAMアルゴリズム

LiNGAM（Linear Non-Gaussian Acyclic Model）は、清水昌平氏らによって開発された、日本発の革新的な因果探索手法である (Shimizu et al.(2006))。DirectLiNGAM はその効率的な実装版である。

従来の統計的手法の多くは、モデルの誤差項が正規分布に従うことを仮定していたが、LiNGAM は非ガウス性を仮定し、非ガウス分布の場合には、観察データのみから因果の方向を一意に特定できるという画期的な性質を持つ。

DirectLiNGAM は、変数の因果的順序を逐次的に同定していく。各ステップで、他の変数から最も独立な変数を選び、それを最上流の変数として確定する。この過程を繰り返し、回帰分析を行うことで、完全な因果グラフを構築し、因果係数の形で因果効果を推定することができる。

3.2 統計的因果プロンプティング（SCP）の詳細

SCP の革新性は、LLM へのプロンプト設計にある。単に LLM に因果関係を尋ねるのではなく、統計的証拠を含めた文脈を提供することで、より信頼性の高い推論を引き出す。

3.2.1 Generated Knowledge Promptingの活用

Generated Knowledge Prompting は、LLM に直接答えを求めるのではなく、まず関連する知識を生成させてから、生成された知識も活用して判断を行わせる手法である。我々は、この手法を因果推論に適用している。

具体的には、変数 A と B の因果関係について、まず「A が B に影響を与えるメカニズムについて、専門知識に基づいて説明してください」と問いかける。LLM は、関連する理論、既知のメカニズム、実証研究の知見などを詳細に説明する。この過程で、LLM は自身の知識をより着実に活用できるようになり、より深い推論が可能になる。

3.2.2 Zero-Shot Chain-of-Thoughtによる推論誘導

Chain-of-Thought（CoT）は、LLM に段階的な推論過程を明示させることで、複雑な問題の解決能力を向上させる手法である。Zero-Shot CoT は、特別な例示なしにこれを実現する。

SCP では、「統計解析の結果も踏まえて、この因果関係の妥当性を、専門家の見地から評価してください」という形で、LLM に推論過程の明示を促す。LLM は、統計的証拠、理論的整合性、既知の事実との一致性などを順に検討し、最終的な判断に至る。

3.2.3 プロンプトパターンの設計思想

我々は、五つのプロンプトパターンを設計し、それぞれの効果を検証している。

パターン 0（ベースライン）では、統計的情報を含めず、純粋に LLM が有するドメイン知識のみで判断させる。これは、LLM の基本的な推論能力を測る基準となる。

パターン 1 では、初期 SCD で得られた因果グラフの構造を提示する。「統計解析では、A から B への因果関係が示唆されています」という形で、因果の方向性の情報を含める。

パターン 2 では、ブートストラップ法で計算した因果関係の出現確率を追加する。「1000 回のリサンプリングで、この因果関係が 75%の確率で検出されました」といった定量的情報を提供する。

また、構造方程式を仮定して因果係数も出力する DirectLiNGAM の場合のみではあるが、パターン 3 では推定された因果係数の値、パターン 4 ではパターン 2・3 の両方の値を、というように、さらに詳細な統計量も含める。これにより、因果関係の強さも LLM に提供される。

3.2.4 トークン生成確率の測定と信頼度評価

LLM の出力の信頼度を定量化するため、トークン生成確率を活用する。変数間に因果性があるか否かという質問に対し LLM が「Yes」または「No」と回答する際の対数確率を取得し、これを因果関係の存在確率として解釈する。

ただし、LLM の出力確率にも揺らぎがあることが知られており、同じプロンプトでも、実行のたびに異なる確率が出力される。この問題に対処するため、我々は同じ質問を 5 回繰り返し、平均確率を採用している。この手法に

より、より再現性の高い評価が可能になる。

3.3 LLMとSCDの統合メカニズム

3.3.1 事前知識行列の生成プロセス

LLM から得られた確率的評価を、SCD アルゴリズムが利用できる形式に変換する必要がある。我々は、確率に閾値を設定し、それを超える因果関係を事前知識として採用する方法を提案している。

具体的には、n 個の変数に対しての $n \times n$ 行列を作成する。行列の (i, j) 要素は、変数から変数への因果関係の有無を表す。LLM の出力確率が閾値（例えば 0.05）を下回る場合は、その辺は絶対に現れないものとして 0 とする。

3.3.2 アルゴリズムの計算複雑度とスケーラビリティ

SCP の計算量は、主に LLM への問い合わせ回数によって決まる。n 個の変数に対して、$n(n-1)$ 個の因果関係を評価する必要があるため、計算量は $O(n^2)$ となる。さらに、各関係について 5 回の測定を行うため、実際の計算時間は $5n(n-1)$ 回の LLM 呼び出し回数に比例する。

この二次の計算複雑度は、大規模ネットワークへの適用において課題となる。例えば、100 変数のネットワークでは約 50,000 回の LLM 呼び出しが必要となり、API 制限やコストを考慮すると現実的ではない。したがって、本手法が適用可能な問題は、変数の数（モデルの規模、考察する範囲）に制約される。

3.4 手法の技術的工夫と革新性

3.4.1 相互参照的な性能向上の仕組み

SCP の最も革新的な点は、LLM と SCD が相互に情報を提供し合うことで、両者の性能の向上が期待される点である。SCD の結果を LLM に提示することで、LLM はデータの統計的傾向を考慮した推論ができる。一方、LLM の知識を SCD に組み込むことで、現実と整合的な因果グラフが得られる。

この相互参照は、人間の専門家が統計解析結果を解釈する過程に似ている。専門家は、統計的に有意な関係を見たとき、それが既知の理論やメカニズムと整合するかを検討する。同時に、統計的証拠が強ければ、既存の理論を修正することもある。SCP は、このような高度な推論プロセスを自動化している。

3.4.2 統計的証拠とドメイン知識の融合方法

従来の研究では、統計的手法はドメイン知識に基づくモデル構築の議論とは独立に発展してきた。SCP は、両者を LLM の出力の確率を介することで統一的に扱うことに成功している。

この融合により、データが少ない場合はドメイン知識に基づいて補完を試みる一方、データが豊富な場合はその統計的根拠に基づいてドメイン知識の確からしさについての定量的な議論にも踏み込むという適応的な推論も可能にな

る。これは、ベイズ統計における事前分布と尤度の関係に類似しており、理論的にも妥当なアプローチである。

4．実証結果とその含意

4.1 ベンチマークデータでの検証

我々は、手法の有効性を検証するため、因果探索の研究で広く使用されるベンチマークデータセットを用いた実験を行っている。これらのデータセットは、因果関係についてある程度、専門家の共通見解が得られているか、あるいは一般常識になっているような変数の組み合わせで、現実世界で実際に取得されたデータであり、因果探索アルゴリズムの性能評価にもよく使われるものである。ベンチマークになる因果関係（ground truth）があるため、因果探索した結果がそのベンチマークにどれだけ近づいているかをもとに、手法の性能を客観的に評価できる。本稿ではそのうち、Auto MPG データ、DWD 気候データでの検証事例を紹介する。

4.1.1 Auto MPGデータ：小規模ネットワークでの安定性

Auto MPG データセットからは、自動車の燃費に関する、重量（weight）、馬力（horsepower）、加速性能（acceleration）、総排気量（displacement）、燃費（miles per gallon）の計五つの連続変数を抽出した。このデータセットは変数の数が少ないため、SCP の基本的な動作を検証するのに適している。

実験結果では、すべてのプロンプトパターンで安定した性能向上が観察された。構造的ハミング距離（SHD）は 2 つの有向グラフ間の構造的な違いを測る指標である。真の因果性を示す有向グラフと推計結果を有向グラフ化したものを比較することで、推計精度を確認することができる。同指標では、事前知識なしのベースラインと比較して 50%以上改善した。興味深いことに、異なるプロンプトパターン間での性能差は小さく、これは変数の数が少ない場合の SCP の頑健性を示している。また、事前知識なしの SCD では、燃費→重量、燃費→馬力といった、不自然な辺を出力する傾向があったが、SCP により、「重い車は燃費が悪い」「馬力が大きい車は燃費が悪い」といった、物理的にも自然な因果関係に、正しく SCD の結果を誘導できることが確認された。

4.1.2 DWD気候データ：中規模での効果検証

DWD データセットは、ドイツ気象局の気候データから作成された 6 変数のデータセットであり、緯度（latitude）、経度（longitude）、高度（altitude）、気温（temperature）、日照時間（sunshine）、降水量（precipitation）などの気象に関連する変数から構成される。

このデータセットでは、SCP の効果がより顕著に現れた。特に DirectLiNGAM を用いた場合、特にパターン 2（ブートストラップ確率を含む）で最大 60%の SHD 改善が見られた。これは、統計的情報が LLM の推論を大きく向上させることを示している。

事前知識なしの SCD では、気温→高度など、地理的に定まっている変数

が他の変数から影響を受けるような不自然な辺が出力されることがある一方、SCP を経ることで、事前知識ありの SCD では緯度・経度・高度が他の変数から影響を受けるといった有向グラフの辺が現れることはなくなり、むしろ「高度が高いと気温が下がる」といった、気象学的にも自然な辺が、SCD に適切に出力されるようになった。

4.2 未公開健康診断データでの実証

ベンチマークデータでの成功は重要だが、公開された著名なベンチマークであり実証研究で数多く利用されているため、LLM が単に学習データを記憶している可能性を完全には排除できない。そこで我々は、GPT-4 の学習データに含まれていない日本の健康診断データを用いた実験を行った。

4.2.1 データの特性と実験設計

使用されたデータは、年齢、HbA1c（糖化ヘモグロビン）、肥満度、中性脂肪値、腹囲、収縮期血圧、拡張期血圧の７変数で構成している。これらは糖尿病リスクに関連する重要な指標である。データは特定の者しかアクセスできないように厳重に管理されており、当然、インターネットにも開示されていないため、LLM が事前に学習している可能性は極めて低い。

実験では、全データから無作為に抽出したサブセットのうち、統計的バイアスを含む結果を、意図的に検証に採用した。これにより、純粋な統計的手法のみでは正しい因果関係を発見しにくい、より現実的な条件での検証が可能となった。

4.2.2 LLMによる因果探索結果の改善の確認

まず、事前知識を組み込まない初期の SCD では、「HbA1c が年齢に影響する」といった、非常に不自然な結果が出力されていた。一方、SCP を進めると、「年齢が上がると HbA1c が上昇する」といった、医学的に確立された因果関係を正しく認識し、その推論結果を事前知識として DirectLiNGAM で再度因果探索した結果、年齢→ HbA1c 経路が正しく推定されるなど、医学的により妥当な因果グラフが得られた。

このように、LLM は未知のデータセットに対しても適切な推論を示した。また、この結果は、SCP が実際の医療データの因果分析においても、有用である可能性を示唆している。特に、観察研究では避けられない選択バイアスや交絡の問題によって、データセットに何らかの偏りが発生してしまう場合でも、LLM の知識が補正効果を持つことが確認された。

4.3 結果の統計的・実務的解釈

4.3.1 性能評価指標の改善とその意味

特にベンチマークデータを活用した実験全体を通じて、複数の評価指標で一貫した改善が見られた。例えば構造的ハミング距離（SHD）の改善は、推定された因果グラフが真の構造により近づいたことを意味する。精度（Precision）

の向上は、誤った因果関係の削減を、精度と再現率（Recall）の両者を総合評価する F1 スコアの改善は、全体的なバランスの向上を示している。

また、統計的妥当性の指標である CFI（比較適合度指標）や RMSEA（近似の二乗平方平均誤差）の極端な悪化はほとんど観測されず、これは SCP により得られた因果モデルがデータにより良く適合することを示している。さらに、BIC の減少は、モデルの複雑さを考慮してもなお、改善されたモデルが優れていることを意味する。

4.3.2 プロンプトパターンの効果分析

プロンプトパターンの分析からは、「適度な」統計情報の提供が重要であることが明らかになった。例えば、パターン 2（ブートストラップ確率のみ）は多くの場合で良好な結果を示したが、パターン 4（すべての統計情報を含む）では必ずしも最良ではなかった。

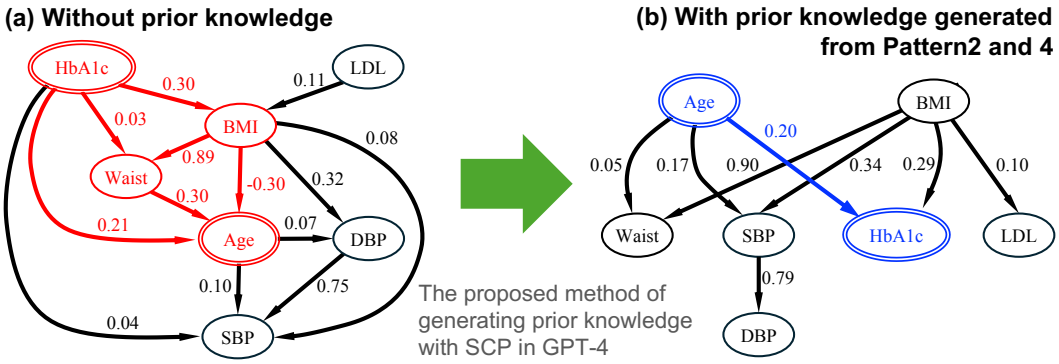
これは、情報過多による LLM の判断の揺らぎ、あるいは統計的詳細がドメイン知識の活用を妨げることもあるという可能性を示唆している。実務的には、使用する SCD アルゴリズムと問題の特性に応じて、適切なプロンプトパターンを選択することが重要である。

図表1　統計的因果プロンプティング（SCP）の概念図

出所）筆者作成

また、図表２からは、LLM が有する情報の活用により、ヘルスケアデータからの因果探索で現れた年齢（age）への矢印（論理的に存在しない）が適切に排除されていることがわかる。

図表2　ヘルスケアデータにおける因果グラフの比較



出所）Takayama et al.(2025)

５．政策的・実務的示唆

SCP の開発は、様々な分野での因果推論の実践に大きな影響を与える可能性がある。ここでは、主要な応用分野における活用方法と、実装時の課題・対策について論じる。

５.１ 医療・健康科学への応用

５.１.１ 疾患要因の探索への活用

医療分野では、疾患の発症メカニズムや危険因子の特定が重要な課題である。SCP は、電子カルテや健康診断データから、より信頼性の高い因果関係を発見するのに役立つ可能性がある。

例えば、生活習慣病の発症要因を分析する際、単純な統計解析では「運動不足→肥満→糖尿病」という直線的な関係しか捉えられないことが多い。しかし、従来の因果探索に加え、SCP を用いることで、現実的には同時にデータの取得の難しい遺伝的要因、環境要因、行動要因の複雑な相互作用も領域知識の観点から考慮した、より現実的な因果モデルを構築できる可能性もある。

特に近年の LLM は、追加学習を行わせなくとも、事前学習の段階で、その学習データのナレッジ・カットオフとなる時点までの医学文献の知識も学習しているため、新しく発見された疾患メカニズムや、稀な疾患についても推論に反映が可能になると期待される。これにより、データ数が限られる希少疾患の研究においても、有用な知見が得られる可能性がある。

５.１.２ 予防医学での因果関係理解

予防医学の観点からは、介入可能な要因の特定が重要である。従来の SCD に加え、SCP を活用することで、どの要因に介入すれば最も効果的に疾患を

予防できるかを、統計的観点と領域知識の両面から明らかにするのに役立つ。例えば、早期介入により薬剤使用を回避できる可能性のある群を特定することで、医療費の削減と患者の QOL 向上の両立が期待できる。

5.1.3 実装時の注意点

医療分野で SCP を実装する際は、SCD を応用した場合の結果解釈と同様、特に慎重な検証が必要である。LLM の推論結果は、ハルシネーションに由来する致命的な誤りがないことを確認するために、必ず医学専門家によるレビューを経るべきである。また、個人情報保護の観点から、取り扱いに配慮が必要なデータは、匿名化や、LLM 利用時に、誤って機微な内容がプロンプトに打ち込まれないようにするための予防的工夫も必要である。

さらに、医療分野特有の倫理的配慮も重要である。因果関係の誤認識が患者の治療方針に影響を与える可能性があるため、SCP の結果は意思決定支援ツールとして位置づけ、最終的な判断は医療従事者が行うべきである。

5.2 経済・金融分野での活用

5.2.1 市場メカニズムの解明

金融市場では、複雑な要因が相互に影響し合っている。従来の SCD に加え、SCP を用いることで、市場変動の背後にある因果メカニズムをより深く理解できる可能性がある。

例えば、為替レート、金利、株価指数の関係を分析する際、従来の計量経済学的手法では見逃されがちな間接的な因果経路を、SCD により発見できる可能性がある。加えて経済理論や過去の市場動向に関する知識を持っている LLM の活用により、統計的に検出された関係が経済学的に妥当かどうかを判断することも可能であり、真に新しい因果経路についても領域知識から議論することができる。

5.2.2 リスク要因の特定

リスク管理において、真の因果関係の理解は極めて重要である。相関関係に基づくリスク評価では、見かけ上の関係に惑わされる危険がある。SCD に加え LLM を活用する SCP は、真のリスク要因を領域知識の観点からも特定し、より効果的なリスク管理戦略の立案を支援する。

例えば、信用リスク評価では、顧客属性と債務不履行の関係を分析する際、社会経済的な背景知識を考慮することで、より公平で正確なモデルを構築できる。これは、人種や住所により与信の可否が左右されるといったアルゴリズミックバイアスの問題を軽減する上でも重要である。

5.2.3 政策効果の事前評価

経済政策の立案において、政策変更が経済に与える影響を事前に評価することは重要である。SCD に加え、LLM を活用した SCP を用いることで、過去のデータから政策変数と経済指標の因果関係を領域知識からも自然な形で推定

し、政策効果をより正確に予測できる。

5.3 実装における課題と対策

5.3.1 計算資源の制約

SCP の実装における最大の課題の一つは、計算コストである。変数の数の二乗に比例して LLM 呼び出し回数が増加するため、大規模な問題では実行時間とコストが問題となる。

この課題に対しては、例えば段階的なアプローチが有効である。まず、ドメイン知識や予備的な統計解析に基づいて変数を絞り込み、重要な変数間の関係に焦点を当てる。また、バッチ処理や並列化により、効率的な実行を図ることも重要である。

5.3.2 専門家による検証の必要性

LLM の出力は完全ではなく、時として誤った推論を行う可能性がある。特に、専門性の高い分野や最新の知見が関わる場合は、注意が必要である。

実務では、LLM の推論結果を専門家がレビューする体制を構築すべきである。SCP の結果を可視化し、推論過程を透明化することで、専門家が効率的に検証できる環境を整える必要がある。

5.3.3 継続的な改善プロセス

LLM 技術は急速に進化しており、新しいモデルが次々と登場している。SCP の性能を最大化するには、最新の LLM を活用し、プロンプト設計を継続的に改善する必要がある。

また、特定分野に特化したファインチューニングや、RAG（Retrieval-Augmented Generation）技術の活用により、より専門的で正確な推論が可能になる。組織内での知識蓄積と共有の仕組みも重要である。

5.4 責任あるAI活用に向けて

5.4.1 透明性の確保

SCP を意思決定に活用する際は、推論過程の透明性が重要である。なぜその因果関係が推定されたのか、LLM がどのような知識に基づいて判断したのかを明確にする必要がある。

プロンプトと応答の記録を保存し、監査可能な形で管理することが推奨される。これにより、結果の再現性を確保し、問題が生じた場合の原因究明も可能になる。

5.4.2 倫理的配慮

因果推論の結果は、政策決定や個人の処遇に影響を与える可能性がある。特に、社会的に敏感な属性（性別、人種、社会経済的地位など）が関わる場合は、慎重な扱いが必要である。

SCP を用いる際も、公平性と説明責任を重視し、結果の解釈と適用において倫理的ガイドラインに従うべきである。また、因果関係の推定が差別や偏見を助長しないよう、継続的なモニタリングが必要である。

6. 今後の展望と課題

SCP は因果推論の新たな地平を開いたが、まだ発展の初期段階にある。ここでは、技術的・理論的な改善の方向性と、社会実装に向けた展望を議論する。

6.1 技術的限界と改善の方向性

6.1.1 大規模ネットワークへの対応

現在の SCP の最大の制約は、計算量が $O(n^2)$ であることである。例えば 100 変数を超えるような大規模ネットワークでは、現実的な時間とコストでの実行が困難である。

この問題に対する有望なアプローチの一つは、階層的な因果探索である。まず変数をグループ化し、グループ間の因果関係を推定した後、各グループ内の詳細な関係を分析する。このようなトップダウンアプローチにより、計算量を大幅に削減できる可能性がある。

また、因果関係の疎性を仮定し、重要な変数ペアのみを選択的に評価する手法も検討に値する。ドメイン知識や予備的な統計解析に基づいて、評価すべき変数ペアを事前にスクリーニングすることで、LLM 呼び出し回数を削減できる。

6.1.2 より高度なプロンプト設計

現在のプロンプト設計は、主に経験的な試行錯誤に基づいている。今後は、日々発展を続けているプロンプト技術を様々に取り入れ、因果推論の目的に応じて最適なプロンプト設計を追究する必要がある。

特に重要なのは、LLM の推論過程をより細かく制御する方法の開発である。例えば、因果関係の時間的順序、介入可能性、交絡要因の存在など、因果推論の重要な概念を明示的にプロンプトに組み込むことで、より精密な推論が可能になるだろう。

また、Few-shot 学習や Chain-of-Thought の発展形など、新しいプロンプティング技術の活用も期待される。分野特有の因果推論パターンを学習させることで、専門性の高い推論が可能になる。

6.1.3 複数LLMのアンサンブル

単一の LLM に依存するのではなく、複数の LLM の推論結果を統合するアンサンブル手法の開発も重要である。異なる LLM は異なる強みを持つため、それらを組み合わせることで、より頑健で信頼性の高い推論が可能になる。

例えば、一般的な知識に強い LLM、医学に特化した LLM、統計的推論に優れた LLM を組み合わせることで、それぞれの専門性を活かした総合的な判断が可能になる。投票メカニズムやベイズ的統合など、複数の推論結果を統合する理論的枠組みの構築が必要である。

6.2 理論的発展の可能性

6.2.1 因果推論理論への貢献

SCP は、統計的手法と知識ベース手法を融合する新しいパラダイムを提示している。この枠組みは、因果推論の理論的発展にも寄与する可能性がある。

特に興味深いのは、LLM の確率的出力を因果推論の不確実性と結びつける理論の構築である。LLM の確信度を事前分布として解釈し、データから得られる尤度と組み合わせることで、ベイズ的因果推論の新しい形式を確立できる可能性がある。

6.2.2 新しい評価指標の開発

現在の評価指標は、主に構造的な一致度に焦点を当てている。しかし、因果推論の実用性を評価するには、より多面的な指標が必要である。

例えば、推定された因果関係の介入可能性、政策的重要性、予測精度への寄与度など、実用的観点からの評価指標の開発が求められる。また、LLM の推論の質を評価する指標も必要である。推論の論理的一貫性、既知の事実との整合性、不確実性の適切な表現など、多角的な評価が重要である。

6.2.3 他の機械学習手法との統合

SCP は、他の機械学習手法と組み合わせることで、さらなる発展が期待できる。例えば、深層学習による特徴抽出と SCP を組み合わせることで、高次元データからの因果発見が可能になるかもしれない。

また、強化学習の枠組みで SCP を活用することで、介入の効果を学習しながら因果構造を発見する、より動的なアプローチも考えられる。これは、実験計画法と因果推論を統合する新しい研究方向を示唆している。

6.3 社会実装に向けたロードマップ

6.3.1 段階的な導入戦略

SCP の社会実装は、段階的に進めることが重要である。まず、リスクの低い領域での小規模な実証実験から始め、徐々に適用範囲を拡大していくべきである。

例えば、第 1 段階では、研究機関や大学での基礎研究への活用から始める。ここでは、手法の改善と検証に重点を置く。第 2 段階では、企業の研究開発部門での活用を進める。製品開発や品質管理など、限定的な領域での応用を試みる。第 3 段階では、より広範な実務への展開を図る。その際、医療、金融、政策立案など、社会的影響の大きい分野への適用を慎重に進める。こうした段

階的な拡張においては、各段階について、フィードバックに基づく改良を繰り返しながら、次の段階での実証に進むかどうかを判断していく。

6.3.2 産学連携の推進

SCP の発展には、学術界と産業界の密接な連携が不可欠である。大学や研究機関が理論的基盤を提供し、企業が実データと実用的課題を提供することで、相乗効果が期待できる。

共同研究プロジェクトの設立、データ共有の枠組み構築、人材交流の促進など、具体的な連携メカニズムの整備が必要である。また、成功事例の共有と普及も重要であり、ワークショップやカンファレンスの開催が推奨される。

6.3.3 国際標準化への取り組み

SCP が広く採用されるためには、手法の標準化が重要である。評価基準、実装ガイドライン、倫理的指針など、国際的に合意された標準の策定が必要である。

また、異なる言語や文化圏での SCP の適用可能性を検証することも重要である。LLM の多言語対応能力を活用しつつ、各地域の文脈に適したカスタマイズを行う必要がある。国際的な研究コンソーシアムの形成により、グローバルな知見の集約と共有を進めることが期待される。

7. おわりに

本稿では、我々が提案した SCP について、その理論的基礎から実用的応用まで幅広く解説してきた。SCP は、大規模言語モデルの知識と統計的因果探索の手法を革新的な方法で統合し、因果推論の新たな可能性を切り開いた。

7.1 本研究の意義の総括

SCP の最も重要な貢献は、これまで独立に発展してきた二つのアプローチ、すなわちデータ駆動的な統計手法と知識ベースの推論を、実用的な形で融合したことである。この統合により、それぞれの手法を相互に補完し、より信頼性の高い因果推論が可能になった。

実証研究では、既知のベンチマークデータセットにおいて顕著な性能向上が確認された。特に重要なのは、LLM の学習データに含まれていない健康診断データでも有効性が実証されたことである。これは、SCP が単なる記憶の再現ではなく、真の知識統合による推論を実現していることを示している。

また、本研究は手法の限界についても率直に議論し、計算量の問題、LLM の不確実性、実装上の課題など、今後解決すべき点を明確にした点も付記しておく。

7.2 因果推論研究の新時代への展望

SCP の登場は、因果推論研究の新しい時代の幕開けを告げている。AI と

人間の知識が協調して複雑な問題に取り組む時代が到来したのである。今後、LLM 技術のさらなる発展により、より高度で専門的な推論が可能になることが期待される。

しかし、技術の進歩と同時に、その責任ある使用についても真剣に考える必要がある。因果推論の結果は、医療、経済、社会政策など、人々の生活に直接影響を与える分野で使用される。したがって、技術の開発と並行して、倫理的ガイドラインの整備、専門家による検証体制の構築、透明性の確保など、社会的な枠組みの整備も不可欠である。

また、SCP の普及には、研究者、実務家、政策立案者など、様々なステークホルダーの協力が必要である。学際的な対話を通じて、技術の可能性と限界について共通理解を形成し、適切な活用方法を模索していくことが重要である。

7.3 読者へのメッセージ

因果関係の理解は、科学の根幹をなす営みである。なぜ物事が起こるのか、どのように介入すれば望ましい結果が得られるのか。これらの問いに答えることは、人類の知的探求の中心的課題であり続けてきた。

SCP は、この古くて新しい問題に対する革新的なアプローチを提供している。しかし、これは決して万能の解決策ではない。むしろ、人間の専門知識と AI の計算能力を組み合わせることで、より深い理解に到達しようとする試みである。

読者の皆様には、この新しい技術の可能性に期待しつつも、批判的な視点を持って接していただきたい。特に、自身の専門分野で SCP をどのように活用できるか、どのような課題があるか、積極的に検討していただければ幸いである。技術の真の価値は、それを使う人々の創造性と責任感によって決まるのである。

最後に、統計的因果探索と大規模言語モデルに関する日本の研究者に敬意を表するとともに、この分野のさらなる発展を期待したい。因果推論の新時代において、日本の研究者が重要な貢献を果たしていることは、誠に喜ばしいことである。今後も、国際的な協力のもと、この革新的な研究が人類の知識の前進に寄与することを願ってやまない。

参考文献

Shimizu, S., Hoyer, P. O., Hyvärinen, A., & Kerminen, A. (2006). A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research, 7*, 2003-2030.

Takayama, M., Okuda, T., Pham, T., Ikenoue, T., Fukuma, S., Shimizu, S., & Sannai, A. (2025). Integrating Large Language Models in Causal Discovery: A Statistical Causal Approach. *Transactions on Machine Learning Research*, 10 May 2025.

オフ方策評価の概要とそのリスクリターンに基づく有用性評価

齋藤 優太｜コーネル大学コンピュータサイエンス学部博士課程在籍

齋藤 優太

コーネル大学コンピュータサイエンス学部博士課程在籍。2021年3月、東京工業大学（当時）にて学士号を取得。主に機械学習分野やデータマイニング分野の主要国際会議に多数論文を出版。また、Netflix、Spotify、DMM、Sony、CyberAgent、LINEヤフーなど国内外のテック企業と連携し、検索・レコメンドの戦略策定やオフライン／オンライン評価手法の構築に従事。また、RecSys、KDD、IBISなどの国内外の会議にて、推薦システムや反実仮想学習に関するチュートリアルを実施。著書に『施策デザインのための機械学習入門』『反実仮想機械学習』がある。WSDM 2022 Best Paper Award Runner-Up、RecSys 2022 ~ 2024 Outstanding Reviewer Award、2021年イノベーション大賞内閣総理大臣賞、2022年 Forbes JAPAN「30 UNDER 30」、孫正義育英財団第5期生選出などの受賞歴。

要約

本稿では、過去のデータを活用して新たな意思決定方策の性能を推定するオフ方策評価と呼ばれる統計技術の基本概念、主要な推定量（DM、IPS、DR）、およびそれらの理論的性質や実験的比較を概説する。また、近年の研究動向として、推定量選択の自動化や大規模行動空間への対応といった課題についても取り上げる。さらに、リスク調整後の性能評価を可能にする最新の指標「SharpeRatio@k」について紹介し、オフ方策評価の実運用における信頼性や安全性の観点を強調する。金融・医療等のリスク感度の高い応用分野における有用性にも言及し、今後の応用・発展の方向性を示す。

1．はじめに

機械学習は、正確な予測を可能にする技術として注目を集めている。多くの論文は、解く意味があるとされているタスクにおいて、より良い予測精度を達成することを至上命題としている。もちろん（天気予報など）機械学習等によって得られる予測値をそのまま活用する場面もあり、学術研究の成果が実務に直結するケースも少なくない。しかし現実世界における応用場面に目を向けると、機械学習による予測値をそのまま用いるのではなく、予測値に基づいて何らかの意思決定を行っていることが多い（Swaminathan and Joachims,2015）。例えば、ECや音楽配信サービスの推薦システムでは、各ユーザーとアイテムのペアごとにクリック確率を予測し、その予測値に基づいてどのアイテムを推薦すべきか決定する。または、過去の株価の系列に基づいて最適なポートフォリオを決定する。あるいは、商品の購入確率予測に基づいてユーザーごとにどの商品の広告を提示するか決定する、などが典型例である（Saito and Joachims,2022）。実際、株式会社ZOZOが運営するファッション通販サイト「ZOZOTOWN」におけるファッションアイテム推薦枠（図表1）では、データ駆動のアルゴリズムに基づいて推薦の意思決定が自動化されていることが知られている（Saito et al.(2020)）。これらの例では、クリック確率や購入確率の予測精度よりも、それに基づく推薦や広告配信などの意思

決定の性能に関心がある。

本稿の主たる興味は、機械学習による予測値などに基づいて導かれる意思決定方策（decision-making policy）の性能を正確に評価することである。意思決定方策の性能評価の方法として理想的なのは、方策を実環境あるいはサービスに実装することでその成果を直接的に計測するオンライン実験（A/B テストと呼ばれることもある）だろう。しかし、オンライン実験には時間がかかることや、多大な実装労力が伴うという欠点がある（Levine et al.(2020)）。また性能の悪い意思決定方策を実験してしまったときに、実験期間においてユーザー体験を害してしまったり、収益を毀損してしまう恐れもある（Chandak et al.(2021)）。したがって、古い意思決定方策により既に蓄積されたログデータのみを用いて、新しい意思決定方策の性能を事前に見積もりたい、というモチベーションが自然に生じる。

このように、新たな意思決定方策の性能を蓄積データのみを用いて推定する問題のことをオフ方策評価（Off-Policy Evaluation; OPE）と呼ぶ。正確なオフ方策評価は、機械学習やデータ駆動の意思決定の実践において多くのメリットをもたらす。例えば、時間を要するオンライン実験を行うことなく方策の性能を正確に見積もることが可能ならば、オフ方策評価の結果をもとに方策の改善サイクルを素早く回すことができるだろう。また、ハイパーパラメータや機械学習アルゴリズムの組み合わせを変えることで多数生成される意思決定方策の候補のうち、どの候補を実装すべきか、あるいはどの候補についてオンライン実験を行うべきなのか事前に絞り込むことも可能になるだろう。オフ方策評価はこれらの実利から、産業および学術界において大きな注目を集めている（Uehara, et al.(2022)）。

以降では、オフ方策評価の標準的な定式化と推定量を紹介する。また、それらの基本的な理論性質や最近の研究動向、今後の課題についても解説する。最後に、最新の研究紹介として、金融におけるリスク・リターントレードオフの考え方を応用したオフ方策評価の推定量の精度評価フレームワークについて、簡単に紹介する。

図表1　ZOZOTOWNのファッションアイテム推薦枠

出所）Saito et al.(2020)、Figure 1

2．オフ方策評価の定式化

本節では、オフ方策評価の問題を具体的に定式化する。まず、意思決定問題における三大要素を導入する。具体的に、特徴量（feature）ベクトルを x、離散的な行動（action）を a、観測される報酬（reward）を r とする。これらに関する具体例を、図表2に示した。

図表2　各種応用における特徴量･行動･報酬の例

応用例	特徴量	行動	報酬
映画推薦	視聴履歴	映画	視聴時間
クーポン配布	購買履歴	クーポンの種類	収益
投薬	過去の検査結果	薬の種類	生存確率
金融	過去の株価推移	ポートフォリオ	投資損益

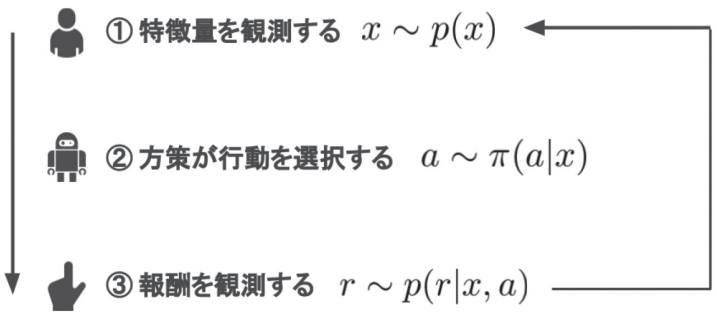
出所）筆者作成

さて意思決定方策とは、特徴量 x で条件付けられた行動空間上の確率分布である。つまり $\pi(a|x)$ は、あるベクトル x で特徴付けられるデータに対して、a という行動を選択する確率のことである。すなわち、$\pi(a|x) \geq 0, \forall(x,a)$ かつ $\sum_a \pi(a|x) = 1$、$\forall x$ が成り立つ。なお、本稿ではより一般的な確率的意思決定方策を主に考えるが、その特殊ケースとして、特徴量 x に対してある行動を確率

1 で選択する決定的な意思決定方策を考えることもできる。

図表３に、意思決定方策による一連の意思決定プロセスを示した。我々はまず、未知の確率分布 $p(x)$ に従う特徴量 x を観測する。これは、YouTube などのプラットフォームにユーザーが訪問したり、直近の株価の変動系列を観測することなどに対応する。次に観測した特徴 x に基づき、意思決定方策が行動を選択する。これは、x で特徴付けられるユーザーに対して推薦すべき動画 a を決めることや、直近の株価系列 x に基づきポートフォリオの構成に変更を加える a といった意思決定を行う場面に対応する。最後に特徴量 x と行動 a の両方に依存する形で報酬 r が未知の確率分布 $p(r|x,a)$ に従い観測される。この意思決定プロセスを大量に繰り返すことで、期待報酬の最大化を目指すのが意思決定方策の役割である。

図表3　方策による意思決定プロセス



出所）筆者作成

ここで、ある意思決定方策の性能（policy value とも呼ばれる）を、次のように定義する。

$$V(\pi) = E_{p(x)\pi(a|x)p(r|x,a)}[r] = E_{p(x)\pi(a|x)}[q(x,a)].$$

なお $q(x,a)=E[r|x,a]$ は、特徴量 x と行動 a を条件付けたときの報酬の期待値であり、期待報酬関数と呼ばれる。

$V(\pi)$ は、方策 π を環境あるいはサービスに実装した際に得られる報酬の期待値である。例えば、報酬がクリックの有無や運用損益で定義されるならば、$V(\pi)$ は方策が達成する期待クリック数や期待運用損益を表し、意思決定方策の性能の定義として妥当だろう。なお、報酬 r の定義は応用先やサービスの目的等によって異なり、実践者が適切に定める必要がある（図表２を参照）。

さて本稿の目的は、方策の性能、すなわち $V(\pi)$ を評価・推定することだった。仮に、方策を実システムに一定期間実装するオンライン実験が実行可能ならば、その期間に観測される報酬の経験平均を計算することで、方策の性能を正確に推定できる。しかし、冒頭で説明したように、オンライン実験はさまざまな理由で実施が難しいことが多い。また、仮にオンライン実験が可能だったとしても、ハイパーパラメータや最適化アルゴリズムなどを変更することで生

まれる大量の方策の候補すべての性能をオンライン実験で推定することは非現実的である。よってオンライン実験の代替として、あるいはオンライン実験を行うべき少数の方策を選択するためのスクリーニングの手段として、方策の性能をオフ方策評価したいというモチベーションが生じる。より具体的には、過去に運用されていたデータ収集方策（logging policy）と呼ばれる意思決定方策 π_0 によって既に収集された以下のログデータ D のみを用いて、π_0 とは異なる新たな方策 π の性能を正確に推定することを考えたい。

$$D = \{(x_i, a_i, r_i)\}_{i=1}^{n}$$

上記のログデータ D は、データ収集方策が環境で稼働していた際に観測された特徴量 x_i と選択された行動 a_i、そしてその結果として観測された報酬 r_i によって構成されている。

オフ方策評価の研究における主たる目標は、データ収集方策 $\pi_0\ (a|x)$ が収集したログデータ D のみを用いて、未だ実装したことのない新たな方策の性能 $V(\pi)$ をより正確に推定できる推定量 $\hat{V}$ を構築することである。ここで推定量 $\hat{V}$ とは、ログデータと評価対象の方策を入力として、真の性能を統計的に近似する次のような関数のことである（Dudik et al.(2014)）。

$$V(\pi) \approx \hat{V}(\pi ; D)$$

推定量 $\hat{V}$ の正確さ（推定精度）は、統計学における典型的な精度指標である平均二乗誤差（mean-squared-error; MSE）に基づき評価されることがほとんどである。

$$MSE\left(\hat{V}(\pi)\right) = E\left[\left(V(\pi) - \hat{V}(\pi; D)\right)^2\right] = Bias\left(\hat{V}(\pi)\right)^2 + Var\left(\hat{V}(\pi)\right)$$

なお、

$$Bias\left(\hat{V}(\pi)\right) = V(\pi) - E[\hat{V}(\pi; D)]$$

$$Var\left(\hat{V}(\pi)\right) = E\left[\left(\hat{V}(\pi; D) - E\left[\hat{V}(\pi; D)\right]\right)^2\right]$$

は、それぞれ推定量 $\hat{V}$ のバイアスとバリアンスである。ここで $E[\hat{V}(\pi;D)]$ は、推定量の期待値を表す。平均二乗誤差はバイアスの二乗とバリアンスの和に等しく、より良い平均二乗誤差を達成するためには、推定量のバイアスとバリアンスを抑えることが重要であることがわかる。

3．標準的な推定量とその性質

本節では、オフ方策評価における標準的な推定量として、Direct Method・Inverse Propensity Score・Doubly Robust という３つの推定量とそれらの基礎的な性質を紹介する。またシミュレーションデータを用いた簡単な実験を通して、これらの推定量の挙動を理解する。

3.1 Direct Method（DM）

DM 推定量ではまず、期待報酬関数 $q(x,a)$ に対する予測モデル $\hat{q}(x,a)$ を事前に得ておく。これは、例えば、ログデータを用いて、次のような回帰問題を解くことに対応する。

$$\hat{q}(x,a) = argmin_{q'} \sum_{i=1} \left(r_i - q'(x_i, a_i) \right)^2$$

ここで $\hat{q}(x、a)$ は期待報酬関数を推定するための予測モデルであり、線形回帰やサポートベクトルマシン、ランダムフォレストなど、よく知られた機械学習の手法を用いて定義できる（Farajtabar et al.(2018)）。

DM 推定量は、事前に得た報酬の予測モデル $\hat{q}(x,a)$ を用いて、次のように定義される。

$$\hat{V}_{DM}(\pi; D) = \frac{1}{n} \sum_{i} E_{\pi(a|x_i)}[\hat{q}(x_i, a)]$$

つまり DM 推定量は、方策の真の性能 $V(\pi)$ の定義において未知である期待報酬関数 $q(x,a)$ を、データから推定された $\hat{q}(x,a)$ で置き換えるというアイデアに基づく。その定義から推察される通り、DM 推定量の推定精度（平均二乗誤差）は、期待報酬関数 $q(x,a)$ の推定精度に大きく依存することが知られている。つまり、予測モデル $\hat{q}(x,a)$ が真の報酬関数 $q(x,a)$ を正確に近似するものであれば、DM 推定量 $\hat{V}_{DM}(\pi; D)$ も方策の真の性能 $V(\pi)$ を正確に推定できる。逆に、$\hat{q}(x,a)$ の推定精度が悪ければ、DM 推定量の推定精度もそれに連動して悪くなる。観測ノイズ、高次元特徴量、特徴量の分布シフトなどさまざまな困難が存在する現実において、真の期待報酬関数 $q(x,a)$ を正確に推定することは往々にして難しい（Farajtabar et al.(2018)）。これらの点から、DM 推定量は平均二乗誤差の推定量のなかでもバイアスが大きくなりやすいという問題を抱えていることが知られている。

3.2 Inverse Propensity Score（IPS）

IPS 推定量は DM 推定量とは大きく異なり、ログデータに含まれる報酬 r_i の重み付け平均により、方策の真の性能 $V(\pi)$ を推定する。同様の推定量のことを、Importance Sampling（IS）や Inverse Probability Weighting（IPW）と呼ぶこともある。

$$\hat{V}_{IPS}(\pi; D) = \frac{1}{n} \sum_{i} w(x_i, a_i)\, r_i$$

ここで、$w(x,a)=\pi(a|x)/\pi_0(a|x)$ は重要度重み（importance weight）と呼ばれる重みであり、「ある特徴量 x に対して、評価対象の方策がある行動 a を選択する確率」と「同じ特徴量 x に対して、データ収集方策が同じ行動 a を選択する確率」の比によって定義される。この重みは、収集されたデータが本来評価したい方策にとってどれだけ「重要」であるか、すなわち評価方策の観点から見た代表性や寄与の度合いを補正するために用いられる。

IPS 推定量は、次の共通サポート（common support）と呼ばれる条件のもとで、不偏性（すなわち、バイアスがゼロ）を持つことが知られている。

データ収集方策 π_0 は、ある評価対象の方策について
$\pi(a|x)>0 \Rightarrow \pi_0(a|x), \forall (x,a)$ であるとき、共通サポートを持つという。

すなわち共通サポートの条件は、ある行動 a が評価対象の方策によってある正の確率で選択されるならば、過去に採られていたデータ収集方策 π_0 もその行動をある正の確率で選択していなければならないことを要求する。これは言い換えると、データ収集方策が評価方策のもとで選択される可能性のある（選択確率がゼロではない）すべての行動の情報を集め、それがログデータとして収集されていることを要求していると理解することができる。

この共通サポートの条件のもとで、IPS 推定量の不偏性（IPS 推定量の期待値が、方策の真の性能に一致すること）を次のように確かめることができる（以下の期待値計算については、読み飛ばしても内容の理解に大きな影響はない）。

$$
\begin{aligned}
&E\left[\hat{V}_{IPS}(\pi;D)\right] \\
&= \frac{1}{n}\sum_i E_{p(x)\pi_0(a|x)p(r|x,a)}\left[\frac{\pi(a_i|x_i)}{\pi_0(a_i|x_i)} r_i\right] \\
&= E_{p(x)\pi_0(a|x)}\left[\frac{\pi(a|x)}{\pi_0(a|x)} q(x,a)\right] \\
&= E_{p(x)}\left[\sum_a \pi_0(a|x) \frac{\pi(a|x)}{\pi_0(a|x)} q(x,a)\right] \\
&= E_{p(x)}\left[\sum_a \pi(a|x) q(x,a)\right] \\
&= V(\pi)
\end{aligned}
$$

すなわち、IPS 推定量は $V(\pi)$ に対する不偏推定量であり、これは平均二乗誤差の構成要素のうちバイアスが完全にゼロであることを意味する。なお共通サポートの条件が満たされないとき IPS 推定量はもはや不偏ではなく、評価対象の方策に依存する量のバイアスを持つことが知られている（Sachdeva et al.(2020)）。

一方で、IPS 推定量のバリアンスは、次のように与えられることが知られている（Dudik et al.(2014)）。

$$
\begin{aligned}
n\,Var\left(\hat{V}_{PS}(\pi;D)\right) &= E_{p(x)\pi_0(a|x)}\left[w(x,a)^2\,\sigma^2(x,a)\right] \\
&+ Var_{p(x)}\left[E_{\pi_0(a|x)}\left[w(x,a)q(x,a)\right]\right] + E_{p(x)}\left[Var_{p(x)}\left[w(x,a)q(x,a)\right]\right]
\end{aligned}
$$

ここで $\sigma^2(x,a)=Var[r|x,a]$ は、x と a で条件付けたときの報酬のバリアンス（ノイズ）である。すなわち、報酬のノイズが大きい（$\sigma^2(x,a)$ が大きい）、言い換えると同一の特徴量と方策の組み合わせに対して得られる報酬の振れ幅が

大きい場合や、データ収集方策と評価対象の方策が大きく異なる（重要度重み $w(x,a)$ の値が大きくなり得る）場合に、IPS 推定量のバリアンスは非常に大きくなってしまう可能性がある。

以上の内容から、DM 推定量と IPS 推定量の間にはバイアスとバリアンスについての明確なトレードオフが存在することがわかる。IPS 推定量は不偏推定量であるため、ログデータに含まれるデータ数 n が大きければ IPS 推定量がより正確であることが多いが、逆にデータが少ないときには DM 推定量の方が正確なことがある（Su et al.(2019)）。

3.3 Doubly Robust（DR）

これまでに紹介した DM 推定量と IPS 推定量には、バイアスとバリアンスについて明確なトレードオフが存在していた。より良い推定量を構築するための自然なアイデアは、異なる利点を持つ DM 推定量と IPS 推定量をうまく組み合わせることだろう。DR 推定量はそれを体現した推定量であり、次のように定義される（Dudik et al.(2014), Jiang and Li(2016)）。

$$\hat{V}_{DR}(\pi; D) = \frac{1}{n}\sum_{i} E_{\pi(a|x_i)}[\hat{q}(x_i, a)] + w(x_i, a_i)\big(r_i - \hat{q}(x_i, a_i)\big)$$

つまり、DR 推定量は期待報酬関数に対する予測モデルを用いる DM 推定量をベースラインとしつつも、第二項において IPS 推定量と類似する方法により期待報酬関数の予測モデル $\hat{q}(x_i, a_i)$、の推定誤差を補正している。このように 2 つの推定量を組み合わせることで、DR 推定量は IPS 推定量の不偏性を保ちつつ、多くの場合 IPS 推定量よりも小さいバリアンスを達成することが知られている。この DR 推定量の利点は、そのバリアンスを実際に計算することで、より正確に理解できる。

$$n\,Var\left(\hat{V}^{DR}(\pi; D)\right) = E_{p(x)\pi_0(a|x)}\left[w(x,a)^2\,\sigma^2(x,a)\right]$$
$$+ Var_{p(x)}\left[E_{\pi_0(a|x)}\left[w(x,a)q(x,a)\right]\right] + E_{p(x)}\left[Var_{p(x)}\left[w(x,a)\,\Delta(x,a)\right]\right]$$

ここで、$\Delta(x,a) = q(x,a) - \hat{q}(x,a)$ は、期待報酬関数の推定バイアスである。ここで導入した DR 推定量のバリアンス表現は、全分散の法則（law of total variance）を 2 度用いることで得ることができる。また、DR 推定量のバリアンスにおいて $\hat{q}(x,a)=0 \Rightarrow \Delta(x,a)=q(x,a)$ とすることで、前節において紹介した IPS 推定量のバリアンスを得ることができる（$\hat{q}(x,a)=0$ としたとき、DR 推定量の定義が IPS 推定量の定義に一致するため）。

さて、IPS 推定量と DR 推定量のバリアンスを比較すると、第三項のみが異なることがわかる。すなわち、期待報酬関数の予測モデル $\hat{q}(x,a)$ が正確であるほど、DR 推定量のバリアンスは小さくなる。一方で、IPS 推定量は $\hat{q}(x,a)$ に非依存であり、そのバリアンスは評価対象の方策および期待報酬や報酬のノイズの大きさのみから決定される。これらが方策の性能 $V(\pi)$ の推定量により大きな悪影響を及ぼすこともありうる。

ここで行った IPS 推定量と DR 推定量のバリアンスの比較は、DR 推定量

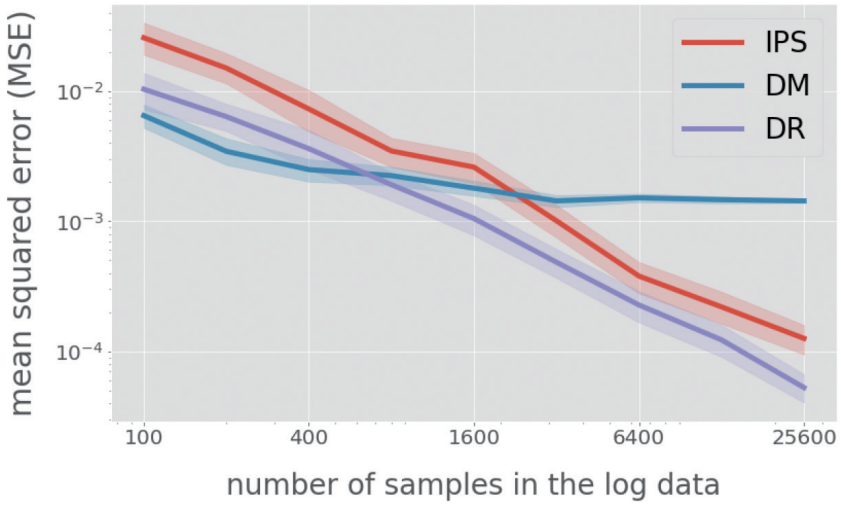
の利点を把握するためのごく初歩的なものである。実際には、先に導入した共通サポートの条件が満たされない状況におけるバイアスなど、より多角的な推定量の比較が可能であり、多くの場合、DR 推定量がほか 2 つの推定量よりも望ましい理論性質を持つことが知られている（齋藤 (2024)）。

3.4 推定量の性能の比較

本小節では、シミュレーションデータを用いて、オフ方策評価における標準的な推定量である DM 推定量・IPS 推定量・DR 推定量の挙動を比較する。実験の実装には、3.6 節で後述する Open Bandit Pipeline を用いた。方策の行動数を 20 で固定し、データ数を 100、200、400、800、…、25,600 と変化させ、事前に用意した評価対象の方策の性能を、DM 推定量・IPS 推定量・DR 推定量によりオフ方策評価を行う。これをデータ数ごとに 100 回繰り返すことで各推定量の平均二乗誤差を算出し、結果を図表 4 に示した（図表 4 は両対数グラフであることに注意されたい）。

まず、データ数が小さいときは DM 推定量が最も正確であることがわかる。これはデータ数が小さい状況では、平均二乗誤差のうちバリアンスが支配的であり、バリアンスが比較的小さい DM 推定量が IPS 推定量や DR 推定量よりも小さい平均二乗誤差を達成するからである。一方で、データ数が大きくなるにつれ、IPS 推定量と DR 推定量はバリアンスを軽減し、徐々に正確さを増すことがわかる。これらに対し、不偏推定量ではない DM 推定量は、データ数に関わらず一定のバイアスを発生してしまうため、平均二乗誤差はほとんど改善していない。このように DM 推定量と IPS 推定量・DR 推定量は、データ数の変化に対して、全く異なる挙動を持つことがわかる。また DR 推定量は IPS 推定量と同様の挙動を示しつつも、そのバリアンス軽減の効果により、IPS 推定量よりも正確なオフ方策評価を達成していることもわかる。

図表4　データ数nが変化したときの推定量の平均二乗誤差の比較



出所）筆者作成

3.5 その他の推定量

本節では、DM 推定量・IPS 推定量・DR 推定量というオフ方策評価において標準的な推定量の定義およびそれらの基礎的な性質について紹介した。また推定量の挙動の違いを簡易的な実験を行うことで理解した。本節で紹介した推定量の定義や性質を把握しておけば、他の最新でより発展的な推定量の構造や理論性質についても困難なく理解できるはずである。より発展的な内容に興味を持った読者は、Switch（Wang et al.(2017)）、More Robust DR（Farajtabar et al.(2018)、Continuous Adaptive Blending（Su et al.(2019)）、DR with Optimistic Shrinkage（Su et al.(2019)）などの推定量について調べてみると、オフ方策評価の実践における選択肢が増えるだろう。

3.6 オフ方策評価の実装

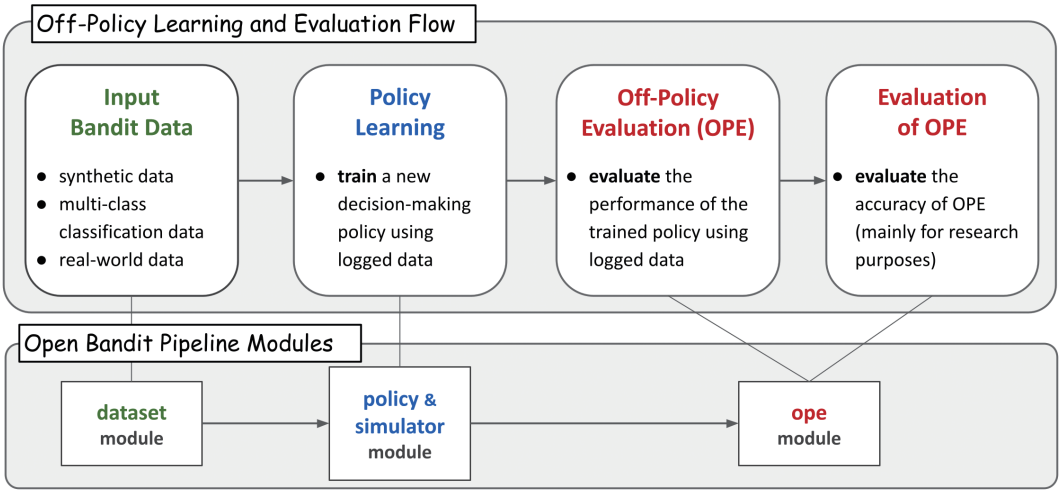
オフ方策評価は非常に実践的なモチベーションに端を発する分野であるが、実際はやや理論偏重的な論文が多数を占め、特に企業の実務家など実践者にとって参入障壁が高い現状がある。誰でも簡単にオフ方策評価の恩恵を受けることができる状況を作り、かつ学術研究における実験評価の再現性も担保すべく、著者らの研究グループは、オフ方策評価のための Python パッケージ「Open Bandit Pipeline（OBP）」を GitHub（https://github.com/st-tech/zr-obp）にて公開している（Saito et al.(2020)）。OBP を活用することで、研究者はオフ方策評価の実装に集中しつつ、再現性のある手順で推定量の精度比較を行うことができる。また現場の機械学習エンジニアやデータサイエンティストは、既に実装されている幅広いかつ最新の推定量を用いて、容易にオフ方策評価を実践できる。OBP は、機械学習において世界的に著名な学会に採択されたオフ方策評価に関する多数の論文の実験で活用されるなど、学界においても一定の地位を得ている。

OBP は、主に以下のモジュールで構成される。

- dataset モジュール：オフ方策推定量の評価に適したシミュレーションデータの生成や多クラス分類データをオフ方策評価で用いられるログデータに変換することで論文執筆用の実験を円滑に行うための機能を提供。
- policy モジュール：標準的なバンディットアルゴリズムやオフ方策学習アルゴリズムを提供。
- ope モジュール： 本節で解説した 3 つの推定量を含む幅広いオフ方策推定量の実装を提供。また、新たにオフ方策推定量を実装するための柔軟なインターフェースも提供。

図表 5 に、OBP の全体像を示した。なお本稿にて行ったいくつかの簡易実験は OBP で実装されており、詳細は Google Colaboratory にて公開している（https://drive.google.com/drive/folders/1YcT7Y0xkVmX_WOyyER3G32bw7DfpEEwA?usp=sharing）。オフ方策評価に関する研究開発や応用に関心を持った読者は、そちらも参考にされたい。

図表5　Open Bandit Pipelineの概要



出所）Saito et al.(2020)、Figure 3

4．近年の研究動向と課題

本節では、オフ方策評価に関連する注目すべき研究動向やビジネスへの実応用を見据える上での課題を紹介する。なお、本節で取り上げる内容は著者の興味・関心に大きく依存していることには注意されたい。

4.1 付随するタスクの自動化

オフ方策評価の推定量には、ハイパーパラメータを持つものも多い。例えば、3節で紹介したDM推定量やDR推定量を用いるには、事前に期待報酬関数を予測する必要がある。ここで、予測モデル $\hat{q}(x,a)$ を得るために用いる機械学習手法やその機械学習手法にまつわるハイパーパラメータは、推定量のハイパーパラメータと見なすことができる。また、より発展的な推定量の中には、バイアスとバリアンスのトレードオフを大きく左右する推定量独自のハイパーパラメータを有しているものも多い。これらの推定量のハイパーパラメータの選択は、推定量の精度に大きな影響を与えることが報告されている（Su et al.(2019)）。しかし、特徴量の分布など取り巻く環境が変化する中で、推定量のハイパーパラメータを実践者の目算やセンスに基づいて適切に定めることは容易ではない。そこで、観測可能なログデータのみに基づいて推定量のハイパーパラメータを選択する方法の研究が徐々に行われるようになっている（Tucker and Lee(2021)）。

また、問題設定に応じて適切な推定量を自動で選択する問題も、オフ方策評価の実応用上重要である（Udagawa et al.(2022)）。すなわち、本稿で紹介したDM推定量・IPS推定量・DR推定量に留まらず多数の推定量が提案されている現状において、実践者は所与の問題設定ごとにどの推定量を用いるべきか？という難しい判断に迫られることになる。推定量の精度は問題設定に依存して大きく変化するため、どの推定量を用いるべきかという判断を適切に下せ

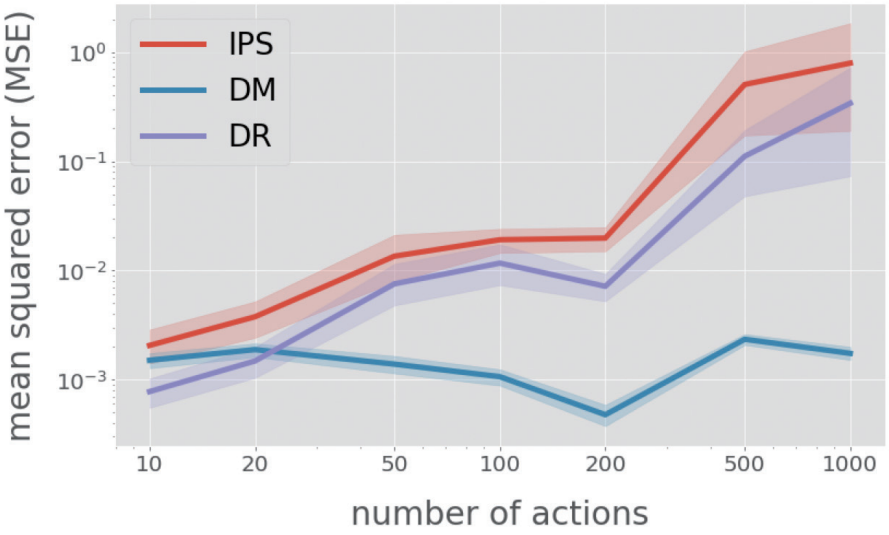
ることは、オフ方策評価の応用上極めて重要である。

この事実は、３節で行った実験の結果（図表４）からも読み取ることができる。データ数 n が小さいときは DM 推定量が最も正確であるが、n が大きくなるにつれ IPS 推定量や DR 推定量の正確さが増している。ここで問題になるのは、現実において図表４に描かれているような平均二乗誤差の値にはアクセスできないという事実である。すなわち図表４を見ると、n=200 以下においては DM 推定量を用い、それ以外の場合は DR 推定量を用いるという推定量の選択が適切であることが一目瞭然なわけだが、推定量の真の平均二乗誤差が未知の現実世界において、このような理想的な推定量選択を実行することは不可能である。また、推定量の正確さはデータ数 n のみならず、行動数や報酬のノイズなど多くの要因の複雑な組み合わせによって決定する。このいわゆる推定量選択の問題は先述のハイパーパラメータチューニングと類似した問題であるように思えるが、ハイパーパラメータチューニングのための既存の手法は、ある特定の推定量が与えられたときにその推定量のハイパーパラメータを選択するタスクにしか適用できない（という仮定に基づいている）(Su et al.(2020))。観測可能なログデータのみに基づいて適切な推定量を選択する問題に関しては、その必要性に反して十分な研究蓄積がないのが現状である。

４.２ 大規模な問題への対応

オフ方策評価の主たる応用に推薦システムや広告配信があるが、これらの問題は往々にして非常に大規模である。すなわち、データ数 n に加えて行動数（施策の選択数）が数千から数百万といったスケールになることが珍しくない。例えば、大規模な顧客にどの商品を推奨するかといった実務上のプラクティスを考えてみると、問題の規模の大きさが判る。図表４を見ると、データ数 n が大きくなるにつれ IPS 推定量と DR 推定量はより正確になっており、大規模な問題におけるオフ方策評価は簡単に思えるかもしれない。しかし、図表４の結果はあくまで行動数を 20 に固定した上での結果である。では、データ数 n を固定した状態で行動数を増やすと、推定量の精度はどのように変化するだろうか。図表６に、データ数を n=3,000 で固定し、行動数を 10、20、50、100、…、2,000 と変化させたときの DM 推定量・IPS 推定量・DR 推定量の平均二乗誤差の変化を示した。この結果から、特に IPS 推定量と DR 推定量は、行動の数が多くなるにつれ推定精度が著しく悪化することがわかる。具体的には、IPS 推定量と DR 推定量の平均二乗誤差は、行動数が 10 から 1,000 に増えたときに 100 倍以上も悪化している。DM 推定量は行動数の増加に対して比較的頑健であるが、図表４で既に確認した通り、データ数 n が増加したときにその恩恵を受けることができない。大規模な問題に対応するためには、データ数 n が増加するにつれて精度が改善し、かつ行動数の増加に頑健な推定量が望まれる。

図表6　行動数が変化したときの推定量の平均二乗誤差の比較



出所）筆者作成

大規模な問題に対処するためのアプローチはいくつか存在する。なかでもユーザーの行動パターンに仮定を課すことで、IPS 推定量や DR 推定量のバリアンスを軽減するアプローチが主流である（McInerney et al.(2020)）。このアプローチは、情報検索の分野で発展してきたユーザー行動モデリングの知見を、オフ方策評価におけるバリアンス軽減に活かしていると理解することができる。例えば、Kiyohara et al.(2022) では、情報検索の分野ではよく知られたカスケードモデルをユーザー行動に仮定することで、DR 推定量を推薦システムのオフ方策評価に活用する際のバリアンスを軽減している。類似のアイデアは、音楽配信サービス Spotify のプレイリスト推薦におけるオフ方策評価に活用されるなど、応用事例が存在する（McInerney et al.(2020)）。

これらのアプローチは、ウェブ産業におけるオフ方策評価の活用において有効であるが、情報検索の知見が活用できない医療や金融、ロボティクスなどの分野においては効力を発揮できない。また、ウェブ応用においても、ユーザーの行動は多様であることが知られており、単一の仮定で全てのユーザーの行動を説明しようとすると、予期せぬバイアスが生じる可能性もある（Kiyohara et al.(2023)）。したがって、Saito and Joachims（2022）や Saito et al.(2023) は、大規模な問題に対応するためのより汎用的な枠組みとして、行動の特徴表現を自然に扱えるようデータ生成過程を一般化し、ユーザー行動モデリングなど特定の分野の知見に依存せずとも巨大な行動空間に対応できる推定量を提案している。

5．リスクリターン指標に基づくオフ方策評価の有用性評価

前節までに述べた通り、オフ方策評価の研究が進展し、現実の意思決定問題への応用が期待されている。その一方、多数の推定量が開発され、現場ごとに

適切な推定量を選択することの難易度が年々増している。既存の研究論文や実践では、主に前述の平均二乗誤差や、後続の方策選択における順位相関およびリグレットといった典型的な精度指標を用いて推定量の正確さの評価がなされてきた。しかし、これらの従来指標は、実運用時におけるオンライン実験を通じて選択された方策が引き起こす潜在的なリスクを捉えるには不十分である。

この問題を解決するために、Kiyohara et al.(2024) は、SharpeRatio@k というオフ方策評価の推定量に対する新たな精度指標を提案した。本指標は、金融におけるポートフォリオ理論に着想を得ており、推定量によって選ばれた上位 k 個の方策群（いわゆる方策のポートフォリオ）におけるリスクリターンのトレードオフを定量化することで、リスクを考慮に入れた推定量の有用性評価を可能にする。

推定量の有用性を測るための新指標である SharpeRatio@k は、以下の数式で定義される。

$$\mathrm{SharpeRatio@k}\left(\hat{V}\right) = \frac{\mathrm{best@k}(\hat{V}) - \mathrm{V}(\pi_0)}{\mathrm{std@k}(\hat{V})}$$

ここで、

- $\hat{V}(\pi; D)$：オフ方策評価における推定量、
- π_0：データ収集方策、
- best@k $(\hat{V})$：推定量 $\hat{V}(\pi; D)$ によって選ばれた上位 k 個の方策ポートフォリオのうち、真の性能が最大の方策の性能値、
- std@k $(\hat{V})$：それら k 個の方策ポートフォリオ内の性能の標準偏差、
- $V(\pi_0)$：データ収集方策の性能。

この指標は、オフ方策評価の「平均的な正確さ」と「ばらつき（リスク）」の比率を取ることで、より実践的かつ信頼性の高い推定量を特定する。たとえば、推定量が選択する方策ポートフォリオの平均性能が高くとも、性能のばらつきが大きければ SharpeRatio@k は低くなり、逆に一貫性が高く低リスクな方策ポートフォリオを形成できる推定量が高評価を得る。

SharpeRatio@k の実用性を示すために、Kiyohara et al.(2024）では、強化学習の分野でよく用いられるシミュレーション環境を通じたデモンストレーションを行なっている。例えば、MountainCar 環境（谷底をスタート地点とし、車を加速させて斜面を登らせるタスクで、物理的な運動量をうまく調整しないと目標に到達できないよう設計された強化学習における古典的ベンチマーク環境）における実験では、SharpeRatio@k がリスクを考慮しない平均二乗誤差やリグレットとは異なる視点から推定量の良さを順位付けする能力が示された。具体的には、平均的な推定精度が高く見えるが、誤って低性能な方策を上位に選出してしまう推定量は、SharpeRatio@k においては低評価となる。

この挙動は、平均二乗誤差などの従来指標が精度のみを基準にしていたのに対し、SharpeRatio@k がリスクという実運用上の視点を併せ持っていることに起因する。

また、SharpeRatio@k はオンライン評価予算（すなわち、方策ポートフォリオの大きさ）の違いにも柔軟に対応できる特性を持つ。ここでいう予算とは、オンラインで実際に評価可能な方策の数に制約があることを意味し、一般に方策数が多くなるほど、各方策を十分に評価するために必要な予算（試行回数や時間的資源、実装に要する手間など）も増加する傾向がある。オンライン評価予算が小さい場合、選ばれる方策ポートフォリオ内のばらつきが大きくなり、1 つの方策の誤選択が全体の成果に大きな影響を与える。一方、オンライン評価予算が大きくなるにつれ、ポートフォリオのリスクが平均化されやすくなり、より安定した評価が可能となる。この性質により、SharpeRatio@k は限られた予算下でのオフ方策評価においても、安定した評価を可能にする推定量を特定できる。

本指標の導入は、オフ方策評価の推定量選択における新たな評価基準の潮流を生み出し始めている。これまで、オフ方策評価に関する研究は主に意思決定アルゴリズムの性能に対する推定精度のみに着目してきたが、大失敗を避けたいことの多い意思決定最適化の現場では、安全性や信頼性も重要な観点である。SharpeRatio@k は、こうしたリスク評価の観点を織り込むことで、オフ方策評価の実用性向上に大きく寄与することが見込まれている。さらに、SharpeRatio@k は新たな推定量設計の基盤ともなり得る。従来の推定量は、報酬の期待値に対する推定精度向上や重要度重みの定義の工夫に注力してきたが、SharpeRatio@k による評価を意識することで、リスクに頑健な推定量の開発が進む可能性がある。例えば、特にリスクの高い方策の悪影響を抑えるような重み付け手法や、方策間の分散を抑制するようなバイアス補正技術が今後ますます台頭するかもしれない。

SharpeRatio@k が提示するリスクとリターンの最適バランスという視点は、これまでより広範な分野への応用も期待される。医療現場では、新しい治療方針を導入する際、その有用性だけでなく副作用などのリスクも重要である。ファイナンス分野においては、投資方針の最適化に際して、リターンとリスクのバランスを定量的に捉えることが求められる。こうした背景から、SharpeRatio@k のようなリスクとリターンの両面に基づく推定量選択は、多彩な応用先が見込まれる意思決定方策の評価において、重要性が高まっていくことだろう。また、SharpeRatio@k の理論的な枠組みを拡張する試みも今後進む可能性がある。例えば、標準偏差の代わりに条件付き VaR（Conditional Value at Risk; CVaR、ファイナンスの先進的なリスク計量の手法の一つ）など、より厳密なリスク指標を用いた拡張を検討することが可能だろう。これにより、リスクに対してさらに慎重な評価が可能となり、安全性が重視される環境において特に有効な評価が期待できる。

6. 総括

本稿では、意思決定方策のデータに基づく性能評価のための統計技法として注目されるオフ方策評価（Off-Policy Evaluation; OPE）を取り上げ、典型的な定式化や推定量の導入、シミュレーションによる比較を行った。また、従来の精度中心の推定量評価が、実運用の観点からは不十分であるという問題を提起した。特に、医療や金融、教育といった領域では、方策が行う意思決定が生み出すリスクに対する定量的評価が不可欠なのである。

その解決策として最近提案されたのが、SharpeRatio@k である。本指標は、金融のポートフォリオ理論における Sharpe Ratio の概念に着想を得ており、リターンとリスクのバランスを考慮する形で推定量の有用性を定量評価する新たな枠組みを提供するものである。単なる推定精度の高さだけでなく、「安全に高性能な方策ポートフォリオを構成できるか」という実務上重要な視点を加える点が革新的である。理論的には、SharpeRatio@k は上位 k 個の推定方策のうち、最も高い実際性能（リターン）とその集合内の標準偏差（リスク）を用いて定義される。この定義は、意思決定に伴う不確実性を定量化する上で有用であり、オンライン実験のような本番環境への導入前段階において、信頼性の高い推定量の選定を可能にする。実験結果からも、SharpeRatio@k は従来の平均二乗誤差とリグレットいった指標と異なる傾向を示し、リスクの低い推定量を適切に評価できることが示された。

加えて、本稿では SharpeRatio@k の金融分野への応用可能性にも言及した。意思決定方策の集合を金融商品のポートフォリオに見立てることで、リスク調整後リターンを考慮するという視点は、投資戦略の評価と対応付けできる。今後の展望としては、SharpeRatio@k を最大化するような新たな推定量の開発、ならびに環境ごとに動的に指標を調整する自律的評価フレームワークの構築が考えられる。また、リスクの性質に応じた派生指標の導入なども考えられる。総じて、SharpeRatio@k は、データ駆動の意思決定最適化やオフ方策評価による性能評価に安全性という軸を加えるための重要な一歩となることが期待されている。

参考文献

Chandak, Y., Niekum, S., Bruno Castro da Silva, Learned-Miller, E., Brunskill, E., & Philip S Thomas. (2021). Universal off-policy evaluation. *Advances in Neural Information Processing Systems,* 34.

Dudik, M., Erhan, D., Langford, J., & Li, L. (2014). Doubly robust policy evaluation and optimization. *Statistical Science,* 29(4):485- 511.

Farajtabar, M., Chow, Y., & Ghavamzadeh, M. (2018). More robust doubly robust off-policy evaluation. In *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, 1447-1456. PMLR.

Jiang, N., & Li, L.(2016). Doubly robust offpolicy value evaluation for reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning*, 48, 652–661. PMLR.

Kiyohara, H., Kishimoto, R., Kawakami, K., Kobayashi, K., Nakata, K., & Saito,

Y. (2024). Towards Assessing and Benchmarking Risk-Return Tradeoff of Off-Policy Evaluation. In *The Twelfth International Conference on Learning Representations*.

Kiyohara, H., Saito, Y., Matsuhiro, T., Narita, Y., Shimizu, N., & Yamamoto, Y. (2022). Doubly robust off-policy evaluation for ranking policies under the cascade behavior model. *arXiv preprint* arXiv:2202.01562.

Kiyohara, H., Uehara, M., Narita, Y., Shimizu, N., Yamamoto, Y., & Saito, Y. (2023). Off-Policy Evaluation of Ranking Policies under Diverse User Behavior. *arXiv preprint* arXiv:2306.15098.

Levine, S., Kumar, A., Tucker, G., & Fu, J.(2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint* arXiv:2005.01643.

McInerney, J., Brost, B., Chandar, P., Mehrotra, R., & Carterette, B. (2020). Counterfactual Evaluation of Slate Recommendations with Sequential Reward Interactions. *arXiv preprint* arXiv:2007.12986.

Sachdeva, N., Su, Y., & Joachims, T. (2020). Off-policy bandits with deficient support. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 965-975.

Saito, Y., Aihara, S., Matsutani, M., & Narita, Y.(2020). Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation. *arXiv preprint* arXiv:2008.07146.

Saito, Y., & Joachims, T. (2022). Counterfactual learning and evaluation for recommender systems: Foundations, implementations, and recent advances. In *Proceedings of the 15th ACM Conference on Recommender Systems*, 828–830.

Saito, Y., Ren, Q., & Joachims, T. (2023). Off-policy evaluation for large action spaces via conjunct effect modeling. In *Proceedings of the 40th International Conference on Machine learning*. PMLR, 29734-29759.

Su, Y., Srinath, P., & Krishnamurthy, A. (2020). Adaptive estimator selection for off policy evaluation. In *International Conference on Machine Learning*, 9196-9205. PMLR.

Su, Y., Wang, L., Santacatterina, M., & Joachims, T. (2019). Cab: Continuous adaptive blending for policy evaluation and learning. In *International Conference on Machine Learning*, vol. 84, 6005-6014.

Swaminathan, A., & Joachims, T. (2015). Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research,* 16(1), 1731–1755.

Tucker, G., & Lee, J. (2021). Improved estimator selection for off-policy evaluation. Workshop on Reinforcement Learning Theory at *the 38th International Conference on Machine Learning*.

Wang, Y., Agarwal, A., & Dudík, M. (2017). Optimal and adaptive off-policy evaluation in contextual bandits" . In *International Conference on Machine Learning*, 3589-3597. PMLR.

Udagawa, T., Kiyohara, H., Narita, Y., Saito, Y., & Tateno, K. (2022). "Policy-Adaptive Estimator Selection for Off-Policy Evaluation." *arXiv preprint* arXiv:2211.13904.

Uehara, M., Shi, C., & Kallus, N. (2022). A review of off-policy evaluation in reinforcement learning. *arXiv preprint* arXiv:2212.06355.

齋藤優太 (2024)『反実仮想機械学習』技術評論社.

暗号資産に投資するのはどんな人？

― 因果推論手法の適用事例 ―

増島 稔｜ SBI 金融経済研究所 研究主幹
難波 了一｜ SBI 金融経済研究所 主任研究員

増島 稔

SBI 金融経済研究所研究主幹・チーフエコノミスト。東京大学経済学部卒業、ノースウェスタン大学 M.A.、埼玉大学博士（経済学）。1986 年経済企画庁（現内閣府）入庁、外務省審議官（経済局、国際協力局担当）、内閣府政策統括官（経済財政分析担当）、同経済社会総合研究所長などを歴任し 2023 年退官。その間、経済財政政策の企画立案、そのための調査研究に従事。専門はマクロ経済、経済政策、財政・社会保障論。滋賀大学データサイエンス・ＡＩイノベーション研究推進センター特任教授。

難波 了一

SBI 金融経済研究所主任研究員。一橋大学経済学部卒業、早稲田大学大学院経済学研究科博士後期課程単位取得退学。2010 年から内閣府経済社会総合研究所勤務。2016 年から中部圏社会経済研究所勤務、2022 年同研究部長・主席研究員。その間、国内の景気分析や地域経済分析に従事。専門はマクロ経済。

要約

SBI 金融経済研究所でも因果推論手法を用いた分析を行っている。ここでは、因果推論手法のうち操作変数法とランダム化比較試験を用いて、個人の暗号資産への投資行動を分析した事例を紹介する。投資行動に影響を与える要因として金融リテラシーが考えられるが、逆の因果性による同時性バイアスや欠落変数バイアスから生じる内生性の影響を排除するため、操作変数法を用いて推計したところ、金融リテラシーが高い人ほど暗号資産に投資することが分かった。さらに、金融情報の有無が投資行動に違いをもたらすか否かを検証するため、ランダム化比較試験を実施したところ、金融情報へのアクセスが暗号資産を購入する可能性を高めることが明らかになった。因果推論手法は日進月歩であり、新しい手法を積極的に取り入れて分析の質を高めていくことが求められる。

1．はじめに

本号の所報でも取り上げているように、因果推論手法の進歩は目覚ましい。経済分析への適用事例も増えており、SBI 金融経済研究でも積極的に取り入れている。本稿では、因果推論手法を実際に適用した事例として、Nakazono, Masujima, Namba, Takahashi, Tango の 5 名による “Who Invests in Crypto Assets?”（SBI-FERI Working Paper Series において 2025 年掲載）の分析の概要を紹介する。同論文は、当研究所が日本、米国、ドイツ、中国の４か国において個人を対象に実施した「次世代金融アンケート 2024」の調査結果を用いて、個人の暗号資産[1] への投資行動を分析し、以下の３点を明らかにした。

① 他国と比較した日本の家計の暗号資産保有状況
② 暗号資産に投資する人の特性
③ 情報介入が暗号資産保有に与える影響の大きさ

1：「次世代金融アンケート 2024」では、新しいデジタル金融商品（暗号資産、ステーブルコイン、セキュリティトークン、非代替性トークン）について質問しているが、本稿ではこれらの金融商品をまとめて「暗号資産」と呼んでいる。

本稿は因果推論手法に関連する②と③を取り上げる。②の投資行動に影響を与える要因としては、性別、年齢、所得、学歴などに加えて金融リテラシーが考えられる。しかし、金融リテラシーについては、投資経験が金融リテラシーを高めるという逆の因果性もありえる。また、モデルに含まれない、投資行動と金融リテラシーの両方に重要な影響を与える変数が存在する可能性もある。逆の因果性による同時性バイアスや、本来モデルに含めるべき重要な変数がモデルから省略されてしまう欠落変数バイアスは、内生性の主な原因であり、通常の回帰分析の結果に歪みをもたらす[2]。そうした内生性の影響を排除するため、操作変数法（IV; Instrumental Variables Method）を用いて推計したところ、金融リテラシーが高い人ほど暗号資産に投資する傾向があることが分かった。

2：同時性バイアス、欠落変数バイアスおよび操作変数法については、Angrist & Pischke(2009)等を参照。

さらに、③の情報介入つまり金融情報の有無が投資行動に違いをもたらすか否かを検証するため、ランダム化比較試験（RCT；Randomized Controlled Trial）を実施した。因果推論手法としてRCTが倫理的・実施的に可能である状況においては、特定の介入の効果を公平かつ厳密に評価する最も信頼性の高い手法となる[3]。検証の結果、投資家の金融情報へのアクセスが暗号資産を購入する可能性を高めることが明らかになった。

3：ランダム化比較試験は「ゴールドスタンダード」とされる。詳細はAthey & Imbens(2017)等を参照。

本稿の構成は以下の通りである。第2節では、因果推論の重要性、操作変数法とランダム化比較試験の考え方を簡潔に説明する。第3節では、分析に用いたデータを解説する。第4節では操作変数法の、第5節ではランダム化比較試験の、それぞれ適用結果を解説する。最後に因果推論手法の有用性について論じる。

2. 因果推論とは

因果関係とは「AがBを引き起こす」という関係である。しかし、その関係は本当に正しいのだろうか。実際には、BがAを引き起こしているのではないか、A以外の別の要因がAにもBにも同時に影響を与えているのではないか、と疑問に思うことも多い。経済学や経営学において、この因果関係を正確に特定することは、経済政策やビジネス戦略を考える上で極めて重要である[4]。本節では、因果関係を科学的に明らかにするための代表的な手法として、ランダム化比較試験と操作変数法の考え方を説明する[5]。

4：例えば、最低賃金の引き上げという経済政策が雇用に与える影響について考えてみる。仮に、ある地域で観測されたデータにおいて、最低賃金の引き上げと雇用の改善が同時に見られたとしよう。しかし、それはたまたま、当該地域の景気が改善しているタイミングで政策が実行されただけなのかもしれない。真の影響は雇用をむしろ悪化させる可能性もあり、表面的な関係から誤って特定された因果関係は政策判断を歪めることにもなりかねない。

5：因果推論の手法としては、この他に差の差分析（Difference-in-Differences）、マッチング法（Matching Methods）などがある。因果推論の概観についてはImbens(2024)等を参照。

2.1 ランダム化比較試験

ランダム化比較試験は最も強力な因果推論の手法である。関心のある介入（治療法や政策など）の効果を測定するために用いられる。ランダム化比較試験では、研究の対象となる人々を完全にランダムに二つのグループに分ける。一つは処置群（Treatment Group）で、効果を検証したい介入を受けるグループである。もう一つは対照群（Control Group）で、介入を受けないグループである。処置群と対照群をランダムに分けることで、両グループは、平均的に見て、介入が始まる時点ではあらゆる点で同質であると期待できる。つ

まり、年齢、性別、収入、学歴、健康状態など、介入の効果に影響を与えうる様々な要因が、両グループで均等になる。その結果、介入後に両グループ間で何らかの違いが見られた場合には、その違いは介入によって引き起こされたものであると、高い確信を持って結論づけることができる。介入以外の要因による影響を排除し、純粋な介入の効果を測定することが可能になるのである。ランダム化比較試験を実施する上で、介入を受けるグループと受けないグループをランダムに分けることが最も重要なポイントとなる所以である。逆に、もし、介入を受けるグループを意図的に選んでしまうと、そのグループの人たちが元々持っている特性（例えば、より裕福である、教育水準が高い、健康意識が高いなど）が、介入の効果と混同されてしまう可能性があるため、介入の効果を正確に計測することはできなくなってしまう。こうしたバイアスは「選択バイアス」と呼ばれている。

2.2 操作変数法

2.1 で説明したランダム化比較試験は、因果関係を特定する上で非常に強力な手法である。しかし、アンケート調査であれば比較的容易に実施できるが、治療法や経済政策などの現実的な課題について実際に実施しようとすると、倫理的な問題があることや莫大な費用がかかることなどから、常に実行できるわけではない。例えば、「大学教育が賃金に与える影響」をランダム化比較試験で検証しようとしても、ランダムに選ばれた一部の人に大学への進学を強制したり、禁止したりすることは現実的ではない。そこで用いられるのが操作変数法である[6]。操作変数法は、実験を行わずに得られた既存の観察データから因果関係を推定する際に、介入と関連するが、結果変数には直接影響を与えず、かつ介入以外の要因とは相関しない操作変数を用いることで、選択バイアスを取り除くことを目的としている。操作変数は、以下の 3 つの条件を満たす必要がある。第一に、関連性（Relevance）である。操作変数（上記の大学教育と賃金の例では親の学歴など）は、検証したい介入（同じく大学進学）と相関している必要がある。第二に、外生性（Exogeneity）である。操作変数は、結果変数（同じく賃金）に直接影響を与えてはならない。第三に、除外制約（Exclusion Restriction）である。操作変数は、介入以外の要因を通して結果変数に影響を与えることはなく、介入を通してのみ結果変数に影響を与える。操作変数法は、ランダム化比較試験が実施できない状況下で因果関係を推定するための重要なツールである。

6：ランダム化比較試験が理想的な実験デザインによる手法であるのに対して、操作変数法は準実験的手法と呼ばれる。

このように、ランダム化比較試験と操作変数法は、どちらも因果関係を明らかにすることを目指す手法だが、アプローチが異なっている。ランダム化比較試験は理想的な実験環境を作り出すことでバイアスを排除するのに対し、操作変数法は観察されるデータの中から偶然の介入に近い状況を探し出すことで因果関係を推定する手法である。

3．データ：次世代金融アンケート調査

本節では、Nakazono et al.(2025) が分析に用いている「次世代金融アンケート 2024」（ＳＢＩ金融経済研究所 (2024b)）について解説する。このアンケート調査はＳＢＩ金融経済研究所が 2022 年以降毎年実施しており[7]、暗号資産やステーブルコイン、セキュリティトークン、非代替性トークンといった新しいデジタル金融商品に焦点を定め、株や債券、FX といった従来のリスク性金融商品と比較しながら、個人の資産選択行動やそれに影響を与える要因を明らかにすることを目的としている。

３回目となる「次世代金融アンケート 2024」は、日本・アメリカ・ドイツ・中国の４か国の 20 歳以上の個人を対象として、2024 年８月末から 10 月初にかけて実施された。サンプル数は、日本が１万人、他の３か国が各４千人、合計２万２千人である。質問内容は、①対象者の属性（性別・年齢・年収等）、②リスク性金融商品についての認知度、投資経験、認識、過去の投資パフォーマンスなど、③新しいデジタル金融商品についての認知度、投資経験、認識、保有額、最近の投資動向、投資目的など、④金融リテラシーやリスク回避度、インフレ率・成長率の見通しとその不確実性、その他の金融資産選択に影響を与える可能性のある要因の４群である。

新しいデジタル金融商品を調査したものとしてはこれまでに例のない規模であり、新しいデジタル金融商品と従来のリスク性金融商品を同時に調査としている点でも、他に類を見ないアンケート調査となっている。

7：2022 年、2023 年 は「 次世代金融に関する一般消費者の関心や利用度に関するアンケート調査」として実施した（ＳＢＩ金融経済研究所 (2022, 2024a)）

4．どんな人が暗号資産を保有しているか？：操作変数法を用いた分析例

以上の「次世代金融アンケート 2024」の調査結果から、どんな人が暗号資産を保有していると言えるだろうか[8]。Nakazono et al.(2025) では、暗号資産を保有している人の特徴を明らかにするため、以下の（１）式を推計している[9]。

$$
\begin{aligned}
&\textit{暗号資産の保有状況}_i \\
&= \beta_0 + \beta_1 \textit{男性ダミー}_i + \beta_2 \textit{年齢}_i + \beta_3 \textit{所得}_i + \beta_4 \textit{学歴ダミー}_i \\
&+ \beta_5 \textit{リスク回避度}_i + \beta_6 \textit{期待インフレ率}_i + \beta_7 \textit{成長期待の不確実性}_i \\
&+ \beta_8 \textit{インフレ期待の不確実性}_i + \sum \gamma_k D_{k,i}^{\text{国}} + \varepsilon_i \qquad (1)
\end{aligned}
$$

添え字の i は個人、k は国を表す。従属変数（暗号資産の保有状況）については２つ検討した。一つめの従属変数としては、暗号資産を保有している人の特徴を明らかにするため、暗号資産を保有している場合は 1、そうでない場合は 0 となるダミー変数（$D^{Holding}$）を用いた[10]。二つめの従属変数としては、金融資産として暗号資産を多く持っている人の特徴を明らかにするため、ポートフォリオに占める暗号資産の割合（$CryptoShare$）を用いた[11]。

調査対象者の属性としては、性別（男性＝ 1）、年齢、所得、学歴（大卒以

8：山沖・副島 (2025) は、2022 年のアンケート調査結果を用いて、日米独英中韓の６か国について暗号資産や株式の投資決定要因を分析している。

9：ここでの推計は標準的な最小二乗法による。なお、Nakazono et al.(2025) は現時点で執筆段階にあるため、本稿で紹介するのは暫定的な推計結果であることに留意されたい。

10：具体的には、暗号資産、ステーブルコイン、セキュリティトークン、非代替性トークンのいずれかを保有している場合に 1 としている。

11：アンケートには「あなたが保有している金融資産の割合をお答えください。」という質問がある。回答者は、合計で 100% になるように「現預金」などの金融資産の割合をそれぞれ回答する。ここでは、その中の「新しいデジタル金融商品」の保有割合を用いている。

上＝1)、リスク回避度を考慮した。また、本アンケート調査では、調査対象者それぞれの持つ期待インフレ率と期待成長率及びその不確実性を調査しており、それらを説明変数に加えた。さらに国別の違いを考慮して、国別ダミー（$D^{国}$）を追加した。

推計結果（図表１）を見ると、男性、若年層、高所得層、低学歴層、リスク回避度の低い（リスク選好度の高い）層が暗号資産に投資する傾向が強いことが分かる。また、インフレ期待が高いほど、インフレ期待と成長期待の不確実性が高いほど、暗号資産に投資する傾向が強いことも分かる。国別の違いを見ると、日本は他国と比較して暗号資産に投資する人が少なく、ポートフォリオに占める暗号資産のシェアも低いことが分かる。

図表１　どんな人が暗号資産を保有しているか?

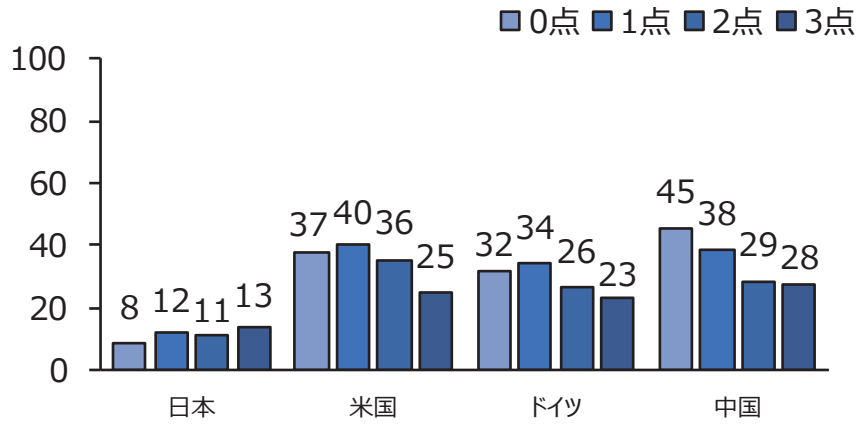
	(1) $D^{Holding}$	(2) *CryptoShare*
男性ダミー	0.057***	0.943***
年齢	-0.035***	-0.518***
所得	0.013***	0.146***
高学歴ダミー	-0.016***	-0.332***
リスク回避度	-0.019***	-0.248***
期待インフレ率	0.051***	1.128**
成長期待の不確実性	0.001***	0.009**
インフレ期待の不確実性	0.043***	0.798**
国ダミー		
米国	0.114***	2.458***
ドイツ	0.117***	2.067***
中国	0.024**	-0.054
観測数	16,428	19,004

出所）Nakazono et al.(2025)

注）従属変数は（1）$D^{Holding}$と（2）*CryptoShare*である。$D^{Holding}$ は、回答者が暗号資産を保有している場合は1となり、そうでない場合は0となるダミー変数である。*CryptoShare*は、各回答者のポートフォリオにおける暗号資産の割合である。学歴はダミー変数であり、回答者が大卒以上であれば1、そうでなければ0を取る。 期待インフレ率は、3つのシナリオ（高、中、低）をそれぞれの主観的確率で加重平均したものである。インフレ期待と成長期待の不確実性は、それぞれ高シナリオ値と低シナリオ値の差の絶対値として計算される予測の不確実性を表す。*** p< 0.01, ** p< 0.05, * p< 0.1。

次に、人々が暗号資産への投資を決定する上では、金融リテラシーも重要な要因となりうる。ここで、金融リテラシーと暗号資産への投資経験の関係を確認すると（図表２）、前者と後者の間には正の相関は見られず、むしろ逆相関が見られる。つまり、金融リテラシーが低い人ほど、暗号資産への投資経験が多くなるという傾向が見てとれる。金融リテラシーと暗号資産の保有の関係については、前者が後者に影響を与える可能性がある一方で、暗号資産への投資経験を積むことで金融リテラシーが高まるという逆の因果性が考えられる。また、投資行動と金融リテラシーの両方に重要な影響を与える何らかの変数が存在している可能性もあり、図表２に見られる表面的な関係性や（１）式のような通常の回帰分析から因果関係を特定することは困難である。

図表2　金融リテラシーと暗号資産への投資経験



出所）SBI金融経済研究所(2024b)

注）金融リテラシーは、単利計算、複利計算、金融資産の実質価値、分散投資の効果に関する４つの質問の正答率から0～3点で点数化した。縦軸は、それぞれの点数における回答者の暗号資産への投資経験割合を示している。

Nakazono et al.(2025) では、こうした逆の因果性による同時性バイアスや欠落変数バイアスから生じる内生性の問題に対処するため、以下の（２）式を操作変数を用いて国別に推計し、金融リテラシーが暗号資産の保有に与える影響を検証している。

$$D^{Holding}{}_i = \beta_0 + \beta_1 \text{金融リテラシー}_i + \varepsilon_i \qquad (2)$$

添え字の i は個人を表す。ここで、$D^{Holding}$ は（１）式と同じく暗号資産の保有の有無を識別するダミー変数である。本アンケート調査では、金融リテラシーを計測するため、単利計算、複利計算、金融資産の実質価値、分散投資の効果に関する質問をしており、金融リテラシーを示す変数として、それらの質問への正答数から算出した指数を用いた。操作変数としては２つの変数を用いた。一つめは、両親または兄弟に投資経験があるかどうかを示すダミー変数であり、二つめは、回答者が金融教育を受けているかどうかを示すダミー変数である。これらの操作変数を用い、内生性を考慮して推計した結果を見ると（図表３）、金融リテラシーが高いほど、すべての国で暗号資産への投資確率が高まることが示唆されている。ここには示していないが、結果の頑健性を検証するため、（１）式同様、回答者のポートフォリオにおける暗号資産の割合（*CryptoShare*）を上記と同じ２つの操作変数法を用いて金融リテラシーに回帰しても、金融リテラシーが高いほど、ポートフォリオにおける暗号資産のシェアが高まることが確認できた。

図表3　金融リテラシーは暗号資産の購入に影響を与えるか？

	(1) 日本	(2) 米国	(3) ドイツ	(4) 中国
金融リテラシー（Ⅳ）	0.219*** (0.019)	0.708*** (0.096)	1.177*** (0.161)	0.874*** (0.108)
定数	0.049*** (0.003)	0.241*** (0.013)	0.229*** (0.021)	0.248*** (0.017)
観測数	9,696	3,305	3,294	2,981

出所）Nakazono et al.(2025)

注）従属変数は$D^{Holding}$である。$D^{Holding}$ は、回答者が暗号資産を保有している場合は1となり、そうでない場合は0となるダミー変数である。金融リテラシーは、単利計算、複利計算、金融資産の実質価値、分散投資の効果に関する質問への正答数から算出した指数である。内生性に対処するため、（1）両親または兄弟姉妹に投資経験があるかどうか、（2）回答者が金融教育を受けているかどうか、の2つの指標を操作変数として用いた。括弧内はロバスト標準誤差。*** p< 0.01, ** p< 0.05, * p< 0.1。

5．情報提供は暗号資産の購入につながるか？：ランダム化比較試験を用いた分析例

本アンケート調査によれば、日本において暗号資産を保有している人は回答者の6%程度に過ぎない。もし、暗号資産に関する情報を持たないために保有していないのであれば、情報を提供することが暗号資産を保有するかどうか、あるいは保有を増やすかどうかの意思決定に影響を与える可能性がある。この仮説を検証し、期待リターンと暗号資産保有との因果関係を明らかにするために、ランダム化比較試験を実施した[12]。上述した通り、ランダム化比較試験は因果関係を特定する上で非常に強力な手法であり、倫理的・実施的に可能である状況においては、特定の介入の効果を公平かつ厳密に評価する最も信頼性の高い手法となる。調査対象者を何も情報を与えない対照群と、ビットコインに関する情報を与える処置群のいずれかに無作為に割りつけた。対照群に割りつけられた人には何の情報も提供せず、一方、処置群に割りつけられた人には、ビットコインの収益率に関して、「ビットコインの価格は過去5年で6倍以上、過去10年で100倍以上になった」という情報を提供した上で、1年後の望ましいポートフォリオを聞いた。

当初暗号資産を保有していなかった世帯が、情報提供の結果、暗号資産を保有するようになるかどうかを検証するために、以下の（3）式を推計した。

$$Y_i = \beta_0 + \beta_1 D^T_i + \sum \gamma_k D^{国}_{k,i} + \varepsilon_i \qquad (3)$$

添え字のiは個人、kは国を表す。ここで、Yは、情報提供前に暗号資産を保有しておらず、情報提供後に1年後の暗号資産の望ましい保有割合が1%を超えた場合に1となり、そうでない場合に0となるダミー変数である。すなわち、情報提供によって暗号資産を保有する意欲を明確に高めた人を識別している。D^Tは、処置群であることを示すダミー変数である。$D^{国}$は国ダミーである。

12：本ランダム化比較試験は、SBI金融経済研究所の倫理審査委員会により承認されており、AER RCT Registry（#AEARCTR-0014458）に登録されている。

推定結果（図表４）を見ると、４か国すべてのサンプルを用いて推計した場合の平均処置効果は有意にプラスであり、3.8％であった。情報提供が暗号資産の購入確率に影響を与えること、つまり、暗号資産を購入する意欲を高めることが示唆されている。ここには示していないが、結果の頑健性をチェックするため、１年後の望ましい暗号資産の保有割合を従属変数として、処置ダミー（D^T）と現在の暗号資産の保有割合を説明変数として回帰した式も推計した。４か国すべてのサンプルを用いて推計した場合の平均処置効果は有意にプラスであり、0.58％となった。情報提供をすることによって暗号資産を保有する意欲が高まることを確認した。

図表４　情報提供は暗号資産の購入につながるか?

	4か国 (1)	日本 (2)	米国 (3)	ドイツ (4)	中国 (5)
D^T（処置ダミー）	0.038***	0.035***	0.025	0.096***	0.003
	(0.008)	(0.009)	(0.024)	(0.023)	(0.023)
国ダミー					
米国	0.221***				
	(0.013)				
ドイツ	0.223***				
	(0.012)				
中国	0.217***				
	(0.012)				
定数	0.104***	0.105***	0.331***	0.298***	0.337***
	(0.006)	(0.006)	(0.017)	(0.016)	(0.016)
観測数	9,858	4,815	1,614	1,679	1,750

出所）Nakazono et al.(2025)

注）従属変数は$D^{Holding}$である。$D^{Holding}$ は、回答者が暗号資産を保有している場合は１となり、そうでない場合は0となるダミー変数である。D^Tは、処置群であることを示すダミー変数であり、回答者がビットコインの実際のリターンに関する情報を提供された場合に1となる。いくつかのコントロール変数の係数値は報告していない。括弧内はロバスト標準誤差。*** p< 0.01、 ** p< 0.05、 * p< 0.1。

６．おわりに

本稿では、SBI 金融経済研究所が因果推論手法を活用している事例として、因果推論手法のうち操作変数法とランダム化比較試験を用いて、個人の暗号資産への投資行動を分析した事例を紹介した。因果推論手法の適用は分析の精度を高め、新たな知見を生む。紹介した分析からは、第一に、操作変数法を使うことによって、金融リテラシーと暗号資産投資の間の内生性を考慮したところ、金融リテラシーが高い人ほど暗号資産に投資することが分かった。第二に、ランダム化比較試験を実施したところ、金融情報へのアクセスの改善が暗号資産の購入可能性と保有比率を高めることが明らかになった。特に前者は金融リテラシーと暗号資産投資の関係について、表面的な関係性や通常の回帰分析からは特定できない因果関係を得た。

これらの事例は、因果推論手法の有用性を示すものだが、同手法には限界もある。ランダム化比較試験は、すでに述べたように、倫理面やコスト面の問題があり、実施のハードルは高い。また、実施できたとしても、結果が出るまでに時間がかかる場合も多く、介入以外の影響が混在したり、脱落者が出たりして因果関係の特定が難しい場合もあろう。操作変数法についても、条件を満たす適切な操作変数を見つけることは実際には難しく、その妥当性には常に議論が伴う。

もっとも、因果推論手法は日進月歩だ。その限界に留意は必要だとしても、新しい手法を積極的に取り入れて分析の質を高めていくことが求められる。SBI 金融経済研究所としてもその努力を続けていく。

参考文献

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist' s Companion*. Princeton University Press.

Athey, S., & Imbens, G.W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives, 31*(2), 3–32.

Imbens, G.W. (2024). Causal Inference in the Social Sciences. *Annual Review of Statistics and Its Application, 11*(1), 123–152.

Nakazono, Y., Masujima, M., Namba, R., Takahashi J., & Tango K.（2025）. *Who Holds Crypto Assets? Demographics, Financial Sophistication, and the Role of Information*. SBI-FERI Working Paper Series, SBI-FERI WP E-1.

SBI 金融経済研究所 (2022)「次世代金融に関する一般消費者の関心や利用度に関するアンケート調査」2022 年 12 月 27 日 https://sbiferi.co.jp/questionnaire/question20221227.html

SBI 金融経済研究所 (2024a)「次世代金融に関する一般消費者の関心や利用度に関するアンケート調査、第 2 回」2024 年 11 月 11 日 https://sbiferi.co.jp/questionnaire/question20241111.html

SBI 金融経済研究所 (2024b)「次世代金融アンケート 2024」2024 年 12 月 24 日 https://sbiferi.co.jp/questionnaire/question20241224.html

山沖義和・副島豊（2025）「暗号資産・株式等の投資決定要因：6 ヶ国検証 ― SBI 金融経済研究所アンケート調査（第 2 回）の個票分析 ―」SBI-FERI Working Paper Series, SBI-FERI WP J-1.

次世代金融インフラのあるべき姿

（2025年3月31日）

次世代金融インフラの構築を考える研究会

「次世代金融インフラのあるべき姿の例示」の公表

問題意識及び第1次提言までの検討経緯

金融 API やブロックチェーン技術、ビッグデータの活用に代表される情報技術の革新によるデジタル化社会の進展に伴い、新しい決済・送金手段、暗号資産などのデジタル金融資産、分散型金融サービス (DeFi) の登場など、金融サービスの提供主体・手段等に変化が生じており、今やデータが収益の源泉となり、情報生産機能の高度化が金融機関の経営課題となっている。

特に、銀行・証券会社・保険会社などの仲介業者を通じた金融サービスの提供、中央銀行と民間銀行による 2 階層型の決済制度などを前提とした現行の金融インフラ（法規制、IT システム、会計ルール、ガバナンス、リスクマネジメント、国際協調など）は、急激な環境変化に対して十分に適応できているとは言い難く、既存の枠組みの中で対処療法的に変更を重ねることの限界も露呈し始めている。新しい金融サービスには新しい金融インフラが必要であり、まずはその将来像を描くことが求められている。

そこで、SBI 金融経済研究所は、このような問題意識の下、将来の金融インフラとして望ましい姿を検討し、これを提言としてまとめることによって、社会的な議論を惹起することを目指して、デジタル金融資産に精通した有識者から成る「次世代金融インフラの構築を考える研究会」を立ち上げ、2023 年 12 月 25 日に第 1 回会合を開催した。なお、金融庁、日本銀行 (金融研究所) からもオブザーバーとして参加している。

本研究会では、「デジタル金融の制度的な枠組みの再構築」をテーマに議論を重ねた結果、国内外の金融サービス利用者、当局を含む金融システムの構成主体に対して第 1 次提言として、昨年（2024 年）7 月 5 日に「新しい金融インフラの構築を考えるに当たっての指針」を取りまとめ、公表した。

同指針では、金融システムの転換期に適応できる次世代金融インフラを構築し、国内外の利用者から選ばれる金融システム・金融センターも目指すという目的を設定した上で、その達成のための視点等として、(A) 新しい金融インフラの構築に当たって必要となる視点、(B) それを考える際の留意事項、(C) その進め方の合計 19 項目に加えて、(D) 当面の課題・懸念事項として 6 項目を挙げている。また、その際、階層（レイヤー）構造を展開した金融インフラの

基盤の上に、従来の経済主体別思考から金融機能別思考に転換した金融サービスのコンポーネントを有機的に機能させるとともに、従前の仕組みが内包する課題を解決するため新旧２つの金融インフラを並走させる方策が有効であると指摘している。

第2次提言の検討経緯

2024 年 7 月下旬からは、国内・海外の事例も踏まえて、「指針を踏まえた次世代金融インフラのあるべき姿」を第 2 次提言としてまとめるべく、検討を重ねてきた。

しかしながら、第 1 次提言の「次世代金融インフラの構築を考えるに当たっての指針」では全部で 25 項目に渡って詳細かつ広範な視点等を示したものの、将来、実現する次世代金融インフラは各国の金融制度の成り立ち、国民性、時代等のほか、特に金融当局による監督のあり方により異なるものであり、一義的に決まるものではない。

例えば、（１）金融仲介業者が次世代金融インフラの中核を担うケースが考えられる一方、（２）ブロックチェーン技術・DeFi 等のデジタル技術を活用して金融仲介サービスが全面的に自動化されるケースも考えられる。

そこで、本研究会では、まず鍵となるコンセプトを次の４つに整理した。①金融機能別思考への転換と金融サービス・提供主体の組換え、②基盤レイヤーの再構築（リビルド）、③非金融分野も包摂した形での横展開、④金融・非金融領域を跨ぐデータの活用による利用者ニーズ等の可視化、それに基づくサービスの自動化・高度化を通じた高付加価値化である。

その上で、上述した２つのケース、すなわち、（１）金融仲介業者が中核を担うケース、（２）デジタル技術によって構築された決済・情報連携システム等が金融仲介機能を代替するケースについて、次世代金融システムの将来像を例示するとともに、次世代金融インフラに求められる構造・機能（例えば、レイヤー構造や各種の金融サービスなど）についても言及することとし、別紙の通り第二次提言として「次世代金融インフラの例示」をとりまとめた。

今回の提言は、あくまでデジタル技術を活用した金融サービスの提供を前提とした金融インフラの姿として想定される２つのケースを例示したものではあるものの、今後、日本が次世代金融インフラを整備するに当たって参考となることを期待する。

今後の進め方

来年度 (2025 年度) には、中期的に取り組むべき具体的なテーマとして３つを選び、それぞれのテーマごとに分科会を設けて検討し、順次、提言をまとめることとしている。

（以上）

「次世代金融インフラのあるべき姿」の例示

目　次

第1部　本提言とりまとめの視点

本研究会では、2024 年 7 月 5 日に公表した「新しい金融インフラの構築を考えるに当たっての指針」に基づき、長期的な視点に立ってデジタル技術を活用した金融サービスの提供を前提とした次世代金融インフラのあるべき姿について検討を続けた。討議では多様な考え方が示され、次世代金融インフラの実現形態には様々な可能性があるという結論に至った。しかし、共通理解として浮かび上がってきたコンセプトを得ることができたため、将来像の一例として取りまとめることとした。

本研究会では「金融インフラ」を「社会を支えるインフラとしての金融サービス全体」と広く捉えており、その中には与信や決済という金融サービスとこれらを支える狭義の金融インフラがレイヤー構造として存在している。用語の混乱を避けるため、ここでは、「狭義の金融インフラ」を「基盤レイヤー」と呼ぶ。また、後述する通り、「金融機能等を組み合わせることによってモジュール化した金融サービス」を「モジュール」と呼ぶ。

1．歴史的経緯と現状認識

過去から現在に至る様々な金融サービスやその基盤レイヤーは、時代環境に即した企業体や業態として誕生し、これに合わせて法規制が整備されてきた。経済、テクノロジー、信認を含む社会統治や文化慣習によって形成される時代環境を背景に、経済合理性など何らかの必然や偶然のもとで、新しい金融サービスが創造され、金融のビジネスモデル・基盤インフラが形成されてきた。例えば、江戸期の三貨制度による分散・分権型金融システムから明治期の通貨統一等による中央集権化、日本銀行設立と民間銀行による二階層型の預金・決済システムへの転換、資金・証券決済システムのデジタル化、派生金融商品の登場、電子マネーの普及と債務性マネーの多様化、分散・分権型金融の再勃興など、時代環境の変化に合わせて歴史的変遷を遂げてきたし、現在も、未来も続いていく。

このような歴史的経緯を俯瞰的に見直すことは、現行の金融システムの成り立ちやその合理性を確認するとともに、環境変化への適合が遅れている箇所を検証し、改善や変革の方向や方法を検討する手がかりにつながる[1]。

こうした視点から我が国の金融制度を金融サービス側に着目して概観すると、金融産業構造の大規模な改革は行われてこなかった（その必要性も高くはなかった）と言えよう。その結果、銀行・証券会社・保険会社等という業態に依拠した思考にとらわれ、業法を中心とした法規制が適用されてきた。金融サービスの提供に当たってもこの思考から脱却できず、新たな金融サービスやその担い手の登場や既存の金融サービスの進化に対応することが難しくなっている。

同様に、決済システム、市場取引基盤、情報通信インフラ、IT システムなど、金融サービスを支えている広範な基盤レイヤーについても、デジタル社会の到来にもかかわらず、産業構造の大きな変革に対応できていない。これらの

1：例えば、当研究所所報の鎮目 (2023.8)、副島 (2024.3)、齊藤達哉 (2024.3)、レポートの高槻 (2024.7・8)、松尾 (2024.9)、脇 (2024.7) を参照。

基盤レイヤーについても、環境の変化に対応して金融サービスとともに将来像を検討する必要がある[2]。

その一方で、分散型台帳技術の発展は、まったく異なる基盤レイヤーの上に新しい金融サービスを展開するという可能性を切り開き、その取組みが世界規模で活発に行われている。中には、ステーブルコインやセキュリティトークンのように市場規模や法的・システム的な安定性を高めつつある領域も登場してきている。また、金融機関が金融・非金融資産や預金のトークナイゼーションを推し進める基盤レイヤーを開発し、これを活用した金融・非金融サービスを展望する事例も増えている。

2．鍵となる４つのコンセプト

上述した現状認識からは、次世代金融インフラの将来像を描く上で重要となる４つのポイントが読み取れる。

一つ目は、経済主体別思考から金融機能別思考への視点の転換であり、これに伴う金融サービスや提供主体の組換えである。金融サービスを機能別に分解（アンバンドリング）した上で、改めて利用者視点に立って個人・企業・公的部門などの利用者から求められる、あるいは、まだ発見すらされていない金融サービスの創造に向けて、金融機能や広義の情報処理機能を組み合わせる（リバンドリング）ことが求められている。こうした金融機能や情報処理機能を組み合わせ、「新結合（組合せによるイノベーション）」が発揮されることによってモジュール化した新たな金融サービスが産み出される。

二つ目は、金融サービス産業をレイヤー構造として捉えた上で、アプリケーション・レイヤーとしての金融サービスの組換えを促す基盤レイヤーの再構築（リビルド）が求められているという点である[3]。

こうした基盤レイヤーにも技術革新による再構築の波が押し寄せており、密結合・一体化していた勘定系システムの疎結合化、クラウドサービスへのシフト、API 公開によるオープンバンキング化、専用回線からインターネットへの通信インフラの拡張、エコシステムビジネス展開に向けた顧客情報台帳の企業内部でのオープン化などが既存の事例である。分散型技術の登場と普及により、金融サービスのアプリケーション・レイヤーとの関係性の見直しを含めた基盤レイヤーの再構築の事例も登場している。

例えば IMF が最近公表した論文ではデジタル金融資産のプラットフォームの相互運用性を促進するために「ASAP モデル」が提唱されている[4]。従来の金融システムでは、証券や預金の台帳は金融資産と不可分であり、台帳に記帳された数字が金融資産そのものであった。トークン化技術の発展は両者を分離することを可能とし、転々流通させることができる金融資産を産み出した[5]。MAS（シンガポール金融管理局）でも、同様の目的を有して、新しい形の金融サービスのレイヤー構造として「Global Layer One（GL1）」を示している。オープンで相互運用可能な分散型技術に基づく基盤レイヤー（GL1）を確立し、その上のアプリケーション・レイヤーにおいては、金融資産とそれに関わるサービスおよびサービスへのアクセス方法が更なる内部レイヤー構造

2：例えば、欧米ではデジタル技術の進歩やそれに伴う法規制等の変化に合わせて、ECN（電子商取引ネットワーク）や ATS（代替取引システム）、MTF（多角的取引システム）が登場し、株式やデリバティブ取引市場の産業構造の大規模な変容と、基盤インフラのオーナーシップを含めた金融市場のグローバル化が進展した。これを契機に、既存の証券清算機関や CSD（集中保管振替機関）の垂直サイロ構造が変容していったが、日本はデリバティブ等の OTC 市場の発展や、取引所の取引システムや清算決済システムの高度化に止まり、産業構造の変化や非金融商品を含めた新しい市場が勃興するには至らなかった。また、メンバーシップによるガバナンスが構造変化を起こりにくくしている一面もあった。後述するように、モノリシックに（一体として）構築されてきた基盤レイヤーにおいてもアンバンドル化や標準化が進展しており、こうした動きへの対応も必要となっている。

3：金融サービス産業がレイヤー構造となっていることは、日本銀行の「決済システムレポート 2024.9」の「図表 1 わが国の決済システムの鳥瞰図」に示されている市場取引から決済までの流れが象徴的である。

4：ASAP モデルでは、アプリケーション、サービス、アセット、プラットフォームという４つのレイヤー構造が想定されている。

5：証券のトークン化では、基本的な所有権は従来システムで管理し、売買などに関わる一部の権利・機能をトークン化で実現するハイブリッド型がその事例である。法的に認められた権利の直接的なデジタル表現としてトークンが機能している。従来の台帳を分散型台帳で完全代替する場合は、資産と台帳プラットフォームは不可分となる。なお、トークン化預金やステーブルコインでは、裏付け資産が従来台帳で管理され、新たに発行されたトークンは分散型台帳で管理されることになる。これも預金金融資産と台帳の分離の一形態である。

をもって実装されるというコンセプトである。これらの提案は、金融サービスのモジュール化をイノベーティブに促進していくためには、基盤レイヤーのモジュール化[6]とその再構築も必要であることを示唆している。

三つ目は、金融サービスと基盤レイヤーにおける再結合（リバンドル）・再構築（リビルド）[7]が金融分野だけにとどまらず、非金融分野も包摂した形で横断的に展開していく点である。金融サービスと非金融サービスの融合は、情報処理における範囲の経済性を活用する形ですでに進展している。FinTech企業がAPIを通じて金融機関の情報システムにアクセスし、顧客からのリクエストを遂行するビジネスモデル、金融機関がEmbedded Financeとして金融サービスやこれに必要な基盤インフラを非金融業に黒子役として提供するビジネスモデルなどが挙げられる。

この場合、金融分野・非金融分野のどちら側からも包摂は起こり得る。機能別にモジュール化された金融サービスと非金融サービスをどのように組み合わせて新しいビジネス領域を拡張していくかは、いずれの産業領域の企業や経済主体（公的部門も含む）でも手掛けることが可能である。アイデアを持つものが容易にビジネス実装できるよう、金融サービスのモジュール化とこれを繋ぐ基盤レイヤーが、非金融分野にも開かれていること、明確な法規制の下、運用が容易な形で安価に提供されること[8]、相互運用性を確保するための標準化を推し進めることが重要となる。

四つ目は、金融・非金融領域を跨ぐデータの活用による利用者ニーズ等の可視化を図り、分断化されているサービス生産プロセスの自動化・高度化を通じた高付加価値化をもたらすことである。金融サービスの領域内で閉じるケースもあれば、非金融領域との連携によって達成されるケースもあり得る[9]。

利用者ニーズの発掘においては、データ基盤等の整備やITシステムへの実装が重要となる。Eコマースの「買う / 売るを簡単にする」取組みや、「買う」の過程で生成された検索履歴情報などを推薦・提案システムに活用し、追加的な需要喚起を図る方法がよく知られている。金融分野においても、リテールを中心にビジネスモデルの見直しが進んでおり、ホールセールにも広がろうとしている。その過程では金融サービスと非金融サービスの境界線を設けることの意味もなくなってきている。

3．リバンドリングを促す要因

「次世代金融インフラのあるべき姿」を実現するためには、金融サービスや基盤インフラのモジュール化とそのリバンドリング・リビルドを促す要因（ドライビングフォース）を考察することが重要である。実現可能性がある姿を描くには、関与する様々な経済主体のインセンティブが互いに成立し合う必要がある。ここでは、リバンドリング・リビルドが価値を産みだす源泉となり得るドライビングフォースとして次の７つに整理した。なお、具体的な内容については、文末のBoxに示した。

① 金融サービスの分解と組合せによる金融機能向上
② 新技術による低コスト・高速化を通じた効率化の追及

6：①データ標準化（資産やそのデータ形態の定義、API等アクセス法の標準化）、②機能のビジネスロジックを既述したライブラリ（資産の発行や移転など業務や契約に係るもの、デジタル・アイデンティティや金融資産など主体別ウォレットで管理運用されるもの）、③ブロックチェーン基盤の３つがモジュール群として整理されている。

7：リバンドルとは結合による再構築を、リビルドは設計図から書き換える再構築を指す。

8：基盤レイヤーの一つである法規制においても、金融サービスのモジュール化に合わせた対応（法規制のモジュール化）が求められる。

9：サービスの自動化については非金融分野において先行して進んでおり、デジタル化社会の特徴となっている。様々な経済活動や社会・自然現象がデジタル情報として記録され、既存の業務プロセスがデジタル情報の処理に最適となるように再設計されつつある。業務プロセスのアルゴリズムを見直す際には、データサイエンス（AI/ 機械学習 / 因果推論等）と新しいITシステムの構築手法が活用され、ビジネス実装と運用結果のフィードバックに基づく連続的改善という高速サイクルが可能となったため、短期間で効率的にサービスの高度化を推し進めることや新サービスを発見することが可能となっている。安価かつ迅速にテスト可能なことは大量のトライアルを必要とする新サービスの発見にとって必須である。例えば、現在進行形で生じている生成AIのビジネス実装が挙げられる。

③　スケール・メリット（規模の経済性）の追及
④　正の外部性への期待
⑤　システム構築・更新・償却にかかるコストの抑制と開発の高速化
⑥　モジュール開発に特化することによる高度化・イノベーションの進展
⑦　グローバル化などビジネス環境変化への対応

これらのドライビングフォースを上手く働かせるためには、相互運用性を確保しつつ、基盤レイヤー・金融サービスのモジュール化やクラウド化・オープン化を図る必要があり、その前提として標準化とデータ整備が求められている。

第２部　「次世代金融インフラのあるべき姿」の例示

１．本提言のアプローチ

第 1 部で述べた通り、様々なドライビングフォースによって、アンバンドリングやリバンドリング、リビルドが進展するものと予想されるが、実際にどのような形で実現するのかを予想するのは困難である[10]。

将来、実現する「次世代金融インフラ」は各国の金融制度の成り立ち、国民性、想定する時期等のほか、特に金融当局による監督のあり方により異なるものであり、一義的に決まるものではなく、様々な可能性があり得る。例えば、誰がサービスの担い手になるかという視点からは、（1）金融仲介業者が次世代金融インフラの中核を担うケースが考えられる一方、（2）IT 技術を活用して金融仲介サービスが全面的に自動化されるケースも考えられる。後者は極端な事例であるが、証券市場における STP 化（=Straight Through Processing、売買から決済までの業務の自動化）はその部分的な実現と捉えることが可能である。分散台帳技術を用いた新しい金融資産では、市場取引から所有の台帳管理までを一気通貫して担う金融インフラが DeFi としてすでに登場している。また、前者（1）のケースについても、決済システムを中央銀行と民間の金融仲介業者がそれぞれの役割を担う二層構造を前提とするケース、中央銀行・民間が一体となって担当する一層構造、中央銀行が存在しない一層構造、民間金融機関のみによる多層構造[11] などを前提とするケースが想定され得る。

そこで、本研究会では、長期的な視点から「次世代金融システムの将来像」を例示することとし、まずは金融機能別思考に転換することによって生じる変化の例として基盤レイヤーとその上に用意される金融サービスのモジュールという構造を示すこととする。

次に、モジュールの提供主体として、上述した 2 つのケース、すなわち、（1）金融仲介業者が中核を担うケース、（2）デジタル技術によって構築された決済・情報連携システム等が金融仲介機能を代替するケースに分けて例示しつつ、利用者が基盤レイヤー上のモジュールを利用する姿を示したい。

10：「未来は既にそこにある。ただ、一様には展開していないだけだ」とは著名な SF 作家の William Gibson が語った言葉である。ここでの文脈で捉えると、様々な要素技術が既に誕生しており、未来はその組合わで実現していく（未来は既にそこにある）が、思いもよらないような形で未来は姿を現すため予断をもって捉えない方がよく、アントレプレナーの思考の闊達さをもって探索を続けていくことが重要という理解になろう。

11：例えば、東京の米ドル決済は、国内の銀行がメガバンクや大手米銀の東京支店に米ドル預金口座を持つことで履行されており、民間銀行の二階層構造となっている。ロンドンにおける円決済も同様であり、国際的な大手行やメガバンク現地支店が担っている。

金融サービスのモジュール化に対応して、法規制を始めとする基盤レイヤーもモジュール化と組合せや組込みの活用が必要となる。その際、規模や範囲の経済性などの外部性の検討、金融システムの安定性やグローバル化への対応なども勘案した議論が必要である。

2．次世代金融インフラの将来像とその実現に向けた論点

（1）金融機能別思考への転換：レイヤー構造下でのモジュール構築

あらゆるビジネスプロセスがデジタル化し、IT によって支えられるようになり、その過程で IT の特徴であるモジュール化とレイヤー構造化がビジネスモデルに入り込んできている。特に、装置産業である金融サービス業ではモジュール化が行われやすい環境にある。実際、金融分野ではアンバンドリング（一体として提供されていた金融サービスを解体）、リバンドリング（解体されたサービス等を組み合わせること）が盛んに行われている。また、金融インフラと金融サービスの関係として捉えられるレイヤー構造や、法規制と適用対象となる金融サービスの関係というレイヤー構造も、IT のレイヤー構造と類似した関係となっている。

こうした金融サービス業の特徴もあり、非金融分野から金融サービス業への参入、さらには、金融分野・非金融分野間におけるデータの利活用を通じた金融サービスの提供が続いており、金融サービス業の担い手にも変化が生じている。例えば、電子マネーの発行、EC プラットフォーマーによる金融サービスの提供、後払い・前払いという与信サービス、BaaS の流れに沿った非金融ビジネスへの金融サービスの埋込み（Embedded Finance）などが挙げられる。

金融を取り巻く環境の激変等に対して、我が国金融制度の成り立ちによる制約もあり、図 1 に示す通り、これまではいわゆる護送船団方式に代表される「銀行・証券・保険・その他の金融業という業態」に依拠した思考に従って業界調整を行うなど、業法を中心とした法規制体系に基づいた金融サービスの提供やそれに対応した監督が行われてきており、新たな金融サービスの登場や既存の金融サービスの変化に対応することが難しくなっている。しかしながら、本来、「同じ活動・同じリスクには同じ規制を適用する」との原則が適用されるべきであり、経済主体別思考から金融機能別思考に転換する必要がある。

そこで、金融サービス産業の現状とその将来像を見通すための新たな視点として次に示す考え方を提示する。

①　金融サービスの構成と金融分野・非金融分野の連携・融合

図 2 に示す通り、金融サービスの提供主体に立った思考から金融機能に立った思考に転換し、金融サービスを機能別に分解した上で、改めて利用者から求められる金融サービスに合わせて金融機能等を組み合わせることによってモジュール化された金融サービスを提供する姿が求められる。

その際、世界的なオープンバンキング化の流れも踏まえ、金融分野・非金融分野で得られる情報データを相互に利活用するなど、非金融分野との連携・

融合を図ることが求められている。この場合、利用者データの権利（CDR = Consumer Data Rights）の法制化など、そのあり方について早急に対応する必要がある。

② 基盤レイヤーにおける共通ルールの設計

具体的には、図 3 に示す通り、デジタル技術を活用した次世代の金融サービスを提供することを前提にして、可能な限り非金融分野も包摂した形で横断的に適用されるルール（法規制、税制、会計ルール、認証制度、ALM/CFT/eKYC などを含むガバナンス、情報セキュリティ、決済システム、国際協調ルール、情報連携ルール、慣行など）とこれらのルールに則ってオペレーションされる IT システムなどが各種のレイヤーとして整備される。なお、各レイヤーは少なくとも金融分野全般に共通して適用されるルールとする必要がある。

③ 金融機能・金融サービス・法規制の対応

これらレイヤー構造の上には、金融分野・非金融分野の垣根や金融業態の枠にとらわれることなく、利用者の求めに応じて、次に示す視点等から分類される金融機能等を組み合わせることによってモジュール化された金融サービスが用意・提供される。なお、この分類はあくまで例示であり、実際に次世代金融インフラを構築する際に、その時々の状況に合わせて改めて見直す必要がある。また、法規制は、業法を中心とした形ではなく、金融サービスに着目した形で設計される必要がある。モジュール化された金融サービスの提供に当たっては、各金融サービスに対応してモジュール化された法規制が適用されることが必要となる。また、法規制については、金融サービスのイノベーション（新たな金融サービスの創造）を最大限引き出す形となるよう留意する必要がある。なお、複数のモジュール（金融サービス）が同時または並行して提供される場合には、モジュール化された法規制が単体で適用されるだけでなく、新たな問題が生じないように対応する必要がある。

【金融サービスを分解する視点】（図 2 参照）

（a）金融機能（交換・決済機能、価値保存機能（調達機能・与信機能・運用機能・期間変換機能）、価値尺度機能、保険機能、情報生産機能など）
（b）金融サービスの利用者属性（保有資産・投資資産の金額、プロ投資家への該当性・情報収集・分析能力など）
（c）リスク性（価格の変動特性、流動性リスクなど）
（d）提供される金融サービスに対応した時間軸（短期、長期など）
（e）市場の種類（リテール市場、ホールセール市場、クロスボーダー市場など）

なお、(e) 市場の種類については、特に (b)・(c)・(d) と密接に関連する。

④ 現行の金融サービスの事例

現行の金融サービスのいくつかを例として取り上げ、その提供のために必要

となる金融機能等との関係を示すと次の通りである。

（1）小口預金
交換・決済機能、流動性リスクを伴う調達機能、主として価格変動リスクなし、リテール、短長期、情報生産機能（預金関連情報）

（2）大口預金
交換・決済機能、流動性リスクを伴う調達機能、主として価格変動リスクなし、ホールセール、短長期、情報生産機能（預金関連情報）

（3）融資
信用リスクを伴う与信機能・資産運用機能・期間変換機能、主として価格変動リスクなし、リテール／ホールセール、短長期、情報生産機能（融資関連情報（融資先の信用情報など））

（4）住宅ローン等
与信機能・資産運用機能・期間変換機能、主として個人向け、長期、情報生産機能（融資関連情報（融資先の信用情報など））

（5）生命保険の募集・引受け
調達機能、主として価格変動リスクなし、主として個人向け、主として長期、保険機能、情報生産機能（保険関連情報）

（6）有価証券の取次ぎ・引受け
交換・決済機能、リテール／ホールセール、短期、価格発見機能、情報生産機能（資金調達者・投資家に関する情報など）

（7）有価証券運用
資産運用機能・期間変換機能、リテール／ホールセール、短長期、価格変動リスクあり、価格発見機能、情報生産機能（資金調達者に関する情報など）

（8）外国為替業務
交換・決済機能、リテール／ホールセール、短期、価格発見機能、情報生産機能（外国為替関連情報）

（9）暗号資産等の交換業務
交換・決済機能、調達機能、主として価格変動リスクあり、リテール／ホールセール、短長期、価値尺度機能、情報生産機能（投資需要、資産売買＜法定通貨預金との入替えを含む＞、送金に関する情報など）

⑤　新たな金融サービス・非金融分野との連携・融合等の事例

デジタル技術の進展に伴い、新たに産み出され、あるいは、変容を遂げた金融サービスを例示すると次の通りである。

（1）電子マネー
交換・決済機能、価値保存機能、主として価格変動リスクなし、リテール、短期、情報生産機能（決済利用状況、送金・入金に関する情報など）

（2）セキュリティトークン
調達機能、資産運用機能・期間変換機能、価格変動リスクあり、リテール／ホールセール、短長期、ファンマーケティング、情報生産機能（資

金調達者・投資家に関する情報など）

（3）ステーブルコイン
交換・決済機能、預金・現金に類似した価値保蔵機能、情報生産機能

（4）ZEDI（全銀 EDI）[12]
交換・決済機能、（非金融分野を含む）情報生産機能

（5）WeChat Pay などのスーパーアプリ[13]
交換・決済機能、調達機能、価格変動リスクなし・あり、リテール、短長期、情報生産機能（預金・与信・保険・非金融分野関連情報）

（6）非金融分野におけるデータの利活用を通じた金融サービスの提供[14]
交換・決済機能、調達機能・与信機能・運用機能・期間変換機能、情報生産機能（預金・与信・保険・非金融分野関連情報）

（7）ポイント経済圏によるプラットフォームへの組込み[15]
リテール、短長期、情報生産機能（非金融分野も含む情報）

（8）クラウドファンディング・サービス[16]
調達機能・与信機能・運用機能・期間変換機能、主として価格変動リスクあり、リテール／ホールセール

（9）実物資産等のトークン化[17]
交換・決済機能、価値保存機能（調達機能・与信機能・運用機能・期間変換機能）、リテール／ホールセール、情報生産 (処理) 機能

（10）統合台帳アプローチ[18]
交換・決済機能、情報生産 (処理) 処理機能

（11）パーパス・バウンダリー・マネー（PBM ＝目的拘束型マネー）[19]
交換・決済機能、情報生産 (処理) 機能

（2）金融仲介機能の担い手： ケース・スタディ

（1）に示したように金融機能等を組み合わせてモジュール化された金融サービスが提供されることとなったとしても、金融仲介機能の担い手については、現行の仕組みと同様に複数の独立した金融仲介業者が担うケースから、決済・情報システムなどの IT システムが金融仲介機能の重要な部分を担い、アルゴリズムによって自動執行されるというケースまで様々な形態が想定され得る。後者は、極端なケースではあるが、ブロックチェーンを活用した自動執行、非中央集権的な IT システムによる介在（IT インフラの提供主体はいても、サービスそのものの提供主体が不在（＝ IT システムが提供主体に該当））という形で一部は実現している。例えば、DEX（Decentralized Exchange）、暗号資産のレンディングサービス、イールドファーミング（流動性マイニング）による運用、ステーブルコインによる送金、DeFi インシュアランスなどが挙げられる。

そこで、上述した通り、ここでは金融仲介機能の担い手として、2 つのケースに分けて例示する。

12：ホールセール決済サービスに情報処理サービス（企業経理）をバンドルしたサービス。

13：決済サービスに預金・投資・ローン・保険などの金融サービスを統合し（包括的リバンドリング）、SNS サービスと連動することで送金機能も高度化。同様に、SoFi（Social Finance）は学生ローンリファイナンスから銀行・投資・保険に拡張。また、Grab Financial（配車サービス）や Apple Pay（通信デバイス）もプラットフォーム・ビジネスへの組込みという視点からは同様なビジネスモデルである。

14：非金融分野における利用者動向のデータ（会計情報・購買履歴・生産活動情報等）の利活用を通じた金融サービスへの展開。

15：E- コマースや通信キャリアなど、個人獲得（ID 獲得）が情報生産活動の基盤。

16：新しい資金調達手段、資金調達者による情報提供手段。

17：所有者情報の管理と所有権の移転の効率化、交換の効率化（市場流動性の向上、クロスボーダー取引の効率性向上）、新しい与信チャンネルの創造、投資機会やファンディング機会の拡大、ユーティリティサービスとの融合、新技術によるアクセス (手段) ・サービス (機能) ・アセット (金融商品) ・プラットフォーム (調達・交換、情報流通の場) の提供が挙げられる。

18：別個に設計・発展してきた資金、証券、実物資産の各台帳基盤を接続する場合、擦り合わせによる連携が個々に行われてきた。新しいアプローチとして、同一プラットフォーム上の台帳に集約することで相互運用性の壁を回避する方法が提案されており、BIS Innovation Hub の Project Agorá では、その POC（概念実証）が進められている。

19：PBM とは支払いに係るプログラマブルな動作をデジタル資産に埋め込むためのマネーであり、IMF・MAS の ASAP モデルで提唱された概念である。支払い用途を限定したマネーとして活用できる。(10) の統合台帳アプローチは、ホールセール決済用途で

【ケース1】 金融仲介業者が中核を担うケース（図４参照）

（A）スキームの概要

１つ目の「金融仲介業者が中核を担うケース」としては、図４に示す通り、可能な限り非金融分野にもまたがった横断的に適用される法規制、決済システムなどが基盤レイヤーとして整備され、金融仲介業者がこれらの基盤レイヤーの上に用意された複数の金融機能等を組み合わせた金融サービスを提供するスキームが考えられる。

このケースにおける重要なポイントは、引き続き金融仲介業者が中核的な役割を担うことを想定しているものの、上述した通り、従来の経済主体別思考から金融機能別思考に転換し、それぞれの金融仲介業者は、金融機能等を組み合わせることによってモジュール化した金融サービスを提供することである。その上で、金融仲介業者は、業態の枠にとらわれることなく、自らの求めるビジネスモデルに沿った形でモジュールを取捨選択して金融サービス業務を遂行する。また、金融当局においても、従前とは異なり、業態の垣根を越えて提供されるモジュールの集合体である金融仲介業者に対して、許認可・登録・届出を含む法規制に基づき監督業務を遂行することとなる。

（B）基盤レイヤーに求められる要件・留意事項

基盤レイヤーには、デジタル技術を活用した金融サービスを提供するために必要とされる要素のうち、可能な限り非金融分野にもまたがって横断的に適用されるルールが求められる。なお、各レイヤーは少なくとも金融分野全般に共通して適用されるルールとする必要がある。

基盤レイヤーとしては、法規制、税制、会計ルール、認証制度、ALM／CFT／KYC などを含むガバナンス、情報セキュリティ、決済システム、国際協調ルール、情報連携ルール、慣行などが考えられる。また、広く捉えれば、これらのレイヤーを律する監督体制や IT システムなども含めて考える必要がある。

（C）モジュールに求められる要件・留意事項

金融仲介業者は、自らのビジネスモデルに沿って取捨選択したモジュール（金融サービス）の集合体として業務を遂行する。そのため、法規制等に従って、モジュールの集合体として、それに見合った体制・組織を整え、金融当局から許認可・登録・届出等を受ける必要がある。なお、モジュールの提供に当たっては、現行の預金サービスと同様に、金融仲介業者が二層構造を前提とした方式（中央銀行が金融機関等を利用する構造など）のほか、統合台帳アプローチのプロジェクトが想定するような中央銀行・民間が一体となってプラットフォーム運営を担当する一層構造（ただし、預金マネーの発行体はいずれでも可）を前提とする方式とした制度とすることも想定し得る。また、CSD（中央証券預託機関）は、直接参加・間接参加の階層構造をとることが多いが、これも CSD または暗号資産台帳のように IT インフラが一元管理することも考えられる。

PBM を一つの台帳インフラで駆動させるものと言える。概ね法定通貨以外のマネーはなんらかのかたちで機能が限定されており、その限定を緩和しマネーを使いやすくすることが支払いに関するイノベーションの一側面であったといえる。これは、金貨などの商品貨幣を除くマネーの多くが債務性マネーとして発行されてきたこと、債務の移転には困難が付きまとうことに起因している。互いに預金口座を持ち合う形のシンプルなコルレスバンキングに対して、滞留資金の不効率性という問題を解消するために二階層型マネーシステムが産み出された。このイノベーションは債務性マネーの限界に対する解法の一つである。ちなみに、上部階層は中央銀行マネーである必要はない。中央銀行が創造される前の米国では、民間銀行の三層構造（Central reserve city banks – Reserve city banks – Country banks）により全国をカバーする資金決済網が展開されていた。

複数のモジュールを組み合わせるビジネスの場合、単純な和集合を超えて、それ以上の外部性を持つ可能性がある。例えば、負の外部性としてはシステミック・リスクや情報の寡占が、正の外部性としては金融機能の効率化や金融包摂、経済成長へのスピルオーバーなどが挙げられる。例えば、ユニバーサルバンキング、さらにはプラットフォーム・ビジネスとの金融の融合については正と負の両方の外部性を伴う。同様に、基盤レイヤーのモジュール化に当たっても、金融サービスのモジュール化に伴う外部性の存在に対応する必要がある。

その典型例が預金サービスである。預金には信用創造機能と資金調達機能の 2 つの側面があり、特に前者の場合は、システミック・リスク防止の観点からも、その決済手段に対して特別な配慮が必要となることに留意する必要がある。

（D）金融仲介業者が中核を担うスキームの効果

このスキームの場合、各金融仲介業者は自らの求めるビジネスモデルに沿った形でモジュール化した金融サービスを取捨選択することとなる。この結果、図 5 に示す通り、EU におけるユニバーサルバンクのように、範囲の経済性を求めて銀行業・証券業・保険業にまたがる形でモジュールを選択する業者も現れることもあれば、ＩＴの発達により大量のデータ処理を効率的に行えることとなったことから銀行業の中の個人向けの小口預金と小口融資に特化した業者（いわゆるナローバンクや個人向け預金･融資業務を行う流通系金融機関など）、逆に業務の選択と集中を徹底する考えから法人等の大口向けの預金・融資等に特化した業者も現れるかもしれない。さらには、個人向けに特化しつつ、範囲の経済による効果を求めて、預金・保険・証券・暗号資産等にまたがる金融サービスを提供する業者が現れる可能性もある。

当然のことながら、現行の金融仲介業者と同様の業務に見合った同じ形（銀行・証券・保険等の業態別）で金融サービスを提供することも可能である。

また、非金融分野においていろいろな情報データを収集している企業が自社のデータ等を金融分野向けに利活用して金融サービスを提供するなど、非金融分野からの参入も想定される。

この場合、上述した通り、当該金融仲介業者は、選択したモジュール（金融サービス）の集合体として法規制等に従う必要がある。特に留意すべきことは、現行の業態を前提とした法規制等とは異なり、適用される法規制等は提供される金融サービスの範囲に限られるということ、また、複数のモジュールを提供する際、前述した外部性に加えて利益相反などの問題に対応する必要があるため、適用される法規制の範囲は提供するモジュールに対する法規制の単なる集合ではなく、広がりのあるものとなる可能性がある。

（E）金融仲介業者が中核を担うスキームのメリット・デメリット

このケースの場合、金融仲介業者がトラブル対応を含めて全ての責任を負うことになることから責任の所在が明確化される。また、利益相反を防止する観点から、ファイヤーウォールやチャイニーズウォールなどの組織対応が可能と

なる。

このほか、金融仲介業者が金融機能等の組合せに当たって新たな金融サービスであるモジュールのアイデアを人為的に産み出すことが可能となる。ただし、新たなモジュールが創出された場合は法規制等に対応する必要があることに留意する必要がある。なお、場合によっては法規制等を見直すことも必要となる。

一方で、人的資源による対応が介在することとなるため、非効率な状況が起こり得る上、かえってトラブルを招くことになる可能性がある。また、利益相反の防止等のルールが遵守されている状況についても金融当局は常に検証・監督を行う必要がある。

【ケース2】 決済・情報連携システム等が金融仲介機能を代替するケース（図６参照）

（A）スキームの概要

２つ目の「決済・情報連携システム等が金融仲介機能を代替するケース」の場合、図６に示す通り、非金融分野にもまたがって横断的に適用される法規制、認証制度、決済システムなどや、これらを支える決済・情報連携システム等の IT システムなどが各種のレイヤーとして整備される必要があることは（２）のケースと同様である。

【ケース１】との違いは、金融分野・非金融分野にまたがるレイヤーとしてデジタル技術によって構築された決済・情報連携システム等が、これらのレイヤーの上に用意された金融サービスのモジュールを自動的に執行する点である。自動化は、アンバンドリングによりモジュール化されたサービス群の再構築（リバンドル）が「組合せによるイノベーション」を促進することから、効率性の向上のみならず、金融サービスの高度化や新しい金融サービスの創造がもたらされる可能性がある。その際、モジュールを繋ぐ機能は、擦り合わせ型開発から API 化に移行していくため、標準化は重要な役割を果たす。また、自動化に当たって、個人や企業、デバイスやシステム構成要素のデジタル・アイデンティティの整備も必要となる。

（B）基盤レイヤーに求められる要件・留意事項

【ケース１】と同様、デジタル技術を活用した金融サービスを提供するために必要とされる要素のうち、可能な限り非金融分野も包摂した形で横断的に適用されるルールが求められる。なお、各レイヤーは少なくとも金融分野全般に共通して適用されるルールとする必要がある。

【ケース１】との違いは、デジタル技術を活用することによって一定のルールに従って金融機能等を自動的に組み合わせてモジュール（金融サービス）を提供する決済・情報連携システム等がレイヤーとして構築されている必要がある点である。なお、決済・情報連携システム等については政府・中央銀行、民間が協力して一元的に管理・運営を担うことも、あるいは、政府・中央銀行の

ほかに、複数の民間機関が担うこともあり得る。また、自動執行するモジュールに見合った執行ルールをあらかじめ整えておくとともに、金融当局からの許認可・登録・届出等を受けておく必要がある。

（C）モジュールに求められる要件・留意事項

モジュールについては、【ケース 1】と同様、業態にとらわれることはないものの、法規制等の適用の観点に立つと一定のルールに従って金融機能等が組み合わされて自動執行されることから、新しい金融サービスが次々と登場するなどの環境変化に対して、どのように法規制を適用し、遵守状況をモニタリングしていくかなど、行政も対応が求められる。

その一例を挙げると、金融仲介業者が介在せず、決済・情報連携システム等を通じてモジュール化された金融サービスが自動的に提供されることから、その提供に当たって発生する予期せぬトラブル等に対して、利用者保護の観点に立って処理するための機関が必要と考えられる。なお、可能な限りトラブル等の発生を抑えるとともに、事前に予想されるトラブル等に対してはあらかじめ決済・情報連携システム等が自動的に処理するようシステム構築が行われている必要がある。

特に、複数のモジュールが同時にあるいは並行して提供される際には、場合によって利益相反が生じるなど、新たな法規制対応が求められる可能性があることから、このような事態が起こらないよう留意する必要があり、また、モジュール同士の利益相反の状況を監視する機関を設ける必要がある。

非金融分野の利用者が決済・情報連携システムを通じて金融サービスを利用する際、その保護を図る観点等から窓口機能・相談機能を果たす機関（ゲートウェイ）が介在することも必要となる。例えば、ＥＵではデジタル・アイデンティティについて、提供・登録・運用に関する法制を一元的に整備しており、その制度やシステムの設計は参考になろう。

（D）決済･情報連携システム等が金融仲介機能を代替するスキームの効果

このスキームの場合、金融サービスの利用者のニーズに応じて、あらかじめ用意された決済・情報連携システムが、業態の枠を超えた形で金融仲介機能を果たし、利用者間で金融サービスが自動的に執行されることとなる。なお、当該決済・情報連携システムについては、１つの共通した基幹システムの場合も、複数のシステムが併存する場合も考えられる。

例えば、中国のアリババ、テンセントの電子マネーを起点としたエコシステムサービスのように、銀行・証券・保険等の業態を超え、さらには非金融分野も巻き込んだ形で金融サービスを主として個人向けに提供するスーパーアプリが考えられる。将来的には、大企業向けにカスタマイズされた決済・情報連携システムを通じて、利用者である当該企業のニーズに合った金融サービスを提供することも考えられる。

例えば、金融資産や実物資産のトークナイゼーションが進展しつつあるが、トークン化されたデジタル金融商品などの売買に当たっても、資金決済と当該商品の決済が一体処理される決済・情報連携システムを通じて、資金とデジタ

ル金融商品の交換が同時かつ自動的に執行されることも考えられる。セキュリティトークンは資金調達・運用の新しいチャンネルであるだけでなく、ファンマーケティングや企業活動モニタリング、会計経理サービスとの連動の可能性を有している。

（E）決済・情報連携システム等が金融仲介機能を代替するスキームのメリット・デメリット

このケースの場合、金融サービスの提供を決済・情報連携システムが担うため、標準化されたルールに従ってモジュールが組み合わされ、金融サービスが提供されることとなる。また、利益相反を防止する観点から、ファイヤーウォールやチャイニーズウォールなどと同様の対応が自動的に図られるようにあらかじめシステム化されている必要がある。ただし、これらの防止装置を含む金融当局による検証・監督については当該金融サービスの提供前に行われれば十分であり、かなり効率的な方法と考えられる。

しかしながら、モジュールの組合せに当たっては、あらかじめ組み込まれた一定のルールに従って行われる必要があることから、法規制等に逸脱する可能性は低くなる一方、その代償として新たな金融サービスのアイデアが生まれる可能性は低くなるかもしれない（逆に、PC やインターネットの普及のように共通化・オープン化によってプラスに寄与するかもしれない。）。利用者サイドから提案されたり、決済・情報連携システムの管理運営機関が利用者利便に立って発案したりすることによって、新たな金融サービスが生まれる可能性は十分にあり、これを促進するような仕組みづくりが必要となる。その検討に当たっては、前述のデジタル化社会における新サービス創造に共通する特徴点が参考となる。

（3）次世代金融インフラ整備に際しての留意事項

次世代金融インフラの整備に当たっては、国際的な動向も踏まえつつルール等の標準化が図られる必要がある。このためには、政府・中央銀行・民間が緊密に協力して、標準化すべき領域を見極めつつ、その推進を図る必要がある。特に、デジタル金融の場合、国境の垣根が低いこともあり、今後、ますます国際標準が重要になってくると考えられることから、我が国として国際標準の取組みに積極的に関与し、世界をリードする必要がある。

（4）利用者に求められる要件・留意事項

非金融分野における利用者においても金融分野における方向性（金融機能別思考に基づく金融サービスのモジュール化など）に対応して、両分野におけるデータの相互利活用等のために業務プロセス・データ共有の自動化などを進める必要がある。

個人や企業の情報をどのように活用し、その過程で個人や企業の権利や利益が保護されるためには、活用のための工夫を考える側だけでなく、個人や企業の側にも新しい手法へ適応していく改革が求められる。ペーパレス化や AI 活用推進におけるハードルは利用者側に存在しているケースが少なくない。デジ

タル化社会のメリットを引き出そうとする新たな金融サービスが社会において受容されていくためには、利用者側の改革も必要である。

3．将来に向けた次世代金融インフラの整備と第2次提言「例示」との関係

今回、とりまとめた「例示」は、ある程度、遠い将来を見通して、2 つのケースを例示したものである。しかしながら、次世代金融インフラのあるべき姿については、本研究会においてすら、各メンバーの思い描く将来像は多種多様である。

本研究会としては、今後も次世代金融インフラのあるべき姿、そこに進むための道筋などについて引き続き検討を進めることとし、当面は、本提言で示した事項を踏まえつつ、中期的に取り組むべき具体的なテーマとして 3 つを選び、それぞれのテーマごとに分科会を設けて検討し、順次、提言をまとめることとしている。

なお、将来、デジタル金融サービスの提供を前提とした次世代の金融インフラが整備される際、本提言に盛り込まれた例示が参考となり、いろいろなバリエーションを描く中でよりよい次世代金融インフラが整備されることを期待する。

（以上）

Box　リバンドリングを促す7つの要因の事例説明

① 金融サービスの分解と組合せによる金融機能向上：　不動産を担保とするリバースモゲージでは死亡時に担保となっている住宅を売却して返済することとなるが、借り手が長生きした場合、与信者は住宅価格の下落リスクに晒される。住宅価格変動リスクをカバーする担保保証サービスと組み合わせたリバースモゲージ型住宅ローンとして組成することにより、元本返済負担を回避するだけでなく、借り手や貸し手が直面する死亡時期の不確実性にも対応することが可能となる。また、クレジットデリバティブや証券化商品では、信用リスクの保証に特化した金融商品やトランチング証券として与信に伴う信用リスクを異なる金融資産モジュールへ組み直している。資金調達にファンマーケティングを加味したセキュリティトークン発行では、転売可能なユーティリティトークンが別モジュールとして発行されている。こうした発想は社債と新株発行権の合成物であるワラント債やステーブルコイン（トークンの発行・流通基盤と価値安定を担保する仕組みの組合せ）にもみられる。以上のような金融機能の分解や組合せは現代に限った話ではない。預金は信用創造機能と決済機能を組み合わせた金融サービスであるが、ナローバンキングのように決済サービスのみに特化することも可能である。一般に発行体債務として提供される電子マネーも決済機能への特化型であり、第二部で紹介した PBM（パーパスバウンダリーマネー）は特定目的に特化したマネーである。以上のように、金融機能の分離と組合せによる高機能化・高付加価値化は金融サービスに広く観察される現象である。これにより、時間や場所の制約、技術の制約、規制等の制約などを自在に乗り越えられる金融サービスの創造が可能となっている。

② 新技術による低コスト・高速化を通じた効率化の追及：　例えば、紙の台帳から電子台帳への進化、オンラインバンキング化、市場取引所の自動発注化、コールセンターのチャットボット化、決済や送金の短期化・低廉化、ERP による仕入れや生産管理、経理の合理化、商流 EDI がもたらした事務プロセスのリアルタイム処理と現状認識・管理の高速化、人間の判断の機械化（AI/ 機械学習）、連続的改善のシステム化（A/B テストほか因果推論や EBPM の DevOps/MLOps による実装[20]）などが挙げられる。デジタル技術の発展により IT システムの構築がモノリシックな密結合システムからレゴブロックを組み合わせるようなモジュールの疎結合開発に変容している。金融サービスや基盤インフラを支える IT システムも同様であり、上記のような新技術による効率化を享受するためには、必然的にモジュールの結合によるサービス構築とならざるを得ない。

③ スケール・メリット（規模の経済性）の追及：　あらゆる産業がソフトウエア産業化・IT 産業化している。こうした産業は、生産拡大における低い限界費用と高い初期コスト（システム開発と実装）の組み合わせが特徴であり、典型的な費用逓減産業となる。初期コストを抑制し、スケール化を容易にするためには、ビジネスの拡大に合わせた伸縮自在な IT システムが求められる。そのためにはモジュールの組み合わせや仮想化技術を活

20：DevOps (Development and Operations) は、開発と運用を連携させ、アプリケーションの開発・テスト・リリースを迅速化する手法。CI/CD (Continuous Integration and Continuous Delivery) のパイプラインを構築し、開発・テスト・運用の自動化を実現する。MLOps は DevOps の改善プロセスに ML(機械学習) を組み込んだもの。機械学習モデルの開発・実装・運用・改善を継続的に管理する手法。

用したクラウドサービスが効果的であり、規模の経済の追及においてもモジュールのリバンドリングによるビジネスモデルの構築が重要となる。

④ 正の外部性の発見：　範囲の経済は、異なるビジネスを同時に手掛けた場合、個別に生産するよりもコスト安になることを意味する。複数のビジネス領域を手掛ける際には、当該企業にとどまらず、財・サービスの高付加価値化において正の外部性が発揮される場合がある。例えば、i）EC グローバル大手が自社ビジネス用に投資したデータセンターとサーバ群、EC 用に開発されたアプリケーション群を、クラウドを通じて一般サービスに転用した事例、ii）業務サービスの一環として派生した顧客データを他の用途に活用することで、需要予測やダイナミックプライシング、推薦システムなどのパーソナライゼーションサービスを進化させて高収益化に繋げるデータ・マネタイゼーション、iii）中国大手 EC 企業が決済サービスとエスクロウサービス（商取引の間に第 3 者が介入し決済にいたるまでの安全な取引を担保する仕組み）から始まったビジネスを、スーパーアプリの展開を通じて生活サポート全般に拡張し、情報の循環回転ビジネスモデルやデータ駆動型システムに発展させていった事例などが挙げられる。様々な標準化事例も公共財としての正の外部性をもたらす効果を有しており、ISO 規格や API、通信規格、インターネットプロトコルなどが該当する。

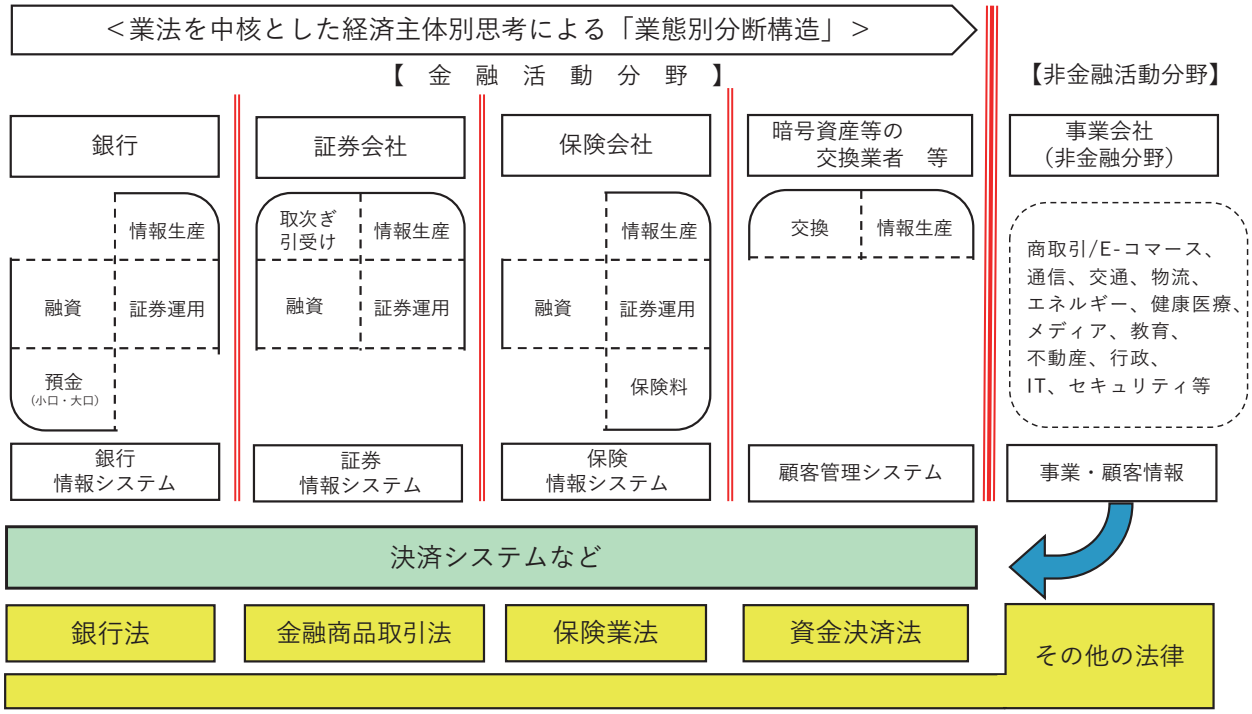
⑤ システム構築・更新・償却にかかるコストの抑制と開発の高速化：　ビジネスの迅速な立ち上げや DevOps の試行錯誤を通じた連続的改善では、システム構築や保守、再構築、ビジネス撤退時の廃棄償却にかかるコスト抑制が必須となる。例えば、オンプレからクラウドへのハードの移行、仮想化コンテナ技術、クラウドのモジュール化されたマネージドサービスの活用（ソフトウエア開発や運営保守でのクラウド移行）、Web アプリケーションによるインターフェイス展開など新しいシステム開発運営手法を駆使することでシステム構築やビジネス撤退時の廃棄償却にかかるコスト抑制が可能となり、ビジネスを迅速に立ち上げたり、連続的な改善を図ったりできるようになる。モジュールの選択によるサービス構築では、必要な機能だけを選択して機敏にシステム構築・廃棄することが可能となる。モジュールを繋ぐ技術も標準化されており、インターオペラビリティの確保にも有益である。例えば、非金融機関が BaaS を活用することにより金融サービスを取り込むケースにおいては、こうしたモジュール化・パッケージ化されたシステムの活用が最適なソリューションとなる。

⑥ モジュール開発に特化することによる高度化・イノベーション進展：　サービス提供に必要なモジュールを各企業が開発するよりも、そのモジュールに特化した専門企業が開発したものを組み込む方が、サービスの高度化、迅速な開発、業界標準への対応、継続的な技術更新やセキュリティ対応などが容易に実行可能となる。例えば、認証認可に特化したサービスや、デジタル・アイデンティティの発行や管理、ネットワークやシステムの負荷分散・セキュリティ、API 連携、支払いサービス、EC 向け物流サービス、ネット情報収集、サブスクリプション管理、ユーザー行動解

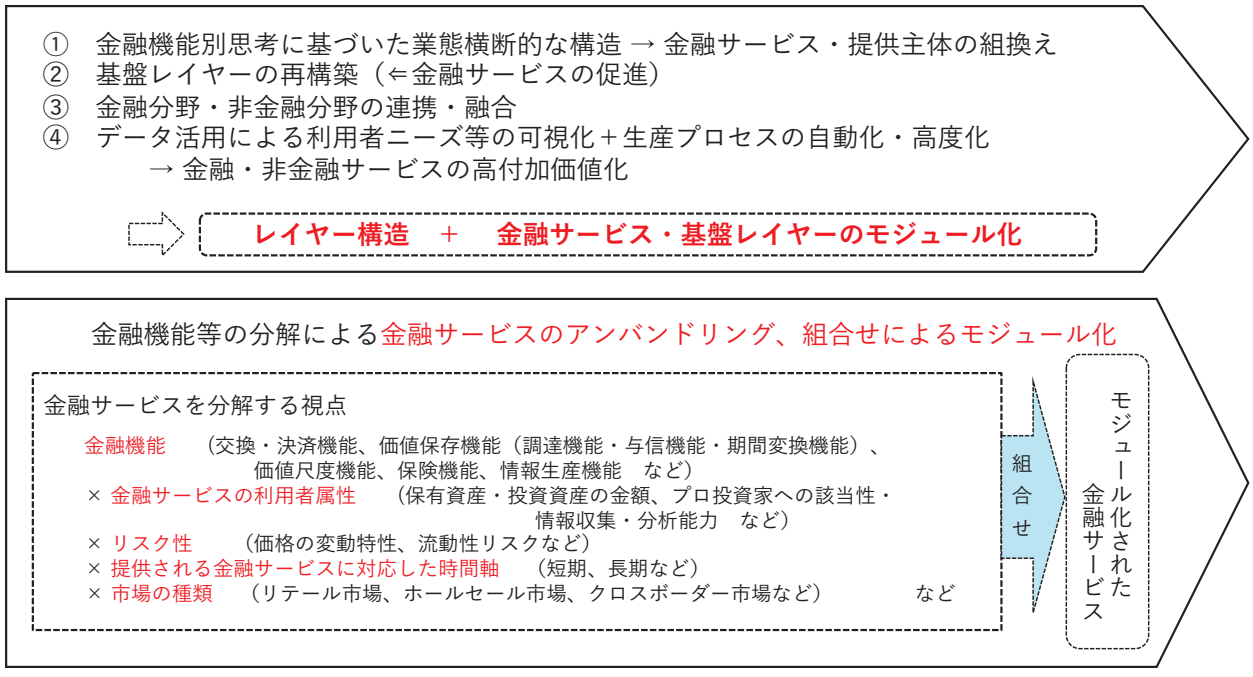
析、データ配信などが挙げられる。近年の生成 AI の開発においても、コアエンジン部分となるLLMの開発に巨額の資金を投入する一部の企業と、これを組み込んだサービス開発に専念する企業との分化が明確となっている。こうしたモジュール化されたサービスの活用により、様々な産業でビジネス戦略設計の自由度が高まっている。EC における販売、物流、決済、マーケティングの分離がその一例である。ほかにも、製造業における設計・製造・販売の分離（ファブレス半導体ビジネスが典型例）、モビリティサービスにおける需要探知・効率的な配車・移動サービス・決済等の分離、リテール電子決済におけるイシュアンス・アクワイヤリング・通信サービス・信用判定・不正利用防止・認証認可・BNPL 与信（後払いサービス）等の分離などが挙げられる。分離して特化開発された優れた業務ロジックやシステムを最適モジュール選択の視点で組み合わせることで、リバンドル化されたトータルサービスを提供することが可能となっている。

⑦ グローバル化などビジネス環境変化への対応：　ビジネスモデルが様々な業務内容の密結合で実現されている場合、グローバル化にともなう標準化や国際規制への対応が困難となりやすい。銀行規制や気候変動対応、AML/eKYC 対応などでは、様々なビジネスラインや国内外の取引先に分散して存在する情報の収集と加工、モニタリングが課題となっている。情報収集や加工・流通・検索を支えるデータ基盤の整備では、変化するビジネス環境に対応するためにも組み換えが容易なシステム開発が鍵となる。基盤レイヤーの整備進展によって情報処理の外部化が進むことへの対応も容易となる。例えば、気候変動対応や AML では、企業を超えた共通データベースが基盤レイヤーとして整備された場合、内部システムとの相互運用性が保てるようなシステム開発が重要となる。また、金融通信メッセージの国際規格である ISO20022 への対応や、デジタル認証とアイデンティティ管理の国際標準化などでも、業務プロセルが疎結合化されていると当該部分の置き換えによる対応が容易となる。

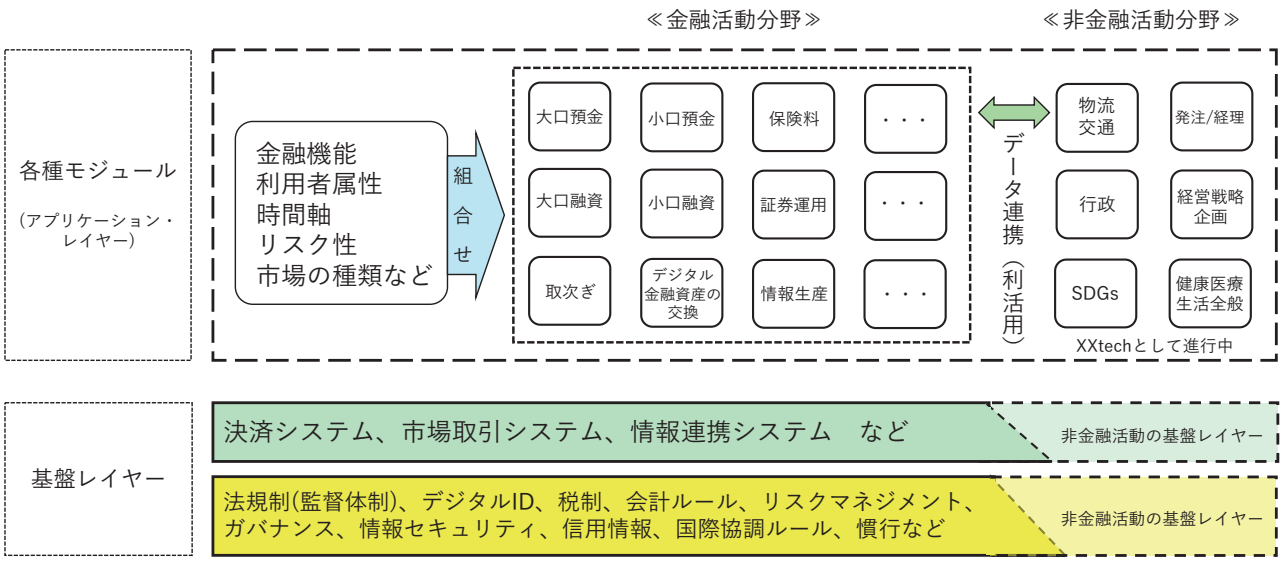
（図 1）
従前の金融インフラ



（図 2）
次世代の金融インフラ



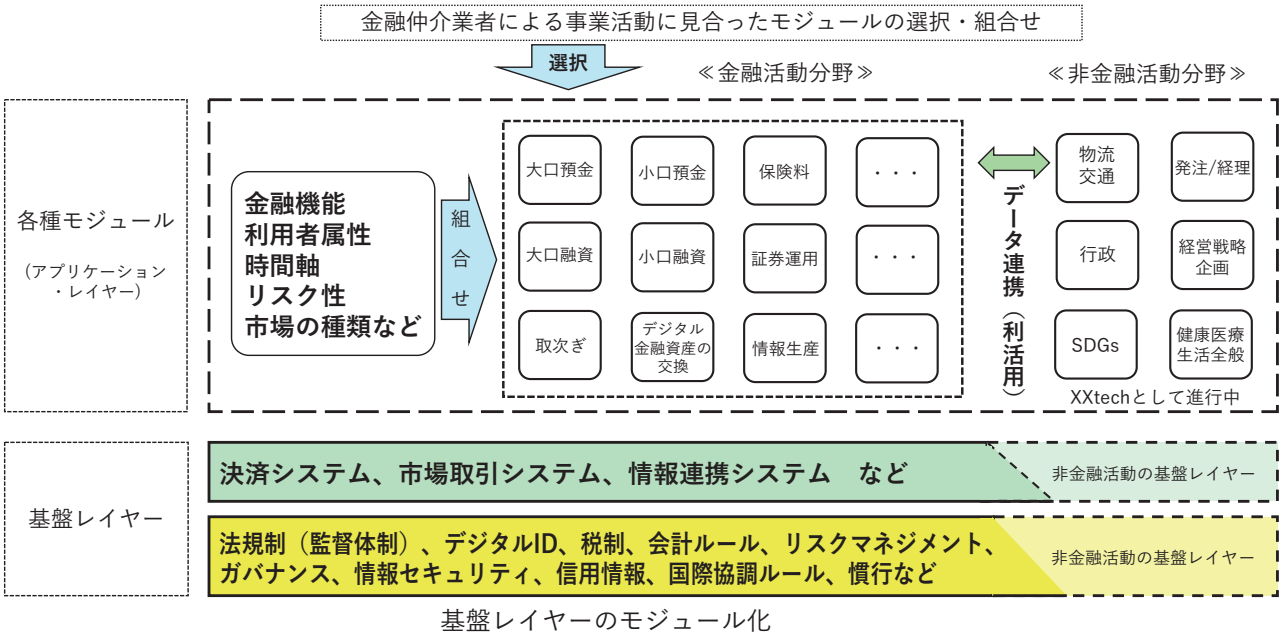
（図 3）
次世代金融インフラの例示：金融機能別思考への転換



基盤レイヤーのモジュール化

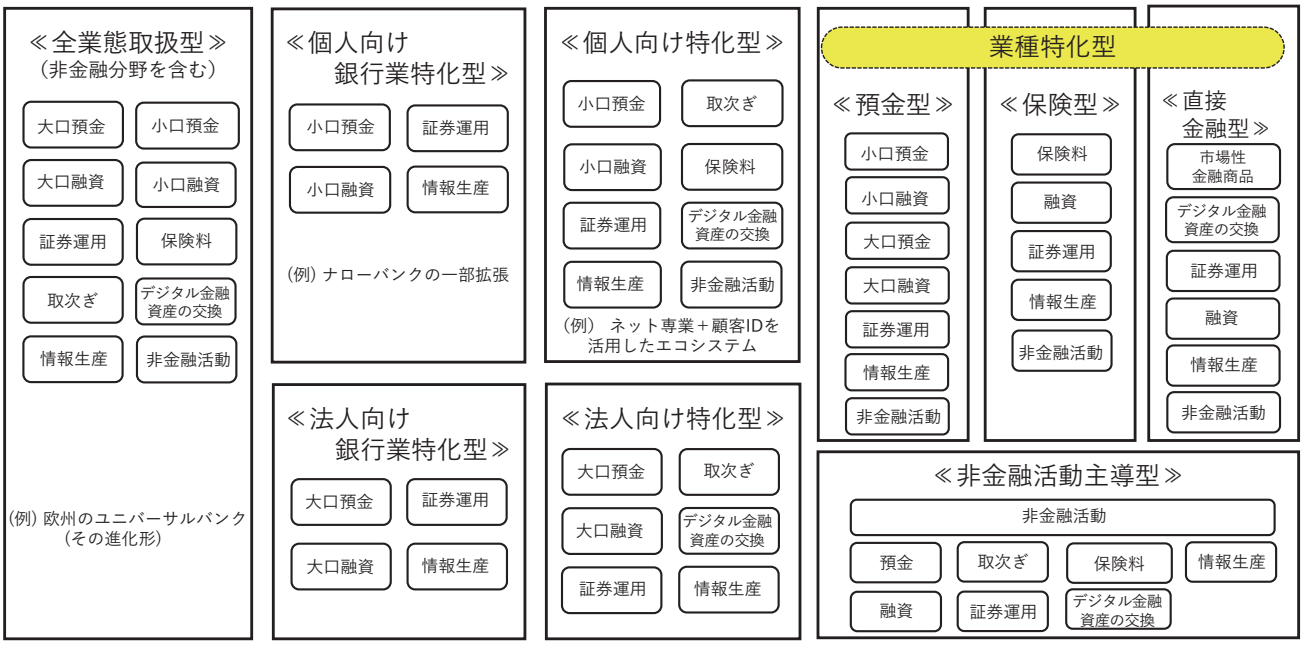
（図 4）
次世代金融インフラの例示：金融仲介機能の主体 ①

【ケース１】　金融仲介業者が中核を担うケース



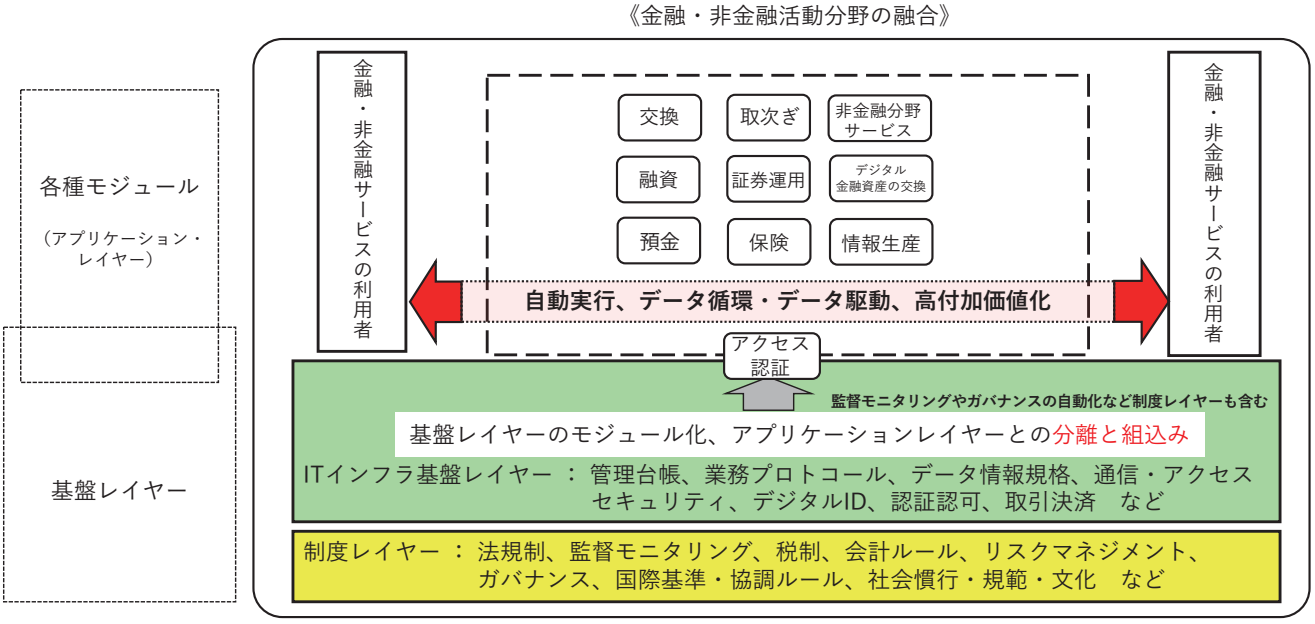
基盤レイヤーのモジュール化

（図 5）
次世代金融インフラにおける「さまざまな金融仲介業者」の例



（図 6）
次世代金融インフラの例示：金融仲介機能の主体 ②

【ケース 2】　決済・情報連携システム等が金融仲介機能を代替するケース



次世代金融インフラの構築を考える研究会

研究会メンバー（50音順）

小田（おだ）　玄紀（げんき）　氏　一般社団法人　日本暗号資産取引業協会　代表理事

廉（かど）　了（さとる）　氏　三菱 UFJ リサーチ&コンサルティング㈱ 調査・開発本部 調査部 主席研究員

小早川（こばやかわ）　周司（しゅうじ）　氏　明治大学政治経済学部 教授

増島（ますじま）　雅和（まさかず）　氏　森・濱田松本法律事務所　パートナー

山上（やまがみ）　聰（あきら）　氏　㈱NTT データ経営研究所　フェロー

若園（わかぞの）　智明（ちあき）　氏　公益財団法人　日本証券経済研究所　主席研究員、理事

（事務局メンバー）

山沖（やまおき）　義和（よしかず）　氏　SBI 金融経済研究所㈱　特任研究員

副島（そえじま）　豊（ゆたか）　氏　SBI 金融経済研究所㈱　研究主幹

（オブザーバー）

金　融　庁

日 本 銀 行（金融研究所）

次世代金融インフラの構築を考える研究会

開催日程

第 6 回会合（2024 年 6 月 11 日）
- 研究会のについて

第 7 回会合（2024 年 7 月 2 日）
- 報告書の公表方法について

第 8 回会合（2024 年 7 月 23 日）
- 今後の進め方

第 9 回会合（2024 年 9 月 27 日）
- ヒアリング（須賀 千鶴氏（経済産業省）、﨑村 夏彦氏（東京デジタルアイディアーズ㈱））

第10 回会合（2024 年 10 月 24 日）
- ヒアリング（西村 友作氏（中国・対外経済貿易大学）、岩崎 薫里氏（日本総合研究所））

第11 回会合（2024 年 11 月 18 日）
- メンバー報告（廉 了 氏、山上 聰 氏）

第12 回会合（2024 年 12 月 16 日）
- ヒアリング（瀧 俊雄氏（マネーフォワード㈱）、千葉 勇一氏（(一社) 全国銀行資金決済ネットワーク）

第13 回会合（2025 年 1 月 31 日）
- 報告書案の提示

第14 回会合（2025 年 2 月 18 日）
- 報告書案に関する意見交換等

第15 回会合（2025 年 3 月 26 日）
- 報告書案の取りまとめ
- 2025 年度の進め方に関する意見交換

SBI 金融経済研究所 所報 vol.8

2025 年 8 月 29 日発行

発行者：SBI 金融経済研究所株式会社

住所　〒106-6013
東京都港区六本木 1-6-1
泉ガーデンタワー 13F

電話　03-6229-1001

制　作：株式会社フクイン

SBI金融経済研究所