



BOJ *Reports & Research Papers*

2020年12月

東京大学金融教育研究センター・日本銀行調査統計局 共催ビッグデータフォーラムの様様

日本銀行調査統計局

本稿の内容について、商用目的で転載・複製を行う場合は、予め日本銀行調査統計局まで
ご相談ください。

転載・複製を行う場合は、出所を明記してください。

東京大学金融教育研究センター・日本銀行調査統計局 共催ビッグデータフォーラムの様相*

■要 旨■

東京大学金融教育研究センターと日本銀行調査統計局は、2020年11月9日、日本銀行本店にて、ビッグデータフォーラムを共同開催した。ここでは、まず、ビッグデータに関する計5本の最新の研究報告が行われた。具体的には、携帯電話等による位置情報、交通移動データ、倒産・延滞データ、ニュース記事などのテキスト情報、クレジットカードの取引データといった様々なビッグデータを経済分析に活用した事例が紹介され、質疑応答も含め参加者間で活発な議論が展開された。次に、学者、エコノミスト、データ作成の実務家といった多様なバックグラウンドを持つスピーカーによるパネルディスカッションも行われ、主として①非伝統的データの長所と短所、および②非伝統的データの将来展望と課題、という論点について議論された。

パネルディスカッションの第一の論点については、非伝統的データの長所として、多くのパネリストから、伝統的データに比べた速報性や頻度の高さ、情報量の豊富さといった点が強調され、これらの長所はコロナ禍における経済情勢の把握に際しても有用であるとの見解が示された。一方、非伝統的データの短所としては、サンプリングの問題やそのバイアス、観測ノイズの存在、データ期間の短さ、および統一化されたデータフォーマットの欠如などが挙げられ、ユーザーはこれらの点を十分に考慮しながらデータを利用すべきとの見方が示された。

パネルディスカッションの第二の論点である非伝統的データの将来展望については、多くのパネリストから、今後、ビッグデータに関する理解が深まるにつれて、政策形成の現場での活用が一段と進んでいくとの意見が示された。現在の課題としては、ビッグデータの提供側、ユーザー側ともに、専門的知識やスキルを有する人材が十分でないことを指摘するパネリストが多くみられた。また、ビッグデータを持続的なかたちで政策形成に活用していくうえで、同データの収集・整備にかかるコストを社会全体でどのように負担していくか、政府や公的機関の役割や関与の在り方も含め、関係者間で議論を深めていく必要があるとの認識が共有された。

* 本稿で示されたフォーラム内での報告・発言内容は発言者個人に属しており、必ずしも日本銀行、あるいは調査統計局の見解を示すものではない。なお、本稿に関する照会先は調査統計局経済調査課経済分析グループ（E-mail: post.rsd18@boj.or.jp）まで。

本稿は、2020年11月9日、日本銀行本店にて開催されたビッグデータフォーラム（東京大学金融教育研究センターと日本銀行調査統計局の共催）の様相を要約したものである。フォーラムでは、計5本の研究事例が報告されたあと、パネルディスカッションが行われた（プログラムは、別添を参照）。なお、今回のフォーラムは、感染防止の観点から、論文報告者・パネリストは日本銀行本店に集まる一方、それ以外の参加者はオンライン視聴するというハイブリッド形式での開催となった。

以下では、プログラムに基づき、冒頭挨拶、セッションごとの研究報告および質疑応答、パネルディスカッション、ラップアップの概要を紹介する。各参加者の所属などは、フォーラム開催時点のものであり、文中では敬称を省略している。

1. 冒頭挨拶

日本銀行理事の貝塚は、近年の急速な情報通信技術の進展を背景に、伝統的な統計調査やアンケート調査では不可能であった大規模なビッグデータの収集が可能となったと述べたうえで、ビッグデータの活用は、マーケティングを含むビジネス分野のみならず、政策形成の現場でも急速に進んでいるとの現状認識を示した。とくに、中央銀行の調査部署では、新型コロナウイルス感染症の流行により、各国の経済情勢が短期間で急激な変動を示しているため、リアルタイムに経済の実態を的確に把握する目的から、速報性の高い高頻度データへのニーズが高まっているとの見解を示した。ただし、ビッグデータを実務に活用していくにあたっては、サンプルの大きさや速報性といったメリットだけでなく、留意点や限界についても同時に理解を深める必要があると述べた。そうした観点から、学者、エコノミスト、データ作成の実務家といった多様なバックグラウンドを持つ参加者が一堂に集まって議論できる本フォーラムは意義深いと述べ、冒頭挨拶を締めくくった。

2. 第1セッション

2-1 「位置情報データによる経済活動の把握」：高橋耕史（日本銀行）、王悠介（日本銀行）

（1）論文報告

高橋・王は、携帯電話等の位置情報データを活用して、経済活動を把握する研究事例を報告した。具体的には、位置情報データに基づく各地点（100m四方のメッシュ）の滞在人口を用いて、対面型サービス業の消費動向や製造業の生産動向をナウキャストイングする手法を紹介した。

まず、対面型サービス業については、滞在人口が来店客数と相関している可能性が高いという特性に言及したうえで、①遊園地（娯楽業）、②ショッピングセンター、③飲食業に関する分析結果を報告した。すなわち、遊園地は、一つ一つの施設の規模が大きく、メッシュ内の滞在人口が他の活動の影響を受けにくいことから、入場者数を高い精度でナウキャストイングできると述べた。ショッピングセンターについても、時間帯別の滞在人口のパターンに基づくフィルタリングを行えば、売上高を比較的高い精度で捉えることができるとの分析結果を示した。一方、飲食業については、飲食店の周りに他業種の店舗が立地することが多いため、ノイズを除去することが難しいものの、クラスタリングといった統計的手法を組み合わせることで、高い精度でのナウキャストイングが可能であることを報告した。

次に、製造業については、鉱工業生産指数と労働投入量の相関度合いが業種によって異なるため、滞在人口により生産活動をどの程度捉えられるかは業種により異なると指摘した。そのうえで、経済センサスの事業所データを用いて工場敷地と想定されるメッシュを幅広く抽出したあと、昼間人口比率などの時間帯別の滞在人口分布を利用して主に工場を含むと思われるメッシュに絞り込むことで、業種別のナウキャストイング指標を作成し、輸送機械や生産用機械の生産動向については高い精度で予測できることを示した。

（２）質疑応答

水門は、飲食業におけるクラスタリングについて、共通のファクターを抽出することが目的であれば、主成分分析の手法を用いるのが自然ではないかと指摘した。また、飲食店に関して分類された各クラスターはそれぞれどの業態（居酒屋、ファミリーレストランなど）と関連が深いのか問うた。川口は、鉱工業生産と人流の相関について、強い業種と弱い業種があるのは、生産関数における資本労働比率が関係しているのではないかと述べた。また、人流に加えて、電力データ等で資本の稼働状況も捉えることができれば、生産活動のナウキャストイングの精度を高めることができるのではないかと提案した。宮川は、今後の分析の拡張として、企業の取引関係など別のビッグデータを組み合わせることにより、人流と物流の関係の分析を深めることを提案した。村上は、フォーキャストの精度に影響を与えているのは、工場が属する「業種」よりも、ノイズの大小を左右する工場の「立地」の影響の方が大きいのではないかと述べた。辻中は、位置情報データと組み合わせて使っている土地利用に関するデータは、その用途が時系列的に変化することは考慮しているのか尋ねた。渡辺は、輸出入や物価といった消費と生産以外の経済活動のナウキャストイングにも、位置情報データを用いることはできるのかと問題提起した。

報告者の高橋・王は、水門の質問に対し、変数が非負値となる制約を課す主成分分析に相当する「非負値行列因子分解」の適用も試みたが、安定的な結果が得られなかったため、今回の報告ではよりシンプルなクラスタリングを利用したと回答した。また、クラスタリングで飲食業の業態別の動向を捉える分析は、未実施であり、今後の課題であると応じた。辻中の問いに対しては、現時点で土地利用の時間変化は考慮できていないと回答した。川口と村上の問いに対しては、ナウキャストの精度が悪い業種は、①生産が資本集約的であり、滞在人口が生産に与える影響度が小さい、②立地によっては、ノイズが大きく、労働投入を正確に捕捉できていない、といった原因が考えられると答えた。渡辺の問いに対しては、物価は滞在人口から計測することは難しいが、物流とも関連する輸出入や在庫については、人の動きとも相関している可能性が高いため、今後分析を進めていきたいと応じた。

2-2 「交通移動データと非負値行列因子分解を用いた都市活動分析と変化検知」：前田高志ニコラス（理化学研究所）

（1）論文報告

前田は、交通ICカードに基づく駅間の移動データを用いて、①各駅において、どのような活動がなされたかを分析する手法、および、②その時系列変化を検出するための手法を提案した。本研究の問題意識として、交通移動は、労働や消費といった経済活動に伴って派生的に生じる事象であり、交通移動の目的を分析することにより経済活動を把握できる可能性を挙げた。そうした観点から、データの正確性が高く、リアルタイムでの取得が可能な交通ICカードに基づく駅の降車客数データは有用であると述べ、その移動の目的別内訳人数を推定する具体的方法として、「非負値行列因子分解」と呼ばれる手法を用いることを提案した。実際の分析事例として、関西圏のICカードデータから取得した降車客数について、時間帯別の分布情報を利用しながら、出勤やレジャーといった9つの目的（アクティビティ）別の因子に分解した結果を紹介した。

次に、各駅での各アクティビティの構造的な変化を捉える観点から、降車客数の分布の形状が日々どの程度変化したかを数値化し¹、構造変化が生じた日を検出する手法を提案した。この手法を、実際に降車データに適用してみたところ、ショッピングセンターの拡張や新規開店のあった日が、大きな変動を示した日として検出されており、経済活動の変化を捉えるために有用な手法であると述べた。

最後に、今後の研究課題として、①新型コロナウイルス感染症による地域ごと

¹ ここでは、ある時点の前後における降車客数分布の差を、「Jensen-Shannon 距離」と呼ばれる統計学的な尺度を用いて計測し、構造変化を定量的に捉えている。

の活動量の変化を分析することや、②移動目的ごとの活動指標を作成し、これをマクロ経済活動のモニタリングに利用することなどが考えられると述べた。

(2) 質疑応答

宮川は、非負値行列因子分解の因子数を何らかの理論や統計量に基づいて決定することは可能か問うた。また、アクティビティごとの変化の検知は有用な手法と評価したうえで、どの程度のデータ期間が存在すれば検知が可能かを尋ねた。渡辺は、移動目的が互いに相関している可能性をどのように考慮しているのか尋ねた。亀田は、駅の乗降車に複数の目的があるケースをどのように考慮しているのか尋ねた。水門は、ICカードを持っていない人が存在することを踏まえると、ICカードデータは、GPSデータとどの程度整合的なのか尋ねた。また、川口は、推定された目的別降車人数が、パーソントリップ調査²の結果とどの程度整合的かを問うた。辻中は、駅周辺地域での消費データやイベント情報など地域別の集計データと組み合わせることで、分析の拡張ができる可能性があるとして述べた。オンライン参加の西村清彦（政策研究大学院大学）は、駅から駅への移動データおよびその時系列的な変化がより重要ではないかとコメントした。

報告者の前田は、宮川の質問に対して、非負値行列因子分解の因子数の決定方法について、因子数を増やしていき、各因子の分布の基本的なパターンとして似たものが現れたら、そこで止めるという処理を現時点では行っていると回答した。そのうえで、因子数の選び方の工夫として、外部データを結びつけて、システムティックに行うことが考えられると応じた。また、変化の検知には、変化日から2週間程度あれば原理的に分析は可能であるが、土日と平日で分布パターンが異なるため、もう少し長い期間のデータが必要であると考えられると応じた。渡辺や亀田の質問に対しては、現時点で乗降目的が相関している可能性や複数ある可能性は考慮できていないと吐露したうえで、例えば、ショッピングエリアに囲まれたビジネスエリアについても、ビジネスに特化した場所とショッピングに特化した場所が必ずあり、その点に関するデータを組み合わせることで、複合的な駅も両者の要素に分解可能となり得ると述べた。水門の問題提起に対しては、ICカードの利用率は場所や鉄道会社、年齢によっても差が生じており、将来的にはこうした地域別・年齢別の利用率の違いも考慮して、GPSデータとの整合性を確認していきたいとした。川口の問いについては、パーソントリップ調査の結果との相関が非常に高かったとしたうえで、この点は、他の手法と比較した今回の分析の優位性を示しているのではないかと述べた。辻中の意見に対して

² パーソントリップ調査とは、国土交通省が取りまとめている都市における人の移動を把握するための調査。本調査では、世帯や個人属性に関する情報と1日の移動をセットで尋ねることで、「どのような人が、どのような目的で、どこからどこへ、どのような時間帯に、どのような交通手段で」移動しているかを把握することが可能となる。

は、駅周辺地域での消費データやイベント情報などを用いることでより正確な分析ができる可能性が高いと応じた。さらなる分析の拡張の可能性として、ソーシャルネットワークのテキストベースのデータと紐付けることも可能ではないかと述べた。西村のコメントについて、乗車駅と降車駅の関係性の分析は非常に有意義であると同意したうえで、とくに、目的地の周りでどのような経済活動がどの程度増えているのか、といった時間的な変化に注目した分析も重要であると述べた。

2-3 「コロナ禍における企業ダイナミクス：企業レベルデータと人流データを用いた実証分析」：宮川大介（一橋大学）

（1）論文報告

宮川は、新型コロナウイルス感染症の拡大を防ぐ目的から導入された外出自粛要請などに伴う人流の変動が、企業倒産に及ぼした影響を実証的に分析した結果を報告した。

まず、Google の人流データと東京商工リサーチの大規模な倒産データを組み合わせて人流と倒産の関係を推計し、感染症拡大以降の人流の減少が企業倒産を有意に増加させていたとの結果を報告した。また、倒産の代理指標として、リース料の延滞データを使用しても、同様の結果が得られることを示した。

次に、こうした関係の背後にあるメカニズムを詳しくみるため、取引ネットワークを通じた人流低下の影響を考慮した回帰式の推計結果を報告した。自社が所在しているエリアの人流の減少だけでなく、取引先が所在している地域における人流の減少も、倒産件数を増やす方向に働いていたと指摘した。また、感染症拡大前の企業業績を説明変数に取り入れた推計結果から、従前から業績が低迷していた企業が感染症の拡大直後に退出する傾向があった一方、政策対応が進むにつれてこうした傾向は弱まったと報告した。さらに、感染症拡大前のデータを用いて推計したパラメータを用いて、感染症拡大以降の倒産件数を予測したところ、実績の倒産件数はモデル予測値よりも抑制されているとの結果を報告し、この間の各種の政策対応が倒産を抑制する効果を有していた可能性があると述べた。

最後に、本報告で紹介した分析枠組みは、経済理論に裏打ちされたものであると強調し、様々なビッグデータを組み合わせて政策効果などを規範的に分析するうえでは、こうした理論的なガイドが重要であると述べた。また、こうした分析を進化させていくためには、これまでのところ活用が進んでいない税務データなどを企業データと組み合わせていくことが有益であるとの見方を示した。

(2) 質疑応答

植田和男（共立女子大学、CARF センター長）は、新型コロナウイルス感染症への対応が進むなかで、今後、これまで対面で行われてきた経済活動がテレワークなどの非対面な活動で代替されるようになると、人流と経済データの相関がこの先低下していく可能性があるのではないかと指摘した。これに関連して、前田は、延滞の動きを説明するための変数として、人流に加えて、テレワークの進展の度合いや、各自治体が企業に対して実施した支援の規模を取り入れてはどうかと提案した。関連して、高橋は、倒産予測には、口座取引データや財務会計アプリのデータなども候補となり得ると指摘した。若田部昌澄（日本銀行副総裁）は、ビッグデータを経済分析に活用していく際には経済理論のガイドが必要であるという主張に同意したうえで、現在実施されている新型コロナウイルス感染症に対する様々な施策について、効果・副作用のトレードオフも含めて、規範的にはどのように評価したらよいかと問うた。貝塚は、税務データの活用がこれまでのところ進んでいない背景について尋ねた。村上は、経済分析の分野で整備を進めるべきビッグデータとして、具体的にどのようなものが考えられるかと尋ねた。渡辺は、ビッグデータを利用した分析において経済理論が有効であるという宮川の主張に対して、一定の同意を示しつつも、情報系・工学系など、経済学とは異なる分野の知見と交流していくことのメリットも大きいと主張した。

報告者の宮川は、植田および前田の指摘に対して、本分析は短期的な視点から企業の倒産に焦点を当てているものの、エリアごとのテレワークの進展度合いを考慮に入れることで、コロナ後の構造変化についても知見が深まるのではないかと応じた。また、若田部の指摘に対しては、現行の資金繰り支援策などは企業倒産をかなり抑制する方向で作用していたと評価することは、現段階までの分析でも可能ではないかと回答したうえで、本分析の結果をより厳密な形で規範的な評価につなげることは、今後の課題としたいと述べた。貝塚の問いに対しては、税務データの利用に対し、具体的にどのような要因が妨げとなっているか詳しく承知していないとしつつも、政策評価の現場では、税の支払い実績などのデータを活用すれば望ましい場面でも、実際にはアンケート調査の結果で代替しているケースがあるのは事実との認識を示した。村上の問いに対しては、税務データに加え、企業の取引金額を把握できるデータが今後充実していくことが望ましいと主張したうえで、高橋が言及した金融機関の口座取引データは、それを近似する指標となり得るかもしれないと述べた。渡辺のコメントに対しては、本報告では、経済学者としての立場を意識してビッグデータ分析における経済理論の有用性を敢えて強調したものの、本報告の内容が、情報系分野の専門家である前田の研究成果と親和性が高いことが示すとおり、他分野の研究者とできる限り協働していくことが望ましい点は、渡辺の指摘どおりであると述べた。

3. 第2セッション

3-1 「因果チェーンとネットワーク学習によるソースごとの COVID19 に関する捉え方の違いの抽出」：久野遼平（東京大学）、坂地泰紀（東京大学）

（1）論文報告

久野・坂地は、様々なニュースに含まれるテキストデータから、人々が経済環境について認識している因果関係を分析する独自の手法「認知因果ネットワーク」を紹介し、その具体的な適用事例として、新型コロナウイルス感染症が拡大するもとの人々の認知因果関係を捉えた研究結果を報告した。

まず、経済のモニタリングを高度化するためには、伝統的データのさらなる活用に加えて、テキストデータをはじめとするビッグデータの活用が重要であるとの問題意識を開陳したうえで、機械学習的な手法（自然言語処理、ネットワーク学習）を用いて、報道記事などのテキストデータから認知因果関係を抽出する本研究の手法を解説した。この際、テキストの主体によって、異なる認知因果ネットワークが構築される一例として、マイナス金利政策を巡る中央銀行と民間金融機関の認識の違いを紹介した³。また、テキストデータ以外の様々な経済データも活用して経済における因果関係を表現する「経済ノレッジグラフ」の手法を紹介した。

次に、その実例として、ダウ・ジョーンズから入手した出版社の異なる大量の報道記事に上記の手法を適用し、①マスクをめぐる様々な混乱が物資不足を生み出し、価格高騰に繋がったという関係や、②感染拡大による人々の恐怖心の高まりが通院さえも拒む外出抑制に繋がったという関係などが、認知因果ネットワークとして抽出されたとの結果を報告した。

最後に、認知因果ネットワークは、経済のストーリーの断片——今次局面のケースで言えば、感染症がどのようなメカニズムを通じて国民の行動変化を引き起こしたか——をうまく捉えることができ、シラー教授の著書⁴でも指摘されているアニマル・スピリットや熱狂、恐怖などナラティブが経済を動かす背景の分析とも親和性が高いと主張した。

³ 認知因果ネットワークを構築するためのデータソースとして、中央銀行のレポートを用いると、「マイナス金利は企業の資金調達コストの低下を通じて、企業の設備投資を増加させる」という因果関係が導かれる一方、民間金融機関のレポートを用いると、「マイナス金利は銀行収益の悪化および貸出姿勢のタイト化を通じて、企業の設備投資を減少させる」と、真逆の因果関係が導かれる。

⁴ Shiller, Robert J. 2019. *Narrative Economics: How Stories Go Viral & Drive Major Economic Events*. Princeton University Press.

(2) 質疑応答

渡辺は、本報告は経済学の言葉で言う「期待の異質性」を捉えようとする試みであるとコメントした。具体的には、人々の期待が異なる原因は、経済学的に考えると、①人々が保有している情報が異なること、②人々が経済に関して思い描く「モデル」ないし「世界観」が異なること、の2つであると整理したうえで、本報告は、経済学の分野で相対的に理解が進んでいない②の視点から、期待形成の異質性への理解を深めようとする分析であるとの解釈を提示した。この指摘を受け、植田は、マネタリーベース増加の物価押し上げ効果を例に、人々が思い描く経済モデルにばらつきがある場合、どのモデルが多数派であるかによって経済への影響が異なり得るほか、経済環境が変化するなかで「モデル」も徐々に修正されていくものであると指摘したうえで、本分析はそのような現象を捉えられるかと尋ねた。若田部は、本分析を意欲的な取り組みと評価したうえで、認知因果ネットワークを規定する情報源には、自分の思い描くストーリーと整合的なものとそうでないものがあるとし、そうした情報源の違いがネットワーク分析にどのような影響をもたらすのかと尋ねた。また、本分析は「経済が認知因果に影響を及ぼす」という点に焦点を当てているが、逆に「認知因果が経済に影響を及ぼす」ということもあり得るのではないかと質した。亀田や辻中は、情報源をマスメディアによる報道記事とした場合、認知因果ネットワークは「マスメディアが考える因果関係」を表現することになるのかと尋ねた。加えて、亀田は、ソーシャルメディアをデータソースとして用いた場合、得られる認知因果ネットワークは大きく変わり得るのかと問うた。宮川は、抽出された認知因果ネットワークを、いくつかの種類に分類・集約することができれば有益であると指摘し、そのような方面での研究動向の有無について尋ねた。桜健一（日本銀行）は、同じ情報であっても、日本語と英語では、言語の違いによって認知因果ネットワークが異なる可能性はないかと尋ねた。川口は、経済学の分野では、捉えるべき因果関係が必ずしも一意に定まらないことがあると指摘したうえで、本分析の手法を応用することで、因果関係の異質性の背後にあるメカニズムを分析することが出来るのではないかと提案した。水門は、経済環境が急激に変化するもとでも、認知因果ネットワークは安定的なのかと尋ねた。

報告者の久野・坂地は、植田の問いに対して、本研究は、人々の思い描くモデルの違いを捉えることに主眼があり、今はそのための基礎技術を開発している段階といえると応じた。若田部の問いに対して、新型コロナウイルス感染症が拡大していく今次局面では、人々が先行きの動向に関して信念（belief）を全く持っていない状況から、次第に多様な信念が形成されていく様子を観察することができると指摘し、この極めて珍しい状況を認知因果の研究にうまく活かしていきたいと述べた。また、認知因果が経済環境に影響を及ぼす可能性に関する研究について

ては、今後の課題であると応じた。亀田や辻中の指摘に対しては、本分析は米国の地方誌など幅広いメディアを網羅しているが、どのソースをデータとして取り込むかによって、得られるネットワークは大きく異なり得るとしたうえで、ソーシャルメディアを用いた分析についても取り組んでいきたいと述べた。宮川の問題提起に対しては、ネットワーク学習の技術を駆使して、抽出された様々な因果チェーンの中から、注目すべきチェーンを自動的に絞り込むことができる手法を構築することが有益であろうと述べた。桜の問いに対しては、最近の研究によれば、日本語・英語どちらも似た手法で因果文を抽出できることが分かっていると述べたうえで、得られる認知因果ネットワークの違いを分析することも本研究の目的の一つであると応じた。川口の指摘に対しては、工学者は認知因果関係の全体的な構造に関心を寄せる傾向にあるが、因果関係の異質性についても研究を進めていきたいと応じた。水門の問いに対しては、認知因果ネットワークは経済環境の急変に対して安定的でないとして述べたうえで、本分析の主眼は、むしろ、新型コロナウイルス感染症前後でネットワーク構造がどのように変化したかを明らかにすることにあるとコメントした。

3-2 「クレカ取引データと位置情報データを用いたステイホーム政策の効果検証」：渡辺努（東京大学）

（1）論文報告

渡辺は、携帯電話の位置情報データとクレジットカードの取引額データを用いて、新型コロナウイルス感染症の拡大に伴い人々の外出自粛が進んだ背景に関する実証分析の結果を報告した。

まず、携帯電話の位置情報データとクレジットカード取引額データを用いて、新型コロナウイルス感染症が拡大するもとで外出自粛の動きが拡がり、それと連動する形で国内消費が落ち込んだ様子を示した。そのうえで、日本政府の外出自粛要請は、米欧等のロックダウンと異なり法的拘束力を伴わないものだったにもかかわらず、これほどまでに外出自粛の動きが拡がったのは何故なのかと問題提起し、この現象は、「情報効果」、すなわち人々の恐怖心の高まりによるものではないかとの仮説を提示した。

こうした仮説を検証するため、外出自粛要請が発出されたタイミングが都道府県によって異なるという事実を識別条件として用いることで、外出自粛の動きが拡がった背景を、①外出自粛要請を通じた「介入効果」と、②感染拡大に伴う恐怖心の高まりを通じた「情報効果」に分解した。具体的に、東京都における外出減少のうち、約4分の3は「情報効果」によって説明されるとの結果を示し、こ

の定量感は米国のデータに同様の手法を適用した研究結果⁵とほぼ同程度であったと報告した。

最後に、本分析の結果を踏まえ、新型コロナウイルス感染症に伴う世界的な消費活動の低迷は、感染症に対する人々の恐怖心のグローバルな伝播により、外出を控える動きがわが国も含め各国で強まったことの影響が大きいと結論づけた。

(2) 質疑応答

若田部は、今次局面の世界的な経済の低迷は、恐怖心の伝播という情報効果だけでなく、サプライチェーンを通じた貿易活動の急激な縮小も影響しているのではないかと指摘した。また、感染情報に対する人々の受け止め方(弾性値)が徐々に変化してきている可能性を検証してはどうかと提案した。川口は、本報告の意義を評価したうえで、緊急事態宣言の発出後に人々の外出自粛の動きが広がった点について、①政府の要請に従うべきとの考えが働いたからなのか、②要請が出たことで恐怖心がおおられたからなのか、背後にあるメカニズムをどのように考えるべきかと尋ねた。植田は、政府による自粛要請が経済に与えた影響を分析するにあたり、政策当局はそのインパクトをかなり正確に予測出来ていたようにみえるとの認識を示したうえで、これには様々な非伝統的データの果たした役割が大きかったのではないかとコメントした。宮川は、勤務先がテレワークをどの程度許容するかによって、自粛要請に対する人々の反応は異なり得るのではないかと指摘した。また、恐怖が伝播していく過程とそれが解消していく過程で何らかの非対称性はないかとコメントした。新家は、日本においては、感染そのものよりも、感染した場合に周囲の視線が厳しくなるとの恐怖感が、外出減少に強く作用した可能性を指摘したうえで、これは本分析では情報効果に分類されるのかと尋ねた。また、感染者数が少ない県の方が(最初の感染者になることを恐れて)自粛する傾向が強かったようにも窺われるが、本分析でそのような傾向は検出されたかと尋ねた。高橋は、人々の恐怖はツイッターや検索サイトのテキストデータなどから計測し得るとし、それを説明変数に取り入れてはどうかと提案した。オンライン参加の河野敏鑑(専修大学)は、本分析の結果を踏まえると、「社会全体として感染症をどのように恐れるのが最適となるか」との問題意識を提示した。

報告者の渡辺は、まず、若田部のコメントについて、グローバルな経済の落ち込みの背景として、貿易などの波及経路の存在を否定するものではないと述べたうえで、世界的に同時発生している個人消費の低迷を説明する重要な要因として、恐怖心のグローバルな伝播があると主張した。また、感染症に対する人々の恐怖心が和らいでいる可能性や、検査体制が拡充されるもとで感染者数の定義が変わ

⁵ Goolsbee, Austan and Chad Syverson. 2020. "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020." NBER Working Paper 27432.

ってきていることも考慮しながら、情報効果の時間的な変化を捉える分析にも、今後取り組んでいきたいと応じた。川口の質問に対しては、本分析では外出自粛要請の効果を統計学的に検出したに過ぎず、その背後のメカニズムまでは明らかにできないと述べたうえで、おそらく政府の要請に従うというメカニズムの方が強かったのではないかと述べた。また、高橋の提案に対しては、人々が抱く恐怖心をテキストデータで捉える分析には今後取り組んでいきたいと応じた。新家の問いに対しては、日本のデータを用いた本分析と、米国のデータを用いた他の研究とでほぼ同様の結果が得られたことは、法的拘束力や周囲の視線の強さといった国ごとの事情とは無関係に、どの国においても、人々の感染症に対する恐怖心が強く働いたことを示唆しているのではないかと述べた。河野の問いに対しては、感染症に関する情報が人々の行動に大きな影響を及ぼしているとの本分析の結果を踏まえると、正しい情報をタイムリーに広めることが社会にとって重要ではないかと応じた。

4. パネルディスカッション

モデレーター：渡辺努（東京大学）

パネリスト：川口大司（東京大学）、新家義貴（第一生命経済研究所）
水門善之（野村證券）、辻中仁士（ナウキャスト）
宮川大介（一橋大学）、村上清幸（インテージ）
須合智広（日本銀行）

パネルディスカッションでは、モデレーターの渡辺が、①コロナ禍で明らかとなった非伝統的データの長所と短所、②非伝統的データの将来展望と課題、という二つの論点を提示した。これらの論点について、パネリストがそれぞれ発言を行ったあと、他の参加者も交えて議論が行われた。

4-1 非伝統的データの長所と短所

新家は、非伝統的データの長所について、速報性、高頻度に加え、伝統的データが十分に整備されていない地方経済の現状把握に有用となる可能性があるとして述べ、今後、政策効果の検証も含め、地方経済の分析が進んでいく可能性に期待を示した。一方、短所については、サンプルに偏りがある、データ期間が短い、季節性・曜日要因などの時系列的な特性が明らかでない、といった点を指摘した。こうした短所は、ユーザーが考慮しなければならないが、その認識は十分に進んでいないのではないかと述べた。

水門は、今回の感染症のように、経済に対して大きな、かつ急激な変化をもたらす状況下では、GPSデータをはじめとする非伝統的データの速報性の高さや

情報量の豊富さなどの長所が際立ったと指摘した。一方、短所としては、①母集団の偏りなどに起因する観測ノイズが存在すること、②観測データと観測主体の経済活動の関係性が変化する場合、データの解釈が難しいこと、③継続的な作成を目的としていないため、観測データに不連続性が生じること、を指摘した。

須合は、新家、水門が指摘する長所に概ね同意したうえで、感染症の経験のようなナラティブが、外生的なショックとして経済を動かす局面では、ニュースやSNSなどのビッグデータから得られる多種多様な情報に価値がある点を、長所として追加した。一方、短所については、伝統的データに比べ、マクロ経済、とくにGDP統計との関係が分かりにくいデータが多いこと、およびデータを加工・分析するためのスキルと時間がなお十分でない点を指摘した。

辻中は、クレジットカード情報に基づくJCB消費NOWやPOSデータに基づく日経CPINowを伝統的データと比較しながら、ビッグデータの長所である速報性や高頻度について具体的に例示した。また、マクロ分析以外にも、本年度前半、有価証券報告書の開示が遅延したなかで、個別企業の分析に有用であった点を長所として挙げた。もっとも、コロナ禍で、「オルタナティブデータ」という言葉が流行語として独り歩きし、正しい理解を欠いたままデータが利用されるケースが増えた点について警鐘を鳴らしたうえで、データ提供者として、バイアスなどデータの特性を丁寧に説明しながら、外部への開示を進めていく必要があると強調した。

村上は、非伝統的データが現状抱えている問題として、第一に、高粒度のビッグデータは、詳細な属性情報を含むため、個人・企業のプライバシー保護の観点からデータ提供者間で共有されにくく、様々なデータが断片的に提供されがちである点を指摘した。第二に、ビッグデータの集計は、適切な設計のもとで行われなければ歪みが大きくなり、データ利用者をミスリードしてしまいかねないと指摘した。

川口は、労働関連の速報性の高いデータとして、ウェブスクレイピングにより作成したハローワーク求人広告のデータを紹介し、労働需要をタイムリーに捕捉できる点を長所として指摘した。もっとも日本では、労働関連のビッグデータが少ない点は短所であるとの見方を示した。米国では、給与計算のクラウドデータを集計・加工した雇用や賃金のデータが、民間研究機関主導で公表されている現状について言及し、日本でも同様に整備されれば、労働経済の研究も進展すると述べた。

宮川は、短所として、単一の非伝統的データに依存して議論するのは危険であると論じた。裏を返せば、複数のデータを組み合わせて分析することによって、

価値が飛躍的に上がる点は、強みであると強調した。この際、複数のデータが、代替的かつ多面的にみるために使われているものなのか、補完的なものなのかという区分けが重要である点を指摘した。また、非伝統的データを利用する際には、ドメイン知識を含めた文脈の理解が重要である点も付け加えた。

以上のパネリストに対する意見に対して、モデレーターの渡辺は、非伝統的データの長所については、多くのパネリストで見解が概ね共有されていると整理した。そのうえで、短所として挙げられたサンプリングの問題やそのバイアスについて、情勢判断をミスリードするような事例がコロナ禍でみられたのか、と問うた。これに対して、水門は、百貨店やアミューズメントパークのように、高頻度データのカバレッジが、自粛や営業時間の短縮など感染症の影響を受けやすい企業に偏っている場合、そのデータの集計値とマクロの消費動向には乖離が生じる可能性を指摘した。また、新家は、コロナ禍における週次のPOSデータは、業態間で異なった動きとなっており、巣ごもり需要を背景に堅調であった食品スーパーなどのウエイトが大きい場合、必ずしもマクロの消費動向とは一致していない可能性があるとして付け加えた。その後、モデレーターの渡辺は、ビッグデータだけでなく公的統計にもバイアスがあるのではないかと問うた。これに対して、須合は、公式統計は統計的手法に基づいてサンプリングを行っているため、非伝統的データに比べればバイアスが少ないのではないかとコメントした。

4-2 非伝統的データの将来展望と課題

宮川は、非伝統的データで経済の実態を的確に捉えていく観点からは、本来、データが一元的に管理されることが望ましいものの、現状では政府や企業が個々にデータを保有・管理しているために、データの利活用が十分には進んでいない点を指摘した。今後、非伝統的データを用いて政策やアカデミックな議論を進めていくうえでは、こうした点を解決していくことが重要であると述べた。また、非伝統的データを使って意思決定を高度化していくためには、人材の発掘・育成・活用も必要であると指摘した。

川口は、非伝統的データが政策形成の現場で活用されるためには、統一的なフォーマットでデータが公開されることが期待されると述べた。その際、経済政策の効果について検証を行うためにも、個社や個人情報保護といった面に注意しつつ、より詳細なクロスセクションのデータの公表が望ましいと付け加えた。こうしたデータ整備には、大学や民間企業が一定の役割を果たしていく可能性があるとして指摘したうえで、その過程では、政府統計との整合性の検証も必要であると述べた。

村上は、非伝統的データを持続的に活用していくためには、どのようなデータ

利活用の「エコシステム」(データが囲い込まれることなく、社会全体として共有されながらも、データ提供者がビジネスとして採算を確保していくシステム)を作り上げていけばよいのか、という点を問題提起した。この際、ビジネスの世界で起きた自社の既存のデータ提供の経験を踏まえつつ、データ提供会社として、他社とも切磋琢磨しつつ、共通のフレームワークを提供することがエコシステム構築の一步になると論じた。こうした考え方を、金融や政策形成の世界にも拡大していけば、非伝統的データの利活用は進むのではないかと述べた。

辻中は、米国における非伝統的データの作成・利用状況について紹介しつつ、日本においては、非伝統的データのエコシステムを支える人材が、他国対比少ない現状を指摘した。そのうえで、コロナ禍で加速した非伝統的データの活用を促進していくため、金融機関や事業会社、政策現場においていかに活用事例を増やしていくのか、また、データ分析や関連するシステム構築に通暁した人材育成をどのように進めていくのか、という点が今後の課題になると述べた。

須合は、今回の感染症の経験やデジタルトランスフォーメーションの進展を踏まえれば、中央銀行が情勢判断に非伝統的データを活用することは、グローバルな潮流であると指摘した。そのうえで、非伝統的データを分析する恒久的なツールボックスを確立していくことが望まれると述べた。これを実現するためには、利用者がデータを丹念に分析すること、および学界・実務家と意見交換を密にすることが必要であろうと指摘した。

水門は、政策形成の場において非伝統的データの活用が進んでいくためには、先行きの情勢判断に必要となる企業の各種計画や見通しなど、フォワードルッキングな情報をいかに抽出していくかがポイントになると指摘した。課題としては、構造的な要因でデータの特性が変化している場合には、解釈が難しい点を挙げたうえで、伝統的データや定性情報、およびヒアリングなどの様々な情報を組み合わせながら総合的に判断していく必要があると論じた。

新家は、非伝統的データについては、景気判断だけでなく、Evidence-Based Policy Making (EBPM) の観点からも、今後、政策効果の測定も含めて利用が増えてくるとの展望を述べた。また、中期的には非伝統的データに関する理解も進むため、利用される比重は伝統的データを上回ると予想し、とりわけ、個別企業の分析などマイクロ分野での活用は一段と進んでいくと述べた。データ整備の観点では、政府も、ユーザー側としてだけでなく、メーカー側として、非伝統的データにどのように関わっていくか検討を進める必要があると論じた。

以上のパネリストに対する意見に対して、モデレーターの渡辺は、非伝統的データが持続的に提供・活用される「エコシステム」をどのように作っていくこと

が望ましいのか、と問うた。これに対して、辻中は、データ提供側の視点から、顧客が利用可能な形でデータをクレンジング・集計・加工する技術とマンパワーが不足している現状を述べた。そして、持続的にデータの顧客を確保するためには、データの特性に応じた効果的な活用事例を増やしていく必要があると付言した。村上は、個々のデータの生成から、それが統計化されるまでの過程において、標準化された枠組みを構築することが、非伝統的データの適正な活用や、エコシステムの確立に向けて重要であるとの見解を示した。宮川は、村上の見解に同意したうえで、適切にプライシングされた形でデータを流通させるシステムを作成することが重要であり、今後データの活用事例が蓄積されることによって、適正な価格評価が行われ、データ提供側のインセンティブも高まってくるのではないかと指摘した。川口は、「公共財」である現行の政府統計のように、コスト面から民間主体で提供するのが難しいものの社会的に重要と位置付けられる非伝統的データについては、公的機関による作成・提供が必要になるのではないかと述べた。こうした議論を受けて、貝塚は、民間主体が提供する非伝統的データの活用を促進するうえで、データの価格設定や公的補助の要否も含め、どのような費用負担構造が社会的に望ましいのかと質問した。これに対して、辻中は、民間データの多くは、元来別の目的をもって存在していることが多いため、それをマクロ分析の目的に転用するために生じる追加的なコストは公的機関が補助していくことで、データ利用が促進される可能性は高くなるのではないかと応じた。

5. ラップアップ

閉会にあたり、亀田は、今回のフォーラムは、エコノミストや経済学者だけでなく、情報工学系の研究者やデータ提供者も参加者に含まれていたことで、多面的な議論が可能となったと述べた。

今回のフォーラムで印象に残ったこととして、第一に、足もとの経済情勢が急速に変化する状況におけるビッグデータの有用性を挙げ、新型コロナウイルス感染症だけでなく、豪雨や台風といった自然災害時においても、景気の短期的な変動を把握するうえで、ビッグデータは有用であろうとの見方を示した。

第二に、ビッグデータの特性を理解したうえで実務に活用していくにあたっては、いわゆる人工知能などによる自動的な処理だけでなく、現状では、エコノミストや学者といった専門家の知恵が必要であることを挙げた。具体的には、パネルディスカッションで取り上げられた統計データのバイアスを例に挙げ、公的統計の性質については専門家が知り尽くしているのに対し、ビッグデータの性質についてはこれから理解を深めていく段階にあり、その過程では専門家のスキルが生きてくると述べた。そのうえで、将来、ビッグデータの特性や限界への知見が

蓄積されていけば、活用が一段と進んでいくであろうと述べた。

最後に、ビッグデータは可能性が広がっている領域であることから、今後も引き続きこの分野の研究動向をフォローしながら、様々な形でオープンな議論を行っていきたいと述べて締めくくった。

以 上

(別 添)



東京大学金融教育研究センター・日本銀行調査統計局共催

ビッグデータフォーラム

□日時 2020年11月9日(月)

□場所 日本銀行本店9階大会議室A

—— 論文報告者・パネリスト以外はオンライン参加

プログラム

司会進行 川本 卓司 日本銀行調査統計局
経済調査課長

13:00-13:05 冒頭挨拶 貝塚 正彰 日本銀行理事

第1セッション

13:05-13:45 位置情報データによる経済活動の把握

報告者 高橋 耕史 日本銀行調査統計局企画役
王 悠介 日本銀行調査統計局主査

13:45-14:25 交通移動データと非負値行列因子分解を用いた
都市活動分析と変化検知

報告者 前田高志ニコラス 理化学研究所
革新知能統合研究センター
因果推論チーム特別研究員

14:25-15:05 コロナ禍における企業ダイナミクス：
企業レベルデータと人流データを用いた実証分析

報告者 宮川 大介 一橋大学准教授

15 : 05 – 15 : 15 休憩

第2セッション

15 : 15 – 15 : 55 因果チェーンとネットワーク学習によるソースごとの COVID19
に関する捉え方の違いの抽出

報告者	久野 遼平	東京大学講師
	坂地 泰紀	東京大学特任講師

15 : 55 – 16 : 35 クレカ取引データと位置情報データを用いたステイホーム政策
の効果検証

報告者	渡辺 努	東京大学教授
-----	------	--------

16 : 35 – 16 : 45 休憩

パネルディスカッション

16 : 45 – 17 : 35	モデレーター	渡辺 努	東京大学教授
	パネリスト	川口 大司	東京大学教授
		新家 義貴	株式会社第一生命経済研究所 主席エコノミスト
		水門 善之	野村証券株式会社 シニアエコノミスト
		辻中 仁士	株式会社ナウキャスト 代表取締役 CEO
		宮川 大介	一橋大学准教授
		村上 清幸	株式会社インテージ 常務取締役 CMO
		須合 智広	日本銀行調査統計局経済調査課 経済分析グループ長
17 : 35 – 17 : 40	ラップアップ	亀田 制作	日本銀行調査統計局長