

ビッグデータの公的統計への貢献と課題

経済調査部 中田 理恵／溝端 幹雄

要 約

近年、公的統計は速報性などの課題に加えて、プライバシー保護や統計回答者の報告負担の重さなど調査の実施自体が世界的に困難となっており、こうした課題解決に民間で活用が進むビッグデータの貢献が期待されている。

公的統計へのビッグデータの貢献は「速報性」「報告者負担の軽減」といった既存統計の補完機能のほかに、大量のデータを用いることによる「詳細性」「網羅性」での改善も期待できる。一方、ビッグデータが抱える「プライバシー」や「継続性」「代表性」「安定性」など解決すべき課題も多い。

海外公的機関では、ビッグデータを扱える専門人材の育成や外部機関とのノウハウのシェア、データの共有・管理に向けた政策的支援や法的枠組みの構築に力を入れている。実際の活用では、カード決済データやウェブスクレイピングによる物価・消費・雇用統計への活用、税務統計などの行政データ、交通データ・位置情報データの活用が見られる。日本も経済産業省などで活用事例はあるが、現状は世界でも公的統計での活用事例は少ない。

ビッグデータ活用のガイドライン策定や統計の民営化の検討に加えて、幅広い分野で経済分析・予測へ応用するなどの経済情報の重層化が望まれる。

目 次

はじめに

1章 経済統計におけるビッグデータ活用の利点と課題

2章 経済分析から見た公的統計の課題と解決への道筋

3章 国内外公的機関におけるビッグデータ活用に向けた取り組み

おわりに

はじめに

ビッグデータは、民間レベルではマーケティング戦略など多方面で積極的に活用されるようになっており、データサイエンティストの活躍の場がますます広がっている。

またビッグデータは、公的統計の課題解決にも大きく貢献するのではないかと期待されている。なぜならば、公的機関により公表されるデータは速報性がないことや、プライバシーや統計回答者の報告負担の重さなどで様々な課題を抱えているからだ。こうした問題意識を受けて、公的統計でのビッグデータの活用に関してはここ数年、海外を中心に活発に議論されてきた。しかし、実際の活用となるとまだ緒に就いたばかりという印象を受ける。

そこで本稿では、今後の公的統計、特に経済統計におけるビッグデータの利用可能性について、これまでの議論と国内外での活用に向けた取り組みを概観する。そして、経済分析におけるビッグデータの活用について、公的統計を通して貢献できそうな点とそうでない点を明らかにしたい。

まず1章では、ビッグデータの特徴を指摘しつつ、経済統計への活用に向けたメリットとデメリットを整理する。続く2章では、現状の経済統計における課題を述べて、経済統計が抱える課題解決のために必要となるビッグデータ活用の留意点を指摘する。そして3章は、海外や国内公的機関等でのビッグデータの活用事例を紹介するとともに、海外事例から得られる日本への示唆を述べる。最後のおわりにでは全体のまとめを行う。

1章 経済統計におけるビッグデータ活用の利点と課題

1. ビッグデータに寄せられる期待

1) ビッグデータの特徴

ここでまず、ビッグデータの特徴を整理しておこう。ビッグデータの定義は明確に定まっているわけではないが、最もよく言及される特徴として3V (Volume、Variety、Velocity) を備えているデータとされている。すなわち、①ソーシャルメディアや携帯端末、センサー等から得られる大量のデータ (Volume)、②通常の数値で表される構造データに加えて、テキスト・画像・動画などの非構造データを含む多様なデータ (Variety)、③データの生成・流通・蓄積・処理・可視化の速度が非常に速いデータ (Velocity) ー という三つの特徴がビッグデータを特徴づけるとされる。

このことから、ビッグデータと言えば新しい種類のデータを想像するかもしれないが、実は政府が日々の行政運営を通じて得られる行政データ (administrative data) もビッグデータの範疇に含まれている。例えば、国連欧州経済委員会 (UNECE)¹ は、公的統計の立場からビッグデータを次の三つのタイプに分類・整理しており、以下で議論する上で参考になる (図表1)。

一つ目は「ソーシャルネットワーク (人間由来の情報)」がある。これはTwitterなどのソーシャルネットワーク (SNS) や、ブログ、Instagramなどの写真、YouTubeなどの動画、Google Trendsなどのインターネット検索、E-Mail が含まれる。

特徴としては非構造データであり、通常のデー

1) <https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>

タとして扱うには、デジタル化や形態素解析にかけるなどの前処理と呼ばれる膨大なデータ整形作業が不可欠となる点がある。さらに、テキストや画像・動画などをベースとしており、統計としての代表性（母集団の特徴を適切に反映した標本）という概念にはなじまない性質を持つ。

二つ目は「伝統的業務体系（処理仲介データ）」に分類される。例えば、政府の行政運営に伴い得られる税・医療記録といった業務統計や、民間企業が日常の業務から入手した様々な取引やクレジットカードでの支払情報などのビジネス関連の統計が該当する。

このデータは構造化されているために比較的扱いやすく、きめの細かい情報が入手できるという意味で詳細性（granularity）に優れている。その一方で、機微に触れるデータが多く、プライバシーの観点から扱いが難しい点、必ずしも代表性を持つデータではない点、そして統計の作成者により手法や定義が多様であり、データの更新が突然中止されるリスクがあるなどの欠点を持つ。

三つ目は「IoT（機器生成データ）」と呼ばれるものである。家庭、気象、交通、監視などのセキュリティといった様々な場面で設置される固定センサーから得られるデータや、スマートフォン・車載センサー・衛星などの移動センサーから得られる位置情報データ、そしてコンピューターやウェブのログ情報などが含まれる。

このデータはセンサー等から大量の情報が得られ、しかも構造化されているというメリットがあるため、詳細性に優れている。しかし、大量の情報を蓄積するストレージの問題や分析に高度な情報処理技術が必要なこと、やはり統計としての代表性に乏しいというビッグデータの持つ欠点を免れない。

図表1 ビッグデータの分類

ソーシャルネットワーク【人間由来の情報】	
ソーシャルネットワーク (Facebook、Twitter、LinkedIn 等)	
ブログ・コメント	
個人書類	
写真 (Instagram 等)	
映像 (YouTube 等)	
インターネット検索 (Google Trends 等)	
モバイルデータの内容 (テキストメッセージ)	
利用者により生成された地図	
E-Mail	
オンラインニュース	
伝統的業務体系【処理仲介データ】	
公的機関により生み出されたデータ	
行政データ (税・医療記録 等)	
業務により生み出されたデータ	
商業取引	
銀行・株式取引の記録	
E-commerce	
クレジットカード	
ビジネスウェブサイト (求人情報 等)	
スキャナーデータ	
IoT【機器生成データ】	
センサーからのデータ	
固定センサー	
ホーム・オートメーション	
天気・汚染センサー	
交通センサー・ウェブカメラ	
科学センサー	
セキュリティ・監視ビデオ・映像	
移動センサー【トラッキング】	
携帯電話の位置 (GPS)	
車	
衛星画像	
コンピューターシステムからのデータ	
ログ	
ウェブログ	

(出所) UNECE、Hammer et al. [2017] から大和総研作成

2) 調査統計の難しさとビッグデータの可能性

そもそもビッグデータは、目的となる事項を政府自らが標本調査や全数調査をして得られる「調

査統計（第一義統計）」とは異なり、特定の業務を遂行する際に得られるデータや、機器の利用に伴って付随的に得られるデータであるため、「業務統計（第二義統計）」に近い性質を持つと言える。さらには、通常のデータの形態とは異なるテキストや画像を統計処理が可能な形に変換しており、一種の「加工統計」の様相も呈している。

近年、プライバシー保護や調査員のなり手不足、統計回答者の報告負担の重さ等の問題もあり、国内外で公的統計の中心的役割を担ってきた調査統計（第一義統計）の実施は年々困難さを増している²⁾。そのため、行政にとって入手しやすいがこれまであまり活用されてこなかった、業務統計（第二義統計）の活用が叫ばれている。ビッグデータも業務統計の範疇にあると考えられるため、公的統計の現状の課題解決のためにその活用が検討され始めている。

2. ビッグデータが抱える課題

経済統計への活用が期待されるビッグデータであるが、従来の統計データとは異なる性質を持つことから、その活用を進めるには以下のような多くの解決すべき課題がある。

1) ビッグデータ自身が抱える課題

(1) 代表性・信頼性

従来の統計指標作成においては、統計学の知見に基づいて必要なサンプル数を計り、データ収集、加工を行っている。一方で、ビッグデータは統計作成を目的とせず、収集されたデータである。このため、サンプルの偏りやサンプル数の大幅な変化などが発生し得るため、対象とするターゲットの全体像を正確に表しているとは限らない。例

えばクレジットカードによる支払いデータはサンプルに占める若年層や高齢者層の割合が低いため、これらの年齢層の購買行動の変化を軽視してしまう可能性がある。また、SNSへの投稿内容では、注目されることを目的とした虚偽の情報が時として広く拡散されることもあり、虚偽の情報の扱いについても考慮せねばならない。

(2) 継続性

民間企業が生成するビッグデータを利用する場合は、企業の意向によりデータ公開が停止される可能性がある。また、継続して公表されているデータであっても、そのサンプル数およびサンプルの特性は時系列に沿って大きく変化し得る。例えば特定のサービスのユーザー情報をサンプルとした場合、サービス開始直後のユーザー数の急増や、代替サービスの登場によるユーザー数の急減が起こり得る。また特定の層に向けたキャンペーンによるサンプルに占める特定のユーザー層の増加なども考えられる。

(3) 比較可能性

従来の統計情報は、国際的な基準を踏まえてデータ収集・加工されているため、国際比較が可能である。一方で、ビッグデータは企業に依存するため、同様の情報であっても国ごとにデータを保有している企業が異なり、データの収集・管理体系が異なっており、国際比較が難しい。

2) ビッグデータの利用で生じる課題

(1) データ取得費用

多くの場合、ビッグデータと呼ばれるデータは民間企業が保有している。もとより外部への提供

2) 例えば、Jarmin [2019] や梅田・宇都宮 [2009] を参照。

を前提としていない場合や、有償での提供を前提としている場合は取得に伴うコストが発生する。

統計は性質上、広く一般向けに無償提供されることが前提となっており、データを保有する企業側としては公的機関に無償で提供する意欲は低い。企業側にデータの共有を促す何らかのインセンティブ付けが必要となる。

(2) 個人情報の扱い

ビッグデータの一部には、スマートフォンの位置情報や、個人の支払いデータなど極めてセンシティブなデータが含まれる。こうしたデータを統計として二次利用する際は、情報倫理に基づいたルールの明確化が求められる。

(3) 専門人材の不足

ビッグデータを扱う場合は、ITインフラの整備や、各種領域に関する専門知識を持った人材が必要となる。一方で、経済産業省〔2019〕によると、2018年時点でIT人材(IT企業及び、ユーザー企業の情報システム部門等に属する職業分類上の「システムコンサルタント・設計者」、「ソフトウェア作成者」、「その他の情報処理・通信技術者」)は22万人不足している状態にあり、2030年には45万人不足する見込みである。

(4) 設備

大規模データを保管・処理するためにはIT環境の増強が求められる。十分な分析環境を整備するために必要となるリソースおよびこれに伴う費用の明確化、予算の確保が求められる。

2章 経済分析から見た公的統計の課題と解決への道筋

1. 日本における公的統計の整備

公的統計は変化していく経済社会の実態を数値で表現するものである。経済政策をはじめ、事業計画の立案や投資家の判断材料など幅広い場面で公的統計は活用されている。こうした公的統計の適切な運用には、経済社会の実態が正確に反映されるように統計の内容を不断に見直すことが必要だ。

政府は公的統計の整備体制を強化するため、従来の統計法(1947年公布)を全面的に見直し、2007年に「新統計法」を公布した(施行は2009年4月)。また2007年には、政府の統計整備の司令塔を担う「統計委員会」が設置された。さらに2009年には「公的統計の整備に関する基本的な計画」(以下、基本計画)を閣議決定し、社会経済情勢の変化に応じておおよそ5年ごとに基本計画を見直すこととされた。直近では2018年3月に「第Ⅲ期基本計画」を策定し、2020年6月には変更されている。

公的統計が対象とする範囲は経済・社会と幅広く、また統計の作成者・協力者や利用者など様々な立場の論点があり得る。しかし、紙幅が限られることもあり、2章では景気動向などを分析する経済統計の利用者の立場に焦点を当てて、新統計法や基本計画にも言及しつつ、公的な経済統計で課題とされている点を整理したい。

2. 経済分析における公的統計の課題

1) 速報性 (timeliness)

経済分析の立場より以前から指摘されるのは、公的統計に速報性がない点である。月次・四半期

の頻度で公表される動態統計（時系列統計）の場合、主に当該月（四半期）の翌月末以降に統計が公表されるため、ほとんどの指標において経済実態の把握に少なくとも1カ月のタイムラグが生じてしまう。例えば、代表的な経済統計である鉱工業生産指数（経済産業省）や消費者物価指数（総務省）の場合は、約1カ月後の公表になる。比較的公表が早い企業物価指数（日本銀行）でも公表は10日～半月後であり、経済統計の中で最も注目される四半期別GDP速報（1次速報、内閣府）の場合には、他の経済統計を加工して作られる（加工統計）ため、1次速報値の公表は当該四半期終了月の翌々月半ばまでずれ込む。

こうした速報性の欠如は、経済実態の把握を遅らせることで、財政・金融政策による迅速な対応の妨げとなるだけでなく、投資判断や経営判断にも影響を与える。さらに今回のコロナ禍で足元の経済実態を早期に把握するニーズが強まる中で、速報性の欠如は公的統計の大きな課題として再認識されつつある。

2) 詳細性 (granularity)

経済全体といったマクロな経済統計は多いものの、地域（都道府県・市区町村）や個人・企業などのよりミクロな経済統計へのアクセスは難しいのが現状である。前者の場合は地域ごとの集計によるサンプル数の減少で集計精度の低下（標本誤差の拡大）の懸念があること、後者の場合はプライバシーの観点からその利用が制限されている。さらに、そもそも既存統計では把握しづらいものもある。例えば、旅行など人々の移動や財・サービスの地域間取引といった、ネットワーク関連のデータは不十分である。

最近では基本計画等を受けて、個票データへのアクセスが幾分改善されつつあるが、プライバシーの問題もありその利用には限界もある。経済主体の異質性や空間的な観点から経済社会を把握するために、より詳細な公的統計の利活用が望まれる。

3) 網羅性 (completeness)

経済社会構造の変化が激しくなると、既存統計ではカバーできない分野が多くなる。例えば、サービス・eコマースに関連する統計、そしてストック関連の統計はその不足が指摘されている。これらは測定ミス (mismeasurement) 問題と呼ばれている。

経済理論上、考慮されるべき変数が統計として把握できない場合、何らかの代理変数や簡易推計が使われるが、それには当然ながら誤差も多くなるため、実態の正確な把握が困難となっている。

4) 報告者負担 (burden on respondents)

利用者側から見た直接的な課題とは言えないが、報告者負担の重さの問題が挙げられる。これは統計作成者側の重要な論点であると同時に、統計の「速報性」や「詳細性」を妨げる原因にもなっており、利用者側からも重要な論点と考えられる。

例えば、複数の類似統計が存在する場合、同じような内容の記入を求められるために、報告者の負担は重くなりやすい。この点に関しては、日本商工会議所・日本経済団体連合会・経済同友会の加盟企業を対象にした行政手続の負担に関するアンケート調査でも³⁾、営業の許認可、社会保険・納税、補助金といった行政手続に次いで、「調査・統計に対する協力」の負担が大きいと指摘してい

3) 内閣府規制改革推進室「事業者に対するアンケート調査の結果の取りまとめ」（平成29年1月19日）

る。

5) その他

その他、共働き世帯の増加で調査を引き受けない世帯が増加していることや、国勢調査のように各調査区に割り当てられる調査員のなり手の減少といった課題も生じている。さらに家計調査のような調査統計では、プライバシーを理由に調査対象者から調査を拒否されるなど、年々、調査統計の実施が難しくなっている。このような経済・社会環境の変化の観点からも、公的統計の在り方を再考する時期に来ている。

こうした背景を受けて、基本計画でも効率的な統計作成のための行政記録情報等の活用や民間事業者の活用が提言されているところである。

3. 経済統計の課題解決におけるビッグデータ活用の留意点

以上の議論を踏まえて、経済統計が抱える課題をビッグデータは解決できるのだろうか？

ビッグデータに最も期待されるのは、「速報性」「報告者負担の軽減」といった既存統計を補完する役割と考えられる。また、大量のデータを用いた「詳細性」「網羅性」にも一定の役割が期待できそうだ。

一方で、ビッグデータが抱える「プライバシー」や「継続性」「代表性」「安定性」といった課題には解決すべき点が多そうだ。これらの課題は官民協力により取り組むことが求められる点である。

それでは現在、海外そして国内の公的機関ではどのような取り組みが行われているのであろうか。次の3章ではそうした事例を紹介する。

3章 国内外公的機関におけるビッグデータ活用に向けた取り組み

1. ビッグデータの活用パターン

ビッグデータを公的統計に活用する形式としては次の三つが考えられる。

まず、①既存統計指標の完全代替としての使用である。既存の統計指標の基となるサンプルデータをビッグデータに置き換える。この場合、データ収集の手間が省かれるため、2章で述べた現在の公的統計が抱える課題の一つである報告者負担を軽減することができる。一方で、既存統計のサンプルと異なり、データは本来の目的と異なる用途での使用になるため、データが既存のサンプルが有していたような代表性を失わないことが求められる。

次に、②既存統計の補完としての使用である。既存の統計指標の基となるサンプルに加えて、ビッグデータから抽出される情報を組み込むという形式である。2章で述べた現在の公的統計が抱える課題の一つである「網羅性」にアプローチすることが可能となる。例えば、既存の物価統計にインターネット上に掲載されている商品の販売価格情報を加えることなどが考えられる。この場合は財・サービス全体ではなく、一部の分野に関してサンプルが十分にその分野全体を代表できればよいため、①の形式と比較してデータの代表性についての要件は緩和される。

そして、③新たな統計指標の作成である。既存の統計指標にない新たな指標の作成にビッグデータを活用することも考えられる。

活用パターンとしては上記の三つが考えられるが、1章で挙げたようなビッグデータの課題(デー

タの代表性、継続性)を踏まえると、②既存統計の補完としての使用、③新たな統計指標の作成—などが現実的な路線となろう。実際に後述する活用例でも②、③のパターンでの活用を企図したものが多。

2. 国内外公的機関での活用事例

では、実際にビッグデータを公的統計に活用している国では、2章で挙げたような課題にどのように取り組み、どのようにデータを活用しているか。下記では、主に英国とEUおよび米国の例とともに日本の例も紹介する。

1) 英国での取り組み

(1) データ利活用に向けた取り組み

英国では国家的なデータ戦略を模索している段階にあり、2020年中に公表を予定している。包括的な戦略は未公表だが、既にデータの利活用に向けた様々な取り組みを行っている。専門人材の獲得には特に注力しており、育成・雇用・ノウハウのシェアという三つの面からアプローチしている。

・人材育成

英国国家統計局は2017年にデータサイエンスキャンパスを設立した。当該機関は、データサイエンスの領域における人材育成、およびビッグデータの活用方法の調査を担っており、国内外の学術機関および民間企業との幅広い協業を目指している。ビッグデータ解析や機械学習分野などで使用されるプログラミング言語であるRやPythonの講座など多様な教育プログラムを開設するとともに、ビッグデータを活用した指標作成に向けて複数のプロジェクトを行っており、進捗やプログラムコードを公表している。

・外部からの雇用による人材獲得

英国政府は専門人材の獲得のための施策として、“Digital, Data and Technology Profession Capability Framework”を作成、ITスキルを有する人材に向けて政府機関が必要とするスキルと職務内容およびキャリアパスを明示している。職務は39部門に細分化されており、職務ごとにキャリアパスにおいて必要とされるスキルが明確化されている。これにより、ITスキルを持った専門人材に対して自身の能力を十分に発揮し伸ばすことができる環境があることをアピールしている。

・ノウハウのシェア

育成や雇用といった手段のほかに、外部とのノウハウのシェアも専門人材の知見を取り込む有用な手段の一つである。英国では各国中央政府に勤めるデータサイエンス領域の専門人材向けにコミュニティ(cross-government data science Slack channel)を開設しており、海外との間でアイデアや課題のシェアを促進している。コミュニティでは毎月データサイエンスに関するニュースレターを発信しており、現在は600名以上の会員が集まっている。

(2) 活用事例

先述した通り、データサイエンスキャンパスではビッグデータを活用した新たな知見の獲得や、新たな経済指標の開発を行っており、それぞれプロジェクトごとに進捗や成果を公表している。

・付加価値税(VAT)申告額による景気動向の早期把握

景気の変化を早期に把握することを目的に、付加価値税(VAT)の申告額という行政データを基にした月次指標を2019年4月より公表開始した。なお、正式な統計指標ではなく、試験的に公

表する指標とのことわりがある。公表されている指標は売上のディフュージョン・インデックス、支出のディフュージョン・インデックス、新規申告者数、申告内容別の申告数などである。いずれも報告対象月から一カ月以内に公開されている。過去12カ月の課税対象取引額が8万5,000ポンドを超える企業は、申告が義務化されているため、データには一定の代表性がある。加えて企業からの申告内容に誤りがある場合は処罰されるため、データに正確性がある。また、25万社からデータを集めているため細分化した分析が可能である。

開発プロジェクトメンバーによる分析ではこれらの指標は景気の大変な変化を迅速に反映する傾向にあるため、景気の転換点の早期把握に有用であるとしている。また、業種別の表示が可能であるため、各業種における指標の変化が指標全体に対しどれほど影響を与えたかを示すことができる。加えて、申告内容別の申告数には経済的なストレスとの関係性も見られた。例えば金融危機が発生した2008～09年は払い戻しの申告等を行った企業が多く、2009年には新規申告者が少ない傾向にあった。

・貨物車の交通状況による物流・貿易等の早期把握

国内の物品取引および貿易の動向を知ることが目的に2019年4月より公表を開始した。なお、VAT申告額と同様に正式な統計指標ではなく、試験的に公表する指標とのことわりがある。

大型車の交通量の推移は複数の経済イベントと一致しており、物流や人の移動が重要となる領域での経済的な変化に対して早期に反応する指標に活用できる見込みがある。公表されている指標は長さ7～12メートルの車両と12メートル超の

車両の交通量の推移である。いずれも報告対象月から2カ月後に公表されているが、統計局ではより早期の公表を実現する方法を模索している。

・財務データ等に基づく金融機関の分類の自動化

金融機関の開示データを基に自動的かつ詳細に業種分類を行うことを企図したプロジェクトである。現況の金融統計上では金融機関の分類は三分類のみであるが、これを詳細化することで監督機関がより正しく金融の安定性やリスクを捉えることができるとの考えに基づくものである。具体的には企業の開示する財務諸表などのデータから詳細な業種分類であるStandard Industrial Classification 2007（S I C 2007）に当てはめることを試みた。一般的に多くの企業はS I C 2007上での分類を定められているが、非常に労力のかかる作業であり、なおかつ適切に分類されているかを再評価する機会は少ない。企業の業態は変化するものであることを踏まえて、定期的かつ機械的に分類することで、金融機関のリスクをより精緻かつタイムリーに把握することに役立てるとしている。分析の結果では、入手可能なデータのみでは完全な自動化は難しいが、特定の企業の分類に対する異常検知は実現可能であることが分かった。

2) EUでの取り組み

(1) データ活用に向けた戦略

EUは、域内におけるデータ利活用のための戦略として“A European strategy for data”を公表しており、戦略の中心として四つの柱を立てている。

一つ目の柱はデータ利活用のための分野横断的なガバナンス体制の形成である。ここでは具体策として3点挙げられている。一点目はEU域内で

共通のデータガバナンスに関する仕組みづくりである。その仕組みは欧州議会が策定した新たな個人情報保護の枠組みであるGDPR（EU一般データ保護規則）に準拠するとされている。二点目は企業、特に中小企業に向けてより高品質な公共部門のデータを公開することである。三点目はデータ共有促進に向けた法的枠組みの模索である。公的機関と企業、および企業間のデータ共有促進のための法的な枠組みの在り方や、特定の状況下においてデータへのアクセスを強制的に可能とすること等が議論されている。

こうした取り組みは1章で掲げた、個人情報保護やデータの信頼性、比較可能性などの課題にアプローチするものと捉えることができる。

二つ目の柱はデータ保管・処理のためのインフラ設備への投資である。2021年から27年までの間に欧州委員会はデータスペースとクラウド設備に対する投資を積極的に行うとしており、プロジェクト全体では総額40億～60億ユーロが投資されると見込んでいる（うち欧州議会は20億ユーロの拠出を見込む）。

三つ目の柱は個人データに関する権利の強化、専門的人材の育成とデータリテラシーの向上などである。

四つ目の柱は戦略的セクターや公共の利益に関する領域における欧州共通のデータスペースの創造である。対象となるセクター、領域については、製造業、グリーンディール、移動データ、ヘルスケアなどが掲げられている。

その他、EU域内での個別国における特徴的な取り組みとしてはフランスが挙げられる。フランスでは、2016年にデジタル共和国法（the Law for a Digital Republic）が公布されて、民間企業が政府にビッグデータを提供する場合にはその

継続的な提供を義務付けている。

（2）活用事例

European Statistical System（ESS）ではESSnet Big Dataプロジェクトと称して、ビッグデータを公的統計に活用することを目的としていくつかのプロジェクトを遂行している。

・オンライン求人情報による雇用統計の拡充等

オンライン求人サイトの情報をウェブスクレイピング（ウェブ上から情報を収集すること）する、または第三者から取得するという二つのアプローチでデータ収集を行い、既存の雇用統計のカバー範囲の拡充に活用できるかを検討したプロジェクトである。結果としてこうした新たなデータは一定の目的において有用であるとされたが、雇用情勢全体を代表することは難しく、従って既存の雇用統計のサンプルの代替とすることは難しいとの結論に至った。しかし、一部の職業、スキルに対する需要を示すことについては有用との結果であった。また、時系列に沿った需要の変化を捉えるには有効であるとの結論から、足元の動向を追うナウキャストニングに応用できる見込みがある。

・AISデータを活用した景気先行指標等の可能性

プロジェクトでは既存の海運統計と比較した船舶自動識別装置（AIS）データの有用性の検討、および新たな指標作成に向けたデータベースの作成などが行われた。結果、AISデータを活用することで既存統計よりも正確に船舶の入出港を捕捉することが可能との結論に至った。AISデータを活用した港内の移動距離および港間の移動距離の計測にも成功しており、今後の有用な統計になり得るとしている。また、AISデータから船

船によるガス排出量などを特定し統計的に計測することが可能となる見込みがあるとしている。新たな統計指標の開発プロジェクトも進んでおり、船舶の入出港を基にした経済の先行きを示唆する先行指標の開発、船舶の運航ルートの変更を基にした規制の変化や港での異常発生の検知を担う指標の開発などが試行されている。

・スマートメーターによる企業・家計動向の把握や空き家の動向に関する新たな指標も検討中

また、進行中のプロジェクトとして、ウェブスクレイピングを基にした企業データベースの作成、スマートメーターのデータ活用などが挙げられる。スマートメーターは欧州議会が導入を推進しており、これらのデータを活用することで、企業活動や家計の動き、空き家などに関する新たな指標の作成を模索している。

・モバイルデータによる観光統計の作成コスト低減・精度向上・速報化

上記のE S Sでの取り組み以外の事例として、エストニアでのスマートフォンの位置情報を活用した観光統計の作成が挙げられる。現在公表している指標としては、入国者数、滞在日数、目的地での宿泊日数などがある。Hammer et al. [2017] が指摘するように、この統計は旅行収支等の推計にも応用が期待される。旅行者の携帯電話の位置情報の活用により、従来のサーベイデータよりもデータ収集にかかるコストが削減できただけでなく、データの質と正確性、速報性の向上が実現されたとしている。

3) 米国での取り組み

(1) データ利活用に向けた取り組み

米国は2016年に大統領の諮問機関である国家科学技術会議が“THE FEDERAL BIG

DATA RESEARCH AND DEVELOPMENT STRATEGIC PLAN”を公表し、ビッグデータの研究開発に関する計画を示した。同計画では七つの方針が打ち立てられた。①ビッグデータに関わる新興技術等を活用し次世代の可能性を創出する、②データおよびそこから得られる知見の信頼性把握のための研究開発支援、③サイバーインフラストラクチャーの増強、④データの共有、管理を促すための政策的支援、⑤プライバシー、安全性および倫理に配慮したデータ収集・共有への理解、⑥ビッグデータに関する教育・訓練の改善による人材育成、⑦国家的なビッグデータイノベーションエコシステムの構築——である。

また、政府保有データの公開・活用に関しては2018年に“OPEN Government Data Act”を制定しており、政府機関がデータを公開可能な場合は、機械可読形式で公開し、オンラインカタログ上に掲載することを必須としている。また、2020年には「連邦データ戦略」が公開されている。三つの領域に対して合わせて10の基本原則を示し、これらに基づいた20のアクションを定義している。また、各アクションについて達成期限と達成状況をウェブサイト上で公開している。基本原則の内容は図表2の通りである。

(2) 活用事例

・カード決済データに基づく地域別消費の早期把握

2019年に連邦準備制度理事会の研究チームから発表されたワーキングペーパー(Aladangady et al. [2019])では、カード決済データに基づく消費統計の作成が試みられている。消費動向を示す統計としては既に小売売上高といった統計があるが、頻度が月次であること、データ公表まで

図表2 連邦データ戦略の基本原則

倫理的ガバナンス	
倫理の維持	連邦データの公開が公衆に与える影響を監視、評価する。
実行責任	効果的なデータ管理の実践。健全なデータセキュリティーの採用、個人情報保護、適切なアクセス管理。
透明性の促進	政府保有データの目的と用途を明確にし、国民の信頼を獲得する。
意識的なデザイン	
関連性の確保	データの品質と整合性の維持。
既存データの活用	優先的な政策課題に取り組む際に必要なデータの特長。必要に応じた追加のデータ取得。
将来的な使用の予期	他のユーザーによる使用可能性に配慮した再利用しやすいデータ作成。
応答性の実証	データ収集・分析・普及の改善。フィードバックプロセスの確立。
学習文化	
学習への投資	データ管理設備と人材への継続的な投資。
データリーダーの育成	連邦政府職員のあらゆるレベルにおいて、公共の利益のためのデータの価値に関してトレーニングを行い、データリーダーを育成する。
責任の実践	責任の明確化、データ管理に関する監査および文書化、必要に応じた変更。

(出所) Federal Data Strategy Development Teamから大和総研作成

2週間かかること、頻繁にリバイスされること、地域別の分析ができないことが課題となっており、カード決済データの活用がこれらの課題を克服できる可能性があるとしている。地域ごとの消費データをリアルタイムで取得することができれば、各地域における経済ショックへの監視が容易となり、ショックが国内全体に広がることを未然に防ぐことにも役に立つ。

データは米国の電子決済サービス大手であるファースト・データ（当時）から取得しており、同社は年間2兆ドルの取引額を誇る。統計的なサンプリングを行ってデータを収集している小売統計と異なり、ファースト・データから提供される取引データは必ずしも全体を表すように構成されていない。なおかつ、加盟店の増減に影響を受け

る。分析では日次ないし月次指標の前年比伸び率を作成することを目標にしており、サンプルは分析対象とする月を含めて14カ月連続で加盟していた店舗に限定して、指標を算出した。なお、14カ月としているのは前年比を取る際に月の途中で加盟した店舗も除外するためである。調整後の取引データから得られた消費の推移は小売統計と高い相関を持つことが確認され、取引データは代表性を十分に有しているとされた。また、取引データによる消費の推移は対象月の終了から最短3日後に取得できるため高い速報性を有する。日次、地域別の消費データの活用方法として、分析ではハリケーンの影響に対する影響の可視化などに有用であることが示されている。

・ビッグデータ活用によるCPI作成コストの低減等

労働統計局では消費者物価指数(CPI)のデータソースにおける民間データの活用を模索している。目的としては、CPIのカバー範囲の拡大、作成コストの削減などが掲げられている。新たなデータソースには様々なものがあるが、大きく分類すると民間企業が保有するデータ、2次データ（民間企業が第三者から収集し整理したデータ）、ウェブスクレイピングなどで独自に収集したデータの3種類である。

既に試験的な公表が始まったものとしてはJ.D.パワー社が保有する新車販売の取引データが挙げられる。従来の新車販売額のデータは回答者の負担の重さ、回答率の低さ、データ収集にかかる費用の観点から民間データとの置き換えを企図されていた。J.D.パワー社の保有データは国内の新車販売取引の3分の1をカバーしており、なおかつ対象とするマーケットがCPIの集計対象と似通っており、ビッグデータの統計活用

の課題の一つであるデータの代表性の課題をクリアできる。分析では新車販売の取引データから、C P I の新車販売額の推移をおおむね再現できることが分かり、またデータソースを J . D . パワー社のデータに置き換えることでデータの収集コストがわずかに低下することが示された。労働統計局は 2019 年 5 月から J . D . パワー社のデータを基にした新車販売額の推移を公表しており、正式なレビューを経た後に、C P I のデータに J . D . パワー社のデータを反映するとしている。またこれに付随してサンプルサイズの拡大、新車販売におけるリアルタイムの支出情報の獲得も達成されるとしている。

4) その他の取り組み

国連は、公的統計でのビッグデータの活用に関する協議の場として Global Working Group (UN GWG) を設けており、各国の統計当局や中央銀行などが参加している⁴。さらに国連のウェブサイトは、各国の公的統計におけるビッグデータ活用の取り組みをまとめ、Big Data Project Inventory を掲載している⁵。

また、マサチューセッツ工科大学 (M I T) による Billion Prices Project (B P P) と呼ばれる取り組みでは、ウェブスクレイピングにより収集した価格データからインフレ率の計測誤差を測っている。このプロジェクトの結果、アルゼンチンのインフレ率には計測誤差があることが明らかとなったが、米国やユーロ圏では計測誤差は非常に小さく、アルゼンチンを除けば、総じてウェブスクレイピングにより物価指数の計測誤差を修正する効果はなかったと結論付けている。

5) 日本での取り組み

(1) データ活用に向けた戦略

まず、総務省により 2016 年に家計調査を含む消費統計の改善を図るべく「家計調査の改善に関するタスクフォース」が開催された。また、総務大臣主宰の「速報性のある包括的な消費関連指標の在り方に関する研究会」(以下、消費研究会)でもビッグデータの活用も含めた新たな消費統計の開発が検討された。そして 2017 年には「消費動向指数研究協議会」が発足し、民間企業が持つビッグデータの消費統計への活用が本格的に議論されている。

内閣府でも、2016 年の「より正確な景気判断のための経済統計の改善に関する研究会」において、S N A の改善に必要な統計の見直しに加えて、行政記録情報を含むビッグデータを使った経済統計の改善や景気動向の把握について検討された。ここでの議論を受けて、2016 年に経済財政諮問会議で「統計改革の基本方針」が決定され、内閣府に「統計改革推進会議」が 2017 年 1 月に設置された。

(2) 活用事例

・消費動向指数 (C T I ミクロ・C T I マクロ) の作成

上記の総務省の消費研究会での検討の結果、公的統計の課題を克服するための新たな経済指標として消費動向指数 (C T I) が開発された⁶。これは、単身世帯も含めた世帯の平均的な消費を月次で捉える「世帯消費動向指数 (C T I ミクロ)」と、G D P 統計のマクロな消費総額を月次で把握する「総消費動向指数 (C T I マクロ)」で構成される。

4) <https://unstats.un.org/bigdata/>

5) <https://unstats.un.org/bigdata/inventory/>

6) 詳細は阿向 [2019] を参照のこと。

C T I ミクロは、単身世帯の月次消費を把握するために「家計消費単身モニター調査」を使っている。この「家計消費単身モニター調査」は、有意抽出のために代表性の問題（バイアス）があるが、これを修正するために傾向スコアという統計的手法を使うという新たな試みを行っているのが特徴である。

また、C T I マクロは時系列分析の一つである状態空間モデルを活用して、GDP 統計の消費動向を月次で推計するものである。これはビッグデータを活用しているわけではないが、新たな統計的手法を公的統計に応用して、より高頻度（ここでは月次）でGDP 統計と整合的な消費を把握する試みと考えられる。

・POSデータ・決済データによる消費動向の把握、SNSを使った鉱工業生産予測指数、人の動きの活発度

一方、経済産業省では、POSデータやTwitter等のSNSによるビッグデータを活用した新たな経済統計として、試験的に「BigData-STATS ダッシュボード（β版）」を公開している⁷。また、内閣官房・内閣府でも、足元の地域経済把握のためのビッグデータ関連の経済統計をまとめた「V-RESAS」を試験的に公開しており⁸、POSデータや決済データによる消費支出の地域ブロックでの把握や、位置情報による人の流れ等をリアルタイムで把握することができる。

3. 海外事例から得られる日本への示唆

1) 海外事例から学ぶべき点

海外の動向を見ると、まずデータ活用に向けた戦略では、どの国・地域もビッグデータを扱える

専門人材の育成や外部機関とのノウハウのシェアに注力している。さらに、データの共有・管理に向けた政策的支援や法的枠組みの構築が挙げられる。日本でもこうした点で本格的な取り組みが求められる。

次に活用事例については、物価統計におけるビッグデータの活用が主流だ。スキャナーデータによるモノの価格・数量の収集や、航空・宿泊などのウェブサイトに掲載される料金をウェブスクレイピングにより収集したビッグデータを物価統計に活用する事例が見られる。また、カード決済データやウェブスクレイピングから消費や雇用統計を充実する方向での検討も多い。

一方、日本への示唆という点では、税務統計などの行政データを公的統計に活用することが今後は重要になるだろう。さらに、交通データや位置情報データを活用した物流・観光統計の充実も検討に値する。これらのデータを公的統計で活用するのは比較的容易であると考えられ、日本でもそうした活用が期待される。

しかし、ビッグデータの公的統計への活用には統計の質や国際比較が困難な面など多くの課題が残されており、海外でもまだ十分活用されていない。

2) 日本での新しい試み

総務省の「世帯消費動向指数（C T I ミクロ）」の例で見たように、傾向スコアを使ってバイアスのあるデータを補正するなどの新しい試みが日本で行われている。阿向 [2019] は、バイアスのあるデータを欠損値のあるデータと捉えるならば、現在、統計学で応用研究が進められている「欠

7) https://www.meti.go.jp/statistics/bigdata-statistics/bigdata_pj_2019/index.html

8) <https://v-resas.go.jp/>

損値の推定」に関する学術的成果を援用できるのではないかと述べており、ビッグデータの課題である代表性の問題に対処できる可能性を指摘している。こうした新しい試みが突破口となり、ビッグデータが様々な公的統計に活用されることが期待される。

3) 今後取り組むべき課題

今後、取り組むべき課題としては、一つには公的統計がビッグデータを利用するためのガイドラインを国際機関や国が提示することが考えられる。例えば、Upadhyaya and Kynclova [2017] では、ビッグデータの活用で国際機関による基準を設け、統計作成のベストプラクティスを導入し、各国の統計当局に対して技術的な支援を行うことを提案している。

柳川・渡辺 [2017] や Hammer et al. [2017] でも指摘されているが、特定の民間提供の統計を政府統計として認定するケースが将来的にあり得る。その際、一定のルールづくりや監督体制づくり、例えばデータの継続的な提供の義務を課す一方で、その見返りにデータの提供に対価を設定することを提言している。このように公的統計が抱える課題を解決するために、民間企業によるビッグデータの提供を一種のビジネスとして成立（統計の民営化）させることで、公的統計として一定水準の質を確保しながらビッグデータを有効に活用できるかもしれない。

また、行政データを経済統計に活用する場合、個人や企業の複数の業務統計を紐づける識別子 (identifier) の整備が必要となる。新たに導入されたマイナンバーや法人番号を活用することで、様々な行政データを統合して、匿名化に十分配慮しながら、行政データの経済統計への一層の活用

が期待できる。その際、美添 [2019] が指摘するように、現状では行政データの電子化が十分に進んでいないという課題がある。現在、政府は行政手続の電子化を含む簡素化を進めているところであるが、公的統計における行政データ活用の面からも、一層それを進めていく必要があるだろう。

その他、川崎 [2019] も指摘しているが、世界では統計当局が省庁間で分散されている国でも、徐々に集中型へ向かう傾向にあると言われている。専門人材の育成や統計作成作業の効率化などメリットが多いためだ。さらに、日本の統計当局の人員が非常に少ない点が指摘されている。日本でもそうした視点からの統計当局の組織改革や、統計人材の拡充および専門人材の育成に力を入れるべきだろう。

4) ビッグデータの貢献は公的統計に限らない

ビッグデータを公的統計に活用する事例は、業務統計の活用を中心に今後ますます増えていくだろう。その一方で、公的統計に限らず、それ以外の経済統計や経済予測の場面でもビッグデータの活用の可能性を秘めている。

例えば Garboden [2019] は、マクロ経済予測に活用できそうなビッグデータを種類別に紹介した上で、実際の活用事例や今後の利用可能性について議論している。さらに、Hammer et al. [2017] は、1章で述べたビッグデータの三つの性質に基づいて、各性質を持つビッグデータがどのような既存および将来的な経済統計へ応用できそうか、その可能性を探っている。

これらビッグデータの経済分析への応用は、公的統計を補完するものとして重要な役割であると考えられる。ビッグデータの公的統計における貢

献が「事実」を明らかにする点に比重を置くのに対して、経済分析における貢献は何らかの「洞察」を得ようとするものである。こうした「洞察」を得るためのビッグデータの活用は、民間による新たな経済統計・予測などの開発につながっていくだろう⁹。

しかし、民間が自ら所有しないビッグデータを活用するには、その費用が高額となる致命的な問題を解決しなければならない。そのため、民間では費用が安価なテキストデータなどを使った分析が多くなりがちである。匿名化や集計化などの工夫により対価を下げることや、データを提供する環境を競争的なものにして、データ提供の対価を引き下げていくことも必要だろう。

ビッグデータの普及により、官民が切磋琢磨しながら様々な経済統計や経済予測を生み出し、経済情報がさらに重層化していくことが望まれる。

おわりに

公的統計におけるビッグデータの活用は、現在、総じて実験段階にあると言える。物価統計や旅行者の移動といった分野で一足先に活用されつつあるものの、その他分野での活用事例は限られている。今後は「伝統的業務体系（処理仲介データ）」や「IoT（機器生成データ）」の活用は進む可能性はあるが、「ソーシャルネットワーク（人間由来の情報）」はそのデータの性質上、公的統計で活用されるには相当の時間がかかるものと思われる。

統計学は社会的要請とともに発展してきた歴史がある。ケトラーやカール・ピアソンらが活躍し

た19世紀の「記述統計学」の時代は、貧困などの社会問題を解決するために、得られたデータそのものの特徴を分析する手法が開発された。20世紀入り後、生物学や物理学からの要請によって、フィッシャーやネイマンらによる「推測統計学」が確立され、確率の手法を用いることで、代表性を担保した手元の限られたデータ（標本）から全体の姿（母集団）を推測することが可能になった。そして20世紀半ば以降、コンピューターの利用が広がり、大量のデータとそれを処理する計算技術や分析手法（機械学習など）の高度化が可能となって現在に至る。

現在のビッグデータによる分析は、必ずしも全体を反映していないデータに基づき予測・分類しているという意味で、19世紀の記述統計学の状況と似ている。しかし、かつての20世紀前半における統計学の発展のように、代表性などの本質的な問題に対して正面から取り組むことで、ビッグデータから母集団を明らかにするための分析手法が次第に開発されていく可能性がある。現在はそうした段階に至るまでの大きな過渡期にあるのではないか。

ただし現時点では、公的統計においてビッグデータを活用するには、速報性や報告者負担の軽減というメリットを享受することが優先されそう。一方、プライバシーの抵触懸念、特に代表性の欠如といったデメリットをどのように軽減していくのかについては、今後、さらなる官民そして学界の知恵が求められる。

9) 例えば大和総研の取り組みとして、日本銀行の「地域経済報告（さくらレポート）」のテキストデータを用いて、日本の地域別の景況感を数値化した経済統計「大和地域A I（地域愛）インデックス」を四半期ごとに公表している。（<https://www.dir.co.jp/report/research/policy-analysis/regionalecnmy/regionalindex/>）

【参考文献】

- Aladangady, A., S. Aron-Dine, W. Dunn, L. Feiveson, P. Lengermann, and C. Sahm [2019], “From Transactions Data to Economic Statistics: Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending”, Finance and Economics Discussion Series 2019-057. Washington: Board of Governors of the Federal Reserve System.
- Blanchet, D. and P. Givord [2017], “Big data, official statistics and measuring the economy”, Insee References, edition 2017.
- Efron, B. and T. Hastie [2016], “Computer Age Statistical Inference: Algorithms, Evidence, and Data Science”, Cambridge University Press.
- Garboden, P. M. [2019], “Sources and Types of Big Data for Macroeconomic Forecasting”, UHERO Working Paper No. 2019-3, July 8, 2019.
- Ghirelli, C., J. Peñalosa, J. J. Pérez, and A. Urtasun [2019], “Some implications of new data sources for economic analysis and official statistics”, Economic Bulletin 2/2019, Analytical Articles 22 May 2019, BANCO DE ESPAÑA.
- Hammer, C. L., D.C. Kostroch, G. Quirós, and STA Internal Group [2017], “Big Data: Potential, Challenges, and Statistical Implications”, IMF STAFF DISCUSSION NOTE, September 2017, International Monetary Fund.
- Jarmin, R. S. [2019], “Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics”, Journal of Economic Perspectives, Vol.33, Number 1, Winter 2019, pp.165-184.
- Kapetanios, G. and F. Papailias [2018], “Big Data & Macroeconomic Nowcasting: Methodological Review”, ESCoE Discussion paper 2018-12, July 2018.
- MacFeely, S. [2018], ‘Big Data and Official Statistics’, “Big Data Governance and Perspectives in Knowledge Management”, Chapter: 2, pp.25-54. IGI Global
- Upadhyaya, S. and P. Kynclova [2017], “Big Data -Its relevance and impact on industrial statistics”, DEPARTMENT OF POLICY, RESEARCH AND STATISTICS WORKING PAPER 11/2017, UNIDO
- 阿向泰二郎 [2019] 「家計調査の改良と消費動向指数（C T I）の開発——公的統計の深化へのチャレンジ」国友直人・山本拓編『統計と日本社会：データサイエンス時代の展開』第 11 章、東京大学出版会
- 梅田雅信・宇都宮浄人 [2009] 『経済統計の活用と論点（第 3 版）』、東洋経済新報社
- 川崎茂 [2019] 「統計制度の国際比較——日本の統計の特徴と課題」国友直人・山本拓編『統計と日本社会：データサイエンス時代の展開』第 12 章、東京大学出版会
- 経済産業省 [2019] 「I T 人材需給に関する調査（概要）」（平成 31 年 4 月）
https://www.meti.go.jp/policy/it_policy/jinzai/gaiyou.pdf
- 小西葉子 [2019] 「ビッグデータは公的統計を変えられるか：海外先進事例に学ぶ」RIETI BBL セミナー、独立行政法人 経済産業研究所
- 中村隆英・新家健精・美添泰人・豊田敬 [1992] 『経済統計入門（第 2 版）』、東京大学出版会。
- 竹内啓 [2018] 『歴史と統計学——人・時代・思想』、日本経済新聞出版社
- 東京大学教養学部統計学教室編 [1994] 『人文・社会科学の統計学』、東京大学出版会
- 柳川範之・渡辺努 [2017] 「民間による経済統計の革新—統計民営化の可能性も含めて—」『統計』2017 年 1 月号、一般財団法人 日本統計協会
- 美添泰人 [2019] 「公的統計の課題と改革」国友直人・山本拓編『統計と日本社会：データサイエンス時代の展開』第 13 章、東京大学出版会

[著者]

中田 理恵 (なかだ りえ)



経済調査部
AI アナリティックリサーチ室
研究員
担当は、内外経済、
データサイエンス

溝端 幹雄 (みぞばた みきお)



経済調査部
AI アナリティックリサーチ室
主任研究員
担当は、経済構造分析
(地域 / 生産性など)