

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（第4回）議事概要

1 日時 平成28年12月27日（火） 13:30～15:00

2 場所 総務省7階省議室

3 出席者

【構成員】

飯塚 信夫	神奈川県経済学部教授
岩村 有広	一般社団法人日本経済団体連合会経済政策本部長
小塩 隆士	一橋大学経済研究所教授
久我 尚子	株式会社ニッセイ基礎研究所主任研究員
座長 国友 直人	明治大学政治経済学部特任教授
熊谷 亮丸	株式会社大和総研調査本部副本部長・チーフエコノミスト
河野 康子	一般社団法人全国消費者団体連絡会事務局長
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
末澤 豪謙	SMBC日興証券株式会社金融財政アナリスト
菅 幹雄	法政大学経済学部教授
関口 博之	日本放送協会解説主幹
平嶋 彰英	総務省参与
座長代理 美添 泰人	青山学院大学経営学部プロジェクト教授
渡辺 努	東京大学大学院経済学研究科教授

【オブザーバー】

茨木 秀行	内閣府政策統括官（経済財政分析担当）付参事官（総括担当） （代理 森脇 大輔 内閣府政策統括官（経済財政分析担当）付参事官（総括担当）付参事官補佐）
長谷川 秀司	内閣府経済社会総合研究所国民経済計算部長 （代理 谷本 信賢 内閣府経済社会総合研究所国民経済計算部国民支出課長）
吉牟田 剛	総務省政策統括官（統計基準担当）付統計企画管理官
山澤 成康	総務省統計委員会担当室長
宇波 弘貴	財務省大臣官房総合政策課長 （代理 森田 稔 財務省大臣官房総合政策課経済財政政策調整官）
上田 洋二	経済産業省大臣官房調査統計グループ参事官（総合調整担当） （代理 高辻 育史 経済産業省大臣官房調査統計グループ調査分析支援室長）
肥後 雅博	日本銀行調査統計局参事役（統計担当）

【発表者】

星野 崇宏 慶應義塾大学経済学部教授

【総務省】

高市総務大臣、原田総務副大臣、島田総務大臣政務官、笹島総務審議官、會田統計局長、長屋官房総括審議官、佐伯官房審議官

【事務局（総務省統計局統計調査部）】

千野統計調査部長、阿向消費統計課長

4 議事

- (1) 高市総務大臣挨拶
- (2) ビッグデータを活用したマクロ消費変動の推計について
－事務局
- (3) データのバイアスの理解と複数データ源からの推論の可能性
－慶應義塾大学経済学部 星野教授
- (4) 需要側統計の改善について
－事務局
- (5) 家計調査の改善・刷新に向けて
－事務局
- (6) 意見交換

5 議事概要

- (1) 高市総務大臣挨拶
 - ・御多用の中での御参集に感謝。
 - ・先月の第3回の研究会では、菅構成員と MasterCard Advisors の星野様から御発表を頂いた。企業の消費が景況感に影響を及ぼし得るという菅構成員の御発表は、私自身の問題意識もそこにあり、大変興味深く伺った。また、クレジットカード利用者の購買行動から公的統計を予測するという星野様の御発表も非常に興味深いものだった。
 - ・第2回と第3回の研究会では、学会や産業界の最先端の事例を学んだが、今回から消費関連指標の開発に向けたより具体的な話に移る。本日は慶應義塾大学経済学部教授の星野崇宏先生をお招きしている。
 - ・前回までの議論を通じ、ビッグデータの持つ癖をいかに除くかが指標開発における非常に重要なステップだと認識。星野先生からは様々な癖を持つビッグデータをどのように貼り合わせるか、不足している情報をどのように補うか、といった手法について御発表頂く。
 - ・構成員の皆様におかれては、本日も多角的かつ活発な御議論をよろしくお願い申し上げます。
- (2) ビッグデータを活用したマクロ消費変動の推計について
事務局から資料2に基づき説明。
- (3) データのバイアスの理解と複数データ源からの推論の可能性
慶應義塾大学経済学部 星野教授から資料3に基づき説明が行われた。主な説明は以下のとおり。
 - ・現在、様々なデータのバイアス補正とその融合をテーマとして、統計局と共同研究を実施している。具体的には、家計調査の個票とインテージ社の SCI、SRI データを利用した融合的な解析を行っており、本日はその一部を報告する。

- ・第3回研究会の事務局資料にあった消費関連指標の開発への三つのアプローチは、全て共通の土台で考えることができる。
- ・統計指標間のかい離が指摘されているが、それは二つの観点に分離して考える必要がある。一つ目はデータの取得対象の違い（選択バイアス）。二つ目はデータの取得方法や変数内容の違い。
- ・家計調査のバイアスについては、既に様々な議論があるが、誤記入や調査疲れに加え、標本の偏り（選択バイアス）も指摘されている。これらは欠測データ解析という枠組みで議論すると分かりやすい。この枠組みでは、単身モニター調査結果との合成やビッグデータとの融合も、一気通貫して議論することができる。
- ・細かい数理的な話は割愛するが、欠測データ解析によって、選択バイアスの除去、統計的な因果推論、複数データの統計的な融合などを統一的に議論することが可能。この手法は、ハーバード大学のRubin教授らを始めとする研究者らによって開発され、現在では社会科学や医学のみならず、マーケティング等にも利用されている。海外では公的統計の分野でも実用化が進み、例えばオーストラリアやカナダの国勢調査、アメリカの経済センサスなどで既に用いられている。
- ・家計調査のバイアスを、家計簿への記入が必要ないスキャンパネルによる購買行動ログデータと比較して、検証した。
- ・今回用いたスキャンパネルデータはインテージ社のSCIというデータで、全国5万人の購買履歴情報が毎日記録されている非常に大規模なもの。アルコール、シャンプー、炭酸飲料など、ほぼ網羅的にSCIデータが得られる品目を対象に、単身世帯に限定して、家計調査とSCIデータの解析、比較を行った。
- ・その結果がP.14~15のグラフだが、購買行動ログデータと比較して、家計調査には誤記入、調査疲れ等による過少記入のバイアスが生じているように見える。
- ・この結果は家計調査の対象者とSCIデータの対象者の結果を直接比較したものであり、調査方法（モード）の違いを議論するためには、対象者の違い（選択バイアス）を分離して考えなければならない。両者を分離するために、欠測データ解析の方法の一つである共変量調整法を利用した。補助変数（共変量）に関しては、対象者の分布が家計調査と同様になるように調整を行った。これにより、家計調査対象者がスキャンパネル式で回答した場合の結果を推定すると、家計調査に比べて高い数値が得られた。
- ・次に、ビッグデータが公的統計に利用できるのか考えてみる。例えば、コンビニやスーパーのトップ企業であっても、そのシェアは多くて3~4割程度。連携ポイントプログラムにおいても同様だろう。また、各社が保有するのは自社のデータのみであり、他社の購買はもちろん、総購買についてもデータがない。つまり、ビッグデータそのものには対象の偏りと変数の違いがあり、そのまま公的統計に利用できる質のものではない。
- ・しかし、公的統計の速報性や精度向上の観点から、そのバイアスを乗り越えて、ビッグデータを活用することには意義がある。その方法論として、欠測データ解析の応用手法として開発された「データ融合」が利用できる。

- ・関心のある全ての変数が同じ対象者から得られている（＝シングルソースデータ）ことが望ましいが、通常は、別々の対象者から分割して得られている（＝マルチソースデータ）。マルチソースデータについて、補助変数をのりしろとしてつなぎ合わせ、シングルソースデータ化することをデータ融合と言う。
- ・学術的な場や政策決定の場など、様々な形でデータ融合は使われている。経済学で利用されている疑似パネルデータ解析もその一種である。
- ・例えば、楽天と Amazon は二大 EC サイトとして、国内では非常に大きなシェアを持っているが、両サイトでは購買層が大きく異なるため、楽天の購買データをそのまま利用して、Amazon の購買データを推測することは難しい。しかし、データ融合を用いれば、楽天のデータから Amazon のデータを推定でき、いかに競合他社から顧客を奪うかという経営戦略の立案も可能となる。このように、データ融合はマーケティング分野では非常に有用で、既に多くの実用化がなされている。
- ・ビッグデータと家計調査の融合的解析の可能性について考える。公的統計が求めるのは国民全体の総購買額だが、各社がいくら多くの顧客データを持っていても、ビッグデータからは自社の購買状況しか分からない。したがって、まずはビッグデータ側で、他社を含めた総購買額を推定する必要がある。非常に困難と思われるが、例えばパネルデータの相関情報（どこで、何が買われていたか）等を用いて、代入又は推測が可能。共通の補助変数がないため、マクロ情報等も活用しながら融合的解析を行うことになるだろう。
- ・今回は、ビッグデータとしてインテージ社の SRI という POS データを用いて分析を行った。SRI データは、スーパー、コンビニエンスストア、ドラッグストア等に関して、地域や業態になるべく偏りがないように、全国 4,000 店舗に協力いただき、日次データの形で集計しているもの。年間で 40 億件以上の購買情報になる。Amazon やデパートなど、SRI に参加していない企業の購買情報は含まれないが、家計調査や SCI データを用いて、これらを含むように補正を行った。
- ・商業動態統計は月次データのみ公表だが、POS データ等のビッグデータであれば、日次でのデータ公開が可能となる。また、商業動態統計は非常に網羅性が高い統計ではあるが、個別の品目ごとの売上額に関しては分からない。POS データなどビッグデータを活用することにより、個別の品目に関しても日次集計が可能になるだろう。
- ・複数データを融合的に利用する場合の問題点は、作業量が膨大であること。今回の解析を例にすれば、家計調査と SCI データでは属性情報の変数定義が異なり、これを共通化するのに非常に苦労した。
- ・購買行動ログデータと POS データとの融合については、今回はどちらもインテージ社のデータを使用したため、変数の共通化は必要なかったが、基本的には企業により商品マスターのコード体系が大きく異なる場合がある。これらを共通化して解析するのは大変な作業量となる。
- ・今回のようなデータ融合の場合、例えば 1 年間のデータで数千万から数億オブザベーションの統計解析が必要となる。データの整形まで含めると、かなりの労力、時

間が必要。ただ、それらの処理を行うことで、ビッグデータの偏りを十分な精度で補正することは可能と考える。

- まとめとなるが、家計調査について指摘されているバイアスは、本当にバイアスなのかどうか精査が必要。これは二つの要素、データ取得対象の違い（選択バイアス）とデータ取得方法や変数の違いとに分けて考えなければならない。
- 今回の解析結果から、データ取得方法を家計簿方式から購入行動ログ形式に変更することで、過少記載が大幅に修正される可能性があることが分かった。
- ビッグデータをそのまま利用するのは明らかに問題。選択バイアスの問題と自社データのみというバイアスもある。欠測データ解析の枠組みとデータ融合という考え方をを使って議論していく必要がある。
- 今回はインテージ社の SRI データと SCI データを用いた家計調査との融合的な解析結果について紹介した。ビッグデータをマクロ指標にする際には、必ずマイクロレベルでのバイアス除去が必要。今後は人的な資源を投入してより詳細な解析を行いたいと考えている。
- 今後、各企業のビッグデータを利用するに当たり、素データの提供を受けるのは難しいだろう。今回紹介した解析でも、SRI の素データの提供は受けていない。私の方から性別、年代、収入、地域等の分布情報を示して、それに適合する形で集計した日次データを提供いただいた。このような方法であれば、企業から提供を受けることが容易になるのではと思う。
- また、政府統計の担う正確な情報の収集や提供は、我が国の行政のみならず、民間の経営効率化にも非常に有用。この公共財としての政府統計を拡充する形で考えていただきたい。

星野教授が説明した後の質疑の時間における主な発言は以下のとおり。

- 傾向スコアの研究は進み、世の中でも様々に使われるようになってきている。発表にあったとおり、消費水準の推計にも役立つだろう。しかし、実際の作業には十分な人手と時間が必要だ。星野先生のような専門家が不可欠であり、政府としても手当を考えなければならない。ビッグデータを簡単に扱えるものだと考えないようにしてほしい。
 - 民間のビッグデータを公的統計に使うに当たり、データの継続性が問題になるのではないか。例えば、民間調査会社のデータであれば、急に調査をやめてしまう、あるいは調査の内容を大きく変えてしまうといったことがあるかと思う。あるいは、ポイントカードのデータであれば、企業努力により会員数が大幅に増える、キャンペーンによりポイント利用が一時的に急増するなど、様々なことが起こりうる。これらに対する補正は可能なのか。
- おっしゃるとおり、シェアの変化やデータ定義の変更等によって、得られる情報は変わる。ただ、一部の企業のみからデータが提供されている状況では、それは確かに大きな問題だが、多くの企業からデータが提供されていれば、そういったバイア

スは取り除けるだろう。それゆえ、データ提供のメリットを企業に打ち出し、快くビッグデータを提供いただく素地ができ上がれば、継続的にデータが得られ、カバレッジも広がるだろう。

- ・ビッグデータの中でも、各企業が決済のためにとっているデータは途切れることがないだろう。企業の合意さえあれば、継続性は担保できると考えている。
- ・最終的にどこまで精度を高めることができそうか。
- まだ解析を始めたばかりであり、現時点で具体的な精度まで保証できないが、現行の家計調査よりも精度をかなり高めることが可能という手応えがある。
- ・大変な作業量だったと、お話の中から伝わってきた。例に挙がっていたアルコールやシャンプーは家計調査の品目の中でも比較的扱いやすいデータではないかと思う。そういう意味で、扱いが難しい品目・分野はどれか。
- 今回は試験的に家計調査と民間の消費者パネルデータできちんと調査されているであろう品目を選んで解析した。耐久消費財に関してはこれほど簡単には行かないかもしれない。家計消費状況調査や全国消費実態調査等、ほかのデータを利用した解析を行う必要があると考えている。
- ・家計調査とインテージの SCI データについて、傾向スコアを用いて分析したということだが、補助変数には何を使ったのか。また、それは各品目で共通か。
- 今回、補助変数に用いたのは居住地域、職業区分、年齢、性別など基本的なものだけを使っているのだから、大きく結果が変わるということにはならない。補助変数は各品目で共通のものを用いている。
- ・楽天と Amazon の購買者を分析・比較していたが、職業や年齢などのほかに、更に細かい補助変数が必要で、これが最終的な意思決定に影響を与えているものと思う。職業等の基本的な情報以外に、実際に必要な補助変数が手に入らない場合には、どうやって補正をかけるのか。
- 楽天と Amazon の分析については、ウェブの閲覧履歴も用いている。どのサイトから流入してきたか、どのサイトを使っていたか、といった情報を補助変数として用いると解析はうまくいった。補助変数として使える情報によって、精度が変わる。今回、家計調査と SCI データを融合する際には、家計調査には様々な調査事項があった一方、SCI 側の方は情報が少なかった。例えば、家計調査と同様の変数に関する調査を別途実施していただければ、そういったものを家計調査と融合することは可能。したがって、調査ベースでやるなら、基本的には調査事項を増やせば良く、補助変数が増えるような形で調査を依頼しデータをとることによって、御質問の点は担保できると考えている。
- ・推計の信頼区間、誤差率などは出せるのか。
- 計算は可能。しかし、今回は大量のデータを扱っているため、信頼区間等は極めて小さく、計算しなかった。
- ・恐らく多くの方が知りたいのは、名目値ではなく、実質値だろう。本日の発表は、名目値の補正に関するものと聞いていたが、名目値の補正と整合的に実質値も補正することは可能か。

→実質化に関しては今回実施していなかった。ご意見を踏まえ今後実施したい。

(4) 需要側統計の改善について

事務局から資料4に基づき説明。

(5) 家計調査の改善・刷新に向けて

事務局から資料5に基づき説明。

(6) 意見交換

各構成員の主な意見は以下のとおり。

- ・「政府の統計は問題がある」といった場合に、デフレーターに問題があるという主張もある。つまり、実質値を作る際に、名目ではなくデフレーターがおかしいという議論は消費に関わらず一般にある。このことを念頭においてみると、資料4で、名目値に関する工夫を色々とされていることは分かるが、それらをデフレートする際には、今までどおり総務省統計局の消費者物価指数を使っているようにみえる。資料4の試算結果が実質値であることを踏まえると、名目値と実質値の両方を直してほしい。名目値とデフレターの両方に問題がある場合に、名目値だけでも直せば、実質値も真実に近づくと思いたくなるが、必ずしもそうではない。名目値に入っている誤差とデフレーターに入っている誤差が相関しているとすれば、片方だけを直しても、実質値が真実に近づく保証はない。名目値、デフレーターをそれぞれ直すということ念頭に置くべきではないか。SCI や ID 付きの POS データからは消費額だけではなく価格のデータもとれるので、星野先生の発表にあった形で補正をかけたければ、名目値のみにとどまらず、デフレーター、実質値もきちんと補正をかけた値を出せるのではないか。そういう観点からのビッグデータの活用の仕方というのも、あり得ると思った。

→本日提示した指標は、消費者物価指数を使って実質化している。ビッグデータの活用については御指摘いただいた点も含めて検討していきたい。

- ・資料5のP.3で、家計簿の記入行数を15行から30行に増やしたと説明があった。行数を増やした場合でも、家計簿に記入された商品に分類をつける作業は、これまで通り、統計局・統計センターで行うのか。

→これまで通りの作業を想定している。

- ・行数が増えるとその分、統計が正確になる反面、時間がかかる可能性がある。異常値をはじく作業などを踏まえると、新方式で調査を正確にやっていくために、平成30年以降はそれなりの時間がかかるという認識をもつ必要がある。
- ・資料4のP.4以降においては、第3回で事務局から提示された資料よりも深掘りされ、新指標のイメージが具体的になり、指標の中身が充実してきたという印象である。
- ・新指標を公表するタイミングについては、家計調査と一体が好ましい。しかし、新

指標の公表にあわせるために、家計調査の公表タイミングが遅れるということであれば、判断に困る。一般的に言うと、マーケットにいるエコノミストは速報性の優先度が非常に高い。新指標の公表のタイミングや意義は、ユーザーとしっかり対話をして、決めていくことが必要と考える。

→発表のタイミングも含めて、慎重に検討して行かなければならない。エコノミストの方々などから、広く御意見を頂戴していきたい。

- ・資料4のP.9のトレンドの抽出について、政府がこういった分かりやすい資料を出すことは、方向性として非常に良いことと思う。他方で、資料4のP.10にある青い線のT系列を公表系列に含めるかどうかは慎重に検討する必要があると思う。理由は、例えばP.6にあるように世帯あたりの消費の傾向は、人口の減少や高齢化など影響で長期的に低下している。トレンドとして落ちているものを景気の動きとして捉えてしまうといった誤解を招く可能性がある。その意味では、資料4のP.10の赤い線のTC系列には周期的な変動が含まれる。日本のその時々々の景気の動向をみるには、TC系列の方が好ましいと考える。

○島田総務大臣政務官感想

- ・御多用の中での御参集に感謝。
- ・本日の議論を踏まえ、家計調査の補完・補強に関する一定の方向性が得られたものと考えている。
- ・ビッグデータのバイアス補正やデータ融合についてお話いただいた星野先生には改めて感謝申し上げる。調査データとビッグデータを相互に補完する画期的な話だったが、非常に大変な作業だと改めて感じた。これが実現すれば、消費関連統計にとどまらず、公的統計全体に新しい道が開けるものと確信した。
- ・ビッグデータは、多くの可能性を秘めた情報源だが、従来型の調査統計が普遍的に正確なデータを提供し続けている意義を改めて申し上げたい。引き続き精力的な検討が進むことを期待する。
- ・今回頂きたいずれの意見も新たな消費関連指標の開発に対する貴重な示唆となった。
- ・皆様方におかれては、本日も改めて御知見を賜り、厚く御礼申し上げます。

○原田総務副大臣感想

- ・闊達な御議論を頂いたこと、誠に感謝。
- ・御発表いただいた星野先生には私からも厚く御礼申し上げます。
- ・これまでの研究会では、ビッグデータのバイアスを除去することの難しさを痛感させられる場面が多く、実現までの道のりの長さを感じていた。星野先生のようにこの難題に挑む先生がいるということは、大変心強く感じる。
- ・初回の研究会において、高市大臣から、国際的にも高く評価され得るような新しい消費関連指標を、という発言があった。幸いにして、我が国には、渡辺先生や星野先生らの世界最先端の知見がある。今後とも御指導いただけたら幸い。