



日本銀行ワーキングペーパーシリーズ

「オルタナティブデータ消費指数」の開発： オルタナティブデータを用いた 個人消費のナウキャストイング

大久保友博*

tomohiro.ookubo@boj.or.jp

高橋耕史**

kouji.takahashi-2@boj.or.jp

稲次春彦*

haruhiko.inatsugu@boj.or.jp

高橋優豊*

masato.takahashi@boj.or.jp

No.22-J-9
2022年3月

日本銀行
〒103-8660 日本郵便（株）日本橋郵便局私書箱30号

* 調査統計局

** 調査統計局（現・総務人事局）

日本銀行ワーキングペーパーシリーズは、日本銀行員および外部研究者の研究成果をとりまとめたもので、内外の研究機関、研究者等の有識者から幅広くコメントを頂戴することを意図しています。ただし、論文の中で示された内容や意見は、日本銀行の公式見解を示すものではありません。

なお、ワーキングペーパーシリーズに対するご意見・ご質問や、掲載ファイルに関するお問い合わせは、執筆者までお寄せ下さい。

無断転載・複製を禁じます。

「オルタナティブデータ消費指数」の開発： オルタナティブデータを用いた個人消費のナウキャストイング*

大久保友博[†]、高橋耕史[‡]、稲次春彦[§]、高橋優豊[#]

2022年3月

【要旨】

近年、マクロ経済分析の分野において、既存統計とは異なる情報源に基づく非伝統的なデータ、いわゆる「オルタナティブデータ」に対する関心が高まっている。本稿では、速報性に優れるオルタナティブデータを用いて、既存統計の公表よりも早いタイミングで、マクロ的な個人消費の動向を捕捉することを試みる。具体的には、3種類のオルタナティブデータ、すなわち、①クレジットカードの利用履歴(JCB消費NOW)、②POSデータ(METIPOS、GfK)、③家計簿アプリの支出記録(マネーフォワード)、を合成した指標「オルタナティブデータ消費指数(ALC)」を構築し、これを用いて日本銀行が作成・公表している消費活動指数のナウキャストイングを行う。ALCは、消費活動指数が公表される約3週間前(対象月の翌月半ば)というきわめて早いタイミングで集計可能であり、速報性に優れている。分析結果によれば、ALCのナウキャストイング精度は総じて良好であり、2020年春以降における新型コロナウイルスの感染拡大に伴う消費活動の大きな変動についても、的確に捉えている。本稿の分析結果は、オルタナティブデータが、マクロの消費活動を迅速かつ的確に捉えることができる点で、経済情勢を把握するための強力なツールであることを示している。

JEL分類番号：C49、E21、E27

キーワード：ナウキャストイング、オルタナティブデータ、個人消費

* 本稿の作成にあたり、亀田制作氏、川本卓司氏、須合智広氏、長野哲平氏、中村康治氏、八木智之氏および日本銀行スタッフから有益なコメントを頂戴した。ただし、本稿のありうべき誤りは全て筆者ら個人に属する。本分析にあたって、マネーフォワード社より、本稿の研究目的のために、個人が特定できない形に加工された統計情報の提供を受けた。なお、本稿に示される内容や意見は、筆者個人に属するものであり、日本銀行、マネーフォワード社の公式見解を示すものではない。

† 日本銀行調査統計局 (tomohiro.ookubo@boj.or.jp)

‡ 日本銀行調査統計局 (現・総務人事局、kouji.takahashi-2@boj.or.jp)

§ 日本銀行調査統計局 (haruhiko.inatsugu@boj.or.jp)

日本銀行調査統計局 (masato.takahashi@boj.or.jp)

1. はじめに

近年、マクロ経済分析の分野において、既存統計とは異なる情報源に基づく非伝統的なデータ、いわゆる「オルタナティブデータ」に対する関心が高まっている。オルタナティブデータとは、クレジットカードの利用履歴に基づく消費データ、スマートフォンの位置情報に基づく人流データ、あるいは各種のテキストデータなど、個人・企業などの様々な社会経済活動に基づく大容量データを、高度な情報技術によって収集・加工したものである。

こうしたオルタナティブデータの多くは、統計作成のためのデータ収集・集計といった一定の段取りを取る既存統計と比較してデータが利用可能になるまでの時間が短い（速報性が高い）という特徴がある。マクロ経済の情勢判断の観点からは、足もとの状況をできるだけ早く捉える「ナウキャストイング」のニーズから、オルタナティブデータを活用する動きは従来からみられてきた。この流れは、新型コロナウイルス感染症の影響により短期的な経済変動が大きくなっている今次局面で加速しており、政策当局をはじめとして、オルタナティブデータを活用した経済情勢のナウキャストイングに関する研究が活発化している。

オルタナティブデータによるナウキャストイングの試みが進んでいる領域の一つが、感染拡大の影響を最も強く受けた、個人消費である。感染症下の個人消費は、公衆衛生上の措置が導入・解除されるもとの、既存の月次・四半期統計では追跡できないほどのペースで、過去に例をみないほど大きく乱高下してきた。こうした中、クレジットカードの利用履歴やスマートフォンの位置情報データなど、日次・週次の頻度でアップデート可能なオルタナティブデータを活用し、既存統計よりも早期かつ的確に個人消費動向を捕捉しようとする試みが、盛んに行われてきた。

例えば、Dunn et al. (2020, 2021) は、米国を分析対象として、クレジットカードの利用履歴の日次データが小売統計の高精度な予測に役立つことを指摘している。他国でも、カード利用履歴のデータが消費のナウキャストイング精度を高めるとの分析結果が報告されるなど、オルタナティブデータの有用性が認められてきている（Chapman and Desai 2021<カナダ>、Bounie et al. 2020<フランス>、Carvalho et al. 2021<スペイン>など）。

日本でも、オルタナティブデータによる個人消費のナウキャストイングが進展している。例えば、日本銀行は、感染症の拡大以降、一部の民間企業が公表を始めた日次の人出データ（東京の繁華街における夜間人口、小売・娯楽関連の人出動向など）が対面型サービスの消費動向と連動性が高いことを確認したうえで、個人消費の情勢判断に活用している。また、実際の消費額を基に構築されたオルタナティブデータとして、クレジットカードの利用履歴データを指数化した「JCB 消費 NOW」や、店頭での購買情報を集計した「METI POS」を、わが国における消費活動の早期把握のために利用している。これらの分析は「経済・物価情勢の展望（展望レポート）」で公表し

ており、一部のデータについては定期的なモニタリング指標として継続的に掲載している¹。そのほかの活用事例として、王ほか（2021）は、スマートフォンの位置情報データを用いて集計した商業施設の周辺人口が、サービス業（遊園地、ショッピングセンター、飲食業）における消費をうまく捉えていると報告している。また、浦沢（2022）は、JCB 消費 NOW がサービス消費のリアルタイム予測に有用であると報告している。

これらの事例から分かるように、オルタナティブデータは個人消費のナウキャストリングで威力を発揮する。ただし、オルタナティブデータは、定義や集計対象が既存統計とは異なるため、データを利用する際には留意が必要である。例えば、クレジットカードの支払データは、集計対象のカードを用いていない個人の消費活動は捕捉できない。また、わが国では、クレジットカードは、比較的高額な財・サービスの支出にはよく使われる一方、日々の少額の取引や、ローンを組むことも多い自動車の購入などには利用されにくい面もある。他方、POS データは、支払手段によらず購買履歴を幅広く捕捉できるものの、サービス消費については利用が難しいほか、財についても集計対象が一部の実店舗での消費に限られる場合もある。また、家計簿アプリのデータについては、集計対象のユーザーの人口分布が、IT リテラシーの高い個人（主に若年層）に偏っている可能性がある。すなわち、オルタナティブデータは、情報源の違いから、特定の業種・品目については高い精度で捕捉できても、それ以外の捕捉精度が高いとは限らないなど、一長一短がある。したがって、マクロの個人消費動向を包括的に捉えるには、単一のオルタナティブデータを用いるだけでは不十分である。この問題を克服するには、それぞれのデータの長所を生かしながら、複数のオルタナティブデータを相互補完的に用いることが有効であると考えられる²。

そこで本稿では、複数のオルタナティブデータを組み合わせて、わが国におけるマクロの個人消費動向を高い精度でナウキャストリングすることを試みる。オルタナティブデータは品目別・業種別の個人消費（対面型サービスなど）のナウキャストリングに活用されてきたが、情報源の異なるオルタナティブデータを合成してマクロの消費動向をナウキャストリングする試みは、筆者らが知る限り、これまでのところ海外の研究を含め存在しない³。ナウキャストリングのターゲットは、日本銀行がマクロの

¹ 日本銀行は、新型コロナウイルスの感染拡大以前から、オルタナティブデータを経済分析に活用してきた。亀田（2021）は、感染症拡大以降の事例も含め、これまで日本銀行が行ってきたオルタナティブデータを用いた分析を紹介している。また、日本銀行では、ホームページにおいて「オルタナティブデータ分析」コーナーを開設し、関連するリサーチを掲載している。

² 2021年11月に開催された東京大学金融教育研究センター・日本銀行調査統計局主催のコンファレンスにおいても、様々なオルタナティブデータを組み合わせて用いることの重要性が指摘されている（日本銀行調査統計局 2022）。

³ 本分析は個人消費のナウキャストリングを対象としているが、GDP のナウキャストリングにオルタナティブデータを活用した研究も存在する。例えば、中澤（2022）は、月次データ（主に公的統計）を用いて日本の GDP を予測するために構築された日本銀行のナウキャストリングモデル（Chikamatsu et al. 2021）について、予測精度が感染症下で悪化していることを指摘したうえで、

個人消費動向を把握するために月次で作成・公表している、消費活動指数（Consumption Activity Index、CAI）である⁴。

ナウキャストに用いるオルタナティブデータは、①クレジットカードの利用履歴に基づく消費指標（JCB 消費 NOW）、②POS データ（METI POS および GfK）、③家計簿アプリの支出記録（マネーフォワード<個人属性別の集計値>）、の 3 種類である。これらのデータを入手するまでの期間は 2 週間程度（例えば、4 月の消費動向のデータが 5 月半ばに利用可能になる）であり、公表までに少なくとも 1 か月程度のラグがある既存統計と比較して、速報性に優れている。また、位置情報データや各種マインド指標などとは異なり、個人の消費額や事業所の売上高を基に作成されているため、個人消費動向を正確に捉えるうえできわめて有用である。本分析では、消費活動指数を構成する品目別指数のそれぞれに対して、フィットが良いオルタナティブデータを上記 3 つの中から選び出し、それらを合成した「オルタナティブデータ消費指数」（Alternative Data Consumption Index、以下、ALC）を作成する。ALC の作成過程では、各データのサンプルの偏りなどに起因するバイアスを可能な限り調整している。

分析結果によれば、多くの業種・品目について、オルタナティブデータは、個人消費動向をうまく捕捉している。そのうえで、捕捉精度が高いオルタナティブデータを品目ごとに選び出して合成した ALC は、マクロレベルの消費活動指数を総じてうまく予測できており、2020 年春以降の感染拡大に伴う消費活動の大きな変動についても、既存統計の公表よりも早いタイミングで、的確に捉えている。

本稿の構成は次の通りである。まず第 2 節で、本稿で用いるオルタナティブデータについて概説する。第 3 節では、ALC の作成方法について概要を述べる。第 4 節では、ALC のパフォーマンスについて報告する。第 5 節は結びである。

2. 本稿で用いるオルタナティブデータ

本節では、本分析で用いる 3 種類のオルタナティブデータについて概説する。

2-1. クレジットカードの利用履歴 —— JCB 消費 NOW

第 1 のタイプは、「クレジットカードの利用履歴」である。本分析では、株式会社ナウキャストが作成・提供している、JCB カードの利用履歴データを指数化した「JCB 消費 NOW」を用いる。これは、無作為に抽出された JCB カード保有者の利用履歴に

Google の検索データや METI POS 小売販売額といった日次・週次のオルタナティブデータをモデルに取り入れると、ナウキャストの精度が改善すると報告している。

⁴ 消費活動指数についての詳細は、中村ほか（2016）、高橋ほか（2021）などを参照されたい。

基づき、消費動向を指数化したものである（支出額そのものではない）。集計対象は、JCB カード（JCB が直接発行しているものに限る）の利用者約 1,000 万人で、性別・年齢・地域など、詳細な属性ごとに消費の動向を把握することができる。また、指数作成にあたっては、人口推計を用いてサンプルの歪みを補正しているほか、上位項目（総合、財、サービス）の集計時には、商業動態統計や消費活動指数の業種ウエイトを勘案しており、マクロの消費動向を的確に捉えるための工夫が施されている。加えて、クレジットカードはインターネット通販での商品購入に利用されやすいため、オンライン消費の捕捉力も高い。オンライン消費は既存統計で十分に捕捉できておらず、JCB 消費 NOW はこの点にも強みがある。

2-2. POS データ —— METI POS、GfK

第 2 のタイプは、「POS データ」である。POS データ（Point of Sales）は、商品の販売時点・価格・販売数量などを記録したものである。先述の JCB 消費 NOW が需要側（消費者側）のデータであるのに対して、POS データは供給側（販売者側）のデータである。

本分析では 2 つの POS データを利用する。一つは、経済産業省が作成している「METI POS」である。これは、小売業の 5 業態（スーパー、ホームセンター、家電大型専門店、コンビニ、ドラッグストア）について、POS データを活用し、品目別（食品、雑貨、家電など）の小売販売額を週次および月次で集計したものである。本分析では、消費活動指数の「飲食料品」、「医薬品、化粧品など」の動きを捉えるオルタナティブデータの候補として、METI POS の月次データを用いる。もう一つの POS データは、GfK Japan が提供している指標である（以下、このデータを「GfK」と呼称する）。これは家電製品についてのみのデータであるが、METI POS より広範な商品群を集計対象としていることから、消費活動指数の「家電」を捉えるための指標の候補として用いる。

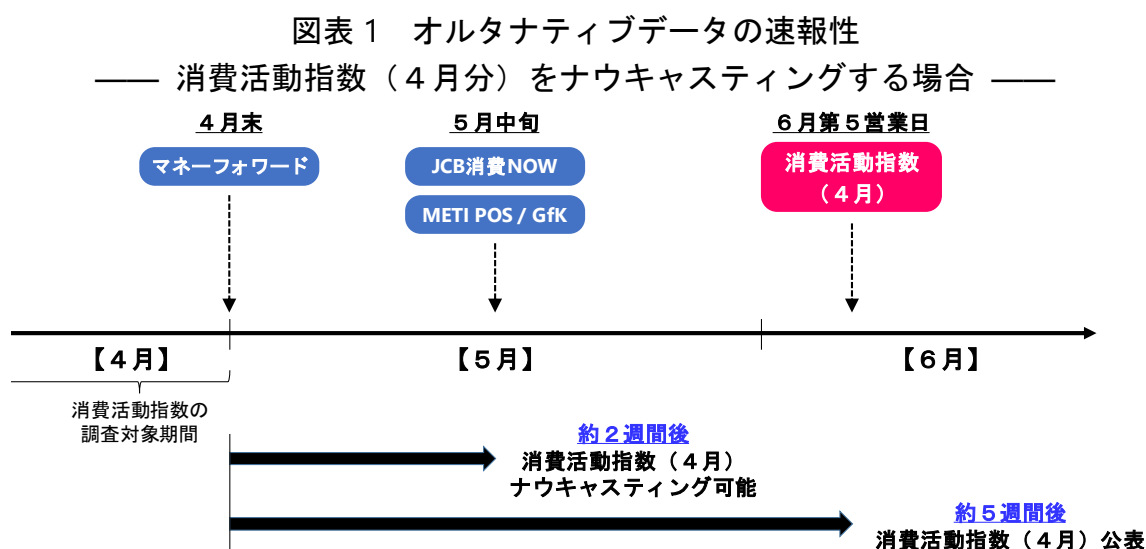
2-3. 家計簿アプリの支出記録 —— マネーフォワード

第 3 のタイプは、「家計簿アプリの支出記録」である。本分析では、株式会社マネーフォワードが提供している家計簿アプリ「マネーフォワード ME」のユーザーの支払履歴データを用いる。当データは、同アプリのユーザーが自ら記録した支払履歴を同社が多岐にわたる調査対象品目ごとに集計したものであり（したがって、このデータで個人を特定することはできない）、需要側のデータと位置付けられる。本稿では、後述のように個人属性が判明しているほか毎月利用があるなど集計対象の要件をかなり限定しているが、それでも集計対象となるユーザー数は約 31 万人と、家計調査の集計対象（約 0.9 万人）を遥かに上回る。ただし、家計簿アプリの利用者はこれまでのところ若年層に偏っており、単純に集計するとマクロの個人消費とは傾向が乖離する可能性がある点には留意が必要である。この点、同データでは、匿名処理された

全てのユーザーの詳細な支出履歴をもとに、年齢や家族構成といった属性グループごとの平均支出額を算出している⁵。本分析では、この利点を活かし、品目別にマクロの支出額を推計する際、グループ別の集計値を日本全体の人口分布と整合的な形で集約する工夫を施し、推計精度を高めている。

2-4. オルタナティブデータの速報性と留意点

これら3種類のオルタナティブデータに共通する特徴は、その速報性の高さである。例えば、4月分の消費に関するオルタナティブデータは5月半ばには出揃う。他方、本分析のナウキャストのターゲットである4月分の消費活動指数は、同月に関する公的統計が出揃い、それらの集計作業が完了する6月上旬まで待つ必要がある（図表1）。オルタナティブデータを用いれば、マクロの個人消費動向を把握するまでの時間を大幅に短縮することができ（約5週間→約2週間）、情勢判断を行ううえでメリットは大きい。



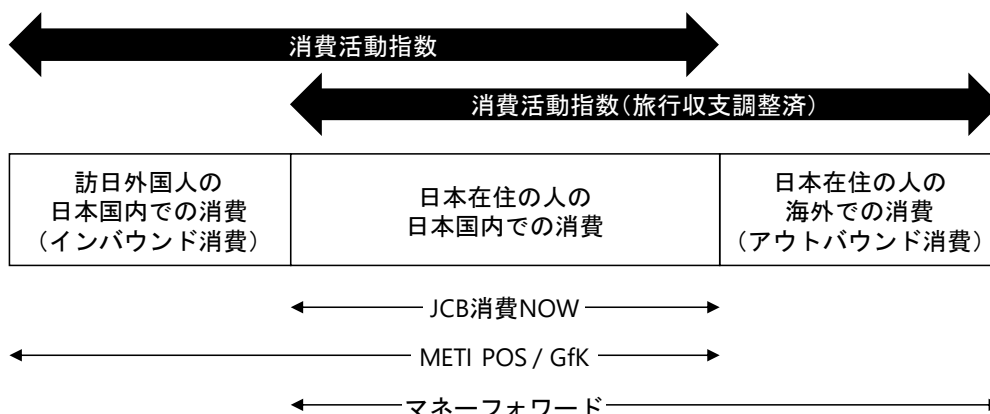
ただし、これら3種類のデータは、情報源の違いから、変動パターンが異なる点に注意が必要である。前述のように、クレジットカードの支払履歴に基づく JCB 消費 NOW は、クレジットカードの利用率が高い品目の消費は精度良くトラックできるものの、クレジットカードの利用率が低い業種の捕捉力は必ずしも高くないと考えられる。他方、POS データは捕捉対象の品目が限られるほか、家計簿アプリのデータはアプリを使用していない人々の消費を捕捉することができないといった制約がある。このほか、インバウンド消費（訪日外国人の日本国内での消費）とアウトバウンド消費（日本在住の人の海外での消費）をカバーしているかといった点も、データの変動に影響し得る（図表2）。後で見るように、この部分の違いが、感染症拡大以降の消費動向（とくに旅行・宿泊）の捕捉に影響を及ぼした面もあるとみられる。

⁵ 筆者らはグループの閾値の設定のみを行い、集計作業はマネーフォワード社に依頼した。

このほか、オルタナティブデータには一定のバイアスが存在し得る点にも留意が必要である。例えば、①キャッシュレス化が進展し、より幅広い商品がクレジットカードを用いて購入されるようになる⁶、または、②家計簿アプリのユーザーの消費行動が、家計簿アプリを用いていない人々と大きく異なる（支出抑制に熱心な人が当該アプリを利用している場合など）、といった状況では、クレジットカードや家計簿アプリの情報を用いて集計された消費額は、実勢以上に大きくも小さくもなり得る。したがって、オルタナティブデータを用いてマクロの個人消費動向を捕捉する際には、データ特有のバイアスにも注意を払う必要がある。

このように、3種類のデータそれぞれについて、消費動向を捕捉しやすい業種・品目があるほか、特有のバイアスを含んでいる可能性があるため、1種類のデータを用いるだけでは、マクロの個人消費動向を的確に捕捉することは難しい。マクロの消費動向をできるだけ正確に捉えるという観点からは、バイアスに注意しながら、複数のデータを組み合わせて用いることが望ましいと言える。

図表2 本分析で使用するオルタナティブデータのカバレッジ（概念図）



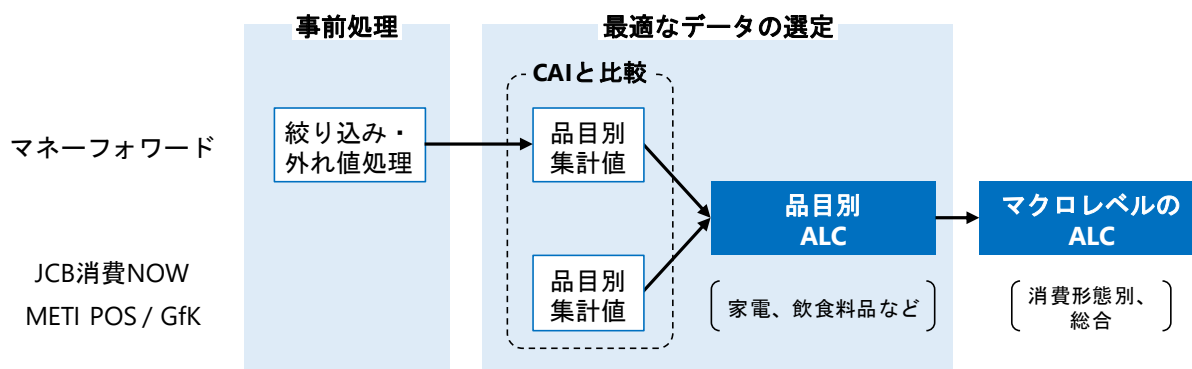
3. ALCの作成方法

本節では、ALCの作成方法について説明する。作成手順は以下の通りである（図表3）。まず、マネーフォワードの個人属性別データをマクロの人口分布と統合的な形で集計した「マネーフォワード指数」を、品目別に作成する。次に、各品目について、消費活動指数の品目別指標との相関係数が最も高いオルタナティブデータ（すなわち、当該品目のALC）を選び出す。最後に、品目別のALCを消費活動指数ウェイトで合

⁶ クレジットカード利用が一定のペースで緩やかに増加する場合、これに伴うバイアスはさほど大きくなく、補正も比較的容易である。しかし、感染拡大に見舞われた今次局面のように、オンライン消費が急激かつ非連続的に拡大する場合などには、主たる決済手段であるクレジットカードの利用増加に伴うバイアスは平時よりも大きくなり、注意深く補正する必要がある。なお、感染症拡大以降のオンライン消費の動向については、中島ほか（2022）を参照されたい。

成し、マクロレベルの ALC を作成する。

図表 3 ALC の作成手順



3-1. 「マネーフォワード指数」の作成

先に述べた通り、本分析で用いる3種類のオルタナティブデータのうち、マネーフォワードのデータは、年齢・性別などのグループ別に集計されたデータである。ALCの算出にあたっては、まず、このデータをマクロの人口分布と統合的な形で集約しておく必要がある。

はじめに、集計対象のユーザーを絞り込む。マネーフォワードのユーザーの中には、アプリの利用頻度が少ないユーザー（たまにしか支出記録を更新しない）、平均からかけ離れているユーザー（毎月の収入・支出額が極端に大きい）、法人と思われるユーザーなどが含まれる。こうしたユーザーは、マクロの個人消費動向を捕捉するうえではノイズになりうるため、集計対象から除外する。また、マクロの人口分布を考慮する観点から、年齢、家族構成が分かるユーザーのみを分析に用いる。本分析では、図表4に掲げる7つの条件を全てクリアしたユーザーのみを集計対象とする。このように厳しい条件で絞り込んだうえでみても、集計対象のユーザー数は約31万人と大規模となった。

図表 4 マネーフォワード指数の作成：ユーザーの絞り込み条件

①	アプリ登録から2年以上が経過している
②	毎月利用履歴がある
③	2つ以上の口座を連携している
④	銀行の法人口座を利用していない
⑤	マネーフォワード社の事業者向けサービスを未利用
⑥	1か月間の支出と収入が1,000万円以下
⑦	個人属性が判明している

次に、ユーザーを絞り込んだ後のデータを用いて、品目ごとに集計値を算出する。具体的には、マネーフォワードデータを年齢・家族構成別に集計したうえで、それらをマクロの人口分布に応じて加重平均する。すなわち、ある年齢・家族構成のグループを g 、このグループにおける時点 t における品目 i への支出の単純集計値を $E_{g,i,t}$ 、グループ g における人口（平成 27 年国勢調査ベース）を POP_g とおくとき、時点 t における品目 i への支出は次式で定義される。

$$MFI_{i,t} = \sum_{g \in G} w_g E_{g,i,t} \quad \left(w_g := \frac{POP_g}{\sum_{g' \in G} POP_{g'}} \right) \quad (1)$$

ここで、式(1)の和は集計対象の全てのグループにわたってとることを意味する。この値が、マネーフォワードデータから示唆される、時点 t における品目 i へのマクロの支出額である。この集計値を、当該品目に対する「マネーフォワード指数 (MFI)」と呼ぶ。

ただし、式(1)の右辺に現れるグループ内集計値 $E_{g,i,t}$ はしばしば急激に変動し、マクロレベルの集計値を大きく歪めることがある。例えば、一部のユーザーが極端な支出を行った場合にそのようなことが起こり得る。こうした変動は、マクロの消費動向から乖離している外れ値と考えられることから、集計対象から除外する。具体的には、次の通り外れ値処理を行う。ある品目 i の MFI の前年同月比 $(MFI_{i,t} - MFI_{i,t-12}) / MFI_{i,t-12}$ において、グループ g の支出変化による寄与は

$$e_{g,i,t} := \frac{w_g (E_{g,i,t} - E_{g,i,t-12})}{MFI_{i,t-12}} \quad (2)$$

である。この寄与度 $e_{g,i,t}$ が分布 $\{e_{g,i,t} : g \in G\}$ の上位 1%点を上回る、または下位 1%点を下回る場合に、該当するグループを集計対象から除外し、式(1)の計算を再度行う（集計対象に含まれるグループの個数が変わるため、人口ウエイトもそれに応じて更新する）。こうして得られた外れ値処理後の MFI を、ALC の作成に用いる。

3-2. ALC の作成

このステップでは、①JCB 消費 NOW、②METI POS または GfK、③MFI の 3 指標それぞれについて、消費活動指数の品目別指数とのフィットを確認し、最もよくフィットするオルタナティブデータを、品目ごとに選び出す⁷。

⁷ 3 種類のオルタナティブデータをすべて用いて重回帰分析を行うことも考えられるが、データの変動が似ている場合に多重共線性の問題が生じることから、本分析では、単回帰分析に基づく方法を採用した。

各データのフィットの度合いについては、単回帰分析を用いて評価する。すなわち、式(3)の通り、各品目について上記3指標の前年同月比をとり、消費活動指数（名目原系列）の前年同月比をオルタナティブデータのそれぞれに回帰し、決定係数が最も大きくなる指標を、当該品目に最もよくフィットするオルタナティブデータとして採用する。選ばれた指標を、当該品目の ALC と呼ぶことにする（水準ベースではなく前年同月比ベースである点に注意）。単回帰分析にあたっては、2018年1月から実績終期までをサンプル期間としている⁸。

$$\left(\frac{\text{CAI}_{i,t}}{\text{CAI}_{i,t-12}} - 1 \right) \times 100 = \alpha_i + \beta_i X_{i,t} + \varepsilon_{i,t} \quad (3)$$

$$\left[\begin{array}{l} X_{i,t}: (1) \text{JCB 消費 NOW} \quad (2) \text{METI POS, GfK} \quad (3) \text{MFI のいずれか} \\ \text{[いずれも前年同月比 (\%)]} \end{array} \right]$$

ただし、自動車については、消費活動指数の基礎統計である新車登録台数がオルタナティブデータより早いタイミングで利用可能になることから、オルタナティブデータを用いる代わりに、新車登録台数を ALC に用いる⁹。また、一部の品目（生命保険、公共放送など）については、3種類のオルタナティブデータのいずれについても消費活動指数との決定係数がゼロ近傍、または回帰係数 β_i がマイナスとなってしまう場合がある。消費活動指数の基礎統計とオルタナティブデータのどちらが実態に近いのか、あるいはなぜこのような乖離が生じているかを論じることは本稿の射程外であるが、消費活動指数をできるだけ正確に予測するという観点からは、当てはまりの良くない品目についてオルタナティブデータを用いるのは適切ではない。そこで、こうした品目については、オルタナティブデータを用いず、消費活動指数の基礎統計を、季節性を考慮した ARIMA モデル (SARIMA モデル) を用いて機械的に補外推計したものを、ALC に用いる。この際、SARIMA モデルの次数は、赤池情報量基準 (AIC) で最適と判定されたものを採用する（モデル判定にあたっては、データが遡れる 2003 年 1 月から実績終期までをサンプル期間としている）。

ここまでで、品目別の ALC がすべて揃うことになる。ただし、ALC は、前年同月比の動きが消費活動指数の品目別指数と似ているとしても、それらの平均や標準偏差は必ずしも一致しない。消費活動指数をできるだけ正確にナウキャストするという目的に照らせば、このズレを補正することが望ましい。そこで、ALC を、式(3)の

⁸ 感染症拡大前（2018年1月～2020年3月）をサンプル期間として推計しても、最適なオルタナティブデータはほぼ変わらなかった。

⁹ オルタナティブデータを用いて消費活動指数の「自動車」をナウキャストすることも可能であるが、オルタナティブデータでは新車登録台数の動きをうまく捉えられない。これは、① JCB 消費 NOW については、自動車購入に占めるクレジットカードの利用率が低い、② マネーフォワードについては、消費活動指数の基礎統計に含まれない項目（中古車購入費や自動車ローンの支払いなど）も集計対象としている、といった理由による。

単回帰で得られた係数を用いて、次式の通りに補正する。

$$ALC_{i,t}^{\text{adjusted}} = \hat{\alpha}_i + \hat{\beta}_i \times ALC_{i,t} \quad (4)$$

ここで、 $\hat{\alpha}_i$ および $\hat{\beta}_i$ は式(3)の単回帰で得られた係数である。以下、ALC は、すべて補正後ベースのものを指すことにする。

最後に、ここまでに算出した品目別の ALC を合成して、消費形態別（耐久財、非耐久財、サービス、総合）の ALC を作成する。これらがマクロレベルの ALC に該当する。このとき、消費活動指数の算出方法に合わせるため、ALC を一度水準ベースに変換してから合成する。具体的には次の通りである。

先に述べた通り、ALC は前年同月比ベースの指数である。これを用いれば、各品目について、消費活動指数の水準をナウキャストリングすることができる。

$$CAI_{i,t}^{\text{nowcast}} = CAI_{i,t-12} \times \left(1 + \frac{ALC_{i,t}^{\text{adjusted}}}{100} \right) \quad (5)$$

耐久財、非耐久財、サービスの3形態については、上式で得られた品目別の指数水準を消費活動指数の品目ウェイトで加重平均し、当該形態の水準ベースの予測値を算出する。式(6)において、和は集計対象の消費形態に属する全ての品目にわたってとることを意味し、 w_i^{CAI} は、その消費形態における品目*i*の構成比（消費活動指数ベース）を表す。この指数の前年同月比（「真の」消費活動指数に対する変化率）を、当該形態の ALC と定義する（式(7)）。総合については、上記3形態の指数水準を国民経済計算の形態別ウェイトで加重平均し、前年同月比をとったものを、ALC と定義する。

$$CAI_t^{\text{nowcast}} = \sum_i w_i^{\text{CAI}} CAI_{i,t}^{\text{nowcast}} \quad (6)$$

$$ALC_t = \left(\frac{CAI_t^{\text{nowcast}}}{CAI_{t-12}} - 1 \right) \times 100 \quad (7)$$

以上が ALC の計算手順である。オルタナティブデータを用いてナウキャストリングを行う消費活動指数の基礎統計の一覧は、図表5の通りである。消費活動指数ウェイトベースで7割強を占める品目について、3種類のオルタナティブデータのみを用いてナウキャストリングすることになる¹⁰。

¹⁰ オルタナティブデータによるナウキャストリングの対象外の品目は次の通り（カッコ内は消費活動指数の品目ウェイト）。生命保険（5.98）、通信業（5.18）、自動車（4.12）、金融（1.83）、介護（1.82）、自動車整備（1.62）、冠婚葬祭（1.38）、学習塾（1.33）、駐車場（1.26）、損害保険（1.16）、

図表5 オルタナティブデータによるナウキャストिंगの対象品目

消費形態	品目名	CAI基礎統計	CAI品目ウエイト
耐久財	家電	商業動態統計	4.73
非耐久財	飲食料品	商業動態統計	19.59
	医薬品、化粧品など	商業動態統計	6.50
	衣料品	商業動態統計	4.85
	電気・ガス・水道	電力調査統計など	4.03
	ガソリン、灯油など	商業動態統計	2.75
	たばこ	鉱工業出荷指数	1.88
サービス	外食	サービス産業動向調査	11.28
	交通	第3次産業活動指数	5.56
	医療	第3次産業活動指数	5.00
	娯楽	第3次産業活動指数	4.79
	旅行・宿泊	サービス産業動向調査など	1.97
合計			72.93

(注) 交通は、鉄道旅客、バス、タクシー、航空旅客を CAI ウエイトで加重平均したもの。
(出所) 日本銀行

4. 分析結果

本節では、上記のプロセスで得られた ALC が、消費活動指数にどの程度フィットしているかを確認する。本稿作成時点で入手したオルタナティブデータのサンプル期間は、2018年1月～2021年4月である。このデータを用いて2021年4月分の消費活動指数のナウキャストिंगを試みる。また、サンプル期間が限られるため幅をもってみる必要はあるが、ALCの有用性をより詳しく確認するため、データが2019年1月から毎月蓄積されていく仮想的な状況を想定し、ナウキャストिंगを逐次的に行う「リアルタイム予測」も実施した¹¹。

4-1. 品目別の ALC

ALCの集計結果をみる前に、まず、原データにあたる3種類のオルタナティブデータの動きを確認すると(図表6)、総じて、消費活動指数の品目別指数の動きを非常によく捉えており、いずれのオルタナティブデータも、品目別の個人消費を捉えるうえで有用であることが分かる。

そのうえで、それぞれの品目について仔細にみると、「家電」はPOSデータ(GfK)

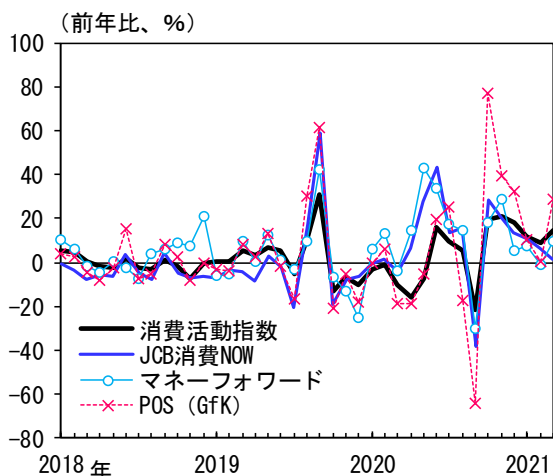
公共放送(0.38)、書籍・雑誌(0.32)、新聞(0.36)、ゲームソフト(0.23)、郵便(0.10)。

¹¹ 厳密には、各データが過去に遡って改定されているため、当時の状況を完全に再現することはできない。ここでの分析は、いわば「疑似リアルタイム予測」と呼ぶべきものである。

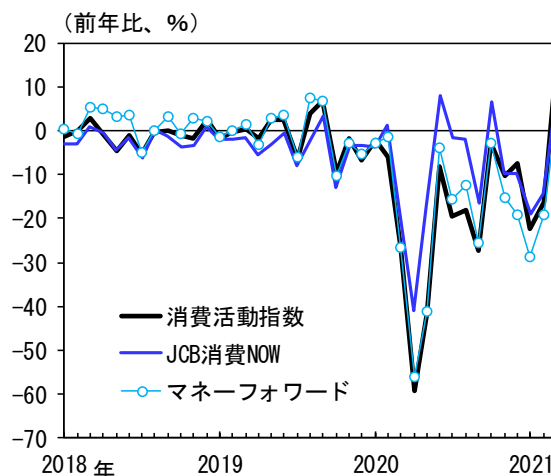
が感染症拡大以降の落ち込みやその後の持ち直しなどをとくに良く捉えている。また、「衣料品」についてはマネーフォワードが、「交通」や「旅行・宿泊」についてはJCB消費NOWが、それぞれ良好なフィットを示している。①「家電」については、(実店舗での購入に限定されるものの)支払手段によらず幅広い商品を集計対象とするGfKの捕捉力が高い、②「衣料品」については、クレジットカードによるインターネット通販での商品購入に加えて現金による店頭での商品購入もカバーしているマネーフォワードの捕捉精度が高いと考えられる。また、③「旅行・宿泊」については、クレジットカードで支出するケースが多いとみられ、JCB消費NOWの捕捉力が高い。これに加えて、消費活動指数と同様にアウトバウンド消費を含んでいないことも、JCB消費NOWが良好なフィットを示した理由であるとみられる(アウトバウンド消費をカバーしているマネーフォワードは、感染症下での旅行・宿泊の減少度合いを消費活動指数対比で大きめに評価した結果、フィットが悪化したとみられる)。

図表6 消費活動指数とオルタナティブデータの比較(品目別)

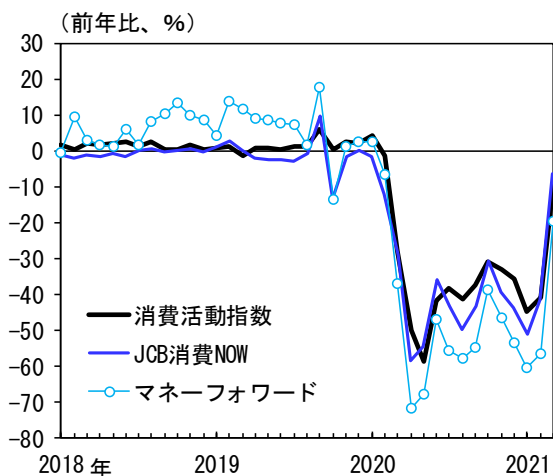
(1) 家電



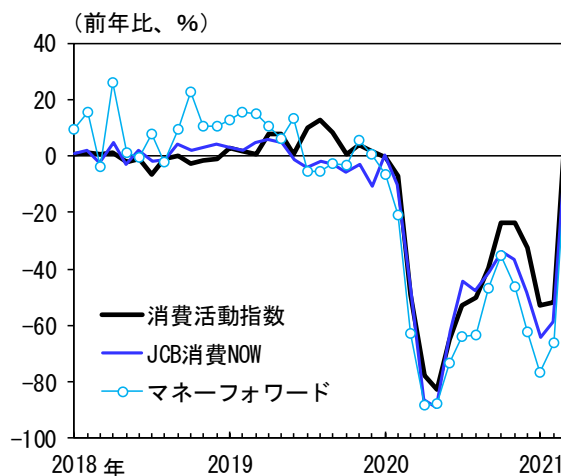
(2) 衣料品



(3) 交通



(4) 旅行・宿泊



(注) 交通は、鉄道旅客、バス、タクシー、航空旅客をCAIウエイトで加重平均したもの。
(出所) JCB/ナウキャスト、マネーフォワード、経済産業省、日本銀行等

上記の品目を含め、オルタナティブデータを用いてナウキャストイングを行ったすべての品目のフィットの状況を、図表7に示す。3種類のオルタナティブデータの中で最も良いフィットを示す系列、すなわち「品目別 ALC」は、ほとんどの品目で消費活動指数の基礎統計との相関係数が 0.8 を上回っており、フィットはきわめて良好である（図表7）。これらの結果から、オルタナティブデータで主要な消費活動指数の基礎統計をうまく捉えることができることが分かる。

図表7 消費活動指数の基礎統計と品目別 ALC の相関関係

消費形態	品目名	CAI基礎統計	ALCとの相関係数
耐久財	家電	商業動態統計	0.886
非耐久財	飲食料品	商業動態統計	0.356
	医薬品、化粧品など	商業動態統計	0.735
	衣料品	商業動態統計	0.957
	電気・ガス・水道	電力調査統計など	0.968
	ガソリン、灯油など	商業動態統計	0.973
	たばこ	鉱工業出荷指数	0.879
サービス	外食	サービス産業動向調査	0.962
	交通	第3次産業活動指数	0.981
	医療	第3次産業活動指数	0.846
	娯楽	第3次産業活動指数	0.971
	旅行・宿泊	サービス産業動向調査など	0.970

(注) 交通は、鉄道旅客、バス、タクシー、航空旅客を CAI ウェイトで加重平均したもの。
(出所) JCB/ナウキャスト、マネーフォワード、GfK、経済産業省、日本銀行等

4-2. マクロレベルの ALC

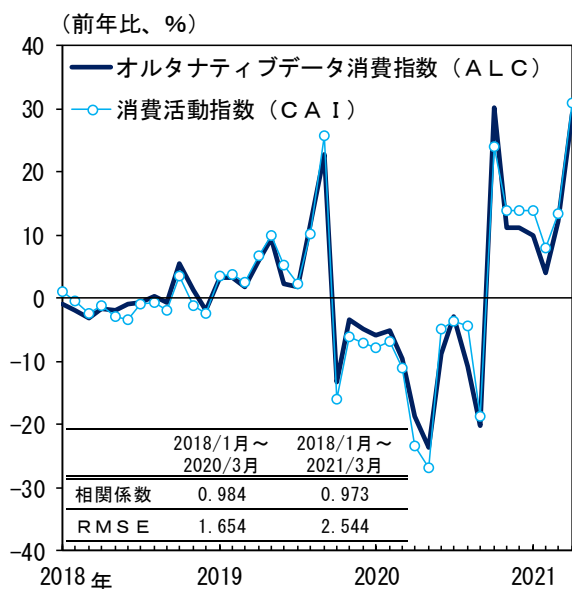
第4-1節で作成した品目別 ALC を消費形態別に集計すると、耐久財・非耐久財・サービスのいずれについても、高い精度で消費活動指数をナウキャストイングできている（図表8(1)~(3)）。とくに、2020年春以降の、感染拡大に伴う消費活動の大きな変動もよく捉えているほか、耐久財については、消費税引き上げ前後の駆け込み・反動もかなり正確にとらえている。統計的にみると、相関係数はどの消費形態も 0.9 を上回っている。また、サンプル期間中の平均的な誤差を二乗平均平方根誤差（Root Mean Squared Error、以下、RMSE）で評価すると、もともと振れの大きい耐久財こそ幾分大きいものの、それ以外の消費形態の RMSE は 0~1%ポイント台と小さい。いずれの消費形態も感染症拡大の影響を強く受けた時期を含めたサンプル期間（2018年1月~2021年3月）の RMSE は、それ以前（2018年1月~2020年3月）の RMSE よりも大きいものの、最大でも 2%ポイント台半ばであり、感染症拡大の前後によらず ALC のパフォーマンスは優れていると評価できる。

また、総合ベースの ALC についてみると、相関係数は 0.9 を大きく上回っているほ

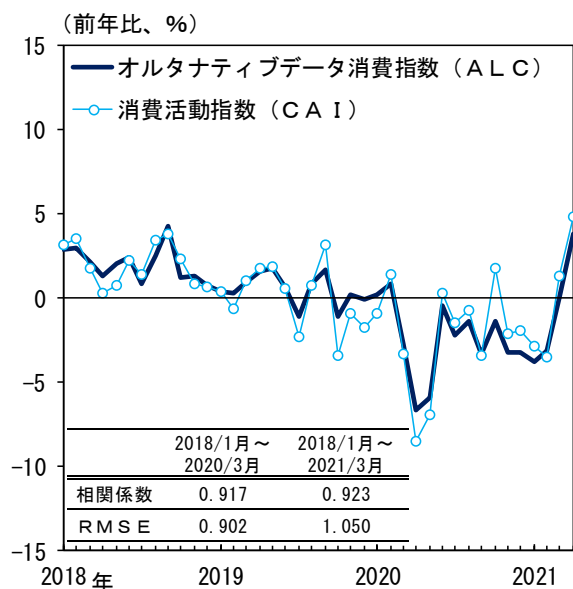
か、RMSEは1%ポイントを下回っており、サンプル期間全体として、消費活動指数とのフィットはきわめて良好である(前掲図表8(4))。とくに、新型コロナウイルス感染症の影響による2020年4~5月の大幅な落ち込みや、その後の昨年夏場から秋にかけての持ち直しを的確に捉えられている。このように、オルタナティブデータを用いれば、感染症下でみられたような消費の急激な変動についても、翌月中旬という比較的早いタイミングで、高い精度でのナウキャストが可能となる。

図表8 マクロレベルのALC

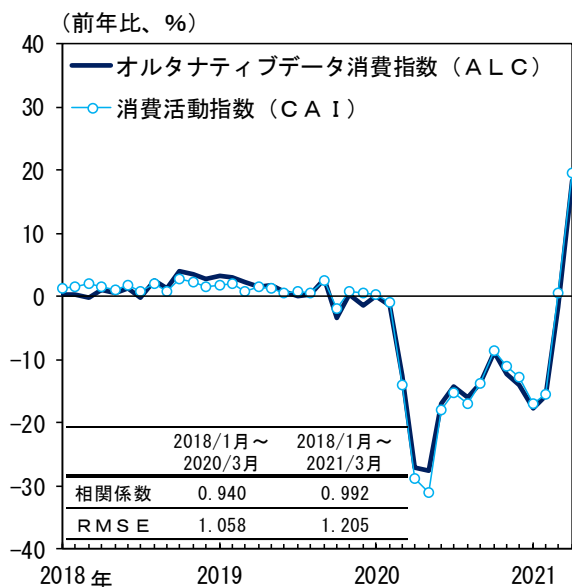
(1) 耐久財



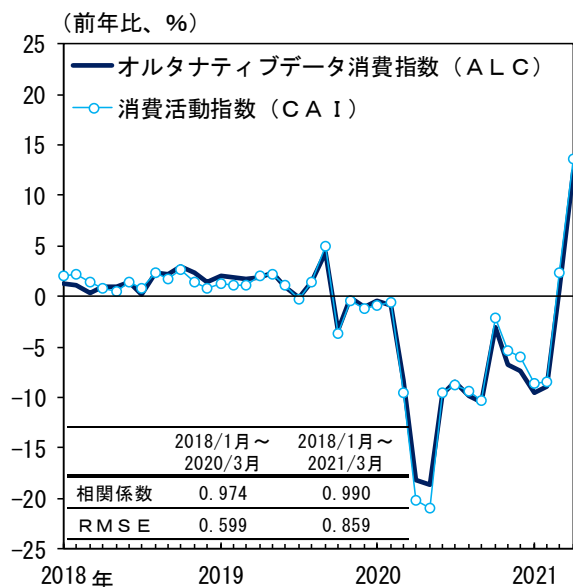
(2) 非耐久財



(3) サービス



(4) 総合



(出所) JCB/ナウキャスト、マネーフォワード、GfK、経済産業省、日本銀行等

4-3. リアルタイム予測に基づく性能評価

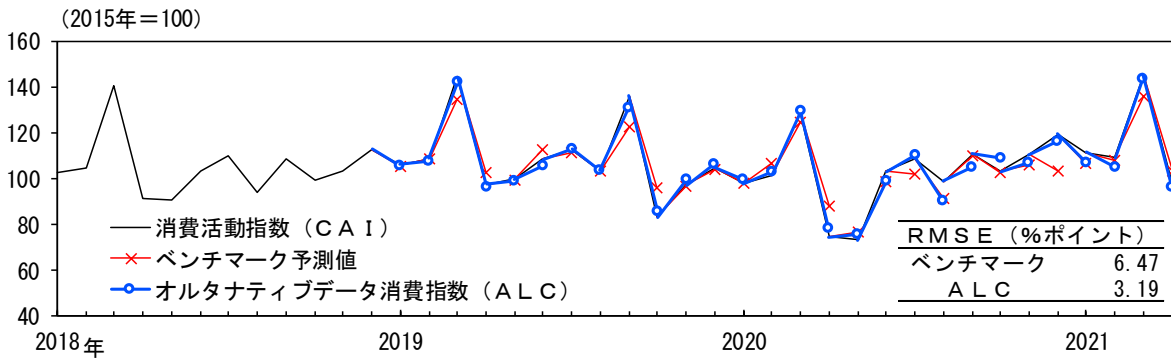
ここまでの議論は、フルサンプル（2018年1月～2021年3月）の推計によって過去のデータへのインサンプルでみた当てはまりの良さを確認したものである。本節では、別の角度から ALC の有用性を確認するため、毎月、オルタナティブデータを入手したタイミングで消費活動指数をナウキャストイングするという仮想的なセットアップの下で「リアルタイム予測」を行う。ALC の性能評価にあたっては、消費活動指数のすべての品目について SARIMA モデルを適用して算出した予測値をベンチマークとする。以下では、テスト期間を 2019 年 1 月～2021 年 4 月とし、各月において、オルタナティブデータが出揃うタイミング（月の中旬）で消費活動指数のナウキャストイングを行う状況を想定する。

リアルタイム予測の結果によると（図表 9）、耐久財、非耐久財、サービスのいずれについても、ALC がテスト期間全体にわたってベンチマーク予測値よりも正確に消費活動指数をナウキャストイングできている。とくに、感染症拡大直後（2020 年春頃）に急激に落ち込んだ時期には、サービスのベンチマークは（過去のトレンドに復する形で）大幅な回復を予測していた一方、ALC は比較的正確に予測できている。総合ベースで見ると、ALC の予測誤差（RMSE）はベンチマークの半分未満となっており、ALC がマクロの個人消費動向をうまく捕捉できている。

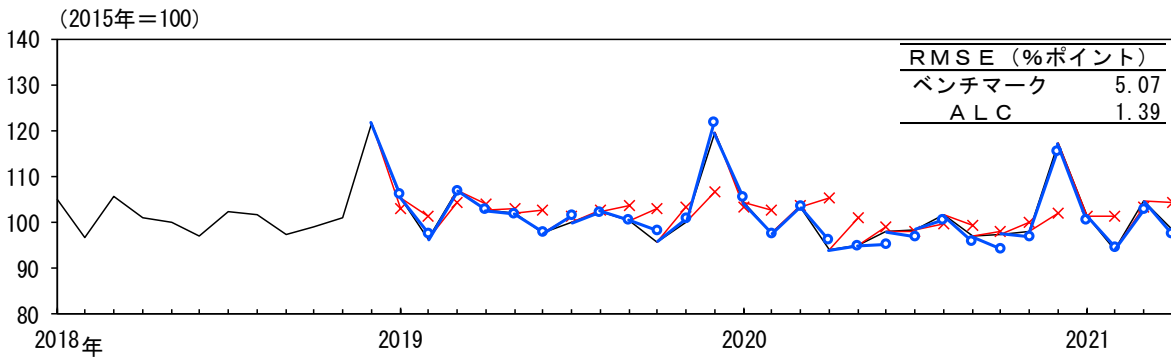
このように、総じてみればリアルタイム予測の精度は良好である。データのサンプル期間が 2018 年 1 月以降と短いため（2019 年初時点では、各品目についてサンプルは 10 か月程度）、結果は幅をもってみる必要はあるものの、ALC は実務的に活用し得る優れた指標であると結論できる。

図表9 リアルタイム予測結果

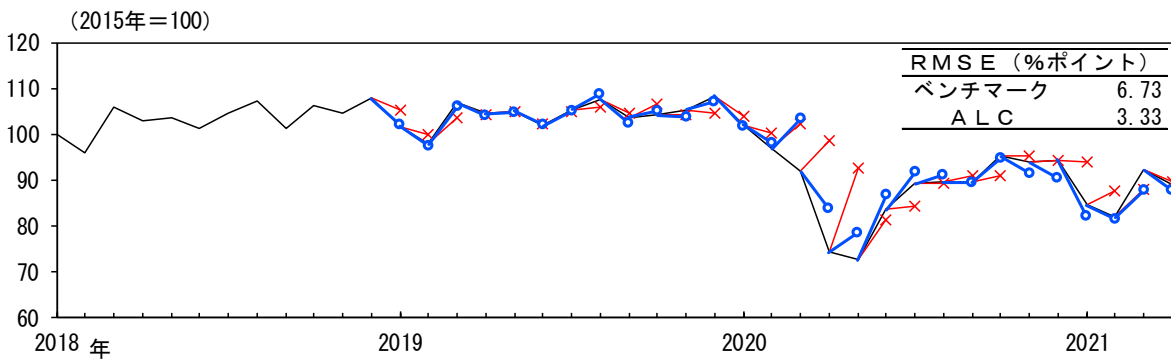
(1) 耐久財



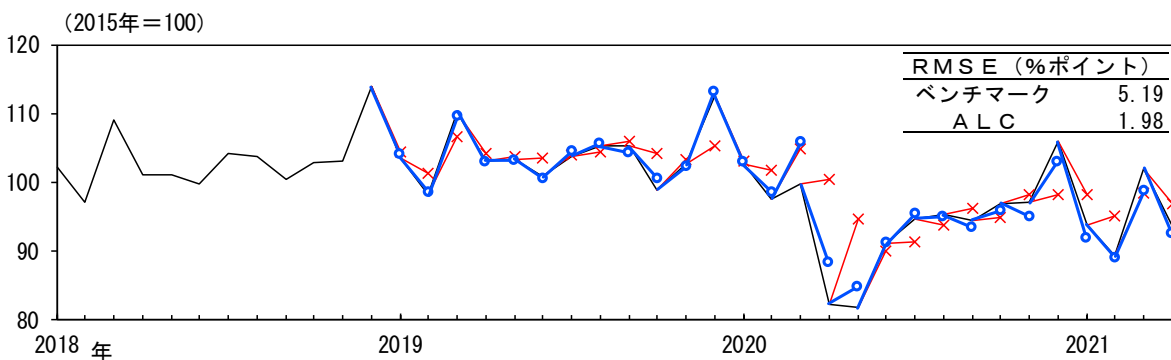
(2) 非耐久財



(3) サービス



(4) 総合



(注) RMSEは前年同月比をもとに算出。サンプル期間は2019年1月～2021年4月。
 (出所) JCB/ナウキャスト、マネーフォワード、GfK、経済産業省、日本銀行等

5. まとめ

本稿では、消費関連の複数のオルタナティブデータを用いて、マクロの個人消費動向をナウキャストした「オルタナティブデータ消費指数 (ALC)」を作成した。ALC は品目別・業種別にみた消費動向の把握のみならず、マクロの消費動向のナウキャストという観点でも良好な予測パフォーマンスを示しており、2020 年春以降の感染拡大に伴う消費活動の大きな変動についても、消費活動指数の公表よりも約 3 週間早いタイミングで、的確に捉えることができた。本稿の分析は、オルタナティブデータをナウキャストに用いる際には、①オルタナティブデータの定義や集計対象は必ずしもマクロ統計と一致しないこと、②同データには外れ値やバイアスが混在しやすいこと、などに注意を払うことが予測精度を高めるうえで重要であることを示している。

なお、本稿はオルタナティブデータの速報性に焦点を当てたが、オルタナティブデータのカバレッジの広さや粒度の高さを生かしたリサーチも有益と考えられる。例えば、JCB 消費 NOW やマネーフォワードは、近年増加しているオンラインでの消費活動（E コマース、ネット予約サイトを介した旅行・宿泊など）も広くカバーしていることから、実店舗が主な調査対象である既存統計では捉えきれない新たな消費トレンドをみるうえで、有用であると期待される。また、詳細な属性別データは、疑似パネルデータとしても有用性が高く、家計の行動様式などに関するより構造的な分析に活かせる面もあろう。オルタナティブデータとそれを用いるユーザーの知見が蓄積されていけば、既存統計とオルタナティブデータのズレについても、掘り下げた分析が可能になるかもしれない。こうしたリサーチを進めていくことは、今後の課題としたい。

参考文献

浦沢聡士 (2022)、「GDP ナウキャスト：成果と課題—オルタナティブデータの活用に向けて—」、AI・ビッグデータ経済モデル研究会、2022 年 2 月 4 日。

王悠介・須合智広・高橋耕史・松村浩平 (2021)、「位置情報データによる経済活動のナウキャスト」、日本銀行ワーキングペーパーシリーズ、No. 21-J-2。

亀田制作 (2021)、「オルタナティブデータを用いた日銀リサーチの紹介」、日銀レビュー・シリーズ、No. 2021-J-12。

高橋優豊・近藤卓司・宗像晃・大久保友博・岩崎雄斗 (2021)、「近年の消費行動の変化を踏まえた消費活動指数の推計方法の見直しについて」、日本銀行調査論文。

中澤崇 (2022)、「オルタナティブデータを用いた GDP ナウキャストモデルの構築」、日本銀行ワーキングペーパーシリーズ、No. 22-J-1。

- 中島上智・高橋優豊・八木智之（2022）、「新型コロナウイルス感染症拡大前後のオンライン消費動向の分析」、日本銀行ワーキングペーパーシリーズ、No. 22-J-5.
- 中村康治・河田皓史・田中雅樹・植前理紗（2016）、「消費活動指数について」、日本銀行調査論文.
- 日本銀行調査統計局（2022）、「東京大学金融教育研究センター・日本銀行調査統計局第9回共催コンファレンス：『ウィズコロナ・ポストコロナの日本経済』の模様」、日本銀行調査論文.
- Carvalho, V. M., J. R. Garcia, S. Hansen, Á. Ortiz, T. Rodrigo, J. V. R. Mora, and P. Ruiz (2021), "Tracking the COVID-19 Crisis with High-resolution Transaction Data," *Royal Society Open Science*, 8, pp. 210–218.
- Chapman, J. T. E., and A. Desai (2021), "Using Payments Data to Nowcast Macroeconomic Variables During the Onset of COVID-19," Bank of Canada Staff Working Paper, 2021-2.
- Chikamatsu, K., N. Hirakata, Y. Kido, K. Otaka (2021), "Mixed-Frequency Approaches to Nowcasting GDP: An Application to Japan," *Japan and the World Economy*, Vol. 57, 101056.
- Bounie, D., Y. Camara, and J. W. Galbraith (2020), "Consumers' Mobility, Expenditure and Online-Offline Substitution Response to COVID-19: Evidence from French Transaction Data," *Econometric Modeling: Microeconomic Models of Household Behavior eJournal*.
- Dunn, A., K. Hood, and A. Driessen (2020), "Measuring the Effects of the COVID-19 Pandemic on Consumer Spending Using Card Transaction Data," BEA Working Paper Series, WP2020-5.
- Dunn, A., K. Hood, A. Batch, and A. Driessen (2021), "Measuring Consumer Spending Using Card Transaction Data: Lessons from the COVID-19 Pandemic," *AEA Papers and Proceedings*, Vol. 111, pp. 321–325.