

民間企業が保有する消費関連データの 消費動向指数（C T I）への利活用について

～ 民間データを用いた総消費動向指数の改善に関する調査研究 ～

令和4年12月28日

総務省統計局統計調査部消費統計課

総務省統計研究研修所統計技術向上支援課

(独)統計センター統計技術・提供部技術研究開発課

I はじめに

1 消費動向指数（C T I）の概要

消費動向指数は、家計調査の結果を補完し、消費全般の動向を捉えるためのデータとして総務省統計局が開発中の参考指標であり、世帯消費動向指数（C T I ミクロ）及び総消費動向指数（C T I マクロ）から成る（図1参照）。

世帯消費動向指数（C T I ミクロ）は、我が国における世帯の消費支出の推移を示す指数であり、世帯消費に関する三調査（家計調査、家計消費状況調査及び家計消費単身モニター調査）によるデータを合成して得られた消費支出（一世帯当たり平均月額）について、基準年を100として指数化したものである。

総消費動向指数（C T I マクロ）は、我が国における世帯全体の消費支出総額（G D P 統計の家計最終消費支出に相当）の推移を推測する指数であり、推定した月次の消費支出総額について、基準年を100として指数化したものである。

総務省では、総務省等（統計局、統計研究研修所及び独立行政法人統計センター）、学識経験者及び設立趣旨に賛同いただいた企業で構成する「消費動向指数研究協議会」（以下「協議会」という。）を、2017年に設立し、消費動向指数の開発に必要となるデータの分析、協議等を開始した。2018年1月分から、既存統計を利用した指数の公表を開始し、民間データの利活用による指数の改善について、研究を進めている。

2 本報告書の目的と構成

本報告書は、協議会の枠組みの下で進めてきた消費動向指数の開発のうち、主にC T I マクロにおける民間データの利活用に向けた研究成果を取りまとめたものである。第Ⅱ章ではその分析手法と分析結果、第Ⅲ章では民間データの入手、分析、管理等の実務を行う中で直面した主な課題について述べる。第Ⅳ章はまとめである。なお、本報告書の分析を行うに当たっては、独立行政法人統計センターにおいて開発したRプログラムを使用している。

本研究を進める上では、参考となる前例が少なかったこともあり、試行錯誤を繰り返すこととなった。特に、本件研究の対象である消費動向は、食料を始めとした多様な品目の動きの集合体であるため、各種の消費関連データのカバレッジなどについて、地道に知見を積み上げる必要があった。また、新型コロナウイルス感染症による消費の急変により、速報性が期待できる民間データの利用ニーズが高まった一方で、予測精度を確保する上での困難さも生じた。以下では、こうした課題に対する検討過程についても触れている。本報告書の内容が、広く民間データの利活用を促進していく上での課題に対して、解決の一助となれば幸いである。

ビッグデータ等を活用し、消費動向をマクロ・ミクロの両面から捉える速報性の高い消費指標の体系：消費動向指数（CTI：Consumption Trend Index）を新たに開発し、

□ 2018年1月分から参考指標として公表開始

□ 2021年7月分公表時に、2020年基準改定を実施

世帯消費動向指数 （CTIミクロ）

世帯の平均消費支出額（10大費目別、世帯類型別など）の月次動向を示す統計指標

- ◆ 家計調査（標本規模：二人以上の世帯約8千、単身世帯約7百）の結果を、
 - 家計消費単身モニター調査（標本規模：2千4百）
 - 家計消費状況調査（標本規模：約3万）
 の結果等と統計的手法によって補正・補強し、標本規模を擬似的に拡大、推計精度を向上



総消費動向指数 （CTIマクロ）

国内経済における個人消費総額（GDPにおける家計最終消費支出）の月次動向を示す統計指標

- ◆ GDP統計（家計最終消費支出）をターゲットとして、最新の動向を推測
- ◆ GDP統計の四半期別公表値では観測できない月次の値を時系列回帰モデルによって推計
- ◆ ビッグデータ利活用成果に関する報告書をリリース

図1 消費動向指数（CTI）の概要

II 民間データを用いた総消費動向指数の改善に関する分析

1 分析の概要

総消費動向指数（C T I マクロ）は、G D P 統計の家計最終消費支出を目的変数、月次で公表される世帯消費動向指数、サービス産業動向調査結果、商業動態統計調査結果、鉱工業生産指数及び第3次産業活動指数を説明変数とした状態空間モデルに基づく時系列回帰モデルにより算出している。

説明変数には、公表される消費関連統計の原数値そのものではなく、状態空間モデルに基づく季節調整法により、毎月最新の原数値から推定したトレンド及びサイクル系列（T C 系列）を使用している。これにより説明変数が過去に遡って改定されるため、C T I マクロの結果数値も、毎月、過去に遡って改定される。

また、説明変数のうち、サービス産業動向調査結果及び第3次産業活動指数（以下「サービス統計」という。）は、C T I マクロの月次結果の算出時に当該月の結果が公表されていないため、これらの値は、時系列モデルを用いた自身のトレンド及びサイクルを延長推定することにより求めて、説明変数に使用している（推定方法の詳細は、参考資料[1]総務省(2022)を参照）。このため、消費動向の変動が大きい場合は、毎月のC T I マクロの結果数値には、当該月のサービス関連の消費動向を十分に取り込めない場合がある。こうした場合、翌月のC T I マクロの結果作成時に、サービス統計の当該月の公表値を取り込むと、C T I マクロの当該月以前の結果が大きく改定されることがある。特に、新型コロナウイルス感染症の流行に伴い、消費動向が極めて特異な動きとなった2020年以降は、結果が大きく改定されることがしばしばあった。

このようなことから、本分析では、直近のサービス関連の消費動向をC T I マクロに反映するため、C T I マクロの公表月のデータが入手できないサービス統計について、民間企業が保有するデータ等により当該月の値を予測推定することを試みた。

具体的には、サービス産業動向調査結果の「サービス産業計」及び第3次産業活動指数の「広義対個人サービス」の直近の値を、民間企業から提供を受けたクレジットカード情報などを活用して予測推定した。また、この予測推定結果を用いて、C T I マクロの結果を2019年9月から2022年1月分まで、名目値29回、実質値28回、計57回算出し、公表結果と比較した。

2 分析手法

本節では、C T I マクロの公表月のデータが入手できない、サービス産業動向調査結果のサービス産業計及び第3次産業活動指数の広義対個人サービスの予測推定に利用する民間データ及び予測推定手順等について解説する。

2.1 民間企業から提供を受けたデータの概要

直近の月の値の予測推定に当たっては、協議会の参画企業から提供を受けたクレジットカード情報による月次の売上高（以下、「クレジットカード情報」という。）を利用した。

このクレジットカード情報は、決済の発生時点における金額を月別に集計した売上高のデータであり、支払の翌月以降に発生したキャンセルの払戻しは、支払月には計上されていない。本来、キャンセルの払戻しについては、支払月と同月に払戻しを計上するなど、予測推定を行う統計の作成方法に即した処理を行うことが望ましい。特に、2020年3月以降は、新型コロナ

ウイルス感染症の流行に伴い、事前に予約・支払を行う消費について相当数のキャンセルが発生したため、クレジットカード情報上、旅行業などの業態においては、払戻しにより売上げがマイナスとして記録されている月もあったことから、払戻しの影響は小さくないと考えられる。

また、クレジットカード情報には、法人企業による支払金額も含まれているほか、カード会員や契約店舗の増減にも影響を受けることに注意が必要である（こうしたデータの特性については、第Ⅲ章(2)を参照されたい）。

C T I マクロは個人消費の動向を推定するものであるため、こうしたクレジットカード情報のデータ特性についても、利用前に適切な処理を行うことが望ましいが、本分析では、C T I マクロの推定時点で入手可能なデータを用いることを優先し、提供を受けたクレジットカード情報をそのまま利用することとした。

なお、実際のデータの利用に当たっては、類似した項目の売上高を幾つかの業態別に統合して集計している。統合に当たっては、日本標準産業分類などを参考に、意味のあるものとなるように配慮した。また、特異な動きを示す系列については除外した。

クレジットカード情報のうち、予測推定に利用した期間は、2017 年 4 月から 2022 年 1 月までの 58 か月分である。

2.2 第 3 次産業活動指数及びサービス産業動向調査結果の予測モデル

第 3 次産業活動指数の算出方法の詳細は公開されていないため、それを完全に再現した予測推定は困難である。また、実務の負担等を考慮すると、簡易な方法が望ましい。

こうしたことから、第 3 次産業活動指数（広義対個人サービス）については、広義対個人サービスに占めるウェイトが比較的高い「小売業」と、同じくウェイトが比較的高く他のサービス業とは消費動向が異なる「医療業」について独立させ¹、小売業及び医療業以外の産業（以下「除く小売・医療業」という。）と合わせて 3 つの産業の値を、それぞれ予測推定することとした。

一方、サービス産業動向調査結果（サービス産業計）においても、サービス産業計に占める医療業の割合が比較的高いが²、その影響は第 3 次産業活動指数（広義対個人サービス）ほど大きくないと判断し、サービス産業動向調査結果のサービス産業計は内訳に分割せず、直接、予測推定することとした。

2.2.1 小売業の予測モデル

第 3 次産業活動指数の小売業は、商業動態統計調査結果における小売業 8 業種の月次販売額及び対応する C P I から作成されている。これらは、C T I マクロの公表月と同時点の結果が利用可能であるため、小売業については、商業動態統計調査結果及び C P I を用いて予測推定することとした。なお、先述のとおり、第 3 次産業活動指数における推定方法の詳細は不明なため、次に示す予測モデルを設定した。

第 3 次産業活動指数の小売業の予測推定は、小売業指数の一階階差を目的変数、商業動態統

¹ 2015 年基準における第 3 次産業活動指数のウェイトをみると、広義対個人サービス（ウェイト 4887.5）に占める小売業（ウェイト 1182.6）の比率は約 24%、医療業（ウェイト 892.6）の比率は約 18%である。

² 2021 年結果をみると、サービス産業計に占める医療業の比率は約 11%である。

計調査結果の売上高の一階階差を説明変数とした重回帰モデルにより行う。推定手順及びモデル式は以下のとおり。

- ①商業動態統計調査結果の小売業8業種の売上高をC P I で実質化
- ②実質化した8業種の売上高を単純加算により4系列に統合³
- ③統合する業種の組合せは、回帰結果における自由度調整済決定係数が最大のものを選定⁴
- ④モデル式：

$$y_t^b = \sum_i \alpha_i z_{i,t}^b + \beta + \varepsilon_t \quad (1)$$

$y_t^b = y_t - y_{t-1}$: 小売業指数 y_t の一階階差 (t は時点)

$z_{i,t}^b = z_{i,t} - z_{i,t-1}$: 商業動態統計調査結果 $z_{i,t}$ の一階階差 ($i(= 1,2,3,4)$ は小売業8業種を統合した分類)

小売業の予測推定に用いる商業動態統計調査結果については、小売業8業種のうち「無店舗小売業」の調査が開始された2015年7月以降の公表値を使用した。これにより、各業種をそのまま説明変数に用いた場合、使用するデータの期間に対して説明変数の数が多いことから、予測モデルの推定精度への影響を考慮し、小売業8業種を統合して重回帰モデルの説明変数とした。

2.2.2 医療業の予測モデル

第3次産業活動指数の医療業は、診療報酬月額支払確定状況(社会保険診療報酬基金及び(公社)国民健康保険中央会)により作成されているが、その月次結果は、第3次産業活動指数よりも後に公表される。このため、毎月公表されている医療業指数の最新月の値は、前月までの医療業指数の結果から何らかの方法により算出された延長推定値となっている。医療業指数の延長推定値は、月次の変動が大きい時期は安定的であるが、変動が大きい時期は、翌月の結果公表時に大きく改定されることがある(図2参照)。医療業指数におけるこうした特性に起因する影響を除外するため、本分析の推定では、医療業指数の最新月の公表値は使わず、別途予測推定したものを用いる。またこのことから、医療業指数は、以下のとおり、前月と最新月の二期分の予測推定を行う(図3参照)。

(ア) 一期目(前月の値)の予測：サービス産業動向調査結果の医療業により一期先予測⁵

(イ) 二期目(最新月の値)の予測：クレジットカード情報の医療業により一期先予測

医療業指数の予測推定は、(ア)一期目(前月の値)及び(イ)二期目(最新月の値)のいずれについても、前年同月比の階差を変数とする重回帰モデル(2)により行う。推定手順は以下のとお

³ 統合後の系列数は、幾つかのパターンについて試算を行い、自由度調整済決定係数などを指標に最適な数を設定した。

⁴ 回帰結果及び選定結果については割愛する。

⁵ 一期目の予測にクレジットカード情報ではなく、サービス産業動向調査結果を利用しているのは、後日公表される医療業指数の確報値との当てはまりが良いためである。なお、サービス産業動向調査結果の公表は、第3次産業活動指数の公表の約2週間後であるが、C T I マクロの推定時に、一期目の予測に使用するサービス産業動向調査結果が利用可能であるため、同方法を採用した。

り。

- ①一期目（前月の値）：第3次産業活動指数の医療業指数を目的変数、サービス産業動向調査結果の医療業を説明変数として予測推定
- ②二期目（最新月の値）：第3次産業活動指数の医療業指数に、①で予測推定した結果を追加した系列を目的変数、クレジットカード情報を説明変数として予測推定
- ③モデル式：

$$y_t^b = \sum_i \alpha_i x_{i,t}^b + \beta + \varepsilon_t \quad (2)$$

$y_t^b = r_t - r_{t-1}$: 目的変数 y_t の前年同月比 ($r_t = y_t/y_{t-12}$) の階差
 $x_{i,t}^b = s_{i,t} - s_{i,t-1}$: 説明変数 $x_{i,t}$ の前年同月比 ($s_t = x_{i,t}/x_{i,t-12}$) の階差 (i は、一期目予測では、サービス産業動向調査結果の医療業、二期目予測では、クレジットカード情報の医療関連業態)

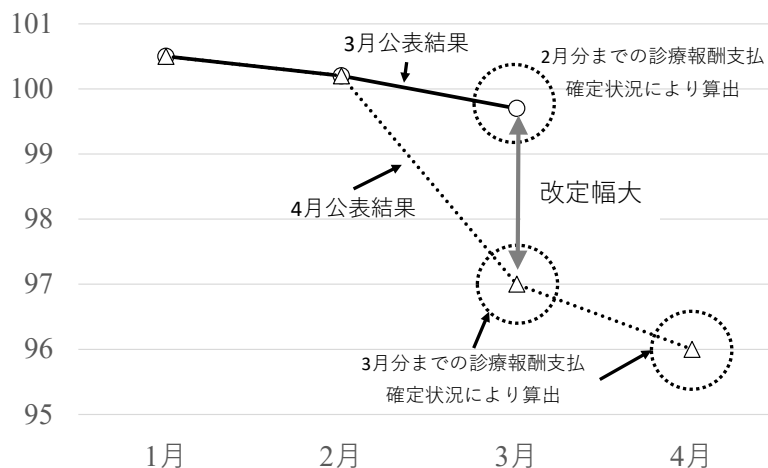


図2 第3次産業活動指数における医療業指数の算出方法イメージ（3月分及び4月分結果作成時）

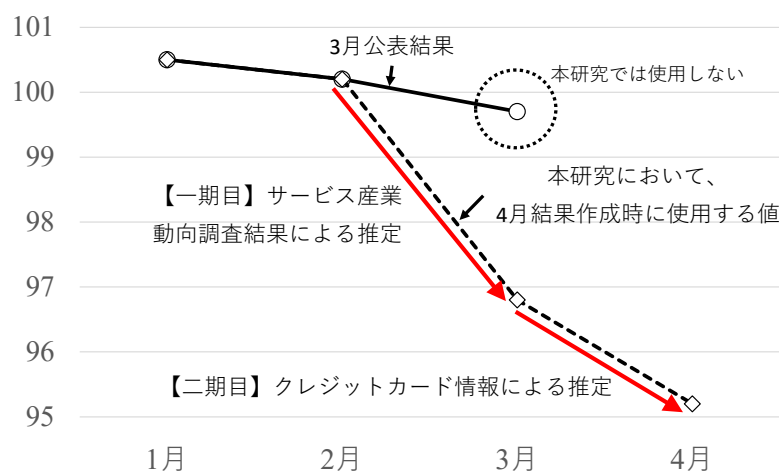


図3 本分析における医療業指数の推定方法イメージ（4月結果作成時）

2.2.3 サービス産業計及び「除く小売・医療業」の予測モデル

サービス産業動向調査結果のサービス産業計及び第3次産業活動指数の広義対個人サービス（「除く小売・医療業」）は、クレジットカード情報により予測を行う。利用できるクレジットカード情報の期間が58か月とそれほど長くないため（第2.1節参照）、予測モデルはシンプルなものとした。なお、目的変数にはサービス統計の公表値を、説明変数にはクレジットカード情報を用いるが、両者の季節変動は一般に異なること、重回帰による推定は定常性を前提とすることから、データの変換方式については、対前年同月比の階差及び対前年同月に対する対数変化率の階差の2種類とし、それぞれについて予測推定を行うこととした。

また、各月の曜日の数、うるう年及び日本型休日数の影響（以下「曜日効果等」という。）を除去するデータの事前処理を行ったものと、行わないもの、それぞれについても予測推定を行うこととした。曜日効果等の除去の手順は次のとおり。

- ① サービス産業計及び「除く小売・医療業」の原系列から、X-13-ARIMA-SEATSにより曜日効果等を推定⁶
- ② 推定した曜日効果等を、クレジットカード情報から除去⁷

サービス産業計及び「除く小売・医療業」の予測推定は、前年同月比及び対前年同月に対する対数変化率の階差を変数とする重回帰モデルにより行う。また、モデル式(3)において、 y_t^b の定義式における第2項(r_{t-1})を右辺に移項すると、 r_t を変数としたAR(1)モデルとなることから、式(4)で定めるAR(1)モデルも検討対象とした。なお、説明変数の候補となる業態の分類が細かく、これをそのまま使用した場合、データ期間に対して説明変数の数が多くなってしまうため、予測モデルの推定精度への影響を考慮し、幾つかの業態を選定・統合して重回帰モデルの説明変数としている。推定手順及びモデル式は以下のとおり。

- ① クレジットカード情報の業態から複数を選定し、3系列に統合⁸
- ② 統合する組合せは、回帰係数 α_i が全て正となる組合せのうち、モデル式(3)では自由度調整済決定係数が最大のものを、モデル式(4)ではAIC(Akaike Information Criterion: 赤池情報量基準)が最小となる分類の組を選定⁹
- ③ 表1に示す、モデル式、データの変換方式及び曜日効果等の除去別の、8種類の推定モデルのうち、推定を行った期間の公表値とのRMSPE (Root Mean Squared Percentage Error :

⁶ X-13-ARIMA-SEATSによる曜日効果等の推定では、内部で用意された曜日変数(tdnolpyear、tdlnolpyear)及びうるう年変数(lpyear)に加え、日本型の休日数を調整するためのユーザー変数を作成し、回帰変数としてReg-ARIMAモデルに取り込んでいる。本分析では、土日以外の法令上の祝日に加え、12月31日から1月3日を社会的慣習に基づく休日(1月1日は法令上の祝日)として扱い、モデルに取り込んでいる。

⁷ 利用するクレジットカード情報の期間が、公的統計と比べて短いことから、目的変数の公的統計から推定した曜日効果等により、クレジットカード情報から曜日効果等を控除した。現時点における曜日効果等は、公的統計から推定したものであり、クレジットカード情報における曜日効果等とは異なる可能性があることに注意が必要

⁸ 統合後の系列数は、幾つかのパターンについて試算を行い、自由度調整済決定係数などを指標に最適な数を設定した。

⁹ 個々の業態(分類)の売上げの伸びは、売上げ全体の伸びに対してプラスの寄与を与えることが自然と考え、回帰係数 α_i を全て正となる組合せに制限している。

平方平均二乗パーセント誤差)¹⁰が最小となるものを選定

④モデル式：

$$y_t^b = \sum_i \alpha_i x_{i,t}^b + \beta + \varepsilon_t \quad (3)$$

$$r_t = \mu + \varphi r_{t-1} + \sum_i \alpha_i x_{i,t}^b + \varepsilon_t \quad (4)$$

$y_t^b = r_t - r_{t-1}$: 目的変数 y_t の前年同月比 ($r_t = y_t/y_{t-12}$) 又は対数変化率 ($r_t = \text{nlg}(y_t/y_{t-12})$ ¹¹) の階差
 $x_{i,t}^b = s_{i,t} - s_{i,t-1}$: 説明変数 $x_{i,t}$ の対前年同月比 ($s_{i,t} = x_{i,t}/x_{i,t-12}$) 又は対数変化率 ($s_{i,t} = \text{nlg}(x_{i,t}/x_{i,t-12})$) の階差 (i は、クレジットカード情報における業種又は業種を統合した分類)

表 1 モデル式、データの変換方式及び曜日効果等の除去別、推定モデル一覧

推定モデル	モデルの説明
RI 1	モデル式=(3)、変換方式=対前年同月比、曜日効果等の控除なし
RLI 1	モデル式=(3)、変換方式=対数変化率、曜日効果等の控除なし
RAR 1	モデル式=(4)、変換方式=対前年同月比、曜日効果等の控除なし
RLAR 1	モデル式=(4)、変換方式=対数変化率、曜日効果等の控除なし
RI 2	モデル式=(3)、変換方式=対前年同月比、曜日効果等の控除あり
RLI 2	モデル式=(3)、変換方式=対数変化率、曜日効果等の控除あり
RAR 2	モデル式=(4)、変換方式=対前年同月比、曜日効果等の控除あり
RLAR 2	モデル式=(4)、変換方式=対数変化率、曜日効果等の控除あり

⑤予測推定期間：

- ・サービス産業計 : 2019 年 9 月～2022 年 1 月
- ・「除く小売・医療業」: 2019 年 10 月～2022 年 1 月

<モデルの選定結果>

付録Aの表 A-1 及び A-3 に、サービス産業計及び「除く小売・医療業」の予測推定結果の、速報値に対する RMSPE を掲載する。これらの結果をみると、サービス産業計及び「除く小売・医療業」のいずれの予測においても、表 1 の推定モデルのうち、RLI 2 の RMSPE が最小となる時点が最も多かった。

表 A-1 のサービス産業計についてみると、2020 年 3 月までは、RI 1 及び RLI 1 の RMSPE が最小となっているが、2020 年 4 月以降、ほぼ全ての時点において RLI 2 の RMSPE が最小となっ

¹⁰ RMSPE (平方平均二乗パーセント誤差) は、以下の式により算出した。

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_t \left(\frac{\text{予測値}_t - \text{公表値}_t}{\text{公表値}_t} \right)^2}$$

¹¹ Neg-log 変換を表す。クレジットカード情報の売上高には負の値が含まれているため、これを考慮して neg-log 変換とした。

た。なお、RI 1、RLI 1、RI 2 及び RLI2 の RMSPE をみると、2021 年 3 月まではおおむね同水準となっているが、RI 1 及び RI 2 について、2021 年 4 月及び 5 月の予測精度が極端に悪いことにより、2021 年 4 月以降、両モデルの RMSPE が拡大した。特に、RI 2 については、2021 年 5 月以降、RMSPE が更に大きく拡大していることが確認できる。

一方で、表 A-3 の「除く小売・医療業」では、RI 2 の RMSPE が最小となる期間と RLI 2 の RMSPE が最小となる期間が交互に表れるといった傾向が 2021 年初頭まで確認された。しかし、2021 年 3 月までの RMSPE をみると、RI 2 と RLI 2 の差は小さいが、2021 年 4 月以降は、RI 2 の RMSPE がやや拡大し、RLI 2 の RMSPE が最小となっていることが確認できる。

同様に確報値との比較も行った。表 A-2 及び A-4 に、サービス産業計及び「除く小売・医療業」の予測推定結果の、確報値に対する RMSPE を掲載する。サービス産業計についてみると、速報値の場合と異なり、RMSPE が最小となる時点が最も多いモデルは、RI 1 と RLI 2 となった。ただし、RI 1 の RMSPE は、2021 年 4 月を境にやや拡大し、直近の 2 か月間では RLI 2 が最小となった。一方、「除く小売・医療業」は、速報値と同様に、RI 2 の RMSPE が最小となる期間と RLI 2 の RMSPE が最小となる期間が交互に表れるといった傾向が 2021 年初頭まで確認された。なお、標本交代及び基準改定の影響により、速報値に対する RMSPE と単純に比較することはできないが、確報値に対する RMSPE と速報値に対する RMSPE の差は小さいことから、速報値及び確報値をおおむね同等の精度で予測できていると考えられる。

以上をまとめると、サービス産業計及び「除く小売・医療業」のいずれにおいても、直近の結果を重視する今回の試算結果においては、モデル RLI 2 による結果が、最も妥当であることが確認できた。しかし、時点によっては他のモデルの RMSPE が最小となる場合もあり、モデルの選定については、引き続き検証が必要といえる。

3 分析結果

3.1 サービス産業計及び広義対個人サービスの予測推定結果

ここでは、本分析における予測モデルのパフォーマンスの確認をする。

付録 A の図 A-1～A-10 に、サービス産業計、広義対個人サービス、内訳である小売業、医療業及び「除く小売・医療業」の、直近の対前年同月増減率（％）に対する予測推定結果と、各統計の速報値及び確報値に対する対前年同月増減率¹²との比較図を掲載する。

<小売業>

図 A-5 及び A-6 をみると、小売業の予測値は、速報値及び確報値のいずれもほぼ正確に予測できていることが確認できる。

<医療業>

速報値（図 A-7）との比較では、一期目予測値及び二期目予測値のいずれも、速報値との乖離が目立っている。特に、値が大きく変化する 2020 年 3 月及び 4 月並びに 2021 年 4 月及び 5 月における乖離が顕著である。これは、第 2.2.2 項で述べたとおり、医療業の速報値が前月までの診療報酬支払確定状況から何らかの方法により延長推定されていることに起因しており、

¹² 「除く小売・医療業」の対前年同月増減率は、公表値を使用し、総務省において算出した値である。

サービス産業動向調査結果やクレジットカード情報により補完して予測を行った結果との相違が表れたものである。

一方、確報値（図 A-8）との比較では、速報値との比較において目立つ乖離がみられた 2020 年 3 月及び 4 月並びに 2021 年 4 月及び 5 月の結果について、予測値との差が縮小していることが確認できる。特に、一期目予測の結果については、比較を行った全時点において、確報値をほぼ正確に予測できており、本分析において用いた簡易的なモデルによっても、比較的高い精度での予測が可能であることが確認できた。二期目予測の結果についても、2020 年 4 月などで一期目の予測推定結果より大きな乖離がみられるものの、全体的にはおおむね確報値と整合的な結果が得られている。

< サービス産業計、「除く小売・医療業」 >

図 A-1～A-4 をみると、サービス産業計及び「除く小売・医療業」共に、その予測値は、試算を行った期間を通して、速報値及び確報値のいずれもおおむね精度良く予測できていることが確認できる。

なお、図 A-1～A-4 に掲載した予測結果は、今回の試算において公表値との RMSPE が最小となる時点が最も多かったモデル RLI 2 によるものである。また、この期間のうち 2020 年 5 月並びに 2021 年 1 月及び 5 月の時点で乖離が大きくなっている要因としては、クレジットカード情報固有の動きの影響や、サービス統計には事業所向けの産業活動が含まれるなど、クレジットカード情報とのカバレッジの相違などが要因として考えられる。一方で、先述の 3 時点は、新型コロナウイルス感染症の流行期において、緊急事態宣言の発令や解除など、大きな転換となる施策が実施された時点であることもあり、施策の実施に対して敏感に反応するクレジットカード情報との相違も要因として考えられる。

< 広義対個人サービス >

小売業、医療業及び「除く小売・医療業」を統合した広義対個人サービスについては、一期目予測の結果（図 A-9 及び A-10 の赤線）をみると、速報値及び確報値のいずれもおおむね精度良く予測できていることが確認できる。また、二期目予測の結果（図 A-9 及び A-10 の青線）についても、医療業（A-7 又は A-8 の青線）の影響を受け、一期目予測より乖離が大きくなる傾向がみられるものの、全体的にはおおむね速報値及び確報値と整合的な結果が得られている。系列ごとの対前年同月増減率の予測結果と公表値の差（表 A-5）をみても、広義対個人サービスの予測において、速報値との差と確報値との差に大きな相違はみられず、医療業の結果が広義対個人サービスの予測に与える影響は全体的に小さいことが確認できる。

ただし、表 A-5 及び図 A-13 に示すとおり、医療業指数の速報値は、2020 年 4 月のように大きく改定される場合がある。こうした時点においては、医療業の予測結果は、広義対個人サービスの予測精度を大きく左右し得ることから、本分析において、簡易なモデルで医療業の確報値をおおむね整合的に予測できたことは、重要な意味を有している。

以上から、クレジットカード情報を活用することで、比較的シンプルな予測モデルによって直近のサービス関連の消費動向をおおむね高い精度で予測できることが確認された。一方で、新型コロナウイルス感染症の影響による消費の急変時期においては、予測精度がやや劣る点も確認された。これへの対応としては、公的統計とクレジットカード情報のカバレッジの相違な

ど、その特性を把握し、適宜、データクレンジングを行うなどの措置が有効と考えられる。

3.2 C T I マクロの推定結果

付録Bの図B-1～B-4に、2019年9月（又は10月）から2022年1月までの各月における推定時点のC T Iマクロの推定結果を同一図内にプロットしたものを掲載する。ここで、同図の下段にある棒グラフは、各月の推定結果における最新月の値と、翌月分の推定結果の当該月の値との差（以下「改定幅」という。）を表しており、これにより、毎月、過去に遡って値が改定されるC T Iマクロの安定性を確認することができる¹³。なお、C T Iマクロは、家計最終消費支出を目的変数として推定していることから、改定幅等をみる際は、四半期ごとに家計最終消費支出の改定の影響が含まれることに留意されたい。

<改定幅の縮小>

図B-1及びB-3はそれぞれ、実質及び名目の公表値をプロットしたものである。新型コロナウイルス感染症の流行により、外出自粛や営業自粛要請などが発出され、国内の消費動向が大きく変動した2020年3月から10月頃までの期間において、最新月の改定幅が大きい様子や、比較的過去の結果も大きく改定される様子が見て取れる。一方、試算結果（実質は図B-2、名目は図B-4）をみると、公表値と比較して、2020年3月から同年10月までの結果の改定が抑えられている様子が確認できる。

<結果の安定性>

次に、直近の結果の安定性をみるため、図B-1～B-4の下段に掲載した、公表値と試算結果における最新月の値の改定幅を比較する。公表値（図B-1及びB-3）では、名目及び実質のいずれについても2020年3月から10月頃までの改定幅が、他の月の改定幅よりも極端に大きいことが確認できる。一方、試算値（図B-2及びB-4）は、名目、実質共に、特に、2020年3月から10月頃までの、公表値における改定幅が極めて大きい時点での縮小効果が高く、改定幅の平均値が大きく改善していることが確認できる。

なお、値が大きく変動する時点においては、比較的過去の結果も大きく改定される傾向は、公表値同様、試算値においても確認された（図B-5及びB-6も参照）。

この他、C T Iマクロの推定結果は、説明変数へ設定する外れ値変数の影響も受けることに注意が必要である。今回の検証では、各月の試算値作成において、毎月の公表結果作成時に設定した外れ値変数と同じものを使用している。しかし、サービス産業計及び広義対個人サービスに設定した外れ値変数は、説明変数である月次の結果がある程度蓄積された段階で、変数の追加・見直しを行っているため、説明変数における直近の大幅な変動を十分に結果に取り込めていない可能性がある。また、家計最終消費支出の改定の影響と同様、外れ値変数の見直しの影響は、当月・前月だけでなく過去の値に対しても生じる可能性があることにも注意が必要である。

¹³ 安定性を評価するための指標として、このほかに、季節調整におけるSliding span分析やRevision history分析で使用される指標を利用することも考えられるが、C T Iマクロにおいては、足下の結果の改定状況を特に重視し、直近の値の安定性をみるための指標とした。

参考までに、各月の試算結果と、公表値の比較グラフを、図 B-7 から B-63 に掲載する。

3.3 C T I マクロの四半期結果と G D P 統計の家計最終消費支出（一次速報）との比較

ここでは、C T I マクロの試算結果と、C T I マクロの公表後に公表される家計最終消費支出（一次速報）との比較を行う。具体的には、家計最終消費支出との近さ（家計最終消費支出との期ごとの残差平方及び RMSE（Root Mean Squared Error：二乗平均平方根誤差））及び動き（前期比及び前年同期比の符号）を確認し、家計最終消費支出との整合性について検証する。また、月次結果と同様に、最新月の値の改定幅を確認し、結果の安定性についても評価する。

<家計最終消費支出との整合性>

表 C-1 及び C-2 に掲載した期別の残差平方をみると、残差平方が小さい期の数が、名目では同数である一方、実質では試算結果の方が、より多くの時点で小さいことが確認できる。ただし、試算を行った期間全体での近さを測る RMSE をみると、名目、実質共に、公表値よりも試算結果の方が小さく、クレジットカード情報を利用することで、家計最終消費支出との差を縮小させる効果が確認できた。ただし、名目については、その効果は極めて小さい。

また、表 C-1 及び C-2 に掲載した試算結果の前期比及び前年同期比をみると、名目及び実質共に、前期比及び前年同期比の符号が家計最終消費支出と一致する時点が公表値と比較して増えていることが確認できる。特に、実質については、試算を行った全期間において、家計最終消費支出と符号が一致していることが確認できる。

以上から、クレジットカード情報を利用することで、家計最終消費支出（一次速報）との整合性を改善する効果がみられた。

<推定精度>

図 C-1～C-18 に C T I マクロ（試算結果、公表値）及び家計最終消費支出（一次速報）の比較図を掲載する。公表値と比較して、特に、新型コロナウイルス感染症の流行の最初期である 2020 年第 1 四半期の結果について、家計最終消費支出との距離が大幅に縮小していることが確認できる。また、その他の時点の結果についても、家計最終消費支出とおおむね整合的な動きをしていることが確認できる。一方で、2020 年第 2 四半期における乖離や、2021 年第 2 四半期における乖離及び前期比の符号の相違などが確認できるが、これは、第 3.2 節でも述べた、目的変数である家計最終消費支出の改定状況や、説明変数の推定の際に設定する外れ値変数の影響を受けたものである可能性もある。

結果の安定性については、表 C-1 及び C-2 から、C T I マクロの公表値及び試算値の、最新期の値の改定幅を確認すると、名目及び実質共に試算結果の方が、改定幅が小さい期の数が多いことが確認できる。ただし、評価期間が、月次の結果と比較して少ないことから、ある程度データがまとまった時期に実施する必要がある。

4 分析結果のまとめ

上記の分析により、クレジットカード情報など早期に利用可能な民間データにより、既存のサービス統計では捉えられていない足下の値を補完することで、直近の消費の変動をいち早く結果に取り込むことが可能となり、C T I マクロの毎月の改定幅を縮小させる効果があること

が確認できた。また、C T I マクロ公表よりも後に公表されるG D P統計の家計最終消費支出（一次速報）との比較において、クレジットカード情報を利用することでRMSEを縮小させる効果や、前期比及び前年同期比の動きとの整合性を高める効果があることが確認できた。

ただし、時点によっては直近の値の改定幅が拡大することや、家計最終消費支出との乖離が大きくなることもあり、C T I マクロ推定の際の外れ値変数の選定を含め、実際の推定に当たっては、更なる改善・検討が必要であるといえる。

また、クレジットカード情報固有の動きの影響や、サービス統計には事業所向けの産業活動が含まれるなど、クレジットカード情報とのカバレッジの相違などにより、一期先予測に差が生じていることが考えられる。利用するデータ特性には十分な理解が必要であるとともに、データの期間により最適な予測モデルが変わるため、推定モデルの選定についても引き続き検証が必要である。

Ⅲ 研究を通じて直面した民間データの利活用に係る課題

研究活動は、データ構造の解析、バイアスの補正及び欠測情報の補完に関する統計数理的方法の研究、収支項目分類の適用・変換方法の検討などの学術的・専門的事項に加えて、データの取扱いに関する民間企業との協議や分析環境の整備等、多岐にわたる。

本章では、これらの活動を通じて新たに直面した課題について、背景事情も含め主なものを列挙する。

(1) 消費関連データの入手

研究当初は、分析用として販売されているPOSデータ以外、民間企業においてどのようなデータが保有されているのか不明であったため、できるだけ多くのデータホルダーのデータの特徴を概観する必要があった。このため、幅広く多くのデータが集まるよう、企業内で一定の集計を経たサマリーデータも含めて、提供を依頼した。その際、データホルダーに新たな負担が生じることを避けるため、既存のサマリーデータも提供の対象に含めることとし、提供のために新たにサマリーデータを作成することは求めなかった。

サマリーデータは、個票レベルのデータと比べて個人情報について考慮する必要が少なくこと及びデータが持つ特性が機密情報を含む場合も一定程度秘匿した形で提供できることから、その提供については多くのデータホルダーから理解を得ることができた。

また、データホルダーが集まる機会が、機密情報が開示される機会につながりかねないと考えデータホルダーに配慮し、各社が一堂に会する会議等は設けず、個別にやり取りをすることとした。協議会設立に向けた連携協力においても、機密情報の取扱い等、データホルダー及び関連団体から聞き取った懸念について必要な対策を検討し、これへの対応を個別に説明した。

このように、複数のデータホルダーとの信頼関係を構築するため、個人情報の保護を考慮することに加え、企業側の負担軽減や情報保護などに対する特段の配慮が必要であった。その点で企業が保有する既存のサマリーデータの利用は有効であったものの、分析の制約が生じた（詳細は(2)で説明）。

今後も民間データの活用拡大を図っていくためには、データホルダー側の負担や情報保護に配慮し、データの形式などの面で柔軟にデータホルダー側と協議していく必要がある。また、本分析において活用した民間データは、分析への利用に同意いただいた民間企業から一時的に提供を受けたものであり、今後も継続して研究や統計作成を行うためには、いかに安定的にデータを入手していくかも課題となる。

(2) データの特性

研究に使用したデータは主に以下の2種類である。

- ① 研究当初において、複数の企業から購入したPOSデータ及びパネルデータ
- ② 協議会参画企業から受領したデータ

これらを公的統計の改善等に利活用した例は少なく、特に消費統計への利活用に際して直接参考になる事例は見当たらなかったため、研究を始めるに当たっては、これらのデータの特性を把握する必要があった。

複数の民間企業から購入したPOSデータ及びパネルデータについて、カバレッジ（家計調査の何品目がカバーできているか）及び指数化した月別支出額の家計調査との相関を確認した

ところ、あるPOSデータの「医薬品」は、カバレッジは一致したが、相関は比較的低かった。原因としては、当該データがドラッグストアの販売データであることから、調剤薬局で処方される薬が含まれていないことなどが考えられた。また、あるPOSデータの「外食を除く食料」では、家計調査との相関は比較的高いものの、毎年12月の支出が家計調査より低くなる傾向がみられた。原因としては、例えば、八百屋などの個人商店での購入やインターネットでの購入等、バーコードで管理されていない商品が当該データに入っていないことなどが考えられた。

協議会参画企業から受領したデータについても、データのカバレッジ、入手可能性及び継続性等について、CTIマクロの説明変数として妥当なものがないか検証を行った。その結果、

- ①家賃、自動車など十分な把握が困難な項目が存在すること
- ②データホルダー間で重複する支出が存在するためダブルカウントの影響を除去する必要があること
- ③データの入手においては、各企業における各種規約等を踏まえる必要や入手のタイミングなど、企業ごとの事情に配慮する必要があること
- ④同質的なデータを過去に遡り長期間にわたり入手することが困難であること

など幾つかの課題が明らかとなった。

また、新型コロナウイルス感染症の拡大に伴い、足下の消費の急変を理解する必要性が高まる中で、分析の対象としたサマリーデータでは情報量が限られるため、キャンセルの影響を検証できない等の制約も判明した。

このほかにも、例えばクレジットカードであれば、キャンペーンによるカード保有者の増加や利用頻度の上昇などのノイズ、保有者に低所得のユーザーが比較的少ないなどのバイアス、現金・クレジットカード等の支払手段の構成比の変動などを取り除くことが、サマリーデータでは困難であることも明らかとなった。

こうした事情に対応するためには、例えば個人情報を含むデータの提供が困難な場合において、世帯分布に合わせて重みをつけた集計をデータホルダーに依頼することも考えられたが、無償でのデータ提供を基本とする中で新たな負担をデータホルダーに求めることは現実的ではなかった。

CTIマクロの説明変数として民間データのみを使う試算なども行ったが、民間データ固有の動きの影響を受けて特異な動きとなる場合があり、やはりそのままで利用は困難であることが明らかとなった。こうしたデータ特性に関する制約に直面する中で、創意工夫を重ねた結果、CTIマクロの算出スキームの大枠を維持した上で、補完的に直近の説明変数を予測するために民間のサマリーデータを使うという分析手法につながることもあった。

以上のように、民間企業が保有するデータは、あらかじめ統計作成のために設計された統計調査によるデータと異なり、そのままでは統計処理に適さない場合も多い。そのため、時系列モデルなど実際の統計分析に差し掛かる前段階で、データの特性把握を丁寧に行う必要があると考えられる。

(3) データ保管・分析環境の整備

協議会参画企業から提供を受けたデータは、一般に販売されているデータとは異なり、データの收受、保管及び利用に当たっては、セキュリティに細心の注意を払う必要がある。また、一方で、提供を受けるデータの多様性及び多量性に鑑みて、効率的なデータ処理及び分析を可

能とする環境の整備も必要である。これらのことから、データの収受、保管及び利用をセキュアに行うことを可能とし、かつ効率的・効果的な研究を行うことを可能とする専用のクラウド環境を構築し、研究活動の円滑な遂行を試みた。

本クラウド環境では、外部からの侵入防止措置を講じた上で、当該クラウド環境に接続してデータを分析するには、ID、パスワード及び物理キーによる認証に加え、データ格納領域が無いシステム専用のパソコンを用いる必要があるなど、データの持ち出しを防止する措置もとることとした。その結果、各種ソフトウェアのバージョンアップその他の運用負荷が発生することとなったことに加え、研究担当者は、通常の行政事務で使用する情報システムと切り離された別の環境でデータを取り扱うこととなり、実務上の一定の不便を受け入れる必要も生じた。また、企業により求めるセキュリティポリシーがまちまちであるため、構築したクラウド環境のセキュリティレベルが、一部企業のセキュリティポリシーを十分に満たしているとはいえず、データ提供に係る円滑な交渉の足かせとなる場合も生じた。このようなことから、今後も、民間データを活用するに当たっては、企業側のセキュリティポリシーや取り扱うデータの性質などを事前に十分に把握した上で、セキュリティ確保や分析環境整備にも負担が発生することを認識し、リソースを確保する必要がある。

IV おわりに

民間データを活用した総消費動向指数の改善については、民間企業の協力の下、第Ⅱ章第3節で示したとおり、クレジットカード情報などの有効活用により、CTIマクロの毎月の改定幅を縮小させる効果があることが確認できた。引き続き、この推定方法の安定性や妥当性、モデルの改善などの検証を行う必要がある。

また、研究の過程においては、データの特性などに関して様々な課題も明らかとなった。新型コロナウイルス感染症の流行により、足下で消費が急変する状況に直面する中、様々な民間データが普及する環境下にある。民間データの更なる有効活用に向けては、データの入手拡大やその特性把握のためにも、引き続きデータホルダーとの連携の下、試行錯誤をしながら取組を継続していくことが必要である。

<学会での報告>

- ・消費動向指数（CTI）の作成への民間企業データ活用に関する検討（2019 年度経済統計学会全国研究大会）
- ・消費動向指数（CTI）の作成への民間企業データ活用に関する検討（2019 年度統計関連学会連合大会）
- ・消費動向指数（CTI）の推計結果の改善に関する検討（2020 年度統計関連学会連合大会）
- ・民間データを活用した総消費動向指数の結果の改善について（2021 年度経済統計学会全国研究大会）
- ・民間データを活用した総消費動向指数の結果の改善について（2021 年度統計関連学会連合大会）

<消費動向指数研究協議会 参画企業（令和4年11月現在）>

株式会社アイディーズ
イオン株式会社
株式会社インテージ
株式会社 NTT ドコモ
株式会社オリエントコーポレーション
株式会社クレディセゾン
株式会社 Zaim
CCCMK ホールディングス株式会社
株式会社ジェーシービー
Segment of One & Only 株式会社
株式会社セブン&アイ・ホールディングス
株式会社 True Data
株式会社パスモ
東日本旅客鉄道株式会社
ビザ・ワールドワイド
株式会社 BCN
株式会社マクロミル
Mastercard
株式会社マネーフォワード
三井住友カード株式会社
三菱 UFJ ニコス株式会社
ユーシーカード株式会社
株式会社ロイヤリティ マーケティング

（五十音順）

<消費動向指数研究協議会 評議員（令和4年11月現在）>

佐藤 整尚 東京大学大学院経済学研究科准教授
高部 勲 立正大学データサイエンス学部教授
星野 崇宏 慶應義塾大学経済学部教授

（五十音順）

参考資料

- [1] 総務省(2017)、消費動向指数（C T I）の開発に向けて、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報告資料
<https://www.stat.go.jp/info/kenkyu/sss/pdf/report.pdf>
- [2] 総務省(2022)、2020年基準 総消費動向指数（C T I マクロ）の推定方法
https://www.stat.go.jp/data/cti/pdf/macro_ref_2020.pdf