

令和4年度

中小企業の実態把握等のためのデータ利活用に関
する委託調査事業

調査報告書

令和5年3月

デロイト トーマツ コンサルティング合同会社

目 次

1. はじめに	3
1-1. 本業務の背景.....	3
1-2. 本業務の目的.....	3
1-3. 本業務の実施概要.....	4
1-4. 本報告書の構成.....	5
2. 中小企業の売上動向把握のための調査・分析.....	6
2-1. 調査方法	6
2-1-1. データの選定.....	6
2-1-2. 指標の算出方法.....	7
2-1-3. 業種選定.....	12
2-1-4. 指標の評価方法.....	14
2-2. 有識者研究会の設置・運営.....	20
2-2-1. 研究会の設置概要.....	20
2-2-2. 各回研究会の議題及び要旨.....	21
2-2-3. 研究会を踏まえた対応.....	24
2-3. 調査結果	24
2-3-1. 指標の評価.....	25
2-3-2. 評価結果を踏まえた指標の課題.....	68
3. 事業者を取り巻く外部環境の構造化及びデータ利活用可能性に関する調査・分析	73
3-1. 因果型共起構造分析.....	73
3-1-1. 分析実施方針.....	74
3-1-2. 分析対象データの選定.....	80
3-1-3. 分析結果.....	83
3-1-4. 構造化データの調査.....	104
3-2. 因果構造分析.....	114
3-2-1. 分析実施方針.....	114
3-2-2. 分析結果.....	119
4. 本事業で得られた示唆と次年度以降のアクションプラン	133
4-1. 本事業で得られた示唆.....	133

4-1-1. データ利活用の可能性.....	133
4-2. 次年度以降のアクションプラン.....	144
4-2-1. 行政における継続的なオルタナティブデータの活用方法の検討.....	144
4-2-2. ユースケースの実現に向けた対応.....	146
4-3. 本事業のまとめ.....	149

- ・ 本報告書に記載されている情報は、調査時点のものであり、公開情報を除き、ヒアリング等で第三者から提供を受けた資料を基礎としております。これら入手した情報自体の妥当性・正確性については、弊社では責任を負いません。
- ・ 本報告書における分析手法は、多様なものがありうる中での1つを採用したに過ぎず、その正確性や実現可能性に関して、弊社がいかなる保証を与えるものではありません。
- ・ 本報告書は、調査委託契約に従って貴庁の政策決定の参考資料として作成されたものです。内容の採否や使用方法については、貴庁自らの責任で判断を行うものとします。

1. はじめに

1-1. 本業務の背景

新型コロナウイルス感染症（以下、「感染症」という。）の流行や原油・原材料価格の高騰など、足下において中小企業を取り巻く環境は激しく変化しており、これまで以上に、可能な限り実態を速報的に捉えたり、今後の教訓として過去の実態を詳細に振り返ったりしながら、必要な政策対応の検討及び実施を迅速かつ的確に行うことが求められている。この点に関して、我が国で伝統的に利用されてきた政府統計は引き続きマクロ経済の動向を捉える上では有用であるものの、中小企業に限定して動向を把握できる経済指標は限られており、特に詳細性や速報性の面でも課題がある。また、こうした政府統計だけでは、外部環境を構成する個々の諸要素間の連関性や、中小企業の経済活動への波及効果といったプロセスが必ずしも明らかになっていない。

これに対して、クレジットカード決済データ、POS データ、位置情報データ、SNS、新聞等のテキストデータといったオルタナティブデータは、感染症流行後に国内外の中央銀行で経済動向を把握・分析する上で活用されるなど、その注目度が高まっている。その活用実態や個々のデータに関する特性（精度、偏り等）、有効性等の検証は途上にあるものの、詳細性や速報性に優れる点から公的統計を補完する役割が期待されている。同様に、中小企業の動向を的確かつ迅速に把握することや外部環境の構造を体系的に理解することにつながる可能性も考えられる。

1-2. 本業務の目的

このような状況を鑑み、本事業では、「中小企業の売上動向を把握するための調査・分析」、「中小企業を取り巻く外部環境の構造化及びデータ利活用可能性に関する調査・分析」などを通じて、これまで以上に充実して中小企業の経済活動や外部環境の変化を把握し、厳しい状況にある中小企業の実態を踏まえた適切かつ迅速な政策的支援につなげられるように、オルタナティブデータの有効活用を検討することを目的として推進した。

1-3. 本業務の実施概要

本事業では、まず「(1) 中小企業の売上動向把握のための調査・分析」として、株式会社ジェーシービー（以下、「JCB社」という。）と株式会社ナウキャスト（以下、「ナウキャスト社」という。）が共同開発したサービス「JCB消費NOW」を活用し、国内のクレジットカードの決済情報データを選定した。ナウキャスト社と共同で、「JCB消費NOW」データから中小企業基本法で定める中小企業を抽出する方法の検討を行い、中小企業の売上動向を定期的に把握するための指標を業種別に試作し、試作した指標の有効性・妥当性の評価を実施した。有効性・妥当性の評価に際しては、研究会を設置・運営し、経済学や統計学、データサイエンス等に知見のある学識者・民間人材から意見を聴取し、指標の更なる精度の向上や指標の活用可能性等に向けた助言・指導等を受けた。また、試作した指標の活用可能性に関して、中小企業の売上動向を注視している地方自治体等へのヒアリング調査を実施し、追加的なユースケースを整理した。

また、「(2) 事業者を取り巻く外部環境の構造化及びデータ利活用可能性に関する調査・分析」においては、社会情勢や経済動向を把握するための重要な情報源と考えられる新聞などのテキストデータのデータベースを利用した。テキストデータを活用することによって、定量的な指標だけでは把握困難な事象間の因果関係を明らかにし、示唆を得ることが期待できる。テキストデータの活用は、企業における特定業務領域では進んでいるところもあるが、行政における活用事例はいまだ少ない。本事業では、行政におけるテキストデータの活用可能性を調査することを目的として、「感染症」「円安」「原材料・資源価格の高騰」「サプライチェーン混乱」「震災」といった具体的な経済・社会上の変動であるトピックを選定し、テキストデータを用いた因果型共起構造分析と因果構造分析の2つのアプローチ手法による分析を行った。因果型共起構造分析においては、新聞のテキストデータから、同時期に発生している複数の事象間の因果関係を考慮してネットワーク的に可視化するための分析を行った。分析の際には、テキストマイニングの専門的知見を有する学識者に意見聴取を行い、事業者の外部環境を構成する個々の諸要素を仮説的にネットワークとして構造化する際の検討に活用した。加えて、ネットワーク化された個々の要素の動向把握に対応した構造化データの存否についても調査を行った。因果構造分析においては、当該分析に関する専門的知見を有する東京大学和泉潔研究室（以下、「和泉研究室」という。）と連携して取り組んだ。また、因果構造分析では、決算短信および新聞のテキストデータに記述された因果関係を示す手がかり表現を基に、特定の経済・社会上のトピックが中小企業の各業種に対して与え得る波及効果

を探索した。これらの仮説的に作成したネットワークに関しては、精度の評価および妥当性の検討を行い、今後の活用方法の提案を行った。

これらの調査を踏まえ、各業務の調査結果を整理するとともに、それらに基づいて導かれる今後の取組に向けた示唆について、本報告書に取りまとめた。

1－4. 本報告書の構成

前項を踏まえ、本報告書は全4章構成とする。

まず第2章では、中小企業の売上動向把握のために試作した指標の内容と指標の有効性・妥当性評価の実施結果を取りまとめた。第3章では、事業者を取り巻く外部環境のネットワーク構造化のために実施した分析の内容について取りまとめた。最後に第4章では、本事業全体を通して得られた示唆と、それを踏まえて次年度以降で取り組んでいくことが望ましい事項について取りまとめた。

2. 中小企業の売上動向把握のための調査・分析

2-1. 調査方法

中小企業の売上動向を的確かつ迅速に把握することを目的として、本調査では、国内のクレジットカード決済データを用いて、中小企業の売上動向を定期的に把握するための指標を業種別に5つ以上試作し、試作した指標の有効性・妥当性を評価した。

本項では、指標の試作で用いたデータや中小企業基本法で定める中小企業の抽出方法、試作した業種、有効性・妥当性評価方法等を示す。

2-1-1. データの選定

本調査の目的に鑑み、中小企業の売上動向を把握することが期待できるオルタナティブデータとして、ナウキャスト社が提供する「JCB 消費 NOW」を選定した。

「JCB 消費 NOW」は、統計対象会員のクレジットカードのトランザクションデータ（JCB カードの実取引データ）を基にした統計データで、インターネットショッピングやキャッシュレス決済を含む幅広い消費の動向を捉えることが可能である¹。約1,000万会員分の豊富な決済情報サンプルを持ち、業種別指数62種と非常に広範なデータを保有している。また、更新頻度は約2週間であり、業種別の消費動向が把握できる地域別の最小単位は都道府県であることから、速報性や詳細性に優れており、本調査の目的に合うオルタナティブデータだと考えられる。加えて、内閣府のV-RESAS²や日本銀行をはじめとする各種金融機関での利活用実績も多い。

¹ 消費者におけるクレジットカードの取引データを活用することで、近年消費者の利用率が高まっているキャッシュレス決済の消費者行動を捉えることが期待できる。一方で、クレジットカードの取引データからは、現金支払いや銀行振込などのキャッシュレス決済以外の支払い方法に関する情報が得られず、また、キャッシュレス決済においても、クレジットカード以外の支払い方法（電子マネー、QRコード決済など）については把握することが困難であるため、これらの問題を考慮したデータ補正が必要である。

² <https://v-resas.go.jp/>

JCB消費NOWとは

JCBカード会員の匿名加工情報を用いた速報性の高い指数を提供

- サンプルデータ
 - <サンプル数>
 - 日本各地のJCBグループカード会員から、約1000万会員を無作為抽出することにより、属性（居住地・年代・性別）に偏りが少ないサンプルを元に、指数算出をしております。
 - <有用性>
 - 匿名加工情報を活用しており、統計情報を用いた指数算出より詳細な分析が可能です。
 - 2016年4月の消費水準を100とし、各月の消費水準を指数化しております。
- 配信指数
 - <速報性>
 - 締日(15日・末日)から約2週間後(1日・15日)に配信いたします。（土日祝除く）
 - <利便性>
 - 店舗業種と消費者属性の掛け合わせたクロス集計が可能なため、目的に応じた消費指数を把握できます。



図 2-1 「JCB 消費 NOW」の特徴

2-1-2. 指標の算出方法

(1) 指標算出の概要

指標算出に際しては、ナウキャスト社と共同で、現行の「JCB 消費 NOW」の算出ロジックを最大限に活用しつつ、トランザクションデータから中小企業分を抽出し、業種別に中小企業の売上動向を代替的に示すと考えられる、消費側（需要側）を基にした指標を試作した。

トランザクションデータから中小企業分を抽出する際に、本調査では、企業（大企業や中小企業含む）の従業員数と資本金データを含むマスタ³（以下、「中小企業分類マスタ」という。）を活用した。中小企業分類マスタとトランザクションデータを突合させ、トランザクションデータに対して、企業の従業員数と資本金データを付与し、中小企業基本法で定める定義に従った中小企業分類を実施した。さらに、全決済と比較してクレジットカードデータに存在するバイアスに対して、現行の「JCB 消費 NOW」の算出ロジックで活用しているバイアス除去やデータ補正を追加的に行い、指標の精度をより高めた。

³ JCB社で保有している法人企業などのデータベースであり、大企業や中小企業の情報が含まれている。

指標算出ロジックの概要

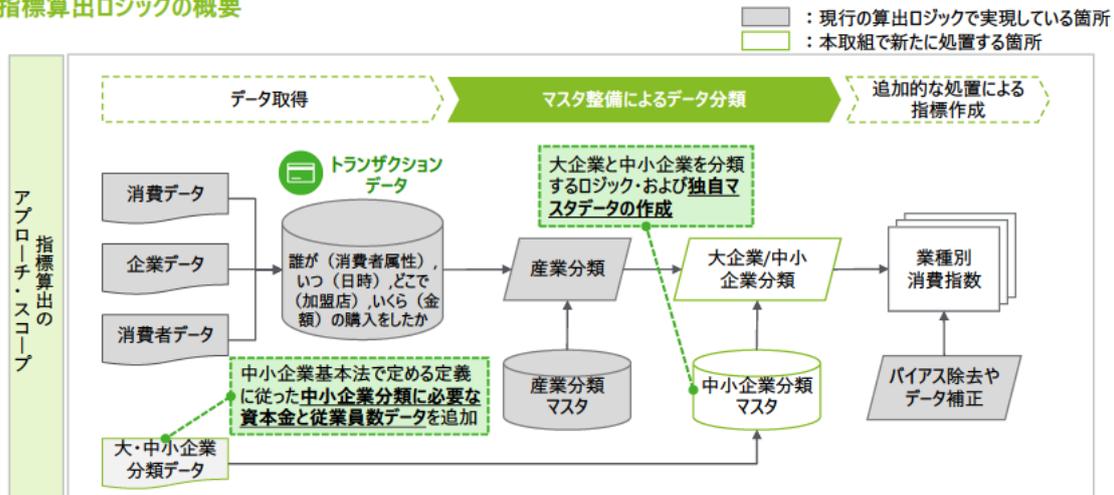


図 2-2 指標算出ロジックの概要

(2) 中小企業分類マスタの特徴

本調査で活用した中小企業分類マスタの特徴を以下に示す⁴。

1) データの網羅性

中小企業分類マスタでは、約 150 万社の上場企業などの与信会社データ、約 450 万社の医療法人や学校法人などの登記法人データを保有している。また、個人事業主や事業所施設などのデータも含めると計 820 万拠点を網羅しており、国内拠点網羅率は 99.7%となっている。

2) データの項目

中小企業分類マスタで保有している主なデータ項目を表 2-1 に示す。

表 2-1 中小企業分類マスタの主なデータ項目

データ項目	備考
契約番号	-
社名	-

⁴ 本内容は、JCB 社からの情報提供を基に記載している。

データ項目	備考
所在地	都道府県名や市区群町村名、字丁目、番地番号、建物名部屋番号
電話番号	-
業種名	-
資本金	資本金の数値ではなく、「1,000万円以上5,000万円未満」や「5,000万以上1億円未満」などの範囲情報
従業員数	従業員数の数値ではなく、「10人以上20人未満」や「100人以上300人未満」などの範囲情報

3) データの整備方法

法人企業等は移転や社名変更等により、その属性情報や財務情報は時点によって異なるため、適宜整備していく必要がある。本調査は試行的な取組であるため、一時点の断面データを活用してデータ分類を実施しているが、中小企業分類マスタがどのようにデータを整備しているかを示す。

中小企業分類マスタのデータは、有価証券報告書や企業ホームページ、登記簿・官報などを基に更新している。各提供元の内容を正誤確認し、データ変更・更新する必要がある場合は、変更内容や提供元の性質により、実際の変更・更新まで数日から1年程度のタイムラグが生じる。主な正誤確認方法とデータ更新までの期間（目安）を表 2-2 に示す。

表 2-2 主な正誤確認方法とデータ更新までの期間（目安）

正誤確認方法	データ更新までの期間（目安）
有価証券報告書の確認	▶ 公開から1か月程度
企業等のホームページ確認	▶ 数営業日
登記簿確認	▶ 1週間程度
官報確認	▶ (Webの場合) 1か月程度、(紙媒体の場合) 1年程度

(3) 中小企業分類マスタを活用したデータ分類

トランザクションデータを中小企業と大企業に分類するため、以下に示す STEP1 から STEP3 までの処理によって、中小企業分類マスタで保有している企業の従業員

数と資本金のデータをトランザクションデータに付与した。なお、トランザクションデータを構成する加盟店マスタには、契約番号や企業名、加盟店名のデータが含まれており、中小企業分類マスタとの突合においては、この加盟店マスタのデータを活用した。

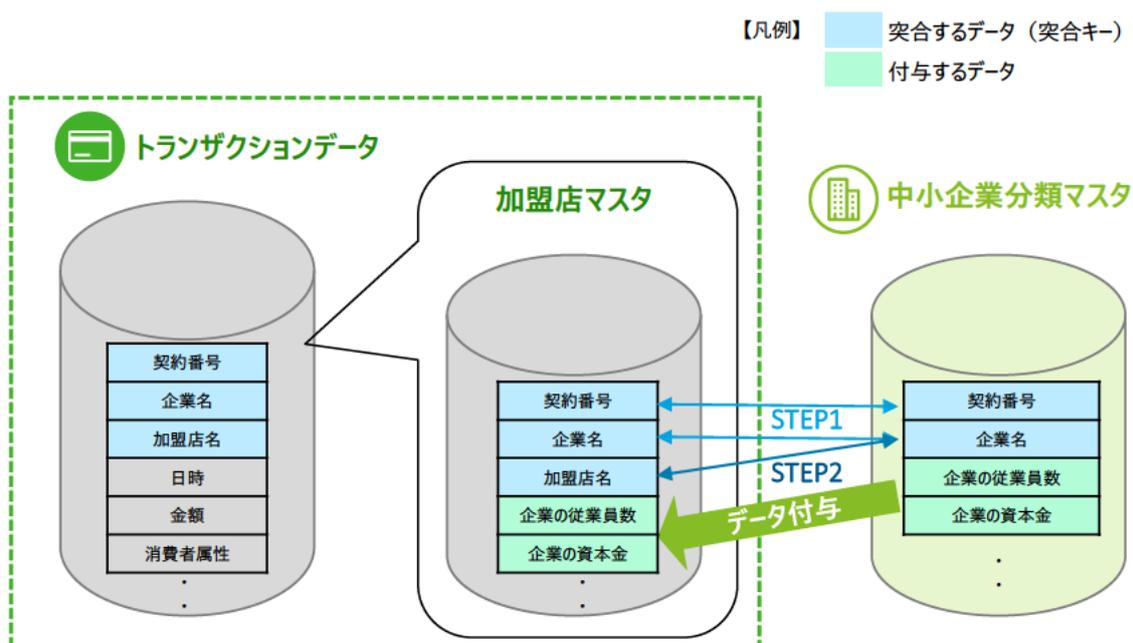


図 2-3 企業の従業員数と資本金データの付与イメージ

1) 契約番号や企業名で突合しデータ分類 (STEP1)

契約番号を突合キーとして、データ付与した。契約番号で突合できなかったものは企業名を突合キーとしてデータ付与した。なお、中小企業分類マスタにおいて、同一企業名で従業員数及び資本金が複数存在する場合は、従業員数及び資本金のそれぞれ最大値のデータを一律で付与した。

2) クレンジングした加盟店名で突合しデータ分類 (STEP2)

加盟店マスタの企業名が「決済企業リスト・プラットフォーム企業リスト」⁵に該当する企業である場合、加盟店名の項目に本来突合すべき企業が記載されている。

⁵ 加盟店マスタの企業名でプラットフォーム運営企業（例：大手 EC 企業、検索サイト）や決済企業（例：〇〇ペイメント）に該当する企業を目検により抽出し、独自作成したリスト。

そこで、中小企業分類マスタが保有する企業情報を最大限に活用するため、加盟店マスタの加盟店名と中小企業分類マスタの企業名を突合した。ただし、加盟店マスタの加盟店の項目には、企業名だけではなく詳細な店舗名まで含まれているケースがあるため、店舗名を除去するクレンジングを実施し、クレンジングした加盟店名を突合キーとしてデータ付与した。

3) 突合不可だったものは全て不明ラベリング (STEP3)

STEP2 でデータ付与できなかったもの (例：中小企業分類マスタに存在しない企業) は、大企業か中小企業か判断ができないため、全て「不明」とラベリングした。

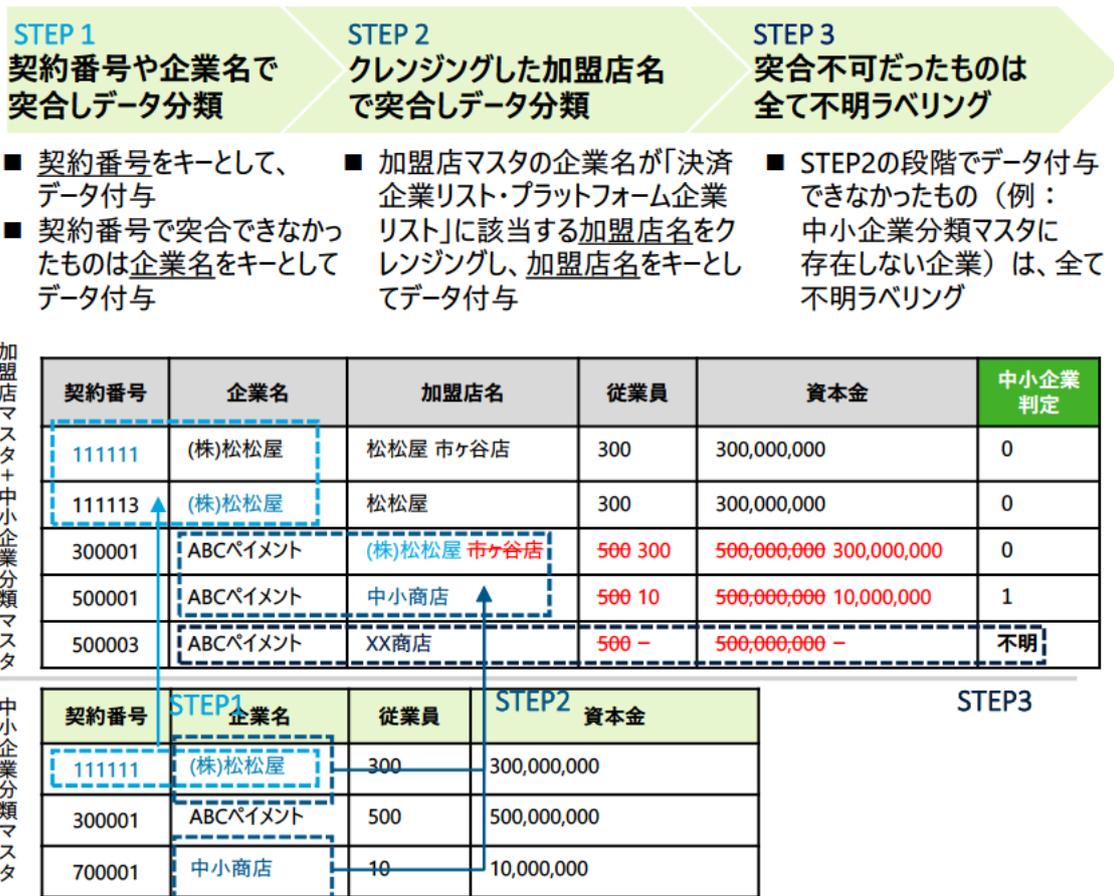


図 2-4 各 STEP におけるデータ分類イメージ

(4) 試作する指標のパターン

中小企業分類マスタを活用したデータ分類を実施したが、データ分類の結果、「大企業」もしくは「中小企業」のラベリングができないデータが見受けられた。中小企業分類マスタに登録されていない企業については、比較的規模の小さい中小企業である可能性が高いと考えられるが、そのようなデータは「不明」とラベリングした。「不明」データの取り扱いについては、今回試作する指標に含めるべきか否かを判断するために、業種別の中小企業指標として「①中小企業のための指標」と「②中小企業及び不明を合算した指標」の2パターンを作成した。なお、企業規模別（大企業/中小企業）の指標を試作することで、指標の評価に活用できることから、業種別の大企業指標として「③大企業のための指標」も作成した。

本調査で試作する指標のパターン

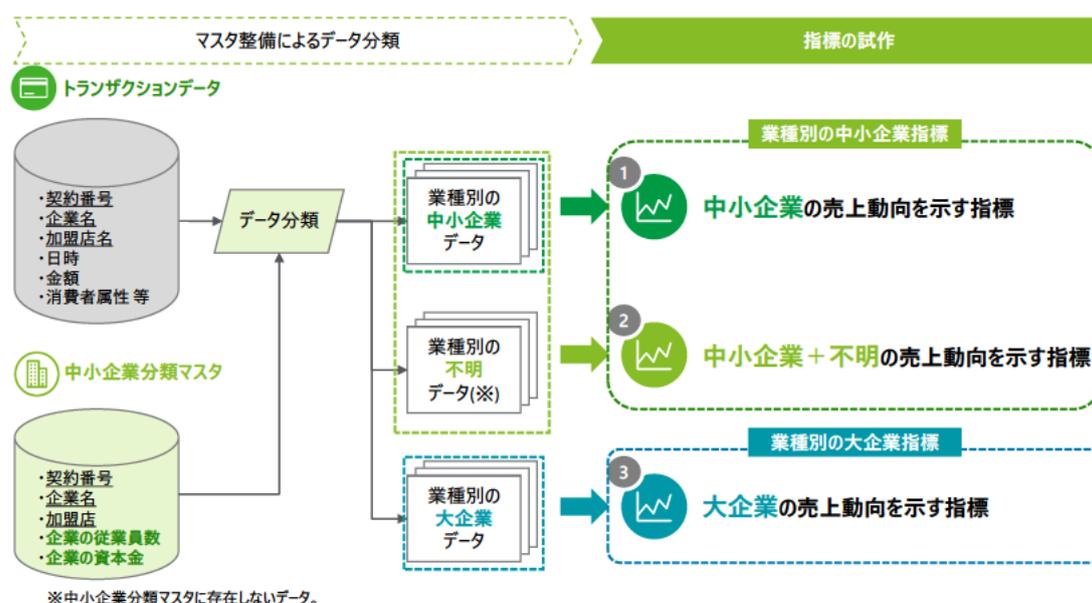


図 2-5 本調査で試作する指標のパターン

2-1-3. 業種選定

本調査では「JCB 消費 NOW」を活用した指標試作を実施するため、「JCB 消費 NOW」の産業分類に基づく業種を試作対象とした。

ナウキャスト社「JCB消費NOW」が提供する国内消費指数の産業分類

マ ク ロ 分 類	各種商品小売業	医薬品 化粧品小売業	医療
	織物 衣服 身の回り品小売業	その他小売業	通信(ISP業、移動通信業)
	飲食品小売業	EC	交通(鉄道旅客、バス、タクシー、航空旅客)
	自動車小売業	電気 ガス 熱供給 水道業	娯楽(映画、スポーツ施設、遊園地)
	機械器具小売業	外食	宿泊(ホテル、旅館)
	燃料小売業	旅行	コンテンツ配信
ミ ク ロ 分 類	百貨店	カー用品	ビジネスホテル
	スーパー	酒屋	ゴルフ場
	コンビニエンスストア	ディスカウントショップ	遊園地
	家具	喫茶店 カフェ	映画館
	医薬品	居酒屋	鉄道旅客
	貴金属	ファミレス	航空旅客
	スポーツ用品	ホテル	デジタルマネー
	ショッピングセンター(実店舗)	子供用品	靴 履物
	化粧品	D I Y	寿司
	タクシー	美容室	焼肉
	カラオケ	高速道路	レンタカー

■ マクロ分類として18業種の分類で消費動向指数を提供

■ ミクロ分類として33業種の分類で消費動向指数を提供

■ EC業種に関しては、EC配下に11業種の分類(EC-各種商品小売業/EC-織物・衣服・身の回り品小売業等)を提供

図 2-6 「JCB 消費 NOW」の産業分類

「JCB 消費 NOW」が対象とするのは全 62 業種であるが、2つの観点から考慮した上で、11 業種に絞り込みを行い、試作する指標の業種を選定した。

業種を絞り込む上で考慮した観点

- 中小企業の実態を踏まえた適切かつ迅速な政策的支援を行う際に、中小企業庁において活用可能性のある業種であるか。
- トランザクションデータを基に指標を試作する上で、現段階では中小企業指数としての精度に大きな懸念がないと想定される業種であるか。

選定した 11 業種

【マクロ分類】織物・衣服・身の回り品小売業、機械器具小売業、外食、旅行、宿泊、EC

【ミクロ分類】居酒屋、焼肉、ゴルフ場、タクシー、酒屋

2-1-4. 指標の評価方法

選定した11業種に対して、中小企業の売上動向を把握するために試作した指標の有効性・妥当性の評価を実施した。

(1) 試作した指標の評価方法

1) 有効性・妥当性の評価

試作した業種別の指標の有効性・妥当性を評価するために、有効性・妥当性の評価軸や評価観点、評価基準を整理した。

有効性評価では速報性と詳細性の観点、妥当性評価では正確性と相関性の観点から評価を実施した。相関性の観点では、既存の統計指標（以下、「ベンチマーク指標」という。）の前提も考慮して整理した。

表 2-3 有効性・妥当性の評価方法

評価軸	評価観点	評価基準
有効性 【定義】 試作する指標が、期待する価値や目的を果たすことができているかどうか	■ 速報性 指標が算出されるまでの期間を確認	▶ 「JCB消費NOW」の通常指標と同様に、毎月15日・末日締めで、2週間程度で算出できる ▶ ベンチマーク指標と比較して、更新頻度の期間短縮が見込める
	■ 詳細性 政府統計との統計粒度の比較	▶ ベンチマーク指標の地域粒度と比較して、より詳細な地域粒度（都道府県別(全47)・地域別(全10)指数）を作成することができる ▶ 既存統計では捕捉できない、より細分化した業種の売上動向把握ができる
妥当性 【定義】 試作する指標が、中小企業に限定した動向把握を適切に測定できているかどうか	■ 正確性 目検による中小企業判定精度を確認	▶ 大企業・中小企業の判定が一定程度、正確であると判断できる（大企業のノイズが少ない）
	■ 相関性 ベンチマーク指標との相関性比較 <u>（ベンチマーク指標の前提）</u>	▶ ベンチマーク指標との相関係数と平均平方二乗誤差が一定水準以上であることを確認できる <u>（ベンチマーク指標の前提）</u>

評価軸	評価観点	評価基準
	試作した指標とベンチマーク指標における前提の違いを確認	▶ 本調査で試作した指標とベンチマーク指標において、算出する前提の違いが大きくないことが確認できる（＝ベンチマーク指標に適している）

2) 各観点の評価における具体的な実施内容

a.速報性

試作した指標が「JCB 消費 NOW」の通常指標と同様に、当月の結果を、翌月 15 日（当月前半の結果は、翌月初日）に、算出できるか確認した。また、ベンチマーク指標の速報性と比較し、どの程度の期間短縮が見込めるか確認した。

b.詳細性

都道府県別・地域別で指数を作成することができる都道府県数(全 47)と地域数(全 10)を確認した。対象となる都道府県と地域については、総務省の地域別表章に関するガイドラインで推奨されている類型 I をベースとし、沖縄県のみ九州に含める形とした総務省統計局の地域区分に従っている⁶。沖縄県を九州に含めるのはサンプル数の観点から単体で切り出すよりも地域にまとめることで地域別の指数としての精度向上が期待できるためである。

都道府県別・地域別で指数を作成可能と判断する際には、指数を構成するサンプル数が十分であるか 2 段階に分けて確認した。1 段階目では、消費者の属性（年齢、性別、居住地）ごとに属性セル（約 1,400 区分＝【年齢：15 区分】×【性別：2 区分】×【都道府県：47 区分】）を作成し、その属性セルにおいて消費者が 5 人以上確保できているか確認した。2 段階目では、30 区分（【年齢：15 区分】×【性別：2 区分】）の属性セルを積み上げて都道府県指数を作成する際に、この 30 区分のうち 20 個以上の属性セルが確認できれば、指数作成可能と評価した。地域別指数の評価方法においても、都道府県別指数と同様のアプローチで評価を実施した。

⁶ 総務省統計局（<https://www.stat.go.jp/data/shugyou/1997/3-1.html>）を参照。

① 消費者(※)が5人以上いるセルを残す

	北海道	青森県	岩手県	・・・	沖縄県
20-24歳男性	2	8	10		43
20-24歳女性	10	42	4		133
・					
80-85歳女性	3	9	31		23

② 都道府県ごとに残存セルを計算

	北海道	青森県	岩手県	・・・	沖縄県
20-24歳男性	2	8	10		43
20-24歳女性	10	42	4		133
・ 12セル残存					
80-85歳女性	3	9	31		23
残存セル数	12	24	17		29

47都道府県のうちx個の都道府県が20セル以上残存セルがある
(こちらのxが運搬したもの)

※ここでは、ユニークユーザーを指す。ユニークユーザーとは、実際にJCBクレジットカードで決済したユーザーであるが、当該ユーザーが複数回決済しても、重複してカウントしない。

図 2-7 詳細性評価方法のイメージ

また、これまで既存の統計では捕捉できなかった業種については、より細分化した業種の売上動向把握が可能であるか確認した。

c.正確性

「JCB 消費 NOW」のトランザクションデータから中小企業分を抽出した過程において、「中小企業」とラベリングした企業に大企業が含まれていないか検証した。

本調査では、中小企業分類マスタを活用し、3段階に分けて「中小企業」及び「不明」とラベリングしたため、各3段階におけるラベリングの正確性を把握することを目的として、それぞれの段階の売上高上位企業を20件ずつランダム抽出し、当該企業のWeb検索を行い中小企業か否かを確認するサンプリングチェックを実施した。その上で、サンプリングチェック結果を各段階の売上寄与度で重みづけした正答率を算出した。なお、各段階の売上寄与度で重みづけし、各段階において売上寄与度が大きい企業を残した結果、当該段階の企業数が5つ以下のものについては、少数の企業による評価は信頼性が高くなく、正答率の算出対象からは除外することとした。

正答率を評価するに当たっては、算出される正答率は過小評価の可能性のあることに留意した。本調査では、完全無作為ではなく売上高上位企業を抽出しており、中小企業の中でも売上高が大きい企業が評価対象となっているためである。

$$\begin{aligned}
 \text{正答率} = & (1 \text{ 段階目で突合した中小企業の正答率} \times 1 \text{ 段階目の売上寄与度}) \\
 & + \\
 & (2 \text{ 段階目で突合した中小企業の正答率} \times 2 \text{ 段階目の売上寄与度}) \\
 & + \\
 & (3 \text{ 段階目 (2 段階目で突合できず「不明」とラベリング) の中小企業の割合} \times 3 \text{ 段階目の売上寄与度})
 \end{aligned}$$

図 2-8 正答率の算出方法

d. 相関性

試作した指標とベンチマーク指標との相関性を確認するために、相関係数 (corr) と平均平方二乗誤差 (RMSE) を評価指標として、比較対象となるベンチマーク指標の更新頻度や業種粒度に合わせる形で算出した。検証期間は、「JCB 消費 NOW」のデータの提供開始時期である 2016 年 4 月から、各ベンチマーク指標における最新公表データ時点までとした。また、比較する際には季節調整を考慮する必要がないように、今回試作した指標の値とベンチマーク指標の値をそれぞれ「前年比」に処理した。

また、本調査で試作する指標とベンチマーク指標の前提が一致、類似しているか比較する上で 3 つの観点を考慮し、今回試作した指標のベンチマーク指標としてどの程度適しているか確認した。

ベンチマーク指標との相関性比較において考慮した観点

- 企業サイズ：中小企業のサイズに違いはないか。
- 業種粒度：業種粒度が揃っているか。
- 売上高指数：売上高そのものを示す指数か。

加えて、ベンチマーク指標との相関性を確認する際には、「ベンチマーク指標との符号比較」と「ベンチマーク指標との原数値（原指数）の比較」も併せて実施した。

ベンチマーク指標との符号比較

相関係数は一定期間における一意の値で評価する指標であるが、相関係数では確認することができない「前年比」で好転、悪化の動きがどの程度一致しているかを

評価することを目的として、試作した指標とベンチマーク指標の「前年比」の値に対して、符号（プラスかマイナスか）を確認した。その上で、ベンチマーク指標の全データポイント⁷における値の符号と、試作した指標の値の符号がどの程度一致していたか符号一致率（割合）を算出した。

ベンチマーク指標との原数値（原指数）の比較

「前年比」に処理したベンチマーク指標との相関性比較だけではなく、売上高の絶対値である「原数値（原指数）」においても相関が取れているか確認することを目的として、「原数値（原指数）」が公開されているベンチマーク指標について、試作した指標とベンチマーク指標の相関係数を算出した。

3) 指標の定性評価

有効性・妥当性評価におけるそれぞれの観点は主に定量的な評価であり、必ずしも相関係数などの数値だけでは評価しきれない部分が存在する。試作した業種ごとの指標にどのような違いがあるのか、それぞれの特徴を見出すために、5つの観点において定性的な評価を実施した。

表 2-4 定性評価の評価観点と評価方法

評価観点	評価方法
キャッシュレス還元政策の影響を受けていないか (2019年終盤から2020年初頭)	▶ 消費増税の前後期間において、加盟店がキャッシュレス還元政策の恩恵を受けるため、クレジットカード決済を導入したと考えられるバイアスが存在していないか確認した
消費増税に入るポイントを捉えているか (2019年10月)	▶ ベンチマーク指標の推移も参考に、消費増税前後の指標の動きについて、急変動が起きているか確認した。なお、消費増税の影響をベンチマーク指標では確認できない業種については、そもそも消費増税の影響を受けていない業種と考えられるため、評価を実施していない

⁷ ベンチマーク指標の更新頻度が高いほど、相関係数を算出する際のデータポイントが多くなる。一方で、更新頻度が低い（例：年次）ベンチマーク指標との比較の場合、試作した指標は2016年4月からのデータであるため、データポイントは少なくなる。

評価観点	評価方法
感染症に入るポイントを捉えているか (2020年4月)	▶ ベンチマーク指標の推移も参考に、感染症前後の指標の動きについて、急変動が起きているか確認した
店舗増加などの影響を受けていないか	▶ ベンチマーク指標の動きと比較して、キャッシュレス化の進展などのアッパーバイアス（過大評価） ⁸ が定常的に存在していないか確認した
ベンチマーク指標と大きく違う動きをしているポイントがないか	▶ ベンチマーク指標の動きと不自然な差分が生じていないか確認した

（２）評価に使用したベンチマーク指標

本調査では、一般的に信頼性が高い政府統計を優先的にベンチマーク指標に採用し、業種ごとに試作した指標との比較によって、有効性・妥当性評価を実施した。

表 2-5 業種ごとのベンチマーク指標

業種	ベンチマーク指標（ベンチマーク指標における業種）
織物・衣服・身の回り品小売業	<ul style="list-style-type: none"> ▶ 中小企業庁「中小企業実態基本調査」（織物・衣服・身の回り品小売業） ▶ 中小企業庁「中小企業景況調査」（小売業（織物・衣服・身の回り品）） ▶ 経済産業省「商業動態統計」（織物・衣服・身の回り品小売業）
機械器具小売業	<ul style="list-style-type: none"> ▶ 中小企業庁「中小企業実態基本調査」（機械器具小売業） ▶ 中小企業庁「中小企業景況調査」（小売業（機械器具）） ▶ 経済産業省「商業動態統計」（機械器具小売業）
外食	<ul style="list-style-type: none"> ▶ 中小企業庁「中小企業実態基本調査」（飲食店） ▶ 中小企業庁「中小企業景況調査」（飲食店、持ち帰り、配達飲食） ▶ 財務省「法人企業統計調査」（飲食サービス業） ▶ 日本銀行「消費活動指数」（外食） ▶ 日本銀行「全国企業短期経済観測調査（以下「短観」という。）」（宿泊・外食サービス）

⁸ 現金決済からクレジットカード決済へと移行する消費者の増加や、クレジットカード決済に対応できる店舗数の増加が指数の結果に影響を与えている可能性のことを指す。

業種	ベンチマーク指標（ベンチマーク指標における業種）
旅行	▶ 日本銀行「消費活動指数」（外食）
宿泊	▶ 中小企業庁「中小企業実態基本調査」（宿泊業） ▶ 中小企業庁「中小企業景況調査」（宿泊業） ▶ 財務省「法人企業統計調査」（宿泊業） ▶ 日本銀行「消費活動指数」（宿泊） ▶ 日本銀行「短観」（宿泊・外食サービス）
EC	▶ 中小企業庁「中小企業実態基本調査」（無店舗小売業） ▶ 経済産業省「商業動態統計」（無店舗小売業）
居酒屋	▶ 日本フードサービス協会「外食産業市場動向調査」 ⁹ （居酒屋）
焼肉	▶ 日本フードサービス協会「外食産業市場動向調査」（焼き肉）
ゴルフ場	▶ 経済産業省「特定サービス産業動態統計調査」（ゴルフ場）
タクシー	▶ 日本銀行「消費活動指数」（タクシー）

2-2. 有識者研究会の設置・運営

2-2-1. 研究会の設置概要

試作した指標の有効性・妥当性評価を効果的に進めるために、経済学や統計学に知見のある有識者計4名を委員として選定し、「中小企業の売上動向把握のための調査・分析に関する意見交換会」を設置した。

本研究会は、第1回（令和4年11月）では取組の概要説明と指標試作に向けた調査設計に関する議論、第2回（令和4年12月）では試作した指標の有効性・妥当性評価に関する議論と次年度以降の取組に向けた議論を行い、全2回の日程で開催した。

<第1回>

日時：令和4年11月7日（月）10：00～12：00

開催方法：Web会議（Teams）

⁹ 民間の統計調査ではあるものの、日本銀行「消費活動指数」の基礎統計指標の1つであり、比較的信頼性が高いとされているベンチマーク指標だと考えられる。

<第2回>

日時：令和4年12月26日（月）13：00～15：00

開催方法：Web会議（Teams）

2-2-2. 各回研究会の議題及び要旨

各回の研究会における議題、討議結果は以下のとおり。

<第1回>

議題：

- (1) 今年度の取組方向性
- (2) 指標試作に向けた調査設計案

表 2-6 第1回研究会における委員からの主なご意見

議題	委員からの主なご意見
(1) 今年度の取組方向性	<ul style="list-style-type: none">▶ 民間企業のデータを基にした公的利用に際しては、指標の算出ロジックの透明性を担保し信頼性を高めることが重要である▶ 試作する指標を政府統計として活用することを検討する際には、現行の「JCB消費NOW」の算出ロジックにおけるバイアス除去やデータ補正についても、可能な限り開示した方が望ましい▶ 中小企業庁における活用方法を具体的に検討する必要がある
計案 (2) 指標試作に向けた調査設計案	<p><u>指標算出ロジック</u></p> <ul style="list-style-type: none">▶ 中小企業判定が困難データ（中小企業分類マスタと突合不可能なデータ）については、中小企業とみなさずに「不明」とラベリングして整理し、より信頼性の高いデータ（中小企業分類マスタと突合可能なデータ）を基に中小企業指標を作成した上で、評価を実施することが望ましい▶ 中小企業指標だけではなく、大企業指標も作成することで、大企業と中小企業の売上動向の差異を確認することが期待できる

議題	委員からの主なご意見
	<ul style="list-style-type: none"> ▶ データに含まれる中小企業の割合が、実態と乖離していないかを確認することも、指標の妥当性を評価する上で必要と考えられる。実態との比較に際しては、中小企業白書のデータを基に比較することが望ましい ▶ (加盟店情報を取得できない出品形式を採っている) 外資大手ECデータについては、大企業として取り扱うのは避けた方が良い。中小企業として扱う、大企業として扱う、除外する、等のパターン別に示して整理することが望ましい ▶ 属性別偏りの精緻化において、年齢・地域・性別のセルごとに算出する場合、各セルのボラティリティは増す可能性があるため、モデルを作成して検討した方が良い <p><u>指標の評価方法</u></p> <ul style="list-style-type: none"> ▶ 大企業指標も試作し、企業規模別(大企業/中小企業)に売上動向を算出できるベンチマーク指標と比較して、算出ロジックの妥当性を確認した方が望ましい ▶ 粒度の細かい業種における精度を担保するために、細かい業種だけではなく、ある程度まとまった粗い業種で比較することも一案である ▶ ユースケース(例:感染症等による急速な外部環境変化を把握できるか)を踏まえて、適切に評価することが重要である ▶ 市場全体におけるJCBクレジットカードの売上割合等を確認し、JCBカードの特性を考慮する必要がある

<第2回>

議題:

- (1) 試作した指標に関する有効性・妥当性の評価の報告
- (2) 次年度以降に向けた検討

表 2-7 第2回研究会における委員からの主なご意見

議題	委員からの主なご意見
<p>(1) 試作した指標に関する有効性・妥当性の評価の報告</p>	<p><u>有効性の評価</u></p> <ul style="list-style-type: none"> ▶ 詳細性の観点においては単に指数を作成する上で、政策部局としてその指標が有用であるかを考慮することが必要である ▶ 速報性や詳細性を満たさない場合でも、既存の統計指標を確認することができるオルタナティブデータとして価値があると考えられる <p><u>妥当性の評価</u></p> <ul style="list-style-type: none"> ▶ ベンチマーク指標との相関性比較の際には「前年比」だけではなく「原数値」で比較することも重要である ▶ 季節調整などの細かいパラメータ処理を実施することで、より相関が高まる可能性がある ▶ ベンチマーク指標との比較において、相関性比較だけではなくターニングポイント（例：感染症の期間中、売上動向が減少傾向から増加傾向に変わる時点）を適切にトラッキングできているかどうかが重要である
<p>(2) 次年度以降に向けた検討</p>	<p><u>指標精度の向上</u></p> <ul style="list-style-type: none"> ▶ 中小企業分類マスタの課題を整理することが望ましい ▶ 指標算出ロジックの更なるバイアス除去やデータ補正を検討していくことが必要である ▶ 公的統計との比較評価だけではなく、各業界（関連団体含む）の方々にて定性評価していただくことも必要である <p><u>指標活用の具体化</u></p> <ul style="list-style-type: none"> ▶ 指標の価値を顕在化するために、可能な限り多くの潜在的なエンドユーザーにヒアリングすることが重要である ▶ 国内クレジットカードデータだけではなくその他オルタナティブデータの活用を検討していくことが必要である ▶ 中小企業は異質性が高いため、中小企業における売上動向の分散を評価することも重要である ▶ EBPMの観点から、本指標の活用により中小企業政策の現状把握が期待できるため、政策の効果検証にも活用可能性がある ▶ 本取組を拡充していくために、今回の取組成果を連携するなど国内の研究機関等と共同研究の機会を探索すべきである ▶ 指標が政策的議論で活用されるように、中小企業庁内に指標を実際に活用できる人員を配置することが重要である

2-2-3. 研究会を踏まえた対応

有識者からの意見を踏まえ、指標の算出や評価、活用に関する検討を実施した。本研究会を踏まえた主な対応内容を表 2-8 に示す。

表 2-8 研究会を踏まえた主な対応内容

議題	主な対応
指標の算出	<ul style="list-style-type: none"> ▶ 大企業/中小企業の判定ができない（中小企業分類マスタに存在しない）データについては、「不明」とラベリングし、「中小企業」のみの指標と「中小企業+不明」の指標の2パターンを試作 ▶ 中小企業白書に掲載している、平成28年経済センサスにおける業種別の中小企業割合との比較を行い、実態との整合性を確認 ▶ EC業種においては「外資大手ECデータを中小企業として扱う」と「外資大手ECデータを大企業として扱う」の2パターンを試作
指標の評価	<ul style="list-style-type: none"> ▶ 企業規模別（大企業/中小企業）の指標を試作し、大企業動向を示すベンチマーク指標との相関性比較を実施 ▶ 相関性比較において、「ベンチマーク指標との符号比較」と「ベンチマーク指標との原数値（原指数）の比較」を実施 ▶ コロナ禍の期間に絞ったベンチマーク指標との比較を実施 ▶ 中小企業分類マスタの課題整理を実施
指標の活用	<ul style="list-style-type: none"> ▶ 中小企業庁における活用方法の検討 ▶ 指標のエンドユーザーとなり得る地方自治体や調査機関に対して、指標の活用可能性やニーズに関してヒアリング調査を実施

2-3. 調査結果

前項で整理した指標の有効性・妥当性評価方法や研究会の内容に基づき、選定した11業種の評価を実施した。また、評価の結果を踏まえて、指標活用にあたっての課題について整理した。

2-3-1. 指標の評価

(1) 評価概要

選定した11業種の有効性・妥当性評価の概要を図2-9に示す。

有効性評価のうち速報性の観点においては、ベンチマーク指標が存在する業種については2～3週間程度の期間短縮が見込めることを確認できた。詳細性の観点においては、一部の都道府県と地域を除き、サンプル数が十分に見込めることから都道府県別・地域別指数を作成可能であることを確認できた。

妥当性評価のうち正確性の観点においては、一部の業種を除き、高い正答率であることを確認できた。相関性の観点においては、ベンチマーク指標に対する懸念事項はあるものの、ベンチマーク指標が存在する業種については少なくとも1つ以上のベンチマーク指標と強い正の相関がみられることを確認できた。

一方で、中小企業割合が実態と乖離している可能性や追加的なバイアス除去・データ補正の必要性なども確認した。課題や懸念事項の詳細内容については、「2-3-2 評価結果を踏まえた指標の課題」を参照されたい。なお、詳細性に関しては、今回実施した範囲は都道府県別・地域別で十分なサンプルサイズを確保できるかという視点が中心であり、都道府県別・地域別の指数を作成した場合、その指数の妥当性については別途十分に確認する必要がある。

※1「中小企業のみ指標」における売上寄与度で重みづけした正答率

※2 強い正の相関($\rho \geq 0.7$)がみられるベンチマーク指標数

評価概要

候補業種	有効性		妥当性		
	速報性	詳細性	正確性 ^(※1)	相関性 ^(※2)	ベンチマーク指標に対する懸念事項
織物・衣服・身の回り品小売業	ベンチマーク指標と比較して、2週間程度の期間短縮	都道府県別指数:47 地域別指数:10	90.7%	2指標	■ 大企業を含むベンチマーク指標との相関、売上DIのベンチマーク指標との相関である
機械器具小売業		都道府県別指数:47 地域別指数:10	94.6%	1指標	
外食	ベンチマーク指標と比較して、3週間程度の期間短縮	都道府県別指数:47 地域別指数:9	79.0%	3指標	■ 大企業を含むベンチマーク指標との相関、業種粒度が異なるベンチマーク指標との相関、売上DIのベンチマーク指標との相関である
旅行		都道府県別指数:37 地域別指数:5	95.0%	1指標	
宿泊		企業の所在地と消費者の居住地が異なる可能性があるため、評価対象外	98.6%	4指標	
EC	ベンチマーク指標と比較して、2週間程度の期間短縮		71.5%	1指標	■ 大企業を含むベンチマーク指標との相関である
居酒屋		都道府県別指数:31 地域別指数:10	100.0%	1指標	
焼肉		都道府県別指数:34 地域別指数:8	94.1%	1指標	
ゴルフ場		都道府県別指数:35 地域別指数:8	75.0%	1指標	
タクシー	ベンチマーク指標と比較して、3週間程度の期間短縮	都道府県別指数:21 地域別指数:7	88.9%	1指標	■ 大企業を含むベンチマーク指標との相関である

図2-9 有効性・妥当性の評価概要

(2) 業種ごとの評価

1) 織物・衣服・身の回り品小売業の評価

a. 指数の推移

研究会の内容を踏まえ、「織物・衣服・身の回り品小売業」においては「中小企業指数」と「中小企業指数(不明含む)¹⁰」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-10 に示す。

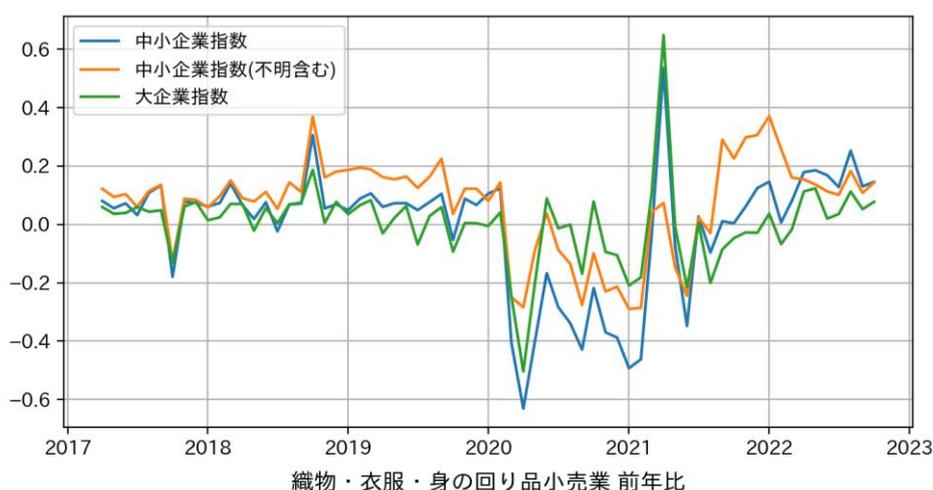


図 2-10 織物・衣服・身の回り品小売業の指数の推移

b. 有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して2週間程度の期間短縮が見込める速報性があり、かつ全ての都道府県別・地域別指数を作成できる詳細性がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、2つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、活用可能性が高い指標と考えられる。

¹⁰ 中小企業分類マスタを活用したデータ分類を実施したが、データ分類の結果、「大企業」もしくは「中小企業」のラベリングができない「不明」データが見受けられたため、この「不明」データを試作する中小企業指標に含めるべきか否かの判断するために試作した指数。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、正確性と相関性の評価結果から、「中小企業のみ指標」と「中小企業及び不明を合算した指標」の2パターンでは、「中小企業のみ指標」が望ましいといえる。

一方で、ベンチマーク指標との比較において、キャッシュレス化の進展などのアップバイアス（過大評価）¹¹といった課題・懸念事項を確認できた。

表 2-9 織物・衣服・身の回り品小売業の有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	<ul style="list-style-type: none"> ▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「商業動態統計(当月の結果を、翌月末に公表)」と比べても、<u>2週間程度の期間短縮が見込める</u>
	■ 詳細性	<ul style="list-style-type: none"> ▶ 消費者属性（性別、年齢区分など）ごとに十分なサンプル数があるため、<u>全ての都道府県別・地域別で指数が作成可能</u>である 都道府県別指数を作成することができる都道府県数(全47)：47 地域別指数を作成することができる地域数(全10)：10 ▶ なお、本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別（地域別）で中小企業の売上動向を把握する際には、旅行等に伴う織物・衣服・身の回り品消費等のノイズを多少含んでいる可能性がある
妥当性	■ 正確性	<ul style="list-style-type: none"> ▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、<u>1段階目と2段階目の精度は高い</u>が3段階目の精度が低い（※サンプリングチェックの対象は売上上位企業） <ul style="list-style-type: none"> ✓ 1段階目の中小企業の正答率：89.5% ✓ 2段階目の中小企業の正答率：93.8% ✓ 3段階目の中小企業割合：52.6% ▶ 売上寄与度で重みづけした正答率は下記のとおりであり、<u>「中小企業のみ指標」の精度が高い</u> <ul style="list-style-type: none"> ✓ 中小企業のみ指標：<u>90.7%</u> ✓ 中小企業及び不明を合算した指標：69.4%

¹¹ 現金決済からクレジットカード決済へと移行する消費者の増加や、クレジットカード決済に対応できる店舗数の増加が指数の結果に影響を与えている可能性のことを指す。このバイアスが存在することにより、特定の業種の売上動向が本来、「前年比」ベースで減少（マイナス）傾向であるのに対し、本指標における指数の動きでは増加（プラス）傾向となってしまう懸念がある。

評価軸	評価観点	評価結果																																																																																																		
	■ 相関性	<p>➤ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。「商業動態統計」及び「中小企業景況調査」と強い正の相関がみられた</p> <table border="1"> <thead> <tr> <th rowspan="2">0内は更新頻度</th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td>中小企業実態基本調査(年次)</td> <td>corr</td> <td>0.24</td> <td>0.07</td> <td></td> <td>◎</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.38</td> <td>0.36</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>商業動態統計(月次)</td> <td>corr</td> <td>0.84</td> <td>0.56</td> <td>0.86</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.12</td> <td>0.17</td> <td>0.08</td> <td></td> <td></td> <td></td> </tr> <tr> <td>中小企業景況調査(四半期)</td> <td>corr</td> <td>0.83</td> <td>0.55</td> <td></td> <td>◎</td> <td>◎</td> <td>○</td> </tr> <tr> <td></td> <td>RMSE(※)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>(※)売上DIは前年比で「好転」から「悪化」を引いた数値であり、RMSEで比較を行うことは不適切であると考えられるため算出しない</p> <p>【corrにおける凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 ■ 誤差が小さい (<0.3) ■ 誤差がやや大きい (>=0.3) ■ 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>➤ 研究会の内容を踏まえ、「前年比」のベンチマーク指標との符号一致率¹²と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <p>✓ 「前年比」のベンチマーク指標との符号一致率</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>商業動態統計</td> <td>65</td> <td>69%</td> <td>66%</td> <td>72%</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>4</td> <td>25%</td> <td>25%</td> <td>-</td> </tr> <tr> <td>中小企業景況調査</td> <td>22</td> <td>32%</td> <td>27%</td> <td>-</td> </tr> </tbody> </table> <p>【凡例】 ■ 符号一致率が25%以下 ■ 符号一致率が50%以下</p> <p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>商業動態統計</td> <td>0.73</td> <td>0.24</td> <td>0.78</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>0.22</td> <td>-0.07</td> <td>-</td> </tr> <tr> <td>中小企業景況調査</td> <td>-</td> <td>-</td> <td>-</td> </tr> </tbody> </table> <p>【凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>➤ 上記の結果より、「前年比」及び「原数値（原指数）」と強い正の相関がみられ、かつおおむね符号一致率が高いことから「商業動態統計」との相関性があると考えられる。一方で、「中小企業景況調査」とは「前年比」のベン</p>	0内は更新頻度		相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	中小企業実態基本調査(年次)	corr	0.24	0.07		◎	◎	◎		RMSE	0.38	0.36					商業動態統計(月次)	corr	0.84	0.56	0.86	△	◎	◎		RMSE	0.12	0.17	0.08				中小企業景況調査(四半期)	corr	0.83	0.55		◎	◎	○		RMSE(※)							ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	商業動態統計	65	69%	66%	72%	中小企業実態基本調査	4	25%	25%	-	中小企業景況調査	22	32%	27%	-	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	商業動態統計	0.73	0.24	0.78	中小企業実態基本調査	0.22	-0.07	-	中小企業景況調査	-	-	-
0内は更新頻度		相関性			ベンチマーク指標の前提																																																																																															
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																																																																													
中小企業実態基本調査(年次)	corr	0.24	0.07		◎	◎	◎																																																																																													
	RMSE	0.38	0.36																																																																																																	
商業動態統計(月次)	corr	0.84	0.56	0.86	△	◎	◎																																																																																													
	RMSE	0.12	0.17	0.08																																																																																																
中小企業景況調査(四半期)	corr	0.83	0.55		◎	◎	○																																																																																													
	RMSE(※)																																																																																																			
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																																																																																
商業動態統計	65	69%	66%	72%																																																																																																
中小企業実態基本調査	4	25%	25%	-																																																																																																
中小企業景況調査	22	32%	27%	-																																																																																																
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																																																																																																	
商業動態統計	0.73	0.24	0.78																																																																																																	
中小企業実態基本調査	0.22	-0.07	-																																																																																																	
中小企業景況調査	-	-	-																																																																																																	

¹² ベンチマーク指標の全データポイントにおける値の符号と、試作した指標の値の符号がどの程度一致していたかその割合を算出することで、「前年比」で好転、悪化の動きがどの程度一致しているかを評価することができる。

評価軸	評価観点	評価結果
		チマーク指標と強い正の相関がみられるが ¹³ 、アッパーバイアス等の要因から符号一致率が低いことを確認できた

c.定性評価

研究会の内容を踏まえ、本業種の中小企業指標に対する定性評価を実施し、その結果を表 2-10 に示す。いくつかの評価観点で懸念があるものの、感染症に入るポイントは適切に捉えていることを確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「商業動態統計」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-11 に示す。

表 2-10 織物・衣服・身の回り品小売業の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	➤ わずかに影響を受けていることを確認でき、少し懸念がある
消費増税に入るポイント (2019年10月)	➤ -
感染症に入るポイント (2020年4月)	➤ 適切に捉えていることを確認できた
店舗増加などの影響	➤ 影響を受けていることを確認でき、懸念がある
ベンチマーク指標と大きく違う動きをしているポイント	➤ 2017年10月頃に確認でき、懸念がある

¹³ 「中小企業景況調査」で公表されている売上DIは、(中小企業へのアンケート調査において)前年比で「好転」から「悪化」を引いた数値であるため、「原数値(原指数)」の相関係数は算出していない。

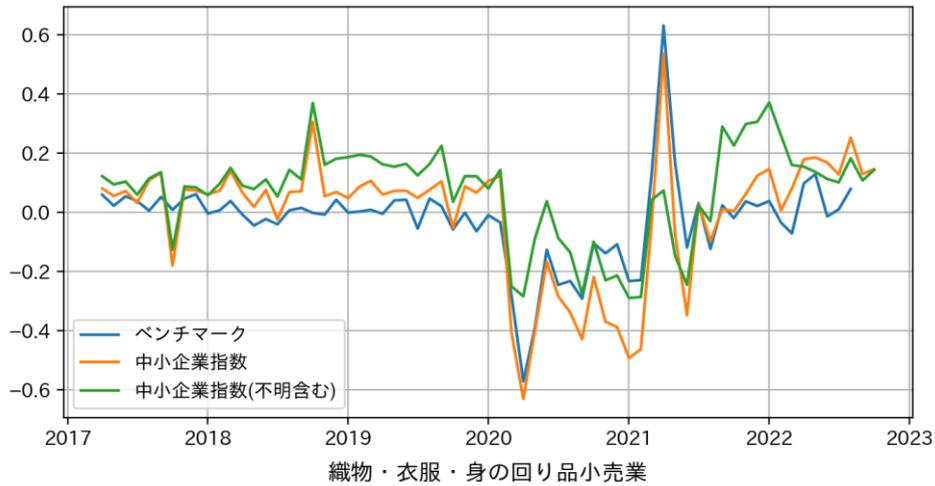


図 2-11 織物・衣服・身の回り小売業の中小企業指数とベンチマーク指標（「商業動態統計」）との比較結果

2) 機械器具小売業の評価

a. 指数の推移

「機械器具小売業」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-12 に示す。

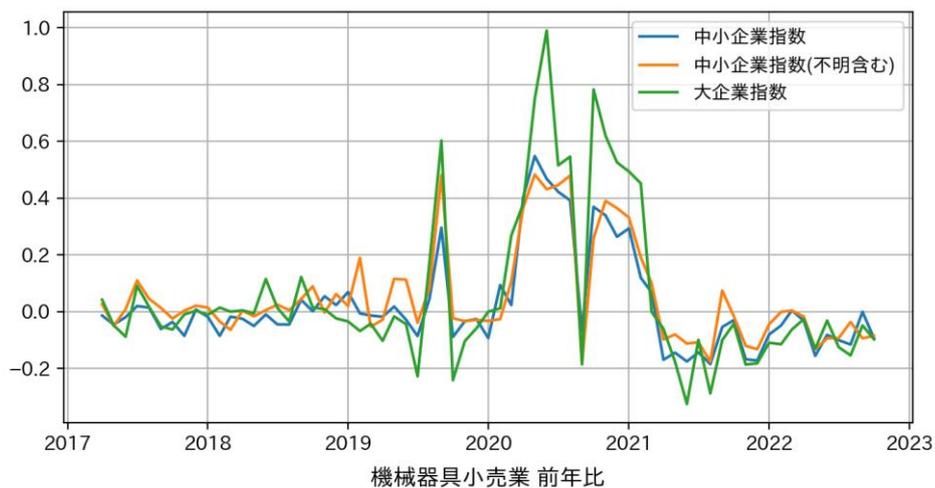


図 2-12 機械器具小売業の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して2週間程度の期間短縮が見込める速報性があり、かつ全ての都道府県別・地域別指数を作成できる詳細性がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、1つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、活用可能性が高い指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、正確性と相関性の評価結果から、「中小企業のみ指標」と「中小企業及び不明を合算した指標」の2パターンでは、「中小企業のみ指標」が望ましいといえる。

一方で、「中小企業実態基本調査」との相関性比較において、強い負の相関がみられたことや、本業種における中小企業割合が実態と乖離している可能性があること（「2-3-2.(1) 中小企業分類マスタの精度」参照）といった課題・懸念事項を確認できた。

表 2-11 機械器具小売業の有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	<ul style="list-style-type: none"> ▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「商業動態統計(当月の結果を、翌月末に公表)」と比べても、<u>2週間程度の期間短縮が見込める</u>
	■ 詳細性	<ul style="list-style-type: none"> ▶ 消費者属性(性別、年齢区分など)ごとに十分なサンプル数があるため、<u>全ての都道府県別・地域別で指数が作成可能</u>である 都道府県別指数を作成することができる都道府県数(全47): 47 地域別指数を作成することができる地域数(全10): 10 ▶ なお、本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別(地域別)で中小企業の売上動向を把握する際には、旅行等に伴う機械器具消費等のノイズを多少含んでいる可能性がある
妥当性	■ 正確性	<ul style="list-style-type: none"> ▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、<u>1段階目と2段階目の精度は高い</u>が3段階目の精度が低い(※サンプリングチェックの対象は売上上位企業) ✓ 1段階目の中小企業の正答率: 94.1% ✓ 2段階目の中小企業の正答率: 100.0%

評価軸	評価観点	評価結果																																																																																																		
		<p>✓ 3段階目の中小企業割合：53.3%</p> <p>➤ 売上寄与度で重みづけした正答率は下記のとおりであり、「<u>中小企業のみ</u>の指標」の精度が高い</p> <p>✓ 中小企業のみ指標：94.6%</p> <p>✓ 中小企業及び不明を合算した指標：61.1%</p>																																																																																																		
	■ 相関性	<p>➤ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。「<u>商業動態統計</u>」と強い正の相関がみられた</p> <table border="1"> <thead> <tr> <th rowspan="2">()内は更新頻度</th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td>中小企業実態基本調査(年次)</td> <td>corr</td> <td>-0.72</td> <td>-0.76</td> <td></td> <td>◎</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.23</td> <td>0.23</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>商業動態統計(月次)</td> <td>corr</td> <td>0.82</td> <td>0.81</td> <td>0.91</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.2</td> <td>0.19</td> <td>0.14</td> <td></td> <td></td> <td></td> </tr> <tr> <td>中小企業景況調査(四半期)</td> <td>corr</td> <td>-0.48</td> <td>-0.39</td> <td></td> <td>◎</td> <td>◎</td> <td>○</td> </tr> <tr> <td></td> <td>RMSE(※)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>(※)売上DIは前年比で「好転」から「悪化」を引いた数値であり、RMSEで比較を行うことは不適切であると考えられるため算出しない</p> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>➤ 「前年比」のベンチマーク指標との符号一致率と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <p>✓ 「前年比」のベンチマーク指標との符号一致率</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>商業動態統計</td> <td>65</td> <td>66%</td> <td>72%</td> <td>68%</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>4</td> <td>25%</td> <td>25%</td> <td>-</td> </tr> <tr> <td>中小企業景況調査</td> <td>22</td> <td>64%</td> <td>45%</td> <td>-</td> </tr> </tbody> </table> <p>【凡例】 符号一致率が25%以下 符号一致率が50%以下</p> <p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>商業動態統計</td> <td>0.66</td> <td>0.76</td> <td>0.64</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>-0.64</td> <td>-0.75</td> <td>-</td> </tr> <tr> <td>中小企業景況調査</td> <td>-</td> <td>-</td> <td>-</td> </tr> </tbody> </table> <p>【凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>➤ 上記の結果より、「中小企業のみ指標」は「前年比」では強い正の相関、「原数値（原指数）」では正の相関がみられ、かつおおむね符号一致率が高いことから「<u>商業動態統計</u>」との相関性があると考えられる</p>	()内は更新頻度		相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	中小企業実態基本調査(年次)	corr	-0.72	-0.76		◎	◎	◎		RMSE	0.23	0.23					商業動態統計(月次)	corr	0.82	0.81	0.91	△	◎	◎		RMSE	0.2	0.19	0.14				中小企業景況調査(四半期)	corr	-0.48	-0.39		◎	◎	○		RMSE(※)							ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	商業動態統計	65	66%	72%	68%	中小企業実態基本調査	4	25%	25%	-	中小企業景況調査	22	64%	45%	-	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	商業動態統計	0.66	0.76	0.64	中小企業実態基本調査	-0.64	-0.75	-	中小企業景況調査	-	-	-
()内は更新頻度		相関性			ベンチマーク指標の前提																																																																																															
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																																																																													
中小企業実態基本調査(年次)	corr	-0.72	-0.76		◎	◎	◎																																																																																													
	RMSE	0.23	0.23																																																																																																	
商業動態統計(月次)	corr	0.82	0.81	0.91	△	◎	◎																																																																																													
	RMSE	0.2	0.19	0.14																																																																																																
中小企業景況調査(四半期)	corr	-0.48	-0.39		◎	◎	○																																																																																													
	RMSE(※)																																																																																																			
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																																																																																
商業動態統計	65	66%	72%	68%																																																																																																
中小企業実態基本調査	4	25%	25%	-																																																																																																
中小企業景況調査	22	64%	45%	-																																																																																																
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																																																																																																	
商業動態統計	0.66	0.76	0.64																																																																																																	
中小企業実態基本調査	-0.64	-0.75	-																																																																																																	
中小企業景況調査	-	-	-																																																																																																	

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-12 に示す。消費増税や感染症に入るポイントは適切に捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「商業動態統計」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-13 に示す。

表 2-12 機械器具小売業の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	▶ 影響を受けていないことを確認できた
消費増税に入るポイント (2019年10月)	▶ 適切に捉えていることを確認できた
感染症に入るポイント (2020年4月)	▶ 適切に捉えていることを確認できた
店舗増加などの影響	▶ 影響を受けていないことを確認できた
ベンチマーク指標と大きく違う動きをしているポイント	▶ 大きく違う動きをしているポイントがないことを確認できた

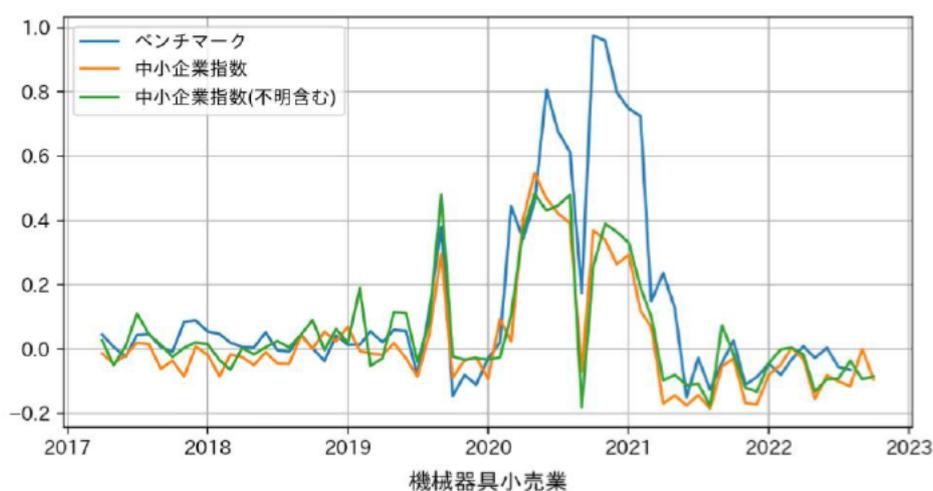


図 2-13 機械器具小売業の中小企業指数とベンチマーク指標（「商業動態統計」）との比較結果

3) 外食の評価

a. 指数の推移

「外食」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-14 に示す。

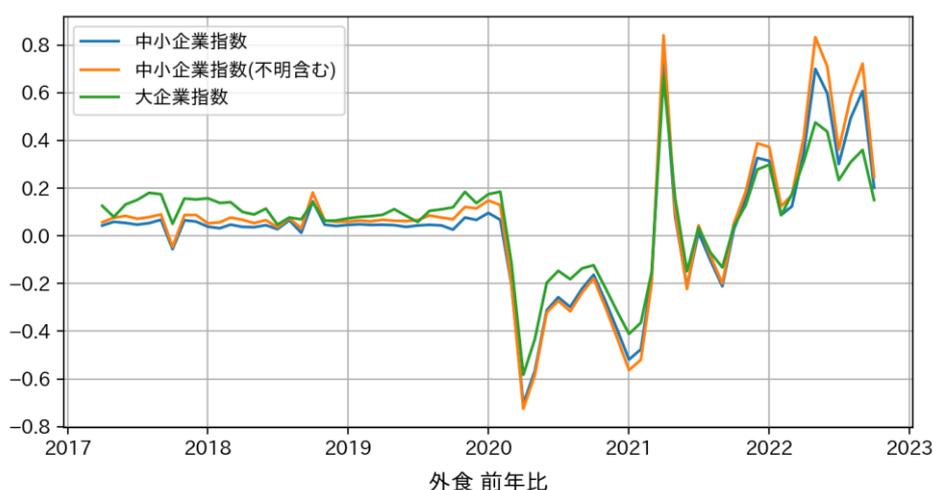


図 2-14 外食の指数の推移

b. 有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して3週間程度の期間短縮が見込める速報性があり、かつほぼ全ての都道府県別・地域別指数を作成できる詳細性がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が低いため正確性の観点は懸念があるが、3つのベンチマーク指標と強い正の相関がみられることから、相関性の観点はおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、相関性の評価結果から、「中小企業のみ指標」と「中小企業及び不明を合算した指標」の2パターンでは、「中小企

業のみの指標」の方が相関係数は高いため、「中小企業のみ指標」が望ましいといえる。

一方で、本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別・地域別指数を作成する際には、旅行等に伴う外食消費等のノイズが含まれている可能性があるといった課題・懸念事項を確認できた。

表 2-13 外食の有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	<p>▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「消費活動指数(当月の結果を、翌々月上旬に公表)」と比べても、<u>3週間程度の期間短縮が見込める</u></p>
	■ 詳細性	<p>▶ 消費者属性(性別、年齢区分など)ごとに十分なサンプル数があるため、<u>ほぼ全ての都道府県別・地域別で指数が作成可能</u>である</p> <p>都道府県別指数を作成することができる都道府県数(全47): 47</p> <p>地域別指数を作成することができる地域数(全10): 9</p> <p>▶ 本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別(地域別)で中小企業の売上動向を把握する際には、旅行等に伴う外食消費等のノイズが含まれている</p>
妥当性	■ 正確性	<p>▶ 1段階目の中小企業の正答率は下記のとおりであり、<u>1段階目の精度が低い</u>(※サンプリングチェックの対象は売上上位企業)</p> <p>✓ 1段階目の中小企業の正答率: 79.0%</p> <p>✓ 2段階目の中小企業の正答率: -</p> <p>✓ 3段階目の中小企業割合: -</p> <p>▶ 本業種は2段階目と3段階目において売上寄与度の大きい企業数は5つ以下であったため、2段階目と3段階目の中小企業の正答率・中小企業割合は算出していない。したがって、売上寄与度で重みづけした正答率は「中小企業のみ指標」「中小企業及び不明を合算した指標」いずれも79.0%であり、<u>精度に懸念が残る</u></p>

評価軸	評価観点	評価結果																																																																																																																																																				
	■ 相関性	<p>▶ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。「消費活動指数」と「日銀短観」、「中小企業景況調査」と強い正の相関がみられた</p> <table border="1"> <thead> <tr> <th rowspan="2">()内は更新頻度</th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td>中小企業実態基本調査(年次)</td> <td>corr</td> <td>0.36</td> <td>0.34</td> <td></td> <td>◎</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.25</td> <td>0.26</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>法人企業統計調査(四半期)</td> <td>corr</td> <td>0.12</td> <td>0.12</td> <td>0.66</td> <td>○</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.34</td> <td>0.36</td> <td>0.21</td> <td></td> <td></td> <td></td> </tr> <tr> <td>消費活動指数(月次)</td> <td>corr</td> <td>0.98</td> <td>0.97</td> <td>0.96</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.06</td> <td>0.09</td> <td>0.11</td> <td></td> <td></td> <td></td> </tr> <tr> <td>日銀短観(四半期)</td> <td>corr</td> <td>0.97</td> <td>0.96</td> <td>0.96</td> <td>○</td> <td>△</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.14</td> <td>0.18</td> <td>0.17</td> <td></td> <td></td> <td></td> </tr> <tr> <td>中小企業景況調査(四半期)</td> <td>corr</td> <td>0.86</td> <td>0.85</td> <td></td> <td>◎</td> <td>◎</td> <td>○</td> </tr> <tr> <td></td> <td>RMSE(※)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>(※)売上DIは前年比で「好転」から「悪化」を引いた数値であり、RMSEで比較を行うことは不適切であると考えられるため算出しない</p> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶ 「前年比」のベンチマーク指標との符号一致率と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <p>✓ 「前年比」のベンチマーク指標との符号一致率</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>48</td> <td>63%</td> <td>63%</td> <td>60%</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>4</td> <td>25%</td> <td>25%</td> <td>-</td> </tr> <tr> <td>法人企業統計調査</td> <td>21</td> <td>43%</td> <td>38%</td> <td>57%</td> </tr> <tr> <td>中小企業景況調査</td> <td>22</td> <td>36%</td> <td>32%</td> <td>-</td> </tr> <tr> <td>短観</td> <td>6</td> <td>67%</td> <td>67%</td> <td>100%</td> </tr> </tbody> </table> <p>【凡例】 符号一致率が25%以下 符号一致率が50%以下</p> <p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>0.89</td> <td>0.85</td> <td>0.59</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>0.18</td> <td>0.21</td> <td>-</td> </tr> <tr> <td>法人企業統計調査</td> <td>0.04</td> <td>-0.02</td> <td>0.57</td> </tr> <tr> <td>中小企業景況調査</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>短観</td> <td>0.92</td> <td>0.87</td> <td>0.75</td> </tr> </tbody> </table> <p>【凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>▶ 上記の結果より、「前年比」及び「原数値（原指数）」では強い正の相関、かつおおむね符号一致率が高いことから「消費活動指数」及び「日銀短観」との相関性があると考えられる。一方で、「中小企業景況調査」とは「前年</p>	()内は更新頻度		相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	中小企業実態基本調査(年次)	corr	0.36	0.34		◎	◎	◎		RMSE	0.25	0.26					法人企業統計調査(四半期)	corr	0.12	0.12	0.66	○	◎	◎		RMSE	0.34	0.36	0.21				消費活動指数(月次)	corr	0.98	0.97	0.96	△	◎	◎		RMSE	0.06	0.09	0.11				日銀短観(四半期)	corr	0.97	0.96	0.96	○	△	◎		RMSE	0.14	0.18	0.17				中小企業景況調査(四半期)	corr	0.86	0.85		◎	◎	○		RMSE(※)							ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	消費活動指数	48	63%	63%	60%	中小企業実態基本調査	4	25%	25%	-	法人企業統計調査	21	43%	38%	57%	中小企業景況調査	22	36%	32%	-	短観	6	67%	67%	100%	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	消費活動指数	0.89	0.85	0.59	中小企業実態基本調査	0.18	0.21	-	法人企業統計調査	0.04	-0.02	0.57	中小企業景況調査	-	-	-	短観	0.92	0.87	0.75
()内は更新頻度		相関性			ベンチマーク指標の前提																																																																																																																																																	
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																																																																																																																															
中小企業実態基本調査(年次)	corr	0.36	0.34		◎	◎	◎																																																																																																																																															
	RMSE	0.25	0.26																																																																																																																																																			
法人企業統計調査(四半期)	corr	0.12	0.12	0.66	○	◎	◎																																																																																																																																															
	RMSE	0.34	0.36	0.21																																																																																																																																																		
消費活動指数(月次)	corr	0.98	0.97	0.96	△	◎	◎																																																																																																																																															
	RMSE	0.06	0.09	0.11																																																																																																																																																		
日銀短観(四半期)	corr	0.97	0.96	0.96	○	△	◎																																																																																																																																															
	RMSE	0.14	0.18	0.17																																																																																																																																																		
中小企業景況調査(四半期)	corr	0.86	0.85		◎	◎	○																																																																																																																																															
	RMSE(※)																																																																																																																																																					
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																																																																																																																																		
消費活動指数	48	63%	63%	60%																																																																																																																																																		
中小企業実態基本調査	4	25%	25%	-																																																																																																																																																		
法人企業統計調査	21	43%	38%	57%																																																																																																																																																		
中小企業景況調査	22	36%	32%	-																																																																																																																																																		
短観	6	67%	67%	100%																																																																																																																																																		
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																																																																																																																																																			
消費活動指数	0.89	0.85	0.59																																																																																																																																																			
中小企業実態基本調査	0.18	0.21	-																																																																																																																																																			
法人企業統計調査	0.04	-0.02	0.57																																																																																																																																																			
中小企業景況調査	-	-	-																																																																																																																																																			
短観	0.92	0.87	0.75																																																																																																																																																			

評価軸	評価観点	評価結果
		比」では強い正の相関がみられるが ¹⁴ 、アッパーバイアス等の要因から符号一致率が低いことを確認できた

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-14 に示す。いくつかの評価観点で懸念があるものの、感染症に入るポイントは適切に捉えていることを確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「消費活動指数」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-15 に示す。

表 2-14 外食の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	➤ 影響を受けていることを確認でき、懸念がある
消費増税に入るポイント (2019年10月)	➤ -
感染症に入るポイント (2020年4月)	➤ 適切に捉えていることを確認できた
店舗増加などの影響	➤ 影響を受けていることを確認でき、懸念がある
ベンチマーク指標と大きく違う動きをしているポイント	➤ 2017年10月頃に確認でき、懸念がある

¹⁴ 同上。29頁。

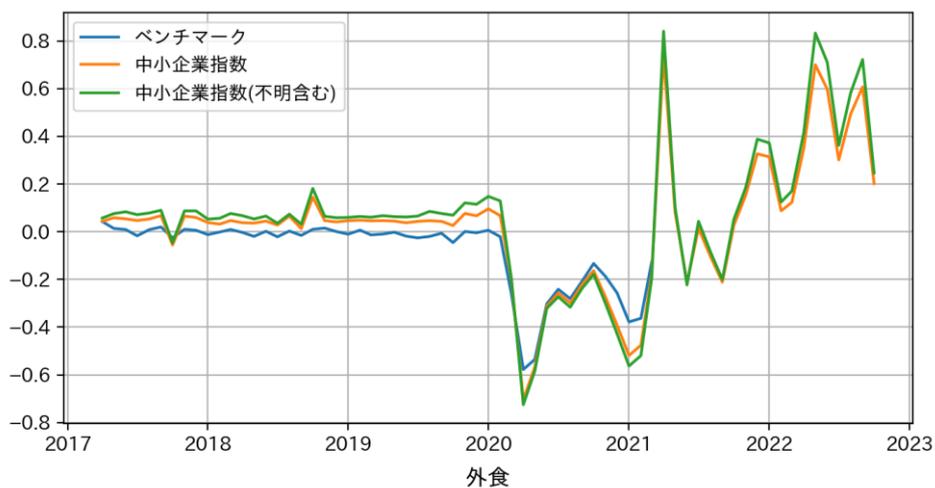


図 2-15 外食の中小企業指数と
ベンチマーク指標（「消費活動指数」）との比較結果

4) 旅行の評価

a. 指数の推移

「旅行」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-16 に示す。

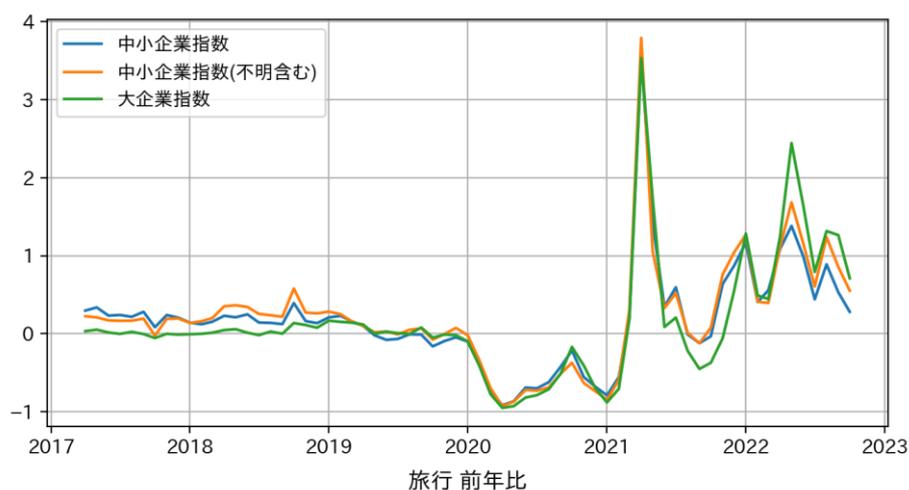


図 2-16 旅行の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して3週間程度の期間短縮が見込める速報性はあるが、特定の都道府県別・地域別指数しか作成できないため詳細性の観点では懸念がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、1つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、正確性の評価結果から、「中小企業のみ指標」と「中小企業及び不明を合算した指標」の2パターンでは、「中小企業のみ指標」が望ましいといえる。

一方で、ベンチマーク指標との比較において、キャッシュレス化の進展などのアッパーバイアス（過大評価）や、本業種における中小企業割合が実態と乖離している可能性があるといった課題・懸念事項を確認できた。

表 2-15 旅行の有効性・妥当性評価

評価軸	評価観点	評価結果																														
有効性	■ 速報性	<p>▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「消費活動指数(当月の結果を、翌々月上旬に公表)」と比べても、<u>3週間程度の期間短縮が見込める</u></p>																														
	■ 詳細性	<p>▶ 都道府県・地域によっては、消費者属性(性別、年齢区分など)ごとに十分なサンプル数が確保できていないため、<u>特定の都道府県別・地域別で指数が作成できない</u></p> <p>都道府県別指数を作成することができる都道府県数(全47)：37 地域別指数を作成することができる地域数(全10)：5</p>																														
妥当性	■ 正確性	<p>▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、<u>1段階目と3段階目はいずれも精度が高い</u>(※サンプリングチェックの対象は売上上位企業)</p> <p>✓ 1段階目の中小企業の正答率：95.0%</p> <p>✓ 2段階目の中小企業の正答率：- (※売上寄与度の大きい企業数は5つ以下であったため、算出していない)</p> <p>✓ 3段階目の中小企業割合：88.2%</p> <p>▶ 売上寄与度で重みづけした正答率は下記のとおりであり、<u>「中小企業のみ指標」の精度が高い</u></p> <p>✓ 中小企業のみ指標：95.0%</p> <p>✓ 中小企業及び不明を合算した指標：91.7%</p>																														
	■ 相関性	<p>▶ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。<u>「消費活動指数」と強い正の相関がみられた</u></p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th rowspan="2"></th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業 + 不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td rowspan="2">()内は更新頻度</td> <td>消費活動指数 (月次)</td> <td>corr</td> <td>0.93</td> <td>0.94</td> <td>0.96</td> <td>◎</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.18</td> <td>0.19</td> <td>0.11</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶ 「前年比」のベンチマーク指標との符号一致率と「原数値(原指数)」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p>			相関性			ベンチマーク指標の前提			中小企業	中小企業 + 不明	大企業	企業サイズ	業種の粒度	売上高指数	()内は更新頻度	消費活動指数 (月次)	corr	0.93	0.94	0.96	◎	◎	◎		RMSE	0.18	0.19	0.11		
		相関性			ベンチマーク指標の前提																											
		中小企業	中小企業 + 不明	大企業	企業サイズ	業種の粒度	売上高指数																									
()内は更新頻度	消費活動指数 (月次)	corr	0.93	0.94	0.96	◎	◎	◎																								
		RMSE	0.18	0.19	0.11																											

評価軸	評価観点	評価結果																		
		<p>✓ 「前年比」のベンチマーク指標との符号一致率</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>48</td> <td>90%</td> <td>85%</td> <td>75%</td> </tr> </tbody> </table> <p>【凡例】 ■ 符号一致率が25%以下 ■ 符号一致率が50%以下</p> <p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>0.84</td> <td>0.79</td> <td>0.88</td> </tr> </tbody> </table> <p>【凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>➤ 上記の結果より、「前年比」及び「原数値（原指数）」では強い正の相関がみられ、かつ符号一致率が高いことから「消費活動指数」との相関性があると考えられる</p>	ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	消費活動指数	48	90%	85%	75%	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	消費活動指数	0.84	0.79	0.88
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																
消費活動指数	48	90%	85%	75%																
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																	
消費活動指数	0.84	0.79	0.88																	

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-16 に示す。いくつかの評価観点で懸念があるものの、感染症に入るポイントは適切に捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「消費活動指数」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-17 に示す。

表 2-16 旅行の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	➤ 影響を受けていないことを確認できた
消費増税に入るポイント (2019年10月)	➤ -
感染症に入るポイント (2020年4月)	➤ 適切に捉えていることを確認できた
店舗増加などの影響	➤ 影響を受けていることを確認でき、懸念がある
ベンチマーク指標と大きく違う動きをしているポイント	➤ 2017年10月頃に確認でき、懸念がある

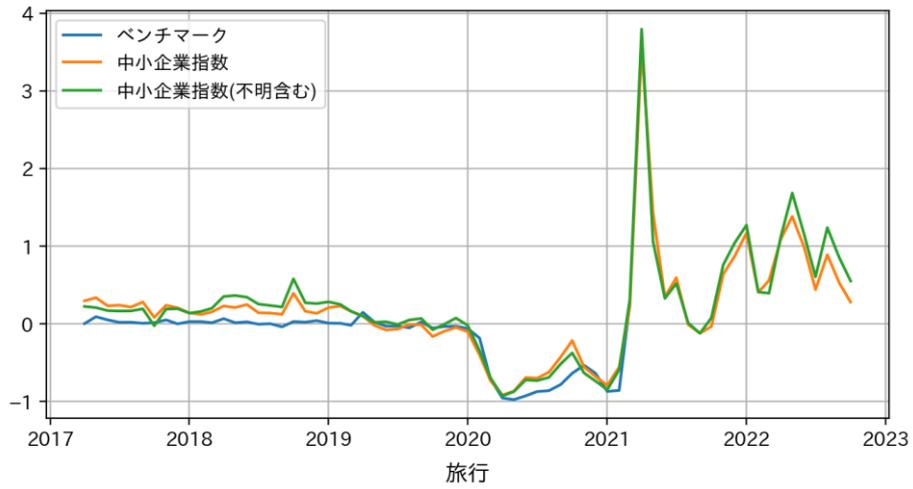


図 2-17 旅行の中小企業指数と
ベンチマーク指標（「消費活動指数」）との比較結果

5) 宿泊の評価

a. 指数の推移

「宿泊」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-18 に示す。

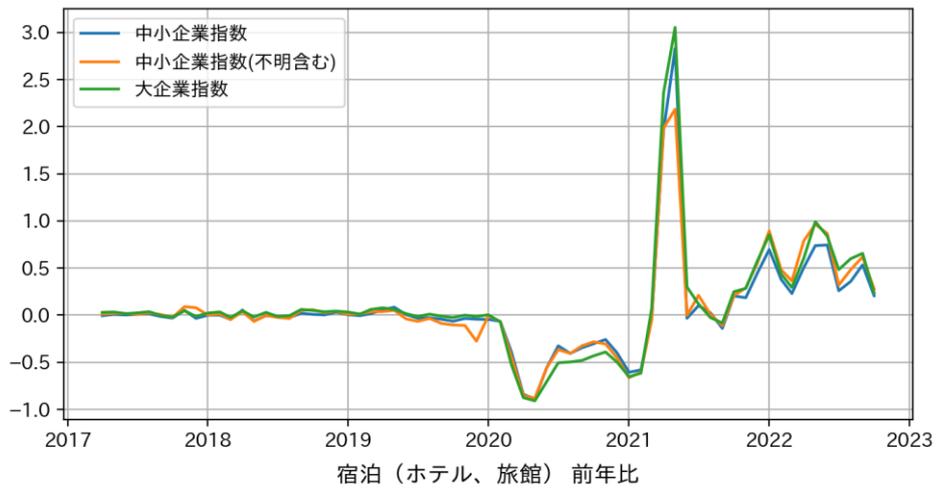


図 2-18 宿泊の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して3週間程度の期間短縮が見込める速報性はある。一方で、本指標は企業の所在地ではなく消費者の居住地を使用しており、宿泊旅行に伴う消費が多く存在すると推察されるため、詳細性の観点での評価は実施していない。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、4つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、活用可能性が高い指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、正確性の評価結果から、「中小企業のみ指標」と「中小企業及び不明を合算した指標」の2パターンでは、「中小企業のみ指標」が望ましいといえる。

一方で、企業の所在地を基にした地域別の売上高変化を捉えることは困難である（「2-3-2.（2）地域別指数の精度」参照）といった課題・懸念事項が確認できた。

表 2-17 宿泊の有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■速報性	<ul style="list-style-type: none"> ▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「消費活動指数(当月の結果を、翌々月上旬に公表)」と比べても、<u>3週間程度の期間短縮が見込める</u>
	■詳細性	<ul style="list-style-type: none"> ▶ 本指標は企業の所在地ではなく消費者の居住地を使用しており、<u>宿泊旅行に伴う消費が多く存在すると推察される</u>。このため、都道府県別・地域別で中小企業の売上動向を把握することが可能であるか評価するこの基準は<u>評価対象外</u>とする ▶ なお、消費者属性（性別、年齢区分など）ごとに十分なサンプル数があるため、ほぼ全ての都道府県別・地域別で指数が作成可能である 都道府県別指数を作成することができる都道府県数(全47)：47 地域別指数を作成することができる地域数(全10)：7

評価軸	評価観点	評価結果																																																																																																																						
妥当性	■ 正確性	<p>▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、<u>1段階目と2段階目の精度は高い</u>が3段階目の精度が低い（※サンプリングチェックの対象は売上上位企業）</p> <ul style="list-style-type: none"> ✓ 1段階目の中小企業の正答率：100.0% ✓ 2段階目の中小企業の正答率：93.8% ✓ 3段階目の中小企業割合：61.1% <p>▶ 売上寄与度で重みづけした正答率は下記のとおりであり、<u>「中小企業のみ</u>の指標」の精度が高い</p> <ul style="list-style-type: none"> ✓ 中小企業のみ指標：<u>98.6%</u> ✓ 中小企業及び不明を合算した指標：86.5% 																																																																																																																						
	■ 相関性	<p>▶ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。<u>「中小企業実態基本調査」と「消費活動指数」、「日銀短観」、「中小企業景況調査」と強い正の相関</u>がみられる</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th rowspan="2"></th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td rowspan="2">中小企業実態基本調査(年次)</td> <td>corr</td> <td>0.91</td> <td>0.94</td> <td>-</td> <td>◎</td> <td>◎</td> <td>◎</td> </tr> <tr> <td>RMSE</td> <td>0.17</td> <td>0.17</td> <td>-</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">法人企業統計調査(四半期)</td> <td>corr</td> <td>0.67</td> <td>0.63</td> <td>0.82</td> <td>○</td> <td>◎</td> <td>◎</td> </tr> <tr> <td>RMSE</td> <td>0.3</td> <td>0.33</td> <td>0.27</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">消費活動指数(月次)</td> <td>corr</td> <td>0.96</td> <td>0.93</td> <td>0.97</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> <tr> <td>RMSE</td> <td>0.07</td> <td>0.09</td> <td>0.08</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">日銀短観(四半期)</td> <td>corr</td> <td>0.97</td> <td>0.96</td> <td>0.96</td> <td>○</td> <td>△</td> <td>◎</td> </tr> <tr> <td>RMSE</td> <td>0.14</td> <td>0.18</td> <td>0.17</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">中小企業景況調査(四半期)</td> <td>corr</td> <td>0.79</td> <td>0.8</td> <td>-</td> <td>◎</td> <td>◎</td> <td>○</td> </tr> <tr> <td>RMSE(※)</td> <td>-</td> <td>-</td> <td>-</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>(※)売上DIは前年比で「好転」から「悪化」を引いた数値であり、RMSEで比較を行うことは不適切であると考えられるため算出しない</p> <p>【corrにおける凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 ■ 誤差が小さい (<0.3) ■ 誤差がやや大きい (>=0.3) ■ 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶ 「前年比」のベンチマーク指標との符号一致率と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <ul style="list-style-type: none"> ✓ 「前年比」のベンチマーク指標との符号一致率 <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>48</td> <td>60%</td> <td>69%</td> <td>71%</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>4</td> <td>50%</td> <td>100%</td> <td>-</td> </tr> <tr> <td>法人企業統計調査</td> <td>21</td> <td>62%</td> <td>67%</td> <td>76%</td> </tr> <tr> <td>中小企業景況調査</td> <td>22</td> <td>63%</td> <td>59%</td> <td>-</td> </tr> <tr> <td>短観</td> <td>6</td> <td>67%</td> <td>67%</td> <td>100%</td> </tr> </tbody> </table> <p>【凡例】 ■ 符号一致率が25%以下 ■ 符号一致率が50%以下</p>			相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	中小企業実態基本調査(年次)	corr	0.91	0.94	-	◎	◎	◎	RMSE	0.17	0.17	-				法人企業統計調査(四半期)	corr	0.67	0.63	0.82	○	◎	◎	RMSE	0.3	0.33	0.27				消費活動指数(月次)	corr	0.96	0.93	0.97	△	◎	◎	RMSE	0.07	0.09	0.08				日銀短観(四半期)	corr	0.97	0.96	0.96	○	△	◎	RMSE	0.14	0.18	0.17				中小企業景況調査(四半期)	corr	0.79	0.8	-	◎	◎	○	RMSE(※)	-	-	-				ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	消費活動指数	48	60%	69%	71%	中小企業実態基本調査	4	50%	100%	-	法人企業統計調査	21	62%	67%	76%	中小企業景況調査	22	63%	59%	-	短観	6	67%	67%
		相関性			ベンチマーク指標の前提																																																																																																																			
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																																																																																																	
中小企業実態基本調査(年次)	corr	0.91	0.94	-	◎	◎	◎																																																																																																																	
	RMSE	0.17	0.17	-																																																																																																																				
法人企業統計調査(四半期)	corr	0.67	0.63	0.82	○	◎	◎																																																																																																																	
	RMSE	0.3	0.33	0.27																																																																																																																				
消費活動指数(月次)	corr	0.96	0.93	0.97	△	◎	◎																																																																																																																	
	RMSE	0.07	0.09	0.08																																																																																																																				
日銀短観(四半期)	corr	0.97	0.96	0.96	○	△	◎																																																																																																																	
	RMSE	0.14	0.18	0.17																																																																																																																				
中小企業景況調査(四半期)	corr	0.79	0.8	-	◎	◎	○																																																																																																																	
	RMSE(※)	-	-	-																																																																																																																				
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																																																																																																				
消費活動指数	48	60%	69%	71%																																																																																																																				
中小企業実態基本調査	4	50%	100%	-																																																																																																																				
法人企業統計調査	21	62%	67%	76%																																																																																																																				
中小企業景況調査	22	63%	59%	-																																																																																																																				
短観	6	67%	67%	100%																																																																																																																				

評価軸	評価観点	評価結果																								
		<p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>0.78</td> <td>0.80</td> <td>0.87</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>0.73</td> <td>0.74</td> <td>-</td> </tr> <tr> <td>法人企業統計調査</td> <td>0.79</td> <td>0.84</td> <td>0.84</td> </tr> <tr> <td>中小企業景況調査</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>短観</td> <td>0.92</td> <td>0.87</td> <td>0.75</td> </tr> </tbody> </table> <p>【凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>➤ 上記の結果より、「前年比」及び「原数値（原指数）」では強い正の相関、かつおおむね符号一致率が高いことから「消費活動指数」及び「日銀短観」との相関性があると考えられる。また、「前年比」では強い正の相関、かつおおむね符号一致率が高いことから「中小企業景況調査」とも相関性があると考えられる。一方で、「中小企業実態基本調査」とは「前年比」及び「原数値（原指数）」では強い正の相関がみられるが、「中小企業指標のみの指標」においては、下記に示すように、「前年比」で若干マイナス傾向である等の要因から符号一致率が低いことを確認できた</p>	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	消費活動指数	0.78	0.80	0.87	中小企業実態基本調査	0.73	0.74	-	法人企業統計調査	0.79	0.84	0.84	中小企業景況調査	-	-	-	短観	0.92	0.87	0.75
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																							
消費活動指数	0.78	0.80	0.87																							
中小企業実態基本調査	0.73	0.74	-																							
法人企業統計調査	0.79	0.84	0.84																							
中小企業景況調査	-	-	-																							
短観	0.92	0.87	0.75																							

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-18 に示す。消費増税に入るポイントは捉えていないものの、感染症に入るポイントは適切に捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「消費活動指数」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-19 に示す。

表 2-18 宿泊の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	▶ 影響を受けていないことを確認できた
消費増税に入るポイント (2019年10月)	▶ 適切に捉えていないことを確認でき、懸念がある
感染症に入るポイント (2020年4月)	▶ 適切に捉えていることを確認できた
店舗増加などの影響	▶ 影響を受けていないことを確認できた
ベンチマーク指標と大きく違う動きをしているポイント	▶ 大きく違う動きをしているポイントがないことを確認できた

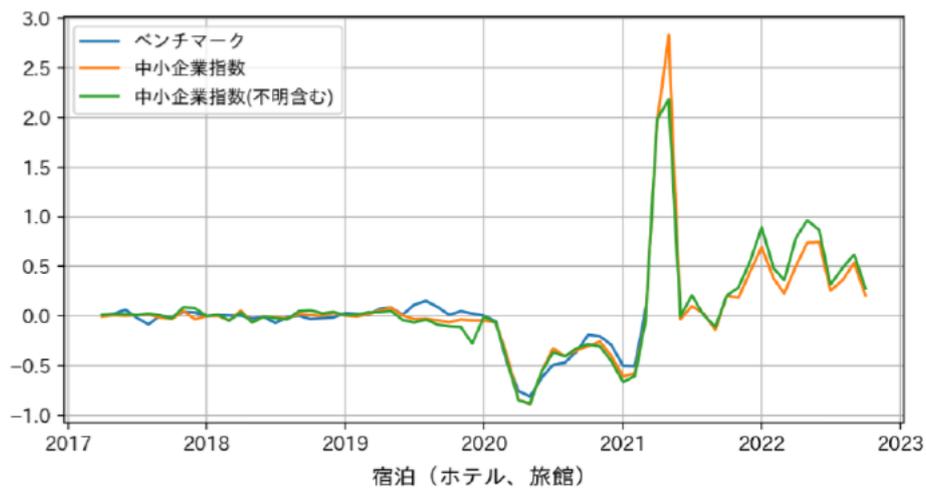


図 2-19 宿泊の中小企業指数と
ベンチマーク指標（「消費活動指数」）との比較結果

6) EC の評価

a. 指数の推移

「EC」においては研究会の内容を踏まえ、「外資大手 EC データを中小企業として扱う」と「外資大手 EC データを大企業として扱う」それぞれのパターンで試作した

¹⁵ため、本業種の中小企業指標として「中小企業指数（外資大手 EC を大企業と扱う）」と「中小企業指数（外資大手 EC を中小企業と扱う）」、「中小企業指数（不明含む、外資大手 EC を大企業と扱う）」、「中小企業指数（不明含む、外資大手 EC を中小企業と扱う）」の 4 パターンを作成した。同様に、本業種の大企業指標として「大企業指数（外資大手 EC を大企業と扱う）」と「大企業指数（外資大手 EC を中小企業と扱う）」の 2 パターンを作成した。試作した本業種における各指数の推移を図 2-20 に示す。

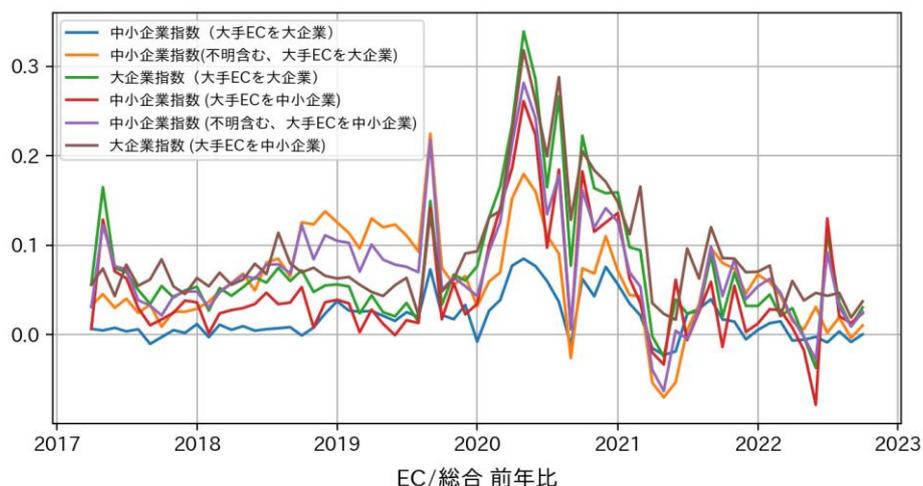


図 2-20 EC の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して 2 週間程度の期間短縮が見込める速報性はある。一方で、本指標は企業の所在地ではなく消費者の居住地を使用しているため、詳細性の観点での評価は実施していない。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が低いため正確性の観点は懸念があるが、1つのベンチマーク指標と強い正の相関がみられることから、相関性の観点はおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断については、正確性と相関性の評価結果か

¹⁵ 外資大手 EC データについては、加盟店情報を取得できない出品形式を採っており、加盟店ごとに大企業あるいは中小企業と判定することが困難であるため。

ら、「中小企業のための指標」と「中小企業及び不明を合算した指標」の2パターンでは、正確性においては「中小企業のための指標」が望ましいといえる。さらに、外資大手ECデータを「大企業」もしくは「中小企業」として扱うべきかの判断について、相関性の評価結果から「外資大手ECデータを中小企業として扱う」パターンが望ましいといえるが、外資大手ECデータをどのように扱うかについては、外資大手ECの実態（全加盟店のうち、中小企業がどの程度の割合を占めるのか）に即した対応を別途検討していく必要がある。

一方で、企業の所在地を基にした地域別の売上高変化を捉えることは困難であることや（「2-3-2.（2）地域別指数の精度」参照）、本業種における中小企業割合が実態と乖離している可能性があるといった課題・懸念事項が確認できた。

表 2-19 ECの有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	<p>▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「商業動態統計(当月の結果を、翌月末に公表)」と比べても、<u>2週間程度の期間短縮が見込める</u></p>
	■ 詳細性	<p>▶ 本指標は企業の所在地ではなく消費者の居住地を使用しているため、都道府県別・地域別で中小企業の売上動向を把握することが可能であるか評価するこの基準は<u>評価対象外</u>とする</p> <p>▶ なお、消費者属性（性別、年齢区分など）ごとに十分なサンプル数があるため、全ての都道府県別・地域別で指数が作成可能である</p> <p>都道府県別指数を作成することができる都道府県数(全47)：47</p> <p>地域別指数を作成することができる地域数(全10)：10</p>
妥当性	■ 正確性	<p>▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、<u>1段階目は精度が高い</u>が、2段階目と3段階目の精度が低い（※サンプリングチェックの対象は売上上位企業）</p> <p>✓ 1段階目の中小企業の正答率：82.4%</p> <p>✓ 2段階目の中小企業の正答率：42.9%</p> <p>✓ 3段階目の中小企業割合：42.1%</p> <p>▶ 売上寄与度で重みづけした正答率は下記のとおりであり、<u>「中小企業のための指標」の精度が高い</u>が、71.5%とやや正答率は低く、中小企業分類マスタの精度に懸念がある</p> <p>✓ 中小企業のための指標：<u>71.5%</u></p>

評価軸	評価観点	評価結果																																																																																												
		<p>✓ 中小企業及び不明を合算した指標：49.4%</p>																																																																																												
	<p>■ 相関性</p>	<p>➤ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。「外資大手ECデータを中小企業として扱う」パターンにおいて、「<u>商業動態統計</u>」と強い正の相関がみられた。</p> <table border="1"> <thead> <tr> <th rowspan="2">()内は更新頻度</th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td>中小企業 実態基本 調査(年次)</td> <td>外資大手 : 中小企業 corr RMSE</td> <td>-0.07 1.1</td> <td>0.3 1.09</td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>外資大手 : 大企業 corr RMSE</td> <td>0.08 1.11</td> <td>0.65 1.08</td> <td>◎</td> <td>◎</td> <td>◎</td> </tr> <tr> <td>商業動態 統計 (月次)</td> <td>外資大手 : 中小企業 corr RMSE</td> <td>0.71 0.15</td> <td>0.6 0.16</td> <td>0.81 0.14</td> <td>△</td> <td>◎</td> </tr> <tr> <td></td> <td>外資大手 : 大企業 corr RMSE</td> <td>0.69 0.19</td> <td>0.28 0.18</td> <td>0.78 0.14</td> <td></td> <td>◎</td> </tr> </tbody> </table> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4) 【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0) 【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>➤ 「前年比」のベンチマーク指標との符号一致率と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <p>✓ 「前年比」のベンチマーク指標との符号一致率</p> <table border="1"> <thead> <tr> <th>パターン</th> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td rowspan="2">外資大手：中小企業</td> <td>商業動態統計</td> <td>65</td> <td>62%</td> <td>62%</td> <td>60%</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>4</td> <td>50%</td> <td>50%</td> <td>-</td> </tr> <tr> <td rowspan="2">外資大手：大企業</td> <td>商業動態統計</td> <td>65</td> <td>67%</td> <td>60%</td> <td>60%</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>4</td> <td>50%</td> <td>50%</td> <td>-</td> </tr> </tbody> </table> <p>【凡例】 符号一致率が25%以下 符号一致率が50%以下</p> <p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>パターン</th> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td rowspan="2">外資大手：中小企業</td> <td>商業動態統計</td> <td>0.64</td> <td>0.81</td> <td>0.92</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>-0.08</td> <td>-0.04</td> <td>-</td> </tr> <tr> <td rowspan="2">外資大手：大企業</td> <td>商業動態統計</td> <td>0.65</td> <td>0.82</td> <td>0.91</td> </tr> <tr> <td>中小企業実態基本調査</td> <td>-0.06</td> <td>-0.01</td> <td>-</td> </tr> </tbody> </table> <p>【凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p>	()内は更新頻度	相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	中小企業 実態基本 調査(年次)	外資大手 : 中小企業 corr RMSE	-0.07 1.1	0.3 1.09					外資大手 : 大企業 corr RMSE	0.08 1.11	0.65 1.08	◎	◎	◎	商業動態 統計 (月次)	外資大手 : 中小企業 corr RMSE	0.71 0.15	0.6 0.16	0.81 0.14	△	◎		外資大手 : 大企業 corr RMSE	0.69 0.19	0.28 0.18	0.78 0.14		◎	パターン	ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	外資大手：中小企業	商業動態統計	65	62%	62%	60%	中小企業実態基本調査	4	50%	50%	-	外資大手：大企業	商業動態統計	65	67%	60%	60%	中小企業実態基本調査	4	50%	50%	-	パターン	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	外資大手：中小企業	商業動態統計	0.64	0.81	0.92	中小企業実態基本調査	-0.08	-0.04	-	外資大手：大企業	商業動態統計	0.65	0.82	0.91	中小企業実態基本調査	-0.06	-0.01	-
()内は更新頻度	相関性			ベンチマーク指標の前提																																																																																										
	中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																																																																								
中小企業 実態基本 調査(年次)	外資大手 : 中小企業 corr RMSE	-0.07 1.1	0.3 1.09																																																																																											
	外資大手 : 大企業 corr RMSE	0.08 1.11	0.65 1.08	◎	◎	◎																																																																																								
商業動態 統計 (月次)	外資大手 : 中小企業 corr RMSE	0.71 0.15	0.6 0.16	0.81 0.14	△	◎																																																																																								
	外資大手 : 大企業 corr RMSE	0.69 0.19	0.28 0.18	0.78 0.14		◎																																																																																								
パターン	ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																																																																									
外資大手：中小企業	商業動態統計	65	62%	62%	60%																																																																																									
	中小企業実態基本調査	4	50%	50%	-																																																																																									
外資大手：大企業	商業動態統計	65	67%	60%	60%																																																																																									
	中小企業実態基本調査	4	50%	50%	-																																																																																									
パターン	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																																																																																										
外資大手：中小企業	商業動態統計	0.64	0.81	0.92																																																																																										
	中小企業実態基本調査	-0.08	-0.04	-																																																																																										
外資大手：大企業	商業動態統計	0.65	0.82	0.91																																																																																										
	中小企業実態基本調査	-0.06	-0.01	-																																																																																										

評価軸	評価観点	評価結果
		<p>▶ 上記の結果より、「前年比」では強い正の相関がみられ、かつおおむね符号一致率が高いことから「商業動態統計」との相関性があると考えられる。一方で、「原数値（原指数）」で強い正の相関がみられたのは、正確性の観点で懸念がある「中小企業及び不明を合算した指標」であり、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断は別途検討していく必要がある</p>

c. 定性評価

本業種の中小企業指標（外資大手データ：中小企業）¹⁶に対する定性評価の結果を表 2-20 に示す。いくつかの評価観点で懸念があるものの、消費増税や感染症に入るポイントは捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「商業動態統計」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-21 に示す。

表 2-20 EC（外資大手データ：中小企業）の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	▶ 影響を受けていないことを確認できた
消費増税に入るポイント (2019年10月)	▶ 適切に捉えていることを確認できた
感染症に入るポイント (2020年4月)	▶ 適切に捉えていることを確認できた
店舗増加などの影響	▶ 影響を受けていることを確認でき、懸念がある
ベンチマーク指標と大きく違う動きをしているポイント	▶ 2017年2月頃と2020年初頭から2021年初頭にかけて確認でき、懸念がある

¹⁶ 「外資大手 EC データを中小企業として扱う」パターンが「前年比」の「商業動態統計」と強い正の相関がみられたため、本紙では当該パターンの定性評価を記載。

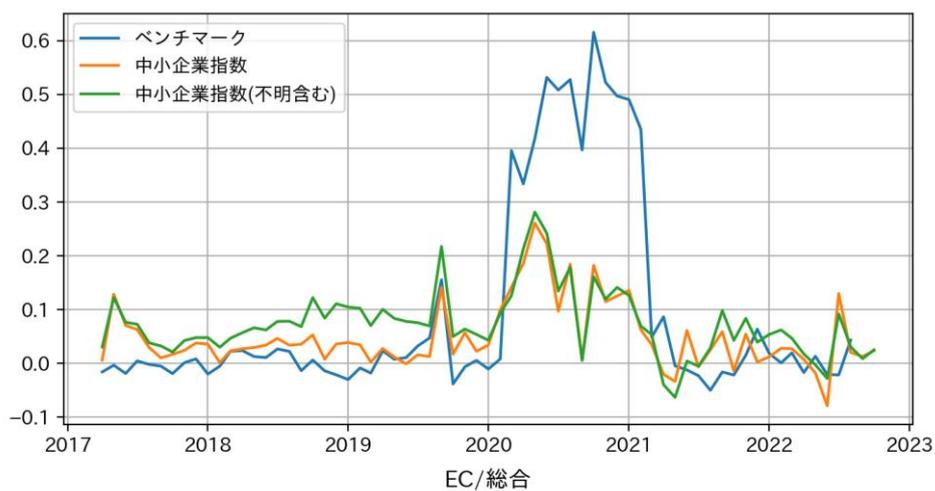


図 2-21 EC の中小企業指数と
ベンチマーク指標（「商業動態統計」）との比較結果

7) 居酒屋の評価

a. 指数の推移

「居酒屋」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-22 に示す。

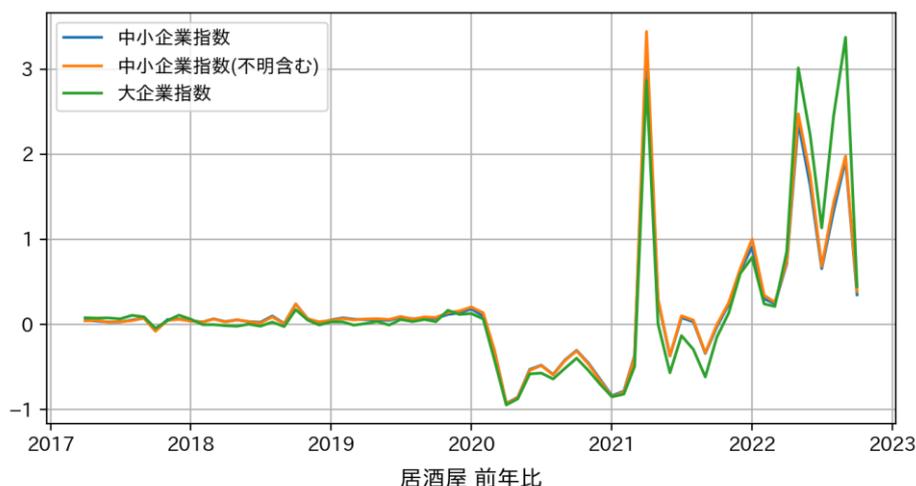


図 2-22 居酒屋の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して2週間程度の期間短縮が見込める速報性はあるが、特定の都道府県別指数しか作成できないため詳細性の観点では懸念がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、1つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、相関性の評価結果では「中小企業のみ指標」の相関係数が若干高いが、本業種は「不明」データが極端に少ないため、「中小企業のみ指標」と「中小企業及び不明を合算した指標」いずれのパターンでも相違ないといえる。

一方で、ベンチマーク指標との比較において、平均平方二乗誤差 (RMSE) の評価結果から、誤差少なく適切に捉えていない可能性があるといった課題・懸念事項を確認できた。

表 2-21 居酒屋の有効性・妥当性評価

評価軸	評価観点	評価結果																																			
有効性	■ 速報性	<p>▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「外食産業市場動向調査(当月の結果を、翌月末に公表)」と比べても、<u>2週間程度の期間短縮が見込める</u></p>																																			
	■ 詳細性	<p>▶ 都道府県・地域によっては、消費者属性(性別、年齢区分など)ごとに十分なサンプル数が確保できていないため、<u>特定の都道府県別で指数が作成できない</u></p> <p>都道府県別指数を作成することができる都道府県数(全47)：31 地域別指数を作成することができる地域数(全10)：10</p>																																			
妥当性	■ 正確性	<p>▶ 1段階目の中小企業の正答率は下記のとおりであり、<u>1段階目の精度が非常に高い</u>(※サンプリングチェックの対象は売上上位企業)</p> <ul style="list-style-type: none"> ✓ 1段階目の中小企業の正答率：100.0% ✓ 2段階目の中小企業の正答率：- ✓ 3段階目の中小企業割合：- <p>▶ 本業種は2段階目と3段階目において売上寄与度の大きい企業数は5つ以下であったため、2段階目と3段階目の中小企業の正答率・中小企業割合は算出していない。したがって、売上寄与度で重みづけした正答率は「中小企業のみ指標」「中小企業及び不明を合算した指標」いずれも100.0%である</p>																																			
	■ 相関性	<p>▶ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。<u>「外食産業市場動向調査」と強い正の相関がみられた</u></p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th rowspan="2"></th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td rowspan="2">外食産業市場動向調査(月次)</td> <td>corr</td> <td style="background-color: #d9ead3;">0.88</td> <td style="background-color: #d9ead3;">0.87</td> <td style="background-color: #d9ead3;">0.97</td> <td rowspan="2" style="text-align: center;">△</td> <td rowspan="2" style="text-align: center;">◎</td> <td rowspan="2" style="text-align: center;">◎</td> </tr> <tr> <td>RMSE</td> <td style="background-color: #d9ead3;">0.37</td> <td style="background-color: #d9ead3;">0.37</td> <td style="background-color: #d9ead3;">0.2</td> </tr> </tbody> </table> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶ 「前年比」のベンチマーク指標との符号一致率と「原数値(原指数)」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <ul style="list-style-type: none"> ✓ <u>「前年比」のベンチマーク指標との符号一致率</u> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>外食産業市場動向調査</td> <td style="text-align: center;">65</td> <td style="text-align: center;">57%</td> <td style="text-align: center;">57%</td> <td style="text-align: center;">71%</td> </tr> </tbody> </table> <p style="text-align: right;">【凡例】 符号一致率が25%以下 符号一致率が50%以下</p>			相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	外食産業市場動向調査(月次)	corr	0.88	0.87	0.97	△	◎	◎	RMSE	0.37	0.37	0.2	ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	外食産業市場動向調査	65	57%	57%
		相関性			ベンチマーク指標の前提																																
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																														
外食産業市場動向調査(月次)	corr	0.88	0.87	0.97	△	◎	◎																														
	RMSE	0.37	0.37	0.2																																	
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																	
外食産業市場動向調査	65	57%	57%	71%																																	

評価軸	評価観点	評価結果								
		<p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>外食産業市場動向調査</td> <td>0.57</td> <td>0.57</td> <td>0.71</td> </tr> </tbody> </table> <p>【凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>▶ 上記の結果より、「前年比」では強い正の相関、「原数値（原指数）」では正の相関がみられ、かつおおむね符号一致率が高いことから「外食産業市場動向調査」との相関性があると考えられる</p>	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	外食産業市場動向調査	0.57	0.57	0.71
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)							
外食産業市場動向調査	0.57	0.57	0.71							

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-22 に示す。いくつかの評価観点で懸念があるものの、感染症に入るポイントは適切に捉えていることを確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「外食産業市場動向調査」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-23 に示す。

表 2-22 居酒屋の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	▶ 影響を受けていることを確認でき、懸念がある
消費増税に入るポイント (2019年10月)	▶ -
感染症に入るポイント (2020年4月)	▶ 適切に捉えていることを確認できた
店舗増加などの影響	▶ 影響を受けていることを確認でき、懸念がある
ベンチマーク指標と大きく違う動きをしているポイント	▶ 2017年10月頃に確認でき、懸念がある

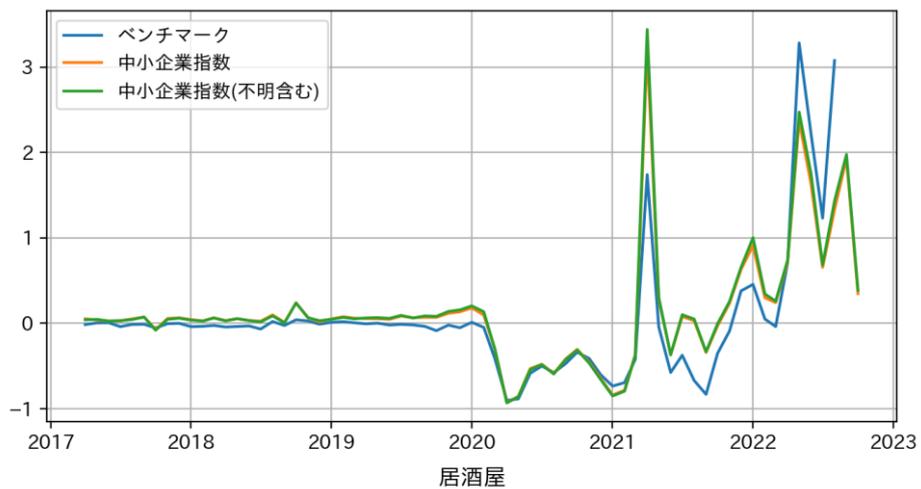


図 2-23 居酒屋の中小企業指数と
ベンチマーク指標（「外食産業市場動向調査」）との比較結果

8) 焼肉

a. 指数の推移

「焼肉」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-24 に示す。

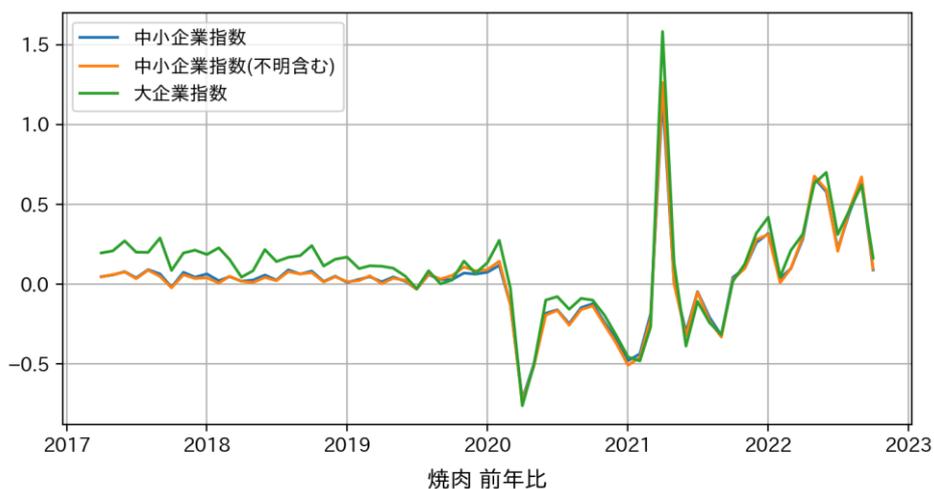


図 2-24 焼肉の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して2週間程度の期間短縮が見込める速報性はあるが、特定の都道府県別・地域別指数しか作成できないため詳細性の観点では懸念がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、1つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、本業種は「不明」データが極端に少ないため、「中小企業のみ指標」と「中小企業及び不明を合算した指標」いずれのパターンでも相違ないといえる。

一方で、本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別・地域別指数を作成する際には、旅行等に伴う焼肉における消費等のノイズが含まれている可能性があるといった課題・懸念事項を確認できた。

表 2-23 焼肉の有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	<p>▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「外食産業市場動向調査(当月の結果を、翌月末に公表)」と比べても、<u>2週間程度の期間短縮が見込める</u></p>
	■ 詳細性	<p>▶ 都道府県・地域によっては、消費者属性(性別、年齢区分など)ごとに十分なサンプル数が確保できていないため、<u>特定の都道府県別・地域別で指数が作成できない</u></p> <p>都道府県別指数を作成することができる都道府県数(全47): 34 地域別指数を作成することができる地域数(全10): 8</p>
妥当性	■ 正確性	<p>▶ 1段階目の中小企業の正答率は下記のとおりであり、<u>1段階目の精度が高い</u>(※サンプリングチェックの対象は売上上位企業)</p> <p>✓ 1段階目の中小企業の正答率: 94.1%</p> <p>✓ 2段階目の中小企業の正答率: -</p> <p>✓ 3段階目の中小企業割合: -</p>

評価軸	評価観点	評価結果																																																
		<p>▶本業種は2段階目と3段階目において売上寄与度の大きい企業数は5つ以下であったため、2段階目と3段階目の中小企業の正答率・中小企業割合は算出していません。したがって、売上寄与度で重みづけした正答率は「中小企業のみ指標」「中小企業及び不明を合算した指標」いずれも94.1%である</p>																																																
	■相関性	<p>▶「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。「<u>外食産業市場動向調査</u>」と強い正の相関がみられた</p> <table border="1"> <thead> <tr> <th rowspan="2">()内は更新頻度</th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td>外食産業市場動向調査(月次)</td> <td>corr</td> <td>0.94</td> <td>0.94</td> <td>0.95</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> <tr> <td></td> <td>RMSE</td> <td>0.1</td> <td>0.1</td> <td>0.11</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4) 【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0) 【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶「前年比」のベンチマーク指標との符号一致率と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <p>✓ 「前年比」のベンチマーク指標との符号一致率</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>外食産業市場動向調査</td> <td>65</td> <td>92%</td> <td>92%</td> <td>94%</td> </tr> </tbody> </table> <p>【凡例】 符号一致率が25%以下 符号一致率が50%以下</p> <p>✓ 「原数値（原指数）」のベンチマーク指標との相関係数</p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>外食産業市場動向調査</td> <td>0.92</td> <td>0.92</td> <td>0.94</td> </tr> </tbody> </table> <p>【凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>▶上記の結果より、「前年比」及び「原数値（原指数）」では強い正の相関がみられ、かつ符号一致率が高いことから「<u>外食産業市場動向調査</u>」との相関性があると考えられる</p>	()内は更新頻度		相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	外食産業市場動向調査(月次)	corr	0.94	0.94	0.95	△	◎	◎		RMSE	0.1	0.1	0.11				ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	外食産業市場動向調査	65	92%	92%	94%	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	外食産業市場動向調査	0.92	0.92	0.94
()内は更新頻度		相関性			ベンチマーク指標の前提																																													
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																											
外食産業市場動向調査(月次)	corr	0.94	0.94	0.95	△	◎	◎																																											
	RMSE	0.1	0.1	0.11																																														
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																														
外食産業市場動向調査	65	92%	92%	94%																																														
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																																															
外食産業市場動向調査	0.92	0.92	0.94																																															

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-24 に示す。消費増税や感染症に入るポイントは適切に捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「外食産業市場動向調査」であり、本業種の中小企

業指数とベンチマーク指標との比較結果を図 2-25 に示す。

表 2-24 焼肉の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	➤ 影響を受けていないことを確認できた
消費増税に入るポイント (2019年10月)	➤ -
感染症に入るポイント (2020年4月)	➤ 適切に捉えていることを確認できた
店舗増加などの影響	➤ 影響を受けていないことを確認できた
ベンチマーク指標と大きく違う動きをしているポイント	➤ 大きく違う動きをしているポイントがないことを確認できた

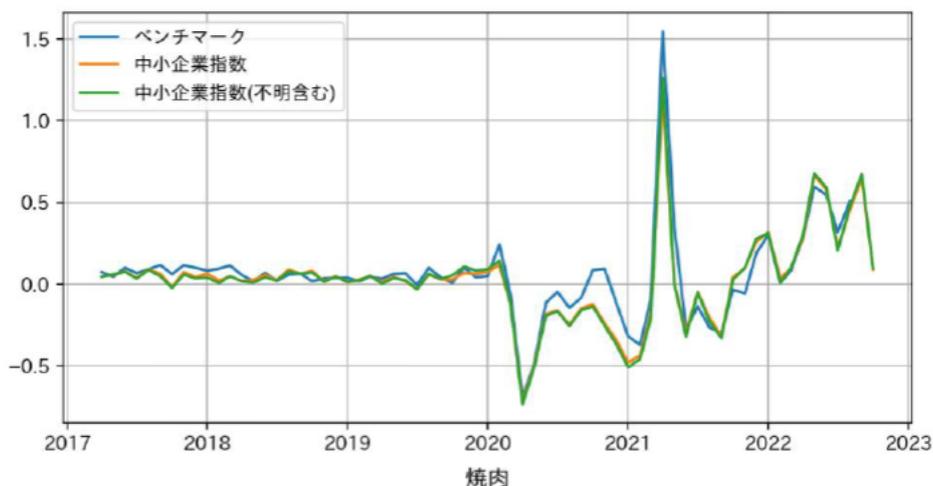


図 2-25 焼肉の中小企業指数と
ベンチマーク指標（「外食産業市場動向調査」）との比較結果

9) ゴルフ場

a. 指数の推移

「ゴルフ場」においては「中小企業指数」と「中小企業指数（不明含む）」の2パ

ターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-26 に示す。



図 2-26 ゴルフ場の指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して3週間程度の期間短縮が見込める速報性はあるが、特定の都道府県別・地域別指数しか作成できないため詳細性の観点では懸念がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が低いため正確性の観点は懸念があるが、1つのベンチマーク指標と強い正の相関がみられることから、相関性の観点はおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、本業種は「不明」データが極端に少ないため、「中小企業のみ指標」と「中小企業及び不明を合算した指標」いずれのパターンでも相違ないといえる。

一方で、本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別・地域別指数を作成する際には、旅行等に伴うゴルフ場における消費等のノイズが含まれている可能性があるといった課題・懸念事項を確認できた。

表 2-25 ゴルフ場の有効性・妥当性評価

評価軸	評価観点	評価結果																																														
有効性	■ 速報性	<p>▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「特定サービス産業動態統計調査(当月の結果を、翌々月上旬に公表)」と比べても、<u>3週間程度の期間短縮が見込める</u></p>																																														
	■ 詳細性	<p>▶ 都道府県・地域によっては、消費者属性(性別、年齢区分など)ごとに十分なサンプル数が確保できていないため、<u>特定の都道府県別・地域別で指数が作成できない</u></p> <p>都道府県別指数を作成することができる都道府県数(全47)：35 地域別指数を作成することができる地域数(全10)：8</p>																																														
妥当性	■ 正確性	<p>▶ 1段階目の中小企業の正答率は下記のとおりであり、<u>1段階目の精度が低い</u>(※サンプリングチェックの対象は売上上位企業)</p> <ul style="list-style-type: none"> ✓ 1段階目の中小企業の正答率：75.0% ✓ 2段階目の中小企業の正答率：- ✓ 3段階目の中小企業割合：- <p>▶ 本業種は2段階目と3段階目において売上寄与度の大きい企業数は5つ以下であったため、2段階目と3段階目の中小企業の正答率・中小企業割合は算出していない。したがって、売上寄与度で重みづけした正答率は「中小企業のみ指標」「中小企業及び不明を合算した指標」いずれも75.0%であり、<u>精度に懸念が残る</u></p>																																														
	■ 相関性	<p>▶ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。<u>「特定サービス産業動態統計調査」と強い正の相関がみられた</u></p> <table border="1" style="margin: 10px auto;"> <thead> <tr> <th rowspan="2"></th> <th rowspan="2"></th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td colspan="8">(内は更新頻度)</td> </tr> <tr> <td rowspan="2">特定サービス産業動態統計調査(月次)</td> <td>corr</td> <td>0.92</td> <td>0.92</td> <td>0.91</td> <td></td> <td></td> <td></td> </tr> <tr> <td>RMSE</td> <td>0.08</td> <td>0.07</td> <td>0.09</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> </tbody> </table> <p>【corrにおける凡例】 強い正の相関 (>=0.7) 正の相関 (>=0.4) ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 誤差が小さい (<0.3) 誤差がやや大きい (>=0.3) 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶ 「前年比」のベンチマーク指標との符号一致率と「原数値(原指数)」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <ul style="list-style-type: none"> ✓ 「前年比」のベンチマーク指標との符号一致率 <table border="1" style="margin: 10px auto;"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>特定サービス産業動態統計調査</td> <td>65</td> <td>86%</td> <td>88%</td> <td>88%</td> </tr> </tbody> </table> <p>【凡例】 符号一致率が25%以下 符号一致率が50%以下</p>			相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	(内は更新頻度)								特定サービス産業動態統計調査(月次)	corr	0.92	0.92	0.91				RMSE	0.08	0.07	0.09	△	◎	◎	ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	特定サービス産業動態統計調査	65	86%	88%
		相関性			ベンチマーク指標の前提																																											
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																									
(内は更新頻度)																																																
特定サービス産業動態統計調査(月次)	corr	0.92	0.92	0.91																																												
	RMSE	0.08	0.07	0.09	△	◎	◎																																									
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																												
特定サービス産業動態統計調査	65	86%	88%	88%																																												

評価軸	評価観点	評価結果								
		<p>✓ <u>「原数値（原指数）」のベンチマーク指標との相関係数</u></p> <table border="1"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>特定サービス産業動態統計調査</td> <td>0.86</td> <td>0.88</td> <td>0.88</td> </tr> </tbody> </table> <p>【凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>➤ 上記の結果より、「前年比」及び「原数値（原指数）」では強い正の相関がみられ、かつ符号一致率が高いことから「<u>特定サービス産業動態統計調査</u>」との<u>相関性がある</u>と考えられる</p>	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	特定サービス産業動態統計調査	0.86	0.88	0.88
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)							
特定サービス産業動態統計調査	0.86	0.88	0.88							

c. 定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-26 に示す。感染症に入るポイントは適切に捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「特定サービス産業動態統計調査」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-27 に示す。

表 2-26 ゴルフ場の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	➤ 影響を受けていないことを確認できた
消費増税に入るポイント (2019年10月)	➤ -
感染症に入るポイント (2020年4月)	➤ 適切に捉えていることを確認できた
店舗増加などの影響	➤ 影響を受けていないことを確認できた
ベンチマーク指標と大きく違う動きをしているポイント	➤ 大きく違う動きをしているポイントがないことを確認できた



図 2-27 ゴルフ場の中小企業指数と
ベンチマーク指標（「特定サービス産業動態統計調査」）との比較結果

10) タクシー

a. 指数の推移

「タクシー」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-28 に示す。

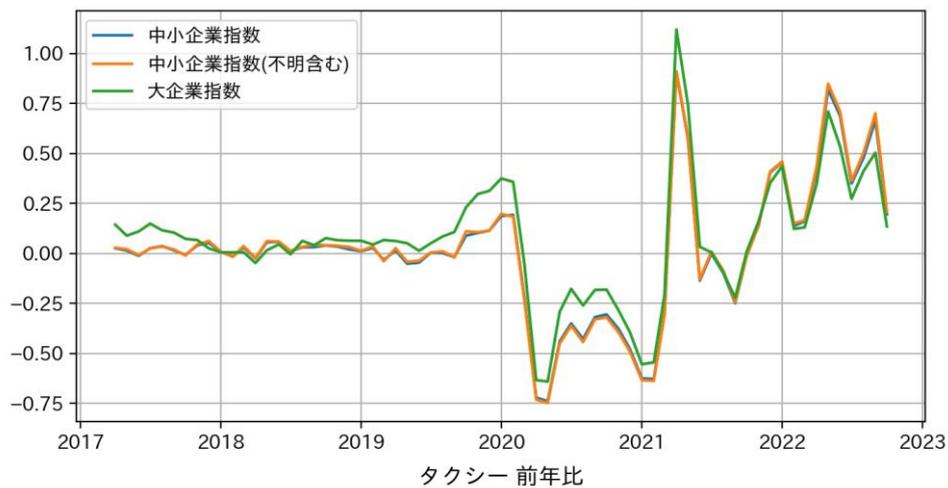


図 2-28 タクシーの指数の推移

b.有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、ベンチマーク指標と比較して3週間程度の期間短縮が見込める速報性はあるが、特定の都道府県別・地域別指数しか作成できないため詳細性の観点では懸念がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、1つのベンチマーク指標と強い正の相関がみられることから、正確性・相関性の観点のいずれもおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、本業種は「不明」データが極端に少ないため、「中小企業のみ指標」と「中小企業及び不明を合算した指標」いずれのパターンでも相違ないといえる。

一方で、本指標は企業の所在地ではなく消費者の居住地を使用しており、都道府県別・地域別指数を作成する際には、旅行等に伴うタクシーにおける消費等のノイズが含まれている可能性があることや、ベンチマーク指標との比較において、キャッシュレス化の進展などのアッパーバイアス（過大評価）といった課題・懸念事項を確認できた。

表 2-27 タクシーの有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できるため、比較的更新頻度が高い既存の「消費活動指数(当月の結果を、翌々月上旬に公表)」と比べても、 <u>3週間程度の期間短縮が見込める</u>
	■ 詳細性	▶ 都道府県・地域によっては、消費者属性(性別、年齢区分など)ごとに十分なサンプル数が確保できていないため、 <u>特定の都道府県別・地域別で指数が作成できない</u> 都道府県別指数を作成することができる都道府県数(全47): 21 地域別指数を作成することができる地域数(全10): 7
妥当性	■ 正確性	▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、 <u>1段階目と2段階目の精度は高い</u> (※サンプリングチェックの対象は売上上位企業)

評価軸	評価観点	評価結果																																															
		<ul style="list-style-type: none"> ✓ 1段階目の中小企業の正答率：88.9% ✓ 2段階目の中小企業の正答率：87.5% ✓ 3段階目の中小企業割合：- <p>▶本業種は3段階目において売上寄与度の大きい企業数は5つ以下であったため、3段階目の中小企業の正答率・中小企業割合は算出していない。したがって、売上寄与度で重みづけした正答率は「中小企業のみ指標」「中小企業及び不明を合算した指標」いずれも88.9%である</p>																																															
	■ 相関性	<p>▶ 「前年比」のベンチマーク指標との相関性の確認結果を以下に示す。「消費活動指数」と強い正の相関がみられた</p> <table border="1" style="margin-left: 20px;"> <thead> <tr> <th rowspan="2">()</th> <th rowspan="2">更新頻度</th> <th colspan="3">相関性</th> <th colspan="3">ベンチマーク指標の前提</th> </tr> <tr> <th>中小企業</th> <th>中小企業+不明</th> <th>大企業</th> <th>企業サイズ</th> <th>業種の粒度</th> <th>売上高指数</th> </tr> </thead> <tbody> <tr> <td rowspan="2">消費活動指数 (月次)</td> <td>corr</td> <td>0.93</td> <td>0.94</td> <td>0.87</td> <td>△</td> <td>◎</td> <td>◎</td> </tr> <tr> <td>RMSE</td> <td>0.1</td> <td>0.1</td> <td>0.14</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>【corrにおける凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>【RMSEにおける凡例】 ■ 誤差が小さい (<0.3) ■ 誤差がやや大きい (>=0.3) ■ 誤差が大きい (>=1.0)</p> <p>【前提における凡例】 ◎：対象が完全に一致している ○：完全一致はしていないが、対象の多くを含む △：対象を一部含む</p> <p>▶ 「前年比」のベンチマーク指標との符号一致率と「原数値（原指数）」のベンチマーク指標との相関係数を算出し、結果は下記のとおりである</p> <ul style="list-style-type: none"> ✓ 「前年比」のベンチマーク指標との符号一致率 <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>ベンチマーク</th> <th>データポイント</th> <th>符号一致率(中小企業)</th> <th>符号一致率(中小企業+不明)</th> <th>符号一致率(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>48</td> <td>69%</td> <td>69%</td> <td>65%</td> </tr> </tbody> </table> <p>【凡例】 ■ 符号一致率が25%以下 ■ 符号一致率が50%以下</p> <ul style="list-style-type: none"> ✓ 「原数値（原指数）」のベンチマーク指標との相関係数 <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>ベンチマーク</th> <th>相関係数(中小企業)</th> <th>相関係数(中小企業+不明)</th> <th>相関係数(大企業)</th> </tr> </thead> <tbody> <tr> <td>消費活動指数</td> <td>0.69</td> <td>0.69</td> <td>0.65</td> </tr> </tbody> </table> <p>【凡例】 ■ 強い正の相関 (>=0.7) ■ 正の相関 (>=0.4) ■ ほとんど相関がない (<0.4)</p> <p>▶ 上記の結果より、「前年比」では強い正の相関、「原数値（原指数）」では正の相関がみられ、かつおおむね符号一致率が高いことから「消費活動指数」との相関性があると考えられる</p>	()	更新頻度	相関性			ベンチマーク指標の前提			中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数	消費活動指数 (月次)	corr	0.93	0.94	0.87	△	◎	◎	RMSE	0.1	0.1	0.14				ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)	消費活動指数	48	69%	69%	65%	ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)	消費活動指数	0.69	0.69	0.65
()	更新頻度	相関性			ベンチマーク指標の前提																																												
		中小企業	中小企業+不明	大企業	企業サイズ	業種の粒度	売上高指数																																										
消費活動指数 (月次)	corr	0.93	0.94	0.87	△	◎	◎																																										
	RMSE	0.1	0.1	0.14																																													
ベンチマーク	データポイント	符号一致率(中小企業)	符号一致率(中小企業+不明)	符号一致率(大企業)																																													
消費活動指数	48	69%	69%	65%																																													
ベンチマーク	相関係数(中小企業)	相関係数(中小企業+不明)	相関係数(大企業)																																														
消費活動指数	0.69	0.69	0.65																																														

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-28 に示す。キャッシュレス還元の影響を受けているものの、感染症に入るポイントは適切に捉えている等を確認できた。なお、本業種の定性評価で比較したベンチマーク指標は「消費活動指数」であり、本業種の中小企業指数とベンチマーク指標との比較結果を図 2-29 に示す。

表 2-28 タクシーの定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	▶ 影響を受けていることを確認でき、懸念がある
消費増税に入るポイント (2019年10月)	▶ -
感染症に入るポイント (2020年4月)	▶ 適切に捉えていることを確認できた
店舗増加などの影響	▶ 影響を受けていないことを確認できた
ベンチマーク指標と大きく違う動きをしているポイント	▶ 大きく違う動きをしているポイントがないことを確認できた

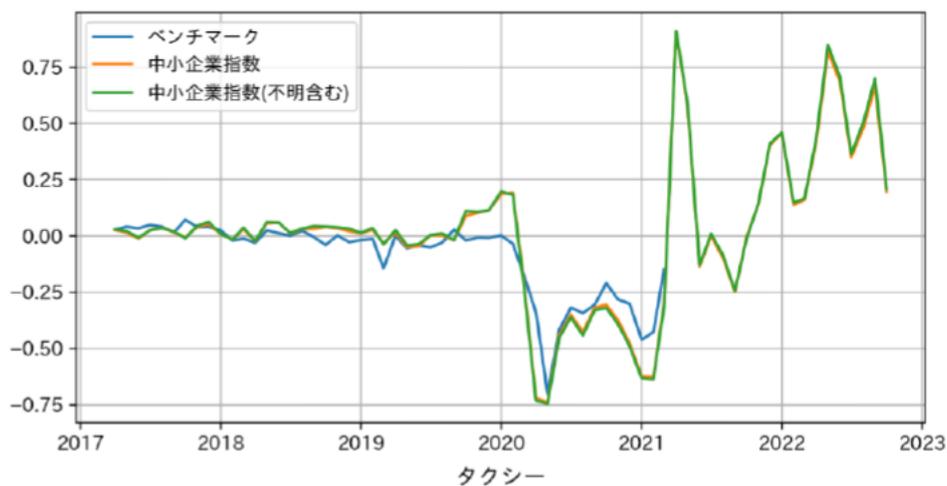


図 2-29 タクシーの中小企業指数とベンチマーク指標（「消費活動指数」）との比較結果

1 1) 酒屋

a. 指数の推移

「酒屋」においては「中小企業指数」と「中小企業指数（不明含む）」の2パターンを作成し、併せて「大企業指数」も作成した。試作した本業種における各指数の推移を図 2-30 に示す。

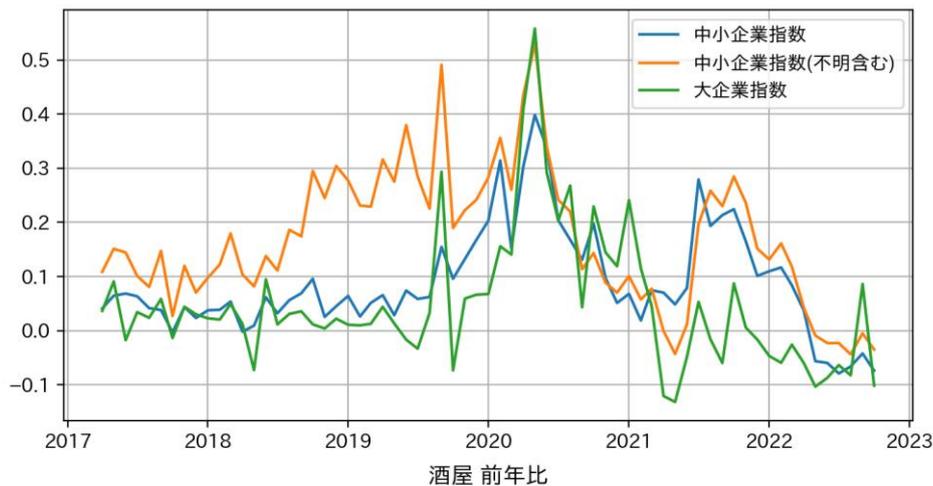


図 2-30 酒屋の指数の推移

b. 有効性・妥当性評価

本業種の中小企業指標の有効性に関しては、約2週間程度で算出できる速報性はあるが、特定の都道府県別・地域別指数しか作成できないため詳細性の観点では懸念がある。妥当性に関しては、サンプリングチェックの結果、中小企業の抽出精度が高いことを確認でき、相関性の観点はおおむね満たしている。したがって、一部懸念があるものの活用可能性がある指標と考えられる。

また、「大企業」もしくは「中小企業」のラベリングができない「不明」データを試作する指標に含めるべきか否かの判断について、正確性の評価結果から、「中小企業のみ指標」と「中小企業及び不明を合算した指標」の2パターンでは、「中小企業のみ指標」が望ましいといえる。

一方で、本指標はベンチマーク指標が存在しないため、既存の統計指標との関係性を示すことができないといった課題・懸念事項を確認できた。

表 2-29 酒屋の有効性・妥当性評価

評価軸	評価観点	評価結果
有効性	■ 速報性	▶ 当月の結果を、翌月15日(当月前半の結果は、当月末日)に公表できる
	■ 詳細性	▶ 都道府県・地域によっては、消費者属性（性別、年齢区分など）ごとに十分なサンプル数が確保できていないため、 <u>特定の都道府県別・地域別で指数が作成できない</u> 都道府県別指数を作成することができる都道府県数(全47)：30 地域別指数を作成することができる地域数(全10)：5
妥当性	■ 正確性	▶ 段階ごとの中小企業の正答率・中小企業割合は下記のとおりであり、 <u>いずれの段階においても精度は高い</u> （※サンプリングチェックの対象は売上上位企業） ✓ 1段階目の中小企業の正答率：93.3% ✓ 2段階目の中小企業の正答率：100.0% ✓ 3段階目の中小企業割合：80.0% ▶ 売上寄与度で重みづけした正答率は下記のとおりであり、 <u>「中小企業のみ指標」の精度が高い</u> ✓ 中小企業のみ指標： <u>95.2%</u> ✓ 中小企業及び不明を合算した指標：89.6%
	■ 相関性	▶ ベンチマーク指標が存在しないため、 <u>相関性評価は実施しない</u>

c.定性評価

本業種の中小企業指標に対する定性評価の結果を表 2-30 に示す。消費増税や感染症に入るポイントは適切に捉えていることを確認できた。

表 2-30 酒屋の定性評価

評価観点	評価結果
キャッシュレス還元の影響 (2019年終盤から2020年初頭)	▶ -
消費増税に入るポイント (2019年10月)	▶ 適切に捉えていることを確認できた

評価観点	評価結果
感染症に入るポイント (2020年4月)	▶ 適切に捉えていることを確認できた
店舗増加などの影響	▶ -
ベンチマーク指標と大きく違う動きをしているポイント	▶ - (※ベンチマーク指標が存在しないため、評価は実施しない)

2-3-2. 評価結果を踏まえた指標の課題

指標の評価や研究会の議論を踏まえ、指標活用にあたっての課題や懸念事項について整理した。

(1) 中小企業分類マスタの精度

正確性の評価において、EC やゴルフ場などでは適切に中小企業分を抽出できていないことが確認できており、一部の業種では中小企業分類マスタを活用したデータ分類の精度に懸念があると考えられる。

研究会の内容を踏まえ、本調査で試作した指標の中小企業割合が既存の統計調査等で公表されている実態と乖離していないかを確認するため、業種別に中小企業割合（企業数ベース・売上額ベース）を算出し、「平成 28 年 経済センサス」（以下、「経済センサス」とする。）における中小企業割合（企業数ベース・売上額ベース）¹⁷と比較した。

¹⁷ 企業数の割合は 2016 年時点、売上額は 2015 年時点である。

指標における中小企業割合と実態との整合性比較

【凡例】 ■：中小企業割合 ■：大企業割合 ■：不明割合
 本調査で試作した指標の業種 経済センサスにおける業種

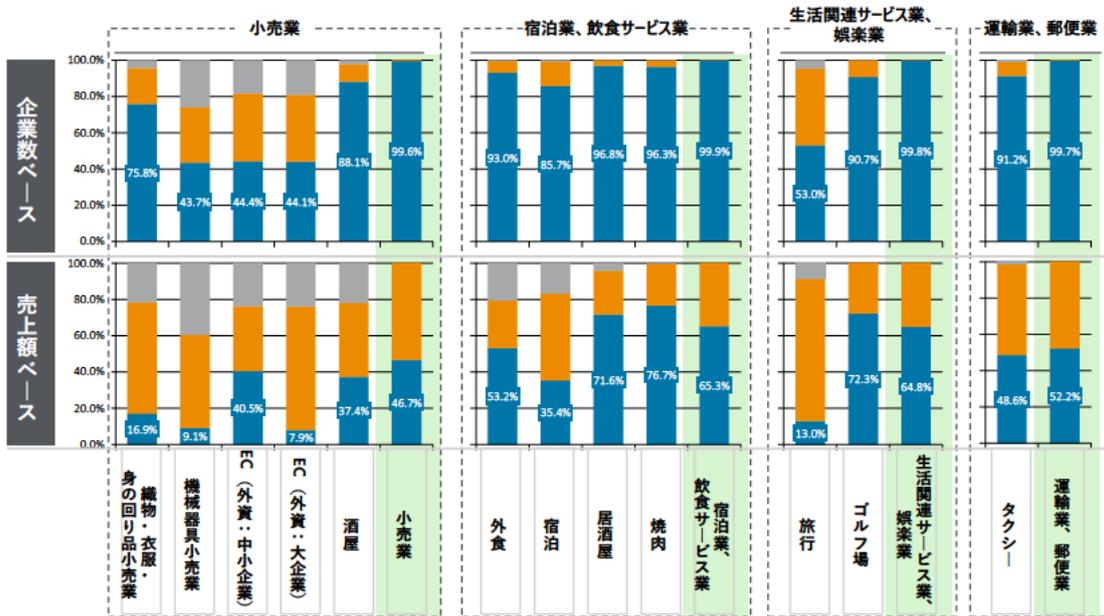


図 2-31 業種別の中小企業割合と

「平成 28 年 経済センサス」における中小企業割合の比較

その結果、業種粒度に違いがあるため一概に判断できないが、一部の業種において実態（「経済センサス」における中小企業割合）と乖離している可能性があることを確認できた。例えば、「経済センサス」における小売業の中小企業割合が、企業数ベースで 99.6%、売上額ベースで 46.7%であるのに対し、機械器具小売業における中小企業割合は、企業数ベースで 43.7%、売上額ベースで 9.1%となっている。乖離している原因を明確化し、可能な限り調整することが望ましいが、現在確認できている中小企業分類マスタを活用したデータ分類に係る課題は 2 つある。

1 つ目は、中小企業分類マスタが保有する「従業員数」と「資本金」それぞれのデータ精度を高めることである。中小企業分類マスタのデータは、様々なデータ提供元の内容を適宜正誤確認し、データ変更・更新されているが、登録されている「従業員数」と「資本金」データが正確ではないケースがある。

2 つ目は、STEP2 における突合精度が低いことである。STEP2 で突合できたのは、突合を試みた全体のうち約 18%であった。突合精度が低い要因として、「企業名」の記載の粒度が揃っていないことが考えられる。特に、外食においては、「企業名」に店舗名や運営企業名が混在しており、突合させることは容易ではない。

(2) 地域別指数の精度

詳細性（都道府県別・地域別の作成可能な指数の数）の評価において、地域の粒度では多くの業種・地域で地域別指数を作成できることが確認できたものの、十分にサンプル数が確保できない業種（例：旅行、酒屋）と地域（例：北陸、四国）も一部確認できた。地域別指数の作成にあたり最低限のサンプル数を確保できているか否かの業種別の確認結果を表 2-31 に示す。確認結果における評価基準（○or ×）の詳細は、「2-1-4.(1) 2) b 詳細性」における具体的な実施内容をご参照いただきたい。

表 2-31 地域別指数の作成にあたっての業種ごとのサンプル数の確認結果

	北海道	東北	南関東	北関東・甲信	北陸	東海	近畿	中国	四国	九州
織物・衣服・身の回り品 小売業	○	○	○	○	○	○	○	○	○	○
機械器具小売業	○	○	○	○	○	○	○	○	○	○
外食	○	○	○	○	○	○	○	○	×	○
旅行	○	×	○	×	×	○	○	×	×	○
宿泊	○	○	○	×	×	○	○	○	×	○
EC	○	○	○	○	○	○	○	○	○	○
酒屋	○	×	○	○	×	×	○	×	×	○
居酒屋	○	○	○	○	○	○	○	○	○	○
焼肉	○	○	○	○	×	○	○	○	×	○
ゴルフ場 ¹⁸	×	○	○	○	×	○	○	○	○	○
タクシー	○	○	○	×	×	○	○	○	×	○

また実際に、地域別指数を作成し指数の動きを確認したところ、都市圏では動きが安定している一方、人口の少ない地域では不安定な動き（例：特定のイベントがない状況で変動幅が大きい動き）を示した。一例として、タクシーの地域別指数の

¹⁸ 北海道と北陸においてサンプル数が確保できないのは、季節要因（冬場は閉場）の影響が考えられる。

結果を図 2-32 に示す。タクシーでは、全国的におおむね似た動きとなっているが、四国地域の指数は欠損があり、北関東・甲信地域と北陸地域の指数は不安定な動きであることが確認できた。

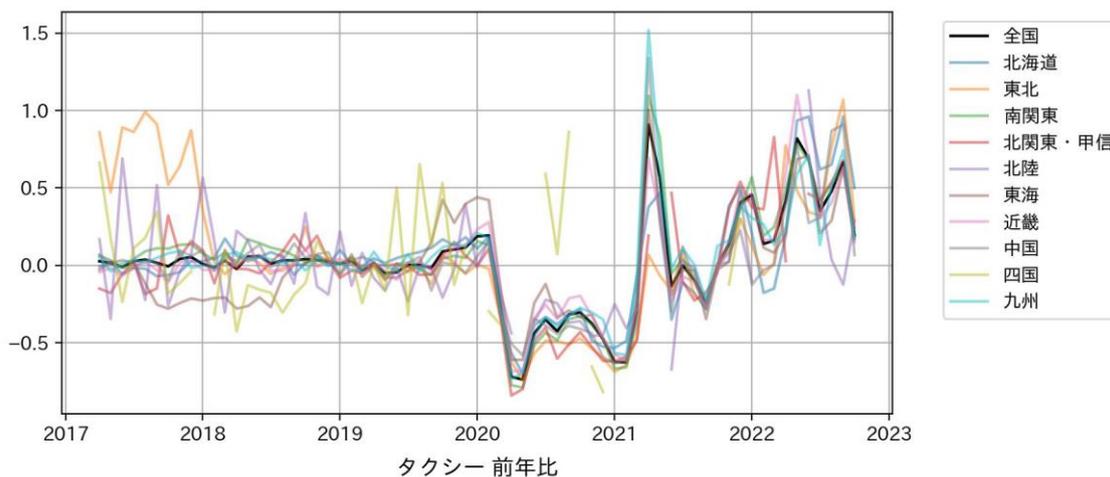


図 2-32 タクシーにおける地域別指数の結果

加えて、本指標は企業側（供給側）の所在地ではなく、消費者側（需要側）の居住地を基にした地域別指数であるため、宿泊や EC といった業種において、企業の所在地と消費者の居住地が異なっている可能性がある。したがって、企業の所在地を基にした地域別の売上高変化を捉えることは困難である。

(3) データの網羅性

本調査では、中小企業の売上動向を把握することを目的として、JCB カードのトランザクションデータを用いて試行的に指標を作成した。JCB カードは消費者の実取引データを基にしているため、商取引でもクレジットカード決済が多い小売業や外食業といった B to C の比重が大きい業種であれば、その業種の売上動向を捕捉することが期待できる一方、B to C 以外の業種では限界がある。

また、企業の売上に関連する決済手段としては、クレジットカードだけではなく、銀行送金や QR コード決済などの複数の手段があるため、これら複数のデータソースで補完して中小企業指標を作成することも有効な選択肢の 1 つである。

(4) 追加的なバイアス除去・データ補正の必要性

外食などの一部の業種におけるベンチマーク指標との比較において、キャッシュレス化の進展などの影響と考えられるアッパーバイアス（過大評価）が確認できた。このバイアスが存在することにより、特定の業種の売上動向が本来、「前年比」ベースで減少（マイナス）傾向であるのに対し、本指標における指数の動きでは増加（プラス）傾向となってしまう懸念がある。また、本調査ではJCBカード決済を導入している全企業を対象に中小企業指標を作成しているが、JCBカードの企業層や消費者を考慮したデータ補正を実施していない。中小企業全体を観測する場合、JCBカード決済を導入している企業の平均的な売上動向だけを把握していても、JCBカード決済を導入していないような比較的小さい企業の売上動向を把握できていないおそれがある。中小企業全体を一括りにせず、企業規模別により解像度を高めた分類を行い、傾向把握と適切なバイアス除去を実施する必要があると考えられる。加えて、銀行送金やQRコード決済などの複数のデータソースで補完することが困難である場合、業種ごとのクレジットカードの決済割合を調査し、クレジットカードデータの決済割合を考慮した適切な補正をする必要がある。

これらを踏まえ、業種別に中小企業の売上動向を示す指標の妥当性を高めるためには、追加的なバイアス除去・データ補正を行い、中小企業指標の精緻化を進めていくことが望ましい。

3. 事業者を取り巻く外部環境の構造化及びデータ利活用可能性に関する調査・分析

社会情勢や経済動向を把握するための重要な情報源と考えられる新聞などのテキストデータを活用することによって、定量的な指標だけでは把握困難な事象間の因果関係を明らかにし、示唆を得ることが期待できる。この点に関して、本章では、特定の経済・社会上の変動をトピックとして選定し、当該変動により連鎖的に発生する事象間の相互連関性や因果関係を捕捉するための実施内容について示す。具体的には、原因と結果の関係をネットワーク形式で可視化する因果型共起構造分析と、原因と結果の関係をチェーン形式で可視化する因果構造分析の2つの分析手法により仮説的な構造化を実施した。

3-1. 因果型共起構造分析

本分析の目的は、新聞のテキストデータから因果に関係する文章を抽出し、事象間の関係性（因果関係・共起関係）の理解に役立てることである。このような活用を想定する場合、同時期に発生している複数の事象間の多面的な因果関係をネットワーク形式で可視化することが有効と考えられる。本書では、このようなネットワーク表現を因果型共起構造と呼ぶこととする。因果型共起構造により、事業者の外部環境を構成する個々の諸要素の因果関係を理解することが期待できる。

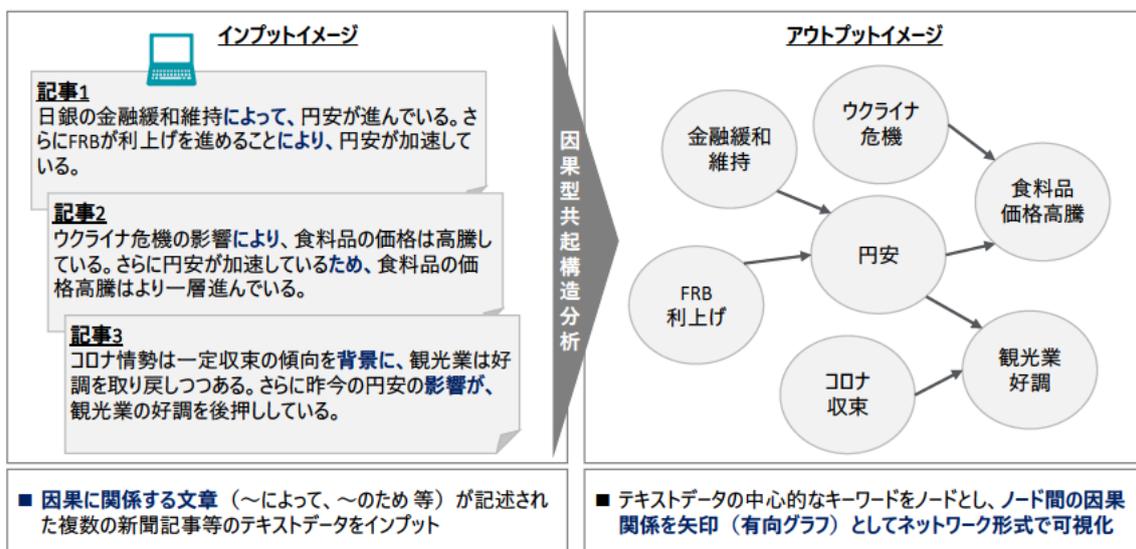


図 3-1 因果型共起構造分析のイメージ

3-1-1. 分析実施方針

(1) アプローチ

本分析の目的に鑑み、どのような因果文書の抽出方法が効果的であるか、因果の抽出に係る確認を行った。また、一般的な単語を基本単位として、共起関係に基づく分析によってネットワークを抽出する際、文章から得られる係り受け関係¹⁹が失われることから、時事の具体的事象間の因果関係を明示的に理解することが困難になると想定される。そのため、因果関係をどのように可視化すると解釈の容易さ（以下、「解釈性」という。）が高まるかといった観点で、ネットワークの可視化を実施した。本分析における解釈性は、テキストマイニングの専門的知見を有する学識者や中小企業庁職員、分析実施者による複数の視点からの定性評価により、「可視化したネットワークについて、人間が理解しやすい形式で提供されているか」を検討した。加えて、本分析における妥当性を「可視化したネットワークに対する解釈の容易さ」として定義した。本分析を進めるに当たっては、テキストマイニングの専門的知見を有する学識者に意見聴取を計5回行い、事業者の外部環境を構成する個々の諸要素を仮説的にネットワークとして構造化する際の検討に活用した。

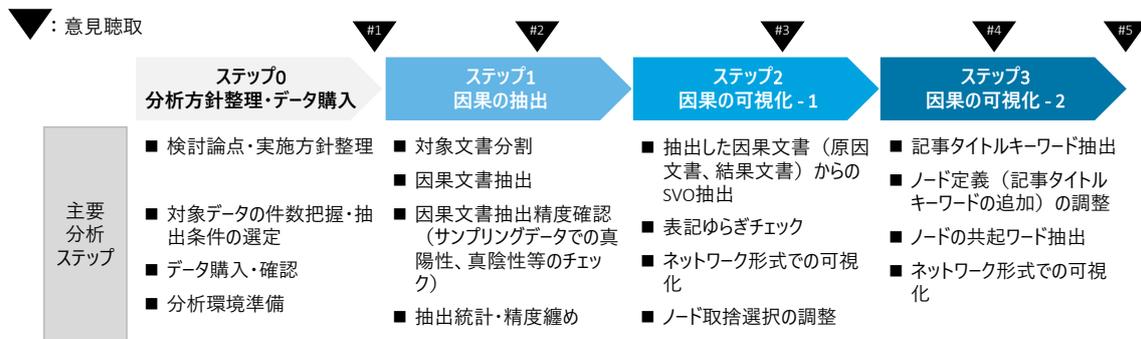


図 3-2 分析の進め方

¹⁹ 文中の言葉同士の文法上の関係性。

表 3-1 主要な検討論点と仮説

主要な検討論点		仮説
因果の抽出 どのような因果文書の抽出方法であれば抽出精度が高いか？	<ul style="list-style-type: none"> ■ 何を手がかりとして因果文書を抽出すると、精度高く抽出することができるか？ 	<ul style="list-style-type: none"> ■ 因果関係を表す文書の目印となる表現（例：～のため）を複数定義し、該当する文書を候補として抽出すると精度が高い ■ 因果関係文書の原因と結果の構文パターンを複数定義し、構文解析手法（係り受け解析）を活用して識別すると精度が高い <p>(実施内容)</p> <p><u>ステップ 1</u></p> <ul style="list-style-type: none"> ➤ 先行研究を参照し、下記の抽出方法で分析を実施する <ul style="list-style-type: none"> ✓ 和泉研究室において研究されてきた因果文書抽出ライブラリ（以下、「因果文書抽出ライブラリ」という。）²⁰を基に、因果情報を示す手がかり表現を目印として因果文書（候補）を抽出する ✓ Izumi et al.(2020)²¹で定義されている因果文書の構文パターンを活用することによって、原因部分と結果部分を識別する ➤ 因果文書抽出ライブラリの手がかり表現を活用し、精度・出現頻度のバランスが良い表現の取捨選択を行う（サンプリングチェックで精度確認を行う）
	<ul style="list-style-type: none"> ■ 因果文書の原因と結果の関係をどのように判断するか？ 	
因果の可視化 因果関係をどのように可視化すると解釈性が高いか？	<ul style="list-style-type: none"> ■ ノードはどの程度の単位（単語、単語群、簡易文書等）で可視化すると解釈性が高いか？ 	<ul style="list-style-type: none"> ■ 単語単位の可視化では解釈性が低く、単語群、ないしは簡易文書（意味チャンク）単位でノードを可視化すると解釈性が高い <p>(実施内容)</p> <p><u>ステップ 2</u></p> <ul style="list-style-type: none"> ➤ 因果文書の原因表現、結果表現のそれぞれをノードとする

²⁰ <https://github.com/tetsuwaka/CausalExtraction/tree/master/sample> (2022/11 時点)

²¹ Izumi, Kiyoshi, Shintaro Suda, and Hiroki Sakaji. "Economic News Impact Analysis Using Causal-Chain Search from Textual Data" In Proceedings of KDF2020: The AAAI-20 Workshop on Knowledge Discovery from Unstructured Data in Financial Services, 2020.

主要な検討論点		仮説	
		<p>▶ 石井他(2009)²²を基に、各ノードは SVOを基本構成単位として可視化する（※SVOが揃っていない場合でも除外せず、SOやVOノードとして定義する）</p>	
<p>■ 解釈可能な形式でネットワークの表現力を高めることができるか？</p>	<p>■ 動的な可視化ツールを活用することで、解釈性を担保しつつ一度に表現できる因果文書の件数を増やすことができる</p> <p>(実施内容)</p> <p><u>ステップ 2</u></p> <p>▶ 因果ネットワークの静的な可視化²³では解釈性に限界があるため、動的な可視化ツールを活用する</p>		
<p>■ どの範囲の文書をネットワーク的に可視化するか？</p>	<p>■ 因果に関係する文書のみを対象に可視化すると解釈性が高い</p> <p>(実施内容)</p> <p><u>ステップ 2</u></p> <p>▶ 原因表現、結果表現のノード化、ネットワーク化を行う</p>	<p>■ 相互連関性のある文書を対象に可視化すると解釈性が高い</p> <p>(実施内容)</p> <p><u>ステップ 3</u></p> <p>▶ 付加的に各ノードと共起度の高いキーワードを無向グラフで付与する</p>	
<p>■ 表現すべき因果文書をより重要度の高い文書に絞り込むことができるか？</p>	<p>■ 単語の重要度（TF-IDF等）によりノードを識別すると解釈性が高い</p> <p>(実施内容)</p> <p><u>ステップ 2</u></p> <p>▶ 下記の方法で絞り込みを実施する</p> <ul style="list-style-type: none"> ✓ 重要度の高い手がかり表現への絞り込み ✓ 重要度の高い文法構造への絞り込み ✓ 重要度の高い単語を含む因果文書への絞り込み ✓ ネットワーク上の重要度が高いノードへの絞り込み 		

²² 石井裕志, 馬強, & 吉川正俊. (2009). SVO 構造を用いた因果関係ネットワーク構築手法について. 研究報告データベースシステム (DBS), 2009(10), 1-8.

²³ 静的な可視化とは、因果ネットワークを視覚化形式で表現する際に、静止画像として表示されることを指す。

主要な検討論点		仮説	
	<ul style="list-style-type: none"> ■ 複数の因果文書間の因果関係をどのように定義・可視化するか？ 	<ul style="list-style-type: none"> ■ 同一ノード及びエッジとして表現される因果関係文書は、一意のノード・エッジとしてマーキングして可視化すると解釈性が高い <p>(実施内容)</p> <p>ステップ2</p> <ul style="list-style-type: none"> ➢ 各ノードのSVOが一致する場合に、同一ノードと見なしマーキングする 	<ul style="list-style-type: none"> ■ テキストデータのメタデータをノード定義に含めて可視化すると解釈性が高い <p>(実施内容)</p> <p>ステップ3</p> <ul style="list-style-type: none"> ➢ ノード定義に付加的な情報を付与し、ノードの個別性を高める

「因果の抽出」における具体的な実施内容

因果文書の抽出では、因果文書抽出ライブラリ²⁴を活用し、因果情報を示す手がかり表現を目印として因果文書（候補）を抽出した。また、Izumi et al.(2020)²⁵によると、因果文書の構文パターンを定義することによって、原因部分と結果部分を識別している研究実績があり、本分析においても Izumi et al.(2020)で定義されている複数の構文パターンを活用して、因果文書の抽出及び原因と結果の識別を実施した。構文パターンの種類を、表 3-2 に示す。

表 3-2 構文パターンの種類

構文パターン	手がかり表現を活用した原因表現と結果表現の因果関係の抽出パターン ※ 【】内は手がかり表現が含まれる文章における原因表現と結果表現の順序を示す
パターンA	➢ 原因表現が前の文にあり、手がかり表現の直後に結果表現が存在する形式【原因表現→手がかり表現→結果表現】
パターンB	➢ パターンAにおいて、結果表現の主語が特定できる形式【結果表現の主語→原因表現→手がかり表現→結果表現】
パターンC	➢ 結果表現が前の文にあり、原因表現が後の文にある形式【結果表現→原因表現→手がかり表現】

²⁴ 前掲。75頁。

²⁵ 前掲。75頁。

構文パターン	手がかり表現を活用した原因表現と結果表現の因果関係の抽出パターン ※【】内は手がかり表現が含まれる文章における原因表現と結果表現の順序を示す
パターンD	▶ パターンCにおいて、文章をまたぐ形式【1つ目の文章：結果表現、2つ目の文章：原因表現→手がかり表現】
パターンE	▶ パターンAにおいて、文章をまたぐ形式【1つ目の文章：原因表現、2つ目の文章：結果表現→手がかり表現】

「因果の可視化」における具体的な実施内容

因果文書の可視化では、単語レベルに留まらず日本語文法における SVO（主語、述語、目的語）に基づく文節レベルでの因果ネットワークを可視化した。対象としているテキストデータが膨大であることから、多量の因果文書の可視化において解釈性を高める工夫が必要と考えられる。そこで、因果ネットワークにおいて表現すべき因果文書については、因果ネットワークの表現力を高めることと重要度の高い文書に絞り込むことが両立できるように、段階的に2つの確認を実施した。「因果の可視化」における実施概要を表 3-3 に示す。

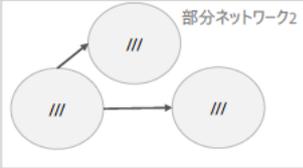
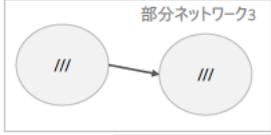
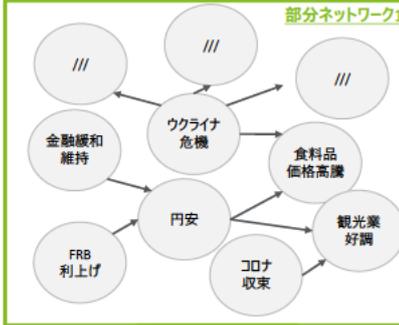
表 3-3 「因果の可視化」における実施概要

確認方法	概要
① 動的な可視化	▶ 因果ネットワークの表現力を高めるために、動的なネットワーク可視化ツール ²⁶ を活用し、動的な因果ネットワークの可視化を行った
② ノード絞り込み	<p>ア) 重要度の高い文書を絞り込むために、下記対応におけるノードの絞り込みを実施し、因果ネットワークを可視化した</p> <ul style="list-style-type: none"> ✓ 重要度の高い単語（TF-IDF²⁷のランキング上位キーワード）を含む因果文書に絞る ✓ 重要度の高い手がかり表現（で、による、のを除く手がかり表現）に絞る ✓ 重要度の高い文法構造（構文パターンD、パターンE²⁸）に絞る ✓ SVOを全て含む因果文書に絞る

²⁶ <https://pyvis.readthedocs.io/en/latest/index.html#>

²⁷ 対象文書内の単語の重要度（重み）を示す手法の1つ。

²⁸ 構文パターンにおける絞り込みにおいて、パターンDとパターンEを採用している理由は、いずれの構文パターンも複数の文書をまたいで記述されており、ノードのSVOが揃っている割合が高いためである。

確認方法	概要
	<p>イ) 因果ネットワーク上の重要度が高いノードに絞る。因果ネットワーク上の重要度が高いノードへの絞り込みの考え方は下図のとおり</p> <div data-bbox="555 436 1337 913" style="border: 1px solid black; padding: 5px;"> <p>絞り込みの考え方</p> <ul style="list-style-type: none"> ■ 多くのノードと繋がりを持つネットワークは、幅広い分野へ影響を与える因果構造を表していると考えられ、巨視的（マクロ）な観点で注目すべきと考える ■ 対象時事、関連トピックに関する記事から得られた因果ネットワーク全体の中から、ノード数の多い部分ネットワークを抽出する <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;"> <p>部分ネットワーク2</p>  <p>部分ネットワーク3</p>  <p>■ 一部のノードとのみ繋がっており、因果の影響範囲は限定的と考えられる</p> </div> <div style="text-align: center;"> <p>部分ネットワーク1</p>  <p>■ 多くのノードと繋がっており幅広い分野へ影響を与える</p> </div> </div> </div> <p>ウ) 特定キーワードに関連する記事に絞り込む際に、記事タイトルキーワードをノードに含めることで、ノードの個別性（因果ネットワークの繋がり の蓋然性）の向上が見込めるため、新聞記事テキストデータが保有する記事タイトルキーワードをノード定義に含める対応を実施した。また、特定事象間の同時発生的なネットワークを把握するために、記事タイトルキーワードを基に共起ネットワーク（共起構造）を可視化した</p>

(2) 妥当性や精度に関する評価

事業者の外部環境を構成する個々の諸要素を仮説的に構造化したネットワークに対して、その妥当性や精度を確認するために、本分析ではサンプリングチェックによる評価を実施した。

精度の評価に関しては、「因果の抽出」において、因果文書抽出ライブラリの手がかり表現を活用し、精度・出現頻度のバランスが良い表現の取捨選択を行いつつ、新たな手がかり表現を発見し、その精度・出現頻度を確認するために、目検によるサンプリングチェックを実施した。

妥当性の評価に関しては、「因果の可視化」において、因果ネットワークにおける複数のノード間の因果関係について、本分析の妥当性を、前述のとおり、解釈の容易さと定義し、テキストマイニングの専門的知見を有する学識者や中小企業庁職員、分析実施者による複数の視点から、確認を行った。

3-1-2. 分析対象データの選定

本分析では、特定の経済・社会上の変動と当該変動により連鎖的に発生する事象間の相互関連性や因果関係を把握することを目的としているため、広く一般的な経済事象や地方経済、企業活動の動向などを捕捉できるテキストデータを対象とすることが望ましいと考えられる。そこで、世界経済や日本経済に関する情報だけでなく、地方経済面も対象としている日本経済新聞社が提供しているサービス（以下、「新聞記事テキストデータ」という。）²⁹を選定した。なお、同サービスでは他紙である日刊工業新聞の記事データも提供しており、本分析においても分析対象とした。

(1) 分析対象データの内容

1) 選定トピック

本分析では、特定の経済・社会上の変動として「東日本大震災」、「感染症」、「円安」の計3つをトピックとして選定した。いずれのトピックも、過去に起きた事象であり、我が国の中小企業の経済活動に大きな影響をもたらした。これらの変動は将来的にも発生し得る可能性があり、テキストデータの政策利活用における有効性を検討する上で適当なトピックと考えられるため、本3トピックを対象とした。

2) 抽出条件

本分析を効果的、かつ効率的に行うために、手がかり表現や新聞の分類、検索キーワードと発行期間による絞り込みを実施し、分析対象とする新聞記事を抽出した。

手がかり表現による絞り込み

本分析で活用する因果文書抽出ライブラリの手がかり表現が少なくとも1つ以上含まれていることを抽出条件に付与した。本分析で活用した因果文書抽出ライブラ

²⁹ <https://nkbb.nikkei.co.jp/nikkei-corpus/>

りの手がかり表現の一覧を図 3-3 に示す³⁰。

手がかり表現			
を背景に	によって	による。	を受けております。
を背景に、	を反映し、	ため	によります。
を受け、	が響き、	ためだ。	によっております。
ため、	ため、	を受けて、	ためであります。
に伴う	を受けて	に伴い	によっています。
に伴い、	から、	ため。	このため、
を反映して	により	が響く	このため
をきっかけに	が響いた。	が響いている	そのため、
により、	ため」	が響いている。	そのため
に支えられて	が影響した。	で、	その結果、
			この結果、

図 3-3 本分析で活用した手がかり表現

新聞の分類による絞り込み

テキストマイニングによる分析を行う上で、本分析目的との直接的な関係性が低いと考えられる下記分類の記事は除外した。

【除外条件】

スポーツ記事、文化記事、数表記事、人事記事、訃報記事、社告・広告記事、インタビュー記事

検索キーワードと記事発行期間による絞り込み

「東日本大震災」においては、震災発生後1年間の記事発行期間を分析対象とした。「感染症」においては、検索キーワードを「新型コロナウイルス AND 影響」と設定した。「円安」においては、円安と一般的に関連性の高いキーワード（例：原材料・資源価格の高騰やサプライチェーンなど）を含めて抽出対象の検索条件とした。

表 3-4 選定トピックにおける検索キーワードと期間

³⁰ 2022年11月時点の因果文書抽出ライブラリから取得している。

トピック	検索キーワード	記事発行期間
東日本大震災	東日本大震災	2011.3.11～2012.3.11
感染症	新型コロナウイルス AND 影響	2020.1.1～2022.10.31
円安	円安 OR 物価 OR インフレ OR 半導体 OR 原油 OR エネルギー OR サプライチェーン OR 資源 OR 仕入 OR 調達 OR 生産 OR 価格 OR 販売 OR 中小	2022.3.1～2022.10.31

3) 記事件数

選定トピック及び抽出条件に基づき、新聞記事テキストデータは全体で 75,661 件であった。記事発行年ごとのトピック別の記事件数を図 3-4 に示す。なお、各トピックにおいて、それぞれの記事発行期間以外の記事発行年の記事も確認されたが、他トピックに該当する記事が抽出されたことが要因である。

件

記事発行年	記事全体*	トピック 東日本大震災	トピック 感染症	トピック 円安
2011年	19,579	18,864	0	10,882
2012年	2,751	2,698	0	1,580
2020年	16,030	422	16,029	9,346
2021年	9,539	177	9,539	5,862
2022年	27,762	322	4,670	26,127
総計	75,661	22,483	30,238	53,797

*複数のトピックに該当する記事があるため、記事全体の件数は一致しない

図 3-4 記事発行年別トピック別の記事件数

今回分析対象とした新聞記事テキストデータには、日本経済新聞の地方経済面と中小企業の動向把握に強い記事である日刊工業新聞も分析対象とした。なお、夕刊は朝刊などの内容とほぼ同様の記事となっている可能性があるため、分析対象から除外した。

件

記事種別	記事全体*	トピック 東日本大震災	トピック 感染症	トピック 円安
日刊工業新聞（小計）	22,468	5,223	8,028	18,863
日刊工業新聞	19,855	5,172	6,877	16,561
日刊工業新聞ニュースイッチ	2,613	51	1,151	2,302
日本経済新聞 地方経済面（小計）	22,646	8,278	9,883	13,434
日本経済新聞 地方経済面	20,560	7,293	8,983	12,855
日本経済新聞 大阪夕刊	932	299	469	411
日本経済新聞 大阪朝刊	395	120	254	56
日本経済新聞 名古屋朝刊	219	122	95	28
日本経済新聞 西部朝刊	138	76	59	25
日本経済新聞 名古屋夕刊	120	119	0	23
日本経済新聞 沖縄朝刊	100	86	8	9
日本経済新聞 西部夕刊	90	88	0	16
日本経済新聞 沖縄夕刊	65	63	0	8
日本経済新聞 北海道朝刊	27	12	15	3
日本経済新聞（小計）	30,547	8,982	12,327	21,500
日本経済新聞 朝刊	30,514	8,970	12,312	21,479
日本経済新聞 朝刊第2部	28	9	13	20
日本経済新聞 朝刊第3部	3	1	2	1
日本経済新聞 朝刊第4部	2	2	0	0
総計	75,661	22,483	30,238	53,797

*複数のトピックに該当する記事があるため、記事全体の件数は一致しない

図 3-5 記事種別 - トピック別の記事件数

3-1-3. 分析結果

(1) 因果の抽出

1) 因果文書の抽出結果

因果文書抽出ライブラリを活用し、新聞記事テキストデータから因果文書を抽出したところ、因果文書の抽出件数は全体で 187,817 件であった。手がかり表現の内訳では、「で」、「ため、」の手がかり表現が、全体の半数以上を占める結果³¹となった。また、構文パターンでは、「パターン A」「パターン B」が9割以上を占める結果となった。

³¹ 「で」が全体割合 46.0%であり、「ため、」が全体割合 10.8%である。

手がかり表現	記事種別 - 抽出した因果文書の件数				手がかり表現	記事種別 - 抽出した因果文書の件数			
	日刊工業新聞	地方経済面	日本経済新聞	総計		日刊工業新聞	地方経済面	日本経済新聞	総計
で、	26,322	24,524	35,623	86,469	をきっかけに	184	300	545	1,029
ため、	6,671	5,482	8,103	20,256	このため、	255	210	277	742
を受け、	4,299	2,352	3,278	9,929	が響き、	145	203	308	656
から、	2,954	2,737	3,621	9,312	が影響した。	113	325	192	630
により	4,816	1,658	2,487	8,961	その結果、	137	52	291	480
に伴う	2,913	1,911	3,419	8,243	が響く	65	119	280	464
によって	2,435	1,459	3,276	7,170	ためで、	67	119	260	446
により、	4,881	860	1,227	6,968	ため」	95	134	147	376
を受けて	1,427	1,435	2,262	5,124	そのため、	187	56	107	350
ためだ。	572	806	3,543	4,921	この結果、	47	24	149	220
を背景に	632	440	1,340	2,412	を反映して	50	36	85	171
に伴い	736	450	640	1,826	を反映し、	60	33	76	169
に伴い、	635	472	640	1,747	ためで	28	28	45	101
が響いた。	351	683	649	1,683	が響いている。	14	33	43	90
を受けて、	738	326	544	1,608	による。	30	16	28	74
そのため	631	176	608	1,415	に支えられて	23	24	23	70
このため	600	211	581	1,392	が響いている	7	9	8	24
を背景に、	350	271	552	1,173	を受けております。	1			1
ため。	341	343	431	1,115	総計	63,812	48,317	75,688	187,817

図 3-6 手がかり表現別 - 因果文書の抽出件数

構文パターン	記事種別 - 抽出した因果文書の件数 (件数割合)			
	日刊工業新聞	地方経済面	日本経済新聞	総計
A	46,051(72%)	35,945(74%)	53,800(71%)	135,796(72%)
B	14,433(23%)	9,338(19%)	14,755(19%)	38,526(21%)
C	250(0%)	293(1%)	1,012(1%)	1,555(1%)
D	1,221(2%)	2,012(4%)	4,108(5%)	7,341(4%)
E	1,857(3%)	729(2%)	2,013(3%)	4,599(2%)
総計	63,812(100%)	48,317(100%)	75,688(100%)	187,817(100%)

図 3-7 構文パターン別 - 因果文書の抽出件数 (件数割合)

2) サンプルチェック結果

a.因果文書抽出ライブラリの精度

因果文書抽出ライブラリの手がかり表現に対して、抽出結果の妥当性や精度を評価するために、サンプリングデータを対象に目検による因果文書抽出の精度を確認

した。因果文書抽出精度の評価指標として、正解率³²、適合率³³、再現率³⁴、f 値³⁵を算出した。また、新聞記事テキストデータにみられる特有の手がかり表現が存在する可能性もあるため、サンプリングデータの中から因果文書抽出ライブラリ以外の新たな手がかり表現も調査した。実施内容は、図 3-8 に示す。

- 因果文書抽出ライブラリによる手がかり表現を基に、**因果文書抽出精度**を確認
 - 確認対象記事数：100件（サンプリング抽出）
 - 検出した文書数：315文書数
 - 評価指標：正解率、適合率、再現率、f 値
- 上記の確認結果から、**新たな手がかり表現を抽出**
 - 偽陰性と判定した因果文書から新たな**16表現**を取得

図 3-8 因果文書抽出ライブラリを基にした因果抽出におけるサンプリングチェックの実施内容

因果文書抽出ライブラリを基にしたサンプリングチェックの結果、2つの要因から全体的な精度があまり高くないことを確認できた。1つ目の要因として、抽出件数が最も多い手がかり表現である「で」の偽陽性³⁶が高いことが挙げられる。2つ目の要因として、因果文書抽出ライブラリ以外の新たな手がかり表現が存在していることから、偽陰性³⁷が高いことが挙げられる。

		因果文書抽出ライブラリ		総計	【評価指標】
		因果文章を抽出できていない	因果文章を抽出できている		正解率： 44.4%
目検	因果文章ではない	11	138	149	適合率： 48.3%
	因果文章である	37	129	166	再現率： 77.7%
総計		48	267	315	f値： 59.6%

図 3-9 因果文書抽出ライブラリを基にしたサンプリングチェック結果

³² 因果文書抽出ライブラリによる抽出結果が正しい割合。

³³ 因果文書抽出ライブラリにより因果文章であるものを抽出した結果に対して、正しいものが含まれている割合。

³⁴ 目検により因果文章であるものに対して、正しいものが含まれている割合。

³⁵ 適合率と再現率の調和平均。

³⁶ 目検により因果文章ではないと判定したものに対して、因果文書抽出ライブラリで因果文章と判定されたものの文章数。

³⁷ 目検により因果文章であると判定したものに対して、因果文書抽出ライブラリで因果文章を抽出できていないものの文章数。

		因果文書抽出ライブラリ		総計
		因果文章を抽出できていない	因果文章を抽出できている	
目検	因果文章ではない	7	93	100
	因果文章である	2	28	30
総計		9	121	130

図 3-10 「で、」に係る因果文書抽出精度

b.新たな手がかり表現を追加した更新版の因果文書抽出ライブラリの精度

因果文書抽出ライブラリの因果文書抽出の精度があまり高くないことを受けて、新たに発見した手がかり表現を因果文書抽出ライブラリに追加し、更新版の因果文書抽出ライブラリを基にして再度因果文書を抽出し、サンプリングチェックを実施した。実施内容は、図 3-11 に示す。

■ 新たな手がかり表現を因果文書抽出ライブラリに追加し、因果文書抽出

➤ 処理実行

■ 上記の、因果文書抽出精度を確認

- 確認対象記事：100件（※「a.因果文書抽出ライブラリの精度」で確認した記事と同じ）
- 検出した文書数：413文書数
- 評価指標：正解率、適合率、再現率、f 値

図 3-11 更新版の因果文書抽出ライブラリを基にした因果抽出におけるサンプリングチェックの実施内容

追加した新たな手がかり表現の一覧を図 3-12 に示す。また、更新版の因果文書抽出ライブラリを活用して、新聞記事テキストデータから因果文書を抽出した結果を図 3-13 に示す。追加した新たな手がかり表現の内訳では、「による」、「の影響で」の手がかり表現が比較的多い抽出件数であった。

因果文書抽出ライブラリの手がかり表現				追加した手がかり表現	
を背景に	によって	による。	を受けております。	ので	に従い
を背景に、	を反映し、	ため	によります。	ので、	を機に
を受け、	が響き、	ためだ。	によっております。	の影響で	を機に、
ため、	ため、	を受けて、	ためであります。	の影響もあり	を手掛かりに
に伴う	を受けて	に伴い	によっています。	もあり、	これを教訓とし、
に伴い、	から、	ため。	このため、	が要因。	こうした影響もあり、
を反映して	により	が響く	このため	この結果から、	
をきっかけに	が響いた。	が響いている	そのため、	この背景には、	
により、	ため」	が響いている。	そのため	として、	
に支えられて	が影響した。	で、	その結果、	による	
			この結果、		

図 3-12 更新版の因果文書抽出ライブラリにおける手がかり表現一覧

手がかり表現	記事種別 - 抽出した因果文書の件数				手がかり表現	記事種別 - 抽出した因果文書の件数			
	日刊工業新聞	地方経済面	日本経済新聞	総計		日刊工業新聞	地方経済面	日本経済新聞	総計
で、	25,491	23,927	34,474	83,892	を機に	291	319	471	1,081
による	16,881	12,563	26,570	56,014	をきっかけに	184	300	545	1,029
ため、	6,671	5,482	8,103	20,256	このため、	255	210	277	742
の影響で	4,093	6,075	5,686	15,854	が響き、	145	203	308	656
を受け、	4,299	2,352	3,278	9,929	が影響した。	113	325	192	630
から、	2,954	2,737	3,621	9,312	その結果、	137	52	291	480
により	4,816	1,658	2,487	8,961	が響く	65	119	280	464
に伴う	2,913	1,911	3,419	8,243	ため、	67	119	260	446
によって	2,435	1,459	3,276	7,170	ため」	95	134	147	376
により、	4,881	860	1,227	6,968	を機に、	119	103	146	368
として、	2,510	1,593	2,374	6,477	そのため、	187	56	107	350
ので	1,489	1,594	2,676	5,759	この結果、	47	24	149	220
を受けて	1,427	1,435	2,262	5,124	を反映して	50	36	85	171
もあり、	1,275	1,618	2,102	4,995	を反映し、	60	33	76	169
ためだ。	572	806	3,543	4,921	ためで	28	28	45	101
ので、	789	566	1,110	2,465	に従い	34	18	48	100
を背景に	632	440	1,340	2,412	が響いている。	14	33	43	90
に伴い	736	450	640	1,826	が要因。	59	18	8	85
に伴い、	635	472	640	1,747	による。	30	16	28	74
が響いた。	351	683	649	1,683	に支えられて	23	24	23	70
を受けて、	738	326	544	1,608	を手掛かりに		5	47	52
そのため	631	176	608	1,415	が響いている	7	9	8	24
このため	600	211	581	1,392	こうした影響もあり、	1		1	2
を背景に、	350	271	552	1,173	を受けております。	1			1
ため。	341	343	431	1,115	総計	90,522	72,192	115,778	278,492

図 3-13 更新版の因果文書抽出ライブラリの手がかり表現別

- 因果文書の抽出件数

更新版の因果文書抽出ライブラリを基にしたサンプリングチェックの結果、再現率が高まったことを確認できた。また、追加した手がかり表現の中において、比較

的抽出件数が多い「による」と「ので」は偽陽性が高く、必ずしも因果文書精度が高くないことを確認できた。

		因果文書抽出ライブラリ		総計	【評価指標】
		因果文章を抽出できていない	因果文章を抽出できている		正解率： 47.2%
目検	因果文章ではない	11	198	209	適合率： 48.2%
	因果文章である	20	184	204	再現率： 90.2%
総計		31	382	413	f値： 62.8%

図 3-14 更新版の因果文書抽出ライブラリを基にしたサンプリングチェック結果

		因果文書抽出ライブラリ		総計
		因果文章を抽出できていない	因果文章を抽出できている	
目検	因果文章ではない	0	46	46
	因果文章である	0	16	16
総計		0	62	62

図 3-15 「による」に係る因果文書抽出精度

		因果文書抽出ライブラリ		総計
		因果文章を抽出できていない	因果文章を抽出できている	
目検	因果文章ではない	0	8	8
	因果文章である	1	2	3
総計		1	10	11

図 3-16 「ので」に係る因果文書抽出精度

3) 因果の抽出における考察まとめ

サンプリングチェック結果を踏まえ、因果文書抽出ライブラリにおいて標準的に準備されている手がかり表現を最大限活用しつつ、かつ新たな手がかり表現を追加し、精度が悪い表現である「で、」「による」「ので」を除去することで、正解率と適合率の向上が確認できた。つまり、因果文書抽出ライブラリに新たな手がかり表現を追加し、「で、」「による」「ので」の手がかり表現を除去した因果文書抽出ライブラリを活用することで、総合的な因果文書抽出精度が高まったといえる。新たな手がかり表現を追加し、「で、」「による」「ので」の手がかり表現を除去した場合のサンプリングチェック結果を図 3-17 に示す。

したがって、因果の可視化を進めていくに当たっては、新聞記事テキストデータ

の文書の特性に応じて手がかり表現を適切に更新することで、因果文書の抽出精度を高めることが可能である。

		因果文書抽出ライブラリ		総計	【評価指標】
		因果文章を抽出できていない	因果文章を抽出できている		正解率： 67.1%
目検	因果文章ではない	4	53	57	適合率： 72.4%
	因果文章である	17	139	156	再現率： 89.1%
総計		21	192	213	f値： 79.9%

図 3-17 新たな手がかり表現を追加し、「で」、「による」「ので」の手がかり表現を除去した因果文書抽出ライブラリを基にしたサンプリングチェック結果

(2) 因果の可視化

1) 分析結果の概要

「因果の可視化」において段階的に実施した2つの確認方法の結果概要を表 3-5 に示す。

- ① 【動的な可視化】静的な可視化³⁸では解釈性が高くないことを受け、動的なネットワーク可視化ツールを活用することで、一度に表現できる因果文書の件数を増やすことができ、因果ネットワークの表現力を高めることを確認できた。一方で、多数のノードが絡まり合う複雑な因果ネットワークであることから、可視化の解釈性という観点では限定的であった。この分析結果を受けて、ノードを絞り込んでいく必要があると考えられ、②ノード絞り込みを実施した。
- ② 【ノード絞り込み】まず、重要度の高い単語及び手がかり表現に絞り込むことで、トピックにおける主要な事象の把握や事象間の関係性について可視化することができ、因果ネットワークの解釈性を高めることを確認できた。次に、因果ネットワーク上の重要ノードに絞り込むことで、巨視的（マクロ）な視点では経済的影響度や社会的注目度が大きい事象（トピック）を確認できたが、微視的（ミクロ）な視点ではノード間の因果関係について、因果関係とは認識できない結果が一部確認できた。また、記事タイトルキーワードを基に共起ネットワーク（共起構造）を可視化した結果、重要度・頻出度の高い単語がノードとして表現された。

³⁸ 同上。76頁。

表 3-5 「因果の可視化」で実施した結果概要

確認方法	特徴や課題
①動的な可視化	<ul style="list-style-type: none"> ▶ <u>一度に表現できる因果文書の件数を増やすことができた</u>が、ノード数が多く、<u>可視化の解釈性は限定的</u>
②ノード絞り込み	<ul style="list-style-type: none"> ▶ 重要度の高い単語及び手がかり表現に絞り込むことで、<u>主要な事象の把握及び事象の関連性の解釈が可能</u> ▶ 因果ネットワーク上の重要度が高いノードに絞ることで、巨視的（マクロ）な視点では影響度の高いトピックや業種を確認できたが、<u>微視的（ミクロ）な視点では一部妥当ではない結果</u>となった（記事タイトルキーワードをノード定義に含める対応により解消）

2) 分析結果の詳細

a.①動的な可視化

動的な可視化を実施する前に、因果ネットワークを静的に可視化したところ、必ずしも解釈性が高くないことを確認できた。図 3-18 は「東日本大震災」トピックについて、2011 年 3 月の新聞記事テキストデータを対象に抽出した因果文書を 100 件（ランダム抽出）で因果ネットワークを静的に可視化した結果である。



図 3-19 因果ネットワーク（動的な可視化、「感染症」トピック、2020年3月の新聞記事テキストデータを対象）

b.②ノード絞り込み

ア) 文書の重要度での絞り込み

図 3-20 は重要度の高い文書を絞り込むために、段階的に抽出した結果である。まず、新聞記事テキストデータから月別に TF-IDF によって抽出した重要度の高い単語を含むノードへ絞り込むことで、注目すべきノードを厳選することが期待できる。次に因果文書としての精度の高い手がかり表現に絞り込むことで、正しく因果関係を記述したノードを厳選することが期待できる。さらに、因果文書が複数の文書にまたいで記述されている構文パターンであるパターン D とパターン E に絞り込むことで、より解釈性が高いノードを厳選することが期待できる。最後に、SVO による絞り込みについて、本分析では SVO をノードの基本単位と定義しているが、日本語の特性上 S や O が省略される文書も多く存在することを考慮し、SVO を全て明示的に含まない因果文書も除外せずに可視化の対象としている。そのため、SVO を全て含むノードに絞り込むことで、さらに文書としての表現力が高いノードへと厳選することが期待できる。実際に抽出した因果文書（原因文書、結果文書）において、SVO を全て明示的に含んでいるノードに限定した場合、ノード数は相当数厳選される結果となった。

抽出した全因果文書		より注目すべき因果へ	より精度の高い因果へ	より解釈性の高い因果へ	より文書として完全な因果へ
記事発行年	因果文書件数	TF-IDFによる絞り込み後の件数*	手がかり表現(で、による、のを除く)による絞り込み後の件数	構文パターン(D, E)による絞り込み後の件数	SVOによる絞り込み後の件数
2011年	61,083	2,038	1,174	116	12
2012年	8,776	153	78	3	0
2020年	69,227	2,571	1,581	91	6
2021年	45,644	1,892	1,022	61	6
2022年	93,762	1,553	673	90	11
総計	278,492	8,207	4,528	361	35

* 月別に原因ノード、もしくは結果ノードがTF-IDFで得られた上位10位の単語を含むノードに絞り込み

図 3-20 因果文書の絞り込み結果

図 3-21 は SVO が全て揃ったノードまでを構成単位として絞り込み、因果ノードを可視化した結果である。可視化するノードを徹底的に絞り込むことで、微視的（ミクロ）に重要なノードを探索し、注目すべき個別の時事情報や具体的な企業名・業種の抽出が期待できる。ただし、ノード数が少ないため連鎖するノードが少なく、巨視的（マクロ）な視点で全体的な波及効果を捉えることには限界がある。

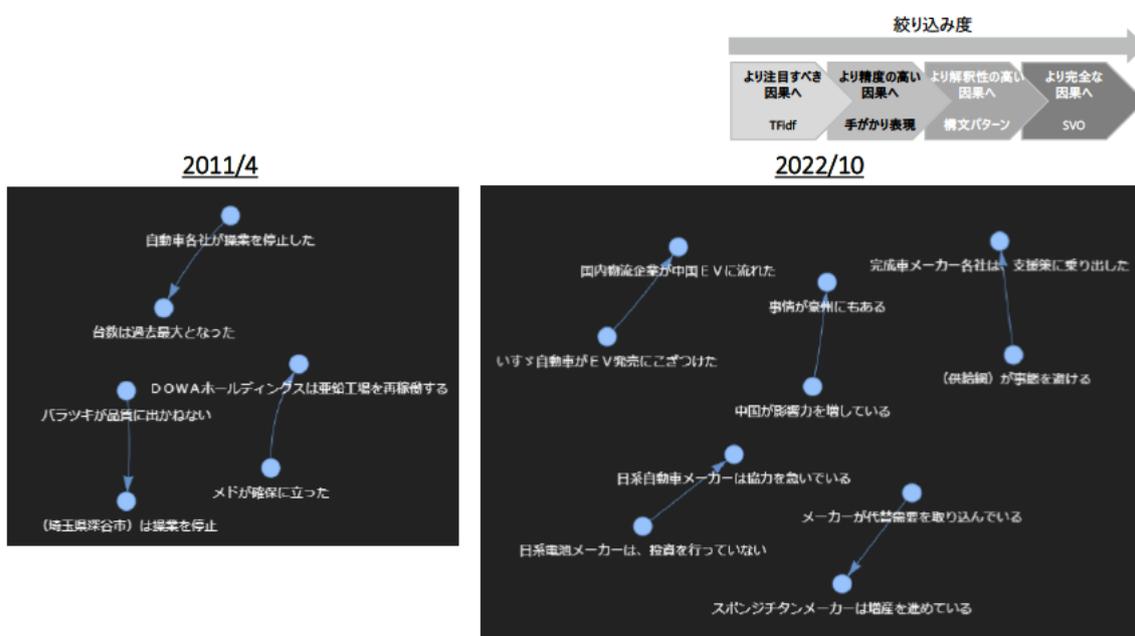


図 3-21 因果ノード例
(因果文書の絞り込み：TF-IDF+手がかり表現+構文パターン+SVO)

図 3-22 は重要度の高い単語及び手がかり表現に絞り込み、因果ノードを可視化した結果である。構文パターンやノードの構成単位を SVO とする絞り込みまでを実施しない場合、連鎖するノード数が多くなり、主要な事象の把握や事象間の関係性について可視化することが期待できる。

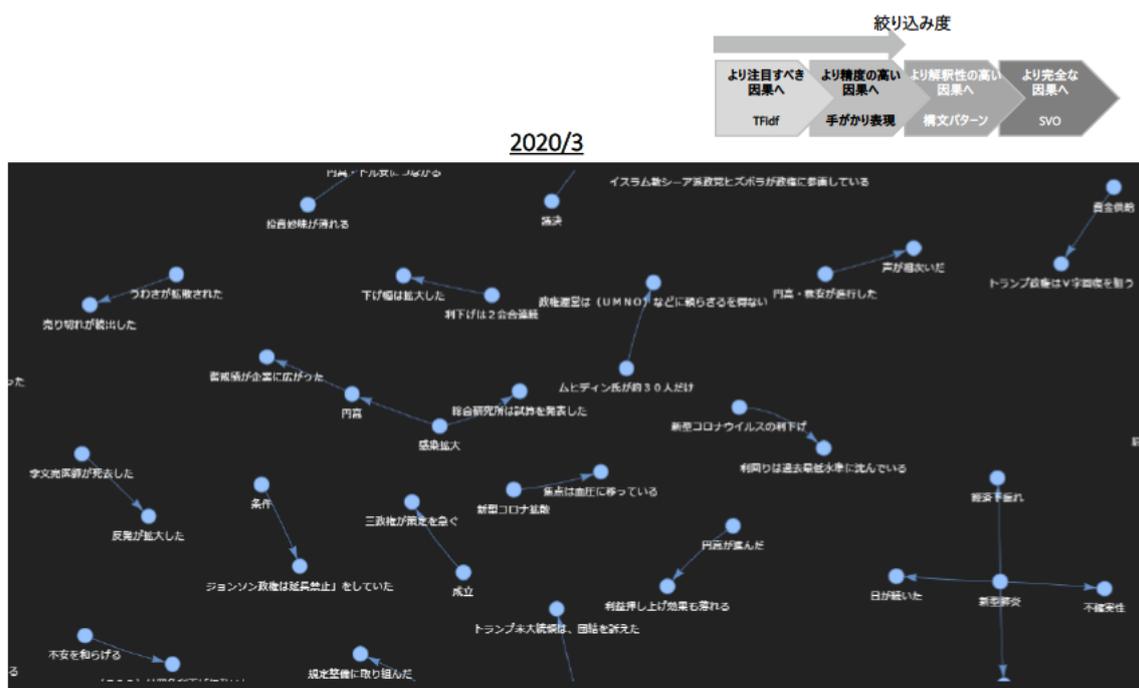


図 3-22 因果ノード例（因果文書の絞り込み：TF-IDF+手がかり表現）

イ) ネットワーク的重要度での絞り込み

因果ネットワーク上の重要ノードとして、連鎖するノードが多いネットワークは、幅広い分野への影響を与え得る因果構造を表現していると考えられるため、連結ノード数の多い部分ネットワークに注目することで、マクロな観点で重要な経済環境を捉えることが期待できる。2011年3月の新聞記事テキストデータを対象にした因果ネットワークから得られた部分ネットワークを図 3-23 に示す。東日本大震災を中心とする非常に広い影響範囲で関連するノード数が多い部分ネットワークが出現していることが確認できる。

部分ネットワークの ノード数	部分ネットワーク数
962 node	1
38 node	1
30 node	1
27 node	1
25 node	1
20 node	1
14 node	1
13 node	1
12 node	1
10 node	2
9 node	1
8 node	4
7 node	3
6 node	6
5 node	22
4 node	40
3 node	133
2 node	1,670
総計	1,890

*手がかり表現(で、による、ので)を除く

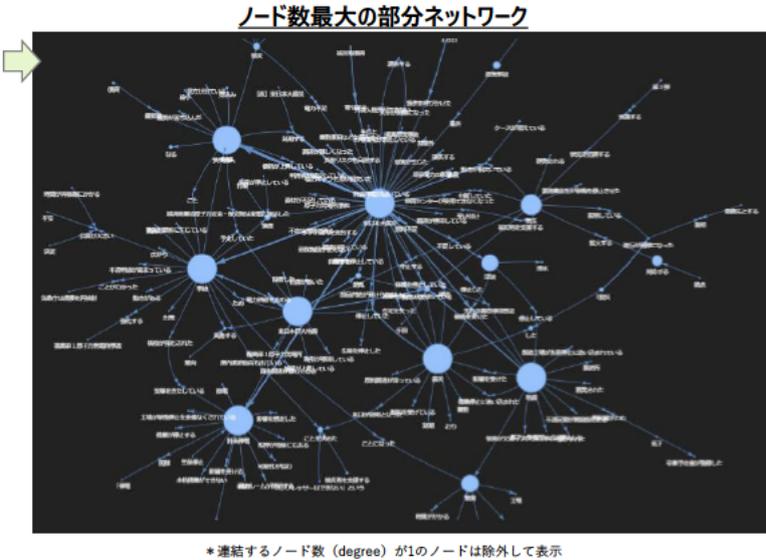


図 3-23 2011年3月の因果ネットワークから得られた部分ネットワーク

部分ネットワークのノード数が最大なネットワーク(以下、「最大部分ネットワーク」という。)のノード数の時系列推移を図 3-24 に示す。図 3-24 では、2011年3月や、2020年3月等の感染症拡大以降では、ノード数が最大の部分ネットワークの規模が、他の部分ネットワークと比較して非常に大きくなる事が分かる。

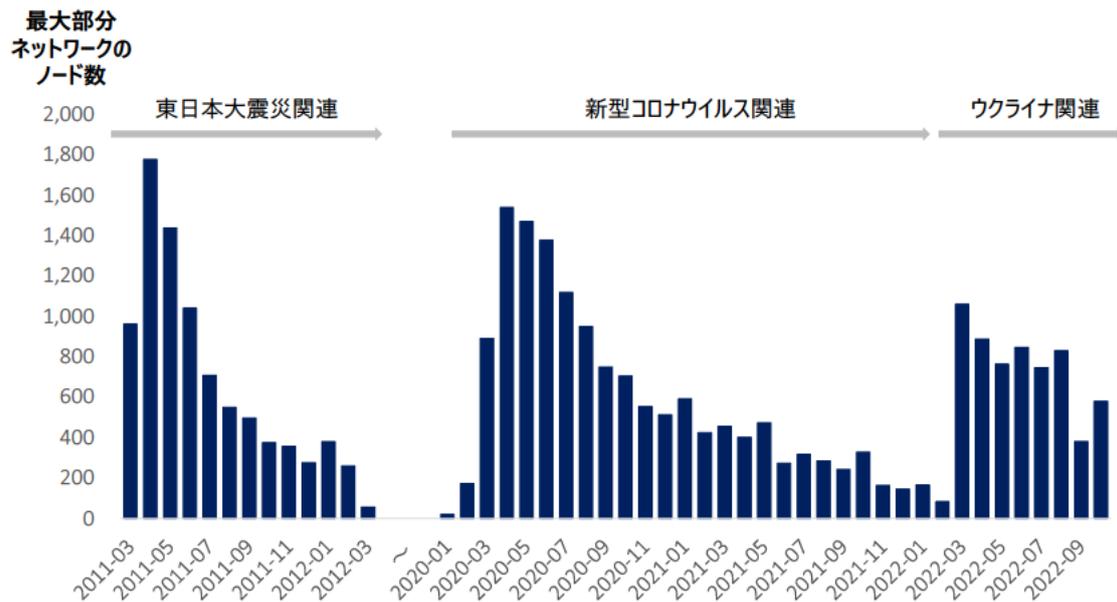


図 3-24 年別・月別における最大部分ネットワークのノード数の時系列推移

最大部分ネットワークの次数（エッジ数）最大のノードを表 3-6 に示す。各トピック（東日本大震災、感染症、円安）の重要度・注目度が上昇（下降）傾向なのかを時系列に捕捉できていることを確認できた。また、円安トピックにおいては、2022年3月以降に「ウクライナ侵攻」というトピックの背後にある注目すべき巨視的な因果関係の中心的なキーワードを特定することができた。

表 3-6 月別 - 最大部分ネットワークのノード数と次数最大ノード

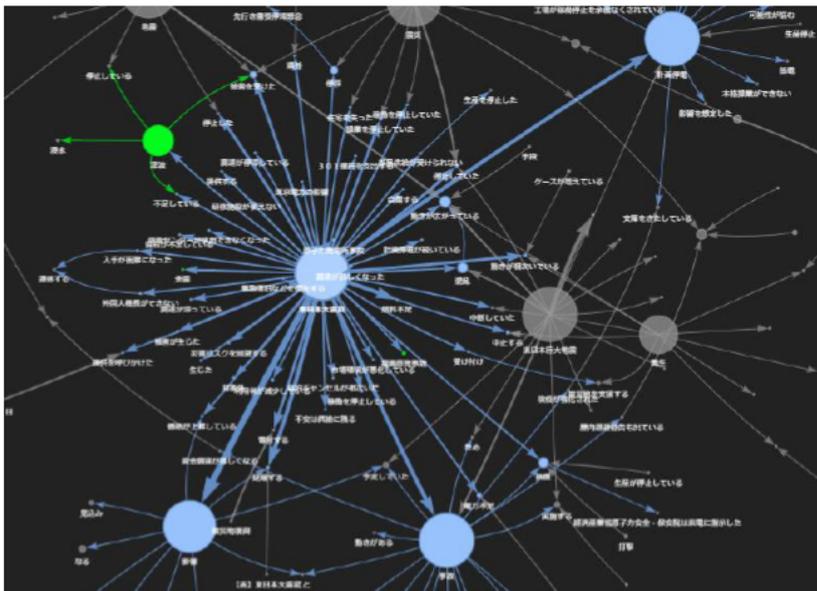
年月	最大部分ネットワークのノード数	最大部分ネットワークの次数最大のノード
2011-03	962	➤ 東日本大震災
2011-04	1,778	➤ 東日本大震災
2011-05	1,438	➤ 東日本大震災
2011-06	1,041	➤ 東日本大震災
2011-07	708	➤ 東日本大震災
2011-08	549	➤ 東日本大震災
2011-09	496	➤ 東日本大震災
2011-10	374	➤ 東日本大震災
2011-11	356	➤ 東日本大震災

年月	最大部分ネットワークのノード数	最大部分ネットワークの次数最大のノード
2011-12	274	➤ 東日本大震災
2012-01	379	➤ 東日本大震災
2012-02	259	➤ 東日本大震災
2012-03	55	➤ 東日本大震災
2020-01	19	➤ 感染拡大
2020-02	172	➤ 感染拡大
2020-03	891	➤ 感染拡大
2020-04	1,539	➤ 感染拡大
2020-05	1,470	➤ 感染拡大
2020-06	1,377	➤ 新型コロナウイルス
2020-07	1,119	➤ 新型コロナウイルス
2020-08	951	➤ 新型コロナウイルス
2020-09	749	➤ 新型コロナウイルス
2020-10	704	➤ 新型コロナウイルス
2020-11	553	➤ 新型コロナウイルス
2020-12	512	➤ 新型コロナウイルス
2021-01	591	➤ 新型コロナウイルス
2021-02	424	➤ 新型コロナウイルス
2021-03	456	➤ 新型コロナウイルス
2021-04	401	➤ 新型コロナウイルス
2021-05	472	➤ 新型コロナウイルス
2021-06	272	➤ 新型コロナウイルス
2021-07	317	➤ 新型コロナウイルス
2021-08	285	➤ 新型コロナウイルス
2021-09	242	➤ 新型コロナウイルス
2021-10	327	➤ 新型コロナウイルス
2021-11	163	➤ 影響
2021-12	145	➤ 新型コロナウイルス
2022-01	165	➤ 新型コロナウイルス
2022-02	82	➤ 新型コロナウイルス

年月	最大部分ネットワークのノード数	最大部分ネットワークの次数最大のノード
2022-03	1,060	➤ <u>ウクライナ侵攻</u>
2022-04	887	➤ <u>ウクライナ侵攻</u>
2022-05	763	➤ <u>ウクライナ侵攻</u>
2022-06	847	➤ <u>ウクライナ侵攻</u>
2022-07	746	➤ <u>ウクライナ侵攻</u>
2022-08	830	➤ <u>ウクライナ侵攻</u>
2022-09	380	➤ 高騰
2022-10	580	➤ <u>ウクライナ侵攻</u>

図 3-25 は「東日本大震災」トピックについて、因果文書の出現頻度をノードサイズやエッジ幅⁴⁰に反映させ、2011年3月時点の因果ネットワークを可視化した結果である。「東日本大震災」を契機として、「津波」や「計画停電」などの事象が波及的に起こり、それらの事象からさらに「工場の稼働停止」などの波及的影響について捕捉できていることを確認できた。つまり、巨視的（マクロ）に重要なノードを探索することができる。「津波」に関しては、「東日本大震災」トピックにおいて重要度の高い単語（TF-IDF のランキング上位キーワード）であり、本分析ではノードとエッジを緑色で着色した。巨視的な観点で重要度が高い最大部分ネットワークにおいて微視的に重要なノードに着色をすることで、因果ネットワークから外部環境を理解する際の解釈性はより高まると考えられる。

⁴⁰ 因果ノード間の矢印を指す。



*TF-IDFの上位ランクキーワードをGreenでハイライト

左記因果ネットワークの一部

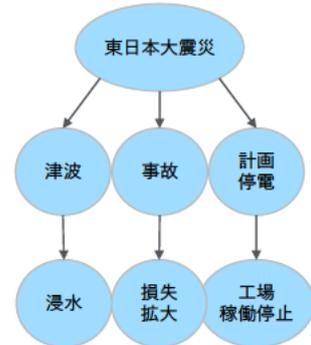


図 3-25 因果ネットワーク（ノード絞り込み（最大部分ネットワーク）、
「東日本大震災」トピック、2011年3月の新聞記事テキストデータを対象）と
因果ネットワーク抜粋

図 3-26 は「感染症」トピックについて、2020年3月時点の因果ネットワークを可視化した結果である。巨視的（マクロ）な視点では「感染拡大」による「休校措置」や「自粛要請」、「株価急落」といった経済的な影響度や社会的な注目度が大きい事象（トピック）を確認できた。一方で、2020年の「感染拡大」による「観光客の減少」と、2016年に発生した「熊本地震」による「観光客の減少」が同一のノードになっており、観光客の減少に繋がる一般的な因果関係としては妥当と考えられるものの、個別事象としての因果関係は無い事象が同一のノードとして表現されている結果が確認された。

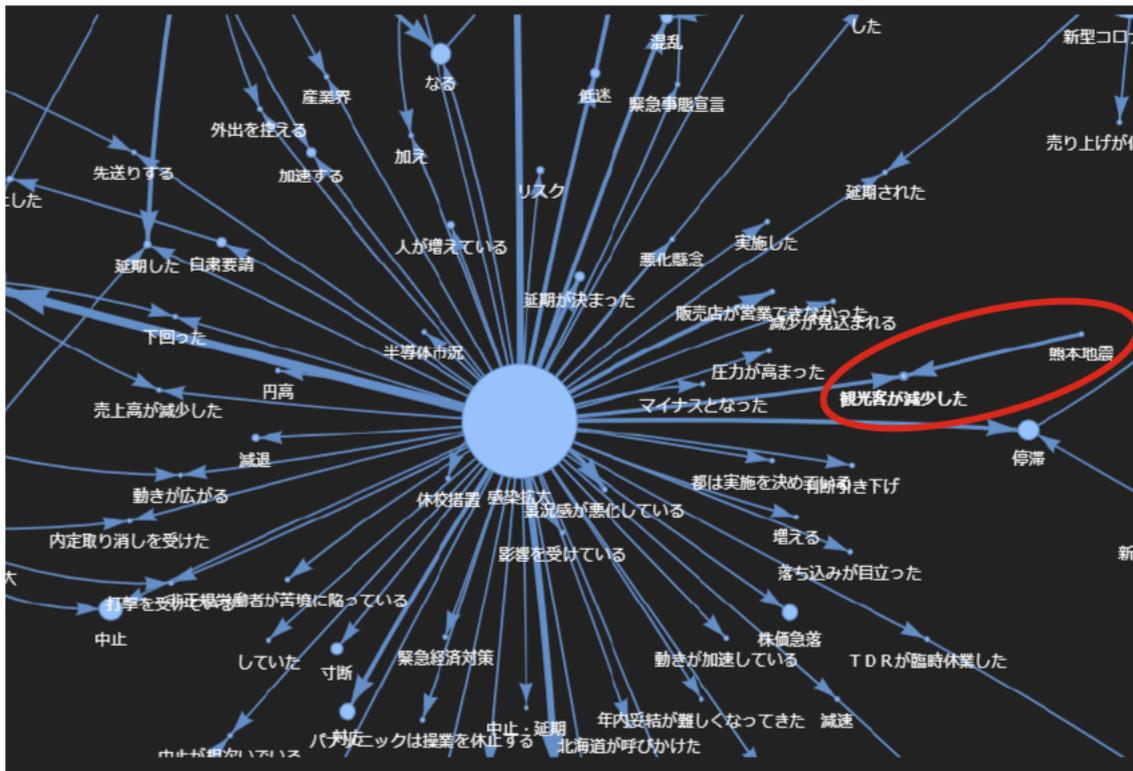


図 3-26 因果ネットワーク（ノード絞り込み（最大部分ネットワーク）、
「感染症」トピック、2020年3月の新聞記事テキストデータを対象）

ウ) 記事タイトルキーワードの活用による絞り込み

図 3-26 の結果を踏まえ、より個別性の高い因果関係を可視化する方法として、新聞記事テキストデータが保有する記事タイトルキーワードをノード定義に含めた結果が図 3-27 である。図 3-26 においてみられた、本来個別事象としての因果関係が無いにもかかわらず同一のノードとなっていたノードが適切に分離され、より明示的な因果関係に基づく因果ネットワークが得られたといえる。

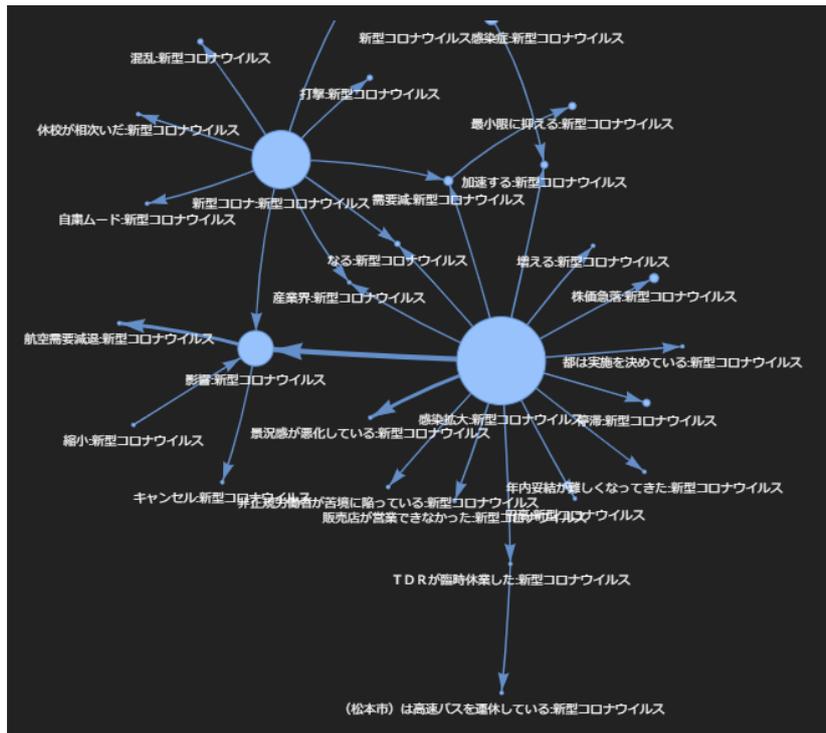


図 3-27 因果ネットワーク（ノード絞り込み（最大部分ネットワーク & 記事タイトルキーワードの活用）、「感染症」トピック、2020年3月の新聞記事テキストデータを対象）

記事タイトルキーワードをノード定義に含めることによって個別事象性は高まるが、部分ネットワークが細分化されてしまい、経済事象の巨視的な関係性に対する情報・示唆が限定的になってしまった。そのため、記事タイトルキーワードを基にした共起ネットワークによって、巨視的な関係性を補完しながら個別事象の高い部分ネットワークと総合的に活用することも一案と考えられる。図 3-28 は「感染症」トピックについて、2020年3月時点の共起ネットワーク（共起構造）を可視化した結果である。「感染症」から「工場」と「ストップ」といった重要度・頻出度の高い単語がノードを確認でき、事象間の同時発生的なネットワークを把握することができた。

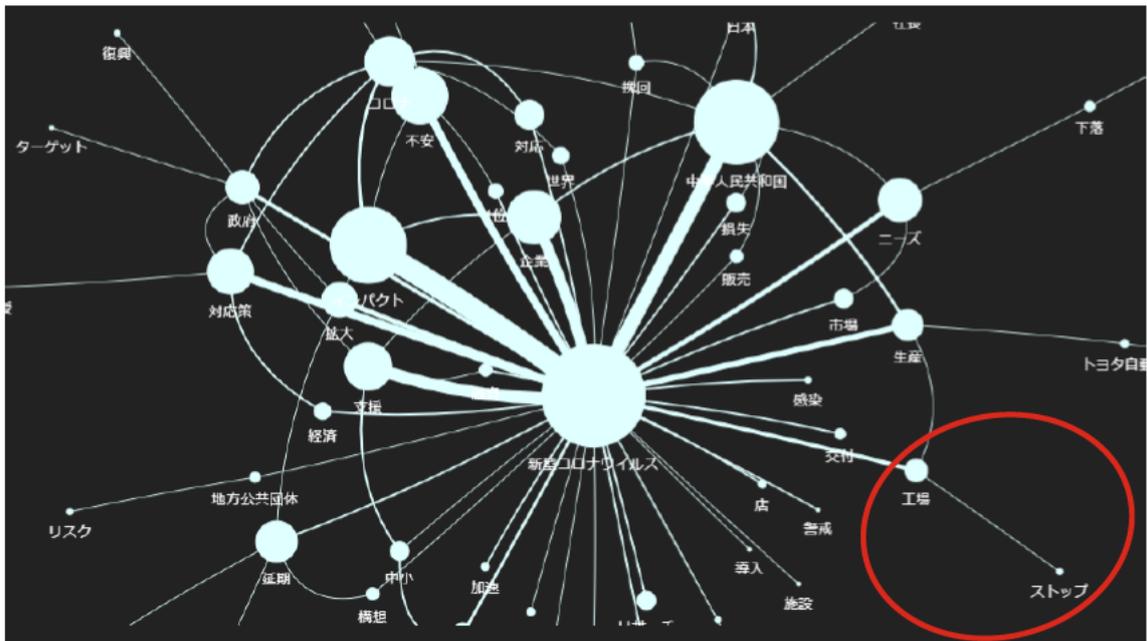


図 3-28 共起ネットワーク（ノード絞り込み（記事タイトルキーワードの活用）、
「感染症」トピック、2020年3月の新聞記事テキストデータを対象）

（3）因果型共起構造分析のまとめ

1) 得られた成果

本分析で得られた結果と示唆を表 3-7 に示す。本分析では、因果の巨視的關係性をネットワーク的に可視化し、時事の注目トピック（ウクライナ侵攻など）を抽出できた。また、特定の注目トピックからの波及的影響によって起こりうる個別事象を捉えることができた。

以上のことから、因果型共起構造分析により、波及的影響の可視化や事象間の因果関係の分析が可能であることが確認できたといえる。

表 3-7 主要な検討論点と確認結果・示唆

主要な検討論点		確認結果・示唆
因果の抽出 どのような因果文書の抽出	■ 何を手がかりとして因果文書を抽出すると、精度高く抽出することができるか？	➤ 因果文書抽出ライブラリの手がかり表現を基にした因果文書の抽出を実施した結果、1記事あたり2.5件の因果文書（合計187,817件）を抽出した

主要な検討論点	確認結果・示唆
方法であれば抽出精度が高いか？	<ul style="list-style-type: none"> ▶ サンプルチェック結果を踏まえ、新たな手がかり表現を追加し、適合率の低い「で、」「による」「ので」の手がかり表現を除去した因果文書抽出ライブラリは総合的な因果文書抽出精度が高い ▶ <u>対象となるテキストデータ特有の手がかり表現や、抽出精度の低い手がかり表現を識別し、手がかり表現の追加・除去を実施することで、因果文書の抽出精度を高めることが可能である</u>
因果の可視化 因果関係をどのように可視化すると解釈性が高いか？	<ul style="list-style-type: none"> ▶ 因果文書の原因表現、結果表現のそれぞれをノードとし、<u>文章として表現力が高いSVOを基本構成単位（※SVOが揃っていない場合でも除外せず、SOやVOノードとして定義する）として可視化することで、解釈性の高い可視化が実現できる</u>
■ 解釈可能な形式でネットワークの表現力を高めることができないか？	<ul style="list-style-type: none"> ▶ 因果ネットワークの静的な可視化では解釈性に限界があるが、<u>動的な可視化ツールを活用することで、多数のノードが絡まり合う複雑な因果関係の解釈を行う上で非常に有効である</u>
■ どの範囲の文書をネットワーク的に可視化するか？	<ul style="list-style-type: none"> ▶ <u>因果ネットワーク（有向グラフ）の方が共起ネットワーク（無向グラフ）と比べて、ネットワークが表現する情報量が多く、ネットワークの解釈性は基本的には高いと考えられる</u>
■ 表現すべき因果文書をより重要度の高い文書に絞り込むことができないか？	<ul style="list-style-type: none"> ▶ 重要度の高い単語（TF-IDFのランキング上位キーワード）を含む因果文書に絞ることで、時事の注目すべき特定事象の抽出へ活用できる可能性がある ▶ <u>因果ネットワーク上の重要ノードに絞り込むことで、巨視的（マクロ）な視点では経済的影響度や社会的注目度が大きい事象（トピック）や業種等を抽出することが可能となる</u>
■ 複数の因果文書間の因果関係をどのように定義・可視化するか？	<ul style="list-style-type: none"> ▶ （単一単語での一致や、文書類似度等ではなく）<u>因果文書のSVO単位で一致する場合に同一ノードと見なしているため因果関係の蓋然性は高く、解釈性が高い因果ネットワークの可視化が実現されていると考えられる</u>

2) 今後に向けた課題

本分析の遂行過程や結果を踏まえ、具体的な活用に向けては「因果文書の抽出精度の向上」が必要である。因果文書の抽出に当たって、本分析では因果文書抽出ライブラリを基に、新聞記事テキストデータの文書の特性に応じて手がかり表現を更新して、因果文書抽出を実施したが、抽出精度をより向上させていく余地があると考えられる。特に今回は、特定の新聞社のテキストデータのみを対象に実施したため、中小企業を取り巻く外部環境を幅広く捕捉する上では、抽出した因果文書の視点が偏っている可能性がある。

3-1-4. 構造化データの調査

前項にて、因果ネットワークを活用することで巨視的・微視的な観点で注目すべきキーワードや特定事象を抽出することが可能であることを示した。本項では、因果ネットワークのそのような特徴の発展的な活用例として、因果ネットワークを注目すべき調査対象の特定に活用した実践例を示す。

(1) 調査目的

新聞記事テキストデータを基に作成した外部環境の変化を、いち早く捕捉するためには経済動向を示す構造化された定量データの存在が不可欠である。外部環境の変化を定量的に把握する上で、過去の有事において影響が大きかった業種・経済動向に関して、現存する構造化データの有無・利用可能性を調査することとした。

(2) 調査対象

過去の有事において影響が大きかった業種・経済動向の選定に、因果ネットワークが活用可能である。新聞記事テキストデータが保有する業界タグ・経済テーマタグを活用し、前項で示した最大部分ネットワークにおいて最も影響が大きい業種・経済テーマを抽出することで、業種・経済テーマの特定を行った。各月の最大部分ネットワークに紐づく記事の業界タグ・経済テーマタグを集計し、最も割合の多い分類、最も上昇傾向⁴¹にある分類を調査対象とした。

東日本大震災関連では、図 3-29 に示すとおり、当初 2011 年 3 月は「電力・ガ

⁴¹ 当該月において、急速に上昇（増加）している業界タグ・経済テーマを指す。

ス」の割合が多く、次月 2011 年 4 月には「自動車・二輪車」業種の「経済活動」が最も割合が多く、かつ最上昇トレンドとなっていることを確認できた。

月	最大部分ネットワーク					
	ノード数	回数最大のノード	最も関連する業種	最も上昇傾向の業種	最も関連する経済テーマ	最も上昇傾向の経済テーマ
2011-03	962 node	東日本大震災	電力・ガス	—	マーケット	—
2011-04	1,778 node	東日本大震災	自動車・二輪車	自動車・二輪車	経済活動	経済活動
2011-05	1,438 node	東日本大震災	公的機関・大学	公的機関・大学	経済活動	輸出入動向
2011-06	1,041 node	東日本大震災	自動車・二輪車	自動車・二輪車	経済活動	景況感
2011-07	708 node	東日本大震災	自動車・二輪車	化学	経済活動	輸出入動向
2011-08	549 node	東日本大震災	公的機関・大学	電力・ガス	経済活動	マーケット
2011-09	496 node	東日本大震災	公的機関・大学	電機・家電	経済活動	生産・投資動向
2011-10	374 node	東日本大震災	自動車・二輪車	自動車・二輪車	経済活動	輸出入動向
2011-11	356 node	東日本大震災	公的機関・大学	公的機関・大学	経済活動	マーケット
2011-12	274 node	東日本大震災	公的機関・大学	文具・家具	経済活動	経済活動
2012-01	379 node	東日本大震災	公的機関・大学	自動車・二輪車	経済活動	生産・投資動向
2012-02	259 node	東日本大震災	公的機関・大学	新聞・放送・出版	経済活動	消費・住宅
2012-03	55 node	東日本大震災	自動車・二輪車	化学	—	—

図 3-29 「東日本大震災」トピックにおける最大部分ネットワークに最も関連する業種・経済テーマ

感染症関連では、図 3-30 に示すとおり、期間を通して「公的機関・大学」の割合が多いが、2020 年 6 月では「旅行・ホテル」業種の「景況感」が最上昇トレンドとなっていることを確認できた。

月	最大部分ネットワーク					
	ノード数	回数最大のノード	最も関連する業種	最も上昇傾向の業種	最も関連する経済テーマ	最も上昇傾向の経済テーマ
2020-01	19 node	感染拡大	鉄道・バス・タクシー	鉄道・バス・タクシー	—	—
2020-02	172 node	感染拡大	公的機関・大学	公的機関・大学	経済活動	—
2020-03	891 node	感染拡大	自動車・二輪車	自動車・二輪車	経済活動	マーケット
2020-04	1,539 node	感染拡大	公的機関・大学	公的機関・大学	経済活動	経済活動
2020-05	1,470 node	感染拡大	公的機関・大学	公的機関・大学	経済活動	GDP
2020-06	1,377 node	新型コロナウイルス	公的機関・大学	旅行・ホテル	経済活動	景況感
2020-07	1,119 node	新型コロナウイルス	公的機関・大学	公的機関・大学	経済活動	景況感
2020-08	951 node	新型コロナウイルス	公的機関・大学	鉄道・バス・タクシー	経済活動	雇用動向
2020-09	749 node	新型コロナウイルス	公的機関・大学	公的機関・大学	経済活動	景況感
2020-10	704 node	新型コロナウイルス	公的機関・大学	自動車・二輪車	経済活動	雇用動向
2020-11	553 node	新型コロナウイルス	公的機関・大学	鉄道・バス・タクシー	経済活動	生産・投資動向
2020-12	512 node	新型コロナウイルス	公的機関・大学	銀行・信用金庫	経済活動	雇用動向

図 3-30 「感染症」トピックにおける最大部分ネットワークに最も関連する業種・経済テーマ

円安関連では、図 3-31 に示すとおり、ウクライナ侵攻が発生した当初 2022 年 3 月において、「石油・鉱業・エネルギー」業種の「マーケット」が最上昇トレンドとなっていることを確認できた。

月	最大部分ネットワーク					
	ノード数	次数最大のノード	最も関連する業種	最も上昇傾向の業種	最も関連する経済テーマ	最も上昇傾向の経済テーマ
2022-01	165 node	新型コロナウイルス	物流・陸運・海運	物流・陸運・海運	経済活動	景況感
2022-02	82 node	新型コロナウイルス	公的機関・大学	ネット事業	経済活動	雇用動向
2022-03	1,060 node	ウクライナ侵攻	公的機関・大学	石油・鉱業・エネルギー	経済活動	マーケット
2022-04	887 node	ウクライナ侵攻	公的機関・大学	公的機関・大学	経済活動	物価
2022-05	763 node	ウクライナ侵攻	自動車・二輪車	非鉄金属・金属製品	経済活動	経済活動
2022-06	847 node	ウクライナ侵攻	自動車・二輪車	電子部品・電池	経済活動	景況感
2022-07	746 node	ウクライナ侵攻	公的機関・大学	公的機関・大学	経済活動	GDP
2022-08	830 node	ウクライナ侵攻	自動車・二輪車	自動車・二輪車	経済活動	金利・債券
2022-09	380 node	高騰	非鉄金属・金属製品	加工食品	経済活動	物価
2022-10	580 node	ウクライナ侵攻	公的機関・大学	電力・ガス	経済活動	景況感

図 3-31 「円安」トピックにおける最大部分ネットワークに最も関連する業種・経済テーマ

上記を踏まえ、各トピックにおける調査対象を表 3-8 に示す。

表 3-8 構造化データ調査対象

トピック	構造化データ調査対象	
	業種	経済指数
東日本大震災	自動車・二輪車	経済活動
感染症	旅行・ホテル	景況感
円安	石油・鉱業・エネルギー	マーケット

(3) 調査結果

実際に選定した業種・経済テーマの構造化データを調査した結果を下記に示す。

調査した構造化データについて、一部非公開（有料）のデータ⁴²が含まれているものの、公開されているものが大半である。これらの過去の有事と関連度の高いデータに対して、より速報性の高い指標開発や公的可用性の向上に取り組むことで、より一層の機動的な政策検討が期待できる。

1) 自動車・二輪車業界

自動車・二輪車業界の経済活動に関する構造化データを表 3-9 に示す。各指標について、中小企業の景況を直接的に把握する指標ではないものの、自動車業界や鉄鋼業界の中小企業の経済活動に大きな影響を及ぼす先行指標だと考えられる。

表 3-9 自動車・二輪車業界の経済活動に関する構造化データ

公表主体	業種	指標名【更新頻度】
経済産業省	自動車・二輪車	▶生産動態統計【月次】
一般社団法人 日本自動車工業会 (jama)	自動車・二輪車	▶自工会データベース【月次】
一般社団法人 日本自動車販売協会連合会	自動車	▶車種別販売台数【月次】 ▶メーカー別販売台数【月次】 ▶ブランド別新車販売台数【月次】 ▶大中型貨物車メーカー別販売台数【月次】 ▶RVタイプ別販売台数【月次】 ▶燃料別販売台数（乗用車）【月次】 ▶乗用車ブランド通称名別順位【月次】 ▶中古車・月別登録台数【月次】
一般社団法人 全国軽自動車協会連合会	自動車	▶新車販売台数【月次】 ▶中古車販売台数【月次】 ▶保有台数【月次】
	二輪車	▶新車販売台数【月次】 ▶中古車販売台数【月次】 ▶保有台数【月次】

⁴² 旅行・ホテル業界の景況感に関する構造化データにおいて、「メトロエンジン株式会社」及び「株式会社 Agoop」、「ノウキャスト社」、「株式会社 hands」が提供する指標が該当する。

公表主体	業種	指標名【更新頻度】
一般財団法人 自動車 検査登録情報協会	自動車・二 輪車	<ul style="list-style-type: none"> ➤ 自動車保有台数【月次/年次】
日本自動車輸入組合 (JAIA)	自動車	<ul style="list-style-type: none"> ➤ 輸入車新車登録台数速報【月次】 ➤ 輸入車中古車登録台数速報【月次】 ➤ 推移（※輸入車の販売台数（登録台数）の推移）【月次】 ➤ 輸入車統計データベース【月次】
国際自動車工業連合 (OICA)	自動車	<ul style="list-style-type: none"> ➤ Sales Statistics - Passengers Cars【年次】 ➤ Sales Statistics - Commercial vehicles【年次】 ➤ Sales Statistics - All vehicles【年次】 ➤ Production Statistics - passenger cars【年次】 ➤ Production Statistics - light commercial vehicles【年次】 ➤ Production Statistics - heavy trucks【年次】 ➤ Production Statistics - buses & coaches【年次】
国土交通省 ⁴³	自動車・二 輪車	<ul style="list-style-type: none"> ➤ 自動車関係統計 - 自動車輸送統計調査【月次】 ➤ 自動車関係統計 - 自動車燃料消費量調査【月次】 ➤ 自動車関係統計 - トラック輸送情報【月次】 ➤ 自動車関係統計 - 交通関連統計資料集（※2023/2/24時点では非公開）【更新無し】 ➤ 自動車関係統計 - 陸運統計要覧（※2023/2/24時点では非公開）【更新無し】 ➤ 自動車関係統計 - 幹線バス旅客流動実態調査（※平成22年、27年の調査結果が公開）【更新無し】 ➤ 自動車関係情報・データ - 自動車運送事業用自動車事故統計年報【年次】 ➤ 自動車関係情報・データ - バス事業者数【年次】 ➤ 自動車関係情報・データ - バスの車両数、輸送人員及び走行キロ【年次】 ➤ 自動車関係情報・データ - 高速バスの運行状況【年次】

⁴³ 国土交通省が提供している指標は、経済活動に直接的には関係しない指標も一部あるものの、自動車に関連する指標として公開されている指標を一覧として整理している。

公表主体	業種	指標名【更新頻度】
		<ul style="list-style-type: none"> ▶ 自動車関係情報・データ - ハイヤー・タクシー事業者数【年次】 ▶ 自動車関係情報・データ - ハイヤー・タクシーの車両数と輸送人員【年次】 ▶ 自動車関係情報・データ - 貨物自動車運送事業 車両数（運輸局・支局別）【年次】 ▶ 自動車関係情報・データ - 貨物自動車運送事業者数（規模別）【年次】 ▶ 自動車関係情報・データ - 貨物自動車運送事業者数（推移）【年次】 ▶ 自動車関係情報・データ - 貨物自動車運送事業者数（運輸局・支局別）【年次】 ▶ 自動車関係情報・データ - 宅配便取扱個数（メール便）【年次】 ▶ 自動車関係情報・データ - 宅配便取扱個数（宅配便）【年次】 ▶ 自動車関係情報・データ - 自賠責保険（共済）の損害別支払保険金（共済金）の推移（会計年度）【年次】 ▶ 自動車関係情報・データ - 政府保障事業の保障金支払状況の推移【年次】 ▶ 自動車関係情報・データ - 認証工場と指定工場の新規・廃止状況の推移【年次】 ▶ 自動車関係情報・データ - 認証・指定工場数及び新規・廃止状況（速報）【月次】 ▶ 自動車関係情報・データ - 優良認定工場数（速報）【月次】 ▶ 自動車関係情報・データ - 自動車整備事業者数【年次】 ▶ 自動車関係情報・データ - 認証・指定・優良認定関係統計【年次】 ▶ 自動車関係情報・データ - 前職別新規認証取得状況及び原因別認証廃業・指定返納状況の推移【年次】 ▶ 自動車関係情報・データ - 指定工場の処分工場数と主な違反内容の推移【年次】

公表主体	業種	指標名【更新頻度】
		<ul style="list-style-type: none"> ▶自動車関係情報・データ - 認証工場及び指定工場数の推移【年次】 ▶自動車関係情報・データ - 自動車保有車両数統計【月次】 ▶自動車関係情報・データ - トラックの不正な二次架装車両の改修進捗状況【四半期】 ▶自動車関係情報・データ - バスの不正な二次架装車両の改修進捗状況【四半期】 ▶自動車関係情報・データ - 自動車検査業務量（月報）【月次】 ▶自動車関係情報・データ - 自動車分解整備業実態調査【年次】 ▶自動車関係情報・データ - 自動車検査業務量（年報）【年次】 ▶自動車関係情報・データ - 数字で見る自動車【年次】

2) 旅行・ホテル

旅行・ホテル業界の景況感に関する構造化データを表 3-10 に示す。各指標について、業種の特性上、一般的には中小企業割合が高い業種だと考えられるため、中小企業の景況を直接的に把握する指標だと考えられる。

表 3-10 旅行・ホテル業界の景況感に関する構造化データ

公表主体	業種	指標名
内閣官房・内閣府	消費者向け 業種全般	<ul style="list-style-type: none"> ▶流動人口データ（V-RESAS名：人流 - 移動人口の動向）【週次】 ▶観光予報プラットフォームデータ（V-RESAS名：宿泊 - 宿泊者数）【週次】 ▶JCB消費NOW（V-RESAS名：消費 - 決済データから見る消費動向）【二週次】 ▶日経CPINOW（V-RESAS名：消費 - POSで見る売上高動向）【週次】

公表主体	業種	指標名
内閣府・経済産業省	消費者向け業種全般	<ul style="list-style-type: none"> ▶ 経路探索条件データ（RESAS名：観光マップ - 目的地分析）【年次】 ▶ 宿泊データ（RESAS名：観光マップ - From-to分析（宿泊者））【年次】 ▶ 宿泊旅行統計調査データ（RESAS名：観光マップ - 宿泊施設）【年次】 ▶ 訪日外国人消費動向調査データ（RESAS名：観光マップ - 外国人訪問分析）【年次】 ▶ モバイル空間統計データ（RESAS名：観光マップ - 外国人滞在分析）【年次】 ▶ モバイル空間統計データ（RESAS名：観光マップ - 外国人メッシュ）【年次】 ▶ 訪日外国人流動データ（RESAS名：観光マップ - 外国人入出国空港分析）【年次】 ▶ 訪日外国人流動データ（RESAS名：観光マップ - 外国人移動相関分析）【年次】 ▶ 多言語乗換案内データ（RESAS名：観光マップ - 外国人経路分析）【年次】 ▶ POSデータ（RESAS名：消費マップ - 消費の傾向（POSデータ））【年次】 ▶ POSデータ（消費マップ - From-to分析（POSデータ））【年次】 ▶ クレジットカード消費データ（RESAS名：消費マップ - 外国人消費の比較（クレジットカード））【年次】 ▶ クレジットカード消費データ（RESAS名：消費マップ - 外国人消費の構造（クレジットカード））【年次】 ▶ 免税利用状況データ（RESAS名：消費マップ - 外国人消費の比較（免税取引））【年次】 ▶ 免税利用状況データ（RESAS名：消費マップ - 外国人消費の構造（免税取引））【年次】 ▶ キャッシュレス・ポイント還元事業における加盟店データ（RESAS名：消費マップ - キャッシュレス加盟店数（ポイント還元事業））【年次】

公表主体	業種	指標名
		<ul style="list-style-type: none"> ▶ キャッシュレス・ポイント還元事業における加盟店データ (RESAS名：消費マップ - キャッシュレス決済データ (ポイント還元事業)) 【年次】
国土交通省 観光庁	旅行・ホテル	<ul style="list-style-type: none"> ▶ 宿泊旅行統計調査【月次】 ▶ 旅行・観光消費動向調査【四半期】 ▶ 訪日外国人消費動向調査【四半期】 ▶ 共通基準による観光入込客統計【四半期】 ▶ 旅行業者取扱額【月次】 ▶ 観光地域経済調査(2015年の調査結果のみ)【更新無し】 ▶ 旅行・観光サテライト勘定(TSA: Tourism Satellite Account)【年次】
国土交通省	旅行	<ul style="list-style-type: none"> ▶ 航空旅客動態調査【年次】 ▶ 全国幹線旅客純流動調査【五年次】 ▶ FF-Data(訪日外国人流動データ)【年次】
法務省 出入国在留管理庁	旅行	<ul style="list-style-type: none"> ▶ 出入国管理統計統計表【月次】
独立行政法人 国際観光振興機構(JNTO)	旅行・ホテル	<ul style="list-style-type: none"> ▶ 日本の観光統計データ【月次】
メトロエンジン株式会社	旅行・ホテル	<ul style="list-style-type: none"> ▶ BASIC PLAN【日次】
株式会社Agoop	消費者向け業種全般	<ul style="list-style-type: none"> ▶ 流動人口データ【—(公開情報無し)】 ▶ マチレポ【—(公開情報無し)】 ▶ Papilio【日次】 ▶ 人流統計レポート【—(公開情報無し)】
ナウキャスト社	消費者向け業種全般	<ul style="list-style-type: none"> ▶ JCB消費NOW【二週次】
株式会社hands	消費者向け業種全般	<ul style="list-style-type: none"> ▶ PERAGARU【日次】

3) 石油・鉱業・エネルギー

石油・鉱業・エネルギー業界のマーケットに関する構造化データを表 3-11 に示す。各指標について、中小企業の景況を直接的に把握する指標ではないものの、産業を構成するインフラコストであるため、ほぼ全ての中小企業に影響を及ぼす先行指標だと考えられる。

表 3-11 石油・鉱業・エネルギー業界のマーケットに関する構造化データ

公表主体	業種	指標名
S&P Dow Jones Indices LLC	エネルギー全般	▶ S&P GSCI (S&P Goldman Sachs Commodity Index) 【即時】
S&P Dow Jones Indices LLC	エネルギー全般	▶ S&P/ASX 200 Energy 【即時】
New York Mercantile Exchange (NYMEX)	原油	▶ WTI原油先物価格 【即時】
ICE Futures Europe	原油	▶ Brent原油先物価格 【月次】
東京商品取引所 (TOCOM)	原油	▶ プラッツドバイ原油先物価格 【即時】
Saudi Aramco	原油	▶ サウジアラビア原油価格 【即時】
クウェート石油公社 (KPC)	原油	▶ クウェート原油価格 【即時】
New York Mercantile Exchange (NYMEX)	天然ガス	▶ Henry Hub Natural Gas 【即時】
Intercontinental Exchange, Inc. (ICE)	天然ガス	▶ Dutch TTF Natural Gas Futures 【即時】
S&P Dow Jones Indices LLC	天然ガス	▶ S&P GSCI Natural Gas Index 【即時】
Australian Energy Regulator (AER)	天然ガス	▶ Gas market prices 【年次】
ARGUS MEDIA LIMITED	石炭	▶ 石炭関連の各種指数 【日次】
Intercontinental Exchange, Inc. (ICE)	石炭	▶ globalCOAL Newcastle Coal Futures 【即時】

公表主体	業種	指標名
globalCOAL	石炭	▶ The NEWC Index 【月次】
S&P Dow Jones Indices LLC	再生可能エネルギー	▶ S&P Global Clean Energy Index 【即時】
S&P Dow Jones Indices LLC / TMX Group	鉱業	▶ S&P/TSX Composite Index 【即時】
NYSE Arca	鉱業	▶ NYSE ARCA GOLD MINERS INDEX 【即時】
MSCI Inc.	鉱業	▶ MSCI ACWI Metals and Mining Index 【四半期】

3 - 2. 因果構造分析

本分析では、因果構造に関する専門的知見を有する和泉研究室の研究を基に、決算短信および新聞のテキストデータに記述された因果情報から、特定の経済・社会上のトピックが中小企業の各業種に対して与え得る波及効果を、テキストマイニングの活用により探索した⁴⁴。なお、本分析は各トピックにおける特定業種の選定やキーワード等の設定後、和泉研究室にて分析を実施した。

本分析は、事業者の外部環境を構成する諸要素として選定したトピックについて、特定の業種における経済活動への因果的波及効果を捉えることを目的として実施した。こうした観点では、和泉研究室が日本語テキストデータから経済事象の因果構造を分析する手法を精力的に探究しており、先行研究結果を多数有している。これを踏まえ、本分析においては、和泉研究室の先行研究⁴⁵を活用した。示唆の導出に当たっては、波及効果の分析対象としたい特定の業種や話題に関する事象の具体性の向上と、テキストマイニングの活用により、事象の連鎖の発見を両立させることを目指し、分析した。

3 - 2 - 1. 分析実施方針

(1) 因果構造分析の分析手法

因果的波及効果を捉える上で、テキストデータを解析し、そこから人間が認知し

⁴⁴ 本節の内容は和泉研究室に所属する小林の卒業論文（以下、小林(2023)とする）の一部であり、当該論文に基づく記載である。出典は「小林涼太郎「出力の多様性を考慮した因果連鎖提示システムによる経済波及効果分析」2022年度東京大学工学部システム創成学科 卒業論文」である。

⁴⁵ 和泉潔, 坂地泰紀. 経済因果チェーン検索のシステム紹介と応用. 第 33 回人工知能学会全国大会論文集, 2019.

た因果関係を抽出する技術が必要となる⁴⁶。本事業では、上記の分析手法の活用及び改良により、事業者を取り巻く外部環境における因果の波及的な事象に関する有益な示唆が得られるかについて検証した。

本分析のアウトプット例を図 3-32 に示す。図中の最上部は分析時に入力したキーワードであり、出力は入力キーワードのノードを起点とする有向グラフとして表現され、図の下方方向に向けてネットワークを拡張していく。破線で示す矢印が、類似性の判定に基づく因果関係同士の接続を示し、実線で示す矢印は、原因表現から結果表現へと結ばれる因果関係を示しており、起点と終点となるノードの単語群はある同一の文中から抽出されていることを表す。表示されている単語群は原因表現と結果表現中の名詞であり、実線矢印の終点となる色付きのノードが結果表現を示している。なお、 n 回目のネットワーク拡張によって得られる接続ノードを「第 n 階層のノード」と呼び、本分析では $n=2$ あるいは $n=3$ の場合について示している。

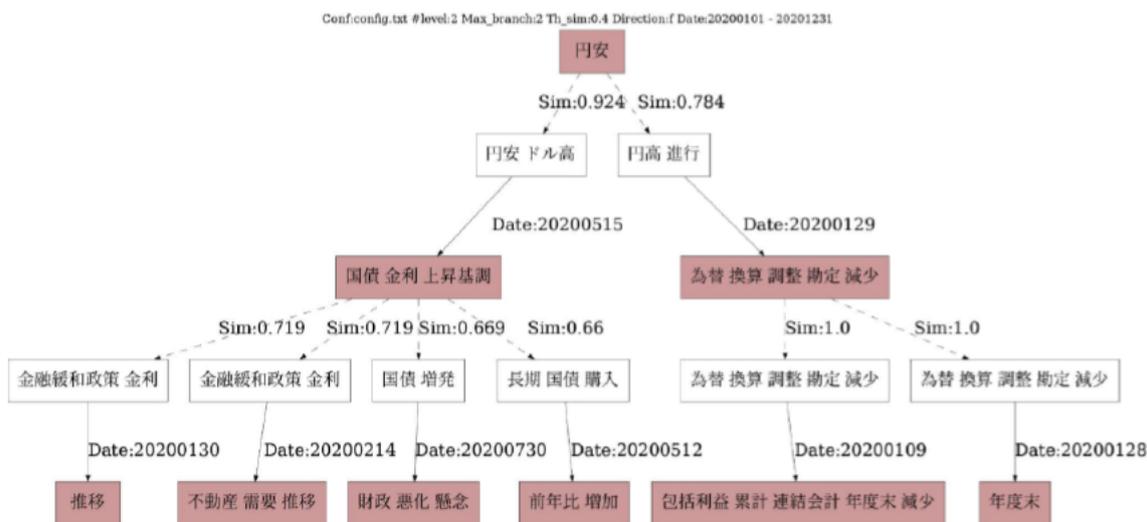


図 3-32 「円安」と入力した場合の出力事例（小林(2023)より引用）

表 3-12 本分析における用語説明

分類	用語	内容
出力図	Date	➤ 因果関係の抽出元となる決算短信が発行された日付

⁴⁶ 和泉潔, 坂地泰紀. "経済因果チェーン検索のシステム紹介と応用." 人工知能学会金融情報学研究会資料 2019.FIN-022 (2019): 67.

分類	用語	内容
	Sim	➤ ノード間の単語分散表現に基づく類似度 ⁴⁷ の値
パラメータ 調整	K	➤ 1ノード・1トピックあたりの最大接続ノード数 ※ K_i : 第 i 階層における 1ノード・1トピックあたりの最大接続ノード数
	θ	➤ 接続判定の際の単語分散表現に基づくノード間類似度Simの閾値 ※ θ_i : 第 i 階層における単語の分散表現に基づくノード間類似度の閾値
	ω	➤ ノード間の類似度の計算における文脈考慮度 ⁴⁸
	$period_i$	➤ 第 i 階層において使用するテキストデータ期間
	$sector_i$	➤ 第 i 階層における対象業種

異なる因果文書間の繋ぎ合わせの際には、和泉研究室で保有する単語の分散表現を算出した学習済みモデルを活用し、結果表現ノードと原因表現ノードの文書類似度を算出した上で、文書類似度が高いノードを繋ぎ合わせている。

(2) 分析のアプローチ

本分析の目的に鑑み、分析手法の改良に際して入力キーワードや対象データなど様々な条件設定や工夫が考えられる。本分析では波及効果の分析対象としたい特定の業種や話題に関する事象の具体性の向上と、テキストマイニングの活用による想定しなかった事象の発見を両立できるように、段階的に4つのアプローチによる検証を行った。アプローチの概要を表 3-13 に示す。

⁴⁷ 類似度 Sim はコサイン類似度 $\text{Cos}(w, w')$ を基に計算する。ここで、 w や w' は各ノードの原因表現や結果表現の単語分散表現を平均したベクトルである。

⁴⁸ 単語分散表現に基づくノード間類似度の計算においては、単に2つのノードの結果表現-原因表現間の類似度を計算しているのではなく、文脈考慮度 ω の割合だけ原因表現-原因表現同士の類似度を足し合わせることで、終端ノードの結果表現のみからは分からない出来事や判断（結果）に至るまでの文脈を考慮している。したがって、ノード間類似度は、単語の分散表現に基づく表層的な一致の計算でありながらも、以前のネットワーク拡張における事象の変遷の過程を反映した値であると考えられる。なお、本分析では「①既存アプローチ」にて $\omega = 0.0$ として実験的に出力を行った。

表 3-13 各アプローチの概要

アプローチ	対象データ	概要
① 既存アプローチ	決算短信テキスト	<ul style="list-style-type: none"> ▶ 従前の分析手法を踏襲し、特段の条件設定や工夫を行わず実施した
② トピック絞り込みアプローチ	データ	<ul style="list-style-type: none"> ▶ 因果的波及効果の分析対象としたい業種や話題を絞り込むために、「サブキーワードの追加」と「対象業種の限定」を行った <ul style="list-style-type: none"> ✓ サブキーワードの追加：「感染症」や「円安」といった比較的大きく幅広い因果的波及効果を生じさせるトピックを「メインキーワード」として、それより具体的あるいは特定の業種に限定的な「サブキーワード」を入力キーワードとして追加することで、出力する結果の絞り込みを行う ✓ 対象業種の限定：特定の業種の企業が発行する決算短信テキストデータのみを使用し、接続される因果関係の対象業種をあらかじめ限定することで、出力する結果の絞り込みを行う ▶ 既存アプローチにおいて、ノード中の名詞が2語以下の場合、話題が発散する傾向にあったため、本アプローチ以降は3語以上となるようパラメータ調整を行った
③ 階層的調整アプローチ		<ul style="list-style-type: none"> ▶ 因果的波及効果の分析対象としたい業種や話題を絞り込む「トピック絞り込み」と、「多様な因果的波及効果の発見」の両立を図るため、階層ごとに使用する対象データの期間や対象業種などのパラメータ調整を行った。具体的には、第1階層のパラメータ調整の閾値を大きい値とし、第2階層以降のパラメータ調整の閾値を小さい値とした⁴⁹
④ 対象データ拡充アプローチ	<ul style="list-style-type: none"> ▶ 決算短信テキストデータ ▶ 新聞記事テキストデータ 	<ul style="list-style-type: none"> ▶ ①から③までのアプローチで対象データとしていた決算短信テキストデータだけではなく、新聞記事テキストデータを併せて利用した ▶ 決算短信テキストデータは主に、企業の決算に直接的に影響を与えるような話題が記載されている傾向にある一方、新聞記事テキストデータは主に、政治や国民に影響を与える一般的な話題が記載されているため、企業への間接的な影響など多様な因果的波及効果の発見が期待できる。そこで、第1階層、第2階層では決算短信テキストデータを利用した因果探索を行い、第3階層では新聞記事テキストデータから抽出した因果に対して第2階層からさらに枝を延ばす

⁴⁹ 具体的には、第1階層の因果を探索する際の類似度の閾値のパラメータ θ を0.8と大きな値にして、類似度が高い因果とのみ接続するように探索条件を厳しくし、第2階層以降では θ を0.4などの小さな値とした。

アプローチ	対象データ	概要
		かたちで、同様に表現の類似度から因果探索を行った。利用イメージを図 3-33に示す

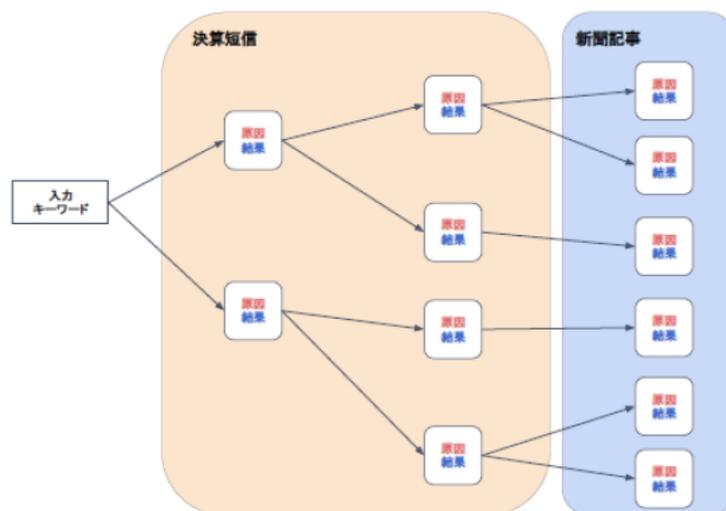


図 3-33 決算短信テキストデータと新聞記事テキストデータの利用イメージ
(小林(2023)より引用)

(3) 選定トピック

本分析では、企業の経済活動に大きな影響を及ぼすと考えられる経済変動事象として、「感染症」「円安」「原材料・資源価格の高騰」「サプライチェーン混乱」「震災」という5つのトピックを入力キーワードとし、各アプローチで検証を実施した。なお、本書では紙面の関係上、各アプローチにおける全5つのトピックの分析結果を記載するのではなく、アプローチごとに特徴的なトピックにおける分析結果を一部記載する。

(4) 妥当性や精度に関する評価方法

事業者の外部環境を構成する個々の諸要素を仮説的に構造化したネットワークに対して、その妥当性や精度を確認するために、本分析ではアプローチごとに中小企業庁職員及び分析実施者による複数の視点からの定性評価を実施した。

3-2-2. 分析結果

(1) 分析結果の概要

各アプローチにおける分析結果の概要を表 3-14 に示す。

- ① 既存アプローチでは、「感染症」と「震災」のトピックについて出力した結果、出力された事象については、各トピックが多岐に渡る一般的な話題であることに加え、条件設定や工夫も特段行っていないことから、「財政」や「広告」といった様々な名詞が導出され、階層を重ねることで話題が散逸する傾向であることが確認された。この分析結果を受けて、トピックについて絞り込みを行っていく必要があると考え、②トピック絞り込みアプローチを実施した。
- ② トピック絞り込みアプローチでは、「円安」のトピックについてサブキーワードの追加や対象業種の限定を行い出力した結果、対象業種に関連した話題の出力が多くなる傾向が確認できた。一方で、特定の業種における決算短信テキストデータの一部のみを使用するため、「売上高」といった業績に関する話題の出力が多く、多様な因果的波及効果を捉えることが困難であった。この分析結果を受けて、より多様な因果的波及効果の発見ができるように、③階層的調整アプローチを実施した。
- ③ 階層的調整アプローチでは、「感染症」、「サプライチェーン混乱」、「原材料・資源価格の高騰」のトピックについて階層ごとにパラメータを調整し出力した結果、「感染症」においては多様な、かつ具体的な内容まで捕捉できていることを確認できた。一方で、決算短信テキストデータは企業に直接的に影響を与える話題が主に記載されていることから、「外食自粛」や「休業」等の発見にとどまるなど、想定しなかった事象の発見という観点では限定的であった。この分析結果を受けて、対象データを拡げていく必要があると考えられ、④対象データ拡充アプローチを実施した。
- ④ 対象データ拡充アプローチでは、「感染症」と「円安」のトピックについて新聞記事テキストデータを活用して出力した結果、「感染症」においては「中部地方の製造業の減産」といった想定しなかった具体的な因果的波及効果を確認できた。「円安」においても、「中部地方の製造業の減産」といった地域限定的な、かつ想定しなかった事象が確認できた一方、感染症に関する話題も多く確認された。この要因として、使用した新聞記事テキストデータでは、感染症に関連する記事が非常に多いことが考えられる。

表 3-14 各アプローチにおける分析結果の概要

	①既存	②トピック絞り込み	③階層的調整	④対象データ拡充
トピック	<ul style="list-style-type: none"> ▶ 感染症 ▶ 震災 	<ul style="list-style-type: none"> ▶ 円安 	<ul style="list-style-type: none"> ▶ 感染症 ▶ サプライチェーン混乱 ▶ 原材料・資源価格の高騰 	<ul style="list-style-type: none"> ▶ 感染症 ▶ 円安
特徴や課題	<p>▶ トピックが多岐に渡る一般的な話題であり、条件設定や工夫を特に行っていないことから、<u>具体的な事象はあまりなく、階層を重ねることで話題が散逸する傾向</u></p>	<p>▶ 因果的波及効果を捉えたい業種の<u>具体的な事象は確認できたが、多様な因果的波及効果の発見が困難</u>（トピックの具体性と多様性はトレードオフの関係）</p>	<p>▶ 一部のトピックにおいて、<u>多様な因果的波及効果が拾えており、具体的な事象まで捕捉できたが、想定しなかった事象の発見は限定的</u></p>	<p>▶ 地域限定的な事象などの<u>想定しなかった事象が確認できた</u></p> <p>▶ 出力される事象は、<u>記事件数が多い話題（例：感染症）に大きく左右される</u></p>

(2) アプローチごとの分析結果

1) 既存アプローチ

a. 感染症

本アプローチにおける「感染症」トピックのパラメータを表 3-15 に示す。

表 3-15 既存アプローチにおける「感染症」トピックのパラメータ

図	入力キーワード	K	θ	ω
図 3-34	新型コロナウイルス感染症	2	0.4	0.0

図 3-34 は「新型コロナウイルス感染症」を入力キーワードとする出力事例である。第1階層のノードでは、「在宅勤務」や「リモートワーク」などの働き方の変化について複数のノードが確認できた。そして、そのような第1階層のノードから、

「CROWD サービス」といった特定の業種への因果的波及効果も確認できた。これらの結果は、一般的な内容であり予測もしやすいという点で、想定しなかった事象とは必ずしも言えない。

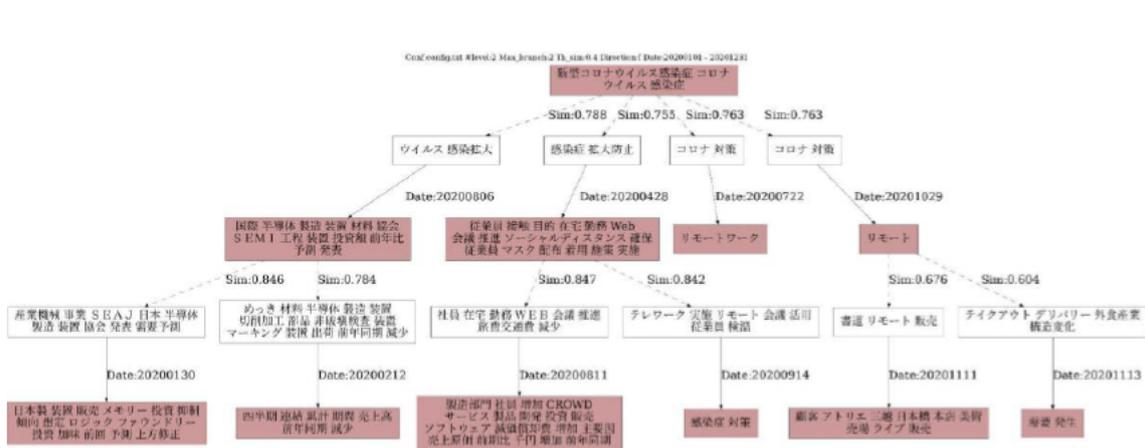


図 3-34 「新型コロナウイルス感染症」と入力した場合の出力事例 (小林(2023)より引用)

b. 震災

本アプローチにおける「感染症」トピックのパラメータを表 3-16 に示す。

表 3-16 既存アプローチにおける「震災」トピックのパラメータ

図	入力キーワード	K	θ	ω
図 3-35	震災	2	0.4	0.0
図 3-36	震災	2	0.4	0.0

図 3-35 は、「震災」を入力キーワードとし、2020 年 1 月から 2020 年 12 月までに発行された決算短信テキストデータを使用した出力事例である。第 1 階層のノードでは「株式」などの資産関連や「消費者」の動向関連、第 2 階層のノードでは「会計」や「財政」、「広告」、「法令」などの様々な名詞が導出され、階層を重ねるごとに話題が散逸してしまう傾向が確認できた。

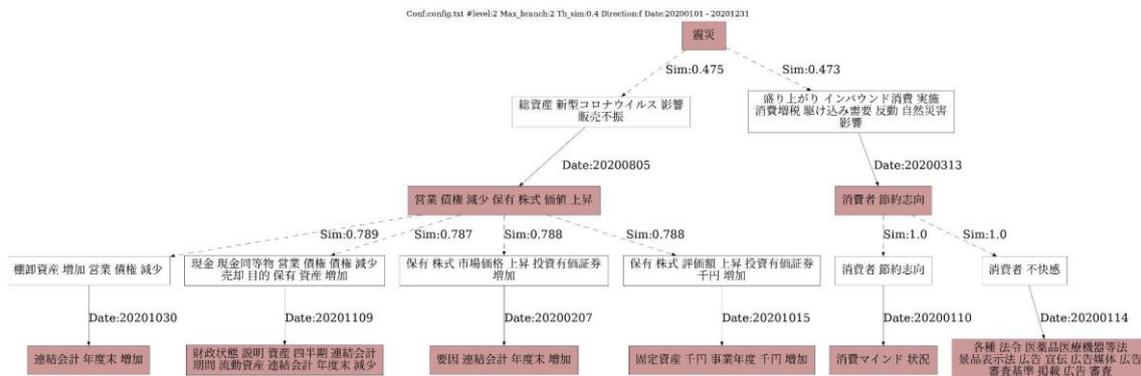


図 3-35 「震災」と入力した場合の出力事例
(データ期間：2020年1月～2020年12月) (小林(2023)より引用)

一方で、図 3-36 では決算短信テキストデータの対象期間を拡大し、2013年1月から2020年12月までに発行された決算短信テキストデータを使用したところ、震災から波及する経済事象として「影響の長期化」といった話題を部分的に確認することはできたものの、具体的な事象であるとは必ずしも言えない。

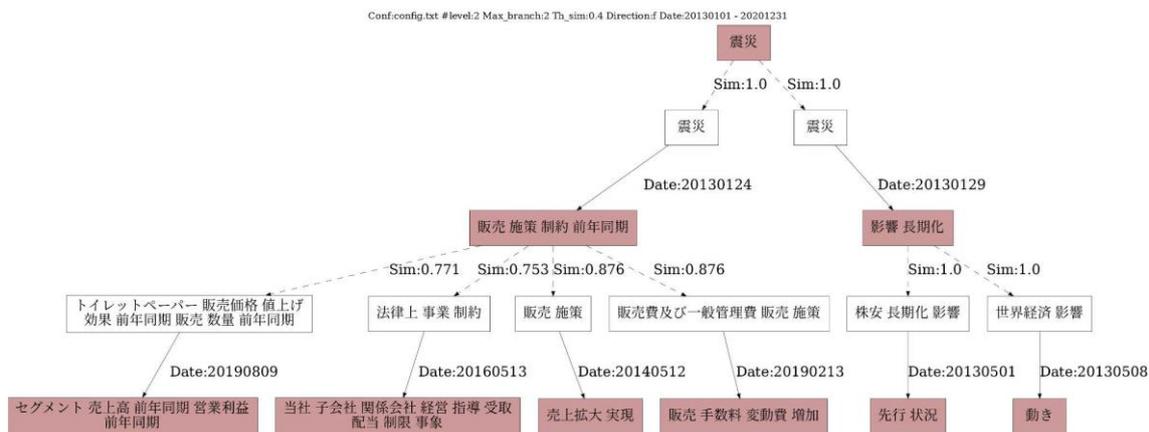


図 3-36 「震災」と入力した場合の出力事例
(データ期間：2013年1月～2020年12月) (小林(2023)より引用)

2) トピック絞り込みアプローチ

a.円安

本アプローチにおける「円安」トピックのパラメータを表 3-17 に示す。

表 3-17 トピック絞り込みアプローチにおける「円安」トピックのパラメータ

図	入力キーワード	対象業種	K	θ	ω
図 3-37	円安	-	2	0.4	0.3
図 3-38	円安	石油・石炭製品	2	0.4	0.3
図 3-39	円安	「製造業」	2	0.4	0.3
図 3-40	円安 自動車	「製造業」	2	0.4	0.3

図 3-37 から図 3-40 では、いずれも「円安」をメインキーワードとしており、2013 年 1 月から 2020 年 12 月までに発行された決算短信テキストデータを使用した。

図 3-37 には、比較のため、トピック絞り込み手法を活用しない出力を示している。第 2 階層のノードでは、「訪日外国人の増加」といった具体的な事象が確認できたが、一般的な内容であり予測もしやすいという点で、想定しなかった事象とは必ずしも言えない。



図 3-37 「円安」と入力した場合の出力事例（小林(2023)より引用）

図 3-38 では、対象業種を「石油・石炭製品」に限定した出力事例である。第 1 階層のノードでは、「利益の損失」といった業種特有の経済事象が確認でき、第 2 階層のノードでは「在庫」や「生産」といった類似するような話題の出力が多い。

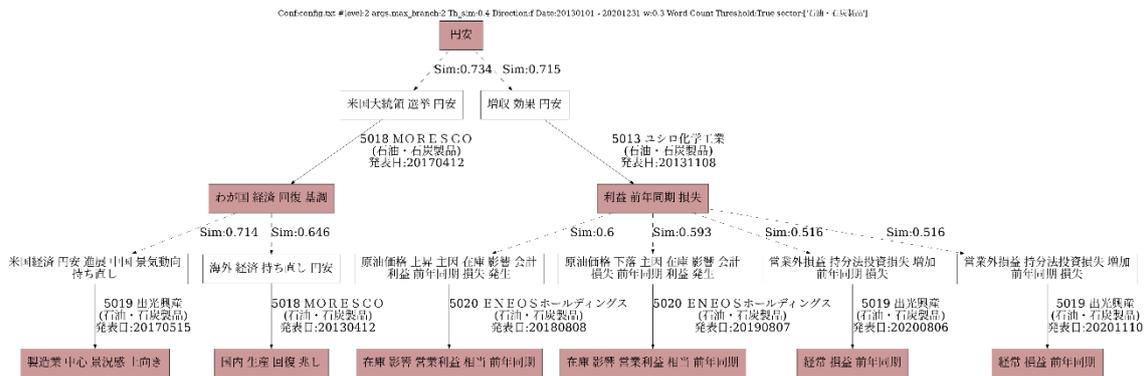


図 3-38 「円安」と入力した場合の出力事例 (対象業種を「石油・石炭製品」に限定) (小林(2023)より引用)

図 3-39 では、大分類で製造業に分類される「金属製品」「非鉄金属」「石油・石炭製品」「輸送用機器」「精密機器」「電気機器」という複数の業種にまたがって対象業種の限定を行った。その結果、図 3-38 と同様に第 2 階層では「売上高」といった業績に関する話題の出力が多く、「タイとベトナムにおける製造」といった「精密機器」の企業に関する話題の出力を確認できた。

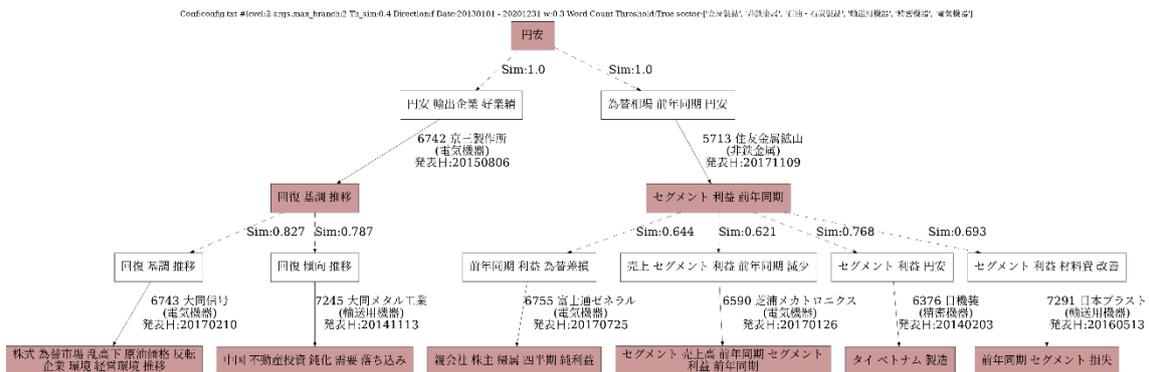


図 3-39 「円安」と入力した場合の出力事例 (対象業種を「金属製品」「非鉄金属」「石油・石炭製品」「輸送用機器」「精密機器」「電気機器」に限定) (小林(2023)より引用)

図 3-40 では、「サブキーワードの追加」と「対象業種の限定」を組み合わせるとピック絞り込みを実施するため、図 3-39 と同様の対象業種に限定し、「自動車」をサブキーワードとして追加した。「円安」と「自動車産業」は関連性が強いキーワードであると考えられるが、両方にまたがる話題の出力はあまり確認できなかった。また、第 2 階層でも「売上高」といった業績に関する話題の出力が多く、特定の業種における決算短信テキストデータの一部のみを使用しているため、多様な因果的波及効果を捉えることが困難であった。

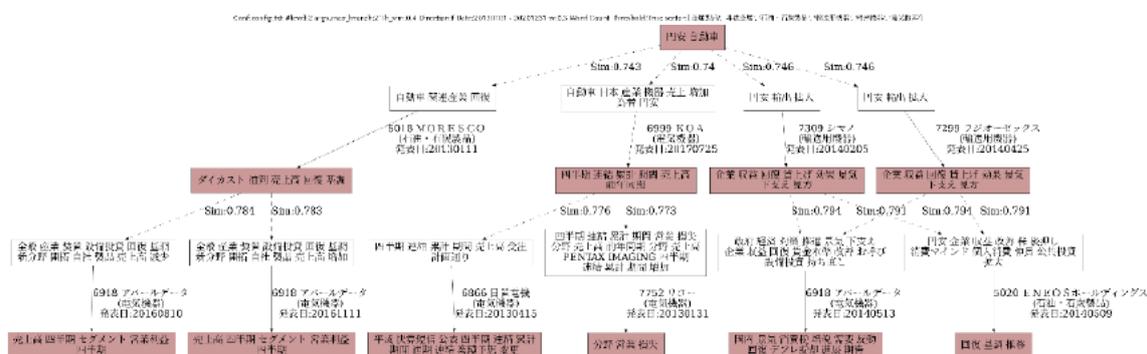


図 3-40 「円安 自動車」と入力した場合の出力事例（対象業種を「金属製品」「非鉄金属」「石油・石炭製品」「輸送用機器」「精密機器」「電気機器」に限定）
（小林(2023)より引用）

3) 階層的調整アプローチ

a. 感染症

本アプローチにおける「感染症」トピックのパラメータを表 3-18 に示す。

表 3-18 階層的調整アプローチにおける「感染症」トピックのパラメータ

図	K_1	K_2	θ_1	θ_2	ω	$period_1$	$period_2$	$sector_1$	$sector_2$
図 3-41	2	2	0.4	0.4	0.3	2020	2020	サービス業	サービス業
図 3-42	2	2	0.4	0.4	0.3	2020	2013-2020	サービス業	-

図	K_1	K_2	θ_1	θ_2	ω	$period_1$	$period_2$	$sector_1$	$sector_2$
図 3-43	5	2	0.4	0.4	0.3	2020	2013-2020	サービス業	-

図 3-41 から図 3-43 では、いずれも「新型コロナウイルス感染症」を入力キーワードとしており、階層ごとに対象業種の限定を変更した。

図 3-41 は、第 1 階層と第 2 階層の対象業種を「サービス業」に限定した出力事例であるが、特に第 2 階層のノードでは一般的な事象が多い傾向にあり、想定しなかった事象とは必ずしも言えない。

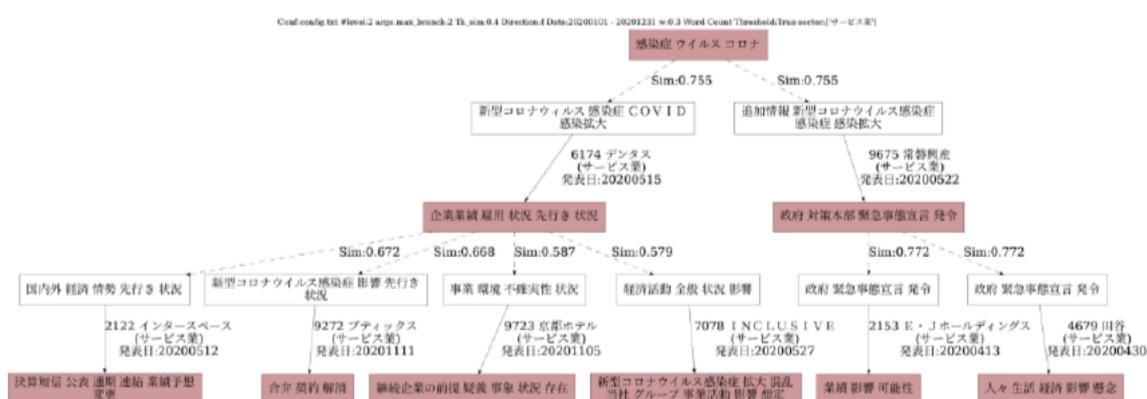


図 3-41 「新型コロナウイルス感染症」と入力した場合の出力事例
(第 1 階層と第 2 階層の対象業種を「サービス業」に限定) (小林(2023)より引用)

図 3-42 は、第 1 階層の対象業種を「サービス業」に限定した出力事例である。第 2 階層のノードでは、「ホテル事業」や「消費マインド」、「個人消費」といった多様な因果的波及効果を把握できた。

conf.config.txt #level:2 max_branch(1st layer):2 max_branch:2 th_sim(1st layer):0.4 th_sim:0.4 direction:f Date(1st layer):20200101 - 20201231 Date:20130101 - 20201231 w:0.3 text:コロナウイルス 感染症 wc:False

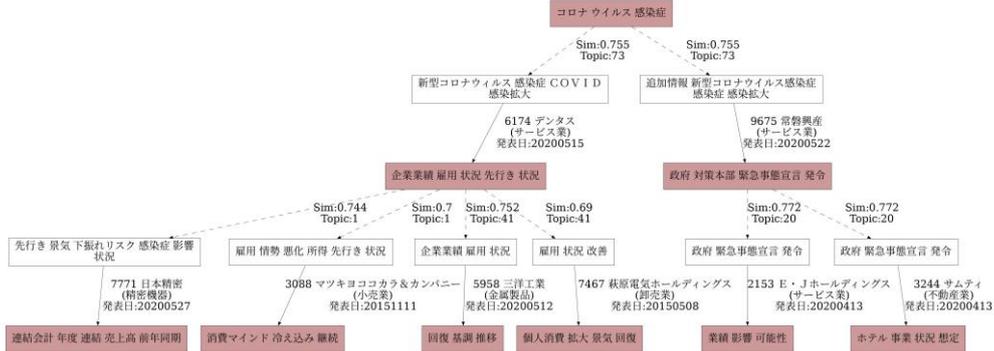


図 3-42 「新型コロナウイルス感染症」と入力した場合の出力事例 (第1階層の対象業種を「サービス業」に限定) (小林(2023)より引用)

図 3-43 は、図 3-42 の出力からさらに、第1階層の接続ノード数を増やした出力事例である。第2階層のノードでは、「消費マインドの冷え込みが継続」や「家庭での内食需要」、「外出自粛によるまとめ買い」といった具体的かつ多様な因果的波及効果を把握できた。一方で、これらの事象はまだ一般的な内容に留まっており、想定しなかった事象の発見という観点では限定的だといえる。

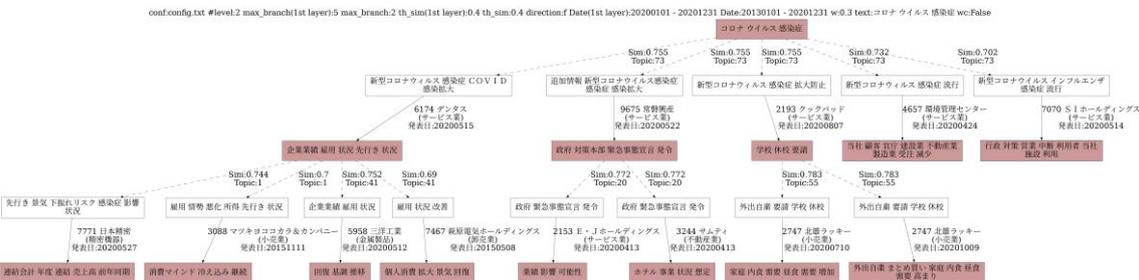


図 3-43 「新型コロナウイルス感染症」と入力した場合の出力事例 (第1階層の対象業種を「サービス業」に限定し、第1階層の最大接続ノード数を増加) (小林(2023)より引用)

b. サプライチェーン混乱

本アプローチにおける「サプライチェーン混乱」トピックのパラメータを表 3-19 に示す。

表 3-19 階層的調整アプローチにおける「サプライチェーン混乱」トピックのパラメータ

図	K_1	K_2	θ_1	θ_2	ω	$period_1$	$period_2$	$sector_1$	$sector_2$
図 3-44	2	2	0.65	0.65	0.1	2020	2020	-	-

図 3-44 は、「需要」というサブキーワードを入力した出力事例である。第1階層のノードでは「流通在庫の増加」や「休業」の事象が確認でき、第2階層のノードでは「プライダル」や「飲食」、「ファッション」といった特定の業種への因果的波及効果を把握できた。データ期間を2020年の一年間だけに絞り込んだことで、対象業種の限定をすることなく、多様な業種への影響を捉えることが可能である。

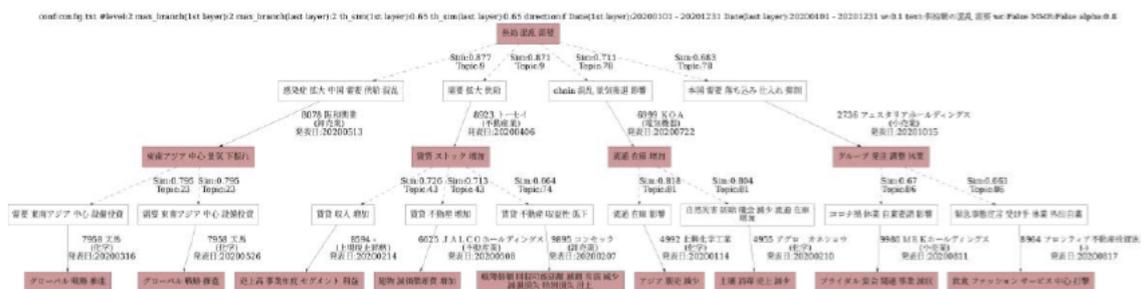


図 3-44 「供給網の混乱 需要」と入力した場合の出力事例 (小林(2023)より引用)

c.原材料・資源価格の高騰

本アプローチにおける「原材料・資源価格の高騰」トピックのパラメータを表 3-20 に示す。

表 3-20 階層的調整アプローチにおける「原材料・資源価格の高騰」トピックのパラメータ

図	K_1	K_2	θ_1	θ_2	ω	$period_1$	$period_2$	$sector_1$	$sector_2$
図 3-45	2	2	0.4	0.4	0.3	2020	2020	-	-
図 3-46	5	2	0.8	0.4	0.3	2020	2013-2020	-	-

本アプローチでは、「第1階層のパラメータ調整を厳しくし、第2階層以降のパラメータ調整を緩くする」といった階層ごとの調整を実施してきたが、第1階層のみ1ノード・1トピックあたりの最大接続ノード数を増やすことによる効果を検証した。図3-45と図3-46では、いずれも「原材料 価格 高騰 製造」を入力キーワードとしており、図3-45は比較のため、階層的調整アプローチを活用しない出力を示している。

図3-45では、第1階層のノードにおいて「ツナ」や「乾物」といった具体的な事象を捉え、第2階層のノードにおいて「食品」や「缶詰」といった特定の業種への因果的波及効果を把握することができた。

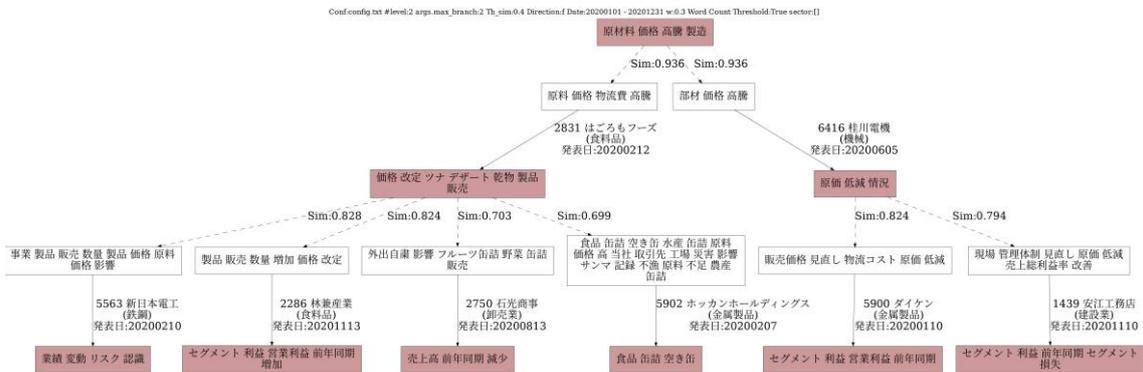


図 3-45 「原材料 価格 高騰 製造」と入力した場合の出力事例 (小林(2023)より引用)

図3-46は、1ノード・1トピックあたりの最大接続ノード数を増加した結果であるが、類似するような話題の出力が多く、可読性の観点においても懸念がある。

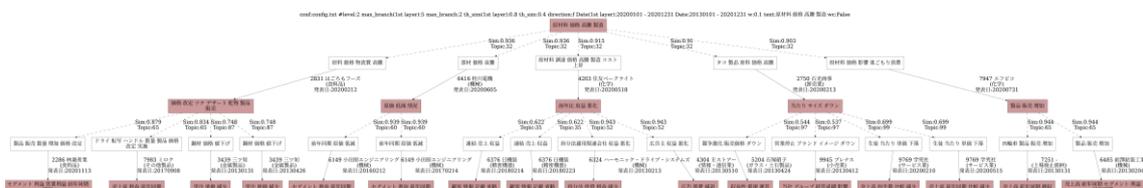


図 3-46 「原材料 価格 高騰 製造」と入力した場合の出力事例 (第1階層における最大接続ノード数を増加) (小林(2023)より引用)

4) 対象データ拡充アプローチ

a. 感染症

本アプローチにおける「感染症」トピックのパラメータを表 3-21 に示す。

表 3-21 対象データ拡充アプローチにおける「感染症」トピックのパラメータ

図	K_1	K_2	K_3	θ_1	θ_2	θ_3	ω	$sector_1$
図 3-47	3	3	2	0.4	0.4	0.6	0.3	サービス業

図 3-47 では、第 1 階層における対象業種を「サービス業」に限定し、「サービス業」業種における「感染症」に関する事象から派生し得る経済事象を出力した。第 1 階層のノードでは「移動や経済活動の制限」や「休業」、「営業活動の縮小」といった事象が確認できた。

これらの事象から続く第 2 階層のノードでは、具体的な波及効果が確認できた。例えば、ポンプ・風水力機械分野で海外事業を拡大している荏原製作所では、「移動制限」による「指導員派遣の影響」といった因果的波及効果を把握できた。また、クリーニング機器の販売を行う蔵王産業においては、取引先の「休業」や「営業活動の自粛」を受け、「実演の機会が減少」しているという波及効果を捉えている。

さらに第 3 階層では、より多様な波及効果も確認できた。例えば、荏原製作所の「指導員派遣の影響」のノードから「HoloLens (ホロレンズ)⁵⁰の作業手順」といったノードが接続され、製造現場における MR 技術の活用推進といった想定しなかった事象も確認できた。

⁵⁰ マイクロソフト社によって開発された MR (Mixed Reality) デバイス。

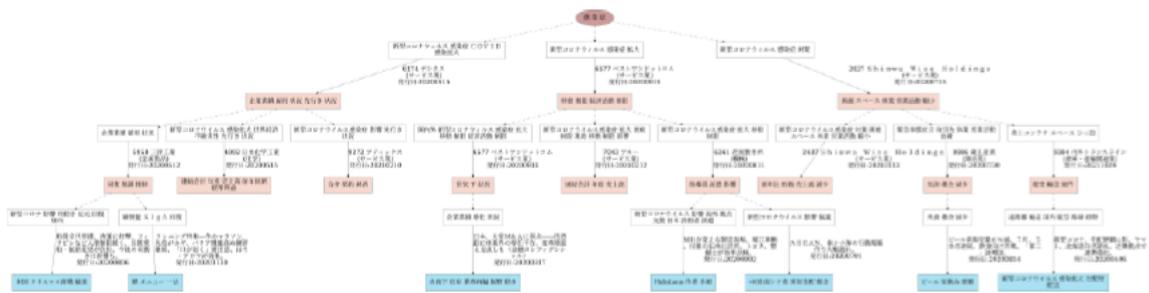


図 3-47 「感染症」と入力した場合の出力事例（第1階層における対象業種を「サービス業」に限定し、第3階層で新聞記事テキストデータを利用）
（小林(2023)より引用）

b.円安

本アプローチにおける「円安」トピックのパラメータを表 3-22 に示す。

表 3-22 対象データ拡充アプローチにおける「円安」トピックのパラメータ

図	K_1	K_2	K_3	θ_1	θ_2	θ_3	ω	$sector_1$
図 3-48	3	2	2	0.4	0.4	0.4	0.3	繊維製品

図 3-48 では、第1階層における対象業種を「繊維製品」に限定し、「繊維製品」業種における「円安」に関する事象から派生し得る経済事象を出力した。第1階層のノードでは「海外経済」や「景気持ち直しの動き」といった事象が確認できたものの、比較的低いノード間類似度で接続されていた。

続く第2階層では、「製造業を中心としてコロナ禍で先送りになっていた投資再開の動きが広がる」といった具体的な事象を確認でき、「円安」から「製造業の投資」への因果的波及効果を把握できた。

第3階層においては、「中部地方の製造業の減産」や「埼玉県内への企業誘致実績」といった地域限定的な事象も確認でき、これらは想定しなかった事象であるといえる。一方で、「円安」トピックにもかかわらず、「感染症」に関するノードが多く見受けられた。この要因として、使用した新聞記事テキストデータでは、感染症に関連する記事が非常に多いことが考えられる。

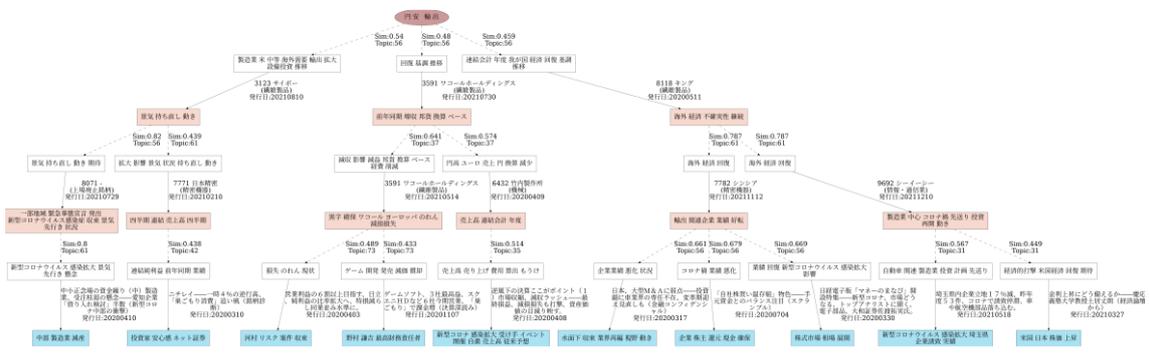


図 3-48 「円安輸出」と入力した場合の出力事例（第1階層における対象業種を「繊維製品」に限定し、第3階層で新聞記事テキストデータを利用）
（小林(2023)より引用）

4. 本事業で得られた示唆と次年度以降のアクションプラン

本章では、前章までにおいて得られた示唆と、それを踏まえたオルタナティブデータの活用に関して、次年度以降で取り組んでいくことが望ましいアクションプランを示す。

4-1. 本事業で得られた示唆

4-1-1. データ利活用の可能性

本事業の遂行過程や有識者からの意見等を踏まえて、クレジットカードデータを基に作成した指標とテキストデータを基に構造化した諸要素ネットワークにおいて、活用可能性が高いと考えられるケース（以下、「ユースケース」という。）について整理した。

(1) 「JCB 消費 NOW」を基に作成した中小企業指標の活用ユースケース

指標の活用可能性については、経済学や統計学、データサイエンス等に知見のある学識者・民間人材で構成される研究会において助言・指導等を受けた。また、中小企業の売上動向を注視している地方自治体等（以下、「想定ユーザー」という。）へのヒアリング調査を実施した。

これらの取組を踏まえ、「既存統計のオルタナティブデータとしての活用」、「中小企業と大企業の比較」、「一人当たり消費金額の変化の把握」、「地方における中小企業の売上動向の把握」の4つを指標のユースケースとして示す。

1) 既存統計のオルタナティブデータとしての活用

既存統計において企業の売上動向を示している業種について、本指標は当該既存統計のオルタナティブデータとしての活用が期待できる。

本指標は既存統計と比較して、速報性の観点でも優れており、足下の状況を踏まえた経済対策の検討に活用できる。速報性については、想定ユーザーからも「既存統計の多くは速報性の観点で課題があり⁵¹、足下のデータを基にした景気判断が困難である」という意見があり、本指標の活用によって、機動的な政策対応の要否を円

⁵¹ 更新頻度が月次である既存統計の多くは、調査対象時期と公表時期のタイムラグは1か月程度である（例：「商業動態統計」の場合、当月の結果を、翌々月上旬に公表）。また、更新頻度が四半期である既存統計は、調査対象時期と公表時期のタイムラグが数か月程度あり、速報性に課題がある（例：「法人企業統計調査」の場合、4～6月期の結果は9月初旬に公表するため、3～5か月のタイムラグがある）。

滑に検討するための材料となり得る。また、中小企業に限定して動向を把握できる既存統計は限られているが、中小企業の売上動向を把握できる本指標の活用によって、中小企業の実態を踏まえた政策対応の要否を円滑に検討するための材料となり得る。本指標と既存統計の両方が存在することで、本指標をモニタリングしながら、既存統計の示す企業の売上動向と比較しながら確認することができるという点で、より中小企業の売上動向把握の蓋然性を高めるデータとしての価値が十分にある。

また、既存統計の「商業動態統計」では、都道府県別・地域別で業種ごとの動向を集計・公表されているが、「対象業種が少ない⁵²ことが課題である」といった想定ユーザーからの意見があり、既存統計は業種のカバレッジに課題がある。一方で、業種のカバレッジに関しては、本事業で活用した「JCB 消費 NOW」が対象としているのは全 62 業種あり（詳細は「2-1-3 業種選定」参照）、幅広い業種を対象とした中小企業指標を作成できる可能性がある。一例として、今回は「酒屋」の中小企業指標を作成したが、「酒屋」自体がこれまで既存統計では売上動向を捕捉できなかった業種である。このように、幅広い業種間で売上動向の差異が確認できることから、これまで以上に中小企業政策における経営支援対象の業種設定に活用できる。業種別で売上動向把握ができる点については、想定ユーザーからも「感染症で注目された宿泊や飲食業などの業種の売上動向を把握できることは、十分に価値がある」といった意見があり、本指標の活用によって、経営支援対象を検討するための材料となり得る。

表 4-1 オルタナティブデータとしての活用が見込める業種ごとの既存統計

業種	強い正の相関がみられた既存統計（既存統計における業種）
織物・衣服・身の回り品小売業	<ul style="list-style-type: none"> ▶ 中小企業庁「中小企業景況調査」（小売業（織物・衣服・身の回り品）） ▶ 経済産業省「商業動態統計」（織物・衣服・身の回り品小売業）
機械器具小売業	<ul style="list-style-type: none"> ▶ 経済産業省「商業動態統計」（機械器具小売業）
外食	<ul style="list-style-type: none"> ▶ 中小企業庁「中小企業景況調査」（飲食店、持ち帰り、配達飲食） ▶ 日本銀行「消費活動指数」（外食） ▶ 日本銀行「短観」（宿泊・外食サービス）
旅行	<ul style="list-style-type: none"> ▶ 日本銀行「消費活動指数」（外食）

⁵² 「百貨店・スーパー」「コンビニエンスストア」「家電大型専門店」「ドラッグストア」「ホームセンター」。

業種	強い正の相関がみられた既存統計（既存統計における業種）
宿泊	<ul style="list-style-type: none"> ▶ 中小企業庁「中小企業実態基本調査」（宿泊業） ▶ 中小企業庁「中小企業景況調査」（宿泊業） ▶ 日本銀行「消費活動指数」（宿泊） ▶ 日本銀行「短観」（宿泊・外食サービス）
EC	▶ 経済産業省「商業動態統計」（無店舗小売業）
居酒屋	▶ 日本フードサービス協会「外食産業市場動向調査」（居酒屋）
焼肉	▶ 日本フードサービス協会「外食産業市場動向調査」（焼き肉）
ゴルフ場	▶ 経済産業省「特定サービス産業動態統計調査」（ゴルフ場）
タクシー	▶ 日本銀行「消費活動指数」（タクシー）

2) 中小企業と大企業の比較

本事業では業種別に中小企業指数と大企業指数を作成しており、それぞれの指数の動きを比較することが期待できる。なお、今回活用した中小企業分類マスタの精度がおおむね正確であることが条件としてあるものの、中小企業指数と大企業指数から見える売上動向の差異が、一部の業種で見られている。一例として、「機械器具小売業」の指数の推移を図 4-1 に示す。「機械器具小売業」の中小企業指数と大企業指数では、2020 年中盤に特徴的な差がみられる。この指数の推移から、「機械器具小売業」の 2020 年中盤において売上は回復傾向にあるものの、中小企業と大企業では違いが生じていることが示唆される。なお、図 4-1 は「前年比」の指数の推移ではなく、2016 年度から 2018 年度の平均値をベースライン（基準）とした「16-18 年度ベースライン比」であり、感染症拡大以前からどの程度売上が回復しているのかを把握可能な指数の推移である。想定ユーザーからも「全体（大企業含む）の動向と、中小企業の動向の乖離を捕捉できる指標があれば有用と考えられる」という意見があり、中小企業指数と大企業指数といった形で細分化することによって、企業規模別の詳細な売上動向を把握し、大企業と比較して中小企業政策を検討するための材料となり得る。

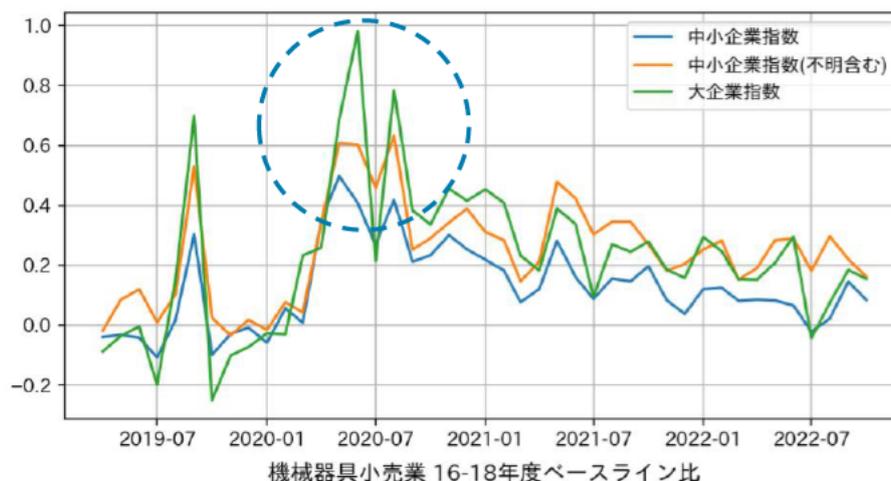


図 4-1 機械器具小売業の指数の推移
(16-18 年度ベースライン比)

3) 一人当たり消費金額の変化の把握

本事業で試作した中小企業指標は、業種・地域単位での消費動向指標であるが、基となるデータは消費者単位の消費金額のデータであり、その積上げベースで指標を算出している。本指標を、一人当たり消費金額の変化を捉える指数（以下、「IM 指数」(Intensive Margin) という。)と消費者数の変化を捉える指数（以下、「EM 指数」(Extensive Margin) という。)に分解することにより、一人当たりの消費金額と消費者数の変化に関して詳細な分析が可能となる。

一例として、「外食」における「IM+EM 指数」を図 4-2 に示す。「外食」では、2017 年から 2020 年序盤までの期間は「前年比」で好転している傾向であるが、その要因は「IM 指数」の示す一人当たりの消費金額の増加ではなく「EM 指数」の示す消費者数の増加であると考えられる。また、2020 年前半から感染症の影響により落ち込んでいた一人当たり消費金額と消費者数はともに、2021 年後半から回復してきていることが確認できる。

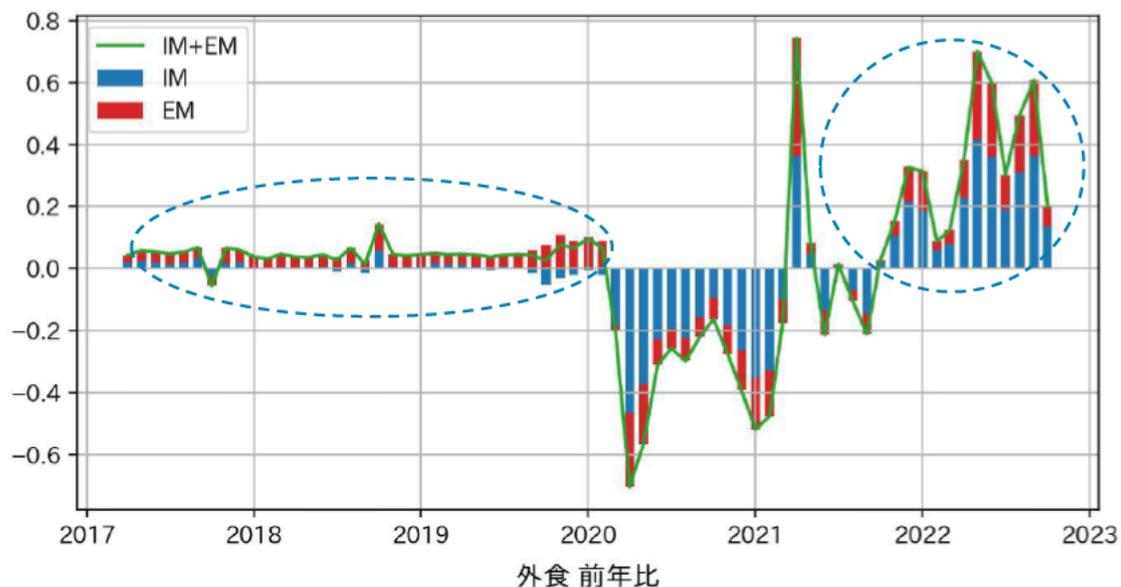


図 4-2 外食における「IM+EM 指数」

一人当たりの消費金額に関しては、想定ユーザーからも「観光事業において、単価（一人当たりの消費金額）を上げるための取組を実施しているため、その成果を確認したい」という意見があり、「IM 指数」と「EM 指数」の分解による分析は、有用であると考えられる。なお、想定ユーザーからは「物価高を考慮した実質的な消費動向の把握ができる」と望ましい」という意見もあり、この「IM+EM 指数」の活用によって、一人当たり消費金額が増加していることを確認できる場合、物価上昇の観点でも示唆が得られる可能性がある⁵³。

4) 地方における中小企業の売上動向の把握

既存統計の多くは全国指数のみが公表されており、地域粒度の細かい指数は公表されていないため、被災（例：地震）等における特定地域の中小企業の売上動向を把握することが困難である。本指標では、全国指数だけではなく都道府県別・地域別指数も併せて作成することができ、一例として、「居酒屋」の地域別指数を図 4-3 に示す。「居酒屋」では、2021 年序盤に全国的に売上が伸びており、特に南関東（埼玉県、千葉県、東京都、神奈川県）で際立っていることが分かる。このよう

⁵³ 本指標は現状、名目的な消費動向しか把握できないことに留意する必要がある。

に、地域別に中小企業の売上動向を把握できることで、特定地域・業種における中小企業の売上動向を確認することが可能となり、経済対策を検討するための材料となり得る。

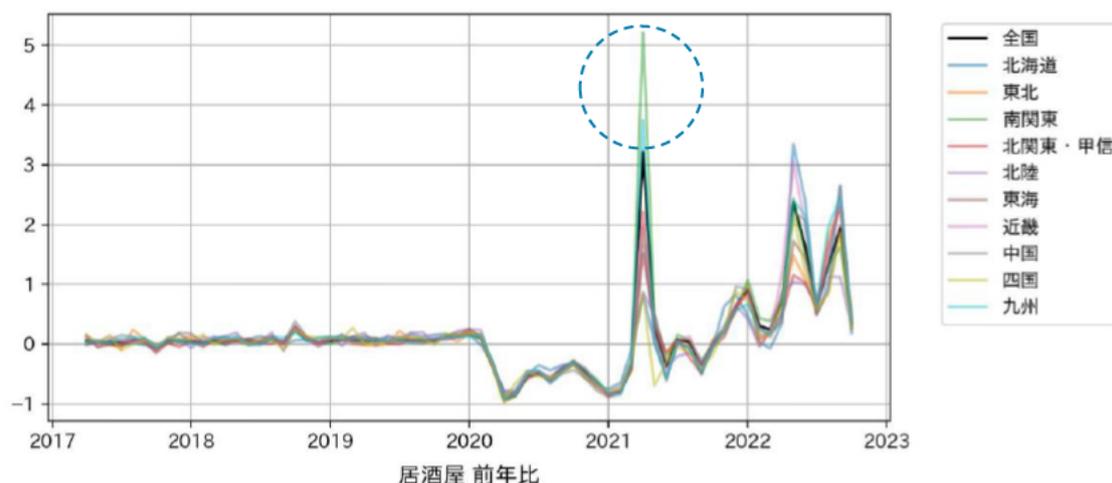


図 4-3 居酒屋の地域別指数

また、全国指数と都道府県別・地域別指数を活用することによって、特定都道府県・地域の中小企業における経営状態を相対的に把握することが可能となる。想定ユーザーからも「全国との比較だけではなく、他県との比較分析をしたい」という意見があり、当該地域の売上動向を相対的に把握することが可能となる。ひいては、当該地域の中小企業が自らの立ち位置を把握し、経営状態を自己診断するための材料となり得る。

なお、都道府県別・地域別に中小企業の売上動向を把握できるため、EBPM（証拠に基づく政策形成）の観点でも重要なデータと考えられる。想定ユーザーからも「過去に実施した施策の効果を検証したい」という意見があり、例えば、ある中小企業支援策のモデル事業を実施する場合、モデル実施地域とモデル未実施地域における売上動向や消費者数の比較をすることで、当該モデル事業の政策の進捗状況や地域間での違い等のモニタリングや、それによる業績評価に活用し得る。

(2) テキストデータを基に構造化した諸要素ネットワークの活用ユースケース

中小企業を取り巻く外部環境の構造化については、本事業では因果型共起構造分

析と因果構造分析を実施した。各分析から得られた示唆を基に「外部環境変化や影響の把握」、「特定経済事象の要因分析」、「潜在的な経済的関係性の検討」、「一般的（普遍的）因果事象の検討」の4つを構造化した諸要素ネットワークのユースケースとして示す。

1) 外部環境変化や影響の把握（因果型共起構造分析（ネットワーク分析））

外部環境の動向をモニタリングする場合や、経済全体への影響度が大きい、あるいは社会的注目度が大きい事象が発生した場合において、注目すべき事象（トピック）や業種を探索的に特定・把握したいときに因果型共起構造分析（ネットワーク分析）が有用である。一例として、「円安」トピックにおける因果型共起構造分析（ネットワーク分析）の分析結果を図 4-4 に示す。ウクライナ侵攻を契機として、「半導体不足」や「原油高」などの事象がみられる。

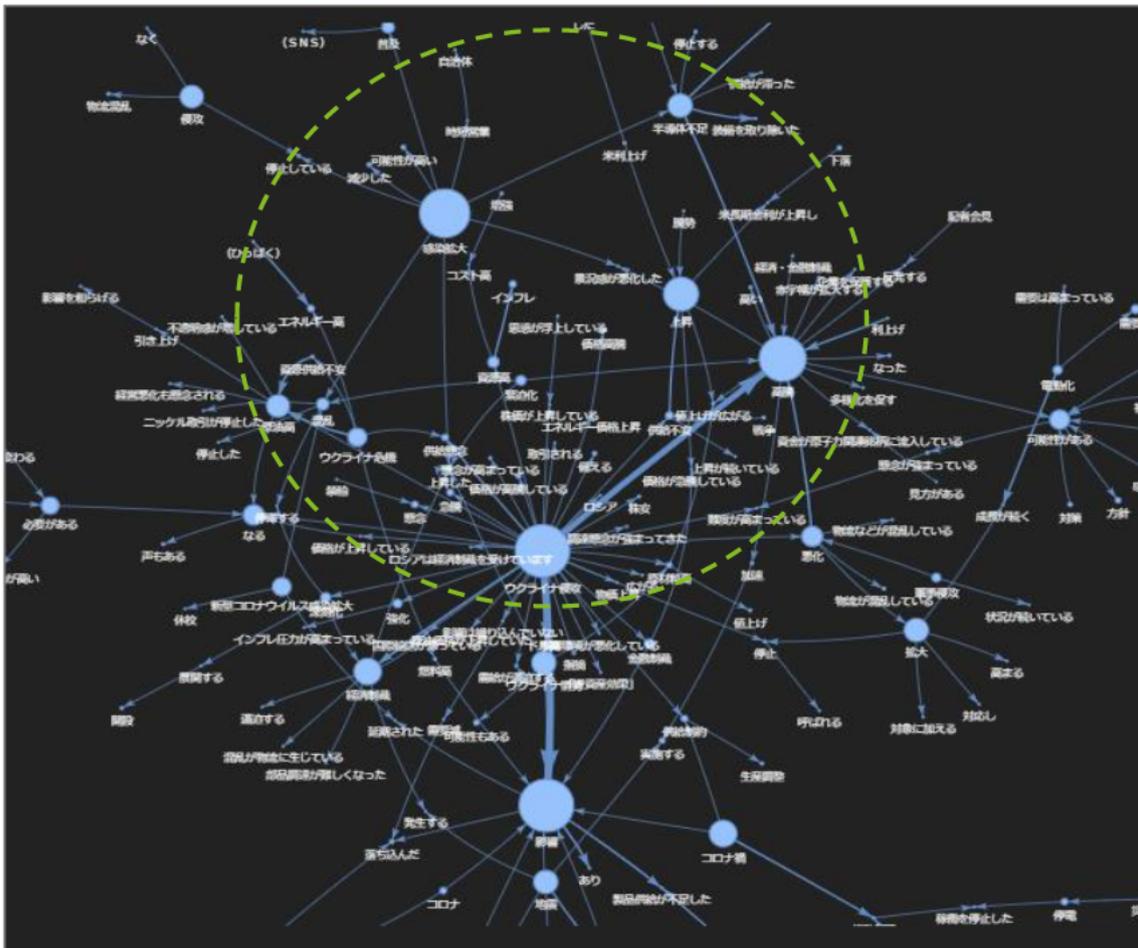


図 4-4 因果型共起構造分析（ネットワーク分析）の分析結果

また、経済的影響度や社会的注目度が大きい事象（トピック）や業種を特定するだけでなく、時系列に沿ってモニタリングすることによって、その影響度や注目度の推移を捉えることもできる。年別・月別における最大部分ネットワーク⁵⁴のノード数の時系列推移を図 4-5 に示す。これを見ると、新聞テキストデータにおける経済の全体像を垣間見ることが可能であり、かつ速報性の高い新聞記事テキストデータを基に、経済全体への影響（因果関係）が想定される注目すべき特定事象（トピック）や業種を抽出することが可能である。行政において、このような分析手法を仕組み化して定常的に活用することで、外部環境の変化をいち早く把握することが可能となり、より機動的な政策検討へと繋がる可能性がある。

⁵⁴ 特定の経済事象や関連トピックに関する記事から得られた因果ネットワーク全体において、連鎖するノード数が多いネットワーク。

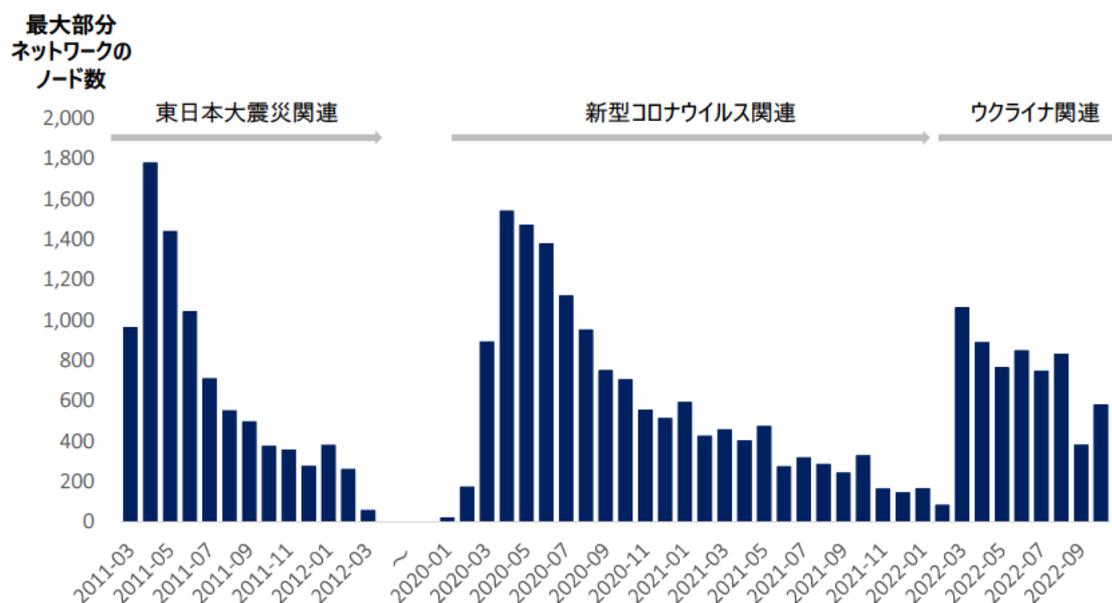


図 4-5 年別・月別における最大部分ネットワークのノード数の時系列推移

2) 特定経済事象の要因分析（因果型共起構造分析（絞り込み分析））

着目すべき特定事象（トピック）を把握している場合、その特定事象（トピック）の詳細な情報を把握したいときに因果型共起構造分析（絞り込み分析）が有用である。特定事象（トピック）の周辺情報として、何がどのような関係性で波及的に影響を及ぼしているのかを捉えることができ、特定事象（トピック）に対して多面的に認識することができる。本事業では、特定事象間の因果関係を明示的に示す解釈性の高い可視化を行うことを目的として、単語レベルに留まらず SVO（主語・動詞・目的語）に基づく文節レベルでのネットワークの可視化を実施した。このような分析手法を用いることで、探索的な因果関係の分析を具体的な事象に紐づいて効果的に実施することが可能になる。そのため本手法は、政策検討の前提となる経済事象の要因の分析に対する、テキストデータの側面からの有効なアプローチとして期待ができる。

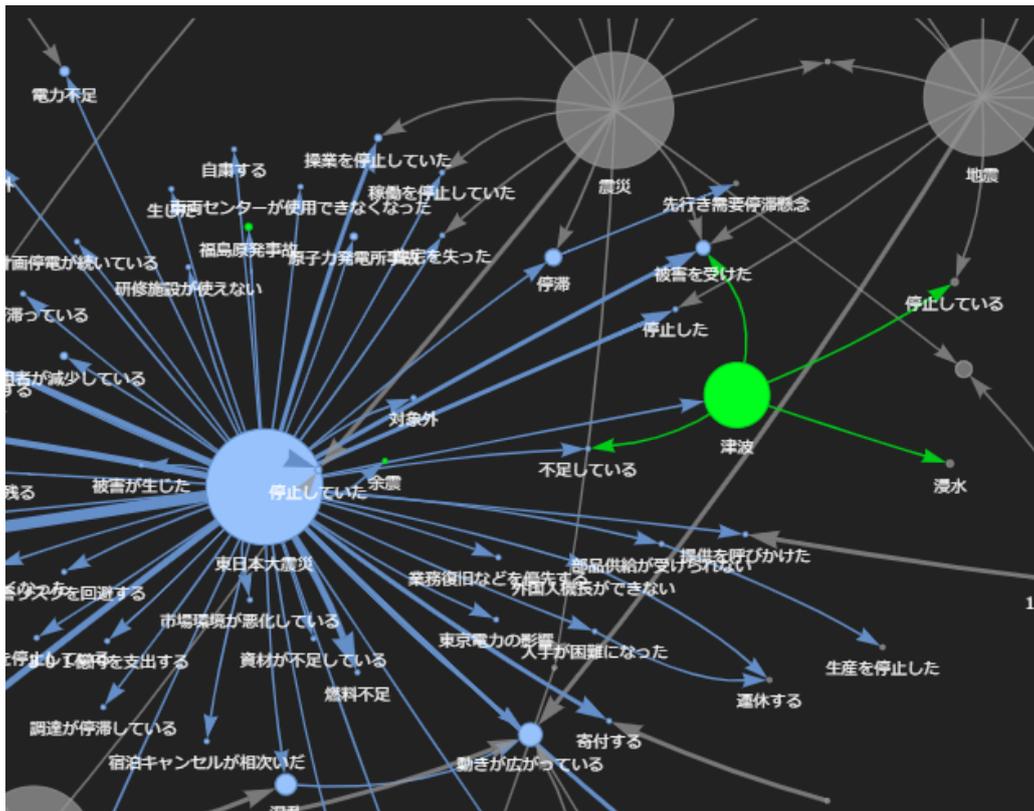


図 4-6 因果型共起構造分析（絞り込み分析）の分析結果

3) 潜在的な経済的関係性の検討（因果型共起構造分析（共起分析））

具体的な特定事象間の因果関係といった明示的な関係性ではない場合においても、特定事象間の同時発生的な、あるいは潜在的な関係性に対する示唆を得たいときに因果型共起構造分析（共起分析）が有用である。一例として、感染症における因果型共起構造分析（共起分析）の分析結果を図 4-7 に示す。因果型共起構造分析（共起分析）の特徴は重要度・頻出度が高い事象のノードが大きくなることであるが、感染症においては「中華人民共和国」や「企業」、「支援」といったノードが際立っていることが分かる。このように特定事象間の関係性を紐解く際に有益な情報が得られる可能性があり、また、特定の経済事象の背後にあるパターンを発見することも期待できる。

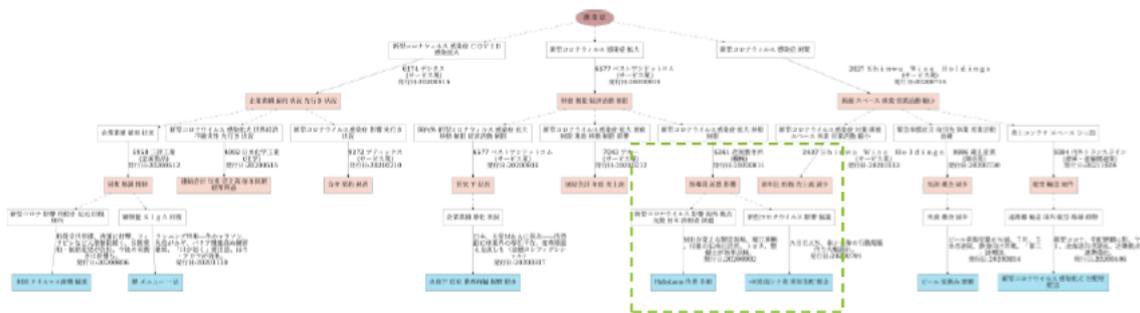


図 4-8 因果構造分析（因果チェーン絞り込み分析）の分析結果

4-2. 次年度以降のアクションプラン

我が国におけるオルタナティブデータの活用を定着させ、適切かつ迅速な中小企業支援のPDCAサイクルを確立するためには、行政における継続的なオルタナティブデータの活用方法を検討していく必要がある。また、前節で整理したオルタナティブデータを活用したユースケースについても、ユースケースを実現していくにあたり顕在化されている課題に対応していく必要がある。そこで、次年度以降のアクションプランとして、「行政における継続的なオルタナティブデータの活用方法の検討」、「ユースケースの実現に向けた対応」の2点を示す。

4-2-1. 行政における継続的なオルタナティブデータの活用方法の検討

オルタナティブデータの台頭や解析技術の発展により、これまでの既存統計等のデータだけでは得られなかった知見が国内外において蓄積されるようになった。行政においてこれらの知見を有効活用し、政策的な議論にオルタナティブデータを継続的に活用するためには、一般に「産学官の連携」「人材の確保・育成」に着目した検討も有用と考えられる。

(1) 産学官の連携

オルタナティブデータの活用は先進的な取組であり、現段階では行政に十分なノウハウや知見が蓄積されていない。そこで、行政において継続的なオルタナティブデータの活用を進めていくためには、データホルダーである民間企業と、分析結果の解釈などの技術的な支援を行う大学・研究機関と密に連携していくことが必要であると考えられる。

データホルダーである民間企業との連携に際しては、当該データの継続性を担保

することが重要である。民間企業は営利企業であり、データ提供がビジネスとして成り立つ必要があるため、データ提供のインセンティブを保持する仕組み化が必要である。また、民間データを基にした公的な利用については、信頼性を担保する必要がある、当該データの特性や課題等について説明責任を果たしながら活用することが求められる。一方で、民間データは企業秘密や競争戦略上の観点から、必ずしもデータの特性や課題を公開できないケースがあるため、情報の秘匿が必要な場合には、得られた情報や知見を取りまとめた行政記録情報の適切な管理など、当該データの信頼性を担保する方法を検討していく必要がある。加えて、クロスセクション方向のカバレッジを高めていくために、クレジットカードデータや新聞などのテキストデータだけではなく、銀行送金や企業会計、電力などの様々なデータを分析に取り入れることで、より中小企業の実態把握に寄与すると考えられる。そこで、これらのデータを保有する民間事業者と連携し、我が国のより有効な政策検討に資する活動を行っていくことが重要であると考えられる。

大学・研究機関との連携に際しては、本事業で得られた知見を共有し、オルタナティブデータの先端的な研究知見を取り入れながら、共同で取組を推進していく体制・枠組も考えられる。本事業においては、有識者の助言・指導等を受けながら遂行したところであるが、行政におけるオルタナティブデータの有効活用を継続して検討する上では、データサイエンティストなどの専門家との議論を重ねていくことが望ましい。

(2) 人材の確保・育成

政策議論におけるオルタナティブデータの活用には、オルタナティブデータの活用に長けた人材の確保が重要であり、外部人材の活用と内部での育成のそれぞれの方法が考えられる。本事業のクレジットカードデータを基に作成した指標とテキストデータを基に構造化した諸要素ネットワークにおいては、高い専門性と経験を有する外部人材を活用したが、その結果を解釈し、機動的な政策判断を実施する場面においては内部での育成人材が有益であることが考えられる。その上で、外部、内部といった専門性の高い人材の確保のあり方によらず、併せて行政内部にオルタナティブデータの活用に係るノウハウや知見を蓄積していく仕組みの定着も重要と考えられる。本事業の成果の共有だけではなく、データ分析に関する教育プログラムを提供することで、組織全体のデータリテラシーも向上させることができる。ひいては、継続的なオルタナティブデータの活用に対する理解を深めることができ、こうした行政におけるオルタナティブデータの知識水準の向上が推進力となって、オ

ルタナティブデータの活用がより進んでいくことも期待される。

4-2-2. ユースケースの実現に向けた対応

(1) 指標のユースケースの実現に向けた対応

2章で整理した課題を踏まえ、指標のユースケースの実現に向けて、指標の精緻化・高度化を進めていく必要がある。

1) 中小企業分類マスタを活用したデータ分類の精度向上

中小企業分類マスタを活用したデータ分類の精度向上に向けて、「中小企業分類マスタの精度向上」と「データ分類方法の改善」の2つの対応が考えられる。

1点目の「中小企業分類マスタの精度向上」については、中小企業マスタが保有する「従業員数」と「資本金」それぞれのデータ精度を向上させるために、経済センサスなどの公的統計や民間調査会社の提供データなどと組み合わせ、中小企業分類マスタに登録されている「従業員数」と「資本金」データを最新、かつ正確な情報として整備していく必要がある。中小企業分類マスタの整備については、本指標の実運用化を見据えて、中小企業分類マスタの点検などの運用体制も適切に検討していくことが望ましい。

2点目の「データ分類方法の改善」については、中小企業判定時の「従業員数」と「資本金」の条件変更や、突合するデータ（突合キー）の工夫などを検討し、中小企業分類マスタが保有する企業情報を最大限に活用していくことが望ましい。

また、「機械器具小売業」などの一部の業種において、業種における中小企業割合が実態（「経済センサス」における中小企業割合）と乖離している可能性があるため、その原因を特定する調査を実施することが重要である。例えば、中小企業分類マスタにおける当該業種の全体に占める売上シェアの確認等を実施し、実態と乖離している原因を明らかにすることで、当該業種における中小企業指標の算出ロジックの修正を行うことも可能となる。

2) 地域別指数の精度向上

地域別指数の精度向上に向けて、「地域別指数の検証」と「企業側（供給側）を基

にした指標試作」の2つの対応が考えられる。

1点目の「地域別指数の検証」については、サンプル数が少ない地域（例：北陸、四国）では不安定な動き（例：特定のイベントがない状況で変動幅が大きい動き）を示している地域別指数を確認できており、全国指数の動きからも大きく乖離している。この課題に対しては、サンプル数の確保や比較的大きい地域企業（加盟店）の登録状況等を確認し、乖離している要因を調査していく必要がある。また、中小企業の売上動向を示す地域別指数に関連する既存統計があまり存在していないことから、地域事情を適切に捉えた動きを示しているかどうかについて、地方自治体や地域金融機関などの関係者による定性評価を実施し、地域別指数の妥当性を検証する必要がある。加えて、地域独自のバイアス⁵⁵が存在している可能性もあるため、必要に応じて対応することが望ましい。

2点目の「企業側（供給側）を基にした指標試作」について、本指標は企業側（供給側）の所在地ではなく、消費者側（需要側）の居住地を基にした地域別指数であるため、宿泊やECといった業種において、企業の所在地と消費者の居住地が異なっている可能性があり、企業の所在地を基にした地域別の売上動向の変化を捉えることは困難である。この課題に対しては、企業側（供給側）のデータを基にした指標を試作することで、企業の所在地に基づく都道府県別・地域別の売上動向の変化を捉えることができる。ひいては、都道府県よりもさらに地域粒度の細かい市区町村別で指数を算出することも可能となる⁵⁶。想定ユーザーにおいても地域別指数は関心が高かったユースケースであるが、「市区町村単位で把握できれば、イベントや支援事業等の効果を測定することが期待できる」という意見もあり、更なる活用可能性が高まると考えられる。

3) データの網羅性向上及び追加的なバイアス除去・データ補正

データの網羅性という観点では、企業の売上に関連する決済手段として、クレジットカードだけではなく、銀行送金やQRコード決済などの決済手段の多様化が進んでおり、データソースを拡充することでカバレッジを上げた中小企業指標の作成を行う方向性が考えられる。また、今回はJCBカードの消費者の実取引データを基

⁵⁵ 地域独自のバイアスとして、QR決済など他の決済手段の加盟店開拓状況が地域ごとに異なるため、決済手段におけるJCBカードの売上シェアが地域ごとに異なるバイアスなどが想定される。

⁵⁶ 「企業側（供給側）を基にした指標試作」に係る課題点として2点挙げられる。1点目は、企業住所を用いる場合、「決済企業リスト・プラットフォーム企業リスト」に該当する企業の住所については住所を振り直す必要がある点である。2点目は、市区町村別にするによりサンプル数の確保が困難になるため、網羅的な作成はできず、地方でも観光地や都市部などに絞った市区町村別指数の作成になる点である。

にした指標であるが、JCB 社が提供する法人カードの実取引データや、銀行送金データなども活用することで、B to B の比重が大きい業種の売上動向把握が期待できる。加えて、現状の JCB カードの情報は国内消費者に限られるが、想定ユーザーからは「インバウンド消費の動向を把握したい」といった意見もあるため、海外の観光客に係る消費データを補完していく方向性が考えられる。

追加的なバイアス除去・データ補正という観点では、キャッシュレス化の進展などの影響と考えられるアッパーバイアス（過大評価）の除去や JCB カードの企業層や消費者を考慮したデータ補正、売上全体に占める JCB カードの決済割合を考慮したデータ補正が考えられる。売上全体に占める JCB カードの決済割合を考慮したデータ補正に関しては、例えば、「中小企業実態基本調査」の代表的な個社群⁵⁷の個票データを活用し、JCB カードのデータと比較することで、当該個社群においてどの程度 JCB カードが売上をカバーしているか検証し、そこから得られた示唆を踏まえたデータ補正を実施することが考えられる。

（2）構造化した諸要素ネットワークのユースケースの実現に向けた対応

構造化した諸要素ネットワークのユースケースの実現に向けて、「分析手法の精緻化・高度化」や「モニタリング体制の構築」に着目し、検討していく必要がある。

1) 分析手法の精緻化・高度化

本事業では、既存の因果文書抽出ライブラリにある手がかり表現を基に、因果文書の判定を実施した。今回、サンプリングチェックにより新聞記事テキストデータにみられる特有の手がかり表現を発見し、既存の因果文書抽出ライブラリに新たな手がかり表現として追加したことで、因果文書の判定ロジックの精度を高めたが、より精度を向上させていく余地があると考えられる。海外では、因果推論に関する自然言語処理データの品質を向上させることを目的として、因果文書の判定パターンを大規模に集積している事例がある⁵⁸。我が国においても、日本語特有の因果関係に関するテキストデータを膨大に収集し、因果文書の基準設定の精緻化・高度化に向けた取組を実施していく方向性が考えられる。

また、分析結果におけるノードの可視化に関して、語形は異なるが意味は互いに

⁵⁷ 「統計法第 33 条の運用に関するガイドライン」において、個々の調査対象である個社に関するデータを活用することが制限されているため、複数社をまとめた形で売上高を比較する必要がある。

⁵⁸ 海外の取組事例である CausalBank とは、因果関係の語彙パターンを収集し、3 億 1,400 万組の原因と結果の文章が含まれているデータセットである。（参照：<https://nlp.jhu.edu/causalbank/>）

類似している表現（例：「東日本大震災」と「震災」）が別ノード扱いとされてしまう。これらの表現をどのように類義語としてまとめていくべきか（例：どの程度一般化するか、個別事象として区別するか）をユースケースに応じて適切に検討・点検を実施していく必要がある。本事業では、既存の類義語辞書を活用したが、類義語辞書の点検や修正を行うことで、ユースケースに応じたより解釈性の高い可視化を実施できる余地があると考えられる。加えて、分析の基となるテキストデータについては、新聞記事テキストデータだけではなく、SNS データなど様々なデータが存在するため、これらのテキストデータへ活用範囲を広げることによって、より広範囲での環境変化を迅速に把握することに活用できると考えられる。複数のテキストデータについてその特性を把握し、分析に活用していくことで、テキストデータの側面から、我が国を取り巻く環境変化の迅速な把握とそれに基づく機動的な行政対応へ資するユースケースの確立が期待できる。なお、新聞記事に関して、本事業では日本経済新聞社のデータを活用した。新聞記事テキストデータを提供している新聞社は日経新聞社以外にも存在し、より客観性の高い分析を行うためには、複数の新聞社の記事を比較しながら活用することも検討する必要がある。

2) 継続的なモニタリングスキームの構築

本事業では試行的に、過去に発行された決算短信や新聞記事のテキストデータを基に分析を実施したが、定常的に外部環境変化を把握する際には、新聞記事データの速報性の高さを最大限に活かすことができると考えられる。そのため、行政においては、リアルタイムにデータを収集・分析を行うモニタリングスキームを構築する必要がある。新聞記事などの速報性の高いテキストデータを継続的に収集する仕組みと運用体制を構築することで行政へのテキストデータ活用を加速させることも、検討に値するテーマではないかと考えられる。

4-3. 本事業のまとめ

本事業では、オルタナティブデータの有効活用により、中小企業の経済活動や外部環境の変化をこれまで以上に充実して把握することを目的として実施した。

「中小企業の売上動向を把握するための調査・分析」においては、中小企業の売上動向を定期的に把握するための指標を業種別に 11 業種試作した。試作した指標については有効性・妥当性の評価結果等を踏まえ、指標の活用に関する見通しを確立

し、それを具体的なユースケースに落とし込むことができた。ユースケースにおける指標の活用にあたっては、一部課題や懸念事項を確認できているため、次年度以降に対応していくことが望ましい。加えて、想定ユーザーへのインタビュー等を実施し、指標の求められる役割やニーズを確認しながら、指標の精緻化・高度化を進めていくことは有用と考えられる。

「中小企業を取り巻く外部環境の構造化及びデータ利活用可能性に関する調査・分析」においては、因果型共起構造分析と因果構造分析においてテキストデータを基に構造化した諸要素ネットワークを試作した。構造化した諸要素ネットワークについては有識者への意見聴取等を踏まえ、具体的なユースケースを整理することができた。試行的な取組であるため、実用化にあたっては継続的なモニタリングスキームをどのように確立するかといった課題はあるが、引き続き行政におけるテキストデータとテキストマイニング手法の有効性を示しながら、より一層のユースケースの探索及び深化を進めていくことができれば、取組の拡大に繋がることが期待される。

上記の取組結果から、クレジットカードデータやテキストデータといったオルタナティブデータは、中小企業の実態把握を行う上で、詳細性や速報性の観点から公的統計を補完する役割が期待できる。また、外部環境を構成する個々の諸要素間の関連性などの体系的な理解にも寄与できるため、オルタナティブデータを活用することは非常に重要であり、その意義は十分に存在することを本事業で確認することができた。今後、オルタナティブデータの活用が進み、売上動向をはじめとする中小企業を取り巻く外部環境変化の影響について、的確かつ迅速な把握可能性の向上が期待される。

以上