

DP/18-1

経済財政分析ディスカッション・ペーパー

POS・テキストデータを用いた消費分析  
—機械学習を活用して—

小寺 信也・藤田 隼平・井上 祐介・新田 堯之

*Economic Research Bureau*

CABINET OFFICE

内閣府政策統括官（経済財政分析担当）付

本稿は、政策統括官（経済財政分析担当）のスタッフ及び外部研究者による研究成果を取りまとめたもので、学界、研究機関等、関連する方々から幅広くコメントを頂くことを意図している。ただし、本稿の内容や意見は、執筆者個人に属するものである。

## 内容

1. はじめに.....	1
2. 関連研究.....	3
2-1. POSデータを利用した関連研究.....	3
2-2. テキストデータを利用した関連研究.....	4
3. POSによる価格・数量変化の需要・供給要因の分析.....	6
3-1. データ.....	6
3-2. 分析手法.....	7
3-3. 分析結果.....	9
BOX 1 需要の強さの程度を表すPOS-DIの開発.....	16
4. POSデータを用いた小売業販売額のナウキャスト.....	18
4-1. データ.....	18
4-2. 分析手法.....	20
4-3. 分析結果.....	22
5. ディープラーニングによるセンチメント指数の開発.....	32
5-1. データ.....	32
5-2. 分析手法.....	34
5-3. 分析結果.....	38
6. まとめ.....	46
BOX 2 チョコレートのPOSデータからみる消費者の嗜好変化.....	47
参考文献.....	50
補論1 POSの品目分類と各系列の記述統計量.....	52
補論2 POSデータに基づく物価指標の作成方法.....	58
補論3. 天気と日次売上高の関係.....	60
補論4. 各主成分の代表品目.....	61
補論5. 分析コード.....	62
補論6. 新聞センチメント指数の分布.....	63
補論7. キーワードに含まれる単語.....	65

## POS・テキストデータを用いた消費分析\* —機械学習を活用して—

小寺 信也<sup>†</sup>・藤田 隼平<sup>‡</sup>・井上 祐介<sup>§</sup>・新田 堯之<sup>\*\*</sup>

### 【要旨】

本稿は、消費分析に対して新しい視点を提供することを目的として、POS・テキストデータをもとに機械学習の手法も活用しながら、3つの論点について検討を行ったものである。主な結果は以下の通りである。(1) 一般的な需要・供給曲線を想定し、POSの価格・数量データを用いて、価格・数量変化が需要要因と供給要因のどちらに起因するのかの要因分解を行った。価格・数量変化の詳細な背景情報を得ることができ、消費者の購買意欲等を把握できる可能性が提示された。(2) POSデータ等の速報性の高いデータから、機械学習により小売業販売額全体の動きのナウキャストを行った。POSデータのみでも、一定程度の精度を持つナウキャストを行うことは可能であったが、天候データも加えることで、ナウキャストの精度が向上するとの結果が得られた。(3) ディープラーニングの手法を用いて、新聞記事の内容がどの程度ポジティブであるかを示す指数を紙面別に作成し、消費者マインド等との相関関係の強さを検証した。全般的な経済情勢の変化や価格変化等の情報が多く含まれている紙面と消費者マインドの間には一定程度の相関関係が確認できたほか、記事の内容がネガティブに変化した場合の方が、消費行動に与える影響が大きい可能性が示唆された。

---

\* 本稿の作成において、内閣府の中村昭裕氏、増島稔氏、林伴子氏、茨木秀行氏、小中進悟氏から有益なコメントを頂いた。東京大学の渡辺努教授、久野遼平特任助教からは、学術的な見地からコメントをいただいた。株式会社大和総研の北野道春氏、渡部裕晃氏、参木裕之氏からは機械学習に関する助言を頂いた。ここに記して感謝を申し上げます。ただし、本稿に残された誤りはいうまでもなく筆者の責に帰すものである。また、本論文で示された見解は筆者の個人的なものであり、必ずしも内閣府の見解を示すものではない。

<sup>†</sup> 内閣府参事官補佐

<sup>‡</sup> 内閣府政策企画専門職

<sup>§</sup> 内閣府政策調査員

<sup>\*\*</sup> 内閣府事務官

## 1. はじめに

機械の性能向上に加え、データ処理に必要なソフトウェアの導入コストが大幅に低下したこと等を背景に、近年、ビッグデータや機械学習に関する研究が進んでいる。新しい分析データや分析手法が利用可能となったことから、これらを経済分析にも活用することで、これまでとは異なるアプローチで経済メカニズムを解明しようとする動きがみられている（例えば、日本銀行（2017）、経済産業省（2017）、内閣府政策統括官（2018）、藤田（2017）等）。

こうした近年の動向を踏まえ、本稿はPOSデータと新聞記事データをもとに、理論的なアプローチに加え、機械学習の手法も活用することで、消費者の購買意欲の把握、消費動向のナウキャスト、消費者マインドの背景等を分析する。POSデータを用いた研究では、渡辺・渡辺（2013）やUeda et al. (2018)等のように物価動向に焦点を当てたものが多く、消費動向をターゲットにした分析は少ない。POSデータには価格情報だけでなく、日々の売上情報が記録されることから、売上動向の情報も活用することで、消費動向をより迅速に把握できる可能性がある。また、Doms and Morin（2004）が指摘するように、新聞記事の内容は消費者のマインドと連動していることが考えられるため、消費動向の早期把握の観点からは新聞記事データも有益な情報となり得る。このようなテキストデータを機械学習により分析した例としては、五島他（2017）やHeston and Sinha(2016)等の金融市場にフォーカスしたものが多く、ここでは新聞記事から作成したセンチメント指数と消費者マインド及び売上高（POSデータ）の関係性を分析した点で既存研究とは異なる。

具体的には、以下の3つの論点から分析を行う。1点目は、渡辺（2016）を参考に、一般的な需要・供給曲線を想定し、POSデータにおける各品目分類の価格及び数量の変化の方向性をセットでみることで、価格・数量変化の背景が需要曲線と供給曲線のどちらのシフトに起因するのかについて分析を行い、その背景を考察する。

2点目は、POSや天候データ等の速報性の高い情報から、小売業全体の動き（経済産業省「商業動態統計」における小売業計の前年比）をどの程度ナウキャストできるのかについて、機械学習を活用することで検証する。分析に利用したPOSデータが小売業計に占めるシェアは限定的であるが<sup>1</sup>、POSデータの動きの中から、小売業計の動きを示唆する情報を機械に学習させることができれば、カバレッジの問題を乗り越え、ナウキャストを行うことが可能であると考えられる。

3点目は、新聞記事の内容が、消費者マインド・消費行動の変化に対してどの程度有益な情報をもっているのかの検証を行う。内閣府「景気ウォッチャー調査」を機械に学習させたモデルにより、18万件以上の各新聞記事の内容がどの程度ポジティブかを判断させ、新聞の「紙面別」に消費者マインドとの関係性を分析する。さらに、新聞記事の内容がプラスに

---

<sup>1</sup> 例えば、今回分析に利用したPOSデータの大半はスーパーにおける食料品であるが、商業動態統計のスーパーにおける飲食料品の販売額が小売業計に占める割合は7%程度である（2017年暦年）。

変化した場合とマイナスに変化した場合とでは、消費行動（POSデータの売上高）に与える影響が異なる可能性についても検証を行う。

本稿では、POSデータを用いたナウキャスト及び新聞記事のテキスト分析に際し、機械学習の手法を利用する。ナウキャストでは、線形回帰のほか、ニューラルネットワーク、ランダムフォレスト、勾配ブースティングといった機械学習の手法を活用し、精度の比較を行っている。特に、ランダムフォレストや勾配ブースティングは、アンサンブル学習と言われており、弱い学習モデルを組み合わせることで高いパフォーマンスを出せることで知られている。POS分析の目的は、データを予測する（ナウキャストする）ことにあるため、こうした予測精度が高い機械学習の手法を用いることが望ましいと考えられる。また、新聞記事の分析では、山本・松尾（2016）等、テキストマイニングの分析で利用されることが多い、長・短記憶（Long Short Term Memory: LSTM）ユニットを持つ再帰型ニューラルネットワーク（Recurrent Neural Network: RNN）（以下、LSTM RNN モデルと表記する。）を用いる。同モデルは、ディープラーニングの手法の一種であり、テキストデータのような要素の並び方（文脈）を考慮した学習を行うことができる。

本稿の主な分析結果は以下の通りである。1点目に、価格・数量変化が需要要因と供給要因のどちらに起因するのかの要因分解を行った。具体的にどのような品目分類が需要要因や供給要因で変化したのかを特定することが可能となり、価格・数量変化に対する詳細な背景を解明することができた。こうした分析を通して、消費者の購買意欲等を把握できる可能性が提示された。

2点目に、POSデータとその他の速報性の高い情報を用いて、経済産業省「商業動態統計」の小売業計の前年比をナウキャストできるかを機械学習の手法を用いて検証した。POSデータのみでも一定程度のナウキャストを行うことが可能であったが、天候に敏感な商品の売上動向を補正するために、天候データも加えて機械に学習させることで、ナウキャストの精度が向上するとの結果が得られた。

3点目に、ディープラーニングの手法を用いて、新聞記事の内容がどの程度ポジティブであるかを示す指数を紙面別に作成し、消費者マインドとの相関を検証したところ、全般的な経済情勢の変化や価格変化等の情報が多く含まれている紙面と消費者マインドに一定程度の相関が確認できた。また、作成した指数とPOSデータとの関係性を分析したところ、記事の内容がネガティブに変化した場合の方が、消費行動（売上高）に与える影響が大きい可能性が示唆された。

本稿の構成は以下の通りである。2章では関連する先行研究を整理する。3章・4章はPOSデータを用いた分析を行うが、3章は価格・数量変化を需要・供給要因に分解する分析、4章は小売業計をナウキャストする分析を行う。5章はディープラーニングの手法を用いたテキスト分析を行い、最後の6章で全体をまとめる。

## 2. 関連研究

### 2-1. POSデータを利用した関連研究

本稿の分析で利用するPOSデータとは、消費者がスーパーなどの小売店で商品を購入する際にレジで読み取られた情報が集計されたもので、ある商品における販売時点や価格・販売数量等の情報が記録されている。経済統計として利用する際には、無作為抽出ではないためサンプルが偏る可能性やノイズを含む可能性がある点には留意が必要であるものの、大規模かつ高頻度で速報性に優れる点は、従来の公的統計調査にはない利点である<sup>2</sup>。

こうしたPOSデータを用いた研究の中には、CPI等の既存の物価指数を補完し、物価動向の把握に役立てるための新指標を開発したものがある。例えば、渡辺・渡辺(2013)はPOSデータを用いて、ほぼリアルタイムで物価動向を観察することができる「日経・東大日次物価指数<sup>3</sup>」と呼ばれる指標を作成した。ただし、同指数は、丸山他(2015)が指摘するように、カバレッジ、品質調整、計算方法等の点においてCPIと異なっていることから、既存のマクロ統計と比較する際には留意が必要となる。その為、Imai and Watanabe(2015)をベースに、CPIと計算手法や銘柄選定基準を同様にした指数の作成も行われている<sup>4</sup>。

POS情報を活用することで、物価の変動要因を詳細に分析した研究も存在する。Ueda et al.(2018)は、1988～2013年のPOSデータを利用して日本における新商品導入時の価格変化の分析を行い、企業は新商品の導入によりデフレ下により低下していた価格水準を引き上げていたことを報告している。渡辺(2016)は、POSデータが利用可能な店舗毎に物価前年比と売上高前年比を計算することで、価格の変化が需要曲線と供給曲線のどちらがシフトすることで起こっているのかを考察している。

上記のように、POSデータを利用した研究には物価動向に焦点を当てたものが多いが、POSデータと景気や消費動向の関係性を分析している研究も存在する。上田他(2016)は、小売店における定価の改定頻度や特売の頻度が、失業率や総労働時間等のマクロ経済指標と相関関係にあることを報告している。藤田(2017)は、POSデータから作成した価格指数と消費者マインド・景気循環との関係性を分析し、家計が小売価格の変化に敏感になった可能性や、定価の改定や特売と景気循環との間には有意な相関関係がみられると指摘している。ただし、併せて、定価の改定や特売等の系列を束ねたヒストリカルDIを作成し、内閣府による景気基準日付との整合性を調べたところ、合致率は6割強にとどまったことも報告している。

さらに、POSデータに対して機械学習を活用した分析もみられており、例えば、Watanabe et al(2016)は、POSデータと天候データの双方を利用して販売数量を予想するモデルを推計している。彼らは、通常の線形モデルに加え、ニューラルネットワークによる推計も行い、

<sup>2</sup> 本稿で扱うPOSデータやテキストデータに関する利点や課題に関する議論については、岡崎・敦賀(2015)も参照。

<sup>3</sup> 現在は、日経CPINow・T指数に移行されている。

<sup>4</sup> 日経CPINow・S指数として月次ベースで作成されている。

後者の推計精度がより高いことを示している。

## 2-2. テキストデータを利用した関連研究

新聞記事等のニュースは消費者マインドに影響を与えることが指摘されており、例えば、Doms and Morin (2004) は、アメリカのデータを利用した実証分析により、ニュース媒体は3つのチャンネル<sup>5</sup>を通して消費者マインドに影響を与えていることを報告している。Hollanders and Vliegthart (2009)は、オランダにおいてネガティブな報道記事が消費者マインドを低下させていることを実証分析により示している。また、Soroka(2006)は、イギリスのデータを利用して、ポジティブなニュースとネガティブなニュースとでは消費者マインドに与える影響が異なり、人々はよりネガティブなニュースに反応していることを報告している。

文章等のテキスト情報を実証分析で使う場合には、何らかの方法でテキストデータを数値化することが必要となる。上記の先行研究でも利用されているように、機械学習が主流になる前は、特定の用語（例えば「不況」、「失業」等）の登場回数や、それらの用語を含む記事数を用いること等で記事のセンチメントを把握することが多かった。しかし、このアプローチの欠点としては、文脈情報が把握できないことにある。例えば、「景気が良いとは思わない」等の表現を「良い」という単語をキーワードに分析すると、この文章はポジティブと判断されてしまう。そこで、近年では、こうしたニュース記事を機械に学習させることでその内容を判断させ、数値化する動きがみられている。饗場・山本 (2018) は、従来型と機械学習（ディープラーニング）の双方の手法のどちらが、内閣府「景気ウォッチャー調査」のコメント正しく分類できるかを比較し、機械学習による改善を確認している。

このような機械学習とテキストデータを用いた分析は、主に金融市場を分析対象とした研究で比較的多く行われている。Heston and Sinha(2016)は、ニューラルネットワークを活用することで、90万件の記事データから株式のリターンを予測できるかを研究し、1週間分のニュースから作成したセンチメントが、1四半期先までの予測力があることを報告している。山本・松尾(2016)は、再帰型ニューラルネットワーク (Recurrent Neural Network: RNN) により内閣府「景気ウォッチャー調査」を学習させたモデルを用いて、内閣府「月例経済報告」や日本銀行「金融経済月報」の指数化を行い、日経平均とある程度の相関があることを確認している。五島他 (2017) は、景気ウォッチャー調査を畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) により学習させたモデルを用いて、150万以上の記事のスコア化を行い、ニュース記事が株・為替・債券等にどのような影響を与えているかを分析した。彼らの分析では、ニュース記事の内容が改善した時と悪化したときとでは株価・ボラティリティに与える影響が異なることも示されている。

---

<sup>5</sup> 彼らが指摘する3つのチャンネルは以下の通り。①最新の経済データや専門家の意見を伝達する。②経済記事のトーンや量が経済状況のシグナルとなる。③経済ニュースの量が消費者マインドをアップデートする。

さらに、機械学習によるテキストマイニングは、ナウキャストやフォーキャストにも活用されている。塩野（2018）は、物価関連記事からC P Iを予想するモデル（ニューラルネット）を作成し、高いパフォーマンスでC P Iのナウキャストを行うことができたことを報告している。また、Shapiro et al.(2018)は、文章の内容が特定の感情分類に分類される確率を計算する機械学習モデルを用いて新聞記事の指数化を行い、同指数を予測モデルに加えることでインフレ率等の予測精度が向上したと報告している。

上記のような機械学習により作成したセンチメント指数を用いて、消費者マインドとの関係性を分析したものとしては、内閣府政策統括官（2018）がある。同分析では、簡易なニューラルネットワークモデルにより景気ウォッチャー調査を機械に学習させたモデルを用いて、インターネット上にある「景気」を含む記事がどの程度ポジティブであるかについて指数化を行い、消費者マインドとのある程度の正の相関がみられたことを報告している。

### 3. POSによる価格・数量変化の需要・供給要因の分析

ここでは渡辺（2016）を参考に、POSデータの品目分類別の価格と数量の双方の情報をを用いて、購入価格や購入数量の変化の背景について分析する。具体的には、購入価格や購入数量の変化が需要ショックによるものなのか、供給ショックによるものなのかを識別することで、消費動向の把握に有用な背景情報を捉えることを目的とする。

#### 3-1. データ

本稿で使用するデータは、株式会社日本経済新聞社が全国のスーパーマーケット約 1200 店舗を対象に収集しているPOSデータ（以下、「日経POS」という。）から作成した売上高や価格指数である。対象となるスーパーは、基本的には株式会社日本経済新聞社と取引のあるチェーン店が対象であるが、一部に地場スーパーのデータも利用されており、ほぼ人口に比して全国均等にカバーされている。対象となる商品は、対象店舗で扱われる食料品（除く生鮮品）及び日用品の合計約 265 万商品である。この各商品を日経POSの大分類に合わせて 217 品目分類に統合している<sup>6</sup>。今回は、この日経POSをもとに、以下の2つの系列を 217 品目においてそれぞれ作成した<sup>7</sup>。

1つ目は、売上高の364日前比である。これは、POSデータをもとに現時点と364日前の両時点において存在する店舗を対象に品目ごとの売上高を合計し、伸びを計算したものである。前年同日ではなく364日前との比較であるのは、売上高が曜日による差が大きいため、同じ曜日を比較するように考慮したためである。

2つ目は、購入価格の前年同日比（T指数<sup>8</sup>）である。これは、POSデータをもとに各商品の価格変化率を求めたうえで、トルンクビスト（Törnqvist）ウェイトと呼ばれるウェイトを用いて統合した指標である。トルンクビストウェイトは現時点と前年同日の両時点における当該商品の平均値として求められる。

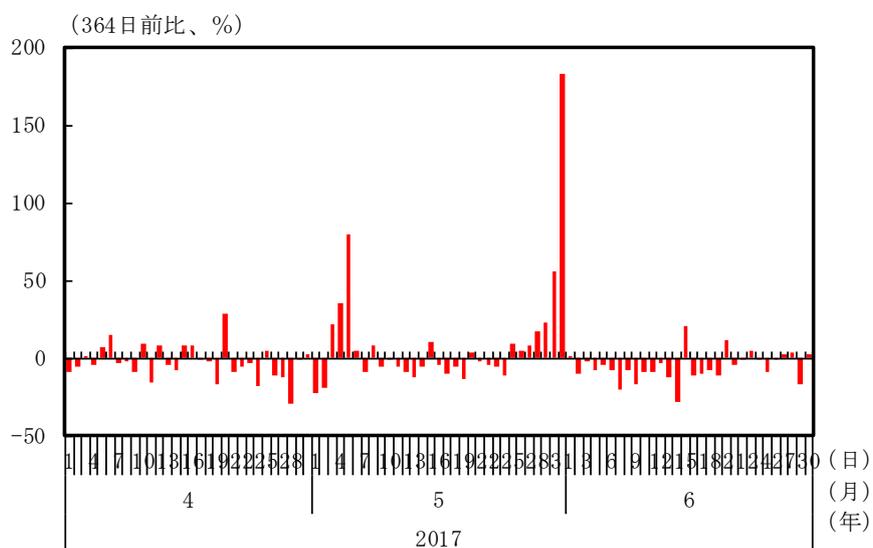
以上2つの系列は、いずれもデータ期間が1989年4月1日から2018年3月31日までの長期時系列であり、日次かつ品目分類別に利用することが可能である。こうした高頻度の強みを活かして、例えば、駆け込み需要等の動きを月次のデータよりも詳細に把握することができる。図1は、2017年の4月～6月におけるビールの日別の売上高を表したものである。ビール販売は2017年6月1日から酒税法改正等による安売り規制が行われたが、5月31日にかけてビールの販売額が大きく増加し、駆け込み需要が生じた様子を見て取れる。また、6月1日以降は364日前比マイナスの日々が数日に渡って続いている様子を確認できる。

<sup>6</sup> 品目の詳細及び各系列の記述統計量は補論1を参照のこと。

<sup>7</sup> 各系列の詳細な作成方法は補論2を参照のこと。

<sup>8</sup> T指数はFinatext傘下の株式会社ナウキャストが公表している指数であり、渡辺・渡辺（2013）による東大日次物価指数の後継指標。

図1 ビールの日次売上高（2017年4月1日～6月30日）



(備考) Finatextにより作成。

なお、以下のPOSデータを用いた分析では、購入価格は月次ベースのデータが利用可能であったため、月次ベースの値を利用した。一方、売上高のデータについては、日次ベースのみが利用可能であるため、以下の処理を行っている。まず、月次化には、単純平均を用い、20日以上がない月は当該月を欠測値とした。売上高については、日次データが364日前比を取っている関係で年末年始の変動が大きくなるため、年末年始の12月28日～1月4日までの計8日間のデータを除いた上で平均している。また、2003年と2004年の11月及び12月の期間は全ての系列が欠測しているため線形補間して用いるが、それ以外の期間における欠測値は補間せず欠測のまま扱うことにする。購入数量は、上記の要領で月次化した364日前比の売上高の値から購入価格の前年比を差し引くことで擬似的に計算した。

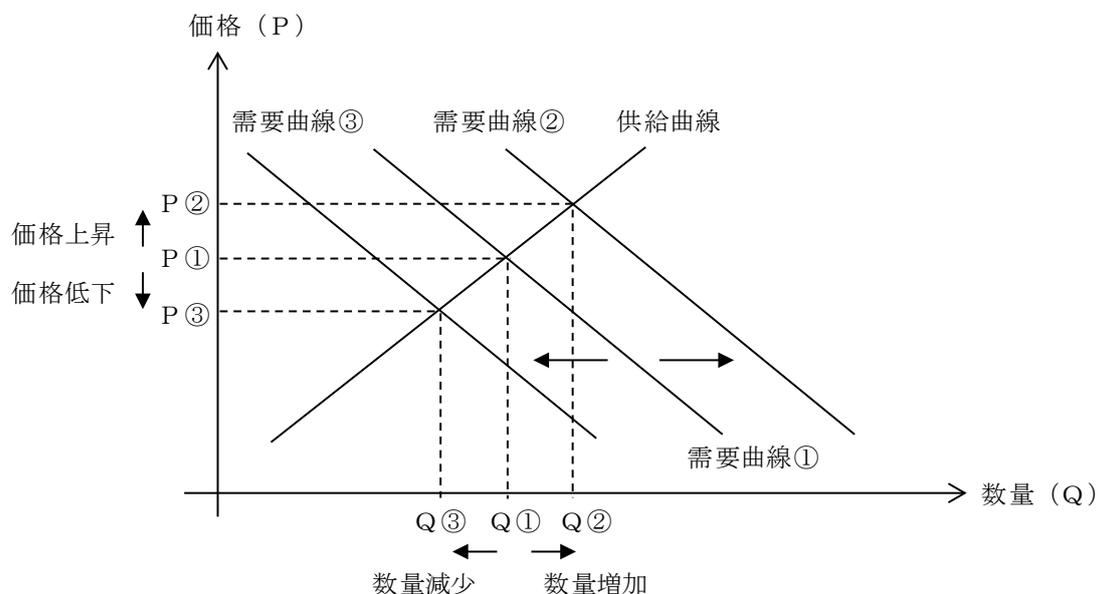
### 3-2. 分析手法

はじめに、価格変化のメカニズムについて、一般的な右下がりの需要曲線と右上がりの供給曲線から考察する。図2は、需要曲線のシフトによる価格変化のメカニズムを表したものである<sup>9</sup>。これをみると、需要曲線が右方シフトすると、数量の増加とともに価格上昇が生じている。逆に需要曲線が左方シフトすると、数量の減少とともに価格低下が引き起こされている。図には示していないが、供給曲線も同様で、供給曲線が右方シフトすると数量の増加とともに価格が低下し、逆に左方シフトすると数量の減少とともに価格は上昇する。まとめると、通常の需給曲線を前提とすれば、「価格上昇かつ数量増加」は需要曲線の右方シフト、「価格低下かつ数量減少」は需要曲線の左方シフトを表し、同様に、「価格上昇かつ数量

<sup>9</sup> ここでは、単純化のため、需要・供給曲線の片方のみがシフトすることを想定する。

減少」は供給曲線の左方シフト、「価格低下かつ数量増加」は供給曲線の右方シフトを表すことになる。

図2 需要曲線のシフトによる価格変化の概念図



こうした需給曲線のシフトは、直接は確認できないものの、POSデータは品目別に価格と数量の情報が揃っているため、ある品目について価格上昇とともに数量が増えたのか、減ったのか、価格低下とともに数量が増えたのか、減ったのかをみることで、間接的に価格変化が需要曲線のシフトによるものなのか、供給曲線のシフトによるものなのかを推察することができる。

数量・価格変化の背景を正確に把握しておくことは、消費動向を把握する上でも非常に重要である。例えば、価格の上昇が観察された場合、それが消費者マインドの改善等の需要要因によるものと、原材料価格上昇等の供給要因によるものとは、得られるインプリケーションが異なる。前者であれば、消費者の購買意欲の増加という景気に対してポジティブな可能性が高いが、後者であれば、コスト増を通じたネガティブな影響が懸念される。

そこで本稿では、POSの217品目分類別の購入価格と数量のデータを用いて、各商品分類を①価格上昇かつ数量増加、②価格上昇かつ数量減少、③価格低下かつ数量増加、④価格低下かつ数量減少の4つの分類に区分する。その上で、①と④を需要要因（需要曲線のシフト）、②と③を供給要因（供給曲線のシフト）と定義し<sup>10</sup>、購入価格と数量の前年比前月差

<sup>10</sup> これは、需要曲線と供給曲線のいずれか片方のみがシフトしたと仮定したものであるが、実際には需要曲線と供給曲線の両方が同時にシフトすることもありうる。もっとも、「価格上昇かつ数量増加」を例にとると、需要曲線の右方シフトに加えて、厳密には供給曲線がシフトして

の変化について、需要要因と供給要因それぞれの寄与度に分解する。なお、POSデータに含まれるノイズを均すため、3か月移動平均の前年比前月差の変化を基本とする。

また、今回利用したPOSデータは、前述の通りその時々を反映するトルンクビストウェイトを用いている。基準日時点で固定されたウェイトを用いる場合よりも、より現実の消費バスケットを反映することができる利点があるが、前年比前月差の寄与度分解を行った場合、各寄与は各品目分類のウェイトの季節性等にも影響を受けることになる。例えば、ある品目分類は、前年比前月差で価格が上昇しているにもかかわらず、季節性によりウェイトが小さくなることで価格全体に対する寄与ではマイナスになってしまう場合がある。この問題を解決するため、本稿では寄与度の算出における各品目分類の統合は常に2015年平均ウェイトを用いることで、ウェイト変化による影響を除去した。なお、この処理を行うことにより全体の購入価格（T指数）の変化と、寄与度分解の合計とは完全には一致しなくなるため、本稿では、両者の乖離があまり大きくない、2005年以降を主な分析の対象とした。

### 3-3. 分析結果

図3は、POSデータに基づく購入価格（T指数）と購入数量をクロスさせた4分類の割合の推移を表したものである。計算にあたっては、価格と数量の前年同月比前月差を利用しており、例えば、価格と数量の前年同月比前月差がともにプラスであれば「価格上昇かつ数量増加」にカウントする。

これをみると、例えば、2008年のリーマン・ショックに端を発する世界金融危機直後は、需要の増加による価格上昇を表す「価格上昇かつ数量増加」の割合が低下し、需要の減少による価格低下を表す「価格低下かつ数量減少」の割合が上昇しているほか、供給要因による価格低下を表す「価格低下かつ数量増加」の割合も上昇している。他にも、消費税率引き上げ前後には「価格上昇かつ数量増加」や「価格上昇かつ数量減少」を中心に需要要因と供給要因ともに大きな変動がみられるなど、時期により特徴的な動きがみ取れる。

続いて、上記の4分類をもとに、購入価格の前年比前月差を需要要因と供給要因の2つに分解したのが図4である。需要要因は「価格上昇かつ数量増加」に分類される品目のプラス寄与と「価格低下かつ数量減少」に分類される品目のマイナス寄与を合計したネットの値である。また、供給要因は「価格上昇かつ数量減少」に分類される品目のプラス寄与と「価格低下かつ数量増加」に分類される品目のマイナス寄与を合計した同じくネットの値である。

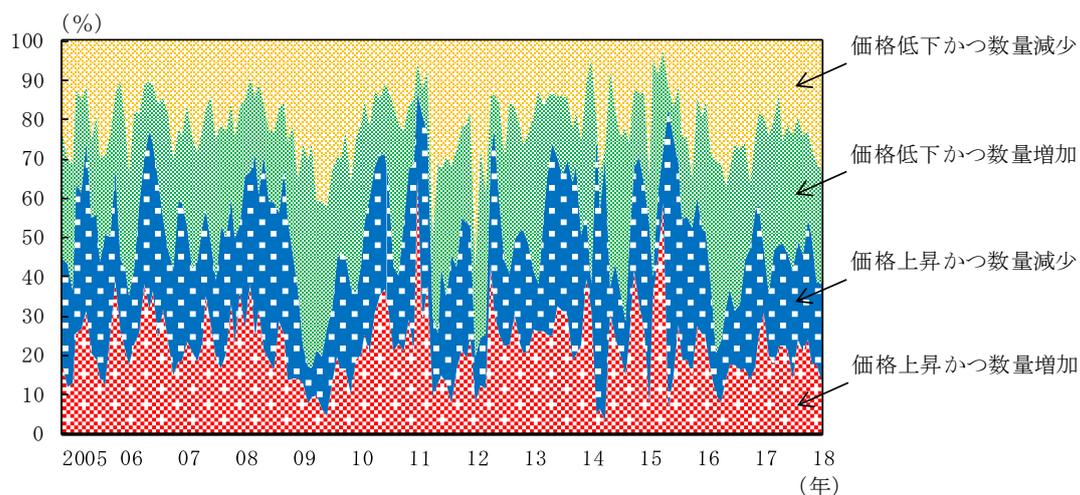
これをみると、例えば、2008～09年にかけての世界金融危機に際しては、需要要因の押し下げ寄与が高まっているのと同時に、供給要因も価格を押し下げていることがわかる。

---

いる可能性もある。もっとも、相対的に需要曲線のシフト幅の方が供給曲線のシフト幅よりも大きくなければ「価格上昇×数量増加」とはならないため、「価格上昇かつ数量増加」が需要曲線のシフトのみによるものとみなしても、エッセンスは失われないと考えられる。これは他の分類についても同様である。

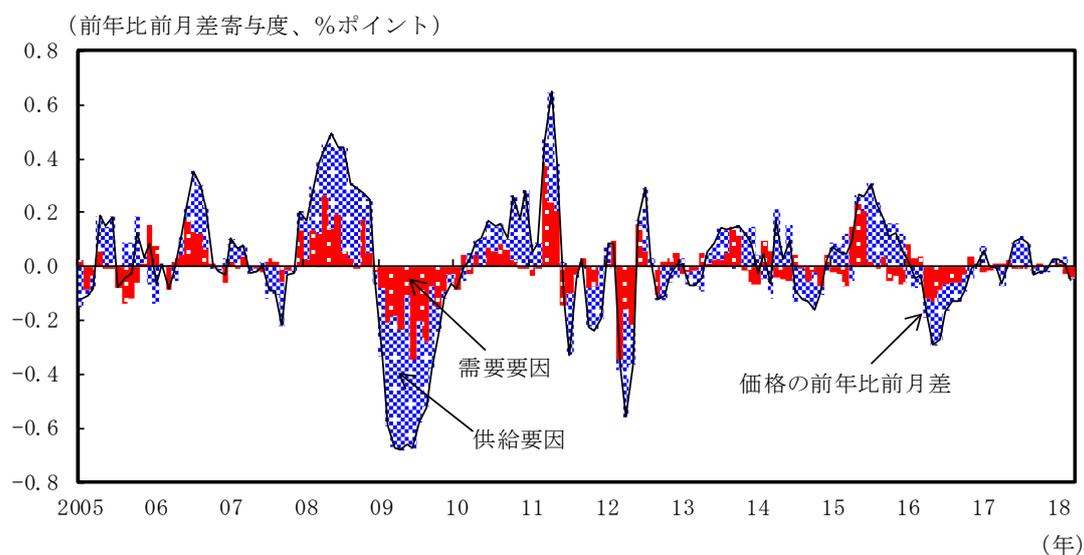
また、2013年以降の動向をみると、供給要因は2015年までは原材料価格や輸入物価の動向等を反映しておおむね押上げに寄与していたものの、2016年には押下げ寄与となっている。需要要因については、2013年～14年初にかけて押上げに寄与した一方、2016年にはマイナス寄与となっている。

図3 購入価格×購入数量の4分類の割合の推移



(備考) 1. Finatextにより作成。  
2. 価格と数量の後方3か月移動平均値をもとに作成。

図4 購入価格の前年比前月差の要因分解（2分類）

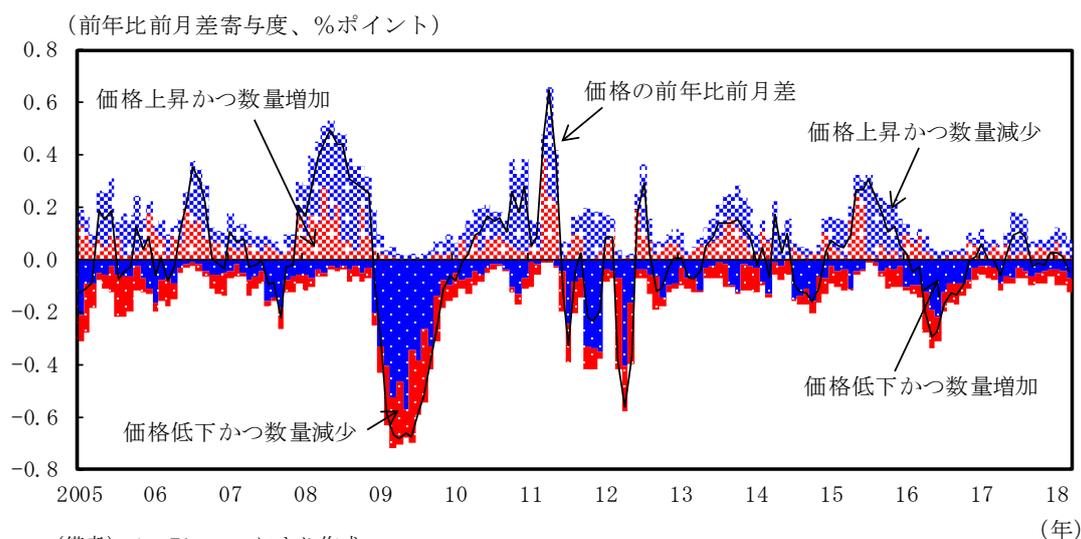


(備考) 1. Finatextにより作成。  
2. 価格と数量の後方3か月移動平均値をもとに作成。

こうした動きについてより詳しく確認するため、図4の2分類をさらに4分類の寄与度に分解したものが図5である。これをみると、例えば価格が大きく上昇した2008年には、「価格上昇かつ数量増加」と「価格上昇かつ数量減少」によるプラス寄与が拡大し、逆に価格が低下した2009年には「価格低下かつ数量減少」と「価格低下かつ数量増加」によるマイナス寄与が拡大している。この期間は価格全体の動きと逆方向の寄与が大きくないことから、多くの品目が似た動きをしていたことがわかる。

他方、2013年以降の動向をみると、2013～14年初や15年などの価格が上昇している期間にも「価格低下かつ数量減少」や「価格低下かつ数量増加」による押し下げ寄与がみられるほか、2017年初や18年初など価格が低下している際にも「価格上昇かつ数量増加」や「価格上昇かつ数量減少」による押し上げ寄与がある程度確認できる。この期間は価格の伸びが比較的安定しているが、その背景には、こうした価格全体とは逆方向に動く品目の存在があるとみられる。

図5 購入価格の前年比前月差の要因分解（4分類）



(備考) 1. Finatextにより作成。  
2. 価格と数量の後方3か月移動平均値をもとに作成。

次に、これまでみてきた2005年以降の購入価格の変化のうち、最も大きな低下幅となっている2008～09年について、特に押し下げ寄与の大きい供給要因に着目して、もう少し詳しく確認してみる。

表1は2009年の供給要因による価格低下品目の中から代表的な品目分類をみたものである。具体的には、2009年1月～12月において「価格低下かつ数量増加」に分類される品目のうち、各月においてマイナス寄与の大きい品目の上位30位内に登場した回数をカウントしてある。これをみると、バターやチーズ等の乳製品（図5の背景緑）やパンや麺

類等の小麦製品（図5の背景黄）が多くランクインしていることがわかる。2009年はバターや小麦粉等の原材料価格の伸びが大きく低下した年であることから<sup>11</sup>、コストの低下によって供給曲線が右方シフトしたことで、最終的な小売価格の低下や数量増加につながったものと考えられる。

表1 2009年に供給要因で価格の下がった主な品目

順位	品目名	登場回数
1	食パン	9
2	チルド半製品	7
2	バター	7
2	プロセスチーズ	7
2	乳飲料	7
2	野菜ジュース	7
2	マヨネーズ	7
2	菓子パン・蒸しパン	7
2	日用紙製品	7
10	つくだ煮	6
10	かまぼこ	6
10	水産練り製品・揚げ物	6
10	ヨーグルト	6
10	みそ	6
10	しょうゆ	6
10	ドレッシング	6
10	即席カップめん	6
10	テーブルパン	6
10	インスタントコーヒー	6
10	生鮮卵	6

（備考）Finatextにより作成

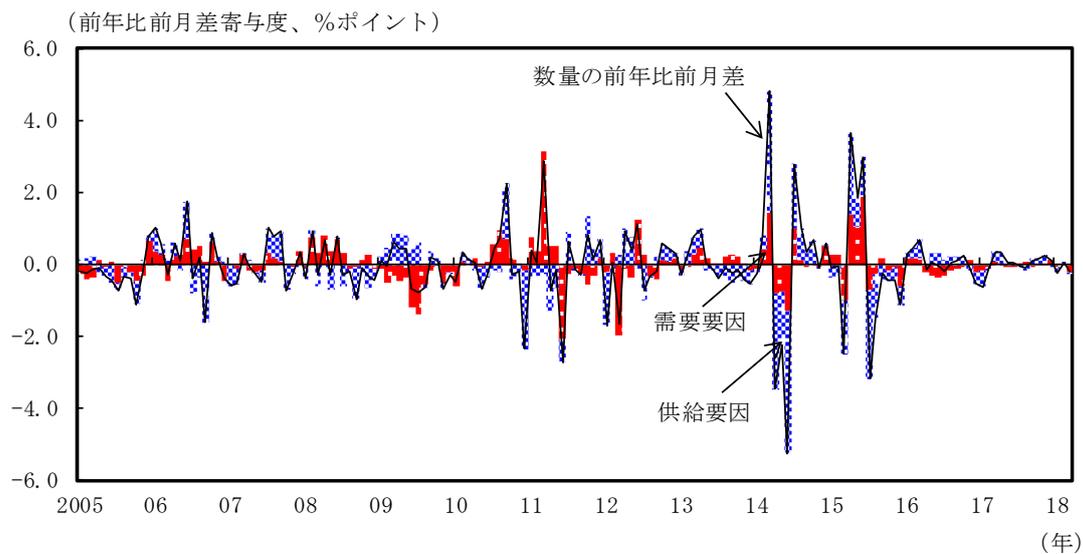
次に、購入価格の前年比前月差についても同様に、需要要因と供給要因の寄与に分解したのが図6である。これをみると、購入価格よりも変化幅は大きいのが、例えば、2011年の東日本大震災後に購入数量が増えたのは主に需要要因によるものであることや、2014年の消費税引き上げに際しては、需要要因だけでなく、供給要因による寄与も大きくなっていることがわかる。また、2016年以降は、需要要因がおおむねマイナス寄与となる中、供給側要因がおおむねプラス寄与となっている。

こうした動きについてより詳しくみるため、価格と同様に、図6の2分類をさらに4分類の寄与度に分解したものが図7である。これをみると、2005年以降多くの期間において常に一定程度のプラス寄与とマイナス寄与の品目分類の存在が確認でき、数量全体の方向感との連動性が購入価格よりも弱い姿をみてとれる。ただし、購入数量が大きく変化した2011年は「価格上昇かつ数量増加」によるプラス寄与が大きく拡大しているほか、2014年の消

<sup>11</sup> 日本銀行「企業物価指数」をもとに前年比前月差の動きをみると、2007～08年頃に生じた世界的な食料価格上昇の一服を受けて、2009年のバター価格は前年比前月差▲7.5%、小麦粉価格は同▲18.4%と大きく低下している。

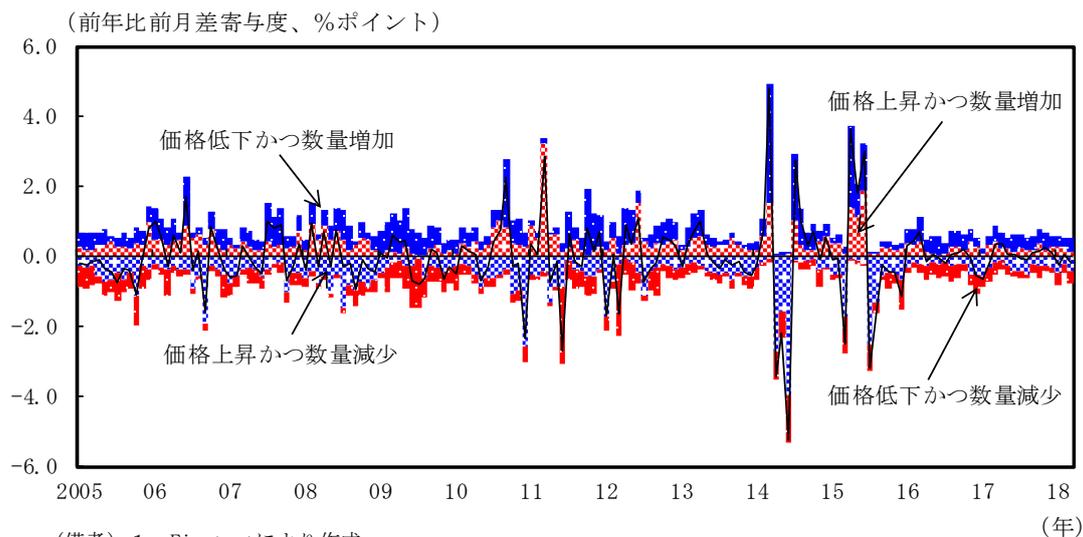
費税率引き上げ前の駆け込み期には「価格上昇かつ数量増加」と「価格低下かつ数量増加」によるプラス寄与が拡大し、逆に消費税率引き上げ後の反動減期には、「価格低下かつ数量減少」と「価格上昇かつ数量減少」によるマイナス寄与が拡大しているなど、購入価格と同様に、大きなショックがあった期間は多くの品目が似た動きを示す様子を確認できる。

図6 購入数量の前年比前月差の要因分解（2分類）



(備考) 1. Finatextにより作成。  
2. 価格と数量の後方3か月移動平均値をもとに作成。

図7 購入数量の前年比前月差の要因分解（4分類）



(備考) 1. Finatextにより作成。  
2. 価格と数量の後方3か月移動平均値をもとに作成。

次に、これまでみてきた購入数量のうち、2005年以降で最も大きく振れている2014年の消費税率引き上げ前後について、特に変化幅の大きい供給要因に着目し、もう少し詳しく考察する。

表2と表3は2014年の消費税率引き上げ前後の供給要因による数量増加及び減少品目の中から代表的なものをみたものである。具体的には、2014年2月～3月において「価格低下かつ数量増加」に分類される品目のうち、各月においてプラス寄与の大きい品目の上位30位内に登場した回数をカウントしたものが表2、2014年4月～6月において「価格上昇かつ数量減少」に分類される品目のうち、各月においてマイナス寄与の大きい品目の上位30位内に登場した回数をカウントしたものが表3である。

これをみると、消費税率引き上げの前と後のいずれにおいても、ビール等の酒類（表2、3の背景黄）やマヨネーズ等の調味料（表2、3の背景緑）のある程度日持ちのする品目が上位につけている。これらの品目は、食料品の中でも比較的駆け込み需要が生じやすい品目だとみられ、コスト要因による供給曲線のシフトというよりは小売店が消費税率引き上げ前の駆け込み需要を取り込むために、それらの品目の価格をあえて下げることで、需要を獲得しようとした可能性が示唆される。

表2 2014年2～3月に供給要因で数量の増えた主な品目

品目名	登場回数
栄養サポートドリンク	2
水産乾物	2
みりん・調理酒	2
食用油	2
マヨネーズ	2
つゆ	2
和風調味料・ソース	2
即席カップめん	2
ビール	2
焼酎類	2
穀類	2
洗髪剤	2
せっけん	2
紙おむつ類	2
洗濯用洗剤類	2
住居洗剤(クリーナー)	2
たばこ・喫煙関連用品	2

(備考) Finatextにより作成。

表3 2014年4～6月に供給要因で数量の減った主な品目

品目名	登場回数
コーヒー飲料	3
しょうゆ	3
だしのもと	3
つゆ	3
即席袋めん	3
水産缶詰	3
日本酒	3
ビール	3
焼酎類	3
発泡酒	3
紙おむつ類	3
女性用メーキャップ化粧品	3

(備考) Finatextにより作成。

以上のとおり、本章では、価格と数量の両方のデータを得られるPOSデータの強みを生かして、購入価格や購入単価の変化の背景について分析を試みた。具体的には、購入価格と購入数量を組み合わせることで、価格と数量の変化を「価格上昇かつ数量増加」と「価格低下かつ数量減少」から成る需要要因と「価格上昇かつ数量減少」と「価格低下かつ数量増加」からなる供給要因の2つに大別し、それぞれの動向を分析した。そのうえで、2005年以降の価格変化の中でも特に動きの大きいリーマン・ショック後の供給要因による価格低下や、数量変化のうち特に動きの大きい消費税率引き上げ前後の供給要因による数量変化を取り上げて、より詳細な考察を加えた。こうした分析により、購入価格や購入数量の変化が、消費者マインドの改善等によるポジティブな要因によるものなのか、コスト増加等によるネガティブな要因によるものなのかを把握することが可能となり、消費動向分析の一助となることが期待される。

ただし、本分析は、POSデータがカバーする店舗の食料品や日用品の動きに基づいたものである点には留意が必要である。今回用いたPOSデータには耐久財等が含まれていないカバレッジの問題があるほか、POSデータが対象としていない店舗や業態の動きを十分に捉えられていない可能性があり、今後の検討課題となるだろう。

### BOX 1 需要の強さの程度を表すPOS-DIの開発

第3章では、通常的需求曲線と供給曲線を前提に、POSデータの購入価格と購入数量の情報を用いて、価格と数量をクロスさせた4分類に識別を行った。4分類のうち、需要曲線のシフトによる価格と数量の変化を表すのは「価格上昇かつ数量増加」と「価格低下かつ数量減少」の2つであり、前者は需要の増えている品目分類の割合を、後者は需要の減っている品目分類の割合を表している。

理論的に、需要曲線のシフトは、消費者マインドの変化に加え、所得の変化や代替・補完関係にある他の財の需給バランスの変化といった複数の要因が複雑に絡み合って生じる。このため、需要曲線のシフトに着目することは、消費者マインドの変化以外の要因も踏まえた広い意味での消費者の購買意欲の把握につながる可能性がある。

そこで今回は試みとして、「価格上昇かつ数量増加」の割合と「価格低下かつ数量減少」の割合の差分を取ったDI（以下、「POS-DI」という。）を考える。需要の増えている品目分類の割合である「価格上昇かつ数量増加」の割合（＝「良い」割合）が、需要の減っている品目分類の割合である「価格低下かつ数量減少」の割合（＝「悪い」割合）を上回っていれば、幅広い品目において消費者の購買意欲の向上がみられると解釈できるため、例えば日銀短観のように両者の差分をとったPOS-DIは「需要の強さの程度」を表す指標とみなすことができる。

BOX図1-1は、購入価格と購入数量の前年同月比前月差をもとに作成した「価格上昇かつ数量増加」及び「価格低下かつ数量減少」の割合とPOS-DIの推移を表したものである。いずれの系列も月々の振れが大きい傾向がやや見にくいものの、例えば、世界金融危機下の2008～09年頃や2011年の東日本大震災後、2014年の消費税率引き上げ前後など、大きな需要ショックが生じた際には、POS-DIを中心におおむね連動した動きを確認することができる。

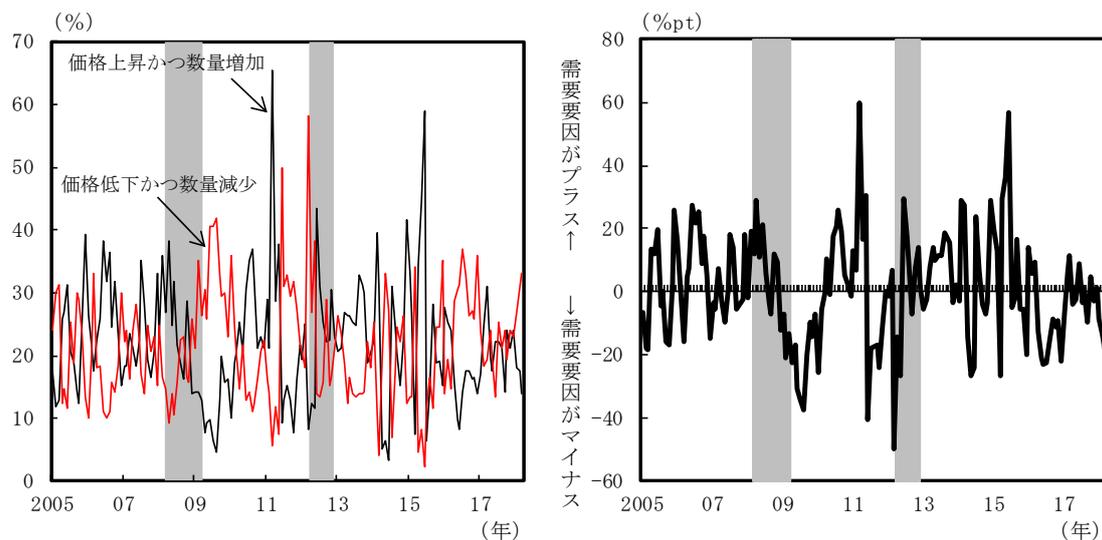
次に、購入価格と購入数量の4分類及びPOS-DIと消費者態度指数、景気一致指数との相関係数を確認する（BOX図1-2）。これをみると、景気動向を表す景気一致指数とはいずれの系列も弱い相関関係が確認できることから、POS-DIが表す「需要の強さの程度」は景気動向とある程度整合的であることが分かる。

他方、消費者マインドを表す消費者態度指数との相関係数はいずれもゼロ近傍となっており、明確な相関関係が確認できない結果となっている。POS-DIは需要曲線のシフトをもとに「需要の強さの程度」を抽出したものであるが、先述のとおり、そもそも需要曲線は消費者マインドの変化だけではなく、所得の変化等によってもシフトするため、POS-DIの変化には所得等のマインド以外の変化による影響が含まれている。また、消費者態度指数についてはコスト増加による値上げ等の供給要因の影響も受ける。このため、需要要因の動きを幅広く捉えたPOS-DIと消費者態度指数とは必ずしも連動するわけではないと考えられる。

POS-DIは消費動向から逆算して消費者の購買意欲を捉えた指標であり、消費者ア

アンケートから作成される既存のマインド指標とは異なる性格を持っている。既存のマインド統計に加えて、こうした独自性を有する指標を補完的に活用することで、消費者の購買意欲の動向をより正確に把握できる可能性がある。

BOX図1-1 各需要要因とPOS-DIの推移



(備考) 1. Finatextにより作成。  
2. シェード部分は景気後退期。

BOX図1-2 消費者態度指数及び景気一致指数との相関係数

	消費者態度指数	景気一致指数
POS-DI	0.09	0.26
割合(価格上昇かつ数量増加)	0.09	0.23
割合(価格上昇かつ数量減少)	▲ 0.01	0.16
割合(価格低下かつ数量増加)	0.01	▲ 0.16
割合(価格低下かつ数量減少)	▲ 0.09	▲ 0.27

(備考) 1. 内閣府「消費動向調査」、「景気動向指数」、「景気ウォッチャー調査」、Finatextにより作成。  
2. 相関係数の計算期間は2005年1月～2018年3月まで。

## 4. POSデータを用いた小売業販売額のナウキャスト

前章では、POSデータを活用し、購入価格や購入単価の変化の背景について分析を試みたが、本章では、POSデータがカバーしていない耐久財等の売上高を含んだ財消費全体の動向が把握できる経済産業省「商業動態統計」の「小売業計販売額」（以下、単に「小売業計」という。）について、同じ日経POSを活用して予測することを試みる。カバレッジが限定的なPOSデータから小売業計を予測するためには、財消費全体の動きと似通った情報をPOSデータの動きの中から抽出する必要がある。ここではこの抽出作業を機械に行わせることで、人間では気づかないようなデータの特徴を機械に把握させ、高い精度の予測を試みる。具体的には、まず、POSデータのみで小売業計を予測する。予測には、複数の機械学習の手法を活用し、予測精度の高さを比較する。その上で、POSデータ同様に速報性のある天候や自動車販売台数といったデータを加え、予測精度が向上するかを検証する。

### 4-1. データ

ここでの予測目的変数は経済産業省「商業動態統計」の小売業計である。同統計は、供給側（販売側）の観点から財消費の動向を捉えるマクロ経済指標の一つとして広く用いられている。統計の公表（速報）は概ね翌月末と1か月程度のラグがあることから、ナウキャストする意義は十分にあると考えられる。今回は、POSデータが前年比であることを踏まえ、小売業計の前年比を目的変数とする。

POSデータは、前章でも利用した日経POSを用いる。小売業計が月次統計であることから、POSデータの売上高の364日前比の日次データを単純平均で月次化する。POSデータは217品目分類があるが、それらは互いに強く関連しており<sup>12</sup>、そのまま全ての品目を学習させると多重共線性<sup>13</sup>が発生し、結果として過学習<sup>14</sup>を起こしてしまう可能性がある。昨今の機械学習の分野では過学習を回避する研究も進んでいるが、本稿ではPOSデータに対し主成分分析を実施し、POSデータの主成分を抽出することで、この問題に対処することとした<sup>15</sup>。具体的には、主成分分析によって、217品目分類のPOSデータから120の主成分を抽出（合成変数を作成）している。以降、本稿の分析で使用するPOSデータは全て

---

<sup>12</sup> 例えば、「歯磨き類」と「台所食器洗剤」の相関は0.89と非常に高い。「歯磨き類」は0.80以上の相関を有する品目を他に7品目もあり、このように互いに強い相関を有する品目がPOSデータには多く存在する。

<sup>13</sup> 変数間で相関が高いと、一方の変数で他方の変数がある程度表せてしまい、回帰分析等をする際に、説明変数の係数が実態に即さない信頼性の低いものとなる現象。

<sup>14</sup> 学習データを学習しすぎた結果、判断の基準が厳しくなり、予測する際に少しでも異なるパターンを誤って判断してしまう状態。

<sup>15</sup> 217品目分類をそのまま使用して過学習を回避する分析も実施したが、本稿の結果よりも精度が低いものであった。分析内容としては、Lasso回帰を使用し、各品目に過学習を抑えるようペナルティ付けをして予測した。しかしながら、後述する課題の通り、ウェイトのない前年比を分析している影響が強く、互いに相関性が高い多くの品目の特徴を差別化しきれないことが、結果して各品目へのペナルティに差が出にくいものとなり、過学習を回避しきれなかったと考えられるため、本稿では主成分を抽出する方法を選択した。

主成分分析の結果を活用したものである。これにより、具体的な品目ではなく、よりスーパーマーケットの高次元的な動向<sup>16</sup>を少ない変数で読み取れることができると考えられる。

また、日経POSはスーパーマーケットの売上動向を表したものであるが、予測したい小売業計には自動車小売業、機械器具小売業（家電等）、織物・衣服・身の回り品小売業等の他の業種も含まれている。このため、POSデータのみで小売業計をナウキャストすることは困難な可能性がある。そこで、速報性が高く、小売業計に影響を与える情報をPOSデータと合わせて機械に学習させることで、予測精度を向上させることが考えられる。この条件を満たすデータとして、ここでは天候と自動車販売台数を使用する。日本銀行（2016）でも示されているように、特に天候は衣類や家電等の売上動向を左右する重要な要素であることが考えられるため、POSデータを補正し予測精度の向上に寄与すると期待される。

天候データは、日本全国の都道府県庁所在地の各天候データを月次で抽出<sup>17</sup>し、2016年の商業動態統計の都道府県別の販売額（百貨店・スーパー別）をウェイトに加重平均した後、前年比を算出した<sup>18</sup>。具体的には、平均気温、最低気温、最高気温、降水量、降水日数<sup>19</sup>、積雪日数<sup>20</sup>の6種類の天候データを抽出し、予測の精度向上を検証した。

自動車販売台数は、新車販売台数が翌月の初旬に公表される速報性の高いデータである。販売金額ではないが販売台数でも概ね同じ傾向を把握できると想定し、月間販売台数の前年比を使用する<sup>21</sup>。なお、軽自動車を含む全車種を対象としている<sup>22</sup>。

なお、本稿の分析は、2007年1月～2016年12月までの10年間を学習期間とし、2017年1月～2018年3月を予測期間とする。POSデータの主成分分析も上記学習期間で分析し主成分を抽出した。機械学習の性質上、データ数は多い方が望ましいものの、POSデータの集計対象となる店舗数の変化や、店舗内での商品構成の変化が考えられるため、時期によってマクロ経済指標と似た動きをする品目や主成分が変化している可能性が考えられる。また、使用したPOSデータに含まれているスーパーマーケットの販売環境の変化等により、主成分の強さが変化していることも確認されている<sup>23</sup>。こうした点を踏まえ、時期によ

---

<sup>16</sup> 例えば、アイスクリームやビール等の動きから、「気温に敏感な商品」の動向という成分を抽出することで、具体的な品目から消費者の購入志向の傾向へ動向の次元を発展させている。

<sup>17</sup> 気象庁「過去の気象データ」を使用。

<sup>18</sup> なお、最大の販売額である東京の天候のみを使用した分析も実施したが、結果は全国ウェイトを使用した天候に及ぶものではなかった。

<sup>19</sup> 日降水量 0.0mm 以上の日数。

<sup>20</sup> 日最深積雪 0cm 以上の日数。

<sup>21</sup> 日本自動車販売協会連合会、全国軽自動車連合会により作成。

<sup>22</sup> 自動車小売業には中古車販売も含まれることから、中古車を含めた販売台数の分析も実施した結果、中古車を含めた方が僅かであるが予測精度が高まった。ただし、中古車は翌月の中旬～下旬に公表される。今回予測する商業動態統計とほぼ同時期の公表となるため、ナウキャストの観点から今回は新車販売台数のみとした。

<sup>23</sup> 例えば、「アルコールテイスト飲料類」「家庭医療用品」「介護・衛生用品」等の品目分類の成分を中心とした主成分が抽出された。1989年4月～2017年12月までの28年間ではこれらの品目は、POSデータの2.4%程度しか説明できない主成分であったが、本稿の学習期間だけでみ

る環境変化も大きくは存在せず、かつ、機械学習にもある程度耐えうる期間と考えられる 10 年間のデータで分析を試みる。

## 4-2. 分析手法

本稿の分析では、大きく 4 つの手法を用いて予測を試みる。従来からある線形回帰、ディープラーニング<sup>24</sup>の原型に当たるニューラルネットワーク、アンサンブル学習として代表的なランダムフォレスト、勾配ブースティングである。

### (1) 線形回帰（重回帰分析：ステップワイズ法）

線形回帰は一般的な重回帰分析を使用する。ただし、多数の説明変数がある場合、重回帰分析は過学習を起しやすく、精度が良い回帰モデルを構築するためには変数選択を行う必要がある。今回は、モデル選択の基準として広く知られる A I C を基準に、ステップワイズ法を用いて説明変数を機械的に選択する。A I C は式 2 で定義される値で、モデルの複雑さとデータとの適合度を評価するものである。

$$A I C = -2 \times \log (\text{最大尤度}) + 2 \times (\text{パラメーター数}) \cdots (\text{式 2})$$

A I C の値が小さいモデルが一般的に良いモデルとされることが多い。ステップワイズ法は、指標（本稿では A I C）が最も小さくなるように、説明変数の数を増減させ、取捨選択を行うアルゴリズムである。

### (2) ニューラルネットワーク

ニューラルネットワークは、主にパターン認識・分類を得意とした手法である。ここで利用したニューラルネットワークは、ディープラーニングを扱った次章と異なり、比較的シンプルなモデルである。モデルは、「入力層－中間層（隠れ層）－出力層」から構成され、入力層に説明変数を入れることで、出力層において予測値の推計結果が出力される。学習は、この出力層における予測値の精度を向上させるように、中間層において特定の値に重みを付けたり、値を偏らせる等の調整が行われる。分析する際は、学習の回数や収束に影響を与える中間層と呼ばれる層の数を設定する必要があるが、今回は統計解析ソフト<sup>25</sup>に、より最適なパラメーターを自動で算出させる。

---

ると 18.8% を説明する主成分となっている。

<sup>24</sup> 現在、多くの画像認識や音声認識で使用される手法。画像コンテストである ImageNet Large Scale Visual Recognition Challenge 等で数年優勝する等、高精度の学習結果を示す。ニューラルネットワークの中間層を複数層にした手法である。

<sup>25</sup> 統計解析ソフト「R」のパッケージ「caret」を使用した。

### (3) ランダムフォレスト

ランダムフォレストはアンサンブル学習の代表的な手法である。アンサンブル学習とは、弱い学習モデルを組み合わせることで高い精度のモデルを構築する学習手法であり、ランダムフォレストは tree (回帰木) と呼ばれる弱い学習モデルをベースにしたアンサンブル学習の手法である。tree とは特徴量を用いてデータを複数に分割し、分割したセグメントの平均値を予測値にする手法であり、線形回帰ではとらえることのできない関係を抽出することが可能な手法である。ただし、データセット間における推定結果にばらつきが大きいという問題点が指摘されている。ランダムフォレストは、この問題に対処するため、復元サンプリングとランダムで選択した変数で tree を独立して大量に作成することを繰り返し、各 tree の予測値を平均値とする手法で、精度が非常に高いことで知られる。これは、複数の tree を組み合わせた結果、説明変数の選択において制約が非常に少なくなり、過学習の影響を排し多くの変数を評価できるためと考えられる。ランダムフォレスト  $RF(X)$  は以下の通りである。

$$RF(X) = \frac{1}{M} \sum_n^M T_n(X) \quad \dots (式3)$$

ここで、 $X$ はランダムに抽出したN個の説明変数からなるベクトル、 $T_n(X)$ は各 tree モデル、 $M$ は復元サンプリングの回数<sup>26</sup>である。なお、本分析では、N個の説明変数からランダムに $\sqrt{N}$ 個の説明変数を選んで学習させている。

### (4) 勾配ブースティング

勾配ブースティングもアンサンブル学習の一つであるが、弱い学習モデルを独立して構築したランダムフォレストと異なり、1つずつ順番に構築していく点に特徴がある。次の学習モデルを前回の学習モデルの残差に適合させようとする手法である。具体的には、弱い学習モデルの推定結果の予測値と被予測値の残差を目的変数にして、次の学習モデルを構築する。まず、誤差の初期値 ( $\gamma_0$ )<sup>27</sup>を目的変数として、予測結果 $f_1(x)$ を算出する。その次の学習モデルの目的変数には式4で計算される誤差 ( $\gamma_1$ ) を使用する。

$$\gamma_1 = \gamma_0 - \lambda f_1(x) \quad \dots (式4)$$

$\lambda$ は学習速度 (learn rate, shrinkage) で  $0 < \lambda < 1$

$\gamma_1$ を目的変数にした学習モデルの予測結果を $f_2(x)$ として算出し、同様に $\gamma_2$ を作成する。以下

<sup>26</sup> 本分析では、1,000回と設定し分析を進めた。

<sup>27</sup> 初期値の $\gamma_0$ は被説明変数 ( $y$ )と等しい ( $\gamma_0 = y$ )。

同様のプロセスを最終モデルまでのN回について、 $\gamma_n$ を更新するステップを繰り返し、最終モデルは $f_1(x) \sim f_N(x)$ のN個のモデルの統合により構築される。なお、学習速度 $\lambda$ は、各モデル構築での学習度合いを小さくするために乗算する値であり、任意に設定する。この値が小さいほど、より多くのモデルを構築する必要があるため精度が高くなる傾向がある反面、モデル構築に時間を要する。今回の分析では、統計解析ソフト<sup>28</sup>に、より最適なパラメータを自動で算出させる。

勾配ブースティングは残差に着目するため、異常値のあるデータや極端なデータがあると過学習を起こす可能性が高まるが、チューニング方法も多様にあり、分析手法としては非常に高い精度を出すことで知られる<sup>29</sup>。なお、勾配ブースティングで一般的に使用される弱い学習モデルはtreeであるが、linear（線形回帰）を使用することもできる。Müller (2018)が指摘するように、両者において明確な優劣はないため、今回は両方を使用する。

### 4-3. 分析結果

#### (1) POSデータのみでの予測

表4はPOSデータのみを使用し、小売業計を予測した分析結果の一覧である。POSデータに対しては主成分分析を行っているが、少ない主成分でどの程度元の217品目分類の情報を集約できているのかについては、Proportion of Variance（寄与度）<sup>30</sup>を計算することで示すことができる。データセット①～④は、主成分の寄与度を降順に並べた上で累積寄与度を算出し、元のPOSデータを主成分でどれだけ説明できるかによってそれぞれ説明変数の数を決定して作成したものである。データセット④（元のPOSデータを90%説明）より少ない数の説明変数で分析した結果は、精度が非常に悪くなるため省略した。なお、本稿の分析は、相関係数とRMSE<sup>31</sup>によって評価している。

改めて表4の分析結果をみると、アンサンブル学習の手法2つが、線形回帰、ニューラルネットワークより高い精度を示す傾向にあることがわかる。ニューラルネットワークはパターン認識・分類に強い手法であることから、本分析のような数量予測ではパフォーマンスが十分に出せなかった可能性がある<sup>32</sup>。勾配ブースティングは、変数が少ないデータセット④で最も相関が高い結果となっている。これは、主成分の寄与度で上位のものほど標準偏差が高いため、変数の数が少なく個々の変動の影響が強いデータセットほど、誤差の補正に強

<sup>28</sup> 統計解析ソフト「R」のパッケージ「caret」を使用した。

<sup>29</sup> 世界的なデータ解析コンテストサイト「Kaggle」（2015年）では、出題問題29問の半数以上で勾配ブースティングを使用したモデルが1位となっている。

<sup>30</sup> 主成分分析では標準偏差に着目するため、各標準偏差の2乗が標準偏差の2乗の合計に占める割合を算出することで寄与度を求めることができる。

<sup>31</sup> Root Mean Square Error は、予測値が正解からどの程度乖離しているかを示す指標であり、0に近い値ほど誤差の少ない予測モデルと評価できる。

<sup>32</sup> 今回、何パターンかのデータセット、チューニングで分析を実施したが、ニューラルネットワークはピークを当てるのが難しく、変動幅の小さい予測をする結果になることが多かった。

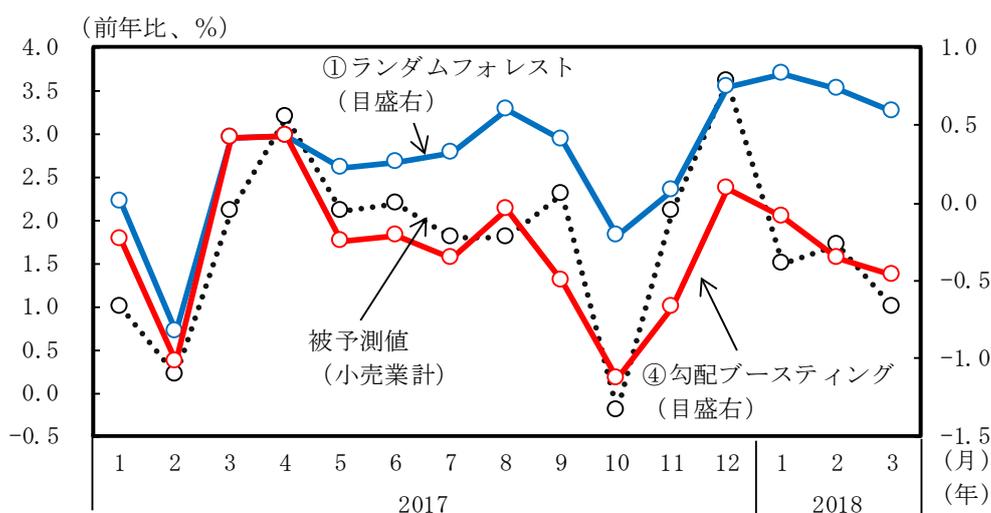
みのある同手法の効果があったと考えられる。ただし、水準感が予測値より大幅に低く、誤差は大きなものとなっている（図8）。なお、勾配ブースティングは、弱い学習モデルとして、tree と linear の両方を試みてみたが、linear の方が精度の高い結果となったため、linear の結果を表4には掲載している。数量予測が目的である上に、説明変数が全て量的変数の連続型の数値データであったため、分類型の tree よりも精度が高まったものと考えられる。この結果を踏まえ、本稿では以降の勾配ブースティングの分析結果は全て linear を弱い学習モデルとして使用した結果で表記する。また、最も誤差が少なかったのがデータセット①のランダムフォレストであるが、勾配ブースティングと同様に水準感は予測値より低く、相関は0.61と方向感もそこまで伸びない結果である（表4）。

表4 分析結果一覧（寄与度が上位の主成分で分析）

データセット	説明変数の数		線形回帰	ニューラルネットワーク	ランダムフォレスト	勾配ブースティング
①99%を説明	63	相関	0.445	0.459	0.605	0.434
		RMSE	2.43	2.14	1.65	1.73
②98%を説明	50	相関	0.361	0.397	0.572	0.442
		RMSE	2.06	1.95	1.71	1.84
③95%を説明	32	相関	0.214	0.252	0.514	0.545
		RMSE	2.06	1.95	1.83	2.00
④90%を説明	20	相関	0.204	0.250	0.638	0.753
		RMSE	2.21	2.09	1.88	2.17

（備考）経済産業省「商業動態統計」、Finatext により作成。

図8 前年比の比較（寄与度が上位の主成分で分析）



（備考）経済産業省「商業動態統計」、Finatext により作成。

さらなる精度向上を目指し、小売業計と相関性のある主成分のみをデータセットにした分析も試みた。そもそも、小売業計との相関で±0.2以上の“弱い相関がある”主成分ですら4系列しかないほど、全般的にPOSデータは小売業計との相関が非常に弱い<sup>33</sup>。先に分析したデータセットは、主成分の寄与度が高い順に変数を選択したが、寄与度が高くても相関性の低い主成分がノイズとなり、水準感を落とした可能性も考えられる。そこで、相関係数の絶対値が高いものから順に抽出してデータセットを作成し分析を試みた（絶対値なので負の相関も含む）。

分析結果が表5であるが、相関性と誤差を踏まえると、データセット⑥<sup>34</sup>のランダムフォレストと勾配ブースティングが当てはまりの良い結果となっている。勾配ブースティングの結果では、先のデータセット④の分析よりも相関性が減少しているが、誤差は改善している。データセット④では、相関性がなくノイズに近い主成分が多かったため、その誤差を補正しようと過学習になっていた可能性が考えられる。相関性がある主成分をデータセットにした分析も、水準感は予測値より低い状態ではあるが、方向感（前年比前月差）は概ね似たような動きになったため、ある程度のナウキャストが行えていると考えられる（図9、図10）。

ここでランダムフォレストと勾配ブースティングにおける説明変数ごとの影響力を確認しておきたい。アンサンブル学習は、変数重要度を用いることで、説明変数が目的変数へどの程度影響したかを評価することができる<sup>35</sup>。表6は、データセット⑥におけるランダムフォレストと勾配ブースティングの分析結果の各変数重要度を比較したものである。各主成分を構成する代表的な品目については補論4を参照されたい。表6からは、ランダムフォレストと勾配ブースティングの双方において重要度が最も高い変数は主成分6（雑酒、殺虫剤・殺鼠剤、モップ等を中心的に構成される主成分）であり、最もPOSデータ全体の動きを説明する主成分1の重要度は必ずしも高くない。また、勾配ブースティングとランダムフォレストの重要度の分布を比較すると、相対的に勾配ブースティングの偏りが大きいことがわかる。この理由としては、勾配ブースティングは、予測誤差の補正に強みがある手法のため、特定の変数に重きを置く傾向があるためと考えられる<sup>36</sup>。

---

<sup>33</sup> 小売業計に最も相関のある主成分で0.406である。

<sup>34</sup> なお、データセット⑥に含まれる主成分のPOSデータに対する累積寄与度は53%である。

<sup>35</sup> アンサンブル学習では、回帰モデルのように説明変数の係数を推定するわけではないので、予測誤差の大きさ等を計測した変数重要度が一般的に評価に用いられる。

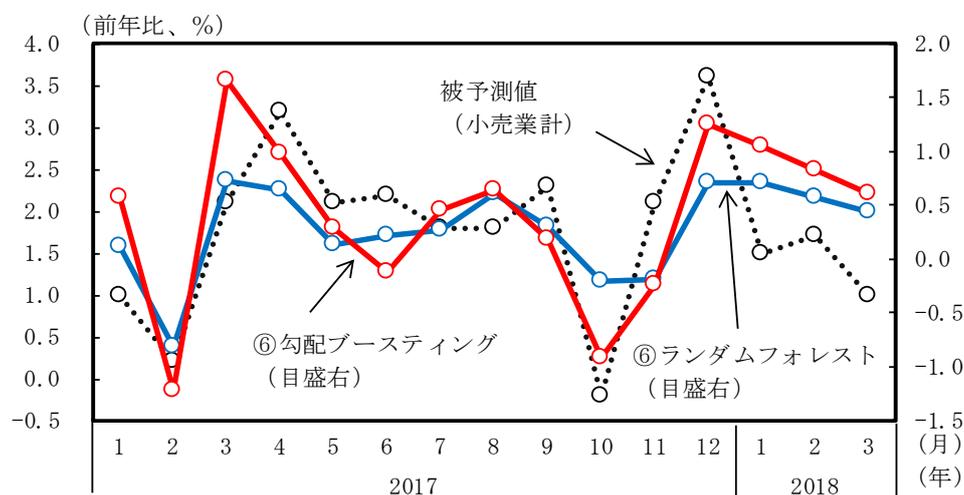
<sup>36</sup> ランダムフォレストと勾配ブースティングとどちらの手法が良いかは一概には言えないものの、先述した通り品目や主成分の強さが時期により変化することを踏まえれば、特定の変数に重きを置く勾配ブースティングより、ある程度重要度が高い変数にばらつきがみられるランダムフォレストの方が、学習期間をアップデートしない場合においては、中期的には安定的な予測になる可能性も考えられる。

表5 分析結果一覧（相関性が高い主成分で分析）

データセット	説明変数の数		線形回帰	ニューラルネットワーク	ランダムフォレスト	勾配ブースティング
⑤相関0.07以上	26	相関	0.375	0.374	0.585	0.403
		RMSE	2.43	2.31	1.54	1.61
⑥相関0.1以上	14	相関	0.367	0.367	0.632	0.634
		RMSE	1.86	1.83	1.66	1.55
⑦相関0.17以上	7	相関	0.362	0.362	0.462	0.425
		RMSE	1.88	1.83	1.69	1.70

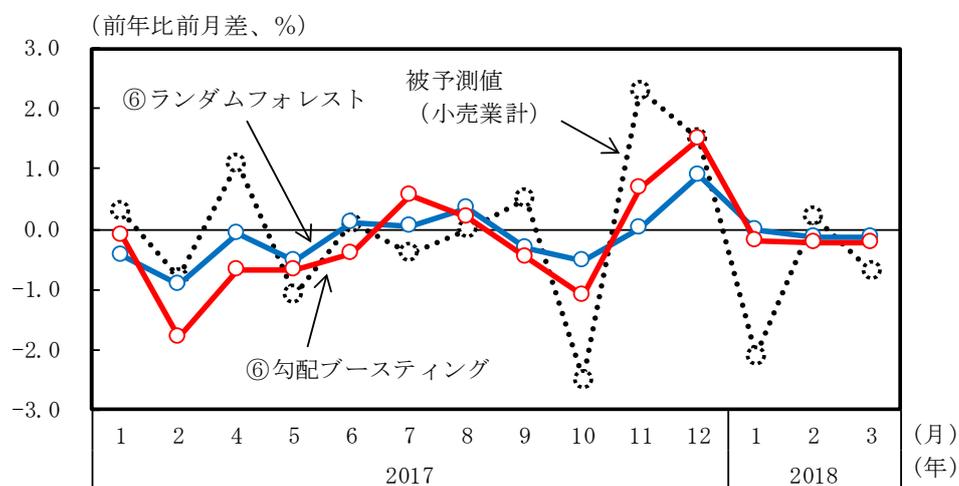
(備考) 1. 経済産業省「商業動態統計」、Finatextにより作成。  
 2. 相関は絶対値であり、負の相関も含む。

図9 前年比の比較（相関性が高い主成分で分析）



(備考) 経済産業省「商業動態統計」、Finatext 提供データにより作成。

図10 方向感の比較（相関性が高い主成分で分析）



(備考) 経済産業省「商業動態統計」、Finatextにより作成。

表6 変数重要度の比較（データセット⑥）

ランダムフォレスト		勾配ブースティング	
変数	重要度	変数	重要度
主成分6	1.000	主成分6	1.000
主成分7	0.769	主成分3	0.658
主成分16	0.571	主成分7	0.503
主成分12	0.539	主成分16	0.194
主成分20	0.475	主成分20	0.163
主成分3	0.281	主成分12	0.134
主成分1	0.254	主成分18	0.074
主成分18	0.246	主成分1	0.068
主成分42	0.174	主成分42	0.029
主成分8	0.132	主成分22	0.015
主成分22	0.101	主成分31	0.015
主成分11	0.080	主成分14	0.010
主成分31	0.048	主成分11	0.001
主成分14	0.013	主成分8	0.000

(備考) 1. 経済産業省「商業動態統計」、Finatextにより作成。  
 2. 重要度は最大値の変数を1とした相対化した値。

## (2) 天候・自動車データを活用した予測

POSデータのみで分析した中では総合的にパフォーマンスが良いと考えられる表5における⑥のデータセットを基本として、天候・自動車の外部データを組み合わせることで精度が向上するかを分析した結果の一覧が表7である。天候データを追加した分析結果がデータセットA・B、自動車データを追加したのがデータセットC、天候データと自動車データを追加したのがデータセットD・Eである。

表7 分析結果一覧（外部データを追加）

データセット	説明変数の数		線形回帰	ニューラルネットワーク	ランダムフォレスト	勾配ブースティング
⑥ POSデータのみ	14	相関	0.367	0.367	0.632	0.634
		RMSE	1.86	1.83	1.66	1.55
A) 天気追加	20	相関	0.208	0.299	0.484	0.603
		RMSE	1.72	1.71	1.52	1.59
B) 気温+積雪日追加	18	相関	0.474	0.463	0.450	0.780
		RMSE	1.76	1.71	1.57	1.57
C) 自動車追加	15	相関	0.301	0.302	0.657	0.591
		RMSE	2.01	1.83	1.54	1.58
D) 気温+積雪日+自動車追加	19	相関	0.541	0.553	0.636	0.343
		RMSE	1.64	1.56	1.29	1.43
E) 天気+自動車追加	21	相関	0.206	0.365	0.581	0.631
		RMSE	1.66	1.59	1.22	1.19

(備考) 1. 経済産業省「商業動態統計」、日本自動車販売協会連合会、全国軽自動車協会連合会、気象庁、Finatextにより作成。  
 2. 気温表記の箇所は、平均気温、最低気温、最高気温を示す。  
 3. 天気表記の箇所は、平均気温、最低気温、最高気温、積雪日、降水量、降水日を示す。

表8 変数重要度の比較

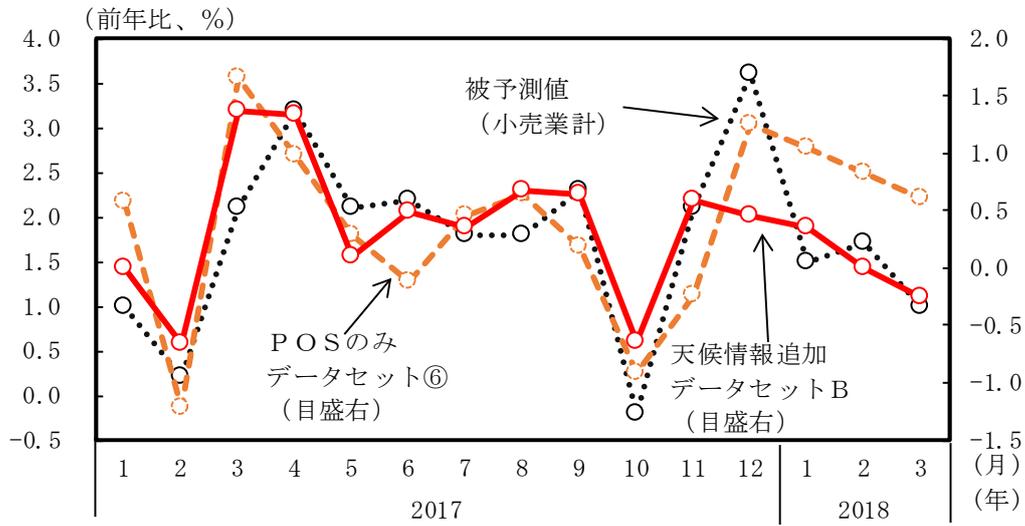
天候データあり		POSデータのみ	
変数	重要度	変数	重要度
主成分6	1.000	主成分6	1.000
主成分7	0.582	主成分3	0.658
主成分3	0.511	主成分7	0.503
主成分16	0.229	主成分16	0.194
主成分20	0.208	主成分20	0.163
主成分12	0.107	主成分12	0.134
主成分1	0.089	主成分18	0.074
主成分42	0.059	主成分1	0.068
主成分18	0.046	主成分42	0.029
主成分31	0.038	主成分22	0.015
平均気温	0.031	主成分31	0.015
主成分22	0.025	主成分14	0.010
積雪日数	0.019	主成分11	0.001
主成分8	0.019	主成分8	0.000
最低気温	0.016		
主成分11	0.012		
主成分14	0.005		
最高気温	0.000		

- (備考) 1. 経済産業省「商業動態統計」、日本自動車販売協会連合会、全国軽自動車協会連合会、気象庁、Finatextにより作成。  
2. 重要度は最大値の変数を1とした相対化した値。

まず、天候データを組み合わせることで、データセットBの勾配ブースティングの結果のように相関性を0.63から0.78へ上昇させることが可能であることがわかる。データセットBには各種の気温と積雪日数を追加しており、各種の気温が気温に敏感な品目の売上動向を補正し、積雪日数が積雪により購買機会が減少したことを補正したと考えられる。ここで、説明変数ごとの重要度を確認しておく。表8は、POSデータのみでのデータセット⑥の勾配ブースティングの分析結果と天候情報を加えたデータセットBの勾配ブースティングの分析結果の各変数重要度を比較した表である。天候データは変数重要度が必ずしも高くないものの、天候データを加えることで、POSデータの変数重要度の強弱関係が、多少なりとも入れ替わっており、これらが補正された結果と捉えることが出来る。分析結果の天候データを加える前後を比較したのが、図11、図12であるが、多くの月で補正されたことが見てとれる。ただし、方向感に関しては、2017年末以降異なる動きとなっている。一方、天候データを全て加えたデータセットAは精度が上がらず、降水量や降水日数がノイズとなってしまっている。これは、各曜日のウェイトを無視して月次データを作成したことで、雨と売上高の関係性を損なわせてしまった可能性が考えられる<sup>37</sup>。

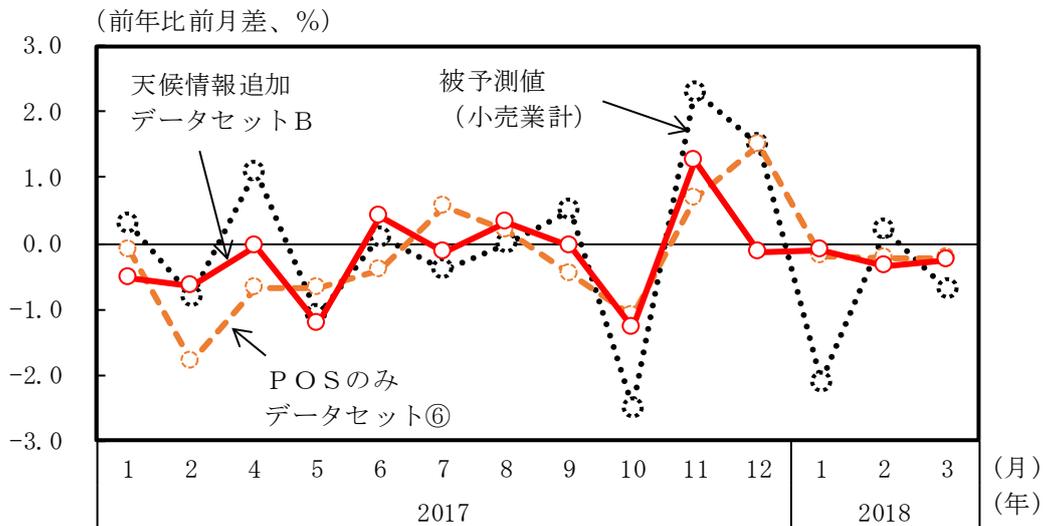
<sup>37</sup> 雨の日は売上高がマイナスになる傾向はあるが、曜日によってその度合いは異なる。詳細は

図 11 前年比の精度向上（天候データの追加）



（備考）経済産業省「商業動態統計」、気象庁、Finatext により作成。

図 12 方向感の精度向上（天候データの追加）



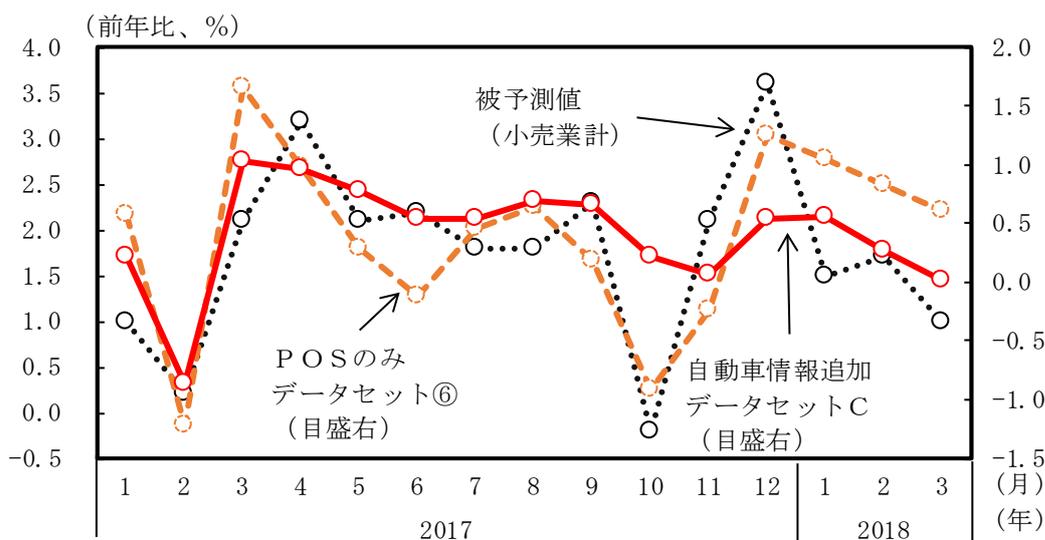
（備考）経済産業省「商業動態統計」、気象庁、Finatext により作成。

自動車データを組み合わせた結果は、表 7 のデータセット C の通りであるが、ランダムフォレストで僅かながら精度向上が確認できる。自動車データを加える前後を比較したのが図 13 である。誤差が小さくなっている月もあるが、全体を通して予測値のボリュームが小さく、予測領域の幅が狭い傾向がみられる。あまり精度が向上しなかった背景としては、自

補論 3 を参照。

動車データが天候データとは異なり、補正として影響するのではなく、POSデータ同様の説明変数として分析されたことが考えられる。POSデータの品目の多くが飲食料品であり、その動きを表すと考えられる飲食料品小売業は小売業計に占める金額比率が31%（2017年）なのに対し、自動車小売業は同13%を占める。説明変数の数を踏まえると、POSデータの主成分14系列で31%を説明するのにに対し、自動車販売台数1系列で13%を説明することから、自動車データに対するウェイトが非常に大きくなったものと考えられる。また、当初同じ傾向であろうと想定した台数と金額の違いも影響している可能性も考えられる。近年では、比較的安価な軽自動車の販売比率が高まる一方、割高な新エネルギー車の普及も進んでいるため、単純な販売台数の前年比以上に販売金額の前年比が大きく振れることが想像される。その結果、販売台数を説明変数に使用した予測結果は予測値のボリュームが小さい傾向になった可能性が考えられる。

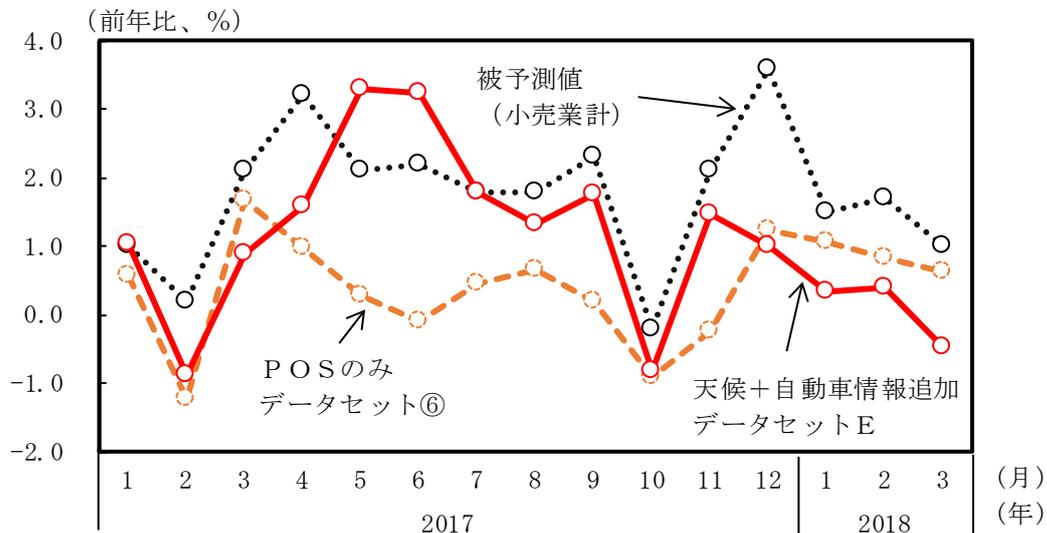
図13 前年比の精度向上（自動車データの追加）



(備考) 経済産業省「商業動態統計」、日本自動車販売協会連合会、全国軽自動車協会連合会、Finatextにより作成。

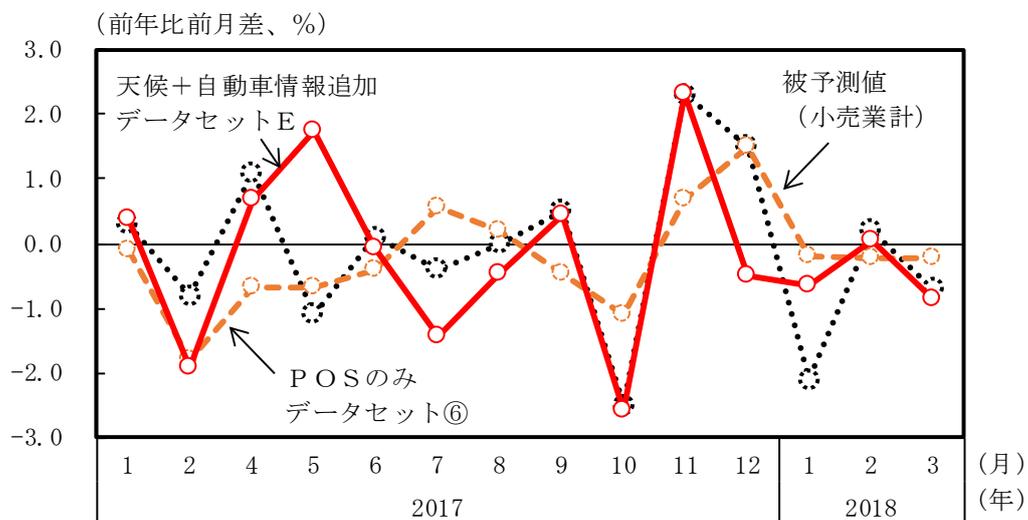
最後に、天候データと自動車データを組み合わせた結果をみると（表7、データセットE）、特に勾配ブースティングにおいて誤差が非常に小さくなっていることが確認できる。これは、天候データが自動車データも補正したことで、水準感を上昇させたのが要因と考えられる（図14）。ただし、相関性は若干下がってしまっており、方向感異なる動きをしている月が多くなってしまっている（図15）。

図 14 前年比の精度向上（天候+自動車データの追加）



(備考) 経済産業省「商業動態統計」、日本自動車販売協会連合会、全国軽自動車協会連合会、Finatextにより作成。

図 15 方向感の精度向上（天候データの追加）



(備考) 経済産業省「商業動態統計」、日本自動車販売協会連合会、全国軽自動車協会連合会、気象庁、Finatextにより作成。

以上の分析より、POSデータを活用することで、おおまかではあるものの、マクロ経済指標のナウキャストを行える可能性が示唆された。併せて、機械学習の手法がPOSデータの高次元な特徴を捉え、従来の線形回帰等の手法よりも高い精度で予測できることも確認できた。

ただし、さらなるナウキャストの精度向上に向けた課題はいくつかある。まず、POSデータの各品目のウェイトが利用できなかったことが挙げられる。本稿で使用したPOSデ

一夕は前年比だけであるため、日常食品である食パンも嗜好品のワインも同じ伸び率としての値であることが、本稿の多くの分析結果で水準感が低かった要因と考えられる。また、月次統計である小売業計を予測するため、日次の前年比を単純平均で月次化したことも、各曜日のウェイトが損なわれ、水準感を低くした要因とも考えられる。このようなウェイトの問題を抱えているにもかかわらず、今回の分析でおおまかな予測が可能であったことを踏まえれば、この問題を解消することで高精度の予測が可能となることが期待できる。

加えて、外部データも改善の余地がある。例えば、気温が小売業計に与える影響は、季節によって高い気温がプラスに働くのか低い気温がプラスに働くのかが異なると想定できる。今回の分析では季節ダミーといった季節性を考慮した変数を使用していないが、天候と季節性を組み合わせることで、予測精度が向上する可能性は十分にあると考えられる。

さらに、今回の分析において、ナウキャストと乖離が起きてしまった部分についてもより詳細な分析を行うことで、より予測精度が向上したモデルを構築することができる。例えば、予測値と実測値に乖離がみられる際に、一定のパターンを確認することができれば、乖離を補正することが可能となり、予測力を大幅に改善することが可能となるだろう。

最後に、月次統計をナウキャストするのに、日次データを説明変数に投入したモデルも考えられる。これにより、各曜日におけるウェイトの違いを考慮できる上、月中での予測もできる可能性が期待できる。POSデータを使用することによるナウキャストの可能性は十分にあると考えられることから、今後も経済分析への活用が進むことが期待される。

## 5. ディープラーニングによるセンチメント指数の開発

本節では、ディープラーニングを利用したテキストデータ分析について記載する。まず、内閣府「景気ウォッチャー調査」における景気判断と判断理由テキストの関係性を学習させ、LSTM RNN モデルを構築する。次に、このモデルを用いて、日本経済新聞（朝刊）のセンチメントを算出し、日次ベースで集約する。最後に、作成された指数と消費者マインド・POS データとの関係を分析し、その結果を考察する。

### 5-1. データ

山本・松尾（2016）や五島他（2017）等と同様に、教師データとしては、内閣府「景気ウォッチャー調査」（期間：2002年1月～2017年12月）を用いた。景気ウォッチャー調査は、地域の景気動向を迅速かつ的確に把握するため、景気に敏感な職場で働くタクシー運転手、百貨店・スーパーの店員、中小企業の経営者・従業員といった人々（景気ウォッチャー）が肌で感じる景気の現状判断及び先行き判断について調査を毎月実施し、公表される。

景気ウォッチャー調査を教師データとした主な理由の一つとして、景気に対する選択式判断、及びこれに対応する日本語のテキストデータを大量に入手可能であることが挙げられる。機械学習は（同種のデータであれば）学習データ数の増加に伴い精度の向上が期待できるため、景気ウォッチャー調査を教師データとして構築したモデルは、新聞記事内に潜む景気の水準をある程度正確に描写できると考えられる。

さらにテキストの質がある程度担保されている点も重要であろう。例えば、商品や映画のレビューサイトの口コミを分析するに当たっては、誤字脱字の修正や求めている情報と無関係なテキストを除外する等のクレンジング作業にかなりの労力を割かざるを得ないこともある。一方、景気ウォッチャー調査に関しては、統計の作成段階を通じてこの種の処理がされているため、上記のようなクレンジング作業の負担は比較的小さいと考えられる。

なお、本分析では、文章がポジティブ・ネガティブのどちらであるかを判断するモデルの構築を行うため、選択式景気判断のうち、①良・やや良、②やや悪・悪、と回答した場合のみを分析対象とした上で、①と②のサンプル数を等しくした<sup>38</sup>。このため、最終的にモデルの構築に使用した判断理由のコメント数は、現状判断は9万3,416件、先行き判断では9万8,122件である。

次に、新聞センチメント指数のベースとなったテキストデータを解説する。今回使用したのは、日経メディアマーケティング（株）から購入した日本経済新聞の朝刊のタイトル・本文であり、テーマが事業戦略、経営管理、市場動向、政策・制度、社会政策、経済、災害・事故、社会問題、消費トレンドに属する記事、あるいは業界が生活・医薬業界、小売・飲食業界、物流・陸運・海運、ネット事業に該当する記事（期間：2013年1月～2017年12月（5

<sup>38</sup> 対象とした期間では、②「やや悪・悪」のコメント数の方が①「良・やや良」のコメント数よりも多いため、①はすべて利用しつつ、無作為に②のコメントを削除することで、②を①のコメント数に合わせた。

年間分)である<sup>39</sup>。なお、記事総数は18万3,897本である。

表9.各紙面の記事タイトルの例

紙面	記事数(本)	記事タイトル及び掲載日の例
一面	6,279	<ul style="list-style-type: none"> <li>■政府、円借款、ドルでも返済、途上国に配慮、インフラ輸出拡大。(2013年1月16日)</li> <li>■ノルウェー政府系ファンド、日本株保有3割増、昨年末4.8兆円。(2015年3月30日)</li> <li>■電子申請、安全・便利に、ブロックチェーン活用、政府、まず入札。(2017年6月29日)</li> </ul>
総合面	18,512	<ul style="list-style-type: none"> <li>■JR北海道に改善命令、改ざん、20年前から——後任不在で経営陣続投。(2014年1月22日)</li> <li>■経済協調に政治リスク、G20閉幕、内向き志向、各国で。(2016年9月6日)</li> <li>■フリーランス失業に保険、対象1000万人、政府・損保、所得を補償。(2017年3月14日)</li> </ul>
政治面	13,217	<ul style="list-style-type: none"> <li>■首相、国際金融市場で発信、浜田参与ら英米に派遣、改革説明、投資呼び込み。(2013年6月4日)</li> <li>■自民、カジノ法案成立へ公明に協力要請。(2015年3月28日)</li> <li>■地方にシェア経済普及、総務省、来年度から助成。(2017年9月27日)</li> </ul>
経済面	14,566	<ul style="list-style-type: none"> <li>■水産加工再生へ、大手と連携支援、復興庁、気仙沼で。(2013年1月21日)</li> <li>■出生率、緩やかに低下へ、官民の少子化対策、急務、人口減社会への対応課題。(2015年6月6日)</li> <li>■訪日客27%増、11月237万人。(2017年12月21日)</li> </ul>
国際面	22,035	<ul style="list-style-type: none"> <li>■EUやIMF、ギリシャ次期融資合意。(2013年4月16日)</li> <li>■ミャンマー、国際会議参加、少数民族「ロヒンギャ」巡り。(2015年5月25日)</li> <li>■米ハリケーン、豪雨続き被害3.2兆円も、製油所15%停止。(2017年8月30日)</li> </ul>
企業面	30,989	<ul style="list-style-type: none"> <li>■13年度粗鋼生産、プラスに期待感、鉄連会長、経済対策や円高修正、追い風。(2013年1月29日)</li> <li>■建設機械出荷額、8月7.3%減。(2015年10月1日)</li> <li>■SEMIと日本半導体製造装置協会、半導体装置受注額24%増。(2017年3月15日)</li> </ul>
マーケット総合面	13,116	<ul style="list-style-type: none"> <li>■国債買いオペ、1～3年の応札9200億円。(2013年8月15日)</li> <li>■円、上値重い展開、米、年内利上げ観測崩れず、東京市場。(2015年4月7日)</li> <li>■仮想通貨関連株が下落、ビットコイン急落受け。(2017年12月26日)</li> </ul>
マーケット商品面	7,840	<ul style="list-style-type: none"> <li>■紙の内需、23年ぶり低水準、12年、経費節減やネットにシフト。(2013年1月22日)</li> <li>■ガソリン6週連続上昇、5カ月ぶり高値。(2015年6月4日)</li> <li>■鋼材値上げ、製品に波及、平鋼や大径角形鋼管、年初比8～9%高、需要増、輸送能力に懸念も。(2017年12月28日)</li> </ul>
社会面	18,954	<ul style="list-style-type: none"> <li>■海・山の事故、4人死亡、11人けが。(2013年7月14日)</li> <li>■ニホンライチョウ、上野動物園でもふ化、絶滅危惧種、富山に続き。(2015年6月29日)</li> <li>■長野南部で震度5強、気象庁「1週間程度注意必要」、2人けが。(2017年6月26日)</li> </ul>

- (備考) 1. 日本経済新聞社により作成。  
 2. 消費者マインドとの関係を分析した紙面のみを掲載。  
 3. 総合面は総合面、総合・経済面、総合・政治面を含む。政治面は政治面、総合・政治面を含む。経済面は経済面、総合・経済面を含む。国際面はアジア Biz 面を含む。

ただし、消費者マインド等とより連動している情報は、新聞記事の中でも特定の情報であ

<sup>39</sup> 換言すれば、スポーツ関係の記事や人事などの景気や消費者マインドとの関係性が薄いと考えられる記事、及び見出しのみの記事は除かれている。

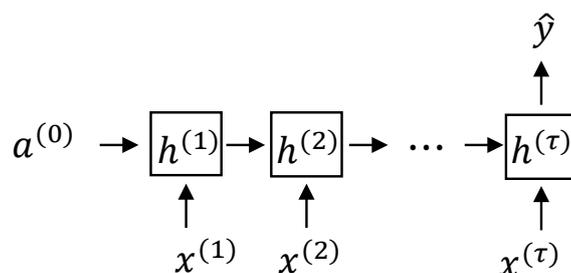
る可能性も考えられるため、対象記事全体を使う分析に加え、記事の「紙面別」に分析を行うことも試みる。対象とした紙面は、データ数や時系列的に利用可能であるとの観点から、①一面、②総合面、③政治面、④経済面、⑤国際面（アジア Biz を含む）、⑥企業面、⑦マーケット総合面、⑧マーケット商品面、⑨社会面、の9種類を対象とした<sup>40</sup>。各紙面の具体的な記事のタイトルと記事数は表9に記載している。

## 5-2. 分析手法

### (1) センチメント指数の構築方法

本稿では、ディープラーニングの手法を用いた LSTM RNN モデルの構築を行う。この手法を利用することの利点は文脈を時系列として評価できるため、テキスト等のデータを分析する際には、より精度の高いモデルの構築が可能となることが指摘されている（饗場・山本、2018）。本稿で行う景気ウォッチャー調査を用いた分析のイメージは図16のように表すことが可能である。通常のニューラルネットワークと同様、RNN についても、「入力層 ( $x$ ) - 隠れ層 ( $h$ ) - 出力層 ( $y$ )」で構成されるが、大きく異なる点は、過去の隠れ層の値が現在の隠れ層に伝達される点である。 $x^{(1)} \sim x^{(\tau)}$ は、ある文章を構成する単語（形態素）であり（イメージとして、例えば「景気が良い」であれば、 $x^{(1)} =$ 景気、 $x^{(2)} =$ が、 $x^{(3)} =$ 良い）、これらを順番にモデルに読み込ませてゆくが、例えば、 $x^{(2)}$ の入力時には、過去の隠れ層の値である $h^{(1)}$ も $h^{(2)}$ に伝達されるため、文脈情報を疑似的に考慮することが可能となる。このプロセスを繰り返し、最終的に当該文章がポジティブである確率 ( $\hat{y}$ ) が推計される。

図16 本稿で使用する RNN モデルのイメージ



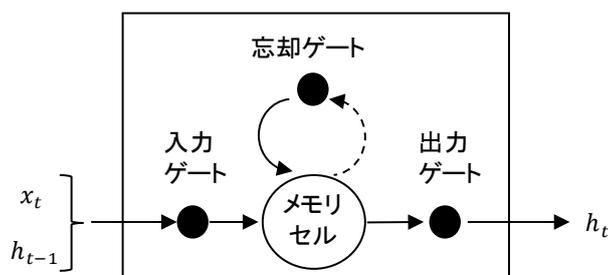
(備考) Goodfellow et al.(2016)を参考に作成。

ただし、単純な RNN が記憶できるのは短期的なものに限られるため、より長期にわたり情報の記憶ができるように考案されたものが LSTM である。基本的な構造は RNN と全く

<sup>40</sup> なお、総合・政治面、総合・経済面に関しては記事の内容に鑑み、前者は「総合面」と「政治面」、後者は「総合面」と「経済面」の双方に分類した。

同じであるが、図 16 の隠れ層の内部に LSTM ブロックと呼ばれる仕組みを導入した点が異なる。ここでは詳細な定義式等は割愛するが<sup>41</sup>、LSTM ブロックのイメージは図 17 のように示すことができ、中央にあるメモリセルと、入力・忘却・出力の3つのゲートから構成される。メモリセルは自己ループの仕組みを持っており、自身が  $t-1$  期に保有している情報を  $t$  期に伝えている。その途中に忘却ゲートがあり、必要に応じて不要な情報が忘却される。入力ゲートは、外部からの新しいデータ ( $x_t$ ) と前期の隠れ層 ( $h_{t-1}$ ) の情報を受け取り、必要な情報をメモリセルに伝達する。また、メモリセルの情報は、適切なタイミングで出力ゲートを通して読み出されている。仮に、忘却も入力も行わない状態であれば、メモリセルの情報は永遠に保持されることになる。この LSTM ブロックの導入により、長い文脈を考慮した高度な推定が行えることが期待できる。以下では具体的なモデルの構築方法について記載していく。

図 17 LSTMブロックのイメージ



(備考) 岡谷(2015)を参考に作成。  
点線は時間遅れを表す。

コンピューターの内部で処理できるのはあくまでも数値であり、日本語をそのまま読むことができるわけではないため、テキストを数値化する(ベクトル化する)作業が必要となる<sup>42</sup>。まず、景気ウォッチャーのコメント(現状、先行き)と新聞記事のタイトルと本文を日本語形態素解析エンジンである MeCab により単語(形態素)ごとに分かち書きを行う<sup>43</sup>。次に、分割された各単語をベクトルに変換するが、ここでは単語埋め込み(word embedding)の手法を利用する。単語埋め込みを利用することで、同義語を同じような単語ベクトルで表現することや、ベクトルの演算により意味の関係が表現できることが期待できる<sup>44</sup>。例えば、前者については、「良い」と「好調」のベクトルは近い値を持つことや、後者については、「男から王様」と「女から女王」への移動が同じベクトルによって可能となる。本稿で

<sup>41</sup> LSTM RNN のより詳細な定義等については、岡谷(2015)等を参照。

<sup>42</sup> テキスト分析の詳細については、Chollet(2017)等を参照。

<sup>43</sup> 分かち書きの辞書は、最新の単語を多数含む mecab-ipadic-NEologd を使用した。また、出力は基本形で行っている。

<sup>44</sup> また、比較的少ない次元数ながら多くの情報を格納できる利点も持つ。

は、東北大学の乾・鈴木研究室が公表している「日本語 Wikipedia エンティティベクトル (2017年2月1日版)」(Word2Vec<sup>45</sup>により日本語版 Wikipedia の本文全文から事前に学習した 200 次元の単語埋め込みベクトル)<sup>46</sup>を活用し、各単語を対応する単語ベクトルに変換した<sup>47</sup>。なお、モデルに読み込ませた単語数はコメント・各記事の最初の 300 個とした<sup>48</sup>。

前述の通り、景気ウォッチャー調査を教師データとして、「良い・やや良い」を 1 (ポジティブ)、「悪い・やや悪い」を 0 (ネガティブ) とした 2 値判断モデルを構築する。また、五島他 (2017) が指摘するように、景気ウォッチャー調査の「現状」から作成した指数と「先行き」から作成した指数とでは異なる情報を有している可能性があるため、現状と先行きは別々に学習を行うこととする。出力層の活性化関数には、シグモイド関数 (sigmoid function) を採用した。

LSTM RNN モデルの主なハイパーパラメータについては、隠れ層の活性化関数は双曲線正接関数 (hyperbolic tangent function)、隠れ層の次元は 250、バッチサイズは 32、ドロップアウトは 0.117、エポック数は 40 である<sup>49</sup>。モデルパラメータの最適化には Adam (adaptive moment estimation) を使用し、patience=4 の Early Stopping を設定した<sup>50</sup>。また、教師データのうちランダムに抽出した 26.5% をテストデータとして利用した。これらのモデルの実装には Keras を用いている。学習結果の正解率 (accuracy) は表 10 に提示した通りであるが、山本・松尾 (2016) 等の先行研究同様、テストデータで 95% 程度の精度で文章のセンチメントを正しく判別できていることがわかる。

表 10 正解率

	学習データ	テストデータ
現状モデル	0.969	0.946
先行きモデル	0.971	0.949

<sup>45</sup> 3層から構成されるテキスト処理を行うニューラルネットワーク。詳細は西尾 (2014) 等を参照。

<sup>46</sup> [http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki\\_vector/](http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/)

<sup>47</sup> イメージとして、例えば「景気が良い」を、分かち書きすると「景気 が 良い」となる。次に、「景気」、「が」、「良い」の 3 つの単語それぞれに対応する 200 次元ベクトルを日本語 Wikipedia エンティティベクトルから取得することで、「景気が良い」の文章は  $200 \times 3$  の行列で表現できる。また、「良い」と「好調」が似たようなベクトルであれば、「景気が良い」をポジティブと学習したモデルに、「景気が好調」の文章を与えると、同じような入力値となることから、同様にポジティブであると判断することが期待される。

<sup>48</sup> 300 個に満たない場合は、足りない分だけ「0」を追加して長さを調整している。

<sup>49</sup> 隠れ層、ドロップアウト、テストデータの分割の設定については、現状の学習モデルにおいて、「ベイズ最適化」により決定した (実装には GPyOpt を用いた)。また、これらのハイパーパラメータは先行きの学習モデルにおいても使用されている。

<sup>50</sup> この結果、現状モデルのエポック数は 9、先行きモデルのエポック数は 10 で学習が止まっている。

上記の要領で構築した現状モデルと先行きモデルそれぞれに、新聞記事の「タイトル+本文」を一括りとして読み込ませ、その内容がポジティブである確率を算出させる。各記事において算出された確率について、紙面別・日次ベースで単純平均を行い、モデル別・紙面別の「新聞センチメント指数」を作成した。また、紙面別に指数を作成した場合、一部の紙面においては該当する紙面が存在しない日が存在するため、前日の値を横置きすることで補完を行った。このように作成した紙面別の新聞センチメント指数と消費者マインドとの関係を確認し、テキストデータとマインドの関係性についての考察を深める。

## (2) 非対称性を考慮したセンチメント指数の分析手法

POSデータ同様にテキストデータは、高頻度であることの強みがあるため、ニュース記事とPOSデータを組み合わせた分析も行う。これにより新聞記事の内容の変化が、消費者マインドの変化等を通して、消費行動（POSの売上高）に影響を与えている可能性を検証する。その際、ネガティブに変化したときの方がマインドに与える影響が大きいとのSoroka(2006)の指摘を踏まえ、新聞センチメントがプラス方向に変化したときと、マイナス方向に変化したときの影響（非対称性）を考慮して分析を行う。

日次データではノイズが大きくなりすぎる懸念があるため、新聞センチメント指数が利用可能な2013年～2017年において、POSデータ及び新聞センチメント指数を週次で単純平均を行い、POSデータの売上高の前年比を被説明変数とした以下のシンプルな回帰分析を行うことで、上記の検証を行う<sup>51</sup>。

$$y_t = \alpha + \beta_0^+ D_t^+ x_t + \beta_1^+ D_{t-1}^+ x_{t-1} + \beta_0^- D_t^- x_t + \beta_1^- D_{t-1}^- x_{t-1} + \sum_{j=1}^3 \gamma_j Z_{jt} + \delta y_{t-1} + \varepsilon_t \dots \text{(式5)}$$

ここで $y_t$ はPOSデータの売上高（前年比）、 $x_t$ は新聞センチメント指数であり、 $D_t^+$  ( $D_t^-$ )は、新聞センチメント指数が前週と比較してプラス（マイナス）になった時に1、マイナス（プラス）になった時にゼロをとるダミー変数である。これにより、新聞センチメント指数がプラス方向に変化した時と、マイナス方向に変化した時とで、売上高に与える影響が異なるのかについて観察することが可能となる。また、 $Z_{1t} \sim Z_{3t}$ は①消費税駆け込みダミー<sup>52</sup>、②消費税反動ダミー<sup>53</sup>、③年末年始ダミーの3つである。前述の通り、POSの売上高は、364日前比を使っている関係から年末年始が異常値となる傾向がある。ここでは12月30～31日・1月1～2日の4日間のデータを除いてPOSの週次の平均を作成しているが、この処置が結果に影響を与える可能性があるため、年末年始ダミーを加えている。

<sup>51</sup> 定式化に当たっては、五島他（2017）を参考にした。

<sup>52</sup> 2014/3/17、2014/3/24の週に+1、1年後の2015/3/16、2015/3/23の週に-1をとるダミー

<sup>53</sup> 2014/3/31、2014/4/7の週に+1、1年後の2015/3/30、2015/4/6の週に-1をとるダミー

### 5-3. 分析結果

#### (1) 新聞センチメント指数のデータの特徴

表11は、現状及び先行きモデルにより作成した新聞センチメント指数の平均値と標準偏差を紙面別に示したものである。新聞センチメント指数はポジティブである確率が高いと1に近くなる指数であるが、データの特徴をみると、現状モデルの平均値は多くの紙面で0.8程度とポジティブである確率が高くなっている。一方、先行きモデルの多くの紙面では0.5程度の数値であり、先行きモデルの数値は比較的ニュートラルなものになる傾向が観察できる。なお、先行きの新聞センチメント指数の分布を示したものを補論6に掲載しているが、多くの紙面において、平均を中心とした正規分布となっていることが観察できる。

表11 新聞センチメント指数の平均値・分散（紙面別）

#### 現状

	記事全体	一面	総合面	経済面	政治面	国際面	企業面	社会面	マーケット 総合面	マーケット 商品面
平均値	0.80	0.81	0.78	0.78	0.82	0.76	0.86	0.69	0.72	0.72
標準偏差	0.03	0.16	0.10	0.11	0.10	0.11	0.07	0.11	0.10	0.14

#### 先行き

	記事全体	一面	総合面	経済面	政治面	国際面	企業面	社会面	マーケット 総合面	マーケット 商品面
平均値	0.57	0.58	0.48	0.54	0.54	0.46	0.75	0.46	0.41	0.45
標準偏差	0.05	0.21	0.13	0.15	0.14	0.13	0.10	0.12	0.12	0.16

#### 現状(平均値)

	記事全体	一面	総合面	経済面	政治面	国際面	企業面	社会面	マーケット 総合面	マーケット 商品面
2013	0.80	0.84	0.78	0.80	0.83	0.75	0.86	0.69	0.71	0.74
2014	0.80	0.84	0.78	0.78	0.82	0.76	0.87	0.69	0.72	0.68
2015	0.80	0.81	0.79	0.79	0.82	0.77	0.87	0.70	0.71	0.69
2016	0.79	0.79	0.76	0.76	0.81	0.76	0.86	0.69	0.72	0.74
2017	0.80	0.80	0.76	0.79	0.83	0.77	0.86	0.70	0.76	0.77

#### 先行き(平均値)

	記事全体	一面	総合面	経済面	政治面	国際面	企業面	社会面	マーケット 総合面	マーケット 商品面
2013	0.59	0.64	0.50	0.55	0.53	0.46	0.76	0.45	0.42	0.45
2014	0.59	0.61	0.49	0.52	0.56	0.48	0.77	0.47	0.42	0.41
2015	0.57	0.59	0.49	0.53	0.55	0.47	0.75	0.46	0.39	0.40
2016	0.55	0.53	0.45	0.51	0.54	0.45	0.74	0.46	0.39	0.46
2017	0.56	0.54	0.46	0.57	0.53	0.46	0.74	0.45	0.42	0.52

- (備考) 1. 内閣府「景気ウォッチャー調査」、日本経済新聞社により作成。  
 2. 紙面の定義はデータの定義部分を参照。  
 3. 図表の「現状」は景気ウォッチャー調査の現状判断から構築したモデルに基づきセンチメントを付与した数値。「先行き」は景気ウォッチャー調査の先行き判断から構築したモデルに基づきセンチメントを付与した数値。

紙面ごとに新聞センチメント指数の特徴を概観すると、数値が相対的に高かったのは企業面であり、その平均値は現状モデル・先行きモデルともに、今回取り上げた紙面の中で最も高かった。この紙面では経営者へのインタビューや新たな製品、サービス、取組などの紹

介記事など、企業のプラス面を中心的に取り上げる記事が時折掲載されており、こうした記事が紙面全体の新聞センチメント指数の高さにつながったと思われる。

逆に、やや値が低い紙面となったのは、社会面とマーケット総合面、マーケット商品面である。社会面の記事では、犯罪・防災対策、ワールドカップなど各種イベントの盛り上がり等の記事が社会面に分類されている例があり、こうした種類の記事が社会面の数値を押し上げている傾向がある反面、事件や不祥事関連の記事がセンチメント指数を押し下げる可能性があると考えられる。

また、標準偏差からは、一面の変動が大きくなる傾向があることがわかる。一面は、特に読者の興味を引くために、ポジティブかネガティブかが比較的明確なものが多くなる傾向にあると推察でき、結果的に変動も大きくなったことが考えられる。また、食料、金属などのコモディティ市場の動向により焦点を当てたマーケット商品面についても、標準偏差が比較的大きくなる傾向がみられており、株式や債券など金融資産の動向を中心に報じるマーケット総合面よりも標準偏差が高くなっている。一般的にコモディティ価格のボラティリティは、金融商品と比較して大きいことが指摘されており<sup>54</sup>、この点が新聞センチメント指数にも反映された可能性がある。

なお、新聞センチメント指数の数値は、対象とした5年間では年ごとに明確な傾向は確認できなかった。新聞センチメント指数を年単位で平均した結果を確認すると、各紙面に共通して、ある特定の年のみ大きい（小さい）との傾向も見当たらないことから、各紙面の変動は、ある程度独立していることが考えられる。

## （2）新聞センチメント指数と消費者態度指数の相関

次に、紙面別に新聞センチメント指数と消費者態度指数<sup>55</sup>の相関係数を確認する。消費動向調査は、調査基準日が毎月15日であるので、相関係数の計算に使用した新聞センチメント指数は後方30日移動平均の毎月15日時点の数値である。

これらの相関係数を示したものが図18である。特に、景気ウォッチャー調査の先行きモデルから作成した「経済面」、「マーケット商品面」の新聞センチメント指数と消費者態度指数の相関は0.6と高いことが確認できる。一方、「記事全体」、「一面」、「総合面」といった、経済や金融以外の様々なジャンルの記事をも含む紙面に関しては、相関係数がゼロ近傍であり、新聞センチメント指数が消費者マインドへ与える影響ははっきりしない。消費者マインドは、経済面に掲載されるような全般的な経済情勢の変化や、商品マーケット面に多くみられる価格変化など、特定の情報の影響を受けやすい傾向にあることが考えられる。また、同じ「経済面」でも現状で学習させると相関が0.25と半分以下となることや、「マーケット

<sup>54</sup> 例えば、諸田（2010）はコモディティ価格の変動の特徴の一つとして、金融資産と比較してボラティリティが高く、その大きな原因は在庫を含む現物の供給制約に起因する部分であり、これが価格水準の急激かつ大幅な上昇として現れることが多い、と述べている。

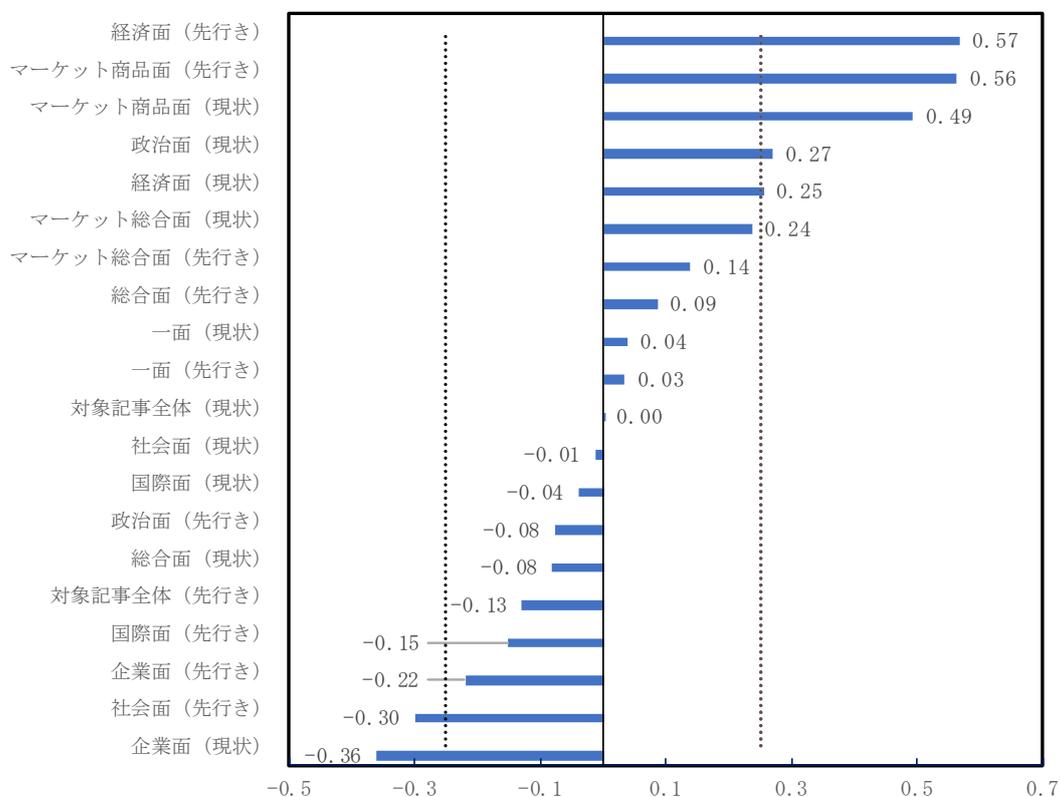
<sup>55</sup> 二人以上の世帯、原数値。

商品面」の相関も現状モデルでは若干低下することから、消費者マインドは見通しも踏まえて形成されている可能性が示唆される。

経済やマーケット面の他に、低いながらも一定程度の相関が確認できたものとして、現状モデルで作成した「政治面」が指摘できる。De Boef and Kellstedt (2004)は、経済的な状況に加え、政治が消費者マインドに影響を与えていることを実証研究により示している。図18の単純な相関からも、経済的な情報と消費者マインドの連動の方が高いものの、政治面も一定程度の相関があることを踏まえると、消費者マインドの変化について、経済的な変化では説明できない部分の一つとして、政治的な要因が示唆されている。ただし、先行きモデルから作成した「政治面」と消費者マインドの相関はほとんどみられないことから、経済面と異なり、政治は現状の観点がより重要である可能性がある。

また、消費者マインドと負の相関係数が確認できる紙面として、「企業面」から構築した指数が指摘できる。この背景は必ずしも明確ではないが、前述の通り、企業面は企業の新しい取組等を中心に取り上げる傾向があるため、仮説の一つとしては、消費者マインドが低下した際、企業は新しい商品・サービスをアピールする（よりポジティブな記事を掲載させる）ことで消費者に関心を持ってもらおうとする傾向がある可能性が考えられる。

図18 新聞センチメント指数（紙面別）と消費者態度指数の相関係数



- (備考) 1. 内閣府「景気ウォッチャー調査」、日本経済新聞社により作成。  
 2. 紙面の定義はデータの定義部分を参照。  
 3. 景気ウォッチャー調査の先行き判断から構築したモデルに基づきセンチメントを付与した。  
 4. 点線は有意水準5%ライン。

上記の結果を踏まえ、とりわけ相関係数が高かった経済面（先行き）、マーケット商品面（先行き）から構築した新聞センチメント指数に対し、より詳細にみていくこととしたい。

まず、上記2紙面に属する新聞記事のうち、新聞センチメント指数が高水準あるいは低水準であった記事タイトルの例を確認する（表12）。経済面の例を確認すると、企業の賃上げ拡大、消費判断の上方修正に関する記事はポジティブであり、国際機関による経済成長見通しの下方修正、介護事業の利益率低下、に関する記事はネガティブと判定された。次に、マーケット商品面の例をみると、オフィス用品の中古価格が上昇している記事及び築地市場でマグロの取引価格の高値が相次いでいることを報じた記事の指数は1にほぼ近く、一方でニッケル価格が6年ぶりの安値圏に入ったことを述べた記事や銀座の店舗の空室率上昇について報じた記事の指数はほぼ0の数値を示した。このように、モデルによるマインドの分類は、直観とある程度整合的だと考えられる。

表12 新聞センチメントの付与例

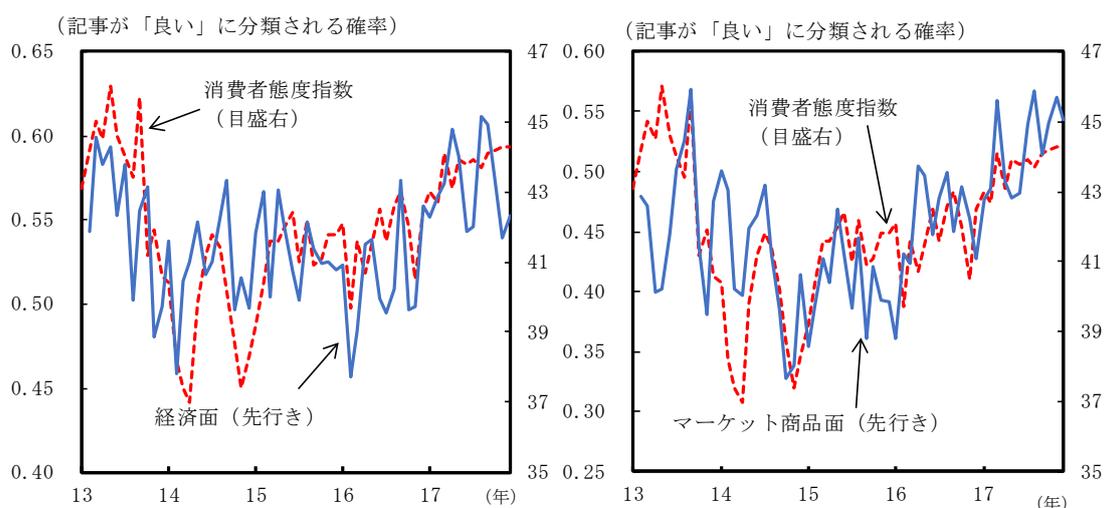
紙面	センチメントの付与例（記事タイトル及び掲載日のみ掲載）
経済面	<ul style="list-style-type: none"> <li>■センチメントが高い例</li> <li>・新設住宅着工戸数、1月、5カ月連続増加。(2013年3月1日) ⇒1.00</li> <li>・企業の賃上げ広がる、好循環へ政府は改革を。(2015年3月20日) ⇒0.98</li> <li>・財務省、7地域の消費判断を上方修正、7月。(2017年8月2日) ⇒1.00</li> <li>■センチメントが低い例</li> <li>・世界経済、足元は鈍化、IMF、今年3.1%成長に下方修正、新興国の減速響く。(2013年7月10日) ⇒0.01</li> <li>・漁獲量7割減で日ロ合意、15年のサケ・マス。(2015年6月12日) ⇒0.06</li> <li>・介護事業、利益率が低下、昨年度3.3%、報酬下げ・人件費上昇で。(2017年10月27日) ⇒0.00</li> </ul>
マーケット商品面	<ul style="list-style-type: none"> <li>■センチメントが高い例</li> <li>・オフィス用品、中古価格上昇、12月、3カ月ぶり。(2013年1月25日) ⇒0.99</li> <li>・国産鶏むね肉、高値続く、飲食店に値上げの動き、消費持続へ食べ方提案も。(2017年12月20日) ⇒0.98</li> <li>・築地、最後の年末で活気、マグロ取引、高値相次ぐ。(2017年12月26日) ⇒1.00</li> <li>■センチメントが低い例</li> <li>・半導体メモリー、相場底入れ感も需要弱く、メーカー供給抑制が支え。(2013年2月7日) ⇒0.05</li> <li>・ニッケル、ファンドの買い縮小、6年ぶり安値圏に。(2015年4月3日) ⇒0.00</li> <li>・銀座の店舗、空室率上昇、9月末、海外ブランド出店鈍く。(2017年11月9日) ⇒0.05</li> </ul>

- (備考) 1. 内閣府「景気ウォッチャー調査」、日本経済新聞社により作成。  
 2. 紙面の定義はデータの定義部分を参照。  
 3. 景気ウォッチャー調査の先行き判断から構築したモデルに基づきセンチメントを付与した。

次に、経済面・マーケット商品面それぞれの新聞センチメント指数と消費者マインドの時系列的な動向について確認する（図 19）。相関の高さからもわかる通り、新聞センチメント指数と消費者マインドの大まかな方向感是一致的である。例えば、2014 年における消費者態度指数は、消費税率が引き上げられた 4 月に向けて低下し、同年半ばには一時的に回復したものの、秋口頃から再び低下に転じ、11 月に再びボトムを付けている。変化幅の程度においては差がみられるものの、経済面、マーケット商品面の新聞センチメント指数についても、同様な動きをしていることが確認できる。

ただし、消費者マインドとの相関の程度は時期に応じて異なっており、例えば、2015 年においてはマーケット商品面の方がより連動している一方、2016 年においては、経済面の方がより相関している。こうしてみると、経済面とマーケット商品面の双方が一方向的に優れているというわけではないため、消費者マインドとの関係性の分析においては両者の情報を総合的にみることが重要であると考えられる。

図 19 新聞センチメント指数と消費者態度指数の推移  
（経済面（左図）とマーケット商品面（右図））



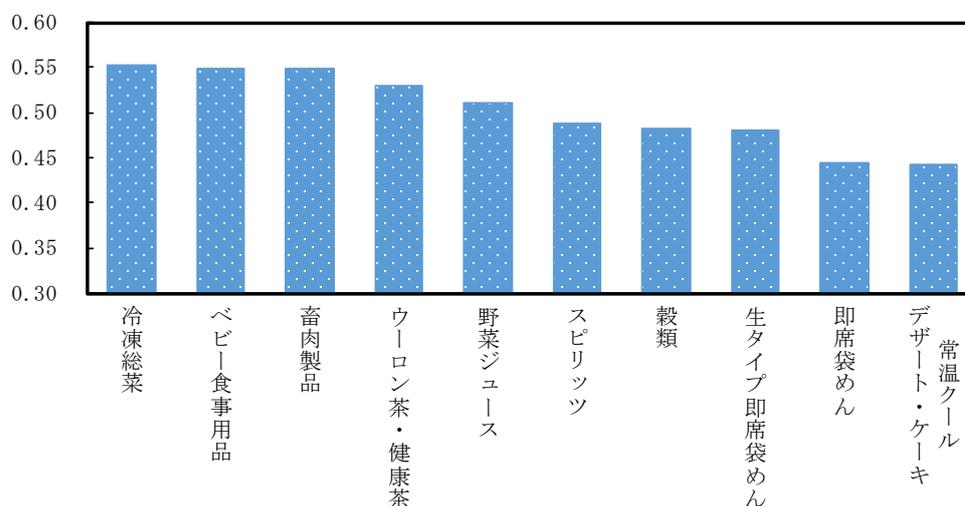
- （備考）
1. 内閣府「消費動向調査」、日本経済新聞社により作成。
  2. 消費者態度指数は二人以上の世帯、原数値。
  3. 消費動向調査の調査期間は毎月 15 日という点を踏まえ、本グラフに用いた新聞センチメント指数は毎月 15 日時点かつ後方 30 日移動平均とした。
  4. 平成 25 年 4 月調査より調査方法等を変更したため、それ以前の数値と不連続が生じている（平成 25 年度以降は郵送調査、平成 24 年度までは訪問留置調査。）。

### （3）非対称性を考慮した回帰分析の結果

最後に、上記で消費者マインドとの相関が確認できた先行きモデルで作成した経済面、マーケット商品面の新聞センチメント指数を説明変数として、POS データとの関係性を週次で分析をする（式 5 を参照）。この分析により検証したいことは、新聞センチメント指数の変化と POS データの関係が非対称的である可能性である。この点をより検証しやすく

するため、被説明変数には、POSデータの217品目分類のうち、推計期間において消費者マインドとの相関が高い品目分類を利用する。新聞センチメント指数は、いわば消費者マインドの代理変数と考えられるため、消費者マインドにより敏感な品目分類を被説明変数につかうことで、非対称性の分析が行い易くなることが考えられる。具体的には、新聞センチメント指数が利用可能な2013年～17年において、消費者態度指数との相関係数が高い上位10品目分類（冷凍総菜、ベビー食事用品、畜肉製品、ウーロン茶・健康茶、野菜ジュース、スピリッツ、穀類、生タイプ即席袋めん、即席袋めん、常温クールデザート・ケーキ）を選定し（図20）<sup>56</sup>、これらを被説明変数とした<sup>57</sup>。

図20 消費者態度指数との相関が高いPOSの上位10品目  
（売上高（前年比）、2013年1月～17年12月）



（備考）内閣府「消費動向調査」、Finatextにより作成。

表13は式5による回帰分析の結果を整理したものであるが、経済面では10品目分類のうち5品目分類（冷凍総菜・畜肉製品、ウーロン茶・健康茶、スピリッツ、穀類、即席袋めん）で新聞センチメント指数とPOSの売上高との間に有意な関係が確認できた。中でも影響の度合いが最も大きかったのはスピリッツであった。また、有意になった5品目のうち穀類以外の4品目は1次ラグの係数が有意になっており、経済面の新聞センチメント指数の影響は1週間のラグを伴って消費行動（POSの売上）に影響を与えている可能性が指摘できる。プラスの方向とマイナスの方向を比較すると、両方の係数が有意になった冷凍総菜、スピリッツにおいては、マイナスの方向の係数がより高い。さらに、マイナスの方向のみが有意になったのは畜肉製品、即席袋めんの2通り、プラスの方向のみが有意になったケース

<sup>56</sup> 3章で作成した月次ベースのPOSデータと消費者態度指数の相関。

<sup>57</sup> 検証したいことは非対称性であるため、本分析において使用する品目分類の解釈についてはここでは省略する。

は穀類のみであった。

同様に、マーケット商品面の結果を確認すると、10品目分類のうち6品目分類（冷凍総菜、ベビー食事用品、畜肉製品、野菜ジュース、スピリッツ、即席袋めん、常温クールデザート・ケーキ）で有意な関係がみられた。影響の度合いは経済面のケースより全体的に大きく、中でもベビー食事用品における影響が大きい。また、有意になった6品目のうち5品目はラグ0の係数が有意になっており、経済面とは異なり、マーケット商品面の新聞センチメント指数の影響はほとんど時間を空けずにPOSの売上に影響を与える傾向が示唆されている。プラスの方向とマイナスの方向を比較すると、スピリッツ以外の5ケースにおいて両方の係数が有意になり、かつマイナスの方向の係数がより大きい<sup>58</sup>。また、スピリッツについては、マイナスの方向のみが有意であった。

表13 新聞センチメント指数が1標準偏差変化した場合における  
POSの前年比への影響（週次）  
（経済面（先行き））

	冷凍総菜		ベビー食事用品		畜肉製品		ウーロン茶・健康茶		野菜ジュース	
	+	-	+	-	+	-	+	-	+	-
同時決定 ( $\beta_0$ )	0.19	0.12	1.60	1.50	-0.18	-0.23	0.38	0.35	0.40	0.25
1期ラグ ( $\beta_1$ )	0.84 **	0.96 **	0.78	1.15	0.44	0.69 *	1.16	1.16	0.71	0.84

	スピリッツ		穀類		生タイプ即席袋めん		即席袋めん		常温クールデザート・ケーキ	
	+	-	+	-	+	-	+	-	+	-
同時決定 ( $\beta_0$ )	0.13	-0.25	1.11 *	1.22	-0.12	-0.20	-0.94	-1.05	0.33	0.43
1期ラグ ( $\beta_1$ )	1.84 ***	2.12 ***	0.35	0.53	0.39	0.55	1.20	1.48 *	-0.08	0.02

マーケット商品面（先行き）

	冷凍総菜		ベビー食事用品		畜肉製品		ウーロン茶・健康茶		野菜ジュース	
	+	-	+	-	+	-	+	-	+	-
同時決定 ( $\beta_0$ )	0.64	0.67	3.05 ***	3.65 **	1.14 **	1.50 **	0.82	0.97	1.57 **	1.65 **
1期ラグ ( $\beta_1$ )	0.27	0.20	-0.57	-0.20	-0.68	-0.56	-0.07	-0.39	0.03	0.14

	スピリッツ		穀類		生タイプ即席袋めん		即席袋めん		常温クールデザート・ケーキ	
	+	-	+	-	+	-	+	-	+	-
同時決定 ( $\beta_0$ )	0.94	0.71	0.97	0.87	0.36	0.34	1.64 *	2.42 **	1.05 **	1.28 *
1期ラグ ( $\beta_1$ )	1.01	1.32 *	1.15	1.10	0.24	0.02	-1.96 **	-1.89 *	-0.50	-0.47

- （備考）
1. 内閣府「消費動向調査」、Finatext、日本経済新聞社により作成。
  2. POSの各品目の売上高（前年比）を被説明変数、新聞センチメント指数などを説明変数とした式5を推計した。
  3. POSデータに対して、 $\beta_0$ は新聞センチメント指数が同時期、 $\beta_1$ は新聞センチメント指数が1期ラグの関係にある係数。+は前週比で新聞センチメント指数が上昇した場合、-は前週比で新聞センチメント指数が下落した場合を指す。
  4. \*\*\*は1%水準、\*\*は5%水準、\*は10%水準で有意であることを示す。

<sup>58</sup> なお、即席袋めんについては、同時決定の係数はマイナスの方向がより大きい、一期ラグの係数は両方が有意でマイナスとなっている。同時期の反応がオーバーリアクションであり、ラグ1でリバーサルが起きた可能性も考えられるが、このような特徴がみられたものは即席袋めんのみであるため、この分析結果から判断するのは困難である。

これらの分析を通じて、新聞センチメント指数の変動は、消費者マインドの変化等を通して、消費行動（売上高）に影響を与えている可能性が示唆されている。プラスの方向とマイナスの方向を比較すると、双方が有意になったすべてのケースにおいて、マイナスの係数の方が高い。また、プラスの方向のみが有意になったのは1つのみだが、マイナスの方向が有意になったケースは3つであり、総合すると新聞記事の内容がネガティブに変化した場合の方が、購買行動に与える影響がより大きい可能性が指摘できる。Soroka(2006)は、ネガティブなニュースの方が消費者マインドに与える影響が大きいことを指摘しているが、ネガティブなニュースは消費者マインドだけでなく、購買行動に与える影響も大きい可能性が指摘できる。

以上の分析を踏まえると、新聞センチメント指数は消費者マインド等の変化に関する情報を含んでいると考えられ、日次レベルのデータが利用可能な点をも踏まえると、消費動向を早期に把握する上で有益なデータの一つになり得ると評価できる。

他方、改良の余地も残されている。第一に挙げられるのは、分析対象データの拡大である。より多様な新聞社の記事をより長期に渡って入手することや、SNSの書き込みやブログの文章等も対象にすることで、より詳細な分析が可能になることが期待できる。また、岡崎・敦賀(2015)が指摘するように、テキストデータの分析は抽出語彙の選択やセンチメントの付与方法等によって分析結果が異なる可能性があるため、様々なパターンで形態素を抜粋することや、良い・悪いの単純な2値モデルではなく、中立（変わらない）を加えた3値モデルを採用すること等が考えられる。機械学習の手法についても、より高度な分析手法を導入することも考えられ、例えば、解釈性の向上に向けて、参木他(2018)のように、LSTM RNNモデルにAttention構造を追加し、形態素ごとに新聞センチメント指数に対する寄与度を算出することも選択肢の一つである。さらに、POSとテキストの分析は、今回は単純な回帰分析であったが、VARモデル等の異なる分析手法を行うことも考えられる。

今後も、新聞などのテキストデータに関しては、テキストに内在する多様な情報をより正確に把握することで、ナウキャストや経済分析に役立つ研究を進めていく必要がある。

## 6. まとめ

本稿では、消費分析に対して新しい視点を提供することを目的に、POS・テキストデータをもとに機械学習の手法も活用しながら、主に3つの論点について分析を行った。主な分析結果は以下の通りである。

(1) 一般的な需要・供給曲線を想定し、POSの価格・数量データを用いて、価格・数量変化が需要要因と供給要因のどちらに起因するのかの要因分解を行った。価格・数量変化の詳細な背景情報を得ることができ、消費者の購買意欲等を把握できる可能性が提示された。

(2) POSデータ等の速報性の高いデータから、機械学習により小売業全体の販売額の動向のナウキャストを行った。POSデータのみでも、一定程度のナウキャストを行うことが可能であったが、天候データも加えることで、ナウキャストの精度が向上するとの結果が得られた。

(3) ディープラーニングの手法を用いて、新聞の内容がどの程度ポジティブであるかを示す指数を紙面別に作成し、消費者マインド等との相関を確認した。全般的な経済情勢の変化や価格変化等の情報が多く含まれている紙面と消費者マインドに一定程度の相関が確認できたほか、記事の内容がネガティブに変化した場合の方が、消費行動に与える影響が大きい可能性が示唆された。

以上のように、POS・テキストデータや機械学習を活用することで、経済分析をよりの確かつ迅速に行うことができる可能性が高い。今回の分析結果は暫定的なものであるが、今後、より分析を発展させていくことで、適切かつ機動的な経済財政運営に資することができると思われるため、引き続き研究を進めていくことが必要である。今後の改善点としては、例えば、今回分析に利用したPOSデータは、日次の前年比データであったため、曜日間や品目分類間のウェイトが考慮できていない等の課題がある。また、本稿で活用した機械学習の手法は、ナウキャストの観点から予測に重点を置いたが、消費の構造的な分析を行う際には、機械学習による因果構造の推定を行うことも必要である。さらに、より長期かつ豊富なテキストデータを利用することで、他の時系列データと組み合わせた分析・予測を行うことや、より高度な機械学習の手法を活用すること等も考えられる。

ただし、このような機械学習を利用する際には、その限界についても十分理解することが極めて重要である。機械は入力・出力されるデータや分析内容が意味することについては、一切理解していないため、分析手法・結果の妥当性を判断するのは人の仕事である。また、機械学習による分析結果は、あくまでも数値データから観測される経験則であるため、経済学等による背景理解が重要となることも忘れてはいけない。

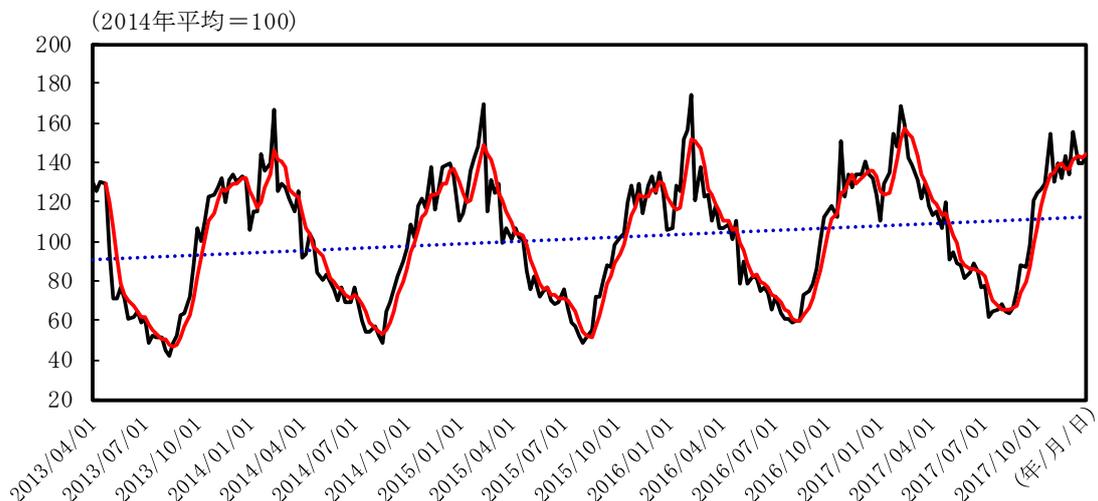
## BOX 2 チョコレートのPOSデータからみる消費者の嗜好変化

ここではチョコレートを題材に、POSデータが保有する売上や数量等の数値データと商品名等のテキストデータを組み合わせた分析によって、消費者の嗜好の変化がどの程度読み取れるかを試験的に分析した。

本分析では本編（価格・数量変化の分析やナウキャスト分析）で使用したPOSデータとは異なり、株式会社 True Data が収集する全国のドラッグストアのPOSデータを使用した。データの対象期間は、2013年4月～2017年12月までの4年9か月の週次データであり、対象店舗のドラッグストア<sup>59</sup>で販売された3,000点以上のチョコレート商品の商品名、購入者数、数量、売上高等がセットで利用可能なPOSデータである。

チョコレートの売上高の週次の推移はBOX図2-1の通りである。売上の山が2月に、谷が8月になる周期を繰り返しながら、年々売上高が増加していることがわかる。1年間のデータが揃う2014年から2017年の売上高の伸び率は+12.0%であった<sup>60</sup>。この売上高の伸び率に対し、購入者数の伸び率は+4.7%とそこまで大きくない。売上高と購入者数、購入品数の関係性と伸び率をBOX表2-1で確認すると、購入者数は増加している反面、一人当たりの購入品数はむしろ減少しており、購入品数で見ると微増にとどまっている。売上高の増加に最も寄与しているのはチョコレート単価の上昇であり、消費者がチョコレートに対して高価格志向となっている可能性が考えられる。

BOX図2-1 チョコレートの売上高の推移



(備考) 株式会社 True Data により作成。

<sup>59</sup> 株式会社 True Data が保有するドラッグストアパネルの内、対象期間に開店している店舗のPOSデータを対象としている。

<sup>60</sup> なお、全日本菓子協会が公表しているチョコレートの小売金額をみると、同期間の伸び率は+13.2%であり、概ね同じ伸び率を示している。

BOX表2-1 売上高・購入者数・購入品数の伸び率

	2014年	2017年	伸び率(%)
購入者一人当たりの購入品数(個/人)	1.32	1.26	▲ 4.3
購入者一人当たりの購入金額(円/人)	151.5	162.0	6.9
購入品数当たりの購入金額(円/個)	114.9	128.4	11.8
売上高			12.0
購入者数			4.7
購入品数			0.2

- (備考) 1. 株式会社 True Data により作成。  
2. 伸び率は2014年から2017年の増加率を示す。

このように消費者がチョコレートに対してより高い価格を払ってもよいと考える背景にはどのようなことが考えられるだろうか。この点を考察するには、POSデータの数値情報だけでは限界があるため、テキスト情報である商品名に着目し、商品名に特定のキーワード(単語)を含む商品の売上高と単価を集約・集計することで考察を行う(BOX表2-2)<sup>61</sup>。

BOX表2-2 売上高・購入者数・購入品数の伸び率

分類	キーワード	総額に占めるシェア(%)			平均単価(円)			総額
		2014年	2017年	差分	2014年	2017年	差分	伸び率(%)
健康志向	カカオ	0.99	5.69	4.70	189.5	202.2	12.6	543.1
	糖質	0.04	0.09	0.05	110.1	247.4	137.3	150.9
	乳酸菌		1.56	1.56		219.5		
味	抹茶	2.52	1.79	▲ 0.74	142.3	187.3	45.0	▲ 20.8
	苺	7.20	7.32	0.12	107.9	110.4	2.5	13.9
趣き	大人	4.78	3.07	▲ 1.72	187.9	206.7	18.8	▲ 28.2
	贅沢	1.76	1.77	0.01	189.2	229.9	40.7	12.5
	濃厚	1.41	2.19	0.78	177.3	200.3	23.1	74.1
	大人数	4.87	4.28	▲ 0.59	215.0	209.2	▲ 5.8	▲ 1.6
時期・季節	限定	0.03	0.62	0.58	75.1	170.4	95.2	2,152.8
	ハロウィン	0.44	0.43	▲ 0.01	169.7	196.5	26.8	9.8
	冬	0.30	0.70	0.41	163.4	150.6	▲ 12.8	165.2
	受験	0.01	0.01	0.00	101.4	154.0	52.5	36.0
全体				114.9	128.4	13.5	12.0	

- (備考) 1. 株式会社 True Data により作成。  
2. 伸び率は2014年から2017年の増加率を示す。

<sup>61</sup> 例えば、「濃厚」というキーワードであれば、「濃密」「まろやか」「深み」等の単語が商品目にある商品を集約した。キーワードに該当する単語の内訳は補論7を参照。

まず、指摘できるのは健康志向の高まりである。「カカオ」がキーワードの商品は、平均単価が全体と比べて1.5倍以上高いにも関わらず、3年でシェアを約5.7倍拡大させ、総額の伸び率は5倍を超えている。「カカオ」には健康志向のイメージが強いことが考えられるが<sup>62</sup>、より健康的な商品であるなら、少し高額でも買ってよいくと考える消費者の嗜好が読み取れる。同様の結果は「糖質」にもみられ、更には、2014年に販売されてなかった「乳酸菌」が2017年には一定のシェア占める等、健康志向への関心が高まっている可能性がある。

また、「限定」がキーワードの商品も、全体に占めるシェアは小さいが総額も含め急激な伸びを示している。平均単価に着目すると、2014年では全体平均よりも安価な額であったのに対し、2017年には2倍以上に上昇し、全体平均よりも高額になっている。「限定」は時期や味、数量といった多様な意味合いで使用されるが、より特別な物を求める消費者の嗜好が強くなっている可能性がうかがえる。

今回の分析はチョコレートだけに絞った限定的なものであるが、このようにPOSデータにテキストデータ等の情報を組み合わせた分析を様々な商品に対して行うことで、消費者の嗜好の変化といった消費者行動の構造的な変化をより詳細に分析することが可能となるとみられる。

---

<sup>62</sup>日本チョコレート・ココア協会によれば、カカオ豆に含まれる「カカオポリフェノール」が健康に良いと紹介している。このような健康に関する発表報道やダイエット効果があると各種メディアで報道されたことが、カカオの健康イメージの定着に繋がったと考えられる。

## 参考文献

- 饗場行洋・山本裕樹 (2018) 「データサイエンスと新しい金融工学」 野村證券
- 上田晃三・須藤直・渡辺広太 (2016) 「POSデータによる「特売」の分析」、渡辺努編『慢性デフレ真因の解明』、p. p97-114、日本経済新聞社
- 岡崎陽介・敦賀智裕 (2015) 「ビッグデータを用いた経済・物価分析について」 日本銀行調査論文
- 岡谷貴之 (2015) 『深層学習』 講談社
- 経済産業省 (2017) 『IoT推進のための新産業モデル創出基盤整備事業 (ビッグデータを活用した新指標開発事業)』
- 五島圭一・山田哲也・高橋大志 (2017) 「畳み込みニューラルネットワークを用いた日次景況感指数の構築と資産価格変動との関連性」 日本ファイナンス学会
- 塩野剛志 (2018) 「人工知能とテキスト・データを活用した数量分析」 日本銀行金融研究所 ディスカッション・ペーパー・シリーズ 2018-J-9
- 内閣府政策統括官 (経済財政分析担当) (2018) 『日本経済 2017-2018-成長力強化に向けた課題と展望-』
- 西尾泰和 (2014) 『word2vec による自然言語処』 O'Reilly
- 日本銀行 (2017) 「企業のインフレ予想形成に関する新事実: Part II-機械学習アプローチ」 日本銀行ワーキングペーパーシリーズ
- 日本銀行 (2016) 『経済・物価情勢の展望』 2016年4月
- 藤田隼平 (2017) 「POSデータを用いた経済分析の試み-小売価格と景気動向との関係性の検証-」 経済財政分析 ディスカッション・ペーパー・シリーズ DP/17-4
- 丸山歩・嶋北俊一・落合牧子・上田聖 (2015) 「CPI と東大指数の乖離の分析について」 統計研究彙報 第72号 2015年3月 (55~78)
- 参木裕之・北野道春・渡部裕晃 (2018) 「Self-Attentive-LSTM への順序ロジットモデル適用」 第32回人工知能学会全国大会
- 諸田崇義 (2010) 「コモディティ価格変動の特徴とプライシング・モデルの展開」 『金融研究』 2010.4. 27-72
- 山本裕樹・松尾豊 (2016) 「景気ウォッチャー調査の深層学習を用いた金融レポートの指数化」 第30回人工知能学会全国大会
- 渡辺広太・渡辺努 (2013) 「スキャナーデータを用いた日次物価指数の計測」 東京大学金融教育研究センターワーキングペーパー CARF-J-094
- 渡辺努 (2016) 「店舗別インフレ率から読み取れること」 ナウキャスト 『マンスリーレポート 2016年2月号』 2016年2月17日
- Chollet, F. (2017), *Deep learning with python*, Manning Publications Co.
- De Boef, S., and Kellstedt, P. M. (2004), "The political (and economic) origins of consumer confidence", *American Journal of Political Science*, 48(4), 633-649.

- Doms, M. E., and Morin, N. J. (2004), "Consumer sentiment, the economy, and the news media", *FRBSF Working paper*
- Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016), *Deep learning*, Cambridge: MIT press.
- Heston, S. L., and Sinha, N. R. (2017), "News vs. Sentiment: Predicting Stock Returns from News Stories", *Financial Analysts Journal*, 73(3), 1-17.
- Hollanders, D., and Vliegthart, R. (2011), "The influence of negative newspaper coverage on consumer confidence: The Dutch case", *Journal of Economic Psychology*, 32(3), 367-373.
- Imai, S., and Watanabe, T. (2015), "Replicating Japan's CPI Using Scanner Data", Replicating Japan's CPI Using Scanner Data", *JSPS Grants-in-Aid for Scientific Research, Working Paper Series No. 072*
- Müller, F. (2018), "XGBoost Tree vs. Linear", STATWORK Blog, 12 January 2018
- Shapiro, A. H., Sudhof, M., and Wilson, D. (2018), "Measuring news sentiment", Federal Reserve Bank of San Francisco
- Soroka, S. N. (2006), "Good news and bad news: Asymmetric responses to economic information", *Journal of Politics*, 68(2), 372-385.
- Ueda, K., Watanabe, K., and Watanabe, T. (2018), "Product Turnover and the Cost of Living Index: Quality vs. Fashion Effects" *Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper No. 337*
- Watanabe, T., Muroi, H., Naruke, M., Yono, K., Kobayashi, G., and Yamasaki, M. (2016), "Prediction of regional goods demand incorporating the effect of weather", In Big Data (Big Data), 2016 IEEE International Conference on (pp. 3785-3791). IEEE.

補説1 POSの品目分類と各系列の記述統計量

品目 コード	品目名	品目例	2015年時点の ウェイト (10万分比)	記述統計量 (日次データ)				
				売上高		購入価格 (T指数)		
				平均値 (%)	標準偏差 (%pt)	平均値 (%)	標準偏差 (%pt)	
001	豆腐・豆腐製品	木綿豆腐等	2293	0.96	19.03	0.18	2.76	10461
002	納豆	粒納豆等	1062	1.63	31.22	-0.25	3.29	10461
003	コンニャク	板コンニャク等	487	-1.38	22.40	0.17	3.31	10461
011	漬物	たくあん漬等	1804	0.42	18.07	0.28	2.64	10461
012	煮豆・きんとん	大豆煮等	319	6.47	210.77	0.28	2.60	10458
013	つくいだ煮	昆布つくいだ煮等	536	3.54	70.38	0.39	1.98	10458
014	総菜・弁当	サラダ、弁当等	1818	9.83	28.14	-0.30	2.36	10461
021	かまぼこ	板かまぼこ等	419	0.97	76.20	0.52	4.38	10461
022	ちくわ	生ちくわ等	402	-0.62	38.36	0.60	5.17	10458
023	水産練り製品・流し物	なると巻等	182	3.16	72.23	0.77	4.45	10459
024	水産練り製品・揚げ物	さつま揚げ等	536	0.15	37.92	0.54	4.70	10457
025	水産加工品	塩辛等	559	4.45	34.11	-0.05	2.10	10459
026	卵製品	茶碗蒸し等	274	2.85	42.98	-0.08	2.34	10456
027	チルド半製品	チルドシューマイ等	1488	0.82	21.74	-0.36	1.84	10461
028	チルド調味料	チルドドレッシング等	78	-0.04	39.45	0.33	2.15	10457
029	生めん・ゆでめん	生中華そば等	1395	0.47	59.49	-0.13	3.10	10461
031	ハム・ベーコン	スライスハム等	1420	2.14	40.39	-0.29	2.42	10461
032	ソーセージ	ウインナー等	1826	-0.10	23.26	-0.49	2.57	10460
033	畜肉製品	焼き豚等	813	3.12	100.36	-0.09	2.25	10458
041	バター	バター等	296	1.78	32.07	1.08	4.49	10459
042	マーガリン・ファットスプレッド	マーガリン等	219	-2.41	24.71	-0.09	3.70	10459
043	ナチュラルチーズ	カマンベールチーズ等	376	3.13	19.89	0.34	3.59	10461
044	プロセスチーズ	箱入りプロセスチーズ等	707	2.11	22.59	0.24	4.37	10459
045	ヨーグルト	プレーンヨーグルト等	2177	3.74	16.59	-0.67	2.03	10461
046	牛乳	一般牛乳等	2510	-1.79	15.10	0.03	2.11	10461
047	乳飲料	コーヒー乳飲料等	629	1.60	22.10	-0.50	2.23	10460
048	乳酸菌飲料	乳酸菌飲料等	424	0.57	16.66	0.02	2.51	10461
049	生クリーム	生クリーム等	140	0.99	31.98	-0.16	1.58	10459
050	豆乳類	豆乳等	387	10.30	31.96	0.11	3.00	10456
061	チルドケールデザート	チルドプリン等	301	-0.14	122.94	-0.70	1.79	10461
062	チルドケーキ	チルドシュークリーム等	401	7.15	32.48	-1.17	3.00	10456
071	コーヒー飲料	缶入りコーヒー飲料等	579	4.59	22.26	-0.94	3.14	10461
072	ココア・チョコレート飲料	缶入りココア等	14	8.04	47.23	-0.34	4.76	10455
073	紅茶飲料	缶入り紅茶飲料等	195	9.21	37.10	-1.34	2.91	10461
074	緑茶飲料	缶入り緑茶飲料等	623	23.39	50.23	-1.52	3.39	10254
075	麦茶飲料	缶入り麦茶飲料等	107	13.45	38.41	-1.24	4.27	10345
076	ウーロン茶飲料	缶入りウーロン茶飲料等	121	1.86	25.25	-1.61	2.72	10461
077	健康茶飲料	缶入り健康茶飲料等	184	44.33	206.86	-1.37	3.03	9624
078	炭酸飲料	缶入りコーラ飲料等	934	3.56	26.24	-1.47	2.55	10461

品目 コード	品目名	品目例	2015年時点の ウエイト (10万分比)	配統統計量 (日次データ)					
				売上高		購入価格 (T指数)			
				平均値 (%)	標準偏差 (%pt)	データ数	平均値 (%)	標準偏差 (%pt)	データ数
079	清涼飲料	缶入り清涼飲料等	408	0.47	21.27	10466	-0.71	1.75	10460
080	果汁100%飲料	缶入り果汁100%飲料等	401	0.14	20.55	10466	-0.83	2.35	10461
081	野菜ジュース	缶入りトマトジュース等	548	7.45	31.42	10466	-0.55	3.06	10461
082	スポーツ飲料	缶入りスポーツ飲料等	309	3.67	27.22	10466	-1.50	2.77	10460
083	希釈飲料	希釈乳酸菌飲料等	51	10.19	64.93	10460	-0.29	7.27	10228
084	栄養サポートドリンク	栄養強化ドリンク等	781	11.61	41.29	10466	-0.39	2.44	10456
085	水	缶入りミネラルウォーター等	422	16.85	51.46	10464	-2.36	4.29	10460
101	ノリ	焼き全型ノリ等	469	0.86	26.33	10466	0.42	2.90	10461
102	水産乾物	ワカメ等	810	0.72	34.53	10466	0.54	2.12	10461
103	粉類	強力小麦粉等	451	-0.29	21.04	10466	0.44	2.72	10461
104	ゴマ	いりゴマ等	124	4.20	30.88	10464	1.11	4.19	10457
105	乾燥豆	乾燥大豆等	69	1.11	43.02	10464	1.30	8.78	10457
106	農産乾物	干しシイタケ等	331	0.73	39.53	10466	0.78	2.51	10460
107	乾めん	干しうどん等	427	2.40	45.79	10465	0.42	3.06	10460
108	乾パスタ	スパゲティ等	252	0.14	19.91	10466	0.60	7.54	10458
111	砂糖・甘味料	白砂糖等	495	-0.45	30.61	10466	-0.18	4.23	10461
112	食塩	精製塩等	145	1.47	30.97	10466	0.60	2.95	10460
113	みそ	一般みそ等	586	-0.23	27.66	10466	0.07	3.67	10458
114	しょうじ	こうじ	16	16.65	81.92	9360	0.57	6.56	9195
115	しょうゆ	濃口しょうゆ等	450	1.17	30.02	10466	0.21	5.05	10461
116	食用酢・酢関連調味料	穀物酢等	309	2.01	38.98	10466	0.15	2.36	10460
117	みりん・調理酒	本みりん等	278	2.41	33.61	10466	-1.05	3.50	10461
118	食用油	サラダ油等	781	3.49	34.96	10466	-0.20	5.91	10460
119	テールソース	ウスターソース等	190	-0.45	22.02	10466	0.06	3.77	10458
120	トマト調味料	トマトケチャップ等	194	-0.23	23.82	10466	-0.87	4.43	10459
121	マヨネーズ	マヨネーズ等	454	-0.26	31.72	10466	-0.33	4.48	10461
122	ドレッシング	ドレッシング等	569	3.22	24.88	10465	-0.23	2.30	10457
123	うま味調味料類	うま味調味料等	27	-2.66	27.72	10466	0.21	2.50	10458
124	だしのもと	和風だしのもと等	554	1.29	30.86	10466	-0.22	2.86	10461
125	香辛料	コシユウ等	479	1.39	22.24	10466	0.19	2.54	10459
126	香辛・調合調味料	オイスターソース等	150	2.62	17.95	10464	-0.10	1.59	10457
127	つゆ	万能つゆ等	540	3.68	53.71	10466	-0.87	2.53	10460
128	和風調味料・ソース	焼肉のたれ等	623	1.38	18.77	10466	-0.26	2.08	10461
129	洋風・その他調味料・ソース	デミグラスソース等	115	2.74	23.19	10465	-0.10	2.00	10457
130	鍋つゆ	おき焼き用割り下等	369	10.64	30.54	10464	-0.80	2.62	10458
131	カレー	カレー等	657	-0.31	16.46	10466	0.11	3.66	10459
132	シチュー・ハヤシ	シチュー等	177	0.94	19.63	10466	0.56	3.57	10457
133	即席スープ	カップスープ等	495	2.16	29.38	10466	-0.34	2.16	10458
134	即席みそ汁・和風汁	即席みそ汁等	240	3.86	25.99	10466	-0.23	2.20	10456
135	パスタソース	パスタソース等	279	1.27	20.34	10466	-0.22	2.14	10459
136	即席袋めん	即席袋中華そば等	515	0.42	31.82	10466	0.65	6.66	10461
137	即席カップめん	即席カップ中華そば等	1104	3.03	30.00	10466	0.44	5.81	10461

品目 コード	品目名	品目例	2015年時点の ウエイト (10万分比)	配述統計量 (日次データ)					
				売上高			購入価格 (T指数)		
				平均値 (%)	標準偏差 (%pt)	データ数	平均値 (%)	標準偏差 (%pt)	データ数
138	即席食品	即席白飯等	451	6.23	23.75	10466	-0.22	2.69	10460
139	ふりかけ・茶漬け	ふりかけ等	289	0.09	23.44	10466	0.23	2.37	10459
140	米関連即席調味料	ちらし寿司のもと等	208	1.16	33.07	10466	-0.12	1.90	10458
141	料理用即席調味料	中華料理のもと等	665	2.42	16.48	10466	-0.29	1.43	10460
142	ケーシング詰め水産練り製品	魚肉ソーセージ等	216	-0.79	39.28	10463	0.19	3.75	10457
144	即席カップ入りスープ・汁	即席カップ入りスープ等	69	12.57	37.63	10456	-0.31	2.83	10430
146	生タイプ即席袋めん	生タイプ即席中華そば等	13	-2.34	31.07	10464	0.86	4.98	10458
147	生タイプ即席カップめん	生タイプ即席カップ中華そば等	11	16.33	100.10	10459	0.40	4.96	10445
151	農産缶詰	コーン缶詰等	138	-1.56	17.49	10466	0.38	2.80	10459
152	果実缶詰	ミカン缶詰等	94	-1.81	33.86	10466	1.54	5.52	10456
153	デザート缶詰	フルーツみつ豆等	25	-1.31	25.50	10465	0.52	3.14	10461
154	水産缶詰	マグロ油漬缶詰等	483	-0.29	23.20	10466	0.62	3.61	10460
155	畜肉缶詰	コロンビー缶詰等	49	1.18	35.69	10462	0.82	4.59	10456
156	総菜缶詰	和風総菜缶詰等	8	2.25	36.45	10456	0.36	4.51	10439
157	農産缶詰	メンマ缶詰等	74	-0.29	35.24	10463	0.50	2.81	10458
158	水産缶詰	ノリつくだ煮缶詰等	149	-0.09	20.70	10466	0.13	2.46	10459
159	畜肉缶詰	畜肉缶詰	1	9.69	54.56	8444	-0.11	6.56	7987
161	食パン	普通食パン等	1592	0.45	13.61	10465	-0.15	3.98	10460
162	テールパン	ロールパン等	761	0.51	18.74	10465	-0.08	2.61	10459
163	菓子パン・蒸しパン	菓子パン等	2355	3.06	11.57	10465	-0.05	3.19	10461
164	調理パン	サンドイッチ等	177	3.92	18.27	10461	-0.28	2.90	10458
165	シリアル類	シリアル等	267	2.22	22.93	10463	-0.43	2.29	10456
166	もち	切りもち等	292	4.53	59.18	10466	0.25	4.59	10461
171	ジャム	ジャム等	227	1.54	21.34	10463	-0.21	2.68	10456
172	スブレッド	ピーナツバター等	84	2.22	22.35	10461	0.15	1.81	10456
173	みつ・シロップ	はちみつ等	170	5.02	43.30	10462	0.74	2.28	10456
174	デザートミックス	プリンのもと等	35	-1.38	24.61	10461	0.52	3.93	10456
175	ブレミックス	ホットケーキミックス等	178	2.29	24.33	10463	0.15	3.12	10456
176	ケーキ・パン材料	香料・エッセンス等	115	2.85	38.48	10459	0.43	1.31	10456
181	レギュラーコーヒー	レギュラーコーヒー等	349	1.90	42.38	10464	-0.73	5.09	10456
182	インスタントコーヒー	インスタントコーヒー等	715	2.67	34.29	10466	-0.60	5.92	10458
183	ココア・牛乳用ドリンクミックス	粉末ココア等	98	1.92	35.68	10460	-0.57	2.85	10456
184	紅茶	紅茶等	126	0.40	20.17	10460	-0.27	3.15	10456
185	緑茶	せん茶等	505	1.83	24.36	10466	-0.16	1.53	10458
186	麦茶	麦茶等	57	1.96	25.79	10457	-0.35	2.62	10456
187	ウーロン茶・健康茶	ウーロン茶等	84	0.55	24.32	10460	-0.44	1.74	10455
188	脱脂粉乳・クリーミングパウダー	クリーミングパウダー、練乳等	175	-3.81	17.92	10466	-0.48	2.08	10461
191	チョコレート	板チョコレート等	753	2.63	31.52	10466	-0.45	1.44	10461
192	チョコレートガム	板ガム等	220	-1.20	16.29	10466	-0.33	1.14	10461
193	キャンデー・あめ菓子	袋入りハードキャンディー等	712	-0.61	13.57	10466	-0.06	1.48	10461
194	スナック菓子	ポテトチップス等	793	0.32	16.03	10466	-0.47	1.37	10461
195	洋焼き菓子	クッキー等	859	0.22	16.95	10466	-0.45	1.45	10461

品目 コード	品目名	品目例	2015年時点の ウエイト (10万分比)	配述統計量 (日次データ)					
				売上高			購入価格 (T指数)		
				平均値 (%)	標準偏差 (%pt)	データ数	平均値 (%)	標準偏差 (%pt)	データ数
196	常温クールデザート・ケーキ	カステラ、常温ゼリー等	553	2.40	16.07	10465	-0.53	2.05	10460
197	せんべい	せんべい、あられ等	1126	-0.50	16.11	10466	-0.06	1.20	10461
198	和菓子	まんじゅう等	1195	3.19	24.16	10466	-0.37	1.68	10461
199	駄菓子	わた菓子、袋入りドーナツ等	344	0.55	17.31	10466	-0.27	1.63	10461
200	玩具入り菓子	玩具入り菓子等	10	2.25	27.75	10466	-0.49	1.53	10457
201	豆菓子・いかり豆	豆菓子等	187	4.24	47.15	10466	-0.10	2.00	10459
202	水産珍味	さきイカ等	292	-0.26	29.38	10464	-0.15	0.94	10457
203	畜産珍味	ジャージャー珍味等	4	6.44	40.52	10454	-0.08	3.64	10413
204	ナッツ	ピーナッツ等	205	5.93	26.84	10462	-0.13	2.51	10457
205	ドライフルーツ	レーズン等	63	5.52	27.52	10463	0.52	2.96	10456
206	各種詰め合わせ菓子	各種詰め合わせ菓子等	22	5.64	59.80	10457	0.02	2.12	10455
211	日本酒	瓶入り清酒等	1014	9.02	31.12	10456	-0.28	1.52	10444
212	ビール	瓶入りビール等	2162	4.69	26.01	10457	-0.18	2.11	10452
213	ウイスキー・ブランデー類	ウイスキー等	366	16.45	167.85	10431	-0.88	7.03	10160
214	焼酎類	甲類焼酎等	1658	13.02	38.60	10446	0.72	3.44	10371
215	ワイン	赤ワイン等	649	13.98	63.90	10452	-0.25	2.28	10244
216	リキュール	梅酒等	121	28.92	248.52	10317	-0.32	1.87	10095
217	スピリッツ	ジン、ウォッカ等	5	6.66	43.98	9856	-0.29	2.70	9057
218	中国酒	老酒等	5	3.05	37.21	9346	0.15	1.82	9062
219	カクテルドリンク類	チューハイ等	776	19.36	70.59	10448	-0.24	2.41	10428
220	雑酒	甘酒、マッコリ酒等	62	16.27	77.13	9997	0.04	3.47	9787
221	発泡酒	缶入り発泡酒等	2828	12.54	36.45	8028	-0.65	2.81	8022
222	アルコールテイスト飲料類	ビールテイスト飲料等	273	21.51	72.11	10447	-0.82	8.35	10313
223	酒類関連飲料	チューハイベース等	9	3.36	31.59	10455	0.05	1.96	10443
231	ベビー・マタニティフード	粉ミルク等	152	5.76	33.69	10460	-0.79	2.65	10456
232	栄養補給食品類	栄養補給菓子等	216	11.60	37.65	10457	-1.02	2.42	10455
233	食品ギフトセット・商品券	調味料・食料油ギフトセット等	1691	16.76	74.03	10423	-0.03	2.41	9802
234	穀類	うるち米等	3655	9.70	38.99	10465	-0.66	8.87	10461
235	生鮮卵	鶏卵等	2175	6.60	31.18	10465	1.21	8.79	10461
236	介護・病時食類	介護・病時食等	13	21.26	100.66	6336	-0.62	5.81	5989
251	冷凍素材	冷凍シーフードミックス等	374	3.43	34.18	10466	0.02	4.83	10457
252	冷凍総菜	冷凍ギョーザ等	2150	3.81	26.73	10466	-0.52	4.65	10458
261	レギュラーアイス	カップアイス等	1082	3.88	24.97	10465	-0.36	2.89	10461
262	プレミアムアイス	パーションナルブレミアムアイス等	130	4.53	35.51	10465	-0.64	3.91	10458
263	水	水	71	6.37	27.89	10465	0.20	4.10	10460
271	加熱用野菜	野菜ミックス等	350	11.94	75.60	10455	0.02	5.03	10437
272	きのこ	生しいたけ等	1199	8.86	25.93	10465	-1.47	7.96	10461
601	洗髪剤	シャンプー、リンス等	332	-0.41	22.69	10466	-2.77	2.57	10458
602	せっけん	固形せっけん、洗顔せっけん等	375	2.28	23.61	10466	-1.99	2.27	10456
603	入浴剤	入浴剤等	98	0.83	28.72	10466	-1.92	3.27	10456
604	歯磨き類	練り歯磨き、デンタルリンス等	255	0.50	23.24	10466	-1.54	2.01	10460
605	歯ブラシ類	一般歯ブラシ等	181	0.74	24.40	10466	-1.48	2.08	10460

品目 コード	品目名	品目例	2015年時点の ウエイト (10万分比)	配述統計量 (日次データ)					
				売上高			購入価格 (T指数)		
				平均値 (%)	標準偏差 (%pt)	データ数	平均値 (%)	標準偏差 (%pt)	データ数
606	口中清涼剤	液体口中剤等	33	4.16	24.01	10457	-0.92	3.09	10434
607	携帯用サニタリーセット	携帯用オーラルケア商品セット等	9	-2.24	20.79	10459	-0.42	1.79	10456
608	生理用品	生理用ナプキン等	236	-2.51	17.93	10466	-2.18	2.27	10457
609	避妊用品	コンドーム等	0	0.58	28.62	10454	0.39	6.09	10251
610	日用紙製品	箱入りティッシュ等	807	0.66	27.84	10466	-0.80	4.78	10460
611	紙おむつ類	ベビー用紙おむつ等	462	4.35	36.39	10460	-3.49	4.28	10455
612	洗濯用洗剤類	一般衣料品用粉末洗剤等	603	1.74	29.50	10466	-2.79	3.47	10458
614	台所食器洗剤	液体食器洗剤等	161	-0.59	25.04	10466	-1.32	2.81	10461
615	住居洗剤 (クリーナー)	トイレ用クリーナー等	253	2.24	47.97	10466	-1.21	1.76	10458
616	消臭・芳香・除菌剤	芳香剤等	196	0.37	27.13	10466	-0.88	1.68	10457
617	除湿剤	除湿剤等	18	1.00	33.48	10462	-1.65	4.10	10456
619	殺虫剤・殺菌剤	蚊取り線香等	139	4.41	35.99	10440	-0.62	3.01	10286
620	防虫剤	衣料用防虫剤等	46	0.25	29.17	10460	-0.90	3.28	10456
621	介護・衛生用品	ベビー用おしり拭き等	39	8.20	39.69	10457	-1.92	4.98	10427
625	入れ歯関連用品	入れ歯洗浄剤等	45	7.71	30.57	10412	-1.65	4.49	10244
631	女性用基礎化粧品	女性用化粧水、女性用パック等	1440	13.66	51.21	10461	-1.13	1.71	10455
632	女性用メーキャップ化粧品	ファンデーション、口紅等	833	8.72	35.47	10459	-0.73	1.54	10455
633	女性用頭髪化粧品	女性用スタイリングスプレー等	255	1.42	26.40	10466	-1.39	2.10	10457
634	フレグランス	香水等	15	7.97	43.69	10444	-1.74	6.27	9670
635	男性用化粧品	男性用化粧水等	23	3.57	27.44	10460	-0.50	2.12	10453
636	化粧品	ハンドクリーム、脂取り紙等	197	5.22	30.20	10466	-0.97	1.85	10458
637	男性用頭髪化粧品	男性用ヘアワックス等	84	0.04	27.06	10462	-1.14	2.59	10455
639	エチケット用品	制汗剤等	56	1.70	24.76	10457	-0.87	2.24	10455
642	かみそり	使い切り男性用かみそり等	36	-0.55	24.17	10466	0.07	1.85	10458
671	家庭医療用品	救急ばんそうこう、マスク等	220	7.27	45.92	10464	-1.02	2.19	10459
674	ベビー食器用品	ベビー用容器等	1	16.03	129.89	7524	0.13	9.36	6937
681	たばこ・喫煙関連用品	たばこ、使い切りライター等	2243	6.23	44.23	10461	2.38	7.54	10456
691	洗面・浴室用品	洗面器、ふろいす等	11	-2.00	30.39	10456	-0.80	4.43	10428
693	洗濯・物干し用品	物干しお、洗濯ばさみ等	37	0.57	28.19	10460	-0.45	4.15	10454
694	清掃手入れ用品	粘着テープ式クリーナー等	124	7.70	65.31	10463	-1.07	2.60	10456
697	雑貨品	マツパ、線香等	99	2.29	24.99	10459	0.86	3.29	10456
699	トイレ清掃用品	トイレブラシ等	4	0.27	32.66	10454	-0.32	4.58	10115
704	調理・キッチン用品	ラッピングフィルム等	551	1.40	32.18	10466	-0.72	2.09	10460
705	流し用品	たわし、スポンジ等	440	0.21	22.59	10464	-0.15	2.11	10457
706	食品容器	調味料保存容器等	0	9.79	71.71	8226	0.25	10.15	5989
710	モップ	モップ	0	8.69	70.82	10372	-0.58	31.08	8921
711	使い切り食器	割りばし、使い切りコップ等	90	-2.20	28.80	10466	-0.26	1.68	10459
713	レジャー食器用品	弁当箱、水筒等	136	14.95	78.25	10348	-3.42	6.32	9412
725	耐久流し用品	洗いおけ、キッチンコーナー等	7	0.63	35.43	10457	-0.54	4.45	10430
731	電池	マンガン電池等	180	4.40	68.28	10466	-0.87	2.37	10457
741	文具紙製品	ノート等	92	2.00	25.47	10460	0.67	3.52	10433
742	日用文具	ホチキス用針、ガムテープ等	66	0.45	21.33	10466	0.00	2.92	10454

品目 コード	品目名	品目例	2015年時点の ウェイト (10万分比)	記述統計量 (日次データ)					
				売上高			購入価格 (T 指数)		
				平均値 (%)	標準偏差 (%pt)	データ数	平均値 (%)	標準偏差 (%pt)	データ数
744	筆記用具	鉛筆、ボールペン等	208	4.67	39.79	10462	-0.06	1.85	10441
745	絵画用品	絵具、クレヨン等	12	14.44	108.31	9839	0.16	5.08	9085
751	OA サブライ	OA 用紙等	92	14.88	92.23	9479	-0.73	3.91	8834
752	書類整理用品	ファイル、カードケース等	22	4.46	56.49	10420	0.01	3.63	10152
765	フック類	吸盤、キッチンフック等	3	-3.54	22.88	10457	0.09	3.26	10443
817	ペットサニタリー用品	ペット用シヤンブー等	97	5.56	24.21	10464	-1.02	3.08	10448
818	ドッグフード	ドライタイプドッグフード等	203	0.87	21.48	10466	-0.92	2.86	10457
819	キャットフード	ドライタイプキャットフード等	272	2.41	17.56	10466	-1.30	3.38	10461
820	ペットフード (犬・猫用を除く)	金魚用えさ等	5	-1.00	19.29	10458	-0.04	2.62	10455
871	消耗家庭用品ギフトセット	バス用品ギフトセット等	48	20.62	193.25	9420	-0.52	7.10	8688

(備考) 日経メディアアマケテイング、Finatextにより作成。データ期間は1989年4月1日から2018年3月31日まで。

## 補論2 POSデータに基づく物価指標の作成方法

以下では、購入価格（T指数）、売上高の作成方法を概観する。なお、購入価格（T指数）の作成方法は渡辺・渡辺（2013）を踏襲していることから、以下の説明においても参考にす

### ①購入価格（T指数）の365日前比の作成方法

POSデータには、ある店舗における、ある商品の売上高と販売数量が日次で収録されており、価格情報は売上高を販売数量で除することによって求めることができる。また、各商品には、日経の定義する下位分類（6桁分類）と、その上位分類（3桁分類）のIDが振られている。これを前提に、T指数は3段階の集計（下位集計・中位集計・上位集計）により作成する。

まず下位集計では、時刻  $t-365$  から  $t$  までの価格変化率を対数差分  $(\ln(p_{t,s,i}) - \ln(p_{t-365,s,i}))$  として計算したうえで、店舗  $s$  における、品目の上位分類  $c$  に属する全ての商品  $i$  について、トルンクビスト（Törnqvist）ウェイトを用いて加重平均し、店舗別・品目（上位分類）別の価格指数  $\pi_{t,c,s}$  を求める（補式1-1）。ここで、トルンクビストウェイト  $(\omega_{t,s,i})$  は、当該商品  $i$  の店舗別の販売額が当該商品の品目分類  $c$  の販売額に占める割合を、比較対象となる2時点で平均したものである（補式1-2）。 $e$  は販売額を表す。なお、計算において、2時点のうちいずれかの時点で存在しない商品、品目、店舗は計算に用いられない（以下、中位集計、上位集計でも同様）。

$$\pi_{t,c,s} = \sum_{i \in I_{[t-365,t],s,c}} \omega_{t,s,i} \ln \left( \frac{p_{t,s,i}}{p_{t-365,s,i}} \right) \quad \dots \text{(補式1-1)}$$

$$\omega_{t,s,i} = \frac{1}{2} \left( \frac{e_{t-365,s,i}}{\sum_{i \in I_{[t-365,t],s,c}} e_{t-365,s,i}} + \frac{e_{t,s,i}}{\sum_{i \in I_{[t-365,t],s,c}} e_{t,s,i}} \right) \quad \dots \text{(補式1-2)}$$

次に中位集計では、店舗別・品目（上位分類）別の価格指数  $\pi_{t,c,s}$  を、店舗  $s$  についてトルンクビストウェイトで加重平均し、ある時点  $t$  における品目別の価格指数  $\pi_{t,c}$  を求める（補式1-3）。ここでトルンクビストウェイトは、ある店舗  $s$  における品目分類  $c$  に属する全商品の合計販売額が全店舗における品目分類  $c$  に属する全商品の合計販売額に占める割合を、比較対象となる2時点で平均したものである（補式1-4）。 $e$  は販売額を表す。

$$\pi_{t,c} = \sum_{s \in S_{[t-365,t]}} \omega_{t,s,c} \pi_{t,s,c} \quad \dots (補式 1-3)$$

$$\omega_{t,s,c} = \frac{1}{2} \left( \frac{\sum_{\substack{i \in I_{[t-365,t],s,c} \\ s \in S_{[t-365,t]}}} e_{t-365,s,i}}{\sum_{\substack{s \in S_{[t-365,t]} \\ i \in I_{[t-365,t],s,c}}} e_{t-365,s,i}} + \frac{\sum_{\substack{i \in I_{[t-365,t],s,c}}} e_{t,s,i}}{\sum_{\substack{s \in S_{[t-365,t]} \\ i \in I_{[t-365,t],s,c}}} e_{t,s,i}} \right) \quad \dots (補式 1-4)$$

最後に、上位集計では、価格指数 $\pi_{t,c}$ を品目分類 $c$ についてトルンクビストウエイトで加重平均し、総合価格指数 $\pi_t$ とする（補式 1-5）。ここで、トルンクビストウエイトは、全店舗における品目分類 $c$ に属する全商品の販売額が全店舗における全商品の販売総額に占める割合を、比較対象となる2時点で平均したものである（補式 1-6）。 $e$ は販売額を表す。

$$\pi_t = \sum_{c \in C_{[t-365,t]}} \omega_{t,c} \pi_{t,c} \quad \dots (補式 1-5)$$

$$\omega_{t,c} = \frac{1}{2} \left( \frac{\sum_{\substack{s \in S_{[t-365,t]} \\ i \in I_{[t-365,t],s,c}}} e_{t-365,s,i}}{\sum_{\substack{c \in C_{[t-365,t]} \\ s \in S_{[t-365,t]} \\ i \in I_{[t-365,t],s,c}}} e_{t-365,s,i}} + \frac{\sum_{\substack{s \in S_{[t-365,t]} \\ i \in I_{[t-365,t],s,c}}} e_{t,s,i}}{\sum_{\substack{c \in C_{[t-365,t]} \\ s \in S_{[t-365,t]} \\ i \in I_{[t-365,t],s,c}}} e_{t,s,i}} \right) \quad \dots (補式 1-6)$$

## ②売上高の364日前比の作成方法

ここでは、売上高の364日前比の作成方法を説明する。売上高は、品目分類 $c$ に属する全商品の合計販売額を $t$ 時点と364日前時点で合計し、その比率を計算する。前年同日ではなく364日前比であるのは、売上高が曜日による差が大きいため、同じ曜日を比較するように配慮したためである。

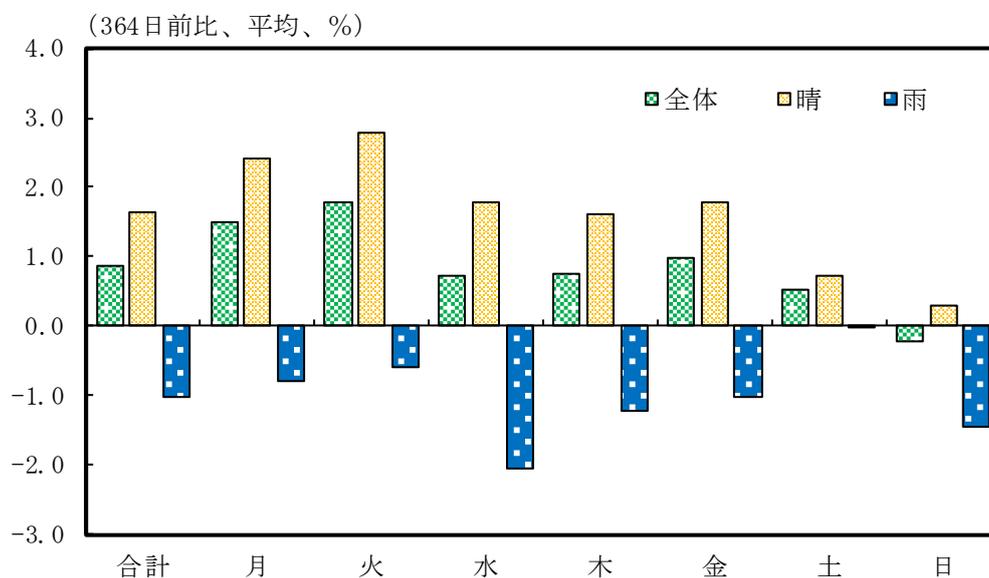
$$\varphi_{t,c} = \left( \frac{\sum_{\substack{s \in S_{[t-364,t]} \\ i \in I_{[t-364,t],s,c}}} e_{t,s,i}}{\sum_{\substack{s \in S_{[t-364,t]} \\ i \in I_{[t-364,t],s,c}}} e_{t-364,s,i}} \right) - 1 \quad \dots (補式 1-7)$$

### 補論 3. 天気と日次売上高の関係

天候が売上高に与える影響について、POSデータを活用して確認する。ここでは天候の中でも天気に着目し、晴れの日と雨の日の売上高の動向を分析する。なお、晴れの日と雨の日の区別は、各都道府県庁所在地において雨が降った日を1、降らなかった日を0としたうえで、2016年における商業動態統計の百貨店・スーパー販売額の都道府県別のシェアをウェイトに加重平均した値が0.5以上であれば雨の日とした。分析データは、POSデータの全体の売上高の364日前比を使用し、対象期間は1989年4月1日～2018年3月31日までとした。

日次データを晴れの日と雨の日に分けて売上高の平均をとると、晴れの日には前年比+1.6%に対し、雨の日には前年比▲1.0%となっており、雨の日には売上が伸びない状況がみてとれる。これは、雨により外出を控える等が背景にあると考えられる。ただし、この傾向も曜日別で度合いは異なる。基本的には全曜日で晴れの日には前年比プラスに、雨の日には前年比マイナスになる傾向ではあるが、土曜日は天気に左右されにくい傾向がうかがえる。同じ休日でも、日曜日は雨の影響を受けやすい等、曜日により天候が与える影響は違うようである。

図 小売店の売上高と天気との関係



(備考) 気象庁、Finatextにより作成。

#### 補論 4. 各主成分の代表品目

ナウキャストの分析で使用したデータセット⑥における主成分の代表品目を示す。  
なお、提示している品目は各主成分への固有ベクトルが絶対値で上位の3品目である。

主成分	小売業計との相関	各主成分における固有ベクトルが上位の3品目（括弧内はベクトルの符号）
主成分1	-0.164	アルコールテイスト飲料類（+）、家庭医療用品（+）、介護・衛生用品（+）
主成分3	0.167	こうじ（+）、家庭医療用品（-）、アルコールテイスト飲料類（+）
主成分6	0.397	雑酒（-）、殺虫剤・殺鼠剤（+）、モップ（+）
主成分7	0.406	家庭医療用品（-）、たばこ・喫煙関連用品（-）、アルコールテイスト飲料類（+）
主成分8	0.169	たばこ・喫煙関連用品（-）、雑酒（+）、電池（-）
主成分11	0.149	食品容器（+）、モップ（+）、電池（-）
主成分12	-0.337	乾燥豆（-）、ココア・チョコレート飲料（-）、生タイプ即席カップめん（-）
主成分14	0.114	リキュール（+）、ドライフルーツ（-）、ココア・チョコレート飲料（-）
主成分16	0.286	食品容器（-）、リキュール（+）、きのこ（+）
主成分18	-0.142	各種詰め合わせ菓子（-）、食品容器（-）、レジャー・食事用品（-）
主成分20	-0.195	豆菓子・いかり豆（+）、介護・衛生用品（-）、ナッツ（+）
主成分22	-0.123	乾燥豆（-）、畜肉瓶詰（+）、各種詰め合わせ菓子（-）
主成分31	-0.183	殺虫剤・殺鼠剤（-）、除湿剤（+）、絵画用品（-）
主成分42	-0.172	口中清涼剤（-）、乾燥豆（+）、入浴剤（-）

## 補論 5. 分析コード

小売業計を予測するナウキャスト分析で実施した、勾配ブースティングの分析コード（統計ソフト R のコード）を示す。

```
library(caret)
XgboostPOS <- train(
  Y~X,
  data = data,
  method = "xgbLinear",
  preProcess = c('center', 'scale'),
  trControl = trainControl(method = "cv"),
  tuneLength = 4)

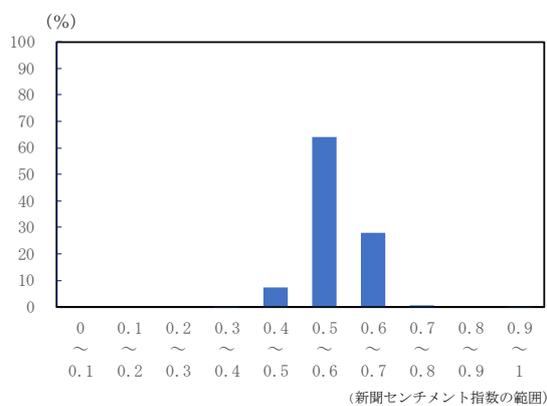
summary(XgboostC)#チューニング結果の確認
> 120 samples
> 18 predictor
>
> Tuning parameter 'eta' was held constant at a value of 0.3
> RMSE was used to select the optimal model using the smallest value.
> The final values used for the model were nrounds = 50, lambda = 0, alpha =
0.003162278 and eta = 0.3.

varImp(XgboostPOS)#変数重要度
```

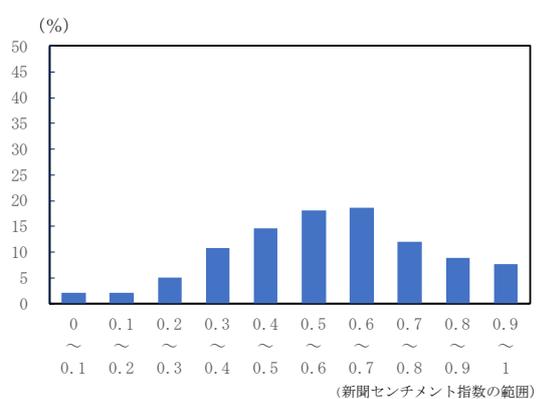
## 補論 6. 新聞センチメント指数の分布

図 新聞センチメント指数の分布 (紙面別)

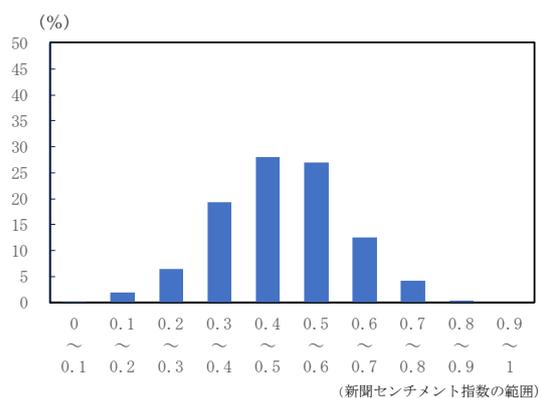
対象記事全体



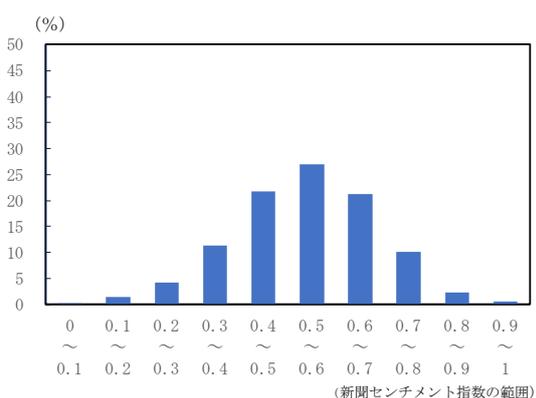
一面



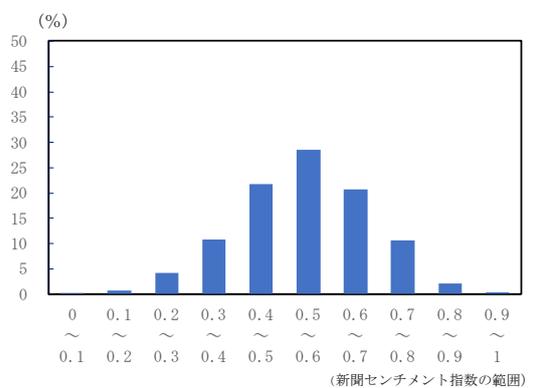
総合面



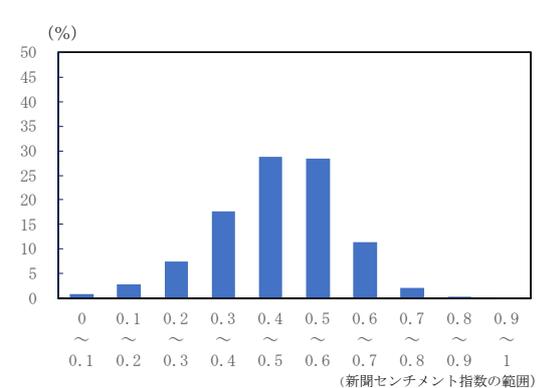
経済面



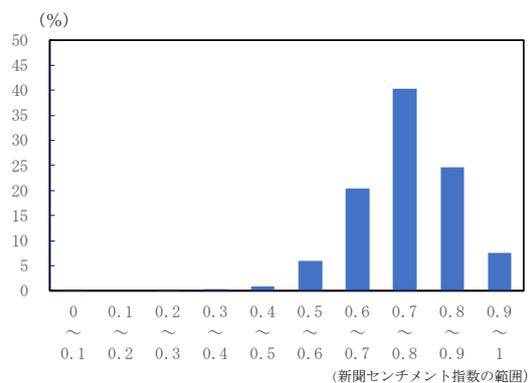
政治面



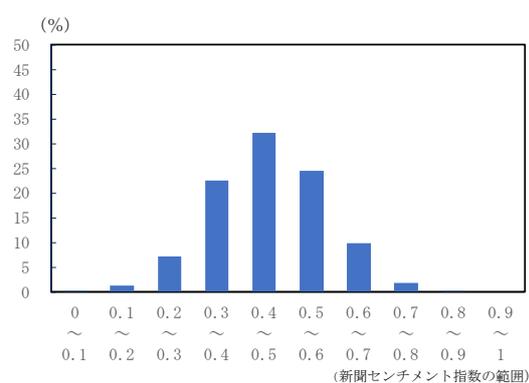
国際面



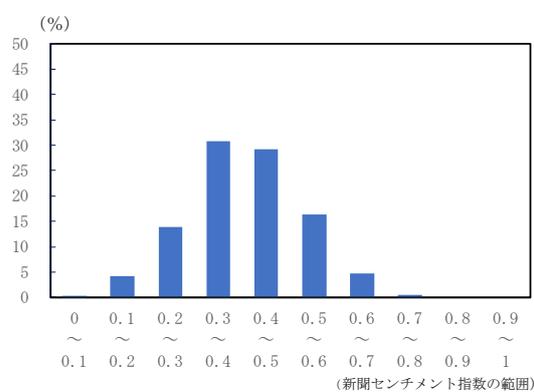
企業面



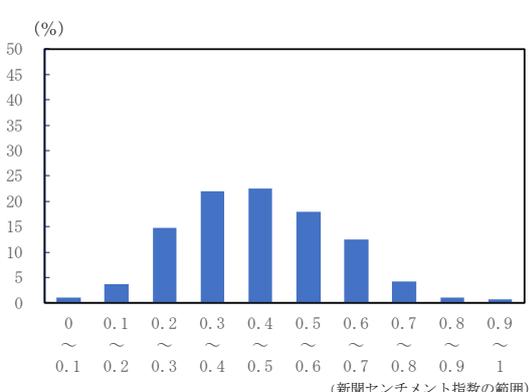
社会面



マーケット総合面



マーケット商品面



- (備考) 1. 内閣府「景気ウォッチャー調査」、日本経済新聞社により作成。  
 2. 紙面の定義はデータの定義部分を参照。  
 3. 景気ウォッチャー調査の先行き判断から構築したモデルに基づきセンチメントを付与した数値。  
 4. 紙面別・日次ベースで単純平均した数値から作成した等の理由により、対象記事全体の数値と各紙面を積み上げた数値は一致しない。

## 補論 7. キーワードに含まれる単語

チョコレートの嗜好変化を読み取るために、商品名にある単語を抽出してキーワードとしてまとめて分析を実施した。ここで、キーワードに含まれる単語の内訳を示す。

分類	キーワード	商品名に記載のある単語
健康志向	カカオ	カカオ
	糖質	糖質、血糖値
	乳酸菌	乳酸菌
味	抹茶	抹茶
	苺	いちご、苺、イチゴ、ストロベリー、あまおう
趣き	大人	大人、オトナ
	贅沢	贅沢、厳選極上、極み、プレミアム、ロイヤル、エクセレント、エクセレンス
	濃厚	濃厚、濃密、とろける、まろやか、深み
	大人数	ファミリー、パーティー、ミックス、バラエティ
時期・季節	限定	限定
	ハロウィン	ハロウィン、ハロウィーン
	冬	冬、雪
	受験	受験、合格

(備考) 株式会社 True Data により作成。