

2 ビッグデータ・AIを活用した消費分析

以下では、小売のPOSデータ及び新聞記事のテキストデータを用いて、AI技術の一つである、コンピュータにデータを解析・予測させる手法（機械学習）を活用した消費動向の分析について、最近の内閣府における取組と暫定的な分析結果を紹介する⁵⁷。

● POSデータは高頻度・早期の利用に強み

スーパーやコンビニ等の小売店では、購入する際に商品をバーコードで読み込み、各商品の販売時点・価格・数量等の情報を採取・記録している。このようなシステムはPOS（point-of-sale）と呼ばれ、売上や在庫管理に役立てられている。ここでは、全国のスーパーマーケット約1,200店舗から収集された日用品・食料品（除く生鮮品）の売上高と価格に関するPOSデータを対象とし、消費動向を試験的に分析する⁵⁸。

POSデータを経済分析に利用することの利点としては、速報性が非常に高いことや、高頻度に観察が可能であることが指摘されているが⁵⁹、マクロ分析に利用するにはカバレッジが限定的である点等には留意する必要がある。今回分析に利用するPOSデータも、217品目分類が日次で利用できること（高頻度）や、その値が早期（数日後）に利用できること（速報性）に強みがある一方、POSデータの大部分を占める食料品等のシェアは消費の2割程度であり⁶⁰、消費全体をカバーできるわけではない。

まず、日次データとしての強みを活かした分析例を紹介する。消費動向を把握する際には、天候がその重要な要因の一つであることが指摘されているが⁶¹、月次で消費と天候の分析を行うのは必ずしも容易ではない。そこで、日次で雨の日と晴れの日に分けてPOSデータの売上高・前年比⁶²を平均すると（第1-3-6図（1））、合計で晴れの日は前年比+1.6%に対し、雨の日は前年比▲1.0%となっている⁶³。曜日別にみても、すべての曜日において雨の日は売上高がマイナスとなる一方⁶⁴、晴れの日は売上高がプラスとなっており、天候により消費者が外出を控えること等のために、売上高（消費額）が左右されていると考えられる。

また、高い速報性は足下の動向を把握するのに非常に有効である。例えば、2017年6月1日にはビールの安売り規制が行われたが、ビールの日次売上高・前年比の推移をみると（第1-

注 (57) 本稿におけるビッグデータ・機械学習を利用した分析は概要のみを記載しているため、データや分析方法等の詳細は小寺他（2018）を参照されたい。

(58) 日経メディアマーケティング株式会社が独自に収集しているPOSデータを利用している。1989年4月以降の日次データ。

(59) 岡崎・敦賀（2015）を参照。

(60) 内閣府「国民経済計算」によると、「食料・非アルコール飲料」及び「アルコール飲料・たばこ」が国内家計最終消費支出に占める割合は2016年で18%程度である。

(61) 例えば、日本銀行（2016）のBOX4等。

(62) 売上高は曜日により大きく金額が異なるため、POSデータの売上高・前年比の計算においては、同じ曜日で比較するために364日前比を用いている。

(63) 雨の日の定義は第1-3-6図の備考2を参照。

(64) グラフからは判別しづらくなっているが、土曜日の雨の日の値は▲0.003%である。

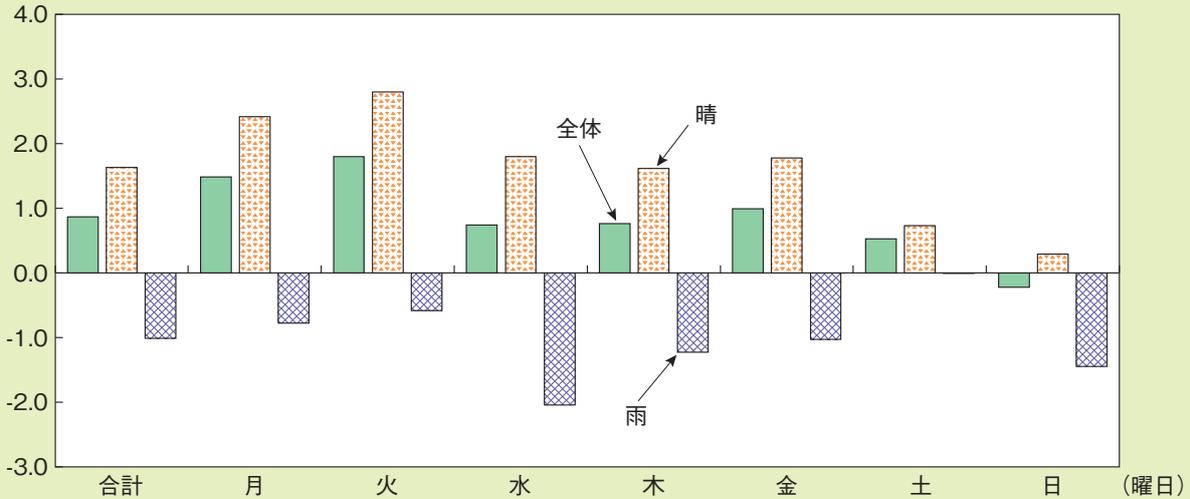
3-6図(2)、2017年5月末にかけて駆け込みがみられ、5月31日には前年比183%となっている。しかし、6月以降の推移をみると、明確な反動減はみられていない。このような動向がほぼリアルタイムで観察可能であるため、制度変更等のイベントの際には、足下の状況を把握するのに非常に有効である。

第1-3-6図 日次データによる分析例

日次のPOSデータにより、詳細な消費活動を把握することが可能に

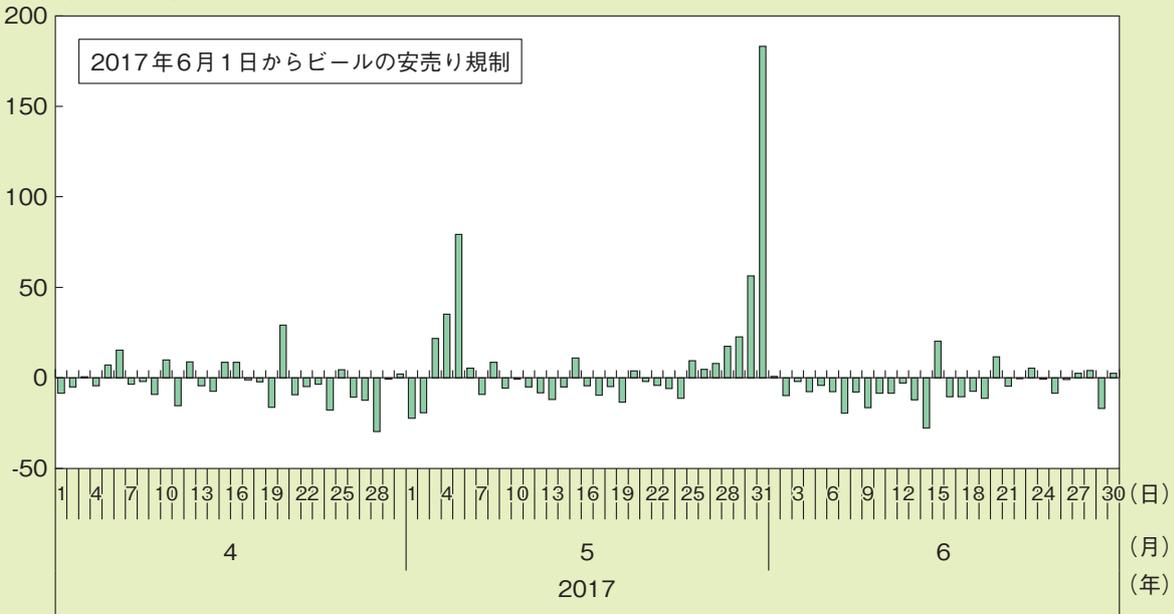
(1) 天気と日次売上高の関係

(364日前比、平均、%)



(2) ビールの日次売上高

(364日前比、%)



(備考) 1. Finatext 提供データにより作成。
 2. (1) は1989年4月1日～2018年3月31日までの平均値(年末年始の各々4日間は除く)。晴れと雨の日の区別は、気象庁のデータをもとに各都道府県の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年における経済産業省「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。

● POSデータを用いて足下のマクロ経済指標を予測する

次に、POSデータの強みを活かした分析例を紹介したい。通常、マクロ経済指標は、その月が終了してから1~2か月後に公表されるものが多いが、迅速な景気動向把握のためには、ある月が終わったら、すぐにその月のマクロ経済指標の値が予測できること（以下「ナウキャスト」という。）が望ましい。この点において、POSデータは速報性という強みをもっているが、カバレッジの低さもあり、単純にPOSデータの値を月次で比較しただけでは、ナウキャストを正確に行うことは困難である。

ただし、マクロ経済指標の動きと似通った動きを、POSデータの動きの中から抽出することができれば、カバレッジ等の問題を解消し、ナウキャストを行うことが可能となる。そこで、売上高のPOSデータを機械に学習させることで、特徴的な動きを抽出させ、経済産業省「商業動態統計」における「小売業計」の前年比をナウキャストできるかを試みた。また、小売業計には家電や衣類等の売上動向が気温等の天候により左右されやすい商品が含まれることや、天候データも速報性の高い情報であることを踏まえ、POSと天候データ⁶⁵を併せて機械に学習させる分析も行った。機械学習の方法は、予測のパフォーマンスが高い学習方法の一つである「勾配ブースティング (gradient boosting)」と呼ばれる手法を用いている⁶⁶。

2007年1月~2016年12月までの10年間の月次のデータを機械に学習させ、2017年1月~2018年3月の小売業計の前年比をナウキャストさせ評価したところ、POSデータのみでは相関係数0.63であったが、天候データを加えると相関係数0.78に上昇した。天候データを使うことで、気温等に敏感な商品の売上高動向等を補正することができたと考えられ、ここでも天候が消費動向に影響を与えていることが確認できる。

このPOSと天候データの両方を使ってナウキャストを行った結果を示したのが、第1-3-7図である。前年比の水準には差がみられるものの、前年比前月差をみると、方向感は概ね似たような動きをしていることから、大まかではあるが、小売業計の前年比について一定程度のナウキャストを行うことができたと考えられる。ただし、今回の分析でも水準の差に加え、2018年以降等で動きが異なる部分もみられることから、精度向上に向けては更なる分析を進めていく必要がある⁶⁷。

注

(65) 具体的には、平均気温、最高気温、最低気温、降雪量のデータを利用した。

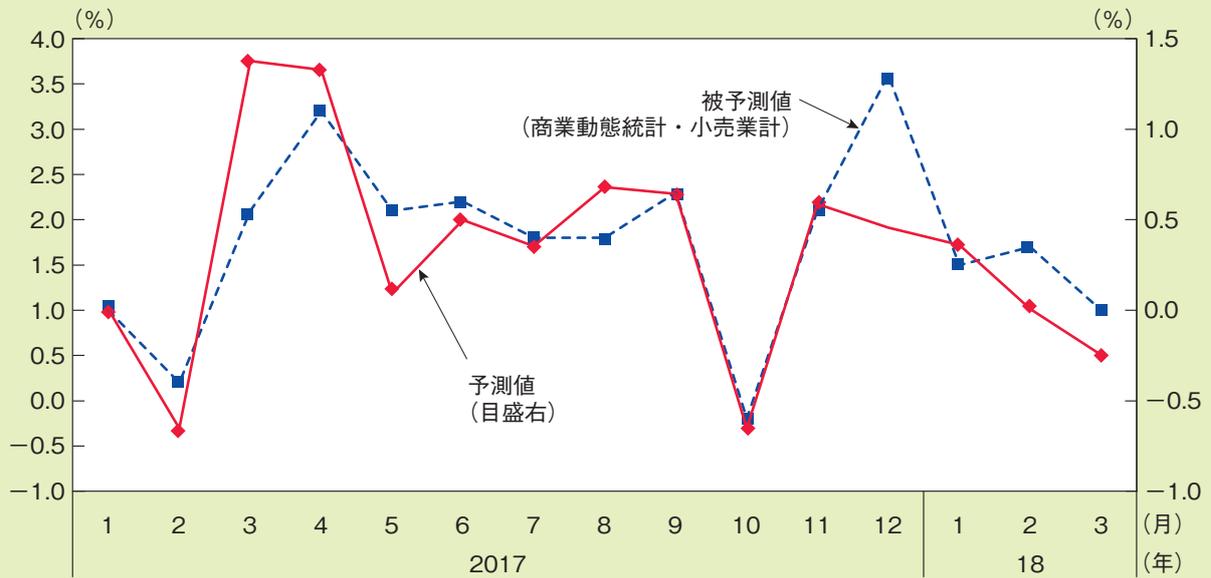
(66) 勾配ブースティングとは、目的変数の誤差に対して弱い予想モデル（ここでは線形モデル）の訓練を繰り返して、推定したモデルの結果を組み合わせることで最終的な予測モデルの推計精度を高める手法である。なお、分析においては、様々な機械学習の手法（ランダム・フォレスト等）を試したが、勾配ブースティングの精度が最も高かった。

(67) 例えば、今回分析に利用したPOSデータは217品目分類の日次の前年比を単純平均により月次化したデータであるため、品目間や曜日間のウェイトが考慮できていない。

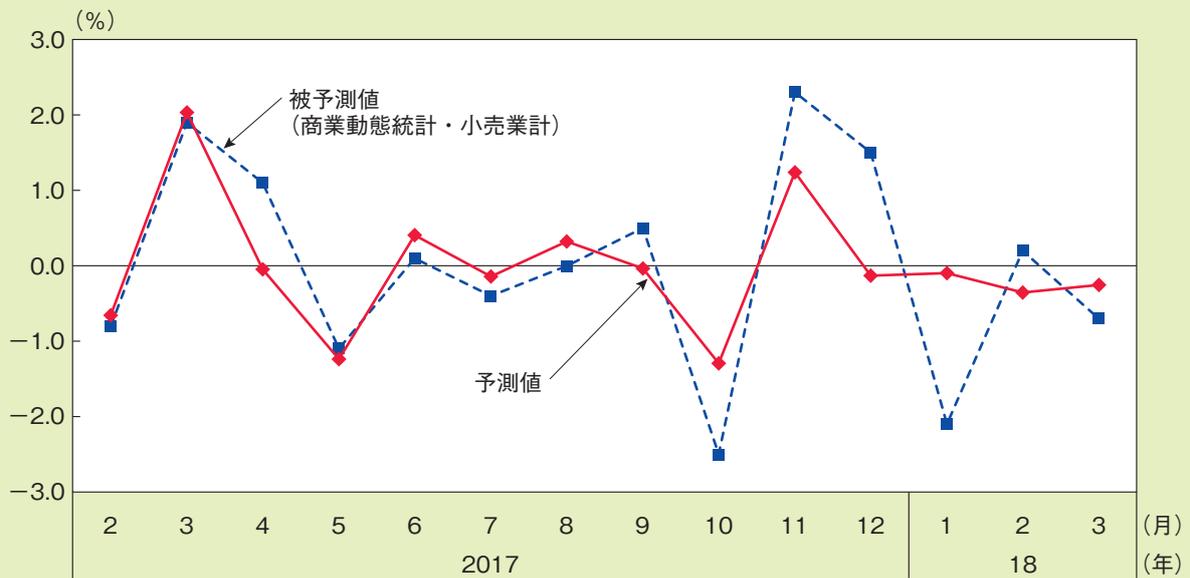
第1-3-7図 POSデータによるナウキャスト

POSと天候データから大まかなナウキャストが可能に
商業動態統計（小売業計）の予測結果

(1) 前年比



(2) 前年比前月差 (%)



(備考) 経済産業省「商業動態統計」、気象庁、Finatext提供データにより作成。
2007年1月～2016年12月までを学習し、2017年1月以降を予測・評価した。
詳細は小寺・藤田・井上・新田(2018)を参照。

● POSデータから需要・供給曲線のシフトが明らかに

また、POSデータの利点の一つとして、217品目分類別に数量と価格の双方の情報がセットで利用可能であることが指摘できる。この利点を生かして、POSデータの価格変化が、需要と供給のどちらのショックにより起こったものかについて簡易的に試算することが可能となる⁶⁸。経済学における一般的な右下がりの需要曲線と右上がりの供給曲線を想定すると、価格上昇は、需要曲線の右方シフト、または供給曲線の左方シフトによって生じる⁶⁹。その際、需要曲線のシフトであれば数量は増加するが、供給曲線のシフトであれば数量は減少する。つまり、POSデータの価格が上昇した際、同時に数量も増えていれば需要ショック（需要曲線のシフト）、数量が減少していれば供給ショック（供給曲線のシフト）とみなすことが可能となる。

POSデータにおける217品目分類の価格・数量の動向⁷⁰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つのカテゴリーに分割し、それぞれの割合をみたのが第1-3-8図(1)である。例えば、2008～09年にかけての世界金融危機の際には、価格上昇かつ数量増加の割合が低下し、価格低下かつ数量減少の割合が増加しているなど、時期により特徴的な動きがみられる。

上記の4つのカテゴリーのうち価格上昇かつ数量増加、価格低下かつ数量減少は需要曲線のシフト（需要要因）、価格上昇かつ数量減少、価格低下かつ数量増加は供給曲線のシフト（供給要因）に対応しているとみなすことができる。そこで、各品目分類それぞれの価格変化率を、需要要因と供給要因のグループに分けて統合することで、POSデータ全体の価格変化に対して、需要要因と供給要因の寄与をそれぞれ算出することができる。

分析結果を確認すると（第1-3-8図(2)）、POSデータの価格変化には需要要因と供給要因の両方が影響していることが確認できる⁷¹。例えば、2008～09年にかけての世界金融危機をみると、需要要因の押し下げ寄与が高まっているが、原材料価格の低下等を反映して供給要因も同時に物価を押し下げていることがわかる。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し上げに寄与していたものの、2016年以降は押し下げに寄与した。需要要因については、2013～2014年初にかけて押し上げに寄与した一方、2016年にはマイナスに寄与した。

このように価格変化と同時に数量変化も確認することで、価格変化の背景が、消費者の購買意欲の変化によるものか、原材料価格の変化によるものか等、POSデータから示唆を得ることが可能となる。ただし、ここで確認ができたのはPOSデータが対象とする日用品・食料品の動きに限られる点には留意が必要である。

注 (68) 分析に当たっては渡辺(2016)を参考にした。また、価格データは、購入価格の前年比(T指数)。指数の作成方法は、渡辺・渡辺(2013)を参照。

(69) ここでは簡素化のため、シフトするのは需要曲線か、供給曲線のどちらか一方のみと仮定する。

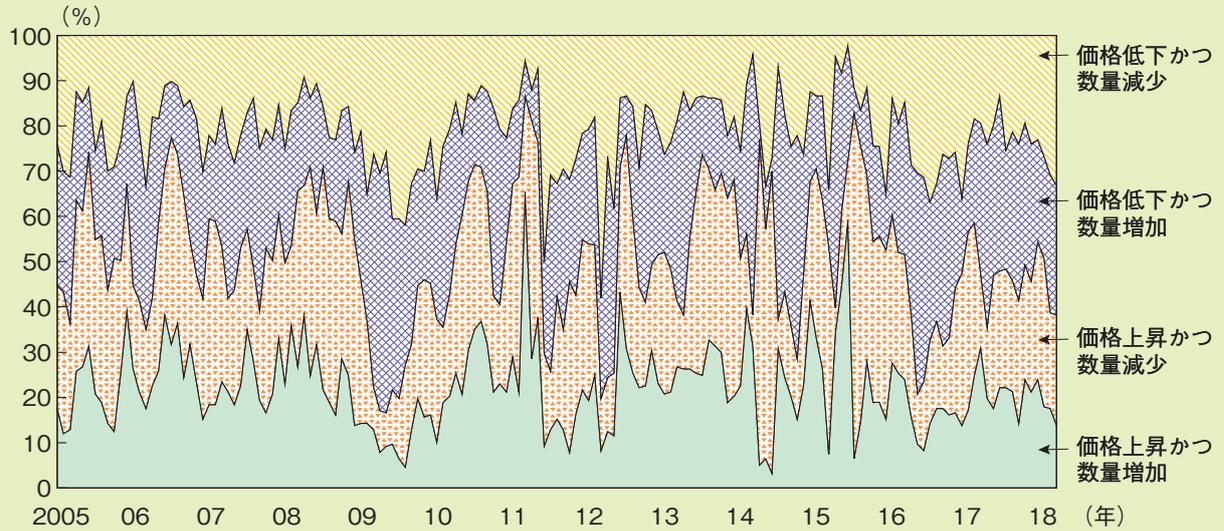
(70) 価格・数量の変化は、3か月移動平均の前年比前月差により判断している。

(71) 2005年1月～2018年3月の期間を通してみると供給要因の寄与が平均6割程度と相対的に大きくなっている。

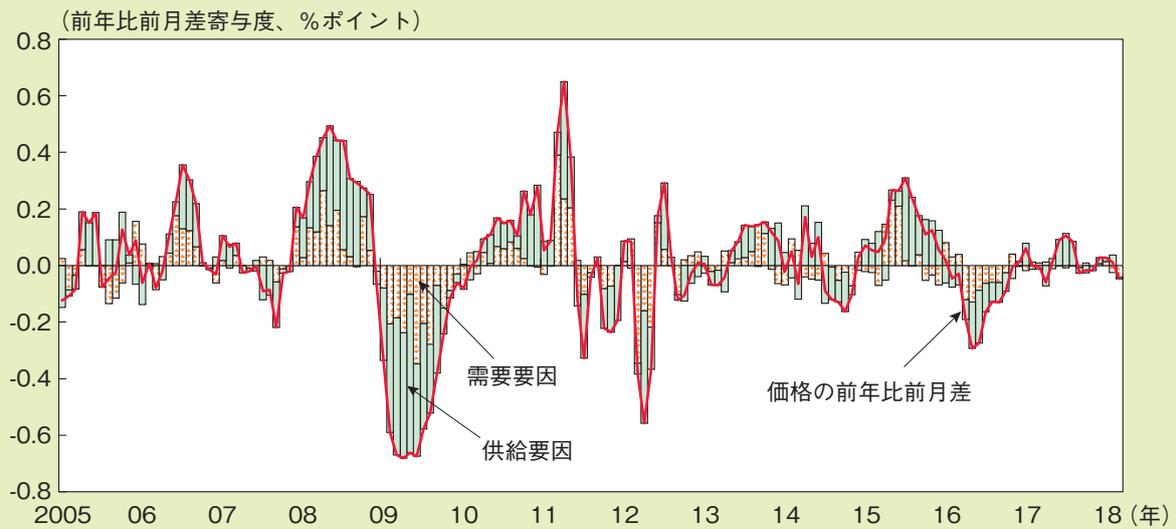
第1-3-8図 POSデータによる価格変化の要因分解

価格と数量の変化を組み合わせることで価格変化の要因を把握することが可能に

(1) POSデータにもとづく価格・数量変化の4分類



(2) 価格の前年比前月差の要因分解 (需要要因と供給要因の識別)



- (備考) 1. Finatext提供データにより作成。
2. 価格と数量の後方3か月移動平均値をもとに作成。
3. 要因分解等の詳細は小寺・藤田・井上・新田(2018)を参照のこと。

● 新聞記事のテキストデータは消費者マインドと連動

近年では、デジタル化されたテキストデータの利用が容易になったことから、テキストを用いた分析も活発化している⁷²。定性的な情報を数値データに変換するため、分析手法によって結果が変わり得る等との課題があるものの、POS同様に速報性の高さ等から、既存の分析を補完する分析が可能となることが期待されている⁷³。

ここでは試験的に、消費者マインドに影響を及ぼすことが指摘されているニュース記事のテキストデータを利用して⁷⁴、消費者マインドとの相関を紙面別に確認するとともに、ニュース記事と消費行動の分析も行うことで、テキスト情報が上記のPOSデータ等の他の分析を補完できる可能性があるのかについて考察する。

具体的には、景気の現状と先行きに対する評価とコメントがセットで利用できる内閣府「景気ウォッチャー調査」を利用し、ディープラーニング（深層学習）の手法として知られるRNN（Recurrent Neural Network）を用いて⁷⁵、機械にどのようなコメントが景気認識に対して良い・悪いコメントであるのかを学習させた⁷⁶。なお、学習は現状と先行きで別々に行い、学習の結果、景気ウォッチャー調査の新規コメントに対して93～95%の確率で正しく分類できるようになった⁷⁷。学習が完了した機械に、2013～17年における18万件以上の新聞記事⁷⁸を読み込ませ、それぞれの記事がポジティブ（ネガティブ）である確率を算出させる。これにより新聞記事のテキストデータを数値化することが可能となり（以下、作成した指標を「新聞センチメント指数」という）、消費者マインドとの相関をみることができる。

経済面、社会面、政治面などの紙面別に新聞センチメント指数を作成し、消費者態度指数との相関をみると（第1-3-9図）、景気ウォッチャー調査の先行きで学習させた「経済面」、「マーケット商品面⁷⁹」の新聞センチメント指数との相関が0.6と高くなっている。同じ「経済面」でも景気ウォッチャー調査の現状で学習させると、相関が半分程度になることから、消費者マインドは見通しも踏まえて形成されていることが示唆される。また、対象記事全体では相関が低いことから、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、マーケット商品面に多く見られる価格変化等、特定の情報に影響を受ける傾向があると考えられる。

注

(72) 例えば、塩野（2018）、五島他（2017）等。

(73) 岡崎・敦賀（2015）。また、Shapiro et al.（2018）は、ニュース記事から作成した指数をモデルに加えることで予測力が向上したと報告している。

(74) Doms and Morin（2004）等。

(75) 具体的には、LSTM（Long Short Term Memory）層を導入したRNNを用いている。詳細は岡谷（2015）等を参照。

(76) 景気ウォッチャー調査における「変わらない」を除き、「良い、良くなっている」をポジティブ、「悪い、やや悪くなっている」をネガティブとして学習させている。

(77) この正答率は山本・松尾（2016）による先行研究と同程度である。

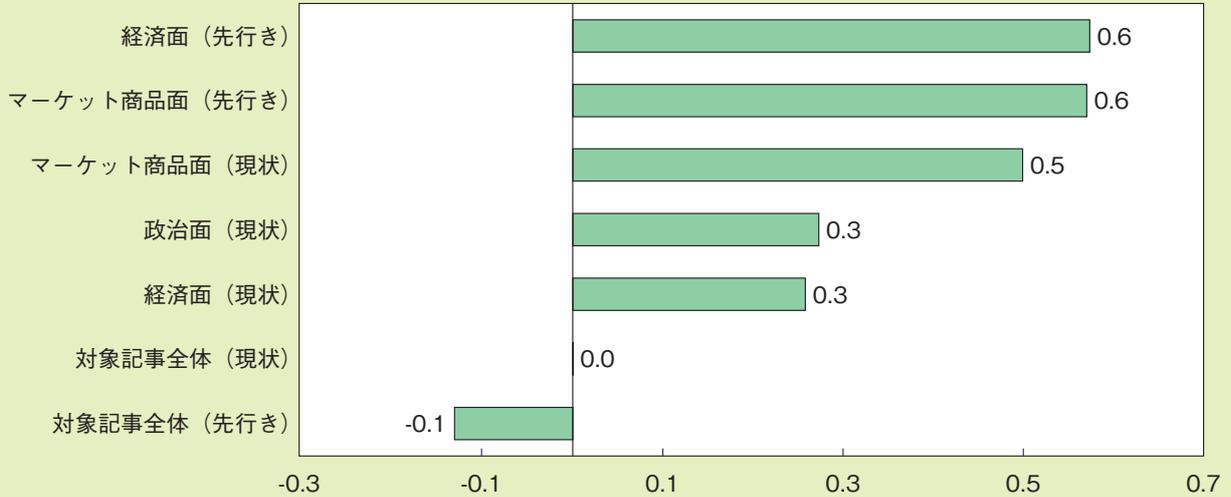
(78) 対象とした新聞記事は日本経済新聞（朝刊）である。ただし、消費者マインドに関係が薄いと考えられる分類の記事は除いている。

(79) 例えば、肉の卸売値やチャーター料の変化等、価格変化の記事が多い。

第1-3-9図 新聞センチメント指数（紙面別）と消費者マインドの関係

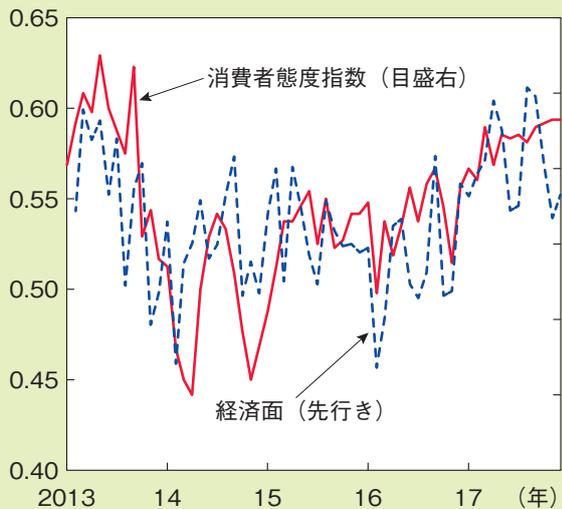
経済面やマーケット商品面から作成した指数と消費者マインドが相関

(1) 新聞センチメント指数（紙面別）と消費者態度指数との相関係数（2013～17年）



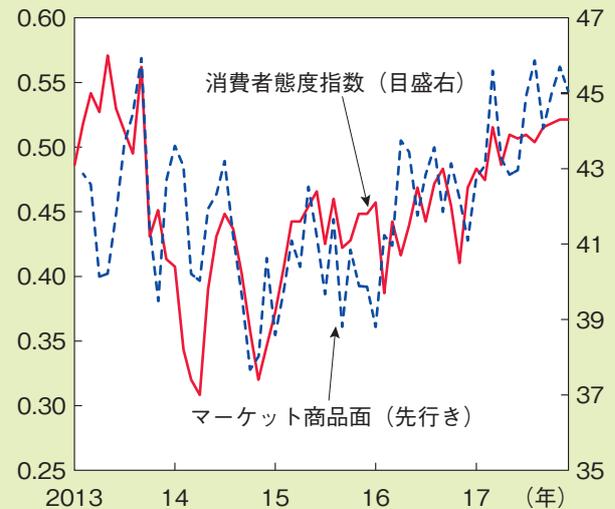
(2) 経済面（先行き）と消費者態度指数の推移

（記事が「良い」に分類される確率）



(3) マーケット商品面（先行き）と消費者態度指数の推移

（記事が「良い」に分類される確率）



- （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. 分析対象とした紙面は①一面、②総合面、③政治面、④経済面、⑤国際面、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。(1)は、消費者態度指数との相関係数が高い上位5項目と対象記事全体を掲載している。
 3. 消費者態度指数は二人以上の世帯、原数値。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後方30日移動平均かつ毎月15日時点とした。
 4. 詳細は、小寺・藤田・井上・新田（2018）を参照。

●センチメントの改善より悪化が消費に大きな影響の可能性

POSデータ同様、上記で作成した新聞センチメント指数の強みとして高頻度であることが挙げられる。この強みを活かし、1週間程度の短期間内における新聞センチメントの変化と消費行動（POSの売上高の変化）の関係性について分析を行う。その際、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対する影響は異なるとの研究結果もあることから⁸⁰、新聞センチメント指数が前週と比較してプラス方向、マイナス方向に変化した際の影響を分けて（非対称性を考慮して）分析する。

まず、217品目分類のPOSデータ（売上高）のうち、消費者マインドの変化に敏感な品目を抽出すべく、新聞センチメント指数が利用可能な2013～17年において、消費者態度指数と相関が高い上位10品目を選定した（第1-3-10図（1））。これら10品目のそれぞれの売上高変化と、非対称性を考慮した新聞センチメント指数の変化を週次ベースで回帰分析を行い⁸¹、新聞センチメント指数が改善した場合と悪化した場合とで、消費行動に与える影響が異なるかを検証する。新聞センチメント指数は、消費者態度指数との相関が高かった景気ウォッチャーの先行きコメントで学習させた経済面とマーケット商品面の2つを使用する。

10品目の推計結果のうち、新聞センチメント指数が1標準偏差変化した時のPOS売上高の変化が有意であった品目をプロットした（第1-3-10図（2）（3））。経済面では10品目のうち5品目、マーケット商品面では10品目のうち6品目で両者の間で有意な関係がみられた。プラスの方向とマイナスの方向を比較すると、両方が有意になったすべてのケースにおいて、マイナスの方向の係数が高くなっている。また、プラスの方向のみが有意になったケースは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高い可能性が示唆されている。

上記の分析を踏まえると、新聞等のテキストデータについても経済分析に対して有用な情報をもっている可能性がある。今後、POSデータに加え、こうしたテキスト情報も活用することで、ナウキャストの精度を向上させ、より迅速かつ的確な景気把握が可能となると考えられることから、引き続き、研究を進めていく必要がある。

注

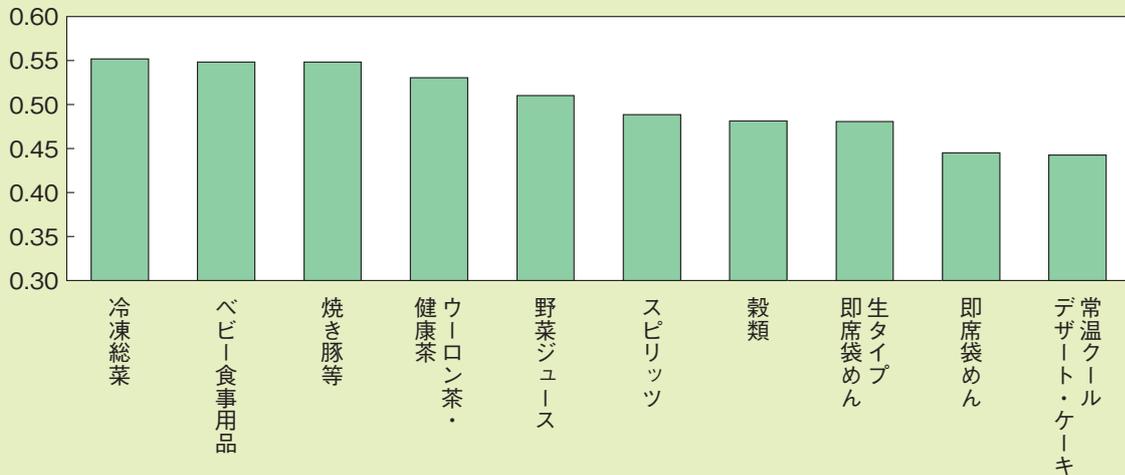
(80) 例えば、Soroka（2006）等。

(81) POSデータの売上高・前年比を被説明変数とし、説明変数に新聞センチメント指数、消費税ダミー等を使用して、1期までのラグ項も加味した回帰分析を行なっている。

第1-3-10図 新聞センチメント指数がPOSデータに与える影響

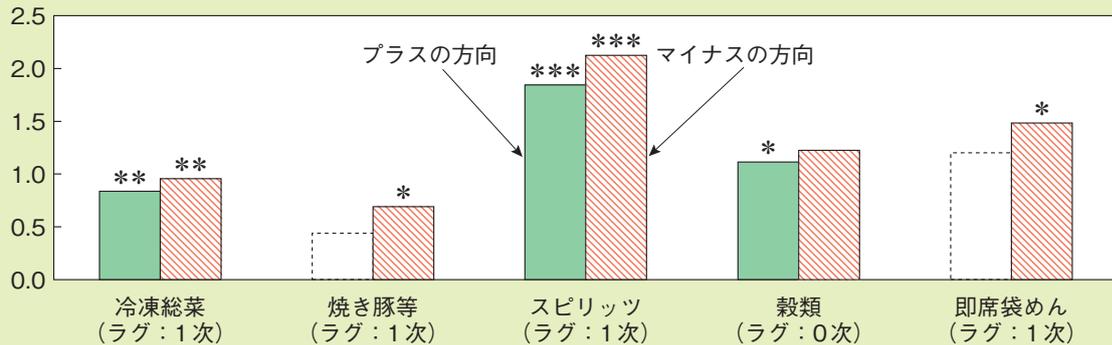
経済面やマーケット商品面から作成した指数と消費者マインドが相関

(1) 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)



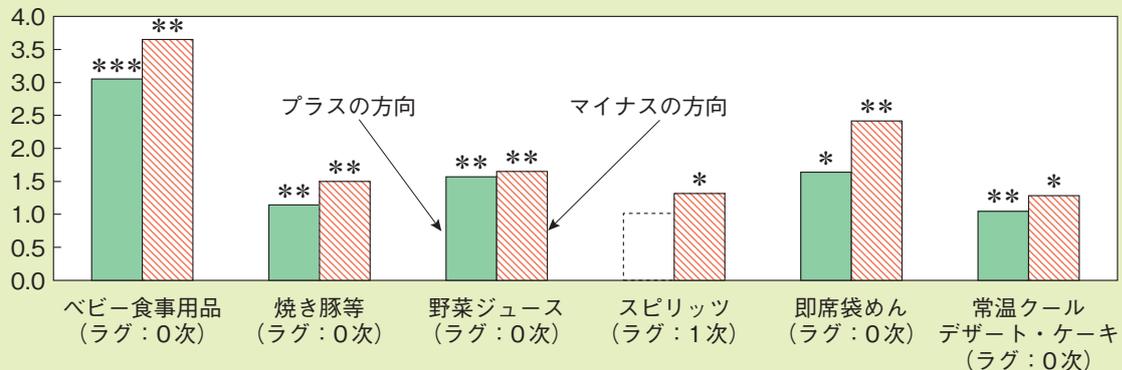
(2) 経済面(先行き)(指数が1標準偏差変化した場合の前年比への影響、週次)

(%ポイント)



(3) マーケット商品面(先行き)(指数が1標準偏差変化した場合の前年比への影響、週次)

(%ポイント)



(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. プラスの方向は前週比で新聞センチメント指数が上昇した場合、マイナスの方向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。
 3. 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推計式に基づく。推計方法の詳細は小寺・藤田・井上・新田(2018)を参照。
 4. ***は1%水準、**は5%水準、*は10%水準で有意であることを示す。