

2024 年 12 月 26 日 全 13 頁

大和物価センチメント指数開発の試み

CPI との間に高い相関および先行性を有する指数の開発に成功

経済調査部 研究員 島本 高志
経済調査部 主任研究員 新田 堯之

[要約]

- 本研究では、消費者物価指数（前年同月比、以下 CPI）に対して先行性を有する「大和物価センチメント指数」の開発を試みた¹。具体的には、内閣府「景気ウォッチャー調査」における「景気判断理由集」の各テキストを「物価上昇」、「物価下落」など 4 種類に分類した結果を基に指数を構築した。
- 各テキストをより高い精度で分類するために、本研究では生成 AI の一種である大規模言語モデル（Large Language Model, 以下 LLM）を活用した。ここでは、LLM への命令であるプロンプトの最適化を通じて、高い精度でのテキスト分類に成功した。さらに、指数の算出式を工夫して、CPI とより高い相関を得ようと試みた。
- 2001 年から約 24 年分の月次データで、大和物価センチメント指数と CPI との相関係数を確認した結果、消費税調整済のコア CPI（除く生鮮食品）との間で最大 0.888（2 カ月先行）、同コア CPI との間で最大 0.817（5 カ月先行）、という先行研究を上回る高い相関が確認された。さらに、指数の振れを抑えるために後方 3 カ月移動平均を施したところ、先行性はやや薄れたものの、同コア CPI との間の相関係数は最大で 0.903（1 カ月先行）までに達した。
- 各種データとの補完的な活用にはきわめて有用である一方、注意点としては、元データが本来は景気に関する調査によるものであることが挙げられる。そのため、先行研究と同様に、定期的な価格改定など景気と無関係な価格動向に弱い可能性は、依然として残る。
- 次回以降の分析では、LLM が物価の上昇や下落を判断する際に用いた単語や、判断の根拠となったテキスト本文を解析することで、物価変動に影響を与えた特徴語を抽出することなどを検討している。

¹ 本研究においては、当社データドリブンサイエンス部の栗山太吾・寺坂壮一郎・参木裕之の各氏から、有益な助言やコメントを頂戴した。なお、本研究内の所見は全て著者らの個人的見解であり、当社の物価見通しを示すものではない。

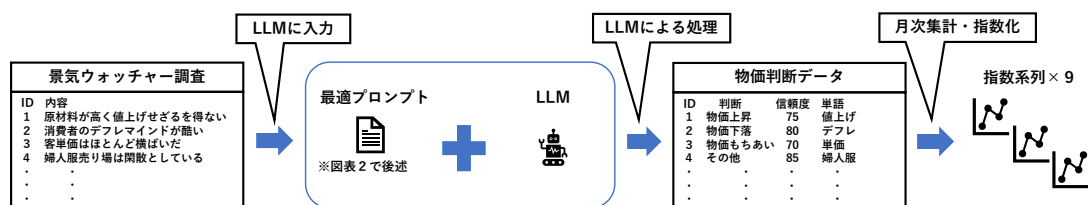
はじめに

日本銀行（以下、日銀）は「物価の安定を図ることを通じて国民経済の健全な発展に資すること」²を使命の1つとしている。そのため、物価の動向について可能な限り早く把握することは、日銀自身や、その一挙手一投足に注目する金融資本市場、景気動向を注視する政府当局、資金調達を考えている事業会社などにとって、重要だ。

2010年代には、いち早く物価の動向を掴むため、小売店の商品販売データ（スキャナーデータ）を利用した物価指数等が開発されてきた³。その一方、同時期には機械学習技術が発展する中、人間が書いたり話したりする言葉をコンピューターで処理する「自然言語処理」と呼ばれる工学分野が台頭した。経済学においても、この技術を利用し、様々な経済指標を再構築しようという試みが進んだ⁴。その一環として、日銀内部の研究である大高・菅（2018）と中島ほか（2021）という論文では「物価センチメント指数（以下、PSI）」という指数が開発されている⁵。

PSIは、マクロ経済学の理論で重要な、予想インフレ率を示す指標の1つとされる^{6 7}。日銀が開発したPSIに関しては、コアコアCPI（除く生鮮食品、エネルギー）の前年同月比に、最大で10カ月ほど先行する性質が知られている。ただし、当時の技術的な問題により、その精度には限界があった。そこで、本研究では図表1のように、大規模言語モデル（Large Language Model, 以下 LLM）を応用し、特に LLM への命令であるプロンプトの最適化と指数処理への工夫を通じて、精度を大幅に向上させた「大和物価センチメント指数」の開発を試みた。

図表1 指標作成の流れの概要



（注）景気ウォッチャー調査および物価判断データの内容はイメージで、実際の内容ではない。

（出所）大和総研作成

² 日本銀行「[日本銀行の概要](#)」（2024/12/11 閲覧）より。

³ 詳しくは、渡辺（2022）などを参照されたい。

⁴ 詳しくは、島本高志「[金融経済分析を変える自然言語処理の力① 従来の自然言語処理が与えたインパクト](#)」（大和総研レポート，2024年09月19日）や、島本高志「[金融経済分析を変える自然言語処理の力② 大規模言語モデルの登場と今後の展望](#)」（大和総研レポート，2024年11月1日）を参照されたい。

⁵ 中島ほか（2021）p.1では「月次のインフレ予想指標のひとつとして、需給ギャップや従来のインフレ予想指標、各種のヒアリング情報などと補完的に活用していくことが考えられる」とされ、中村豊明（2022）「[わが国の経済・物価情勢と金融政策——山梨県金融経済懇談会における挨拶——](#)」にも用いられた。

⁶ 中島ほか（2021）より。なお、現代的なマクロ経済学理論では、人々のインフレ予想が実際のインフレ率に影響するメカニズムが理論化されてきた。詳しくは、渡辺（2022）、渡辺（2024）などを参照されたい。

⁷ 人々が実際にどのようなインフレ予想を持っているかは把握が難しい面があり、日銀は長期にわたりインフレ予想の推計に尽力している。詳しくは、西野ほか（2016）や、長田・中澤（2024）、日本銀行（2024）「[経済・物価情勢の展望（2024年4月）](#)」などを確認されたい。

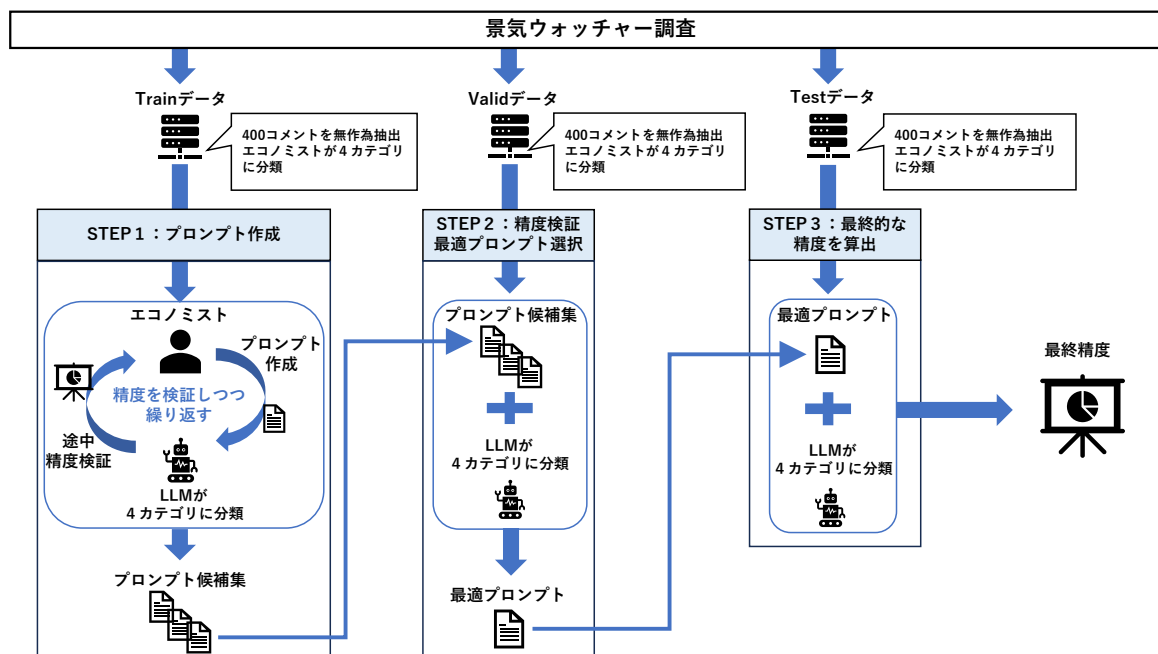
大和物価センチメント指数の開発

今回の研究では、先行研究と同じく、内閣府「景気ウォッチャー調査」における「景気判断理由集」の各テキストを「物価上昇」、「物価下落」など4種類に分類し、この結果に基づき指数を構築した⁸。分類の部分に関しては OpenAI 社の GPT モデル⁹を使用し、先行研究より精度を高めつつ、独自に PSI の計算を行った。その際、本稿2ページの図表1のような流れで開発を行った。まずは、この最適プロンプトの作成について解説する。

最適プロンプトの開発と分類精度検証

最適プロンプトの開発と分類結果の検証の流れは、図表2の通りである。

図表2 最適プロンプトの開発および分類精度検証の流れ



(出所) 大和総研作成

本稿では、プロンプトを開発する主体は人間だが、一般的な機械学習モデルの開発手順を参考

⁸ この調査は、内閣府が景気に関する市中の体感を数量的に把握するために、景気に敏感な職種の人々に対して毎月行っているアンケート調査で、直接的に物価に関して行った調査ではない。だが、大高・菅（2018）は、このデータに物価に関する様々なコメントが入っていることに着目し、簡素な分類を通じて物価に関する指標の構築に成功した。

⁹ Azure OpenAI Service REST API 経由で GPT 4o-mini の 2024-07-18 版を使用した。その理由は、約 66 万件ものデータに対して、高速に計算するためには高性能な GPU が必要となるローカル LLM や、あるいは非常に高額な予算が必要になる他の LLM のモデルを使用することは、実務的な観点から難しいと考えられたためである。

に、プロンプトの性能を客観的に評価し、過剰適合¹⁰を防ぎながら最適なプロンプトを選択した。

まず、研究時点で最新だった、景気判断理由集の2024年9月分までの659,837個のデータを内閣府のウェブサイトから入手し、分析に必要な形へ整形する。このデータセットから、①プロンプトの初期的な改善や試行錯誤に用いる訓練 (Train) データ、②過剰適合を防ぎながらプロンプトの表現や構成を調整し、最適なプロンプトを選択するための検証 (Validation) データ、そして最後に③最終的に選択されたプロンプトの精度評価を行うテスト (Test) データ、の3種類を抽出する。景気判断理由集には「現状」と「先行き」の2種類があるため、それぞれランダムに200個を2種類ずつ3回、合計1,200個のデータを抽出する¹¹。

次に、大高・菅 (2018) の図表に掲載された24個の分類例を参考に、1,200個のテキストデータの分類を手で行う。具体的には、先行研究と同様、「物価上昇」、「物価下落」、「物価もちあい等¹²」、「物価への言及無し¹³」の4種類へと分類する^{14 15}。

それから、図表2のSTEP1のように、最適なプロンプトの構築を試みる。詳細としては、LLMに対して、プロンプトと訓練データを与えて分類を試み、最低限、正しい回答形式で分類を行えるプロンプト候補を作成した。また、分類の自信度合を聞くことを意図し、LLMに分類の信頼度を0~100の数字で回答させたほか、判断において最も重要視した単語を回答させた¹⁶。

次に、過剰適合を防ぐため、図表2のSTEP2のように作成したプロンプト候補集と検証データを合わせてLLMに与えて分類させることで、精度を再検証し、最適なプロンプトを選択する。

最後に、未知データへの予測精度を調べるべく、図表2のSTEP3のように、作成した最適プロンプトとテストデータを合わせてLLMに与え、分類を行わせる。この際、LLMの回答には多少の振れが存在することから¹⁷、再現性担保のため同一LLMに同じデータを3回与え、3回分の多数決を取ることで回答を選択する¹⁸。結果、最適プロンプトでは、Testデータの正答率が90.25%、精度指標の1種である加重F1スコアという指標でも0.901を記録した^{19 20}。

¹⁰ モデルが訓練データに適合しすぎるあまり、未知データへの予測精度がかえって下がってしまう現象のこと。

¹¹ 先行研究と比べても、サンプルサイズでは土井ほか (2024) の約3倍を処理しており、担当者1人当たりのデータ処理数でも大高・菅 (2018) と比べて遜色ない。実務上、妥当な水準だったと考えられる。

¹² これは、大高・菅 (2018) で用いられた「物価に言及はあるが方向感なし (物価横ばい)」の別称である。

¹³ 脚注12の文献の一部では、「物価への言及無し」の表記が「その他」となっている。

¹⁴ この物価動向に関する評価判断が人間にも難しく、分類データの正しさが制約されることは、大高・菅 (2018) の時点で言及されている。よって、最適プロンプトが同一ならば最終的な相関係数の再現性は確保されるが、精度検証では、仮に類似の手法をとった場合でも、必ずしも同一の結果になるとは限らない。

¹⁵ この際、増税に関するコメントに関しては、物価の基調と関係ないため、物価上昇とみなさなかった。

¹⁶ 最終的に信頼度で重みづけた指数も計算したが、結果はほぼ同じだった。

¹⁷ 少なくとも現時点のGPT 4o-miniでは、出力結果を固定するためSeedパラメータを固定し、Temperatureパラメータをゼロとしても、モデルの計算特性上、同じプロンプト文を入力した場合に回答が常に一致するとは限らない。

¹⁸ 3回とも違う回答が出た場合は分類不能と分類する予定であったが、結局、分類不能は出なかった。

¹⁹ なお、精度を測るその他の指標である加重適合率は90.51%、加重再現率は90.25%であった。

²⁰ これらの指標算出元の混同行列は補論図表2として末尾に掲載されているため、必要に応じ参照されたい。

分類と指数計算における工夫

次のステップでは、最適プロンプトを用いて、研究中の最新データであった景気判断理由集の2024年9月分までの、659,837個のテキストデータを分類する。

具体的には、最適プロンプトと全データを LLM に与えて、再現性担保のため3回の分類を行い回答間の多数決を取ることで、最終的な指数計算用の物価判断データを作成した²¹。

次に、この物価判断データを、元となった景気ウォッチャー調査の景気判断理由集における「現状」と「先行き」の2種類各々のコメント数に関して、月次で集計した。

また、先行研究において「現状」と「先行き」に関して両者から作成した指数の傾向は大まかに同じであったと指摘されていることや、データが多いほど指数系列が安定する可能性が考えられたことから、両者のコメント数を合わせた「現状と先行きの総合（以下、現先総合型）」についても月次で集計し、合計3種類の集計を作成する。

指数に関しては、日銀の元の計算式（以下、日銀式）だけでなく、大和1式、大和2式と称する2種類の新たな計算式を考案した。

【日銀式】

$$\frac{\text{物価上昇コメント数} - \text{物価下落コメント数}}{\text{物価上昇コメント数} + \text{物価下落コメント数} + \text{物価もちあい等コメント数}}$$

【大和1式】

$$\frac{\text{物価上昇コメント数} - \text{物価下落コメント数}}{\text{物価上昇コメント数} + \text{物価下落コメント数}}$$

【大和2式】

$$\frac{\text{物価上昇コメント数} - \text{物価下落コメント数}}{\text{物価上昇コメント数} + \text{物価下落コメント数} + \text{物価もちあい等コメント数} + \text{物価への言及なしコメント数}}$$

新計算式考案の背景には、先行研究で「物価もちあい等」コメントの分類が不十分だったという問題意識がある。そこで、以下の仮説に基づき新計算式を考案した。

大和1式は、先行研究の分類が不完全でも成功していることから、「物価もちあい等」コメントは分母に含めなくても指数の計算には問題がない、との仮説に基づく。

²¹ ただし、433個のデータに関してはエラーにより分類ができなかった。これは、今回用いた Azure OpenAI のコンテンツフィルターが誤作動したためだと思われる。このエラー数はデータ全体の僅か 0.067% であることから、指数の精度に影響を与えたとは特に考えられない。

大和2式は「物価への言及なし」コメントも、物価変動を感じていない意思表示と捉え、分母に含むべきとの仮説に基づく。景気ウォッチャー調査は物価に関する調査ではないため、物価変動を強く感じない限りコメントされない可能性がある²²。そのため、これらを除くと情報が欠落する恐れがある。また、これらの指数計算式では、分母の物価関連コメント数が指数の変動を抑えるための安定化の役割を果たす。だが、分母が小さい時期の指数は不安定になりがちである。つまり、全回答数から「物価への言及なし」コメントを除くと、指数が不安定化する可能性も考えられる。

大和物価センチメント指数と消費者物価指数との関係

最高で約0.9の相関（2カ月先行）を確認

こうして構築された大和物価センチメント指数とCPIなどとの関係を確認する。ここでは参考までに、日銀式の指数算出法で計算した「日銀式」と称する指数の結果も、併せて掲載する。

図表3 日銀式・大和1式・大和2式でのPSI（原系列）の時差相関一覧表

	種類	CPI	コアCPI	コアコアCPI
日銀式【参考】	現先総合	0.831（2カ月先行）	0.838（3カ月先行）	0.774（7カ月先行）
	現状	0.773（2カ月先行）	0.780（2カ月先行）	0.746（7カ月先行）
	先行き	0.813（3カ月先行）	0.819（4カ月先行）	0.728（7カ月先行）
大和1式	現先総合	0.788（2カ月先行）	0.794（2カ月先行）	0.730（7カ月先行）
	現状	0.728（1カ月先行）	0.735（2カ月先行）	0.703（7カ月先行）
	先行き	0.770（5カ月先行）	0.776（4カ月先行）	0.681（8カ月先行）
大和2式	現先総合	0.878（2カ月先行）	0.888（2カ月先行）	0.817（5カ月先行）
	現状	0.829（1カ月先行）	0.837（2カ月先行）	0.794（5カ月先行）
	先行き	0.847（2カ月先行）	0.856（2カ月先行）	0.768（6カ月先行）

（注1）スペースの都合により、時差相関を測った上で、最も相関が高かった時期の数値のみを記載している。また、相関がどの時期に一番高く出たかを、括弧内に併記している。

（注2）黒字太線は、各指標（CPI・コアCPI・コアコアCPI）に対して最も相関が高かった相関係数が載っている。なお、ここで相関係数を測るのに用いられた各指標の前年同月比は、消費税調整済の系列である。

（出所）総務省統計より大和総研作成

図表3では、日銀式・大和1式・大和2式の指数について、2001年1月から2024年9月までの月次ベースで²³、CPIなどと時差相関を測った。その上で、最も高い相関係数と、対応する時差の期間を示した。図表3からは、大和物価センチメント指数（大和2式、現先総合型）とコアCPIの前年同月比の間の相関が一番良好であり、その相関係数は0.888と0.9に近い²⁴。また、

²² つまり、回答の際に「景気の調査で物価に関して言及したいほど、そもそも物価変動を感じているか否か」という閾値が存在し、それを超えない限り回答が為されない一種の閾値モデルとなる。

²³ 指数の計算は2000年から可能であるが、中島ほか（2021）の説明と同じく、2000年はコメント数が少なすぎるため省いた。

²⁴ 脚注15で言及したように、今回の研究では消費税増税に関しては「物価上昇」とはみなさない形で行っており、消費税調整済の指数系列で相関係数が高いのは想定通りである。大高・菅（2018）で相関が測られたCPI

コア CPI の前年同月比との間にも、約半年のラグを経て 0.82 ほどの相関が確認できる。

この精度は、金融市場で CPI の先行指標として広く用いられ、全国版 CPI よりも約 20~30 日早い当月末頃に出る東京都区部 CPI のコア CPI と、全国版コア CPI の間の、同じ月での相関係数である 0.974 には劣る。だが、大和物価センチメント指数は、翌月 10 日ごろに発表される景気ウォッチャー調査を基に算出され、2 カ月先行する。つまり、全国版 CPI より約 70 日早い。その点も含めると、東京都区部 CPI よりも 40~50 日ほど早く動向が判明することになり、先行指標として有用である可能性がある。

図表 4 で確認すると、大和物価センチメント指数（大和 2 式，現先総合型）が、物価指数の基調をとらえる先行指標として期待できることが、改めて確認される^{25 26}。

一方、日銀式の指数（現先総合型）も、3 カ月先行でコア CPI の前年同月比に対して 0.838 の相関係数を記録している。LLM で最適プロンプトを開発して分類を行ったことと、大和 2 式による指数計算法の改善の、双方が相関の改善に寄与したことも見て取れる^{27 28}。

物価の基調を正確にとらえる上でも有益

さらに、最も精度に優れていた大和物価センチメント指数（大和 2 式，現先総合型）と、CPI およびコア CPI の前年同月比の動きを**図表 4** で比較²⁹すると、基本的にはほぼ一致している。ただし、直近（2020 年から 2021 年頃）においては、大和物価センチメント指数と CPI およびコア CPI の前年同月比との間に方向性の乖離が発生してしまっているように見える。だが、この期間に関してはむしろ、大和物価センチメント指数が、コア CPI 以上に物価の基調を正確にとらえていた可能性すら考えられる。

その理由として、当時の CPI 変化率の下落に関しては、CPI の計算法の問題から携帯電話通信料引き下げ政策が影響したことが、岸川（2021）や、肥後（2022）で指摘されている。具体的には、岸川（2021）では、携帯電話通信料などの特殊要因を除いた場合のコア CPI について、「前年比+0.1%と、わずかながらもプラス」と推計されていた。そして当時の試算では、2020 年 8 月から 2021 年 3 月のコア CPI の前年同月比が、ほぼゼロ近傍になっている様子が見える³⁰。

関連指標も、消費税調整済の系列である。本来、物価の基調を見るためには、一時的な政策要因は取り除かれるべきだが、消費税分を調整しないコア CPI を使用した場合も、含意は大きく変わらなかった。

²⁵ 一方で、企業物価指数との相関は冴えず、日銀式（先行き型）が、消費税を除く円ベースでの企業物価指数に対して、最大で 0.747 の相関係数を記録するのみに留まった。企業物価指数に先行する指数を作成する場合は現状、相関係数が 0.793 と報告されている Suzuki, Sakaji (2024) の手法が最適であると考えられる。

²⁶ なお、先行研究と同様に大和 2 式（現先総合型）の PSI と 12 カ月先の CPI やコア CPI の前年同月比に対してグレンジャー因果性検定を行ったが、先行研究に近い結果で特に問題はなかった。詳しくは補論図表 1 を参照されたい。

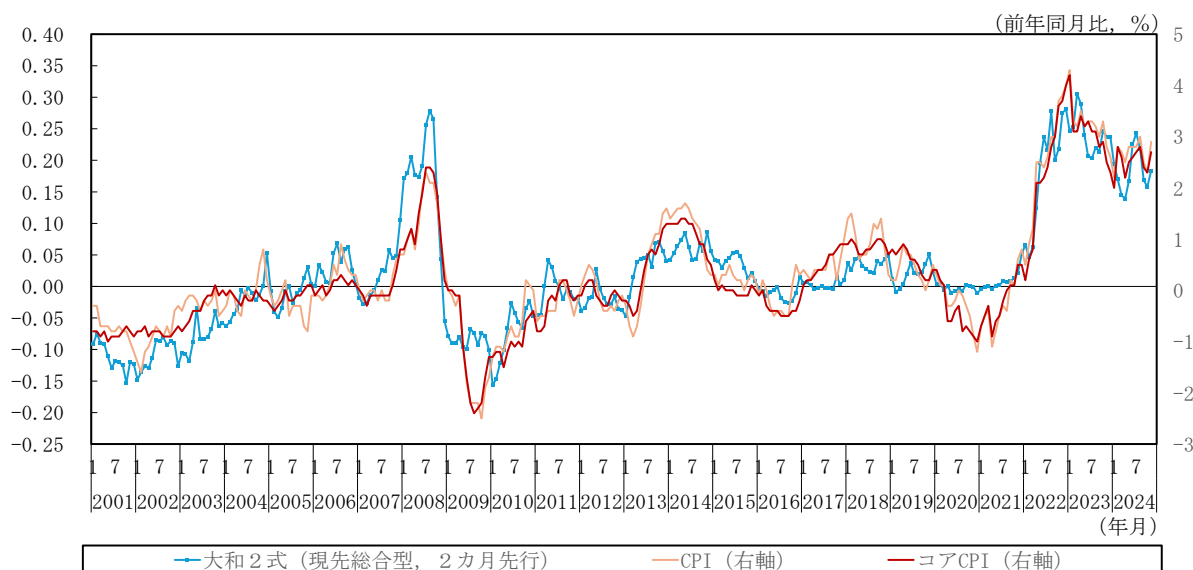
²⁷ 一方で、やはり分母が小さく指数が不安定になるためか、あるいは物価への言及が無いデータを±ゼロと見なす考えが正しかったからか、大和 1 式の指数は、全般的に振るわなかった。

²⁸ 参考として、日銀式で算出した指数の推移を補論図表 3 に掲載したほか、大和物価センチメント指数とコア CPI の動きに関しても、補論図表 4 に掲載した。必要に応じてご覧いただきたい。

²⁹ ここでは、大高・菅（2018）などと同様に、消費税調整済の指数を使用している。

³⁰ 岸川和馬「[携帯電話通信料の引き下げと CPI への影響](#)」（大和総研レポート，2021 年 05 月 11 日）。ただし、岸川（2021）は、同じ CPI の前年同月比でも 2015 年基準での試算であることに留意が必要である。

図表4 大和2式（現先総合型）のPSI（原系列）とCPIおよびコアCPIの推移



（注1）CPIおよびコアCPIについては、消費税調整済の系列の前年同月比を使用している。

（注2）大和2式（現先総合型）については、コアCPIおよびCPIに先行する月数分を調整して掲載している。つまり、2カ月先行のため、2024年11月のところには、2024年9月の指数が載っていることになる。

（出所）総務省統計より大和総研作成

その点において、大和物価センチメント指数がゼロ近傍を推移していることは、むしろ物価の基調を判断する指数としての有用性を示す可能性もある。

先行き判断を急がない場合は後方3カ月移動平均が有効か

また、物価の基調を見るためには、後方3カ月移動平均を施した方が有効な可能性がある。

図表5 日銀式・大和1式・大和2式でのPSI（後方3カ月移動平均）の相関係数一覧

	種類	CPI	コアCPI	コアコアCPI
日銀式【参考】	現先総合	0.847（2カ月先行）	0.857（2カ月先行）	0.792（6カ月先行）
	現状	0.786（1カ月先行）	0.795（1カ月先行）	0.762（6カ月先行）
	先行き	0.835（3カ月先行）	0.841（2カ月先行）	0.747（6カ月先行）
大和1式	現先総合	0.803（1カ月先行）	0.814（2カ月先行）	0.748（6カ月先行）
	現状	0.742（1カ月先行）	0.751（1カ月先行）	0.720（6カ月先行）
	先行き	0.792（4カ月先行）	0.798（1カ月先行）	0.700（7カ月先行）
大和2式	現先総合	0.893（1カ月先行）	0.903（1カ月先行）	0.835（4カ月先行）
	現状	0.841（0カ月先行）	0.852（1カ月先行）	0.811（4カ月先行）
	先行き	0.863（1カ月先行）	0.871（1カ月先行）	0.786（5カ月先行）

（注1）スペースの都合により、時差相関を測った上で、最も相関が高かった時期の数値のみを記載している。また、相関がどの時期に一番高く出たかを、括弧内に併記している。

（注2）黒字太線は、各指標（CPI・コアCPI・コアコアCPI）に対して最も相関が高かった相関係数が載っている。なお、ここで相関係数を測るのに用いられた各指標の前年同月比は、消費税調整済の系列である。

（出所）総務省統計より大和総研作成

たとえば、後方3カ月移動平均を施した大和物価センチメント指数（大和2式、現先総合型）の精度を確認すると、コアCPIの前年同月比に対して、1カ月先行で0.903の相関係数を記録した。また、コアコアCPIでも、後方3カ月移動平均で調整する前に対し、相関係数が4カ月先行で0.835と高まっている。したがって、特に先行き判断を急がない場合には、先行性が少し薄れるかもしれないが、指数の振れを抑制する3カ月移動平均を用いた方が有効な可能性がある。

おわりに

以上、大和物価センチメント指数の開発手法の説明および、消費者物価指数との関係について、解説してきた。

結果としては、大和物価センチメント指数（大和2式、現先総合型）とコアCPIの前年同月比の間の相関係数は、約0.9に達し、2カ月の先行性があることが判明した。また、2020年8月から2021年3月に、当時の携帯電話通信料の引き下げ政策がコアCPIに影響してしまっていたのに対して、大和物価センチメント指数（大和2式、現先総合型）はその影響を取り除いた物価の基調を正確に捉えている可能性が示唆された。さらに、後方3カ月移動平均を施した方が、物価の基調を見る上ではより有効な可能性がある点も分かった。このように、大和物価センチメント指数が、物価の短期的な先行きを見るために、優れた指標であると考えられる。

一方、大高・菅（2018）でも言及されるように、指数の基となった景気ウォッチャー調査は、本来は景気に関する調査によるものであるため、定期的な価格改定など景気と無関係な価格動向に弱い可能性は依然、課題として残っているだろう。各種データとの補完的な活用にはきわめて有用だが、一方で、この指数を絶対的な未来を映す魔法の鏡として過信するのは危険だ。また、この指標はCPIの予測制度を最大化するために組まれたモデル等ではなく、あくまで先行指標として開発されたことにも留意が必要だ。

今回の研究は、あくまで新しい試みの一環である。そのため、詳細について、より深く分析していくことで、その性質などを明らかにしていく必要があると考えられる。よって、物価センチメント指数が持つ性質について、今後も追加的な分析を実施することを検討している。

たとえば、LLMが上昇下落などの判断を行う際に併せて回答した単語などは、今回は利用されていない。だが、それらのLLMが抽出した単語や、あるいは「物価上昇」や「物価下落」の判断がなされたテキスト本文を解析することで、物価の上下に影響した特徴語の抽出や、企業の短期インフレ予想に影響するキーワードの探索等ができる可能性もある。

また、今回の総合型指数は、あくまで景気判断理由集の、現状と先行きに関してコメント数を統合するにとどめたものとなっているが、現状型と先行き型の指数について、CPIの前年同月比にフィットする最適な配分の比率を探索する等の、発展的な処理も考えられる。

以上

補論 1. 物価センチメント指数 (PSI) の沿革

従来型の PSI の算出法

大高・菅 (2018) では、以下の手法で PSI が計算されている。

まず、当時の景気判断理由集の全データである約 40 万個のテキストから、景気判断理由集には「現状」と「先行き」の 2 種類のコメントがあるため、1,500 個×2 種類のサンプルを無作為に抽出する。次に、同論文の著者らがサンプルを人手で、「a. 物価上昇」、「b. 物価下落」、「c. 物価もちあい等」、「d. 物価への言及無し」の、4 種類へ分類する。そして、これを学習データとして、ナイーブベイズ分類器という手法を用いて、機械学習モデルを作成する。

この際の分類精度は、正答率で、現状のサンプルでは 74.1%、先行きのサンプルでは 76.1% だったと報告されている。その後は、モデルに全データを分類させてから、月ごとに DI³¹として、本論で言及した「日銀式」の計算法で、指数を計算する。

結果としては、現状のデータから作成した指数とコアコア CPI の前年同月比³²との間で、7 カ月先行して 0.76 の相関係数が報告されている³³。加えて、現状のデータから作成した指数へ後方 3 カ月移動平均の処理を行い、コア CPI (除く生鮮食品) の前年同月比との相関係数を測った結果では、2 カ月ほど先行して 0.8 程度の相関係数があることが、論文内の図から見て取れる。

他方、日銀が補強的に出した論文である中島ほか (2021) では、PSI に関して更なる精緻化が図られ、「ナロー版」と称する派生指数が開発された³⁴。

具体的には、分類スコアの作成に使われる単語を厳選し、物価の基調からの乖離を抑えようとした。そのため、単語数は解析結果等を踏まえ、大高・菅 (2018) の手法で作られた指数 (以降、中島ほか (2021) に従い「ブロード版」) の約 3,000 個から、物価の基調との関連性が深いと考えられる 20 個へと減らされている。

結果としては、コアコア CPI の前年同月比との間の相関係数について、ブロード版では、8 カ月先行で 0.75 の相関係数が報告されている。一方、ナロー版については、10 カ月先行で 0.80 の相関係数が報告された。また、コア CPI の前年同月比との間の相関係数も、ブロード版との相関係数が 3 カ月先行で 0.61 だったことに対し³⁵、ナロー版との相関係数は 3 カ月先行で 0.71 と、より高かったことが報告されている。他にも、日本銀行の全国企業短期経済観測調査 (以下、日銀短観) で計測されている、各種 DI 等との相関が報告されている。その相関係数は、最大のものと、販売価格判断 DI の先行きとの間で 0.927 と報告されている。

³¹ プラスの回答数から、マイナスの回答数を引いた景況感指数のこと。日銀短観の指数などが有名。英語の Diffusion Index (直訳すると「拡散指数」) の略で、景気の拡大縮小を示すため、この名がついている。

³² 大高・菅 (2018) の図表 12~14 の注釈に従う場合、消費税調整済の系列の前年同月比が使用されている。

³³ ただし、具体的な数値は大高・菅 (2018) には見当たらず、中島ほか (2021) p. 5 で言及されている。

³⁴ 数値の報告は、景気判断理由集の「現状」から作成した指数に関するのみ行われている。だが、「先行き」から作成した指数の傾向もほぼ同じだったことについても、言及されている。

³⁵ 大高・菅 (2017) の図から見て取れる 0.75 前後の相関係数から大きく下がっているが、直後で後述のように、コロナ禍の期間の新語に対応できなかったことなどが影響したとみられる。

従来型の PSI の限界と LLM を用いた PSI の推計

このように、日銀の PSI には性質および精度から強い期待が寄せられたが、分類精度の面では問題を抱えていた。さらに、大高・菅（2018）を確認すると、添付されている表³⁶からは、特に「物価もちあい等」の分類について正答が皆無であることが分かる。つまり、先の数式のうち「物価もちあい等」のコメント数については、ほぼ機能していなかったことになる。

また、中島ほか（2021）では、コロナ禍の期間に使用例のない新語が増加したことに伴い、過去の景気ウォッチャー調査から学習した分類器が、物価の先行き動向を十分に捉えられていない可能性が指摘されている。実際に、同論文では、先述のようにコロナ禍の期間のデータも含む形で相関係数を測ったが、ブロード版の精度が落ちていることが確認された。同論文では精度の改善に成功した面もあったが、しかし、同時に「ブロード版を作成するために使用した多くの単語を見落としている可能性もある」と、研究の限界を指摘している。

そして、このことは、ナロー版において単語を絞ったことが精度に影響したように、指数の目的である CPI との相関に悪影響を及ぼしている可能性がある。そのため、この点が改善できれば CPI との相関係数が高まる可能性は高く、より優れた指数を作成できると考えられた。

一方、近年は LLM を初めとする自然言語処理技術の向上が著しく、文書分類に関しても、精度が著しく向上している³⁷。さらに LLM には、文脈を考慮できる利点や新語への強靱性があると考えられる。そのため、LLM を用いれば、指数そのものの精度向上が可能であると考えられる³⁸。

補論 2. その他の補論図表

補論図表 1 グレンジャー因果性検定の結果

帰無仮説：AはBに対してグレンジャーの意味での因果関係はない		F値	P値
A	B		
大和物価センチメント指数（大和 2 式，現先総合型）	→ CPI	5.875	0.000
CPI	→ 大和物価センチメント指数（大和 2 式，現先総合型）	1.020	0.429
大和物価センチメント指数（大和 2 式，現先総合型）	→ コア CPI	7.118	0.000
コア CPI	→ 大和物価センチメント指数（大和 2 式，現先総合型）	1.033	0.417

（注 1）グレンジャーの意味での因果関係はないとする帰無仮説は P 値が 0.05 以下ならば 5 % 有意水準で棄却。

（注 2）図の作成方法や、12 期のラグの選択などについては、大高・菅（2018）の手法になった。

（出所）大和総研作成

³⁶ 機械学習の用語を用いれば、いわば「混同行列」を示す表である。

³⁷ 詳しくは島本高志「[金融経済分析を変える自然言語処理の力② 大規模言語モデルの登場と今後の展望](#)」（大和総研レポート，2024 年 11 月 1 日）の、土井ほか（2024）や、Lee et al.（2024）や Guo et al.（2024）への言及部分を参照されたい。

³⁸ ナイーブベイズ分類器では単語同士の関係による文脈は考慮されない一方、LLM はモデルが広範なジャンルの文章から、文脈を含む形で作成されており、新語への耐久性などが高いと考えられる。

³⁹ 実際に、直近に同じアイデアに基づく研究である Suzuki, Sakaji（2024）という論文も発表された。本研究と同じく LLM を用いて PSI を推計する研究だが、①使用したプロンプト、②「物価上昇」、「物価下落」、「物価もちあい等」、「物価への言及無し」の 4 種類への分類を、まずは物価動向に言及があるか分類してから再分類するなど 2 段階で行っている点、③複数 LLM の使用、④景気判断理由集の「現状」のコメントのみの使用、⑤より細分化した PSI の開発、といった点などで、本研究とは大きく異なる。Suzuki, Sakaji（2024）での最終的な相関係数の向上は一定程度に留まり、CPI 関連指標との相関係数は、最大でも約 0.724 に留まっている。また、新たに相関が測られた企業物価指数との間で約 0.793 の相関係数が報告されている。

補論図表 2 混同行列

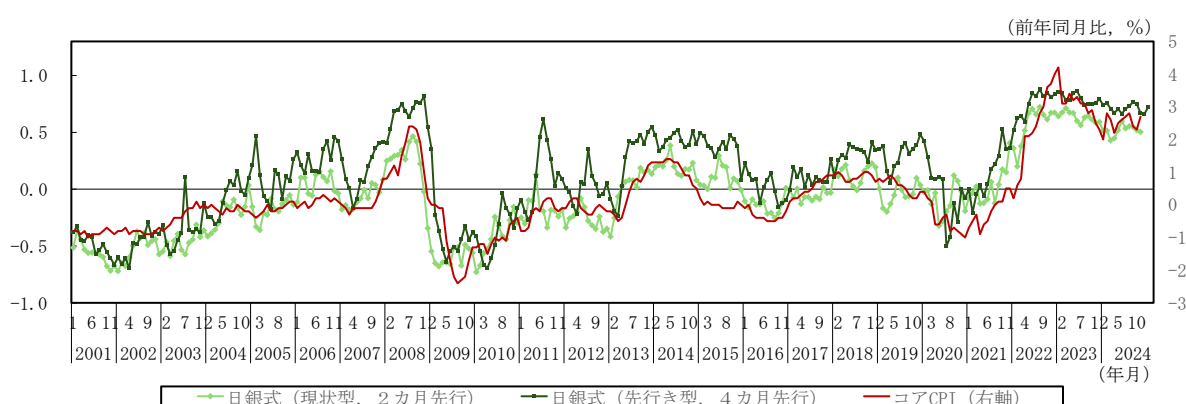
混同行列	物価上昇	物価下落	物価もちあい等	その他
物価上昇	17	1	3	3
物価下落	0	21	4	11
物価もちあい等	1	1	3	6
その他	4	1	4	320

(注1)「物価もちあい等」とは、物価への言及はあるが、状態が横ばいであるなど、方向感を欠いている場合のことを指している。

(注2) 指標についてはそれぞれ、正答率は、90.25%。加重適合率は90.51%。加重再現率は90.25%。加重F1スコアは0.901だった。

(出所) 大和総研作成

補論図表 3 日銀式で算出された指数系列とコアCPIの推移

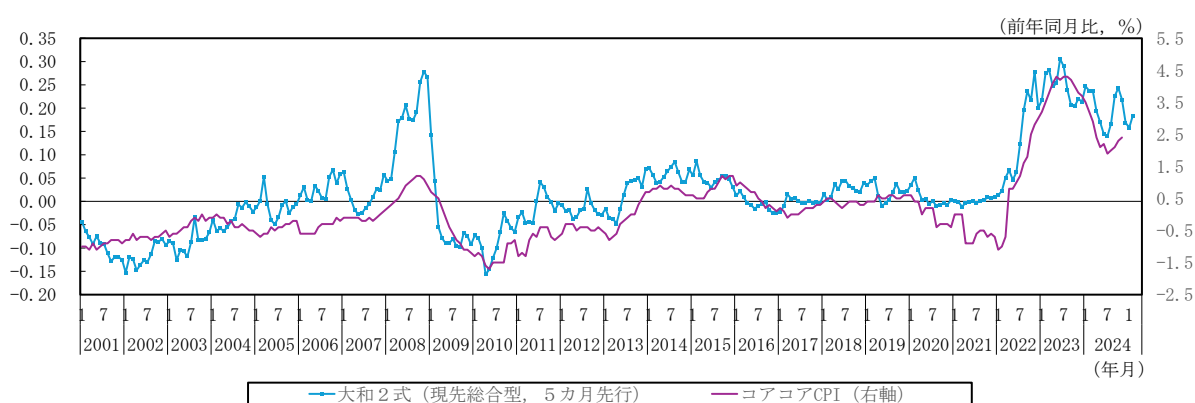


(注1) コアCPIについては、消費税調整済の系列を使用している。

(注2) 日銀式の2種類の指数については、CPIの前年同月比に先行する月数分のみ調整して掲載している。つまり、4カ月先行ならば、2024年12月のところには、2024年7月の指数が載っていることになる。

(出所) 総務省統計より大和総研作成

補論図表 4 大和2式 (現先総合型) のPSI (原系列) とコアコアCPIの推移



(注1) コアコアCPIについては、消費税調整済の系列を使用している。

(注2) 日銀式の2種類の指数については、CPIの前年同月比に先行する月数分のみ調整して掲載している。つまり、5カ月先行ならば、2024年12月のところには、2024年6月の指数が載っていることになる。

(出所) 総務省統計より大和総研作成

参考文献

- Guo, Y., Z. Xu, and Y. Yang (2023) “Is ChatGPT a Financial Expert? Evaluating Language Models on Financial Natural Language Processing,”
<https://arxiv.org/pdf/2310.12664>
- Lee, J., N. Stevens, S.C. Han, and M. Song (2024) “A Survey of Large Language Models in Finance (FinLLMs),” <https://arxiv.org/pdf/2402.02315>
- Suzuki, M., H. Sakaji (2024) “Refined and Segmented Price Sentiment Indices from Survey Comments,” <https://arxiv.org/pdf/2411.09937>
- 大高一樹・菅和聖 (2018) 「機械学習による景気分析—『景気ウォッチャー調査』のテキストマイニング—」, 日本銀行ワーキングペーパーシリーズ, No. 2018-J-8, 日本銀行
- 土井惟成・小田悠介・中久保菜穂・杉本淳 (2024) 「大規模言語モデルを用いたゼロショットテキスト分類による TCFD 推奨開示項目の自動判定」, JPX ワーキング・ペーパー, Vol. 43, 日本取引所グループ,
https://www.jpx.co.jp/corporate/research-study/working-paper/JPXWP_Vol43.pdf
- 長田充弘・中澤崇 (2024) 「多角的レビューシリーズ 期間構造や予測力からみたインフレ予想指標の有用性」, 日銀レビュー, No. 2024-J-5, 日本銀行
- 中島上智・山縣広晃・奥田達志・香月信之輔・篠原武史 (2021) 「景気ウォッチャー調査のテキスト分析からみた企業の短期インフレ予想」, 日本銀行ワーキングペーパーシリーズ, No. 2021-J-12, 日本銀行
- 西野孝佑・山本弘樹・北原潤・永幡崇 (2016) 「『総括的検証』補足ペーパーシリーズ① 『量的・質的金融緩和』の3年間における予想物価上昇率の変化」, 日銀レビュー, No. 2016-J-17, 日本銀行
- 肥後雅博 (2022) 「SNA 推計ならびに SNA 推計に用いる基礎統計に関する課題」, 企画部会第1ワーキンググループ会合説明資料, 資料1, 総務省統計委員会
- 渡辺努 (2022) 「物価とは何か」, 講談社
- 渡辺努 (2024) 「物価を考える デフレの謎、インフレの謎」, 日本経済新聞出版社