



## HUB-FS Working Paper Series

FS-2020-J-002

### 時系列分解を施したニュース指標の投資戦略への 応用とテキスト情報への探索的アプローチ

中山 純

一橋大学大学院経営管理研究科

横内 大介

一橋大学大学院経営管理研究科

First version: 2020年7月9日

All the papers in this Discussion Paper Series are presented in the draft form. The papers are not intended to circulate to many and unspecified persons. For that reason any paper can not be reproduced or redistributed without the authors' written consent.

# 時系列分解を施したニュース指標の投資戦略への 応用とテキスト情報への探索的アプローチ

中山 純\*      横内 大介†

## 概要

近年、ニュース等のテキスト情報と株価の関係に関する研究に注目が集まっている。当分野の先行研究は、主に 1 単位期間における全てのニュースデータを機械的に 1 つの値に要約することでニュース指標を作成し、株価との相関性等を議論する分析であり、各個別のニュースの持つ意味まで考慮する研究は少ない。本研究の目的は、ニュースデータを用いた投資戦略において、テキスト情報の内容や属性等を探索的に解析することで、パフォーマンス改善効果が見込めるか否かを実証することにある。具体的には、QUICK 端末より配信された各ニュースに付与されたポジティブ・ネガティブを表すスコアを用いて作成したニュース指標を活用し、株価指数との相関性の検証や時系列分解手法を適用した投資戦略への応用を行った。分析の結果、ニュース指標の投資戦略における有用性を示唆するとともに、探索的なアプローチによりテキスト情報の重要度を投資戦略に取り入れることでパフォーマンスの向上が期待できることを示した。

## 1 はじめに

我々はニュースに触れる時、意識的かどうかに関わらず、ニュースの情報としての重要性を速やかに判別し、当該ニュースの影響を即座に判断している。近年、大量のニュース情報を大規模データとして取得・蓄積し、定量分析を行うことで株価変動を予測しようという動きが活発になってきているが、これらの研究において、まるで人間が情報を取捨選択するように、ニュース内容まで勘案してその影響を吟味するような分析は少ない。実際、ニュース記事や SNS 情報等のテキスト情報が株価に対して本源的な情報を持つことを期待して、ニュースデータより作成されたニュース指標と株価指数の変動率の関係に関

---

\* 一橋大学大学院経営管理研究科博士後期課程、野村アセットマネジメント株式会社。本研究で示された内容は、野村アセットマネジメント株式会社としての見解をいかなる意味でも表さない

† 一橋大学大学院経営管理研究科

する研究が多数行われている。ニュースデータを扱う限り、人間がニュースに触れた際に情報を整理するフローを意識してニュースの持つ意味や重要性を考慮した分析をすることが重要と考える。

当分野での研究において、ニュースのテキストがポジティブかネガティブかをスコアの形で定量化してニュース指標を作成するのが一般的なアプローチであるが、ほとんどの先行研究では個々のニュースの内容や重要度まで吟味することなく、1 単位期間に配信された全てのニュースデータを使用してニュース指標の時系列データを作成している。しかしながら、実際の証券市場において投資家は、触れた情報の良し悪しだけでなく、その情報の新規性等から各々の重要度を判断した結果として投資行動を起こすと考えられる。このような投資家の情報整理に関するフローを前提とすると、その日の全てのニュースを盲目的に採用して定量化するのでは不十分だと考える。なぜなら、配信されるニュース全てを使用して日々のニュース指標を作成することは、株価にすでに反映されているであろう、過去の事実のみを伝えるニュースや何度も繰り返されるニュース等のノイズを多く含んでしまう可能性が高いためである。

本研究は、この問題意識に対するアプローチとして、個別のニュースデータに対して探索的なデータ解析を行っている。探索的データ解析 (Exploratory data analysis) とは、J.W.Tukey により提唱された、データの持つ価値を発見するためには肝要なデータ解析手法のことである。Tukey(1977) で示された探索的なデータ解析は、特定のデータ分析に役立つ技術群のようなものではなく、どのような姿勢でデータを科学的に扱うべきかを指し示すものであると言える。J.W.Tukey は探索的データ解析のポイントとして、例えば、何かしらのモデルを機械的にデータに当てはめるのではなく、データの分布や外れ値等を慎重に検証することで隠されたデータ構造を明らかにし、これらの特徴に基づいた適正な仮説を立てて仮説検定を行う、といった方法論を採ることの重要性を説いている。また、データの特徴を一目で理解しやすくするための多様なグラフィカル表現を用いてデータ構造の検証にあたり、モデルをデータに当てはめた後の残差を入念に精査することでモデル自体の精度をあげていく、といった丁寧な解析手法も探索的データ解析の特長と言える。本研究では、ほとんどの先行研究で行われたような機械的な処理に基づく分析を行うのではなく、個別のニュースデータに焦点を当てた探索的なアプローチを採ることで、より効果的なテキスト情報活用の可能性を探る分析を行っている。

ニュースと株式市場の関係を分析した先行研究のうち、Tetlock (2007) は、ウォール・ストリート・ジャーナルの記事を基に作成したニュース指標を用いて情報理論を検証した当分野における近年の代表的な研究である。本研究は、ニュース指標と株式リターン、出来高からなる 3 変量 VAR モデルに基づいて情報理論を検証し、市場観測コラム記事の中

に含まれる悲観的な語句の出現度合いが、翌日以降の米国株価指数リターンに一時的な影響を与えていることを明らかにした。しかしながら、株価はその後短時間で一時的な影響を打ち消すようリバウンドしていることが確認され、テキスト情報が株価に対して本源的な情報を持っているという証拠は少ないと結論付けている。また、ニュースの株価への影響に対して懐疑的な主張を行った研究として、Roll(1988)があげられる。Roll(1988)は、米国上場株式の月次リターンを用いて分析を行い、個別銘柄レベル、もしくは、業種レベルにおいても、ニュースに株価に対する説明力はほとんどないことを、 $R^2$ を用いて示している。

一方、ニュース指標が翌営業日の株式リターンに有意な説明力を持つことを確認し、ニュースが株価に対して本源的な情報を持つことを示唆した研究としては沖本・平澤(2014)がある。沖本・平澤(2014)はQUICK 端末より配信されるニュースに基づくニュース指標を作成しニュース情報による国内株式市場の予測可能性について検証したもので、情報理論の検証を国内株式市場に適用した代表的な研究である。主な分析結果としては、作成したニュース指標は翌営業日の株式リターンに有意な説明力を持つことが報告されている。また、国内株式市場に対する同分野の研究として、石島他(2013)は、日本経済新聞の朝刊記事の言語解析により作成した経済状態を表す指標を用いて、ニュース記事が株式リターンに与える影響を研究している。リーマンショック後の期間において、作成した指標は3日後の株式リターンを有意に説明・予測し得ることを確認しており、沖本・平澤(2014)と同様な結論となっている。株式リターンの影響に3日のラグが生じた理由として予想されることは、ニュースによる市場参加者の感情が伝播するにはある程度の時間がかかることと、金曜日の取り扱いが影響を与える曜日効果の影響をあげている。さらに、ニュースの日本株式市場に対する予測力を示す研究として、前川他(2015)はBloombergへ掲載される日本語ニュースのテキスト情報を用いて、日経平均株価の将来水準を予測するモデルを作成し、その有効性を確認している。蔵本他(2012)は、和泉他(2011)が提案したCPR法<sup>\*1</sup>という手法を用いて、日本経済新聞のテキスト情報から株価指数の騰落率を予測し高い正答率を示した。また、五島・高橋(2016)は、ディープラーニングを用いてニュース情報の評判分析を行い株価指数との関係を分析した研究で、ニュース指標が株価指数に対して後追いで反応性している可能性があることを示している。

米国株式市場におけるニュース情報の株価への影響を示唆した研究としてBoudoukh et al.(2012)があげられる。当研究は、対象とするニュースの内容がポジティブかネガ

---

<sup>\*1</sup> CPR法は和泉他(2011)が提案した共起解析、主成分分析、回帰分析の三段階からなるテキスト情報の解析手法で、膨大なテキスト情報から有益な情報を抽出することにより経済動向の効果的な予測が期待できる。

ティブかというトーンを勘案してテキスト情報の米国株式の株価への影響を計測し、Roll(1988)とは反対の結論を主張した研究である。ニュースのトーンを考慮したニュース指標を作成することで、一定程度株価を説明しうることを、 $R^2$ を用いて確認した。また、Tetlock(2011)は、複数のニュースにおける類似性を活用して、投資家の投資行動に関する検証を行ったもので、上記であげた先行研究では見られなかったニュースの新規性に着目した研究である。テキストマイニングの技術を用いてニュースに含まれる単語の出現率を計算し、一連の複数ニュースの類似性の算出に活用することでニュースの新規性の判定を行っている。当研究の主な分析結果としては、特に個人投資家は、しばしば新しいニュースと既出のニュースの区別がつかないまま売買している傾向があることを確認している。これらの先行研究の成果を鑑みるに、新規に発表された情報価値のあるニュースの内容まで加味して適切に使用することができれば、ニュースデータから作成したニュース指標をさらに効果的に活用できる可能性があると考えられる。本研究では、ニュース内容のトーンを加味して定量化されたニュース指標を作成するだけでなく、前述したとおりのニュース情報に対する探索的なアプローチによって、ニュース内容が過去の事実を繰り返す類のニュースの影響を加味した投資戦略の分析、検証を行っている。

本研究の特徴としては、このような個別ニュース内容まで勘案してニュース指標を活用した投資戦略を提案するだけでなく、先行研究では活用できなかったニュース配信時間の情報を利用して実現性の高い out-of-sample での分析を行ったことがあげられる。本研究のニュースデータとしては、沖本・平澤(2014)と同様、QUICK 端末で配信されるニュースを利用しているが、沖本・平澤(2014)では活用できなかったニュース配信時刻を示すタイムスタンプを活用して、株式市場取引終了後のニュースと取引時間中のニュースを区別し、投資タイミングの意味で実現性のある分析を行っている。また、本研究が提案する投資戦略については、投資開始タイミングを判断する手法として Nakayama and Yokouchi(2018)<sup>\*2</sup> で用いられた時系列分解手法を適用している。具体的には、作成したニュース指標の株価指数に対する一定の相関性を背景に、ニュース指標と株価指数の一時的な乖離が短期間で収束する作用を活用した投資戦略を提案している。Nakayama and Yokouchi(2018)の時系列分解手法を応用することで、2つの時系列データの乖離における長期トレンドを除去し、短期的な変動を捉える意味で効果的な計測方法を採用している。

本稿の構成は以下のとおりである。次章は、本研究で用いたデータについて触れ、第3

---

<sup>\*2</sup> Nakayama and Yokouchi(2018)は、時系列分解手法をインデックスファンドにおけるポートフォリオ構築手法に応用した研究で、株価指数構成銘柄と株価指数間の類似度に基づく運用手法に一定の有効性があることを示している。

章では分析手法、第 4 章では分析結果を記す。第 5 章は本研究が用いた探索的アプローチについて議論し、第 6 章は本分析のまとめとする。

## 2 使用したデータ

### 2.1 ニュースデータについて

本研究では、2010 年 12 月 1 日から 2019 年 6 月 28 日までの期間に QUICK 端末より配信された 200 万件を超えるニュースデータを使用して分析を行った。配信ニュースの対象となった国内上場企業の銘柄数は 3,831 社にのぼる。QUICK 端末で配信されるニュースは多くの市場関係者がリアルタイムで閲覧していると思われる、市場への情報伝達力が高いニュースということができる（沖本・平澤 (2014)）。

株式会社 QUICK より購入し提供頂いた当ニュースデータには、各ニュースが初めて配信された時刻のタイムスタンプや、ニュースの対象銘柄名、ニュース内容のイベント属性情報等の基本的な情報とともに、株式会社金融工学研究所のビッグデータ解析技術を用いて付加されたポジティブ・ネガティブを示すスコアが含まれている。各ニュースのスコアは、株式会社金融工学研究所が自然言語処理技術と独自の推計ルールを用いて定量化されたもので、0 に近いほどマーケットに対するネガティブ反応を、100 に近いほどポジティブ反応を期待することを意味する。本研究では、このスコアを利用して作成したニュース指標の株価指数との相関性の検証や、ニュース指標を活用した投資戦略の提案を行った。使用したニュースデータにおけるスコアの平均値は 54.2、中央値は 51、不偏標準偏差は 31.2 であった。本研究では、当スコアが 50 を上回るニュースをポジティブ、下回るニュースをネガティブと判定し、後述する方法でニュース指標を作成した。なお、ニュース指標の作成にあたり、付与されたスコアが 50 に等しいニュースは除外して算出している。

ニュース内容を表すイベント属性毎の配信件数上位 5 イベントのスコア統計値を表 1 に示した。配信ニュースの特徴として、「市場\_上昇」や「市場\_下落」のように、市場において過去に発生した事実を伝えるニュースが多く含まれるのが分かる。具体的には、「市場\_上昇」は、直前の市場である銘柄の株価が上昇した、という事実を伝えるものである。

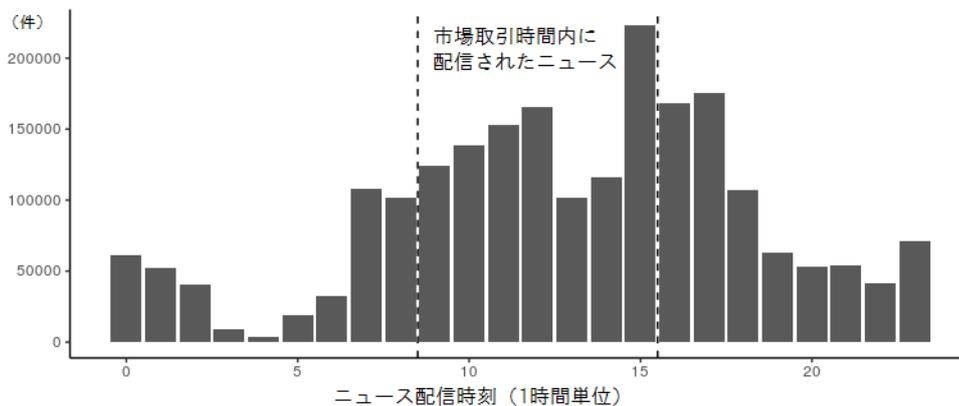
図 1 は、分析対象として全配信ニュースの発生時間帯の分布を示した。ニュース配信件数では、9:00~15:00 の取引時間中のニュース配信が多く、大引けの 15:00 で配信件数は最大となった。また、取引終了後は徐々に投稿は減少傾向となることが確認できる。実際、15:00 ちょうどに配信されたニュースを受けて株式市場内での個別銘柄の取引を行

表 1 ニュース配信件数上位 5 イベント属性毎のスコア統計値

No.	イベント属性	平均	中央値	最大	最小	不偏標準偏差
1	市場_上昇	77.5	91	99	14	20.7
2	市場_下落	24.4	11	85	1	20.9
3	大量保有_増	49.0	48	96	11	11.8
4	大量保有_減	38.1	37	86	17	5.6
5	利益_上昇	62.2	65	96	4	13.7

うことは不可能であり、このような特徴を持つニュースソースを用いる投資戦略の提案を行う上では、本研究が採用したようなニュース配信時刻が取引時間中か否かを明確に区別した分析を行うことが重要と考える。なお、本研究で使用した株価指数である TOPIX のデータ等、ニュースデータ以外のデータは Bloomberg より取得して使用している。

図 1 全配信ニュースの発生時間帯の分布



## 2.2 配信ニュース内容と株価動向の関係

ニュース内容とその直後の株価のパフォーマンスの関係を検証し、本研究で使用したニュースイベント属性毎の株価リターン傾向を把握するため、配信件数上位 30 イベント属性におけるイベント属性毎のニュース配信直後の TOPIX に対する平均超過リターンを付録 A に記した。表 8 は市場の取引時間中に配信されたニュースを対象に、ニュース対象銘柄を当日の大引けの株価で買い付けた場合の翌営業日寄り付き付までの対 TOPIX 超過リターンを O/N、翌営業日大引けまでを D1、3 営業日後の大引けまでを D3、5 営業

日後の大引けまで D5 として示している。さらに、株価に対する配信ニュースの同時点での影響を検証するために配信日当日の日中リターンである当日寄り付きから大引けまでの超過リターンを INT として表している。INT の大きさでは、過去に実現された株価変動を伝える「市場\_上昇」・「市場\_下落」のニュースを除けば、「株式分割」や「自社株買い」、「利益\_ポジ\_対 Q コンセンサス」がプラスで、「格下げ」や「利益\_ネガ\_対 Q コンセンサス」、「不祥事」がマイナスのパフォーマンスとなる傾向が見て取れる。また、全体的な傾向としては、初動の INT の影響がその後も緩やかながらも持続するモメンタム効果の傾向も確認できる。特に、No.24 の「株式分割」は 5D でも大きなプラスの超過リターンが持続している。

また、表 9 は市場の取引時間終了後に配信されたニュースを対象に、上記表 8 のタイミングよりも半日遅い翌営業日の寄り付きの株価で買い付けられた場合の翌営業日大引けまでを INT、翌々営業日寄り付きまでを D1、3 営業日後の大引けまでを D2.5、5 営業日後の大引けまで D4.5 として算出した。これに加えて、上記と同様、配信ニュースそのものの同時点での影響を検証するために配信日当日の大引けから翌営業日寄り付きまでの O/N も示している。ニュースそのものの影響 (O/N) の大きさでは、「増配」等の配当関連のニュースが大きくプラスのリターンとなった一方「公募増資」は大きくマイナスとなった。さらに、「株式格付け\_格上げ」のニュースは 5 営業日後でも持続している傾向があった。「利益\_ポジ\_対 Q コンセンサス」や「利益\_ネガ\_対 Q コンセンサス」は 5 営業日後でも順方向のパフォーマンスとなった傾向は表 8 の大引け買付の場合と同様であった。表 8 と表 9 により、平均値としてポジティブな内容のニュースにはプラス、ネガティブなニュースにはマイナスとで反応するという自然な現象を確認するとともに、イベント属性毎のパフォーマンス格差の存在から、これらを詳細に分析することでイベント属性に応じたシンプルな投資戦略に活かせる可能性を示唆していると考えている。

付録 B では各ニュースに付与されたスコア水準の 10 分位分析による平均的な対 TOPIX 超過リターンを示している。表 10 は大引け買い付け、表 11 は寄り付き買い付けでの結果を示している。どちらの場合でも、概ねスコア通りのパフォーマンス、つまり高いスコアは正の超過リターン、低スコアは負の超過リターンとなる傾向が示されている。表 10 の大引け買い付けでのパフォーマンスでは、D5 でもプラス継続したのは Q6 と、Q9、Q10 で、明確なマイナス傾向となったのは Q1 から Q3 の 3 分位であった。Q8 が O/N からマイナスリターンであったは意外な結果となった。表 11 は寄り付き買い付けでのパフォーマンスであるが、Q9、Q10 の特徴として O/N で大きくプラスに反応した後は低調なパフォーマンスとなった。この結果からは、取引時間終了後に配信されたニュースは直後の寄り付きで当該材料が概ね株価へ織り込まれ、寄り付き後からのパフォーマンス

への影響が薄かった、という推移が表されていると言える。

### 3 分析手法

本研究では、QUICK 端末で配信されたニュースデータを元にニュース指標を作成し、これを活用した投資戦略を提案する。以下の節でニュース指標の作成方法や具体的な投資戦略案について説明する。

#### 3.1 ニュース指標の作成方法

本研究で使用したニュース指標は、Tetlock et al.(2008) と沖本・平澤 (2014) を参考に、ポジティブ・ネガティブを表すスコアが付与され、かつ、上場銘柄に紐づけされた<sup>\*3</sup>配信ニュース数データを用いて以下の式 (1) により日次で算出した。

$$NI = \frac{P - N}{P + N + 1} \quad (1)$$

ここで、 $NI$  はニュース指標、 $P$  はスコアが 50 より大きいポジティブニュースの数、 $N$  はスコアが 50 より小さいネガティブニュースの数である。ニュース指標作成法において、沖本・平澤 (2014) と本研究とは、何時までに配信されたニュースを当該日のニュース指標作成に含めるか、という点に違いがある。沖本・平澤 (2014) では 24 時までに配信されたニュースを当日分として使用したが、本研究では株式市場の取引時間中に配信されたニュースとして、15 時までに配信されたニュースを当日分としてニュース指標作成に使用している<sup>\*4</sup>。取引時間中に配信されたニュースは株価に反映されている可能性が高いため、ニュース指標と株価指数の関係を捉えた分析、投資戦略の提案を行う上では、厳密にニュース配信時刻で対象ニュースを区切ってニュースの株価への影響をより効果的に捉える分析を行うことが重要と考える。

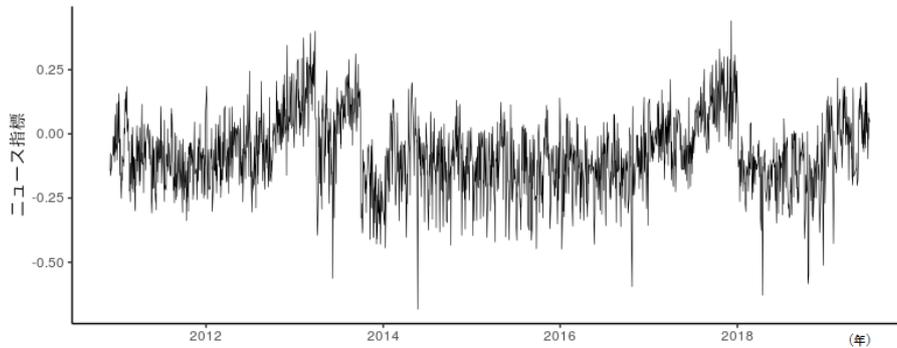
図 2 に作成したニュース指標の推移を示している。本指標は式 (1) のとおりポジティブなニュース数がネガティブニュース数と比較し、相対的に多くなるほど高い水準となるものであり、本研究の全分析期間における平均値は-0.076、最大は 0.44 で、最小は-0.68 となった。

---

<sup>\*3</sup> 先行研究と同様、国内上場銘柄に紐づけされていないニュースは分析の対象外とした。

<sup>\*4</sup> 沖本・平澤 (2014) ではデータに含まれなかったニュース配信時刻を示すタイムスタンプの情報を本研究では利用可能であったことで、株式市場の取引時間終了前後で配信ニュースを区別したニュース指標作成が可能となった。

図2 ニュース指標の推移



### 3.2 時系列分解によるトレンド抽出

本研究は 3.3 節で後述するように、ニュース指標と株価指数との一定の相関性を前提に、ニュース指標と株価指数の間に発生する乖離を捉えて投資タイミングを計る戦略を提案している。本研究の分析期間である 2010 年 12 月から 2019 年半ばという期間、分析対象とした株価指数の TOPIX は年率 +9.4 % と大きく上昇<sup>\*5</sup>しており、明確なプラスの長期トレンドを持っていた。このため、株価指数の持つ長期的なトレンドが邪魔をし、株価指数とニュース指標の間の短期的な乖離を効果的に捉えられないことが懸念される。そこで本研究は、Nakayama and Yokouchi(2018) で用いられた時系列分解手法を適用し、分析対象時系列データから長期トレンドを抽出・除去することでニュース指標と株価指数の短期的な乖離を捉える手法を採用した。

一般的に、ある時系列データのトレンドの抽出を行う際、事前にどんなサイクルで何のモデルがあてはめるのが適切かを特定するのは難しい。したがって本研究では、Nakayama and Yokouchi(2018) と同様、何かしらの経済的サイクルに対する仮説や、ARIMA 等の特定の時系列モデルの前提を必要としないノンパラメトリックなトレンド抽出法を採用した。我々が採用した手法は“Lowess” (局所回帰平滑化) と呼ばれるもので、当初は Cleveland(1979) により提起され、Shibata and Miura(1997) が初めて金融時系列データに応用した平滑化手法である。Lowess により事前に特定のモデル等をあてはめることなく、移動平均法等の伝統的な平滑化手法に比べてより自然な平滑化が可能となる。

先行研究と同様、本研究もオリジナルの時系列データに対してこの平滑化処理を施して

<sup>\*5</sup> 2010 年 11 月末から 2019 年 6 月末までの TOPIX (配当込み) の年率パフォーマンス。

長期トレンドを抽出し、さらにその残差に対して2回目の平滑化処理を施すことで短期トレンドを抽出した。2回目の平滑化処理を行った後の残差はランダムな定常過程のイレギュラー項となり、オリジナルの時系列データ  $Z(t)$  に対する Lowess による時系列分解は以下の式 (2) として表される。ここで、 $t$  は時点、 $L(t)$  は長期トレンド、 $S(t)$  は短期トレンドで、 $I(t)$  はイレギュラー項である。

$$Z(t) = L(t) + S(t) + I(t) \quad (2)$$

図3に、TOPIX (配当込み) の時系列分解例を示した。オリジナルの株価指数 (TOPIX) が、大きな流れを捉える長期トレンドと、より細かな変動を捉える短期トレンド、2つのトレンド抽出後のイレギュラーに分解されているのが分かる。本研究ではニュース指標と株価指数の短期的な乖離を効果的に捉えるため、それぞれ基準化したこれら2つの時系列データに Lowess による時系列分解処理を行うことで抽出した短期トレンド間の乖離を活用して投資戦略に応用している。

### 3.3 投資戦略

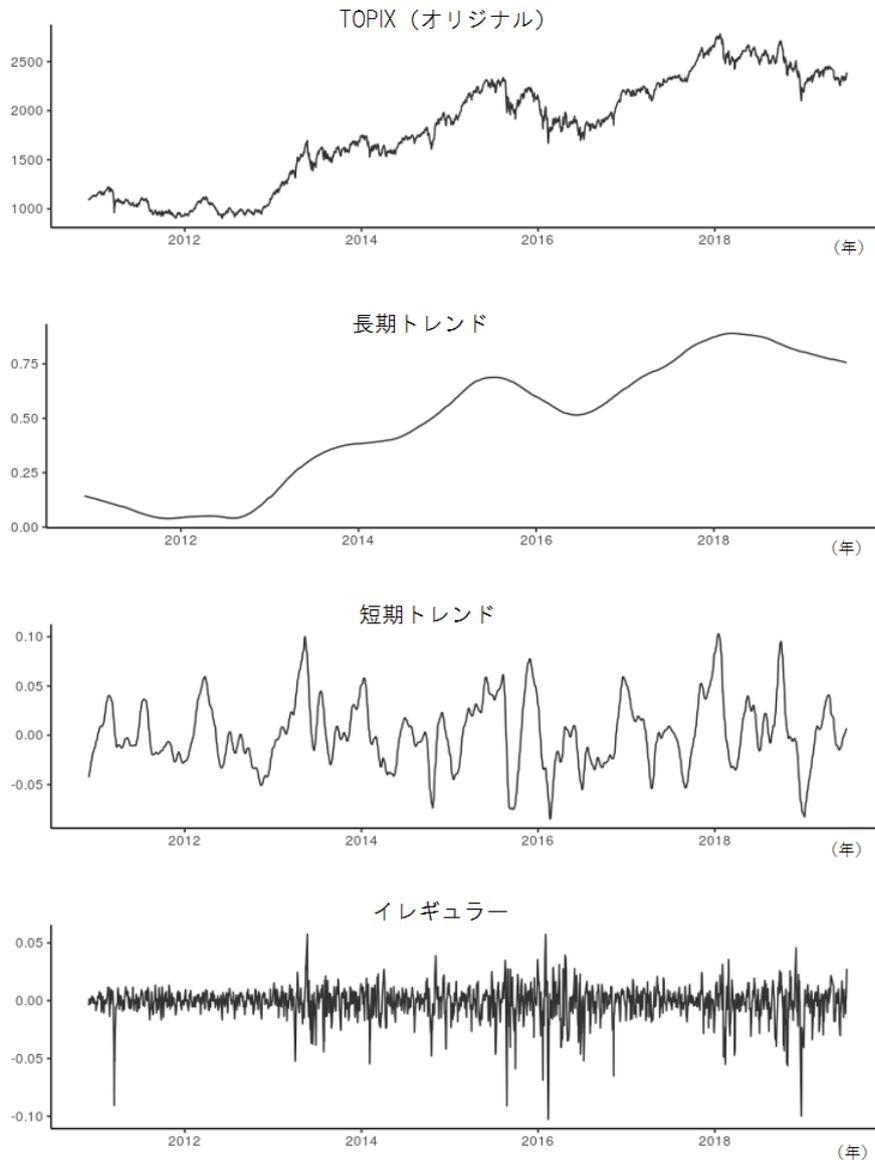
本研究が提案するニュース指標を活用した投資戦略は、上記の手法で抽出したニュース指標と TOPIX の短期トレンドの乖離の絶対値が設定した閾値よりも大きくなった際に、TOPIX 先物 (直近限月) に投資することで正の収益を上げることを目指す戦略である。具体的には、直近営業日のニュース指標の短期トレンドが TOPIX の短期トレンドを大きく上回った時に翌営業日の寄り付きで TOPIX 先物を買って建て、また、ニュース指標の短期トレンドが TOPIX の短期トレンドを大きく下回った時には翌営業日の寄り付きで TOPIX 先物を売り建て、10 営業日後の清算価格でポジションを解消するというものである。4.1 節で示すようなニュース指標と TOPIX の一定の相関性を前提に、長期トレンドを除去した短期トレンド\*<sup>6</sup>で見る2つの時系列データ間の一時的な乖離を捉え、その乖離が収束する作用を活用した戦略と言える。なお、一度先物に投資を開始すると10日間ポジションを保持するため、投資開始のトリガーとしては、先物に投資を行っていない直近10営業日の間で初めて乖離の絶対値が閾値を超えた日、としてシミュレーションを行っている。

図4は、短期トレンド推移の例として、2019年6月末までの3年間におけるニュース

---

\*<sup>6</sup> Nakayama and Yokouchi(2018) を参考に、本研究でも直近3年間 (750 営業日) のデータを使用して時系列分解処理を行った。

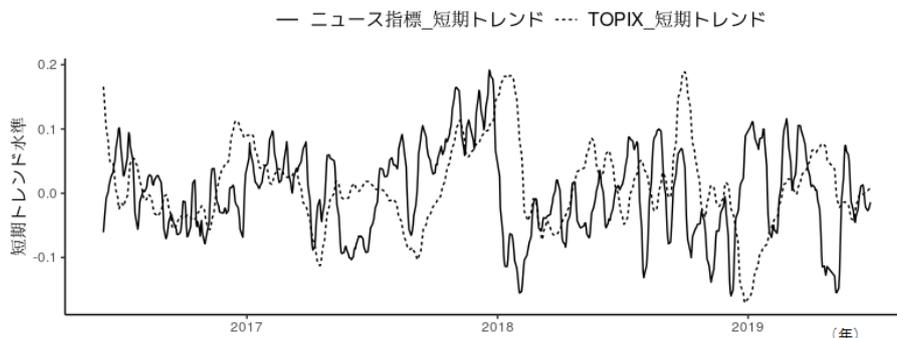
図3 TOPIXの時系列分解



指標と TOPIX の短期トレンドの推移を示している。実線で示したニュース指標の短期トレンドと破線で示した TOPIX の短期トレンドが、一時的に乖離しながらもある程度の相関性を持って推移していると言える。当投資戦略は、図 4 に示された短期トレンド同士の比較を日々更新し続けたデータを用いて行い、直近営業日の乖離に着目することで投資タイミングを計るという戦略である。

なお、当投資戦略で活用するニュース指標は当日の株式市場の取引終了時までには配信さ

図4 時系列分解処理がされた短期トレンドの推移



れたニュースで作成され、取引時間終了後のニュースは翌営業日のニュース指標算出に用いられる。短期トレンド間における乖離の発生を把握した後、翌営業日寄り付きの先物への投資タイミングまでの時間は十分に確保されており、データ取得タイミングと投資タイミングのどちらの時点に関しても実現性の高い投資戦略だと言える。

## 4 分析結果

### 4.1 ニュース指標の株価指数との相関性

本節では、本研究が提案する投資戦略の前提となっているニュース指標と TOPIX の相関性について示す。同時点のニュース指標と TOPIX 対数収益率の相関性を検証するために行った単回帰分析の結果が図5に示されている。推定された単回帰式の切片、回帰係数はともに1%有意水準で有意で、それぞれ0.002と0.019と推定され、2つのデータに正の相関関係があることが示唆された。

さらに、ニュース指標がどの時点の TOPIX に対して影響しているかを検証するため、ニュース指標に5次までのラグを付けた式(3)で表される回帰モデルで分析を行った。ここでは、ニュース指標の TOPIX への影響は1週間以内に消えると想定し、5営業日前まで遡った回帰モデルで分析した。ここで  $TOPIX$  は TOPIX (配当込み) の対数収益率で、 $NI$  はニュース指標である。

$$TOPIX_t = \alpha_t + \sum_{k=1}^5 \beta_k NI_{t-k} + \varepsilon_t \quad (3)$$

図5 ニュース指標と TOPIX の相関性

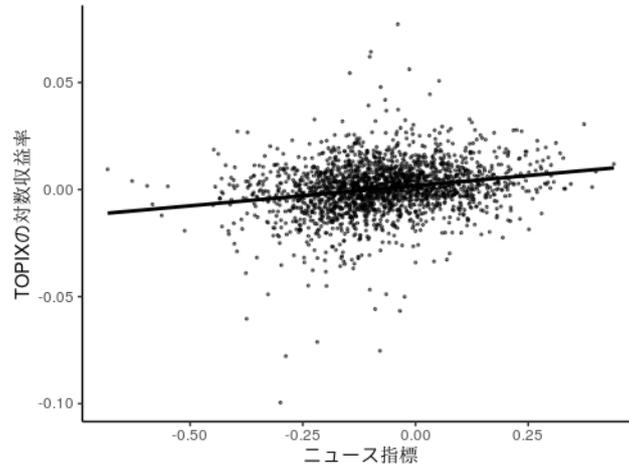


表2 にニュース指標の各回帰係数の推定結果を示しているとおり、 $\beta_1$  が 10% 有意水準で正の関係で有意となり、ニュース指標が翌営業日の TOPIX 収益率に正の影響を与える可能性があることを示唆している。2つの時系列データの乖離に着目した投資戦略において、この関係を前提に一時的な乖離が収束する作用を狙った投資タイミングを計る着想点となっている。

表2 式(3)の推定結果

	係数推定値	t 値
$\beta_1$	0.0041*	1.70
$\beta_2$	0.0004	0.16
$\beta_3$	0.0019	0.70
$\beta_4$	-0.0016	-0.60
$\beta_5$	0.0005	0.19
調整済み $R^2$	0.0098	

注:\*, \*\*, \*\*\*はそれぞれ 10%、5%、1% の有意水準で有意であることを表す。

## 4.2 投資戦略のパフォーマンス

本節は、3.3 節で示した投資戦略のシミュレーションの結果を示す。表 3 は、本研究の分析期間における 5% 刻みの閾値毎の投資回数や先物の買い建て売り建ての内訳、平均期間リターン<sup>\*7</sup>、不偏標準偏差、勝率である。まず、投資回数については、閾値が大きくなるほど投資開始のトリガーが発生し難くなるため、分析結果に示されているとおり閾値の水準が大きくなるほど投資回数が減少するのは自然な現象と言える。平均期間リターンや勝率の水準を見ると、閾値が 15%<sup>\*8</sup>を超えると当投資戦略のパフォーマンスが良好になるのが分かる。閾値が 15% 以上の 3 パターンのパフォーマンスは、勝率は概ね 60% 程度で、平均期間リターンは +1% を超えるものとなった。投資回数の意味では、閾値が 15% の設定で投資回数 49 回というのは、1 カ月に約 0.8 回の投資を行う水準となっており、これは約 20 営業日ある 1 カ月の取引日の中で、7.6 日間のポジションを取っているという戦略<sup>\*9</sup>である。

表 3 投資戦略のシミュレーション結果

閾値	投資回数	買建/売建回数	平均期間リターン	$\sigma$	勝率
5%	70	25/45	0.43%	3.63%	55.71%
10%	67	30/37	0.14%	3.22%	49.25%
15%	49	25/24	1.16%	3.22%	61.22%
20%	32	18/14	1.04%	4.46%	59.38%
25%	16	8/8	1.48%	3.46%	62.50%

注:直近 3 年間のデータを用いた時系列分解により抽出した短期トレンドを用いた投資戦略を作成しているため、実際の投資戦略のシミュレーション期間は 2013 年 12 月 17 日から 2019 年 6 月 28 日となる。 $\sigma$  は期間リターンの不偏標準偏差。以下に同じ。

3.2 節で言及したように本研究の分析期間において TOPIX 自体が大きく上昇していたことで、投資タイミングの巧拙に関わらず、先物を一定期間買い建てていけば正のリターンが得られたと考えられる可能性がある。つまり、機動的に先物を取引する当投資戦略の

<sup>\*7</sup> 3.3 節で示した投資開始のトリガーが発生した翌営業日の TOPIX 先物寄り付きから 10 営業日後の先物清算価格までの平均期間リターンのことで、約 9.5 営業日分の平均リターンを表している。

<sup>\*8</sup> 時系列分解処理後のニュース指標の短期トレンド水準と TOPIX の短期トレンド水準の直近営業日における乖離幅の絶対値が 15% を超過した際に TOPIX 先物への投資を開始することを表す。

<sup>\*9</sup> 1 回あたり 9.5 日間のポジションを取る戦略のため、1 か月に 0.8 回であれば  $9.5 \times 0.8 = 7.6$  として算出。

特性上、ポジションが買い建てに偏っているのはニュース指標を活用した当投資戦略の有効性に疑義が生じる可能性がある。これに応えるため、表3に投資回数に占める買い建て、売り建ての回数を示している。これは、例えば閾値が15%の設定では、合計49回の投資のうち、ニュース指標の短期トレンドがTOPIXの短期トレンドを上回った場面で先物買い建てが25回、逆にニュース指標がTOPIXを下回った場面で先物売り建てが24回であったことを表している。相対的に高いパフォーマンスを示した閾値15%以上の設定において、買い建てと売り建ての回数に大きな偏りは見られていないことが分かる。

さらに、先物買い建てと売り建て時のパフォーマンスについて詳細に比較したのが表4である。特に高いパフォーマンスとなった閾値が15%以上に注目すると、平均期間リターンにおいては先物買い建ての方が高いリターンとなった。閾値15%の設定では、買い建ての+1.83%に対して売り建ては+0.46%であった。市場の上昇傾向が買い建て時のパフォーマンスを後押しした可能性もある一方、売り建て時のパフォーマンスも正の収益を得られており、勝率も原則売り建ての方が高い。少ないサンプル数ながら、表3、表4の閾値15%の設定でのパフォーマンスを鑑みるに、ニュース指標を効果的に活用した投資戦略の一定の有効性が示されていると言える。

表4 先物買い建てと売り建て時のパフォーマンス比較

閾値	買い建て			売り建て		
	投資回数	平均期間リターン	勝率	投資回数	平均期間リターン	勝率
5%	25	0.88%	72.00%	45	0.19%	40.00%
10%	30	0.02%	56.67%	37	0.23%	45.95%
15%	25	1.83%	68.00%	24	0.46%	70.83%
20%	18	1.37%	66.67%	14	0.62%	85.71%
25%	8	2.83%	75.00%	8	0.13%	75.00%

## 5 探索的なテキスト情報の解析と活用

4.2節より、ニュース指標とTOPIXの乖離に着目した当投資戦略の有効性が期待できる結果であると言えるが、この戦略は1単位期間において取得したニュースデータ全てを機械的に使用して作成したニュース指標を活用した戦略である。本研究の主旨の一つである“個別ニュース内容や意義を精査するテキスト情報の解析”を行う探索的なアプローチを取ることでさらに当投資戦略のパフォーマンス向上が期待できると考え、本章ではこ

のアプローチによる投資戦略の改良施策について議論し、そのパフォーマンス改善効果を示す。

前述の通り、ニュースデータからニュース指標を作成し、その有用性を議論する先行研究は存在するものの、ニュース内容まで踏み込んでその情報を活用した研究は少ない。そこで本研究では、投資開始のタイミングを計る上で、全体的なポジティブ・ネガティブのトーンを捉えるニュース指標を活用するだけでなく、個別ニュース内容の意味や重要度で投資を開始するか否かの判定を行う手法を導入した。つまり、配信されたニュースのイベント属性毎の特徴や配信件数、さらに、一日に配信されたニュース数に占める特定のイベント属性ニュースの比率等、ニュースデータを探索的に解析することで、先物取引を開始するかどうかの判断における下記の仮説を立てて検証を行った。

2.1 節の表 1 で示した通り、本研究で使用したニュースデータで最も配信件数の多かったイベント属性は「市場\_上昇」で、次点は「市場\_下落」である。既に市場で実現した株価の動きを伝えるニュースが将来の株価与える本源的な影響を鑑みるに、この属性のニュースが多いほどニュース指標と株価指数の乖離に対する信頼感は低下すると考えた。ニュースデータの探索的解析により立てた仮説は、配信された全てのニュース数に占める“既に実現した株価変動を伝える市場ニュース”の割合が高い程、3.3 節で説明した投資開始のトリガーはフェイクで、期待したリターンが得られないのではないかと考えた。具体的には、ニュース指標を活用した投資開始のトリガーが発生した場合でも、「市場\_上昇」と「市場\_下落」の配信数比率が一定以上の日は投資を行わない、という“ストップトリガー”を導入することでパフォーマンスの向上が期待できると考えた。言い換えると、ニュース指標の投資開始トリガーについて、当日のニュース内容が株価に与える影響が高いと考えられる時のみ、先物への投資を開始する戦略と言える。実際の投資戦略への実装方法としては、全配信ニュースに占める上記の株価変動ニュースの配信数比率が平均的な水準よりも 5% 以上超過した日は、たとえニュース指標と TOPIX の乖離による投資開始のトリガーが発生していたとしても先物への投資を行わない、とした。

このパフォーマンス向上施策を導入した後のシミュレーション結果を表 5 に示している。閾値 15% の設定で、平均期間リターンは +2.10% と導入前の +1.16% を大きく上回り、勝率も 72% 程度と 61% から上昇している結果となった。また、表 6 にストップトリガーが発動した日、つまり、株価変動ニュースの配信数比率が高く投資を止めた日のパフォーマンスを示しているが、平均期間リターン、勝率、ともに低調なパフォーマンスであったことが分かる。サンプル数は少ないものの、この結果は当投資戦略のパフォーマンス向上施策として立てた上記の仮説の有効性を示している可能性があると考えられる。

表5 パフォーマンス向上施策導入後のパフォーマンス

閾値	投資回数	平均期間リターン	$\sigma$	勝率
5%	48	1.14%	3.68%	56.25%
10%	48	0.37%	3.02%	54.17%
15%	36	2.10%	3.17%	72.22%
20%	22	2.26%	4.17%	72.73%
25%	10	3.19%	3.69%	80.00%

表6 ストップトリガー発動時のパフォーマンス

閾値	ストップ回数	平均期間リターン	$\sigma$	勝率
5%	22	0.20%	3.59%	54.55%
10%	19	-0.17%	3.75%	36.84%
15%	13	-0.60%	2.74%	30.77%
20%	10	-1.38%	4.32%	30.00%
25%	6	-0.08%	2.61%	33.33%

表7に、投資戦略シミュレーションの全期間におけるパフォーマンス向上施策導入前後のパフォーマンス比較を示している。これは、閾値15%の設定でのストップトリガー導入による年率パフォーマンスの変化を表している。ストップトリガーの導入で年率リターンは+7.89%と導入前から1.4%程度上昇し、年率リスクは9.40%と1.35%程度低下している。シャープ比で見ると、導入後は0.84と0.24程度上昇し、上記仮説に基づくパフォーマンス向上施策は期間全体で見ても狙い通り、パフォーマンスを引き上げる効果を発揮できたと言える。

表7 パフォーマンス向上施策導入前後の閾値 15% の設定でのパフォーマンス比較

	ストップトリガー導入前	ストップトリガー導入後
投資回数	49	36
年率リターン	6.49%	7.89%
年率リスク	10.75%	9.40%
シャープ比	0.60	0.84
最大月次リターン	12.07%	12.07%
最小月次リターン	-7.94%	-7.64%

注:シャープ比は、無リスク利子率として無担保コール翌日物を用いて算出した。

## 6 まとめ

本研究では、QUICK 端末より配信されたニュースに付与されたポジティブ・ネガティブを表すスコアを用いてニュース指標を作成し、株価指数との関係を検証するとともに、時系列分解手法を適用して長期トレンドを除去した短期トレンドを用いて実現性の高い投資戦略への応用を行った。分析の結果、ニュース指標と株価指数の間に一定の相関性が示されるとともに、1 単位期間に配信された全てのニュースデータを機械的に採用したニュース指標を活用した投資戦略において、良好なパフォーマンスが期待できることが示唆された。さらに、テキスト情報の探索的なアプローチにより、イベント属性を適切に識別し株価への影響の意味で重要性の低いニュースデータが多く含まれる日には投資戦略を実行しないというパフォーマンス向上施策の分析を行い、その有効性が示された。

本研究はニュース指標を活用した投資開始のトリガーが発生した場合に翌営業日の寄り付きで先物取引を開始し、その 10 営業日後の先物ポジションを閉じる、という投資戦略を提案したが、この約 10 営業日というポジション保有期間に対しては未だ改善の余地があると考えられる。例えば、ニュース指標と株価指数の乖離状況によってポジション保有期間を変更する等の最適な投資期間の考え方の検証が必要である。この点に加えて、日次単位ではなく 1 分単位、1 秒単位等、より高頻度に短期間のニュース指標の作成と、その高頻度ニュース指標と株価指数との関係、投資戦略への応用の可能性についても考えられ、これらについては今後の課題としたい。

## 付録 A イベント属性毎のパフォーマンス

表 8 大引け買付での対 TOPIX 平均超過リターン

No.	イベント属性	INT	O/N	D1	D3	D5
1	市場_上昇	1.78%	0.63%	0.29%	0.25%	0.10%
2	市場_下落	-1.32%	-0.32%	-0.47%	-0.67%	-0.61%
3	大量保有_増	0.06%	0.05%	0.10%	0.12%	0.11%
4	大量保有_減	-0.13%	0.04%	-0.09%	-0.07%	-0.08%
5	利益_上昇	0.31%	0.07%	0.03%	-0.16%	-0.22%
6	利益_上昇:市場_上昇	1.04%	0.28%	0.15%	0.13%	0.16%
7	利益予想_上昇:市場_上昇	1.31%	0.71%	0.41%	0.52%	0.36%
8	株価格付_格上げ	0.44%	0.30%	0.45%	0.49%	0.62%
9	株価格付_格上げ:市場_上昇	1.33%	0.13%	0.17%	0.19%	0.21%
10	売上_増加	0.07%	0.03%	0.08%	0.12%	0.11%
11	利益_低下	-0.31%	-0.26%	-0.41%	-0.65%	-0.57%
12	株価格付_格下げ	-0.47%	-0.31%	-0.51%	-0.59%	-0.50%
13	利益予想_低下:市場_下落	-0.78%	-0.53%	-0.75%	-0.99%	-0.99%
14	利益_低下:市場_下落	-0.78%	-0.20%	-0.32%	-0.49%	-0.51%
15	利益_ネガ_対Qコンセンサス	-0.28%	-0.09%	-0.27%	-0.54%	-0.62%
16	株価格付_格下げ:市場_下落	-1.12%	-0.14%	-0.11%	-0.13%	0.01%
17	売上_減少	-0.13%	-0.03%	0.01%	0.03%	-0.05%
18	利益_上昇:市場_下落	-1.48%	-0.08%	-0.35%	-0.69%	-0.84%
19	不祥事	-0.31%	-0.59%	-0.61%	-1.15%	-1.13%
20	自社株買い	2.45%	0.10%	0.18%	0.22%	0.29%
21	利益予想_上昇	0.39%	0.47%	0.36%	0.38%	0.57%
22	自社株買い:市場_上昇	0.61%	0.06%	-0.15%	-0.13%	-0.05%
23	市場_上昇:売上_増加	1.69%	0.43%	0.35%	0.31%	0.15%
24	株式分割	3.97%	1.30%	1.16%	1.22%	1.57%
25	利益_ポジ_対Qコンセンサス	0.88%	0.34%	0.47%	0.75%	0.88%
26	利益予想_低下	-0.22%	-0.53%	-0.80%	-1.28%	-1.01%
27	虚偽・偽装	-0.25%	-0.58%	-0.71%	-0.70%	-0.26%
28	利益予想_上昇:市場_下落	-1.73%	0.07%	-0.27%	-0.40%	-0.53%
29	市場_上昇:市場_下落	-0.40%	0.00%	-0.19%	-0.40%	-0.27%
30	利益_低下:市場_上昇	2.05%	0.12%	-0.01%	-0.36%	-0.39%

注:取引時間中に配信されたニュースに対して、INT はニュース配信日の寄り付きから大引け、O/N は大引けから翌営業日寄り付き、D1 は大引けから翌営業日の大引け、D3 は大引けから 3 営業日後の大引け、D5 は大引けから 5 営業日後の大引けの平均リターンとして算出。「:」で区切られたイベント属性は複数のイベントが付与されたことを表し、スコア算出はポジティブ・ネガティブの評価が複数発生したとして処理される。以下に同じ。

表9 寄り付き買い付けでの対 TOPIX 平均超過リターン

No.	イベント属性	O/N	INT	D1	D2.5	D4.5
1	市場_上昇	2.10%	0.09%	0.06%	0.28%	0.34%
2	市場_下落	0.30%	0.08%	0.12%	0.14%	0.32%
3	大量保有_増	0.20%	-0.09%	0.09%	-0.14%	-0.14%
4	大量保有_減	-0.60%	-0.14%	-0.02%	-0.15%	-0.15%
5	利益_上昇	0.70%	-0.17%	0.01%	-0.26%	-0.28%
6	利益_上昇:市場_上昇	3.80%	-0.12%	0.04%	-0.32%	-0.43%
7	利益予想_上昇:市場_上昇	0.30%	0.00%	0.04%	0.07%	0.05%
8	株価格付_格上げ	6.10%	-0.13%	0.00%	-0.32%	-0.21%
9	株価格付_格上げ:市場_上昇	0.90%	-0.18%	0.05%	-0.26%	-0.26%
10	売上_増加	-0.70%	0.00%	-0.05%	-0.12%	-0.12%
11	利益_低下	1.70%	-0.02%	-0.17%	-0.26%	-0.33%
12	株価格付_格下げ	6.20%	-0.41%	0.92%	0.17%	0.36%
13	利益予想_低下:市場_下落	-0.50%	-0.14%	-0.02%	-0.28%	-0.31%
14	利益_低下:市場_下落	8.90%	-0.25%	0.18%	-0.40%	-0.43%
15	利益_ネガ_対Qコンセンサス	-0.20%	-0.02%	-0.02%	-0.05%	-0.12%
16	株価格付_格下げ:市場_下落	1.90%	-0.21%	0.11%	-0.17%	-0.13%
17	売上_減少	-0.90%	-0.02%	-0.12%	-0.18%	-0.45%
18	利益_上昇:市場_下落	5.80%	-0.08%	0.15%	-0.50%	-0.62%
19	不祥事	-2.40%	0.10%	-0.16%	-0.18%	-0.18%
20	自社株買い	-3.90%	0.16%	-0.07%	-0.09%	-0.73%
21	利益予想_上昇	1.40%	-0.15%	0.08%	-0.21%	-0.15%
22	自社株買い:市場_上昇	-0.80%	0.19%	-0.04%	0.09%	0.07%
23	市場_上昇:売上_増加	-0.90%	0.25%	-0.02%	0.17%	0.13%
24	株式分割	1.80%	0.00%	0.01%	0.16%	0.29%
25	利益_ポジ_対Qコンセンサス	-0.60%	0.01%	-0.04%	0.17%	0.51%
26	利益予想_低下	0.20%	-0.90%	0.14%	-1.26%	-1.17%
27	虚偽・偽装	5.30%	-0.40%	0.32%	-0.21%	0.03%
28	利益予想_上昇:市場_下落	-0.20%	-0.12%	0.05%	-0.09%	0.11%
29	市場_上昇:市場_下落	2.70%	-0.17%	0.11%	-0.29%	-0.45%
30	利益_低下:市場_上昇	0.00%	-0.21%	-0.05%	-0.35%	-0.37%

注:取引時間終了後に配信されたニュースに対して、O/N はニュース配信直前の大引けから翌営業日の寄り付き、INT はニュース配信直後の寄り付きから大引け、D1 はニュース配信直後の寄り付きから翌営業日の寄り付き、D2.5 はニュース配信直後の寄り付きから 3 営業日後の大引け、D4.5 はニュース配信直後の寄り付きから 5 営業日後の大引けの平均リターンとして算出。

## 付録 B スコア水準毎のパフォーマンス

表 10 大引け買付での対 TOPIX 平均超過リターン

スコア分位	INT	O/N	D1	D3	D5
Q1	-1.42%	-0.36%	-0.52%	-0.79%	-0.82%
Q2	-0.77%	-0.42%	-0.64%	-0.88%	-0.71%
Q3	-0.11%	-0.39%	-0.52%	-1.09%	-1.03%
Q4	-0.09%	-0.03%	-0.11%	-0.15%	-0.14%
Q5	0.19%	0.03%	-0.12%	-0.28%	-0.38%
Q6	0.13%	-0.01%	0.10%	0.12%	0.09%
Q7	0.37%	0.06%	0.15%	-0.05%	-0.06%
Q8	0.30%	-0.17%	-0.22%	-0.32%	-0.48%
Q9	1.09%	0.23%	0.17%	0.07%	0.04%
Q10	1.89%	0.66%	0.37%	0.37%	0.31%

注:取引時間中に配信されたニュースに対して、INT はニュース配信日の寄り付きから大引け、O/N は大引けから翌営業日寄り付き、D1 は大引けから翌営業日の大引け、D3 は大引けから 3 営業日後の大引け、D5 は大引けから 5 営業日後の大引けの平均リターンとして算出。

表 11 寄り付き買い付けでの対 TOPIX 平均超過リターン

スコア分位	O/N	INT	D1	D2.5	D4.5
Q1	-2.09%	0.01%	-0.22%	-0.27%	-0.40%
Q2	-0.70%	-0.07%	-0.10%	-0.34%	-0.39%
Q3	-0.25%	0.02%	-0.04%	-0.12%	-0.16%
Q4	-0.21%	-0.07%	-0.01%	-0.18%	-0.17%
Q5	1.09%	-0.07%	-0.01%	-0.20%	-0.21%
Q6	0.31%	0.08%	0.02%	0.06%	0.10%
Q7	0.59%	-0.13%	-0.02%	-0.18%	-0.19%
Q8	0.86%	-0.10%	0.00%	-0.20%	-0.22%
Q9	4.69%	-0.11%	0.17%	-0.07%	-0.01%
Q10	3.52%	-0.13%	0.27%	-0.04%	0.03%

注:取引時間終了後に配信されたニュースに対して、O/N はニュース配信直前の大引けから翌営業日の寄り付き、INT はニュース配信直後の寄り付きから大引け、D1 はニュース配信直後の寄り付きから翌営業日の寄り付き、D2.5 はニュース配信直後の寄り付きから 3 営業日後の大引け、D4.5 はニュース配信直後の寄り付きから 5 営業日後の大引けの平均リターンとして算出。

## 参考文献

- Boudoukh, J., Feldman, R., Kogan, S. and Richardson, M. (2012), “Which News Moves Stock Prices? A Textual Analysis” *NBER Working Paper 18725*.
- Cleveland, W. S., Douglas, M. D., and Terpenning, I. J. (1978), “SABL: A resistant seasonal adjustment procedure with graphical methods for interpretation and diagnosis”, in: A. Zellner (1979). *Seasonal Analysis of Economic Time Series*, NBER Books, National Bureau of Economic Research, Inc, 201-241.
- Cleveland, W. S. (1979), “Robust Locally Weighted Regression and Smoothing Scatterplots”, *Journal of the American Statistical Association*, Vol. 74, No. 368, 829-836.
- Nakayama, J. and Yokouchi, D. (2018), “Applying Time Series Decomposition to Construct Index-Tracking Portfolio”, *Asia-Pacific Financial Markets*, Vol. 25, Issue 4, 341-352.
- Roll, R. (1988), “R2”, *Journal of Finance*, Vol. 43, No. 3, 541-566.
- Shibata, R. and Miura, R. (1997), “Decomposition of Japanese Yen Interest Rate Data Through Local Regression”, *Financial Engineering and the Japanese Markets*, 4, 125-146.
- Tetlock P. C. (2007), “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”, *The Journal of Finance*, Vol. 62, No. 3, 1139-1168.
- Tetlock P. C. (2011), “All the news that’s fit to reprint: Do investors react to stale information?”, *Review of Financial Studies*, Vol. 24, No. 5, 1481-1512.
- Tetlock P. C., Saar-Tsechansky M. and Macskassy S. (2008), “More Than Words: Quantifying Language to Measure Firms’ Fundamentals”, *The Journal of Finance*, Vol. 63, No. 3, 1437-1467.
- Tukey J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley Publishing Company.
- 石島博・數見拓朗・前田章 (2013), 「日次データを用いた市場センチメント・インデックスの構築と株価説明力の分析」, 『第 11 回人工知能学会金融情報研究会資料』, SIG-FIN-011-06.
- 和泉潔 (2012), 「大規模テキストを利用した経済指標分析手法に関する研究」, 『学際大規模情報基盤共同利用・共同研究拠点平成 24 年度共同研究中間報告書』 12-DA05.

- 和泉潔・後藤卓・松井藤五郎 (2011), 「経済テキスト情報を用いた長期的な市場動向推定」, 『情報処理学会論文誌』 Vol. 52, No 12, 3309-3315.
- 沖本竜義・平澤英司 (2014), 「ニュース指標による株式市場の予測可能性」, 『証券アナリストジャーナル』 52(4), 67-75.
- 蔵本貴久・和泉潔・吉村忍・石田智也・中嶋啓浩・松井藤五郎・吉田稔・中川裕志 (2012), 「新聞記事のテキストマイニングによる長期市場動向の分析」, 『人工知能学会論文誌』 Vol. 28, No 3, 291-296.
- 五島圭一・高橋大志 (2016), 「ニュースと株価に関する実証分析ーディープラーニングによるニュース記事の評判分析ー」, 『証券アナリストジャーナル』 54(3), 76-86.
- 前川浩基・中原孝信・岡田克彦・羽室行信 (2015), 「大規模ニュース記事からの極性付き評価表現の抽出と株価収益率の予測」, 『オペレーションズ・リサーチ』, Vol. 58, No. 5, 281-288.