

# ヘッドラインニュースが債券市場に与える影響に関する分析

山 下 泰 央\*・上 瀧 弘 晃\*\*・高 橋 大 志\*\*\*

Analyzing the Influence of Head-line News on Bond Market in Japan

Yasuo YAMASHITA\*, Hiroaki JOHTAKI\*\* and Hiroshi TAKAHASHI\*\*\*

This research analyzed the relationship between headline news and bond market in Japan through text mining technology. As a result of intensive analyses, we found the significant relationship between interest rate changes and the negative (or positive) news classified by keywords. At the same time, we also found that a measure of constructing keyword list has a significant impact on performance of classification.

**Key Words:** finance, asset management business, text-mining technology, bond market

## 1. はじめに

資産運用に有用な情報には、企業のファンダメンタルに関する情報やマクロ経済に関する情報など数多くの種類が存在する。金融機関などをはじめとする投資家は、超過収益源泉の探索のために数多くの分析を行なっている。とりわけ、数値情報を対象とした分析については、取り扱いが比較的容易であることなどの要因から、従来より数多くの報告が行なわれている<sup>1), 2)</sup>(注1)。有価証券報告書やブローカーから配信される報告書などのテキスト情報も数値情報と並び資産運用において有用な情報源の一つと捉えることが可能である。テキスト情報には数値情報に反映されていない情報が含まれている可能性があり、その意味で、テキスト情報の分析を通じ、超過収益の源泉となる有用な情報を獲得できる可能性がある。

テキスト情報と金融市場の関連性について分析した報告<sup>3)~5)</sup>も行なわれており、たとえば、新聞の内容と株価の関連性<sup>6), 7)</sup>、企業の年次報告中のリスクなどに関する単語と企業業績、株価リターンの関連性<sup>8)</sup>、CEOへのインタビューが株価リターン出来高に与える影響の分析<sup>9)</sup>、TwitterとDow Jones指数との関連性の分析<sup>10)</sup>などが行なわれている。Antweiler

and Frank<sup>11)</sup>は、米国の株式市場を対象とした分析を行ない、掲示板の情報と取引高に有意な関係があるとの主張を行なっている。日本においても、新聞の記事から抽出したキーワードと株価との関連を分析した報告<sup>12), 13)</sup>や、ルールベース分類アルゴリズムによりテーマを付与した新聞記事と株価との関連を分析した報告がある<sup>14)</sup>。また、高橋ら<sup>15)</sup>は、日本の株式市場を対象とした分析を行なっており、ファンスマネージャーが投資判断する際の重要な情報源の一つであるヘッドラインニュース(注2)と株価変動に有意な関係があるとの結論を得ている。また、上瀧ら<sup>16)</sup>は、クレジット市場を対象とした分析を行なっており、CDSと社債のスプレッドで、ヘッドラインニュースへの反応が異なることなどの指摘を行なっている。テキスト情報は、多様な情報を含んでいる可能性があるとの同時に数値情報と比べ分析が複雑となる傾向があるため、分析における誤差が大きくなる懸念がある。それらの懸念に關し、Loughran and McDonald<sup>17)</sup>は、テキスト情報の分析においては、各分野ごとに固有の単語のリストをつくる必要性があるとの主張を行なっている。

情報が価格に与える影響に関する実証分析は、市場の効率性に関する議論をはじめとし、これまで数多くの報告が行なわれているが、それらの多くは数値データをもとにしたものである。テキスト情報を用いた分析は、そのような従来より行なわれてきた実証分析に新たな情報源を提供するものと捉えることができる。とりわけ、ヘッドラインニュースは、機

\* 三井住友信託銀行パッシブ・クオンツ運用部  
東京都千代田区丸の内1-4-1

\*\* 三井住友信託銀行債券運用部 東京都千代田区丸の内1-4-1

\*\*\* 慶應義塾大学大学院経営管理研究科 横浜市港北区日吉4-1-1

\* Passive and Quantitative Investment Department, Sumitomo Mitsui Trust Bank, 1-4-1 Marunouchi, Chiyoda-ku, Tokyo

\*\* Fixed Income Investment Department, Sumitomo Mitsui Trust Bank, 1-4-1 Marunouchi, Chiyoda-ku, Tokyo

\*\*\* Keio Business School, Keio University, 4-1-1 Hiyoshi, Kouhoku-ku, Yokohama

(Received February 17, 2012)

(Revised June 25, 2012)

(注1) たとえば、金利の期間構造分析は、債券の分析において最も基礎的な項目の一つであるが、数多くの分析手法が提案されている。また、企業の信用リスクの分析においても、財務データを用いた手法や金融市场の価格変動を用いた手法など数多くの手法が提案されている。

(注2) ヘッドラインニュースとは、ニュース本文の内容を簡単なまとめた情報端末上の1行で表示される数十から数百文字で構成される文字列のことである。

関投資家のファンドマネージャーが触れる主要な情報の一つであり、それら情報が金融市場の価格形成に影響を与えていく可能性は高い。

これらを背景とし、本稿では、ヘッドラインニュースと債券市場の関連性について分析を行なう。本分析ではとりわけ、金利水準の変化に焦点を当てた分析を行なう。ヘッドラインニュースと債券市場の関連性を分析した研究はこれまで、上瀧ら<sup>16)</sup>をはじめとし、いくつかの報告が行なわれているが、ヘッドラインニュースと金利水準との関連性を分析した研究は報告されていない。また、ヘッドラインニュースと日本の金利水準について、分析の際に用いるキーワードのリストの違いについて分析した研究も報告されていない。本研究では、ヘッドラインニュースと金利変化の間に統計的に有意な関連があること、また、キーワードのリストが分類精度に大きな影響を与えることを明らかにする(注3)。

次章において、分析方法について説明した後、3章で分析結果について説明を行なう。4章はまとめである。

## 2. 分析方法

本章では、分析方法について説明を行なう。はじめに、テキスト情報の分析手法について説明を行なった後、分析データについて説明を行なう。

### 2.1 テキスト情報の分析手法

本分析では、日本の債券市場を対象とし、ヘッドラインニュースと金利変化の関連性について分析を行なう。ヘッドラインニュースにより投資家に伝えられるテキスト情報には、その内容により、市場にポジティブな影響を及ぼすものとネガティブな影響を及ぼすもののが存在すると考えられる<sup>15)</sup>。本分析ではポジティブなキーワードとネガティブなキーワードをあらかじめ選定し、そのキーワードを利用することでポジティブなニュース、ネガティブなニュースの判別を行なう。

分析に用いるキーワードは、金融・経済のキーワードである必要があることから、Loughran and McDonald<sup>17)</sup>らの報告にて、米国の財務諸表分析のために用いられたキーワードを採用した。なお、本分析では、日本語のヘッドラインニュースを分析することから、それらキーワードを日本語に訳したもの用いている(「キーA」とする)。

「キーA」において用いられるキーワードは、財務諸表分析に用いられたものであり、その意味で当分析のキーワードとして採用するのは適切と考えられる。その一方で、さらに細かい点に着目すると、本分析は、金融市場を分析対象とするものであることから、財務諸表分析とは異なる性質をもつ可能性がある。そこで、本分析では、金融市場に関するキー

ワードを選別するために、日本経済新聞の金利に関するコラム(MARKETSCRAMBLE: 2008年4月から2010年12月までのテキストを利用)の頻出回数上位の単語と「キーA」の両方に含まれる単語を抽出し、さらに「キーA」に含まれていないコラムにおける出現回数の多い単語を追加した、キーワード(「キーB」とする)を用いた分析を実施した。

「キーA」は米国の財務諸表に含まれるテキストデータをもとにネガティブとポジティブとして抽出された単語を日本語に訳し作成しているが、日本語に訳す際、英単語としては区別されるが日本語の単語としては同じとみなせるものがあることから、もとのキーワード数よりは少なくなっている。また、「キーA」に含まれるキーワードには、訴訟関係と思われる単語が少なからず含まれており、日本の債券市場との関連を分析するために利用するには多少不都合の生じることが予想される。一方、「キーB」は、「キーA」から絞り込んだ単語に、金利関連の新聞記事の頻出上位の言葉で「キーA」に含まれていない単語を追加したものである。「キーA」に比べればキーワード数は少ないものの、より債券市場に関連の深いキーワードが抽出されていると考えられる。

本分析では、2種類のキーワードリスト(「キーA」および「キーB」)を用い、ヘッドラインニュースを、ポジティブなニュースおよびネガティブなニュースに分類する。分類における判定条件は、ヘッドラインニュースにキーワードが1回以上含まれていることとする。たとえば、ネガティブなキーワードが、1語以上含まれていれば、ネガティブなニュースと判定する。ヘッドラインニュースで、ポジティブとネガティブ両方のキーワードを含む場合はポジティブ、ネガティブそれぞれの場合から除いてある。ポジティブ、ネガティブと判定されたヘッドラインニュース件数に対し、ポジティブとネガティブ両方のキーワードを含む場合の件数は平均的には4分の1程度で、ポジティブとネガティブ両方のキーワードを含む場合を除いても影響は大きくないと考えている。本研究では、ネガティブ・ポジティブ両方のキーワードが含まれるヘッドラインニュースの影響を除くため、ヘッドラインニュースにキーワードが1回以上含まれることを判定基準とする方法を採用しているが、ネガティブ・ポジティブ両方のキーワードが含まれる場合にも対応した分析は今後の課題である。情報の抽出法として1日のネガティブ、ポジティブ・キーワードの出現回数を直接集計する方法がありうるが、その方法では1件のヘッドラインニュースにネガティブ、ポジティブの両方のキーワードが含まれる場合を分離することができないため、1件ごとのヘッドラインニュースをネガティブ、ポジティブに分類して集計する方法を探った。

### 2.2 データ

本稿では、資産運用の実務において最もよく用いられているメディアの一つであるQuick社の情報端末において利用可能なヘッドラインニュースを対象として分析を行なった。分析対象期間は2009年6月1日から2011年5月31日までの2年間であり、ヘッドラインニュースの数は、137万件程度

(注3) テキストデータとして本分析では、資産運用ビジネスの実務において最もよく用いられているメディアの一つである、Quick社のヘッドラインニュースを対象として分析を行なっている。また、本分析では短期的に債券価格の変動から利益を獲得することに焦点をあてており、長期的な債券価格の変動については今後の課題である。

Table 1 A breakdown of headline-news

Label	(0-9, 15-24)	(9-15)	All Day
Newspaper	24.1%	11.1%	35.2%
Broker	4.9%	5.3%	10.2%
News Agency	13.3%	11.2%	24.5%
Rating Agency	0.8%	0.2%	1.1%
Others	20.3%	8.6%	28.9%
Total	63.5%	36.5%	100.0%

である。ヘッドラインニュース一件あたりの文字数は、数十から数百であり、比較的短い長さの文章により構成されている。Quick 社のヘッドラインニュースは、債券市場関係者に加え、株式市場関係者や為替市場関係者にも重視されている情報ソースである。そのため、ヘッドラインニュースにおいて取り扱われる内容は債券市場に加え株式市場や為替市場に関するものも存在する。また、新聞各社の記事などのヘッドラインニュースも含まれる。

Table 1 は、ヘッドラインニュースの内容についての内訳を示したものである。「Label」はヘッドラインニュースに含まれる配信元のラベルをもとに、新聞、ブローカー、通信社、格付機関、そのほかに分類したものである。1行目の「0-9, 15-24」は、それぞれ、営業時間外に配信されたニュースであり、「9-15」は営業時間内に配信されたニュースであることを表す。Table 1 における数値は、全ニュース件数に対する割合である。配信元では、新聞や通信社のニュースが多くなっている。営業時間内と営業時間外で比較をすると、営業時間外において多くのニュースが配信されていることを確認できる。個別の項目でみると、通信社から配信されるニュースは、営業時間外であまり差がないのに対し、新聞社から配信されるニュース件数は営業時間外のほうが営業時間内のニュース件数の倍程度に多くなっていることを確認できる。格付機関から配信されるニュースは全体のニュース件数に対する割合は小さいが、営業時間外に配信されるニュース件数が営業時間内に配信されるニュース件数に比べ4倍程度と多くなっている<sup>(注4)</sup>。

Fig. 1 は、月ごとのヘッドラインニュース件数の推移を示したものである。「0-9, 15-24」のラベルは営業時間外に配信されたニュース件数の推移であり、「9-15」のラベルは営業時間内に配信されたニュース件数の推移である。営業時間内のニュース件数には、あまり変化がないが、営業時間外のニュース件数は3月、6月、9月、12月に少なくなっている。Fig. 1 を注意深くみると営業時間内のニュース件数にも同じ傾向がみられる。これは、四半期決算を3月、6月、9月、12月に実施する企業が多いためと考えられる。つまり、決算関連情報を決算月の翌月、翌々月に配信するために、決算月のニュース件数が少ないようみえているものと思われる。

(注4) これらは、市場に与える影響を配慮して市場が開いている営業時間内に配信するニュースを抑制しているのではないかという、市場関係者の一般的な見方を裏付けるものであり興味深い。

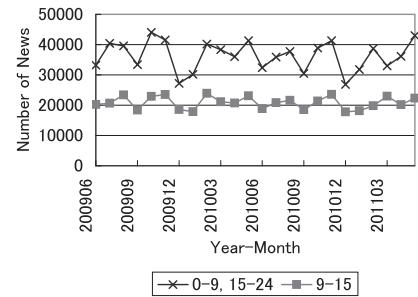


Fig. 1 Number of headline-news (monthly)

本分析では、ポジティブ・ニュースおよびネガティブ・ニュースに分類したヘッドラインニュースの件数を日次で集計し、日次の金利変化との関連について分析を行なう。金利変化を算出するための金利は、残存年数10年の国債ペーレートを利用した。ペーレートの推計に使用した国債価格の時刻が15時であることから、ヘッドラインニュース件数の日次の集計は、前日の15時以降当日の15時までの間隔で集計している。債券市場が休場である土日祝日の件数は、直近翌営業日の15時までのヘッドラインニュース件数の集計に加算するものとした。また、ヘッドラインニュースの件数を営業時間内に限定した場合を分析するために、当日の9時から15時までの間隔でもヘッドラインニュース件数の日次の集計を行なった<sup>(注5)</sup>。

次章において、当キーワード(キーA, キーB)を用いた分析結果について説明する。

### 3. 分析結果

本章では、テキスト情報の分析結果について説明する。はじめに、各キーワードの特徴について分析を行なった後、テキスト情報と金利変化との関係について分析結果を示す。そして、テキスト情報を用いた長期国債先物のシミュレーションの結果を示す。

#### 3.1 基礎分析：キーワードの件数

Table 2 は、それぞれ「キーA」と「キーB」に含まれるキーワード数を示したものである<sup>(注6)</sup>。Table 2 の左列は「キーA」に含まれる negative および positive のキーワード数であり、右側は「キーB」に含まれるキーワード数である。キーAに含まれるネガティブおよびポジティブのキーワード数は、それぞれ951件と214件であり、キーBに含まれるネガティブおよびポジティブのキーワード数は、それぞれ109

(注5) テキストデータの処理は、数理システム社のText Mining Studioを利用した。テキスト処理に使用したパソコンは、CPUがIntel core i7 2600kで、OSはWindows7の64bit版を利用した。計算に要した時間は、たとえば、137万件のヘッドラインニュースの分かち書きに4時間程度の時間を費やしている。また、キーワードによるポジティブ、ネガティブの判定もText Mining Studioを利用している。キーワード生成において金融コラムの単語頻出数をカウントするためにText Mining Studioを利用している。

(注6) 「キーB」については、ネガティブ・キーワードとポジティブ・キーワードを付録に掲載する。

**Table 2** The number of keywords

	key A	key B
negative	951	109
positive	214	70

**Table 3** Statistics of negative (or positive) news

	keyword	mean	standard deviation	max	min
negative	key A (15-15)	214	89	764	78
	key B (15-15)	160	76	627	41
	key A (9-15)	134	36	349	48
	key B (9-15)	50	17	163	6
positive	key A (15-15)	108	29	311	47
	key B (15-15)	118	34	349	47
	key A (9-15)	103	28	240	41
	key B (9-15)	68	21	218	7

件と 70 件である。

### 3.2 金利変化とテキスト情報との関係

「キー A」および「キー B」のキーワードをもとに判定したネガティブ・ニュースおよびポジティブ・ニュースの日次件数の統計値を **Table 3** に示す。最左列の「negative」はネガティブ・キーワードによりネガティブ・ニュースと判別されたニュースの数であり、「positive」はポジティブ・キーワードによりポジティブ・ニュースと判別されたニュースの数である。表の左から 2 列目にある「(15-15)」の表記は、前日 15 時から当日 15 時までの期間のヘッドラインニュースを 1 日分とする場合の結果であり、「(9-15)」の表記は、当日 9 時から 15 時までの期間のヘッドラインニュースを 1 日分とする場合の結果である。

平均件数をみると、ネガティブ・ニュースにおいては、「キー A」で判別されたニュースが、214 個であり「キー B」で判別されたニュース (15-15) は 160 個となっており、「キー A」で判別されたニュースのほうが、「キー B」で判別されたニュースを上回っていることを確認できる。他方、ポジティブ・ニュースに関しては、「キー A」で判別されたニュースが、108 個であり「キー B」で判別されたニュース (15-15) は 118 個となっており、逆に、「キー B」で判別されたニュース件数のほうが上回っていることを確認できる。また、営業時間外にに関してみると、営業時間外に配信されたニュースを含む「key B (15-15)」に比べ、営業時間内に配信されたニュースのみの「key B (9-15)」は、ネガティブとポジティブ合わせると平均的には半分以下のニュース件数となっている。ネガティブと判別されたヘッドラインニュース件数のうち、「key A (15-15)」と「key B (15-15)」ともにネガティブと判別されたニュース件数の割合（一致率）は、「key A (15-15)」では 17% で、「key B (15-15)」では 37% である。ポジティブと判別されたヘッドラインニュースのうち、「key A (15-15)」と「key B (15-15)」ともにポジティブと判別されたものの割合（一致率）は、「key A (15-15)」では 42% で、「key B (15-15)」

**Table 4** Correlation coefficient of interest rate change and the ratio of negative (or positive) news

keyword	negative	positive
key A (15-15)	-0.03	0.06
key B (15-15)	-0.06	0.04
key A (9-15)	-0.21**	0.04
key B (9-15)	-0.27**	0.15**

\*\*: It is significant at 1% level.

では 33% である。半数以上が、「key A (15-15)」と「key B (15-15)」において、それぞれ独自に判定されていることが確認できる。

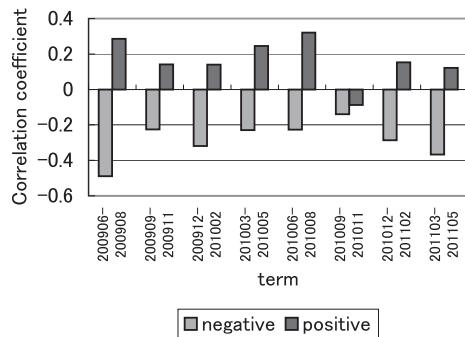
「キー A」と「キー B」のキーワードで判定したネガティブ（またはポジティブ）・ニュースの割合と金利変化との相関係数を **Table 4** に示す。金利変化を計算するための金利としては、金利の水準を示す代表的な指標である 15 時における残存年数 10 年の国債パレートを採用し、前日 15 時から当日 15 時までの国債パレート変化を金利変化とした。この金利変化とネガティブ・ニュース（またはポジティブ・ニュース）の件数割合との相関係数を計算した。ヘッドラインニュース数は日によって変化することから<sup>(注 7)</sup>、ネガティブ（またはポジティブ）ニュースの件数そのままの値ではなく、1 日のヘッドラインニュース全件数でネガティブ・ニュースおよびポジティブ・ニュースの件数をそれぞれ除した割合にて分析を行なっている。

**Table 4** 最左列の「(15-15)」の表記は、前日 15 時から当日 15 時までの期間のヘッドラインニュースを 1 日分とする場合であり、「(9-15)」の表記は、当日 9 時から 15 時までの期間のヘッドラインニュースを 1 日分とする場合を表わしている。真ん中の列は、ネガティブ・ニュースと金利変化の相関、右の列はポジティブ・ニュースと金利変化の相関を示している。

**Table 4** より、「キー A」をキーワードとしてネガティブ、ポジティブ・ニュースを判定した場合の相関はそれぞれ、-0.03, 0.06 であり、金利変化との関連はみられないことが確認できる。「キー B」をキーワードとしてネガティブ、ポジティブ・ニュースを判定した場合でも、債券運用者の営業時間外のニュースを含む場合では、相関係数はそれぞれ、-0.06 と 0.04 であり、金利変化との関連はみられないことを確認できる。しかし、「キー B」の場合で、ヘッドラインニュースを営業時間内 (9-15) に限定すると、ネガティブ・ニュースおよびポジティブ・ニュースとの相関係数はそれぞれ、-0.27 および 0.15 であり、金利変化との相関が高いことを確認できる<sup>(注 8)</sup>。統計的にも景況感や株式市場の動向は、金利変化に影響を与える主要な要因のひとつとして挙げられるが、一般に、景況感や株式動向と金利変化は、逆相関となる傾向にあることから、ネガティブ・ニュースと金利変化は負の相関係

(注 7) たとえば、Fig. 1 などからもニュースの数が時期により変動することを確認できる。

(注 8) ポジティブ・ニュースおよびネガティブ・ニュースのいずれにおいても、1% の水準で統計的に有意な差が認められる。



**Fig. 2** Correlation coefficient of interest rate change and negative (or positive) news ratio

数、ポジティブ・ニュースと金利変化は正の相関係数になっていると考えられる。また、ヘッドラインニュースを営業時間内(9-15)に限定した場合の「キー A」においては、ポジティブ・ニュースと金利変化との関連性が確認できなかったが、「キー A」のキーワードは、財務諸表に使用されているテキストから抽出されたものであり、マーケットデータである金利変化の分析には適していない可能性がある<sup>(注 9)</sup>。その意味で、当分析結果は、キーワードリスト構築の重要性を指摘するものであり、その意味で意義の大きい分析結果である。

**Fig. 2** は、相関関係の安定性をみるために、3ヶ月ごとに、「キー B」のキーワードで判定した営業時間内のネガティブ(またはポジティブ)・ニュースの割合と金利変化との相関係数を計算した結果である。Fig. 2 より、全体的には、ネガティブ・ニュース(ポジティブ・ニュース)と金利変化の相関はマイナス(プラス)の傾向を確認はできるものの、厳しくみると、両者の関係は安定しておらず、時期によって変わっていることが確認できる<sup>(注 10)</sup>。本分析では、ネガティブ、ポジティブを判定するキーワードを固定としているが、判定するキーワードを変えると良い分析結果が得られる可能性がある。これら詳細な分析は今後の課題としたい。

本節では、ニュースと金利変化の関連性に焦点をあてた分析を行ない、金融市場と密接な関連をもつキーワード(「キー B(9-15)」)において、統計的に有意な関連性のあることを見出した。次節において、これら関連性に基づく投資戦略の有効性について分析を行なう。

### 3.3 長期国債先物によるシミュレーション

「キー B」をキーワードとするネガティブ(またはポジティブ)・ニュースの件数情報と金利変化に相関関係がみられることから、ニュース件数情報を用いたシミュレーションを実施した。シミュレーションの運用対象資産は、日本の長期国債先物とした。運用戦略は、前日の9:00から当日の9:00まで<sup>(注 11)</sup>のニュース件数情報を用いて当日の9:00から

(注 9) これらの結果は、Loughran and McDonald<sup>17)</sup>の指摘と整合的な結果である。

(注 10) たとえば、Fig. 2 中の 2010/09-2010/11 期におけるポジティブ・ニュースの相関係数はマイナスとなっている。

(注 11) 平日は 1 日分のニュース件数情報となるが、休日をまた

**Table 5** Correlation coefficient of JGB future return and the ratio of negative (or positive) news

Strategy	term	negative	positive
Strategy 1	2009/06 - 2010/05	0.12	-0.04
	2010/06 - 2011/05	0.08	-0.08
	2009/06 - 2011/05	0.10	-0.06
Strategy 2	2009/06 - 2010/05	-0.03	0.08
	2010/06 - 2011/05	0.07	-0.23
	2009/06 - 2011/05	0.03	-0.09

18:00までの長期国債先物リターンを予測する戦略(Strategy 1)と、当日9:00から当日11:00までのニュース件数情報を用いて当日12:30から当日18:00までの長期国債先物リターンを予測する戦略(Strategy 2)<sup>(注 12)</sup>の2種類の戦略について有効性の検証を行なった<sup>(注 13)</sup>。

はじめに、ポジティブ(ネガティブ)・ニュースと長期国債先物との関連性について分析を行なった。Table 5 は、ネガティブ・ニュースとポジティブ・ニュースの1日の件数に対する割合と、長期国債先物リターンとの相関<sup>(注 14)</sup>を示したものである。Strategy 2 の 2009/06 から 2010/05 の期間以外では、長期国債先物リターンとネガティブ・ニュースの件数割合は正の相関、ポジティブ・ニュースの件数割合とは負の相関であることが確認できる。

これらの結果より、簡単な投資戦略としてネガティブ・ニュースの件数割合がある閾値より高く、ポジティブ・ニュースの件数割合がある閾値より低い場合に、長期国債先物の買いポジション(Long)をとり、逆に、ネガティブ・ニュースの件数割合がある閾値より低く、ポジティブ・ニュースの件数割合がある閾値より高い場合に、長期国債先物の売りポジション(Short)をとるとの戦略が考えられる。本分析では、当投資戦略のパフォーマンスを計測し、投資戦略の有効性について分析を行なう。上記投資戦略に基づき、長期国債先物のポジション<sup>(注 15)</sup>を日次でリバランスすることにより投資戦略

ぐ場合は複数日分のニュース件数情報となる。

(注 12) 当日の午前のニュース情報により、午後の長期国債先物の予測をするということ。なお、ポジション形成のための計算には時間がかかると仮定しているが、11:00から12:30まで時間があるので実現可能な想定である。

(注 13) 本研究は、絶対収益型投資商品などで利用される、リスクフリーレートに対する超過収益の源泉となる一つの情報としての有効性を分析するものである。従来の、数値情報分析を加味する分析については今後の課題としたい。

(注 14) Strategy 1 では、前日の9:00から当日の9:00までのネガティブ・ニュース(またはポジティブ・ニュース)件数割合と、当日の9:00から18:00までの長期国債先物リターンの相関を計算した。Strategy 2 では、当日9:00から当日11:00までのネガティブ・ニュース(またはポジティブ・ニュース)件数割合と、当日12:30から当日18:00までの長期国債先物リターンの相関を計算した。

(注 15) ポジションは、実現可能な方法として期近限月の長国物をレバレッジはかけないで1投資単位ロング(またはショート)し、Strategy 1 では、当日の9:00から18:00までの期間保有し反対売買して決済する。Strategy 2 では、当日12:30から当日18:00までの期間保有し反対売買して決済する。

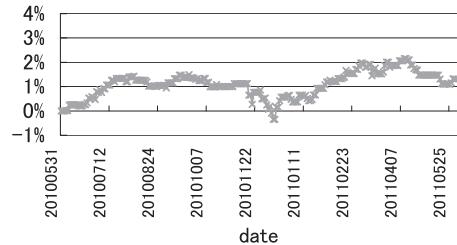


Fig. 3 Transition of cumulative return (Strategy 1).

Table 6 Win-loss standings (Strategy 1)

	Win	Loss	Total
Long	32	23	55
(ratio)	13%	9%	22%
Short	29	34	63
(ratio)	12%	14%	26%
Long + Short	61	57	118
(ratio)	25%	23%	48%
Neutral			126
(ratio)	—	—	52%
Total			244
(ratio)	—	—	100%

の有効性の検証を行なうものとした。なお、戦略において必要となる閾値としては、過去のネガティブ・ニュースとポジティブ・ニュースそれぞれの件数割合の平均値<sup>(注16)</sup>を採用するものとした。なお、上記に示した条件を満たした場合には、Long または Short のポジションをとるが、条件を満たさない場合は Long (または Short) ポジションはとらない (Neutral) ものとする。

Fig. 3 は、Strategy 1 の戦略を 2010/06 から 2011/05 の 1 年間日次で実施したときの累積リターン<sup>(注17)</sup>の推移を示したものである。1 年間の運用において 1% 程度の累積超過リターン<sup>(注18)</sup>が獲得できることを確認できる。2010/12 頃、一時的に収益率が低下してきているが、期間を通じ、おおむね累積超過リターンは正であり、ネガティブとポジティブのニュース件数割合の情報が、長期国債先物リターンの予測に利用できる可能性がある。より精緻な分析は今後の課題としたい。

Table 6 は、Strategy 1 のシミュレーションにおける Long

(注 16) 具体的には、閾値はネガティブ・ニュースとポジティブ・ニュースそれぞれの 2009/06 から 2010/05 までの 1 年間の日次の件数割合の平均とした。

(注 17) ここでの累積リターンは、リスクフリーレート (無担保コールレート) に対する超過リターンである。長国先物の取引手数料は、0.0003% 程度 (マーケットインパクト無視) であるので、影響はほとんどないものとみなせる。

(注 18) キーワード全体から重複なく 1 割ずつのキーワードを除いて 10 回のシミュレーションを行なうクロスバリデーションを実施したところ、累積リターンの平均は 0.40%、標準偏差は 0.86% であった。すべてのキーワードを使用した場合の累積リターンが 1.31% であり特に使用しているキーワードに大きな偏りはないことを確認している。

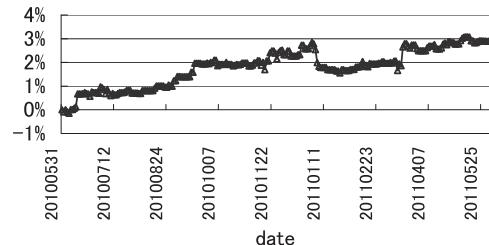


Fig. 4 Transition of cumulative return (Strategy 2).

ポジションと Short ポジションの予測の勝敗を示したものである。表は上段からそれぞれ、買いポジション (Long)、売りポジション (Short)、買いポジションと売りポジションの合計、売りも買いもしないポジション (Neutral) の値を示している。表の左から予測が正解だった場合 (Win)、不正解であった場合 (Loss)、両者の合計 (Total) の数を示している。なお、表には、件数 (上段) および全体に占める割合 (下段) がそれぞれ記してある。表より、たとえば、買いポジション (Long) に関しては、その期間の予測が正しかった (長期国債先物リターンが正であった) 日が 32 日あり、その割合が 13% であることを確認できる。一方、予想が外れた (長期国債先物リターンが負であった) 日が、23 日あり、割合では 9% であった。また、1 年間の 244 営業日のうち、ポジションをとらない Neutral な日が 126 日で割合として 52% であり、1 年間のうち半分程度ポジションをとっていたことがわかる。全体でみると (Long + Short)、予測が正しい数のほうが 2% 多くなっている。正の累積リターン獲得に貢献していることを確認できる<sup>(注19)</sup>。

Fig. 4 は、Strategy 2 の戦略を 2010/06 から 2011/05 の 1 年間日次で実施したときの累積リターンの推移を示したものである。図より、1 年間で 3% 程度のリターン<sup>(注20)</sup>を獲得していることを確認できる。Strategy 2 に関しては、分析期間である 1 年間を通じて、累積超過リターンは正であり、Strategy 1 と比較して頑健な結果が得られていることを確認できる<sup>(注21)</sup>。これらの結果は、ネガティブおよびポジティブ・ニュースに関する情報が、長期国債先物リターンの予測に有効であることを示すものであり、実務的観点からも興味深い結果である。ただし、金利が大きく変動する場合に効果があるかなど、1 年間だけではなく他の期間のデータを使ったシミュレーションが必要と考えられるが、それは今後の課題としたい。

(注 19) Long の場合の予想の勝敗の差は、割合では 4% であるのに対し、Short の場合の予想の勝敗の差は -2% (= 12 - 14) である。Long の正解の割合が多かったことから全体的に予測が正しい数のほうが多くなっている。

(注 20) キーワード全体から重複なく 1 割ずつのキーワードを除いて 10 回のシミュレーションを行なうクロスバリデーションを実施したところ、累積超過リターンの平均は 3.53%、標準偏差は 0.72% であることを確認している。すべてのキーワードを使用した場合の累積リターンが 2.90% であることから、本分析において使用しているキーワードに特に大きな偏りはないと考えられる。

(注 21) これらの結果は、Table 4 の結果とも整合的である。

**Table 7** Win-loss standings (Strategy 2)

	Win	Loss	Total
Long	44	44	88
(ratio)	18%	18%	36%
Short	35	28	63
(ratio)	14%	12%	26%
Long + Short	79	72	151
(ratio)	32%	30%	62%
Neutral		93	
(ratio)	—	38%	
Total		244	
(ratio)	—	100%	

**Table 7** は、Strategy 2 のシミュレーションにおける Long ポジションと Short ポジションの予測の正確さを示したものである。表より、1 年間の 244 営業日のうち、ポジションをとらない Neutral な日が 93 日で割合として 38% であり、1 年間のうち 6 割程度ポジションをとっていたことがわかる。Short ポジションに関してみると、予測が正しかった（長期国債先物リターンが負であった）日は、35 日（割合でみると 14%）あり、予想が正しくなかった（長期国債先物リターンが正であった）日は、28 日（割合では 12%）あることを確認できる。これより Short ポジションの場合の予想の正解と不正解の比率の差は、割合では 2% であることを確認できる。一方、Long ポジションに関してみると、予想の正解と不正解の比率の差は、0 % (= 18 – 18) であり、予測が正しかった場合と正しくなかった場合の比率は同じであった。このように、Short ポジションにおいて、予測が正しかった割合が多いことより、Strategy 2 でも 1 年間の累積リターンが正となる結果になっていることが確認できる。

Strategy 1 および Strategy 2 の比較をしてみると、Strategy 1 は、前日のニュース件数情報により当日の長期国債先物リターンを予測する戦略であるのに対し、Strategy 2 は、当日午前のニュース件数情報により当日の午後の長期国債先物リターンを予測する戦略であり、より新しい情報を投資戦略に生かした手法となっている。利用するニュース件数情報の時間帯と、予測する長期国債先物リターンの組み合わせを工夫したり、ポジションをとるための閾値を工夫することで予測の精度を向上できる可能性があるが、それらについては、今後の課題としたい。また、一定数の正解データを作成し、機械学習による分類を考慮することで精度向上を図ることも今後の課題である。

#### 4. まとめ

本稿では、テキスト情報分析を通じ、ヘッドラインニュースと債券市場の関連性について分析を行なった。分析の結果、(1) ヘッドラインニュースと金利変化の間に統計的に有意な関連がみられること、(2) ニュースを用いた投資戦略により正の超過収益を獲得できる可能性のあることを見出した。ま

た、分析を通じ、(3) 分析の際に用いるキーワードのリストが分類精度に大きな影響を与えることも確認した。これらの結果は、債券運用における超過収益の源泉に關し興味深い結果を示すものであり、学術的および資産運用の実務的観点からも意義の大きい結果である。

本稿においては、テキストの情報を 1 日単位にて集計し金利変化との関連性について分析を行なったが、時間の分割を 1 時間ごとに細分化した分析など、より詳細な分析については、今後の課題として挙げられる。

#### 参考文献

- 1) R.C. Merton: On The Pricing of Corporate Debt: The Structure of Interest Rates, *The Journal of Finance*, **29**-2, 449/470 (1974)
- 2) F.J. Fabozzi: *The Handbook of Fixed Income Securities*, McGraw-Hill (2005)
- 3) 松村、三浦：人文・社会科学のためのテキストマイニング、38、誠信書房 (2009)
- 4) S. Takahashi, M. Takahashi, H. Takahashi and K. Tsuda: Analysis of the Validity of Textual Data in Stock Market through Text Mining, *WSEAS Transactions on Business and Economics*, **3**-4, 310/315 (2006)
- 5) S. Takahashi, H. Takahashi and K. Tsuda: Analysis of the effect of Headline News in financial market through text categorisation, *Int. J. Comput. Appl. Technol.*, **35**-2, 204/209 (2009)
- 6) P.C. Tetlock: Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance*, **62**-3, 1139/1168 (2007)
- 7) P.C. Tetlock, M. Saar-Tsechansky and S. Macskassy: More Than Words: Quantifying Language to Measure Firms' Fundamentals, *The Journal of Finance*, **63**-3, 1437/1467 (2008)
- 8) F. Li: Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?, University of Michigan Working Paper (2006)
- 9) F.J. Meschke: CEO interviews on CNBC, working paper, Arizona State University (2004)
- 10) J. Bollen, H. Mao and X. Zeng: Twitter mood predicts the stock market, *Journal of Computational Science*, **2**-1, 1/8 (2011)
- 11) W. Antweiler and M.Z. Frank: Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *The Journal of Finance*, **59**-3, 1259/1294 (2004)
- 12) 内田、後藤：債権放棄の経済的機能：イベント・スタディによる実証分析、*日本経済研究*, 45, 128/146 (2002)
- 13) 秋吉、広瀬：銀行のエクスボージャーと債権放棄における企業銀行間交渉：イベント・スタディによる検証、RIETI ディスカッション・ペーパー 06-J-037, *経済産業研究所* (2006)
- 14) 小川、渡部：株価データと新聞記事からのマイニング、*情報処理学会研究報告*, FI, *情報学基礎*, **2001**-20, 137/144 (2001)
- 15) 高橋、高橋、津田：株式市場におけるヘッドラインニュースの効果についての研究、*ファイナンス学会第 15 回大会予稿集*, 374/383 (2007)
- 16) 上瀧、高橋、高橋：クレジット市場におけるヘッドラインニュースの効果、*ファイナンス学会第 17 回大会予稿集*, 113/122 (2009)
- 17) T. Loughran and B. McDonald: When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *The Journal of Finance*, **66**-1, 35/65 (2011)

**Table A.1** Positive keywords (key B)

安心感	急騰	上値
安定	急反発	進む
安定化	強い	進行
安定操作	強まる	人気
一段高	業績	全面高
印象づける	景気-> づく	続伸
印象深い	堅調	大幅高
印象的	効果	大幅続伸
影響力	好感	大幅反発
越える	高い	値上がり
越す	高値	値上がり銘柄数
円高	高値引け	超える
押し上げ	高値更新	追い風
押し上げる	刺激的	底堅い
押し目買い	小反発	買い
欧米株安	小幅続伸	買い戻し
回復	小幅反発	買い戻す
改善	上げ	買う
拡大	上げる	反発
期待感	上げ幅拡大	米株安
急激-> づく	上回る	膨らむ
急伸	上昇	戻す

**Table A.2** Negative keywords (key B)

さえる+ない	急	小幅反落
とどまる	急激	消える
ドル高	急落	伸び悩む
悪	救済措置	清算
悪い	傾く	全面安
悪意	景気後退	続く+ない
悪意-> 持つ	経営-> 失敗	続落
悪化	警戒	損失
悪事	警戒感	大幅安
悪評	欠く	大幅続落
安い	欠ける	大幅反落
安値	嫌	値下がり
安値更新	嫌気	値下がり銘柄
一段安	懸念	値下がり銘柄数
一服	見える+ない	値下げ
引っ張る	見送る	遅れ
越える	減少	遅れる
円安	効果+ない	低水準
延期	効果+ない-> 性質	低調
押し下げる	高値警戒感	軟調
下がる	支える+ない	売り
下げ	止める	売る
下げる	資金+ない	売る+できない
下げ幅拡大	資源開発	判決-> 下す
下回る	失業	判断-> 誤る
下方	失望	反落
下落	弱まる	否定
過度	取り戻す+できない	否定的
回収+できない	需給悪化懸念	不安
回収性+できない	重荷	不透明感
回復+しにくい	重要性	米国株高
回復+できない	縮小	崩す
拡大	小さい	乏しい
格下げ	小型化	利益確定
割り込み	小反落	利益確定売り
緩慢	小幅続落	

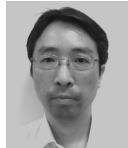
## 《付 錄》

**A. キーワードリスト：「キー B」**

本章では、分析に用いたキーワードリストを示す。Table A.1 および Table A.2 は、それぞれ、本文中において分析に用いた「キー B」のネガティブ・キーワードとポジティブ・キーワードを示したものである<sup>(注 22)</sup>。表中の「+」は、単語と単語をつないで一語扱いにする記号であり、「->」は、係り受けの関係を記述するための記号である<sup>(注 23)</sup>。

## [著者紹介]

## 山下泰央



1991 年東京大学工学部卒業、同年三井信託銀行(株)入社、現在、三井住友信託銀行在籍、2001 年筑波大学大学院修士課程修了、2011 年東京工業大学大学院博士課程修了、日本証券アナリスト協会検定会員、博士(工学)。

## 上瀧弘晃



1997 年慶應義塾大学理工学部卒業、同年三井信託銀行(株)入社、現在、三井住友信託銀行在籍、2004 年一橋大学大学院国際企業戦略研究科修了、日本証券アナリスト協会検定会員。

## 高橋大志(正会員)



1994 年東京大学工学部卒業、94~97 年富士フイルム(株)研究員、97~2005 年三井信託銀行(当時)シニアリサーチャー、2002 年筑波大学大学院修士課程修了、2004 年同大学院博士課程修了、2005~2008 年岡山大学准教授、2007 年キール大学客員研究員、2008 年より慶應義塾大学経営管理研究科准教授、博士(経営学)。

(注 22) キーワードの中には、債券価格を上昇させる要因と下落させる要因が混在しているが、そうした要因の混在を解消した分析は今後の課題である。

(注 23) 表中のキーワードの中にみられる「+」や「->」といった記号は、Text Mining Studio の記法を準用したものである。