

クレジット市場におけるヘッドラインニュースの効果

Analyzing the influence of Headline News on Credit Markets

上瀧弘晃^{*1}
Hiroaki Jotaki

高橋悟^{*2}
Satoru Takahashi

高橋大志^{*3}
Hiroshi Takahashi

^{*1,2} 中央三井アセット信託銀行 ^{*3} 慶應義塾大学大学院経営管理研究科
Chuo Mitsui Asset and Banking Company, Limited Graduate School of Business Administration, Keio University

Abstract: This paper analyses the influence text information gives to the credit market, focusing on Headline News, a source of information that has immediate influence to the money market, and also which is regarded as one of the important information when making investment decisions. Text automatic classification algorithm has been applied for analysis. After classifying the headline news into several types using this algorithm, verification were made whether there were any characteristic changes in the credit market. As a result, (1) It is possible to build a headline news algorithm by an accuracy of 80% using the automatic classification algorithm. (2)The information extracted from the headline news gives influence to the bond market, similar to the equity market. (3)Difference could be observed in the market before and after the headline news was broadcasted.

1. はじめに

金融市場においては、株式や債券などをはじめ、数多くの金融資産の取引が行われている。その中でも、債券市場は時価総額が 560 兆円にのぼるなど株式市場と並び主要な金融市場の一つである。債券市場は国債が中心であるが、信用リスクを伴う社債の時価総額は 80 兆円と年々規模を拡大しており、また CDS 市場は近年、グローバルで拡大している。こうしたクレジット(社債, CDS)取引を行うためには、企業の会計情報など、数多くの情報を用いて企業のクレジットを適切に評価することが求められる。

クレジットの評価については、これまで学術分野および実務の分野において、膨大な量の分析が行われており、数多くの有用な分析結果が報告されてきた[2][3]。しかし、これまで数多くの分析が行われているものの、分析に用いられてきた情報の多くは、会計データなどの数値化された情報であり、数値情報以外のデータについて取り扱った分析は、十分な数があるとはいえない。その理由の一つとして、数値情報以外のデータは、データの取り扱いが難しいことがあげられる。

資産運用の実務においては、機関投資家は、ニュースなどの数多くのテキスト情報も利用し、投資判断を行っている。加えて、このようなテキスト情報は、数値情報とは異なる情報を伝えるとの報告も株式市場を中心に行われており[1][5]、例えば、高橋／高橋／津田[1]は、ファンドマネージャーの投資判断における重要な情報源の一つであるヘッドラインニュースに焦点をあて、株式価格に与える影響について分析を試みており、ヘッドラインニュースから抽出した情報は株式価格の形成に影響を与えているとの結論を得ている。このようにテキスト情報を用いたこれまでの研究の多くは、株式市場に焦点を当てたものであるが、近年重要性が増している債券市場、とりわけ評価が難しいクレジット市場において、数値情報以外の情報が与える影響を明らかにすることは、学術的および実務的観点からも意義が大きい。本稿では、このような議論を背景に、代表的なテキスト情報の一つであるヘッドラインニュースに焦点を当て、ヘッドラインニュースがクレジット市場に与える影響について分析した。

2. 分析データ

2.1 ヘッドラインニュースについて

本稿では、ヘッドラインニュースに対するクレジット市場の反応について分析を行う。ヘッドラインニュースとは、クイックやロイターといった情報端末によって、タイムリーに提供される情報であり、ファンドマネージャーにとって最も即時性のある情報源のひとつである(以下ヘッドラインニュースをニュースと表現)。分析に用いるニュースデータは、時事通信社から提供されるニュースを用いるものとした。

ニュースには市場全体に影響を与える情報や、個別企業の業績に関する情報などの数多くの情報が含まれている。本稿では、これらのニュースの中から、「企業不祥事」、「増資等の資本政策」や「経営者の発言とそれに対するコメント」といった、個別企業の信用リスクに影響を与える可能性の高いニュースを抽出し、そのニュースが発信された前後に、ニュースの対象となった企業の社債のスプレッド、及び CDS のスプレッドがどのように変動したかを分析する。

ニュースを収集した期間は、2006 年 8 月 10 日から 2006 年 11 月 24 日の約 4 ヶ月半であり、この期間を分析対象期間とした。期間中に時事通信社から発信されたニュースの総数は、346,975 件と膨大な数に上る。高橋／高橋／津田[1]は、分析対象企業を、分析期間中に東証 1 部に上場していた企業とした。本稿では、同期間中に東証 1 部に上場していた企業のうち、社債を発行している企業、185 社を対象とした。これらの企業を対象とし、ニュースの統計量を計測した。企業関連ニュースのうち、当該企業を対象としたニュースは、6,172 件あり、全ニュースの約 1.8%を占めていた。また、1 日あたりのニュースの発信数で見ると、平均で 83 件発信されている。

もう一つのクレジット市場である CDS 市場は市場が発達したのが最近であり、分析対象期間にデータがある企業数は社債より少ない。分析対象とした社債を発行している企業のうち、CDS が取引されている企業は 152 社、また当該企業を対象としたニュースは、4,597 件あった。

2.2 市場データについて

社債スプレッドは社債の複利利回りから、対応する残存年限の国債の複利利回りを引いて算定した。時価はインデックス(NOMURA-BPI)を構成する銘柄のBPI時価(終値)を用いた。銘柄は最も発行銘柄の多い、償還までの残存年数が3年以上5年未満の社債を対象とし、同期間に複数銘柄が存在する場合は最も残存年数の長い銘柄とした。この条件をもとに、1つの発行体に対し、1つの社債を選択した。

ニュースの効果を分析するために、各ニュースが発信された時点を基準とし、その前後のスプレッドの変化を計測した。図2.1で示したように、ニュースにはタイムスタンプが刻まれている。実際の投資行動は、ニュースを受信した後に行うので、タイムスタンプを利用して、社債スプレッドの変化を計測する基準時点を設定した。具体的には、ニュースを受信した日が営業日であり、かつ受信時間が午前中の場合は、その日から投資を行えるものと仮定し、ニュース受信日の時価を基準として社債スプレッドの変化を計測した。また、それ以降の時間に受信したニュース、および休日に受信したニュースは、翌営業日以降に投資を行うものとして、翌営業日の時価を基準として社債スプレッドの変化を計測した。

社債スプレッドは日次で計測し、計測期間は、ニュースを受信して投資行動をする基準日から、前後30営業日とした。また、市場の変動による影響を排除するために、全銘柄のスプレッドを単純平均し、その平均スプレッドからの超過スプレッド(以下社債超過スプレッド)を分析対象とした。

CDSスプレッドは、Itraxx Japanが提示している日本国内のクレジット・デリバティブ市場の動向を示す代表的なクレジット・デフォルト・スワップ指数を用いた。残存年数は主要な取引年数である5年を用いた。

社債の場合と同様の手順で、CDSスプレッドの変化を計測した。市場の変動による影響を排除するために、全銘柄のスプレッドを単純平均し、その平均スプレッドからの超過スプレッド(以下CDS超過スプレッド)を分析対象とした。

3. 分析方法

本稿では、ニュースが発信されたことによって、その情報がクレジット市場に影響を与えるかどうかを分析する。ニュースの内容によっては、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があると考えられる。そこで、本稿では、最初にニュースを3つのタイプに分類し、次にそれぞれのタイプのニュースが、市場にどのような影響を与えるか分析を行う。

3.1 ナイーブベイズによるテキスト分類アルゴリズム

ニュースの分類方法として、テキストマイニングで使用される「ナイーブベイズ」を用いる。ナイーブベイズは、テキスト分類においては「決定木」、「最近隣法」や「ニューラルネット」と並んで、幅広く使われているアルゴリズムのひとつである[4][6]。ナイーブベイズは、分類結果を直感的に理解しやすく、かつ分類精度が高いアルゴリズムである。さらに、クラス分けされているデータとされていないデータが混在する場合、理論的な拡張が容易であり、本稿で取り扱うような、曖昧な分類結果を含むニュースの分類に適している。

以下で、ナイーブベイズの分類方法について説明する。まず、テキストデータの特徴付けるキーワードが、 n 個あると仮定する。このキーワードを使い、テキストデータをベクトル表現を用いて、と表現する。はテキストデータの中に、 i 番目のキーワードが存在するかどうかをあらわしたもので、単純に(0,1)で表現される場

合と、TF-IDFといったキーワードの重要度をあらわすウエイトで表現する方法がある。分類するテキストのクラスを c とすると、テキストがどのクラスに属するかは、以下の事後確率を最大化するクラスを求めれば、もっとも精度の高い分類を得ることができる。

$$\hat{c} = \underset{c}{\operatorname{argmax}} (P(c|\mathbf{x})) = \underset{c}{\operatorname{argmax}} \left(\frac{P(\mathbf{x}|c)P(c)}{P(\mathbf{x})} \right) = \underset{c}{\operatorname{argmax}} (P(c)P(\mathbf{x}|c))$$

… (1)

式(1)は、ベイズの定理および、がすべてのクラスについて一定であることを展開している。式(1)から、テキストが与えられたときに、各クラスの出現確率と、各クラスにおけるテキストベクトル \mathbf{x} の出現確率の積がもっとも大きくなるクラスが、そのテキストが分類されるクラスとなる。ここで、クラス c が与えられたときに、テキストベクトル \mathbf{x} の各要素が独立であると仮定すると、は以下のように展開される。

$$P(\mathbf{x}|c) = P(x_1, x_2, \dots, x_n | c) \approx \prod_{i=1}^n P(x_i | c)$$

… (2)

式(1)、式(2)より、テキストの属するクラスは、

$$\hat{c} = \underset{c}{\operatorname{argmax}} (P(c|\mathbf{x})) = \underset{c}{\operatorname{argmax}} (P(c) \prod_{i=1}^n P(x_i | c))$$

… (3)

と表現される。従って、各クラスの出現確率と、キーワードが各クラスにおいて出現する確率を求めることにより、テキストデータの属するクラスを求めることができる。ここで、とは、学習データを選び、各クラスの相対頻度、および各クラスにおけるキーワードの相対頻度から算出する。上記の式の展開において、テキスト内の単語の出現が互いに独立であるという仮定が、素朴(naïve)であることが、ナイーブベイズの名前の由来である。しかし、クラスを特徴付ける複数のキーワードの有無が分類に影響を与えるため、直感的な理解がしやすい。さらに、直感的な理解が困難であるニューラルネットと同等の分類精度を持っている。

次に、学習データから構築したアルゴリズムの分類精度について、テストデータを用いて検証を行う。本稿では、クラスとして、「Good News」、「Bad News」と「Neutral News」の3パターンを設定した。また、ナイーブベイズによる分類アルゴリズムを学習するために、企業に関連するニュース(13,402件)から、学習データとテストデータとして、それぞれ分析対象データの約3%に当たる400件をランダムに取り出した。

4. 分析結果

4.1 ナイーブベイズによるテキスト分類アルゴリズム

まず学習データとして、ランダムに400のニュースを抽出した。抽出したニュースに対して、Good News, Bad News と Neutral News の教師ラベルを与えるために、実際に人手によって内容を読み、カテゴリー分けを行った。分類結果を表4.1に示す。学習データのうち、Good Newsが49%と半分近くを占めていることがわかる。

表4.1 学習データのニュース分類

ニュースのタイプ	Good	Bad	Neutral
ニュースの個数	185	90	125
割合	46.25%	22.50%	31.25%

学習データに対して、形態素解析とパターンマッチングおよび、実際に人が読むことによりキーワードを抽出した。最初に抽

出したキーワードは 1,360 個であった。そこから, [4], [6]における分析方法を参考とし, キーワードの出現頻度および, 3.1 節における, 学習データによるインサンプルの分類結果から, 205 個までキーワードを絞り込んだ。学習データのインサンプルによる分類結果を, 表 4.2 に示す。インサンプルの分析結果ではあるが, 約 80% の分類精度を得ることができた。

表 4.2 学習データでのインサンプルの分類結果

ニュースのタイプ	Good	Bad	Neutral
ニュースの個数	185	90	125
割合	46.25%	22.50%	31.25%

次に, テストデータとしてランダムに 400 個のニュースを抽出する。ナイーブベイズアルゴリズムの分類精度を検証するためにこれらのテストデータには, 教師ラベルを与える必要がある。そこで, 各ニュースを手で読むことにより, Good News, Bad News と Neutral News にカテゴリ分けを行った。分類結果を表 4.3 に示す。テストデータについても, Good News が 46% ともっとも多く割合を示している。

表 4.3 テストデータのニュース分類

		実際のタイプ					実際のタイプ		
		Good	Bad	Neutral			Good	Bad	Neutral
ベイズ 学習 結果	Good	179	24	24	ベイズ 学習 結果	Good	90.86%	24.74%	22.64%
	Bad	2	66	9		Bad	1.02%	68.04%	8.49%
	Neutral	16	7	73		Neutral	8.12%	7.22%	68.87%

テストデータに対して, 学習データを用いて作成したナイーブベイズのアルゴリズムを適用して自動分類を行った。分類結果を表 4.4 に示す。表 4.4 から約 78% の分類精度を得られたことがわかる。

表 4.4 テストデータでのアウトオブサンプルの分類結果

		実際のタイプ					実際のタイプ		
		Good	Bad	Neutral			Good	Bad	Neutral
ベイズ 学習 結果	Good	168	27	35	ベイズ 学習 結果	Good	90.81%	30.00%	28.00%
	Bad	3	58	5		Bad	1.62%	64.44%	4.00%
	Neutral	14	5	85		Neutral	7.57%	5.56%	68.00%

さらに, このアルゴリズムを使い, 企業に関連するニュース (13,402 件) に対して自動分類を行った。結果を表 4.5 に示す。

表 4.5 企業に関連する全ニュースにおけるテキスト分類結果

ニュースのタイプ	Good	Bad	Neutral
ニュースの個数	10098	2418	4463
割合	59.47%	14.24%	26.29%

表 4.5 から, Good News が全体の約 59% を占めており, 学習データ, テストデータより割合が若干高くなっていることがわかる。これは, 表 4.4 から読み取れるように, 作成したテキスト分類アルゴリズムは, 全体では約 80% 弱の分類精度を持つが, 実際のラベルは Bad News であったニュースを, Good News と判別してしまう傾向があることから, 全データの分類結果が, Good News に若干偏ってしまうためと考えられる。さらに高い精度を持つ分類ルールを作成することが, 今後の取り組みとして挙げられる。

4.2 社債を用いた分析

ニュースの対象となっている各企業について, ニュースの発信日前後における社債超過スプレッドを計測する。次に, 4.1 節で分類したニュースのタイプ毎に, 社債超過スプレッドの平均値を求め, Good News, Bad News および Neutral News において, ニュースが発信された前後の社債超過スプレッド変化に有意な差が存在するかどうか, 仮説検定を行う。検定の結果, 社債超

過スプレッドの大きさは, Good News < Neutral News < Bad News の順であり, その間には有意な差が存在するとの結果を得た。本節では, 特に Good News と Bad News に着目して分析を行う。表 4.6 に, Good News と Bad News の社債超過スプレッド格差について, Welch 検定により分析した結果を示す。

表 4.6 全データに対する Welch 検定の結果

	Sprd-30	Sprd-20	Sprd-10	Sprd+10	Sprd+20	Sprd+30	
平均スプレッド	Good News	-0.057%	-0.056%	-0.051%	-0.071%	-0.117%	-0.159%
	Bad News	0.158%	0.154%	0.156%	0.056%	0.201%	0.280%
	Diff	-0.214%	-0.210%	-0.206%	-0.127%	-0.317%	-0.440%
Welch検定	T値	-19.91	-14.39	-7.55	-5.06	-7.32	-10.88
	検定結果	差有り	差有り	差有り	差有り	差有り	差有り

表 4.6 から, Good News を公表した企業は, Bad News を公表した企業よりも社債超過スプレッドが縮小していることがわかる。基本的に, スプレッドの縮小は, 正のリターンが得られることを示し, スプレッドの拡大は, 負のリターンとなる。これは, ニュースを利用し, 社債超過スプレッドを獲得する可能性を示しており, ニュースに有効な価値が存在することを示唆している。またこの結果は, 株価と同様, 社債市場でもナイーブベイズ分類法が有効に機能していることを示している。

さらに表 4.6 からは, Good News の社債超過スプレッドが Bad News の社債超過スプレッドを下回る傾向が, ニュース発信の前後において共通していることがわかる。これは, ニュースが発信される前から, 市場がニュースを織り込んでいっていると考えることができる。この理由として, 2.1 節で示したように, 一日当たりに発信されるニュースは平均で 83 件もあるため, 同じようなニュースが連続して発生し, ニュースの効果が重なって評価されていることが理由として考えられる。アナリストレポートをテキストマイニングした分析においても, 本稿と同様の分析結果が確認されている [1]。

次に, より詳細な社債超過スプレッドの推移を見るために, 日次で社債超過スプレッドの推移を計測した。ニュースのタイプ毎に社債累積超過スプレッドの平均値を求め, グラフ化した (図 4.1)。

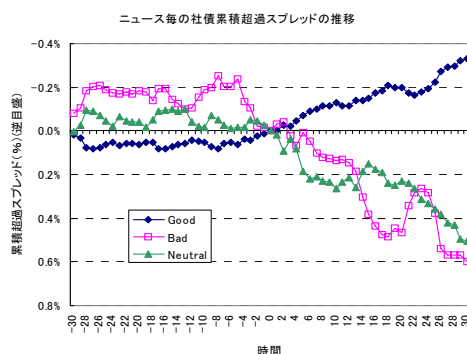


図 4.1 ニュース毎の社債累積超過スプレッド

図 4.1 から, 表 4.6 と同様に, 社債スプレッドの大きさは, Good News < Neutral News < Bad News の順に並んでいることがわかる。特にニュース発表前に比べて, ニュース発表後の変化が大きいことがわかる。これらから社債はニュース発信後に反応しやすく, 超過収益の源泉となる可能性があることを示している。

また Neutral News の社債超過スプレッドは Bad News と同じような動きをしていた。これは Neutral News が何らかの情報を含んでいる可能性がある。この点はナイーブベイズ分類法の改良が必要であり, 今後の課題である。

4.3 市場間のリターン分析

本節では、クレジット市場(社債, CDS)と株式市場のニュースに対する反応の違いについて比較する。具体的には、2.1 節で分析対象とした企業のニュース、6,172 件について、社債, CDS, 株価リターンについて比較を行った。ニュースのタイプ毎に、株価超過リターンの平均値を求めニュース発信前後における超過リターンを計測した。ニュース発表前、発表後(30日)の Good News と Bad News の累積超過リターンの差を図 4.2 に示す。

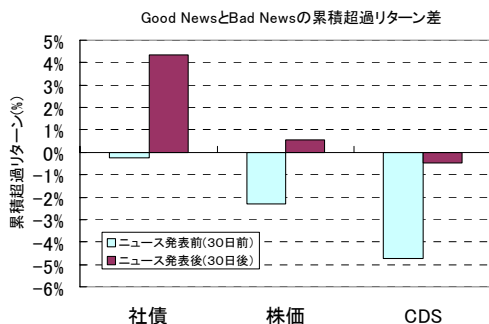


図 4.2 Good News と Bad News の累積超過リターン差

図 4.3 から、株式価格と CDS が類似した挙動をし、社債スプレッドは株式、CDS と異なる動きをしていることを確認できる。また、社債スプレッド変動は、ニュース発信後の累積超過リターンがニュース発信前よりも大きいことを確認でき、これらの結果は、ヘッドラインニュースに含まれる情報を用いた社債投資により、超過収益を獲得できる可能性を示唆するものであり、実務的観点からも興味深い。また、図 4.3 より株価の方が社債よりもニュースに反応するタイミングが早いことが確認できる。次節において、社債と株価の関係について分析を行う。

4.4 社債と株価の関係

本節では、社債のニュース発表後の超過スプレッドがニュース発表前の株価の超過リターンにて、どの程度説明できるか、分析を行う。

ニュース発信日から30日後の社債累積超過スプレッドを被説明変数とし、ニュース発信日の30日前からニュース発信日までの株価累積超過リターンを説明変数に単回帰分析を行う。その回帰した結果を表 4.7 に示す。

$$ExcessSpread_{社債,+30day} = \beta \cdot Excess Return_{株価,-30day} + \epsilon \quad \dots (4)$$

表 4.7 単回帰分析の結果

説明変数	β	t
累積株価リターン	-0.0032	-22.2
R-square	0.07	
Adjusted R-square	0.07	
N	6172	

同回帰分析の残差項に対し、前節までと同様の分析を行い、Good News と Bad News の社債累積超過スプレッド差を、格付別に集計した(図 4.3)。図 4.3 から、A 格付に該当する累積超過スプレッド差は、ニュース発表前の株価累積超過リターンでおおよそ説明されるのに対し、BBB 格以下は、株価累積超過リターンの影響で調整しても、社債累積超過スプレッドの値が十分に残存していることがわかる。これらの傾向は、高格付企業の社債スプレッドは、信用リスクは比較的安定的で、直前の株価のみに対応して変動するが、低格付企業の社債スプレッドは、

信用リスクが不安定な場合が多く、株価よりも、それ以外の信用リスクに関連する様々な情報に敏感に反応しやすいことが影響していると推測される。またこれらの結果は、株価累積超過リターンにて調整を行っても、ヘッドラインニュースの効果が社債リターンに残っている可能性を示すものであり、興味深い結果である。

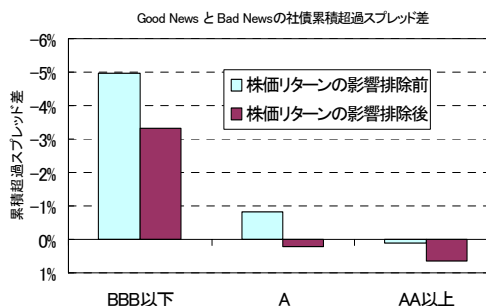


図 4.3 Good News と Bad News の社債累積超過スプレッド差

5. まとめ

本稿では、金融市場において最も即時性があり、投資判断における重要な情報源の一つである、ヘッドラインニュースに焦点を当て、テキスト情報がクレジット市場に与える影響について分析を行った。分析には、自然言語処理の分野において応用されている、テキスト自動分類アルゴリズムを用いた。テキスト自動分類アルゴリズムによりヘッドラインニュースを自動分類し、分類されたタイプ毎にスプレッドの変動に特徴があるかどうか検証した。分析の結果、(1) 自動分類アルゴリズムにより分類精度が約 80% のヘッドラインニュースのアルゴリズムを構築可能であること、(2) ヘッドラインニュースから抽出した情報は有効であること、(3) ニュース発信前後で社債, CDS, 株式間のリターンに違いが見られ、特に社債はニュース発表後に反応する傾向にあること、等の結果を得た。長期間のデータを用いた分析、取引コストを考慮した分析等については、今後の取り組みとしたい

参考文献

- [1]高橋悟, 高橋大志, 津田和彦, “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, pp.373-383 (2007.6).
- [2]大山慎介, 杉本卓哉, “日本におけるクレジット・スプレッドの変動要因”, 日銀ワーキングペーパー(2007.1)
- [3] 森平爽一郎, “信用リスクの測定と管理 第 3 回: オプションモデルによる倒産確率推定: 基礎”, 証券アナリストジャーナル(2000.1) pp.85-100
- [4] Antweiler, W., and Frank, M. Z., “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards,” The Journal of Finance, Vol. 59, pp.1259 - 1294 (2004)
- [5] Takahashi, S., Takahashi, H., Tsuda, K., Terano, T., “Analyzing Asset Management Knowledge from Analyst’s Reports through Text Mining,” International IPSI-2004 (2004.11)
- [6] 金明哲, 村上征勝, 永田昌明, 大津起夫 and 山西健司, “言語と心理の統計,” 岩波新書(2003)