

Title	研究開発活動のニュースが株価に及ぼす影響について
Sub Title	
Author	岡崎, 華奈(Okazaki, Haruna) 高橋, 大志(Takahashi, Hiroshi)
Publisher	慶應義塾大学大学院経営管理研究科
Publication year	2022
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2022年度経営学 第3956号
Genre	Thesis or Dissertation
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002022-3956">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002022-3956</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

学位論文（ 2022 年度）

論文題名

研究開発活動のニュースが株価に及ぼす影響について

主 査	高橋 大志
副 査	小幡 績
副 査	齋藤 卓爾
副 査	

氏 名	岡崎華奈
-----	------

## 論文要旨

所属ゼミ	高橋研究会	氏名	岡崎 華奈
(論文題名)			
研究開発活動のニュースが株価に及ぼす影響について			
(内容の要旨)			
<p>研究開発活動は技術系企業にとって重要な役割を果たしている。なぜならば、研究開発活動を行うことによって企業価値上昇に繋がる技術の開発や特許の申請などを行うことが可能になるからである。しかし、研究開発活動は外部からは実際にどのような活動を行なっているのか目に見えづらく、実際にその研究開発活動が事業の売上に寄与したかどうかを測ることも難しい。</p> <p>外部からの企業活動の動向を追う手段として企業から発表される有価証券報告書やニュースなどが存在する。これら公表されている情報は、タイムリーな情報と実際の活動と公表の間の期間が空いている情報などが存在する。例えば、有価証券報告書であれば年に数回ほどの発表しかなく必ずしもタイムリーな情報とはなっていない。</p> <p>企業動向は日を追うごとに変化を伴うものであり、実際の動向から期間が空いてしまう情報源は、既に企業活動が価格に反映されている可能性もあり、情報発信と資産価格変動の直接的な関連性を見出し難い可能性がある。一方、即時性の高い情報発信は価格変動と密接な関連を見出せる可能性があり、ニュースリリースは企業にとっても投資家や市場にとっても重要な情報源である。その一方で、未だタイムリーな情報源から研究開発活動などの企業動向を追うための手法や指標は確立されておらず、それらの企業動向を評価するための手法や指標を提案することが求められている。</p> <p>これらを背景として、本研究ではタイムリーな情報源であるどのようなニュースを発信すれば市場からどのように反応が返ってくるかについての指標を作成することを目指す。</p> <p>具体的な指標作成は、(1) BERTopic を使用してニュースデータをトピック分類し、その結果より対象企業を選出、(2) 単語抽出を使用して選出企業のニュースデータのみを抽出、(3) 抽出したニュースを BERTopic を使用してトピック分類を実施し、対象企業からどのようなニュースがリリースされているのかを調査、(4) リリースされているニュースをトピックごとに分類しそのトピックごとにニュースリリース日を特定、(5) リリース日をイベントデーとしイベントスタディを行って、ニュースと市場に関係性を分析、との手順にて実施した。</p> <p>本分析では、電気業界に属する東芝（株式コード&lt;6502.T&gt;）を対象企業とした。また、イベントスタディでは特徴的なトピックが含まれるリリース日だけでなく、様々なトピックがリリースされた際のイベントスタディをそれぞれ行った。</p> <p>分析の結果、特定企業のニュースから複数のトピックを抽出することができ、それらトピックと企業活動について一定の関連性を見出すことができた。これらの結果は、ニュースデータの抽出方法やトピック抽出方法などの改善することで、投資に有効な企業活動に関する指標を作成できる可能性があることを示唆するものである。</p> <p>一方、本分析では、ニュースと市場価格の関連性については、直接的な関連性を見出すことはできなかった。本分析では、特定企業および特定期間を対象とした分析を実施したが、詳細な分析は今後の課題である。</p>			

## 目次

1. はじめに .....	4
2. 先行研究 .....	5
3. 本研究の目的 .....	7
4. 使用するデータ .....	8
4.1 使用するマーケットデータ .....	8
4.2 使用するニュースデータ .....	8
5. 分析手法 .....	9
5.1 分析手法の概要 .....	9
5.2 BERTopic .....	9
5.3 BERT とは .....	11
5.4 対象企業の選定方法 .....	11
5.5 イベントスタディ .....	12
6. 分析結果 .....	13
6.1 企業選定の結果 .....	13
6.2 特定企業の結果 .....	14
6.3 イベントデーの決定 .....	17
6.4 イベントスタディの結果 .....	20
6  まとめと今後の課題 .....	24
7  おわりに .....	25
参考文献 .....	26
Appendix .....	27
その他のイベントスタディの結果について .....	27

## 1. はじめに

技術系企業にとって研究開発活動は企業を存続していくために重要な活動である。研究開発活動は、環境変化に対応するためにも、競争力を高めていくためにも必要な活動である。

また、企業の活動を存続していくためには対外的にも研究開発活動を公表していく必要がある。公表することにより、投資機会を得ることが出来たり、ブランドイメージの確立などが可能になったりと様々なメリットが存在するからである。このような企業内の動向は重要な評価軸となっているため、企業はニュースリリースなどを行い自社内の活動を外部に伝達している。

しかし、外部に研究開発活動の情報を伝達しているにも関わらず、外部からすると研究開発活動は無形資産であり、また研究開発活動が将来の事業に必ずしも結びつくとは限らないため研究開発活動を正しく評価されることは難しい。

そのため、研究開発活動を正しく評価する手法を提案することが求められている。

## 2. 先行研究

研究開発のような無形資産を評価する先行研究として Baglieri ら[1]の研究があり、この研究では研究開発の評価を行うためのフレームワークを提案している。しかし、このフレームワークは特定の企業や活動分野にしか適応することができず、また、研究開発活動の遂行と貢献度に強い関係性がある場合にのみ適応可能なフレームワークとなってしまう。

また、長谷部[2]の論文では無形資産を評価する別の方法として財務会計から着目し、コストアプローチ、インカムアプローチ、マーケットアプローチの観点から考察している。しかし、これらのアプローチの中で使用される貢献度と割引率に関して課題があるとされている。貢献度は将来性に基づいて試算されるが、将来性を正しく評価するためには個々の無形資産の状況を考慮し、整理した上で評価する必要性があり、実際に計算することは難しい。また割引率は、無形資産のリスクを反映して計算する必要があるが、リスクは特許の破棄や権利の無効など自社内だけで収まらないリスクが無形資産の場合は含まれてしまうため、計算することが難しい。よって現在の無形資産の評価方法は、貢献度、割引率を正しく設定した上で、無形資産の価値を財務目的で利用するのか、そして内部管理目的で使用するのかなど、どのように利用するかに着目しなければ正しく価値評価できないとしている。

外部への伝達手段としてのニュースが市場に与える影響としての先行研究として、Chen[3]の研究では悪いニュースがリリースされた後、株価が下落していることを証明し、ニュースと株価の関連性を実証している。

そして Antweiler[4]の研究では、インターネットの掲示板などの書き込みによって取引量が増える可能性があることを示唆している。

また、五島ら[5]の研究では、ニュースのテキスト情報から株価を予測することを試みている。数値情報だけでなく、ニュースのテキスト情報は数値に表れない情報を表すことが可能であるとされており、株価を予測する上でテキスト情報は重要な一つの試算価格の決定要素として位置付けられている。そしてニュースの内容にポジティブとネガティブを付与することによりディープラーニングの手法によってニュース記事の資産価格の分析を行なっている。

同じく五島ら[6]の研究ではディープラーニングを使用し、ニュースの内容を分類することで株価との影響を調査し、実際にニュース記事と株価が影響していることを実証しているが、ニュース指標による株価への影響は後追いが生じている可能性があることに言及している。

そしてニュースの内容がどのように株価に影響を与えるかにおける先行研究として Ryan ら[7]の研究が存在する。この研究では、ポジティブなニュースは

ポジティブに市場から捉えられ、ネガティブなニュースはネガティブに捉えられることを実証している。

また、特定のニュース記事に注目した先行研究として、趙ら[8]の研究がある。この研究では情報セキュリティ事故というニュース記事に絞って、株式市場の反応を見ることにより、ニュース記事の内容によって株価の変動が起こることを実証している。

研究開発活動と株式市場の先行研究として Zhang ら[9]の研究が存在している。この先行研究では、研究開発比率が株価に影響していることを証明しており、その中で研究開発活動の情報開示を適切に行うべきだと示唆している。

そして、BERTopic を使用し、ニュースから企業の部門ごとの活動を追うことができる可能性を示唆した論文も存在する。[10]

### 3. 本研究の目的

先行研究をまとめると、投資家が投資を行う際にニュースは重要な情報源かつ、ニュースというものは企業行動を知らせてくれるものであり、今後の企業行動を予測することができるということがわかる。またニュースに対する世間の反応を見ることによって、どのように株価が変動するかを自分自身で予測し、投資活動に活かすことができる。そのためニュースは投資家にとっても自分自身の投資行動を決断するために必要な情報源となっていることがわかる。

そして、昨今では投資家がニュース記事を参考にし、投資行動を決定するという事象からニュースを自然言語処理技術によって解析し、ニュース記事が株価変動にどのように影響しているかを明らかにする研究が増えてきている。その上、特定のニュース記事に焦点を当てた研究もされてきている。しかし、研究開発活動に焦点を当てた研究は少なく、また株価予測をする上で実際の株価の値を使用している研究は少ない。

これらの先行研究を踏まえると、ニュース記事と株価変動に関して関連性があることが示されており、特定のニュース記事や分野においてもそういった関連性があることがわかっている。また、研究開発活動は株式市場から注目されていることもわかっているがニュース記事を利用した研究はいまだ少ない。

また、研究開発活動は自社内で行われることに留まらず、提携、買収、合併などさまざまな方法で実施される。この特性により、ニュースの内容の中でもどのような方法を使用した研究開発活動によってポジティブな内容のニュースであったとしてもネガティブな反応が返ってくる可能性がある。

そのため、どのような内容のニュースリリースを行うことが出来れば研究開発活動の提携や自社内などの方法に左右されず、市場に評価してもらえるかを企業は知る必要がある。そこで、本研究では研究開発活動に関するニュースがどのように市場から評価されるかを調査する。そしてどのようなニュースをリリースすればどのような市場の反応が返ってくるのかについての指標を作成することを目指す。



## **4. 使用するデータ**

### **4.1 使用するマーケットデータ**

使用する株価のデータは TOPIX の 1996 年度から 2019 年度までのものを使用する。また株価の値としては終値を使用する。これらの株価データは日経 NEEDS からダウンロードしたものである。マーケットデータは日経 NEEDS の東証株価指数 1 部総合(終値)を使用している。また、対象企業の株価も同様に日経 NEEDS からダウンロードを行なっている。

### **4.2 使用するニュースデータ**

使用するニュースデータとしては、トムソンロイターニュースの 1996 年度から 2019 年度までのニュースデータを使用する。ニュースデータには様々な情報が記載されている。例えば、ニュース記事のヘッドライン、内容、ニュースがリリースされた日付などが要素として記載されている。

抽出方法としては、全ニュース記事の中から日本株式コードが含まれているニュースを抽出する。その後、トピック分類を行い、分析対象とする企業を選定する。

## 5. 分析手法

### 5.1 分析手法の概要

分析手法の一連の流れを図1にまとめる。

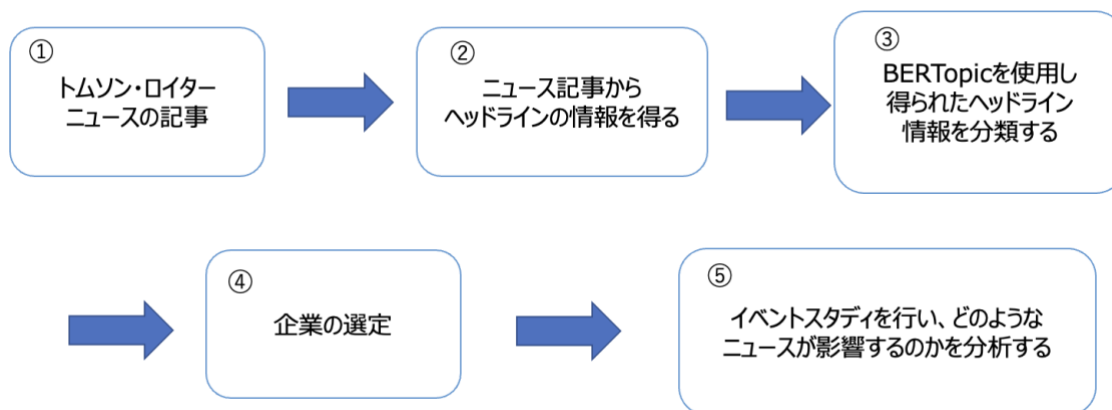


図1 本研究の分析手法の流れ

まず、分析を行う企業を選定するために用意したトムソン・ロイターニュースをBERTopicにてトピック分類する。その後、トピック分類した結果からニュース記事が多く、研究開発が盛んな電気業界に該当する企業を選定する。

次に、選定した企業のニュースのみを抽出し選定した企業がどのようなニュースリリースを行なっているかをトピック分類によって抽出する。そしてそこからトピックごとにイベントスタディを行い、どのようなニュースリリースであればどのような反応が市場から返ってくるのかを分析する。

次の節から具体的な使用技術や使用する設定などについて説明を行う。

### 5.2 BERTopic

BERTopicはMaarten Grootendorst[11]によって提唱されたトピックモデルの1つである。トピックモデルとは、文書中に現れる単語の出現度合いと意味などによってトピックを推定するモデルのことであり、自然言語処理においてはこのモデルを使用することによってニュース記事などのテキストデータをトピックに分類することを行ってきた。

元来のトピックモデルとは違い単語の出現度合いと意味でトピックを推定するのではなく、BERT(Bidirectional Encoder Representations from Transformers)という自然言語処理モデルを使用し文書をベクトル化することによってトピックを推定することを可能としている。従来のトピックモデルとしてLDA(Latent Dirichlet

Allocation) [12]などが挙げられる。この2つのモデルは文書中に単語がどれくらい現れるかによって確率的にトピックを推定するモデルであり、昨今では様々な場面で使用されてきた。

この従来のモデルと BERTopic の大きな違いは対象となる文章をどのように解釈するかが異なっている。従来のモデルでは対象となる文章の単語を確率的に計測し文章の評価を行うが、BERTopic では対象となる文章を自然言語処理によって解析を行い、ベクトル化することで文章の評価を行う。

具体的には、文書を元来のモデルでは形態素解析し、そこで得られた単語の頻出度によって分類していたものを BERT という自然言語処理によって対象となる文章につき 768 次元の特徴量を算出しベクトル化している。BERT とは Devlin [12]によって提唱された自然言語処理モデルの一つである。

BERTopic はこの BERT を使用することで元来の確率的なトピックモデルから、文章の内容をベクトル化し、精度を上げることを可能とした。以下の図 2 に他のトピックモデルと BERTopic との精度比較を行なった図を示す。20 NewsGropus、BBC News、Trump はそれぞれのどのテキストデータを分類したかを示しており、左側の LDA、NMF、T2V-MPNFT、T2V-Doc2Vec、CTM、BERTopic-NPFNET はそれぞれ使用したトピックモデルを表している。また、TC はトピックの一貫性、TD はトピックの多様性に関するスコアを示したものである。

	20 NewsGroups		BBC News		Trump	
	TC	TD	TC	TD	TC	TD
LDA	.058	.749	.014	.577	-.011	.502
NMF	.089	.663	.012	.549	.009	.379
T2V-MPNET	.068	.718	-.027	.540	-.213	.698
T2V-Doc2Vec	.192	.823	.171	.792	-.169	.658
CTM	.096	.886	.094	.819	.009	.855
BERTopic-MPNET	.166	.851	.167	.794	.066	.663

図 2 BERTopic と他のトピックモデルの比較 (出典[11])

この図 2 からわかるように BERTopic は全てのスコアに関して 1 番ではないものの幅広くスコアの精度を上げることを可能にしたモデルであることがわかる。

本研究では、BERTopic を使用することでニュースをトピックごとに分類し、どのようなニュースがどのように市場に影響をもたらすのかについて指標を作

ることを目指す。

### 5.3 BERT とは

BERT (Bidirectional Encoder Representations from Transformers)とは、前述した通り Devlin ら[13] によって提唱された自然言語処理モデルの1つである。自然言語処理の手法として、従来は単語の出現度合いなどから類似文章を探索する手法が存在する。また、極性辞書と呼ばれるそれぞれの単語に点数をつけた辞書をあらかじめ用意し、文章を形態素解析することによって文章をそれぞれの単語に分けて、各単語のスコアを合算することにより、その文章がポジティブな文章なのかネガティブな文章なのかを判別するという手法などが存在した。しかし、これらの手法の欠点として、文脈を考慮した文章の評価ができないという欠点があった。例えば、「この花は綺麗だ。しかし、あの花の方が綺麗である。」という文章があるとすると、「綺麗」という単語と「しかし」という単語の頻出度合いより、極性辞書の単語に対するスコアの付け方によっては、この文章はポジティブに捉えられてしまう。

このような欠点を改善するために提案された手法がBERTである。BERTでは事前学習において、MLM(Masked Language Modeling)とNSP(Next Sentence Prediction)を行うことにより精度を大幅に上げることを可能にした。MLMでは、1つの文章の中に穴を作っておき、その穴を埋める短度を予測する学習を行うことである。NSPは2つの文章が連続しているかを当てる学習のことである。これらを行うことでBERTは文脈を考慮した文章の評価を行うことができるモデルとなった。

BERTは幅広く利用されており、実際にニュース記事と株価情報を結びつける際的手段として実際に井口ら[14]の論文でもBERTが使用されている。

### 5.4 対象企業の選定方法

今回企業動向を追う企業の対象方法としてBERTopicによる選定方法を使用する。具体的には1996年度から2019年度までのトムソンロイターニュースをBERTopicによりトピック分類を行う。その際、トピック分類を行う対象としてニュースのヘッドラインを対象としてトピック分類を行う。そしてその結果より、上位にトピックが上がる企業、かつ、研究開発が盛んとされている電気業界に含まれる企業を対象とする。また、本研究では各企業が研究開発を行っていることが前提とされており、ニュースのトピックには製品名などが出てくると予想されるため、製品と研究開発活動が結びついていることが多い電気業界を対象とした。

## 5.5 イベントスタディ

イベントスタディとは、企業活動において何らかの情報発信が行われた際に株価変動などを見ることによりどのようにそのイベントや情報が影響を与えたかを分析する手法のことである。イベントスタディではイベントが発生しなかった場合の時の正常リターンとイベントが発生した際の異常リターンを累積し、この累積異常リターンがどれだけ正常リターンと差があるかによってそのイベントの影響度合いを図るものである。

Campbell[15]らの研究では企業の金融分野においてイベントスタディは実証研究として有効であることが示されている。そして、イベントスタディを用いた研究において、イベントデーを特定することが重要であると結論づけられている。

本研究では、イベントデーをニュースが発信された日とする。その際、土曜や日曜などの株式市場が取引を行っていない日に出たニュースに関してのイベントで一日は翌営業日と定義している。そしてイベントウィンドウを前後19日と設定している。またイベントスタディの方法としては先行研究である趙ら[8]の方法を参考にして行なっている。

## 6. 分析結果

### 6.1 企業選定の結果

5.4 節にて説明した通り BERTopic を使用して企業選定を行う。以下の図 3 は 1996 年度から 2019 までのトムソン・ロイターニュースを対象とし、そのニュースをトピックに分類した結果である。

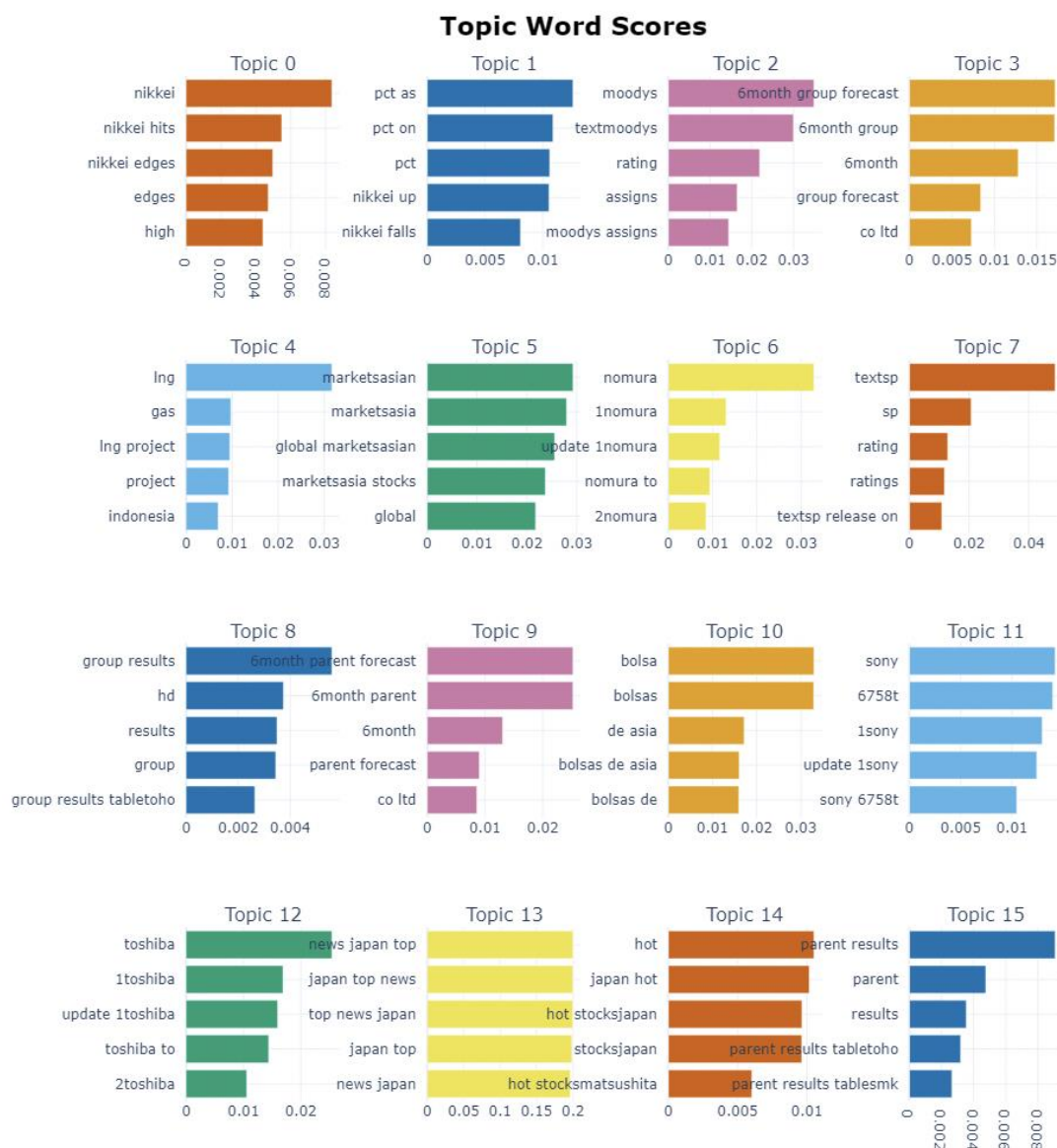


図 3 1996 年度から 2019 年度までのトムソン・ロイターニュースを分類した結果

この図 1 は全体のニュースにおけるヘッドラインを抽出しトピック分類した結果のうち上位 20 トピックに分類されたものを表している。これらの Topic0 から Topic19 は全体のニュースを分類した中で、トピックにカウントされたものが多

い順に並んでいる。また、横軸の単語はそのトピックを表す単語を上から順に並べている。この横軸の単語群を組み合わせることによって、そのトピックが具体的にどういうトピックを表しているのかを理解することが可能である。

このトピック群の結果より、前述した 5.3 節の企業選定の方法を使用し今回は東芝（株式コード<6502.T>）を対象企業とした。東芝（株式コード<6502.T>）は電気業界を代表する非常に有名な企業であり、今回の研究に適切な企業である。

## 6.2 特定企業の結果

選定した企業である東芝（株式コード<6502.T>）のみのニュースを抽出する。1996 年度から 2019 年度のトムソン・ロイターニュースの日本銘柄だけを抽出したニュースデータ群からヘッドライン、またはニュース内容事態に株式コード<6502.T>が含まれているかどうかによってニュースを抽出した。この時の対象となるニュースは 26,058 件であった。この 26,058 件のニュースをもう一度 BERTopic にて分類し、どのようなニューストピックがリリースされているのかを検証する。以下の図 4 に東芝（株式コード<6502.T>）のみのニュースで分類された際の上位 20 に関するトピックを示す。

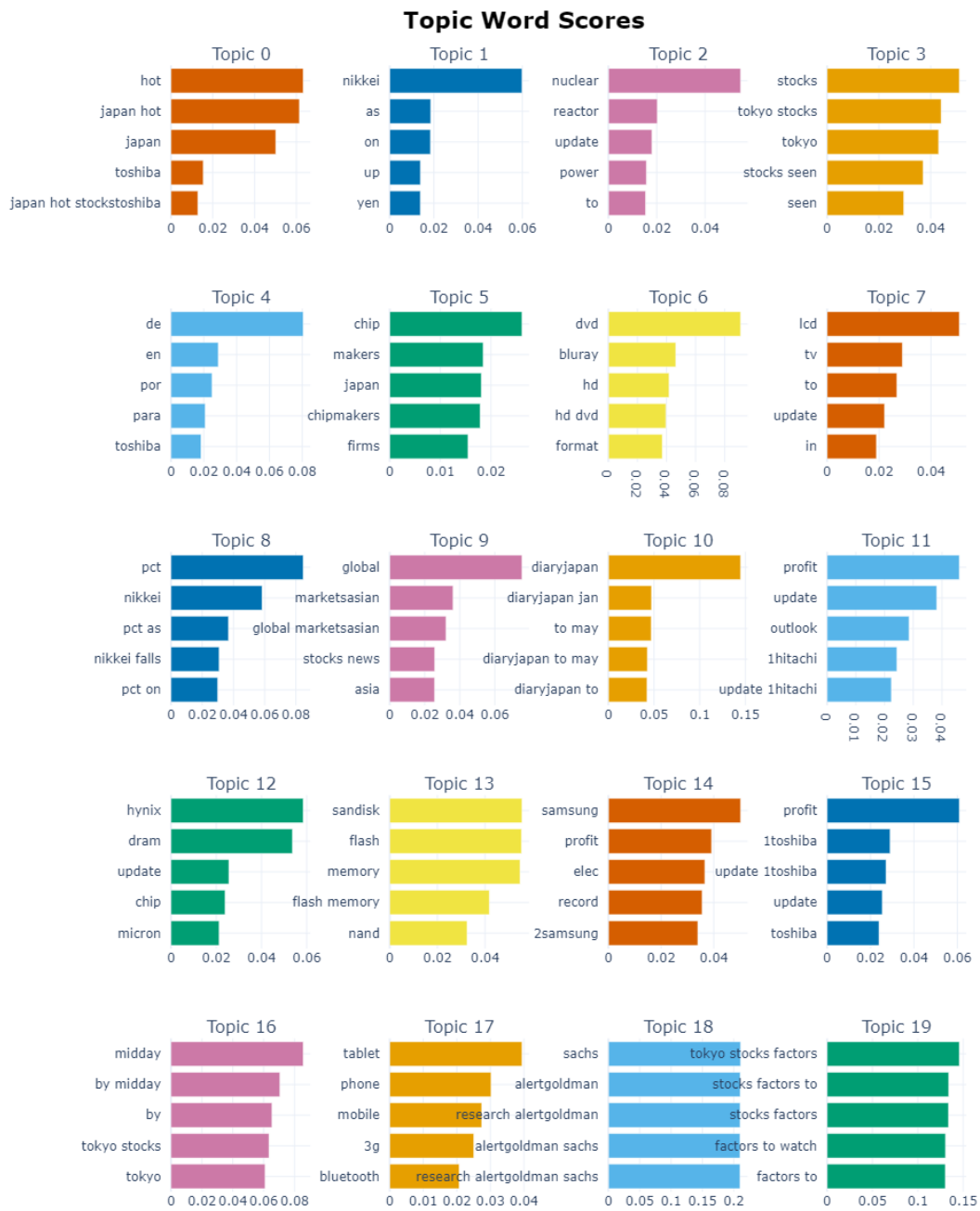


図 4 BERTopic にて東芝(株式コード<6502.T>)のニュースのみで分類した結果

また以下の表にそれぞれのトピックを表す最も代表的な単語を示す。



表1 図2の結果より各トピックを最も表す単語の一覧

Topic	トピックを最も表す単語	カウントされた数
0	hot	3986
1	nikkeis	1775
2	nuclear	1472
3	stocks	1240
4	de	855
5	chip	623
6	dvd	503
7	lcd	396
8	pct	373
9	global	341
10	diaryjapan	329
11	profit	339
12	hynix	284
13	sandisk	268
14	samsung	256
15	profit	252
16	midday	236
17	tablet	210
18	sachss	184
19	tokyo stocks factors	182

この表の結果よりそれぞれのトピックがどのようなトピックを表すかを簡略的に理解することが可能である。しかし、Topic11 と Topic15 では同じ「profit」という単語が最もそのトピックを表している単語となっており、2 番目の単語を考慮しなければどのようなトピックであるかを区別することができない。また、他のトピック 4 の「de」やトピック 8 の「pct」など、その一つの単語だけでは何を指しているのかがわかりづらいものが最も代表的な単語であるという結果になってしまっており、事前のデータにおいて副詞や形容詞などの単独では意味をあまりなさない単語を除くなどの改善が必要があると考えられる。

今回の研究においてはそれぞれのトピックを表す単語を表 1 にある単語とする。これは 2 番目以下の単語も含んだニュースをそれぞれのトピックに分類さ

れたニュースであると仮定すると、条件に合致するニュースをイベントスタディを行う特定の期間内にて抽出することが難しい場合が多いからである。

### 6.3 イベントデーの決定

今回、イベントスタディを行う期間として任意の1ヶ月での実験を試みた。今回選択した期間は2007年の3月である。6.2節で前述した結果を元にどのようなニュースが株価に影響を与えているのかを検証するためにまずイベントデーを選出する。このイベントデーを選出する方法として、6.2節の表1で表されている通り、各トピックを表す単語を1つ決定しその単語が内容に含まれているニュースをそのトピックのものであるとする。ただし、Topic11とTopic15では同じ「profit」が最もトピックを表す単語として算出されてしまったため、例外的にこの2つのトピックのみ2番目の単語まで採用している。そのためTopic11を表す単語は「profit」と「update」とし、Topic15を表す単語は「profit」と「1toshiba」としている。

各トピックのニュースを抽出した後、そのニュースがリリースされた日を調べる。そしてニュースが証券取引所が休みでなければその当日、休みであれば翌営業日をイベントデーとしてイベントスタディを行う。以下の表2に2007年3月のそれぞれの日にどのようなトピックのニュースがリリースされたのかを示す。

表2 2007年3月にリリースされたニュースのトピック

日付	曜日	該当したトピック数	該当したトピック
2007/3/1	木	2	Topic4, Topic9
2007/3/2	金	8	Topic1, Topic3, Topic4, Topic5, Topic7, Topic8, Topic9, Topic18
2007/3/3	土	0	なし
2007/3/4	日	0	なし
2007/3/5	月	0	なし
2007/3/6	火	3	Topic3, Topic4, Topic9
2007/3/7	水	3	Topic4, Topic6
2007/3/8	木	10	Topic0, Topic1, Topic3, Topic4, Topic5, Topic8, Topic9, Topic11, Topic14,

			Topic18
2007/3/9	金	0	なし
2007/3/10	土	4	Topic1, Topic3, Topic4, Topic9
2007/3/11	日	0	なし
2007/3/12	月	2	Topic4, Topic9
2007/3/13	火	0	なし
2007/3/14	水	4	Topic1, Topic2, Topic4, Topic9
2007/3/15	木	13	Topic0, Topic1, Topic2, Topic3, Topic4, Topic5, Topic6, Topic7, Topic8, Topic9, Topic12, Topic14, Topic18
2007/3/16	金	7	Topic0, Topic1, Topic2, Topic3, Topic4, Topic5, Topic18
2007/3/17	土	0	なし
2007/3/18	日	0	なし
2007/3/19	月	4	Topic4, Topic5, Topic8, Topic18
2007/3/20	火	8	Topic1, Topic3, Topic4, Topic5, Topic12, Topic14
2007/3/21	水	10	Topic0, Topic3, Topic4, Topic5, Topic7, Topic9, Topic11, Topic12, Topic13, Topic14
2007/3/22	木	15	Topic0, Topic1, Topic2, Topic3, Topic4, Topic5, Topic6, Topic7, Topic8, Topic9, Topic11, Topic12, Topic13, Topic14, Topic18
2007/3/23	金	2	Topic4, Topic9
2007/3/24	土	0	なし

2007/3/25	日	0	なし
2007/3/26	月	0	なし
2007/3/27	火	6	Topic2, Topic4, Topic5, Topic9, Topic14
2007/3/28	水	0	なし
2007/3/29	木	3	Topic2, Topic4, Topic7
2007/3/30	金	4	Topic1, Topic2, Topic4, Topic9
2007/3/31	土	0	なし

この表 2 の結果より、トピックの中でもどのようなトピックか直感的にわかりやすい Topic6 (最も代表的な単語が「dvd」) が含まれている日と該当するトピック数が 5 個未満の日をイベントデーとして設定する。それぞれ Topic6 が含まれる日は 2007/03/07、15、22 であり、Topic6 が含まれる日として挙げられた日程以外の日かつ該当するトピック数が 5 個未満の日は、2007/03/01、06、10、12、14、19、23、29、30 である。このとき、2007/03/10 は土曜日であり証券取引所が休みのため、翌営業日の 2007/03/12 とみなし、2007/03/12 と同じイベントスタディとして行う。表 2 の日付が赤い日に関しては該当するイベントデーとして設定された日付である。5 個未満に設定した理由としては同時点で発信されるニュースを考慮してのことである。同時点に発信されるニュースが多ければ多いほどお互いに干渉してしまいそのニュースが株価に影響しているか検証しづらいためである。

## 6.4 イベントスタディの結果

それぞれのイベントデーのイベントスタディ結果を、AR をそれぞれの日の異常リターン、CAR を累積異常リターンとして示した図と共に考察を行う。横軸の数字はイベントデーからどれほど離れているかを示す。マイナスの数値はイベントデーより前の日であり、プラスの数値はイベントデーより後の数値を示している。以下の表3に各イベントデーと市場からの反応をまとめたものを示す。市場からの反応の中で「×」は市場からの反応が見られなかったものであり、「+」は市場から好ましい反応が得られたものであり、「-」は市場から好ましくない反応が得られたというものである。この章ではそれぞれの反応ごとに結果を示す。好ましい反応として2007/3/7の結果、反応が認められない例として2007/3/15の結果、好ましくない反応の例として2007/3/29の結果を示す。

表3 市場からの反応の結果

日付	該当したトピック	市場からの反応
2007/3/1	Topic4, Topic9	×
2007/3/6	Topic3, Topic4, Topic9	(—)
2007/3/7	Topic1, Topic4, Topic6	(+)
2007/3/10	Topic1, Topic3, Topic4, Topic9	(—)
2007/3/12	Topic4, Topic9	(—)
2007/3/14	Topic1, Topic2, Topic4, Topic9	(+)
2007/3/15	Topic0, Topic1, Topic2, Topic3, Topic4, Topic5, Topic6, Topic7, Topic8, Topic9, Topic12, Topic14, Topic18	×
2007/3/19	Topic4, Topic5, Topic8, Topic18	×
2007/3/22	Topic0, Topic1, Topic2, Topic3, Topic4, Topic5, Topic6, Topic7, Topic8, Topic9, Topic11, Topic12, Topic13, Topic14, Topic18	(+)
2007/3/23	Topic4, Topic9	(—)

2007/3/29	Topic2, Topic4, Topic7	(—)
2007/3/30	Topic1, Topic2, Topic4, Topic9	(+)

(1) 2007/03/07 の結果

以下の図 5 に 2007/03/07 のイベントスタディを行った結果を示す 2007/03/07 は Topic1 と Topic4 と Topic6 が観測された日である。イベントデー当日の AR は -0.0148997 であり、前日は 0.0012125、翌日は 0.02690224 となっている。イベントデー当日から翌日にかけての異常リターンは上昇しており、Topic4、Topic6 のいずれかのトピックを含むニュースの影響によって市場から好ましい反応が返ってきた可能性が存在する。

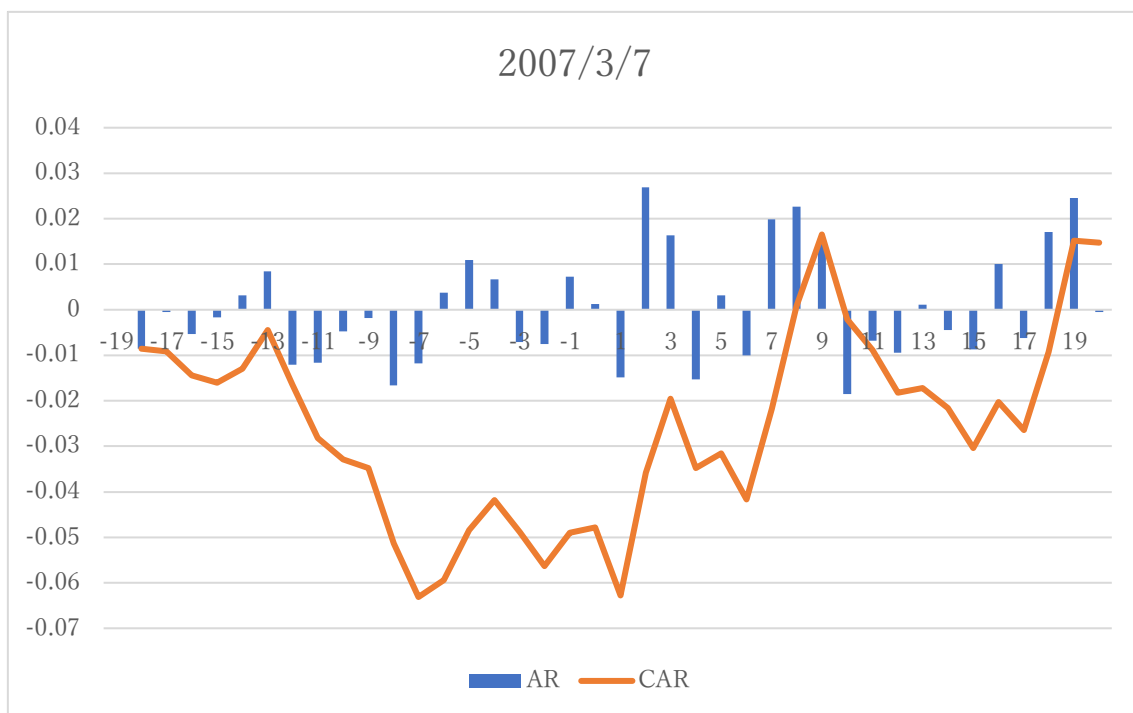


図 5 2007/03/07 のイベントスタディの結果

(2) 2007/03/15 の結果

以下の図 6 に 2007/03/15 のイベントスタディ結果を示す。2007/03/15 は Topic0、Topic1、Topic2、Topic3、Topic4、Topic5、Topic6、Topic7、Topic8、Topic9、Topic12、Topic14、Topic18 が観測された日である。イベントデー当日は 0.01828739、前日は-0.0109394、翌日は 0.02147569 であった。前日から当日、そして当日から翌日にかけてどちらとも異常リターンは上昇しており、イベントデーが起因となって異常リターンが上昇したとは結論し難い。

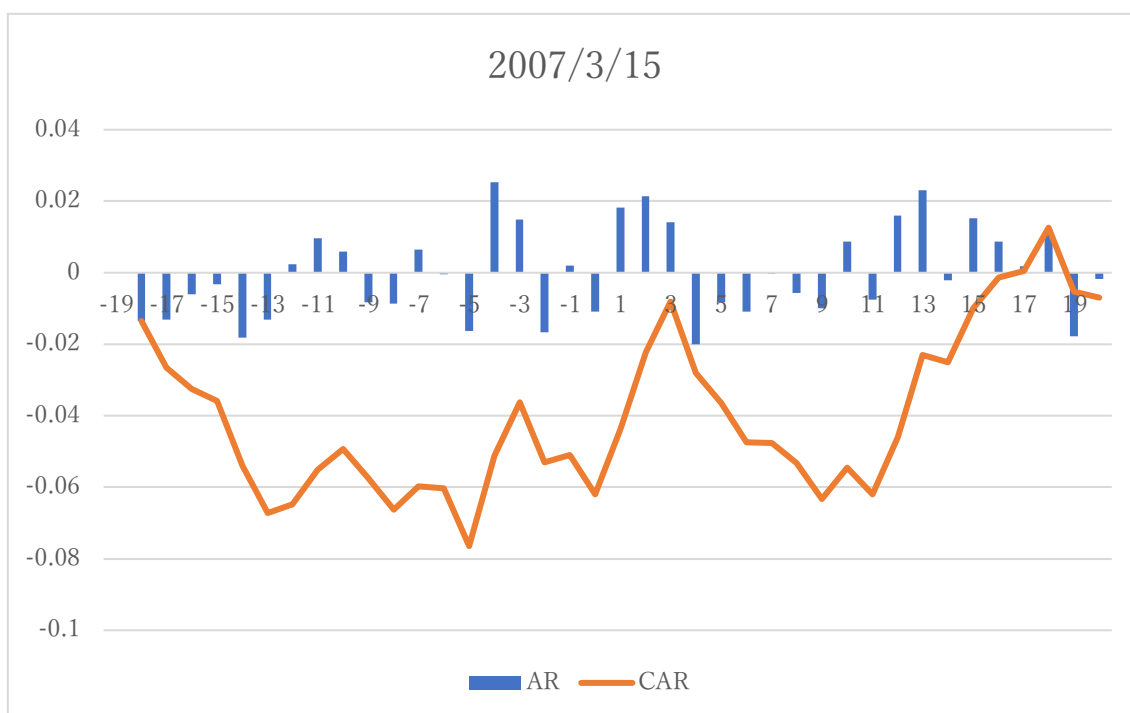


図 6 2007/03/15 のイベントスタディの結果

### (3) 2007/03/29 の結果

以下の図7に2007/03/29をイベントデーとしてイベントスタディを行った結果を示す。このイベントデーで観測されたトピックはTopic2、Topic4、Topic7である。イベントデー当日の異常リターンは0.00742285であり、前日は-0.0114598、翌日は-0.0088251である。イベントデーの前日から当日にかけて異常リターンは上昇しており、当日から翌日にかけては異常リターンが減少している。このことから、Topic2、Topic4、Topic7のいずれかのトピックが市場から好ましくない反応をされた可能性が示唆される。

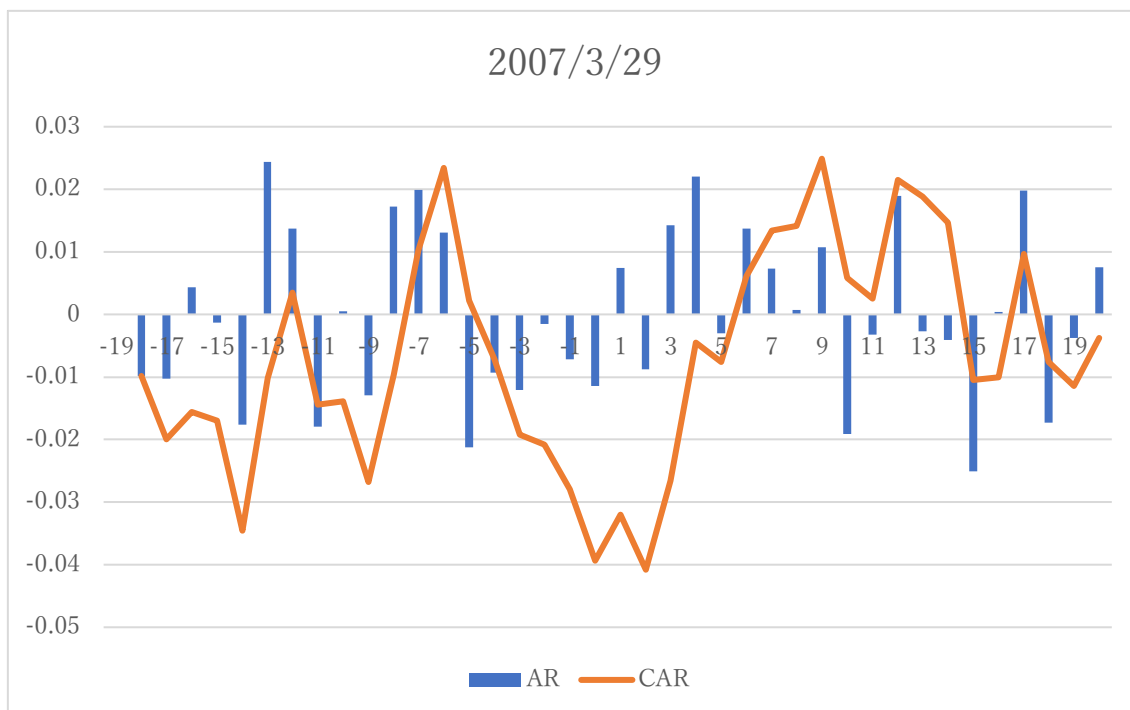


図7 2007/03/29のイベントスタディの結果



## 6 考察

6 章にてニュースの抽出からイベントスタディを行った結果までを示した。これらの結果より BERTopic を使用することで、ニュースがどのように市場から評価されるかについての指標を作成する可能性があることを示唆した。

しかし、未だに細かくどのようなトピックがどのように市場に影響を与えるのかについて言及することは難しい。このことを解消するためには、データの工夫や追加調査が必要である。

まず、ニュースをトピックに割り当てる際の割り当て方法である。本研究ではニュースをトピックに割り当てる際はトピックの最も代表的な単語がニュースの内容に含まれているかどうかで該当ニュースを抽出し、抽出されたニュースをそのトピックのニュースであると設定した。しかし、この方法をとってしまうと 1 つの単語依存の抽出方法になってしまい、該当するニュース数も多くなってしまう。そのため、事前に他の方法で抽出したニュースをもう一度割り当て直すなどの工夫が必要である。今回の研究では該当するトピックが多すぎるが故にどのようなニュースが影響するかを断定することが難しかった。そのため、該当するニュース数をさらに絞ることができるような割り当て方法や抽出方法を考案する必要が存在する。

また、イベントスタディを行う期間に関しても工夫が必要である。今回は限定された 1 ヶ月間みのイベントスタディにて結果を求めたが、さらに正確性を求めるならば追加の期間でイベントスタディを行うべきである。また複数社のイベントスタディを比較することも行うべきである。

そして、イベントデーの設定についても考え直すべきである。今回はニュースがリリースされた時間を考慮せず、リリースされた日をそのままイベントデーとして設定した。しかし、ニュースリリースされた時間が取引所の営業時間後であれば次の日に反応が見られる可能性が大きい。そのためイベントデーの設定をも見直すべきである。

## 7 まとめと今後の課題

本研究では、研究開発活動を評価する手法の1つを示し、その手法の妥当性を評価した。また、本研究において提案された手法は、研究開発活動を評価するための1つの指標になり得る可能性があることを示した。研究開発活動は不透明な部分が多く、また、実際に売上が経つまでにラグがどうしても発生してしまう。このような特殊な活動を主にする電気業界などに属する企業にとってこの手法を使用することによって、自社製品や自社の活動がどれほど市場と乖離があるかを測ることに繋がり、どのようなニュースをリリースすれば市場に正しく評価されるかを導くことに寄与することができる。しかし、そのためにはさらにこの手法を改善していく必要性がある。

## 参考文献

- [1] Enzo Baglieri, Vittorio Chiesa, Alberto Grando and Raffaella Manzini: "EVALUATING INTANGIBLE ASSETS: THE MEASUREMENT OF R&D PERFORMANCE", Research Division Working Paper No. 01/49, Available at SSRN: <https://ssrn.com/abstract=278260>, 2001.
- [2] 長谷部智一郎 : "無形資産価値評価に関する一考察", 特許研究(67), 2019, p.35-46.
- [3] Wesley S. Chan: "Stock Price Reaction to News and No-News: Drift and Reversal After Headlines", Journal of Financial Economics, vol.70, no.2, 2003, pp.233-260.
- [4] Werner Antweiler and Murray Z. Frank, "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," The Journal of Finance, Vol. 59, No. 3, Jun 2004, pp. 1259-1294.
- [5] 五島圭一, 高橋大志, 寺野隆雄, "ニュースのテキスト情報から株価を予測する", 第29回人工知能学会全国大会, 2015.
- [6] 五島圭一, 高橋大志, "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, 2016.
- [7] Basilone Ryan: "The Impact of News on Stock Market Investors", SSRN, <https://ssrn.com/abstract=3908662>, August 20, 2021.
- [8] 趙健森 and 高橋大志: "情報セキュリティ事故に対する株式市場の反応に関する研究: 日本および中国株式市場を対象とした分析", 第19回経営課題にAIを! ビジネスインフォマティクス研究会, 2019.
- [9] Xinzhu Zhang, Yu Zhang and Yuxin Mao: "Corporate Research and Development Strategy and Stock Price Crash Risk", Proceedings of the 2021 3rd International Conference on Economic Management and Cultural Industry, 2021.
- [10] H.Okazaki, H.Takahashi, "Nowcasting of Corporate Research and Development trends through news article analysis by BERTopic: the case of Japanese electric company", Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) 16-18 November 2022, Maldives(to be published)
- [11] M. Grootendorst: "Bertopic: Neural topic modeling with a class-based tf-idf procedure", Available at <https://doi.org/10.48550/arXiv.2203.05794>, March 2022.
- [12] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, p.993-1022, 2003.
- [13] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv:1810.04805v2, 2018.
- [14] 井口勝太 and 湯本高行, "BERT を用いた株価への影響力のあるニュース記事の推定", DEIM, 2022.
- [15] John Y. Campbell, Andrew W. Lo and Archie Craig Mackinlay: "The econometrics of financial markets", Princeton University press, 2022.

## Appendix

### A.1 その他のイベントスタディの結果について

(A-1) 2007/03/01 の結果

以下の図 A に 2007/03/01 のイベントスタディを行った結果を示す 2007/03/01 は Topic4 と Topic9 が観測された日である。イベントデー当日の AR は -0.00475897 であり、前日は 0.00894184、翌日は -0.00517656 であった。イベントデー前日から当日にかけて AR が下がっており、翌日も下がっている。しかし、イベントデーの前日から異常リターンが下がっているため、このイベントデーに該当するトピックに属するニュースが株価に影響があったかどうかを断定することは難しい。

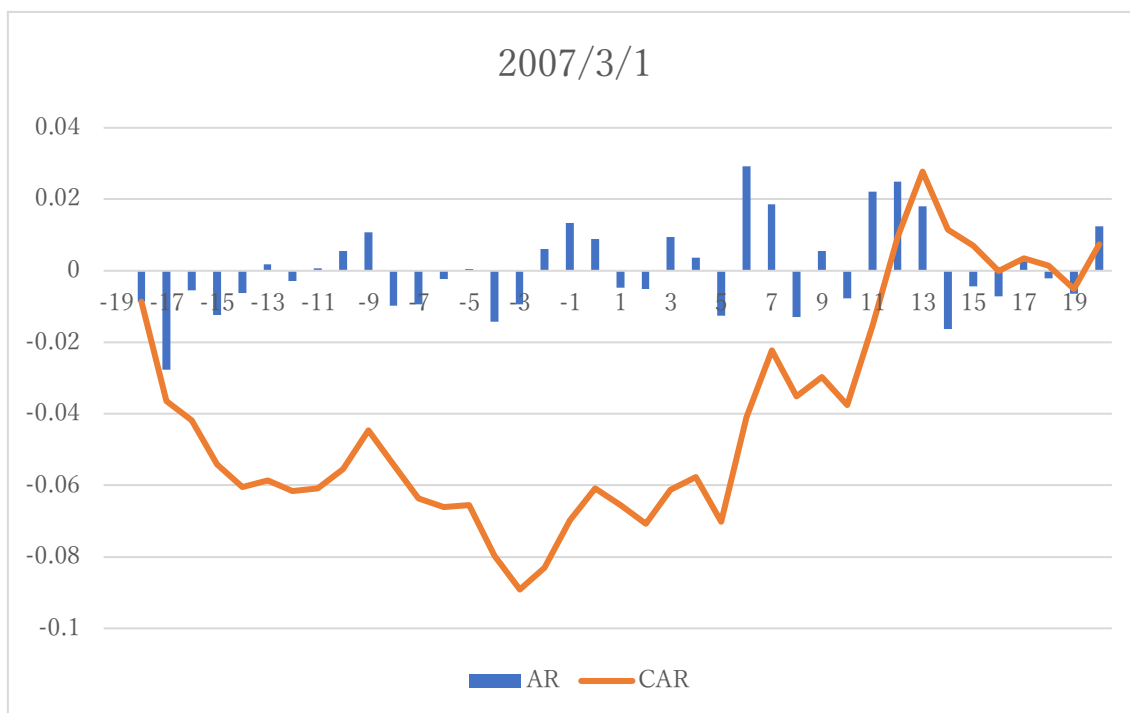


図 A 2007/03/01 のイベントスタディの結果

(A-2) 2007/03/06 の結果

以下の図 B に 2007/03/06 のイベントスタディを行った結果を示す 2007/03/06 は Topic3 と Topic4 と Topic9 が観測された日である。イベントデー当日は 0.00103564、前日は 0.00811125、翌日は-0.014695 である。前日から当日、当日から翌日にかけて異常リターンは減少しており、Topic3、Topic4、Topic9 のいずれかのトピックを含むニュースが市場に影響を与えたかどうかについて言及することは難しい。

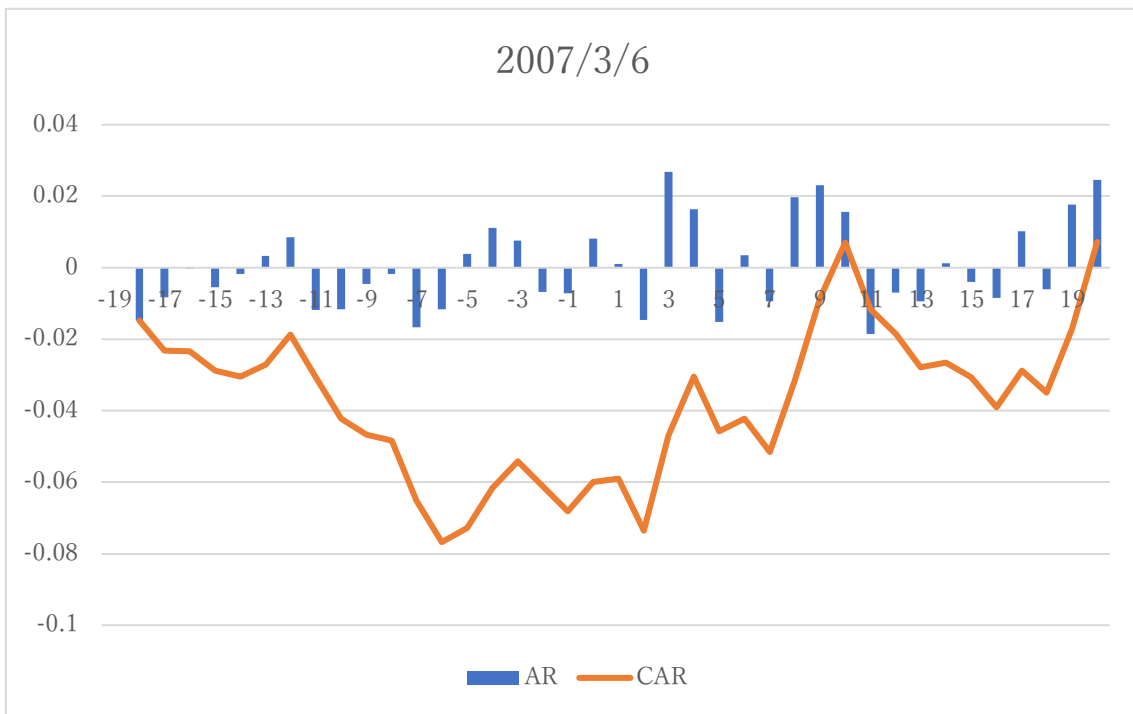


図 B 2007/03/06 のイベントスタディの結果

(A-3) 2007/03/10 と 2007/03/12 の結果

以下の図 C に 2007/03/12 のイベントスタディを行った結果を示す。前述した通り、2007/03/10 は土曜日であったため 2007/03/12 とみなしたため、ここでは 2007/03/10 と 2007/03/12 を同じ結果として示す。2007/03/12 にて観測されたトピックは Topic1、Topic3、Topic4、Topic9 である。イベントデー当日の異常リターンは -0.0166282、前日は 0.01496376、翌日は 0.00188013 である。前日から当日にかけては異常リターンが減少しているのにも関わらず、当日から翌日は異常リターンが上昇している。つまり、イベントデーを起因として異常リターンに変動があった可能性がある。このことより、Topic1、Topic3、Topic4、Topic9 のいずれかのトピックを含むニュースが市場に好ましい影響を与えた可能性が高い。

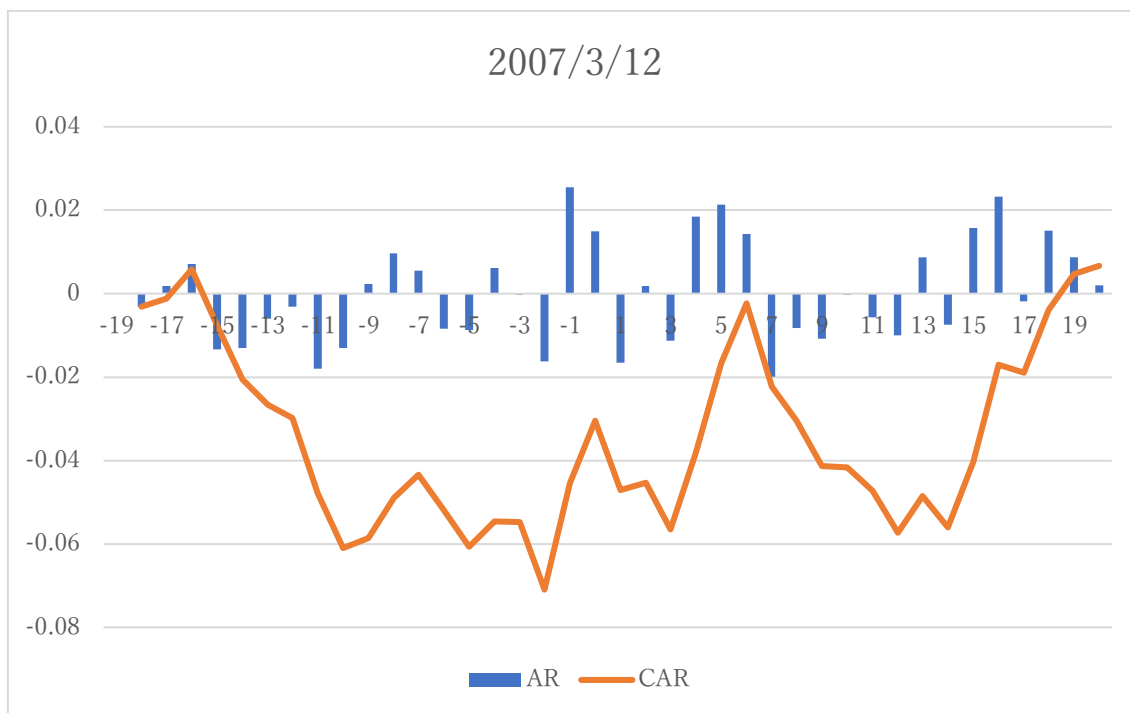


図 C 2007/03/10 と 2007/03/12 のイベントスタディの結果

(A-4) 2007/03/14 の結果

以下の図 D に 2007/03/14 のイベントスタディを行った結果を示す。2007/03/14 は Topic1、Topic2、Topic4 と Topic9 が観測された日である。イベントデー当日の異常リターンは-0.0110443、前日は 0.00187627、翌日は 0.01820573 であった。イベントデー前日から当日にかけての異常リターンは減少しているが、当日から翌日にかけての異常リターンは上昇している。このことから、Topic1、Topic2、Topic4 と Topic9 のいずれかのトピックを含むニュースが市場に好ましい影響を与えた可能性が高いことがわかる。

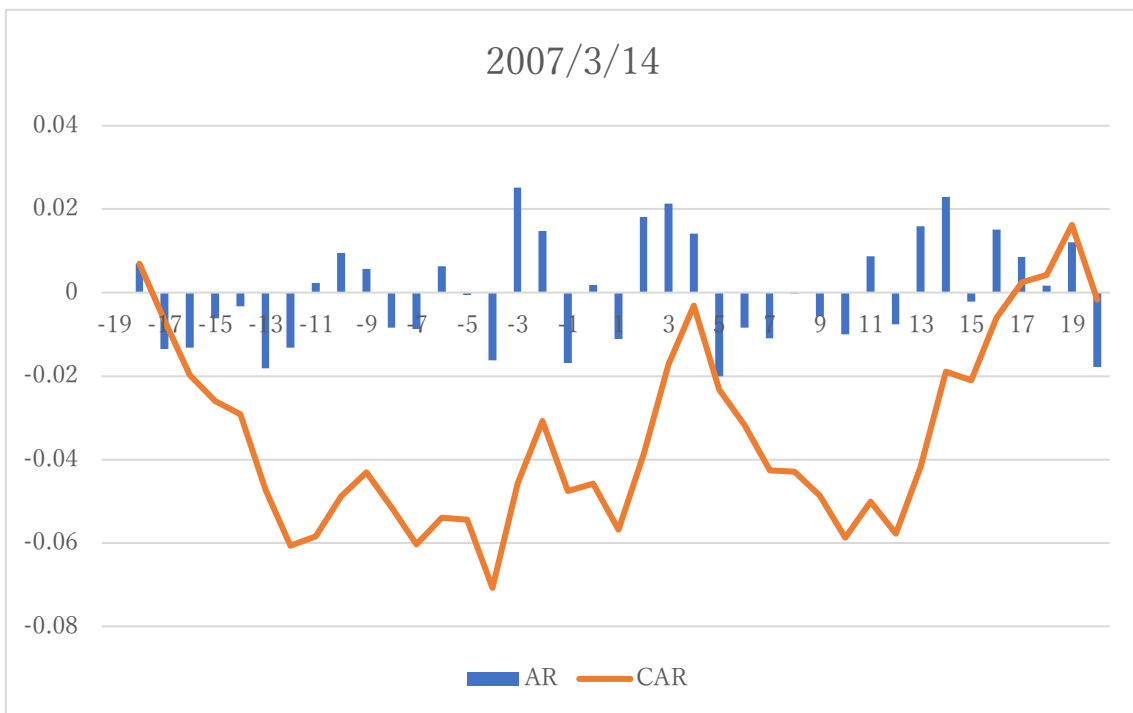


図 D 2007/03/14 のイベントスタディの結果

(A-5) 2007/03/15 の結果

以下の図 E に 2007/03/15 のイベントスタディ結果を示す。2007/03/15 は Topic0、Topic1、Topic2、Topic3、Topic4、Topic5、Topic6、Topic7、Topic8、Topic9、Topic12、Topic14、Topic18 が観測された日である。イベントデー当日は 0.01828739、前日は-0.0109394、翌日は 0.02147569 であった。前日から当日、そして当日から翌日にかけてどちらとも異常リターンは上昇しており、イベントデーが起因となって異常リターンが上昇したとは結論し難い。

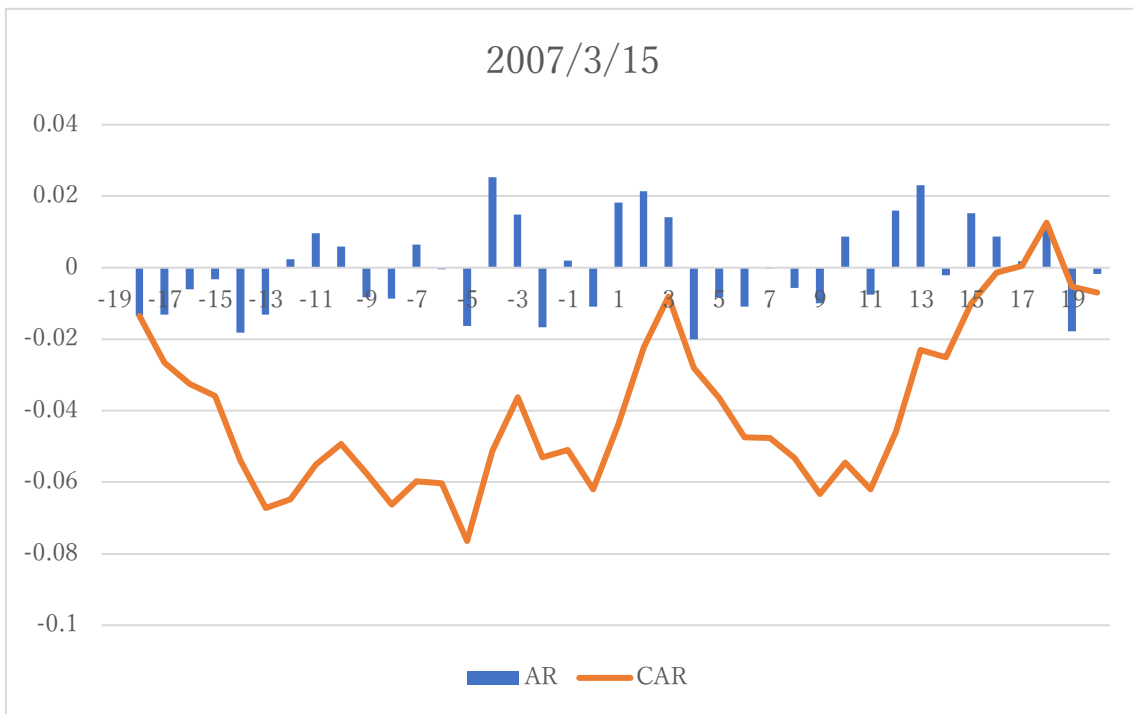


図 E 2007/03/15 のイベントスタディの結果



(A-6) 2007/03/19 の結果

以下の図Fに2007/03/19のイベントスタディを行った結果を示す。このイベントデーに観測されたトピックはTopic4、Topic5、Topic8、Topic18である。イベントデー当日の異常リターンは0.01408194、前日は0.02037016、翌日は-0.0202551である。前日から当日にかけて、そして当日から翌日にかけていずれの期間も異常リターンは減少しており、イベントデーが市場に影響を与えたとは考え難い。

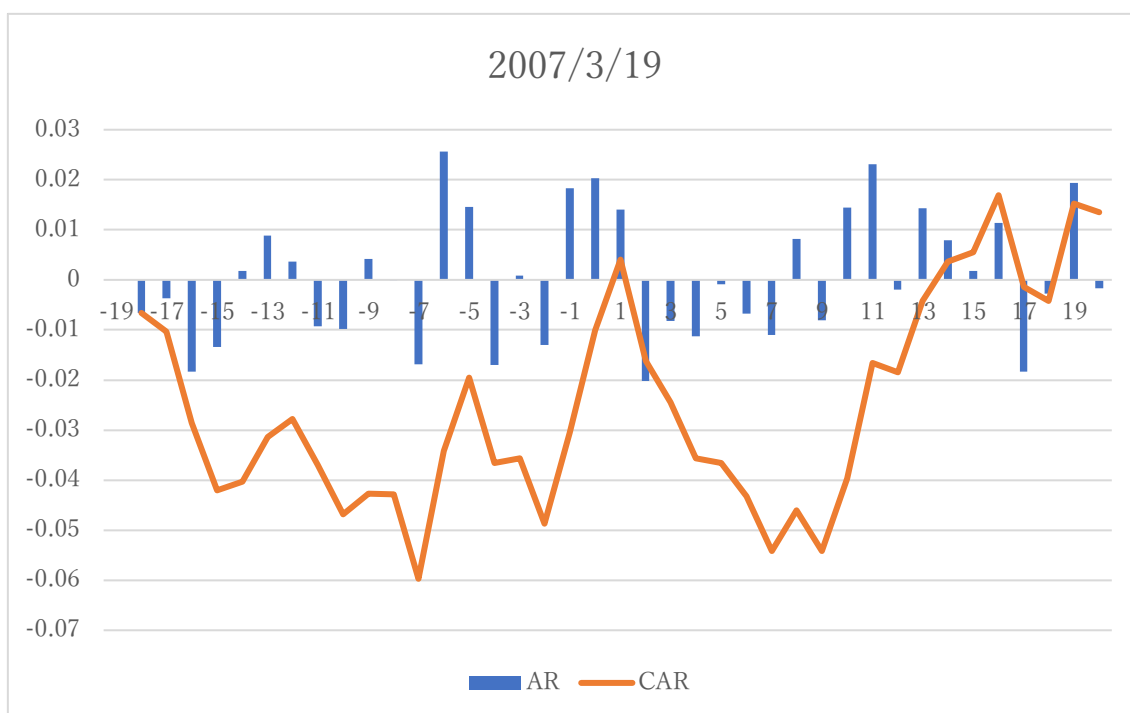


図 F 2007/03/19 のイベントスタディの結果

(A-7) 2007/03/22 の結果

以下の図 G に 2007/03/22 をイベントデーとしたイベントスタディの結果を示す。このイベントデーに観測されたトピックは、Topic0、Topic1、Topic2、Topic3、Topic4、Topic5、Topic6、Topic7、Topic8、Topic9、Topic11、Topic12、Topic13、Topic14、Topic18 である。イベントデー当日の異常リターンは-0.0077533 であり、前日は-0.0198664、翌日は-0.0108972 であった。前日から当日にかけての異常リターンは上昇しており、また当日から翌日にかけての異常リターンは減少している。このことからイベントデーが起因となって市場に好ましい影響を与えた可能性が高い。しかし、このイベントデーで観測されたトピックは 15 個に分かれており、どのようなトピックが市場に影響を与えたかを特定することは難しい。

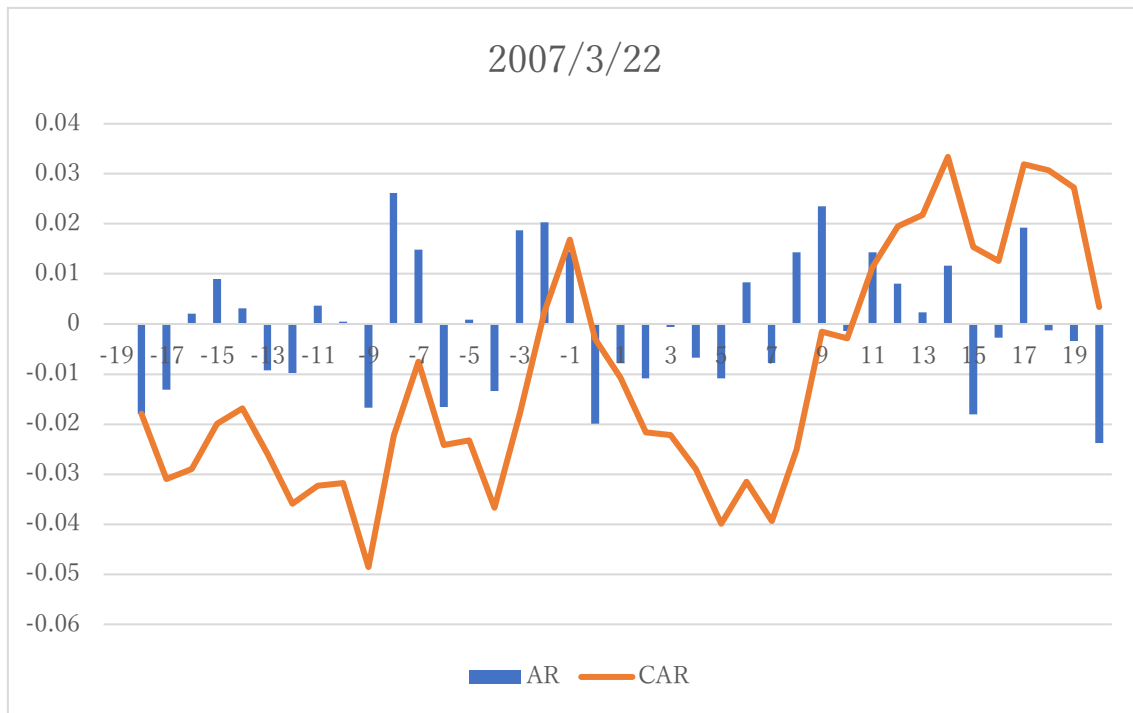


図 G 2007/03/22 のイベントスタディの結果

(A-8) 2007/03/23 の結果

以下の図 H に 2007/03/23 をイベントデーとしたイベントスタディの結果を示す。このイベントデーに該当したトピックは Topic4 と Topic9 である。イベントデー当日の異常リターンは-0.0110028 であり、前日は-0.0078649、翌日は-0.000746 である。前日から当日にかけては異常リターンが下がっており、当日から翌日にかけては異常リターンの値が上昇している。このことから、Topic4、Topic9 のいずれかのトピックが市場から好ましい反応を誘因した可能性がある。

また、(A-1)の結果では同じ Topic4 と Topic9 のみが該当していたにも関わらず、(A-1)の結果では異常リターンは下がる傾向にあり、結果の相違が見られる。この 2 つの結果より、Topic4 と Topic9 は不確定要素が多く、市場から好ましい反応が返ってくるのか、あまり好ましくない反応が返ってくるのかが時と場合によって左右されるトピックである可能性が存在することがわかった。

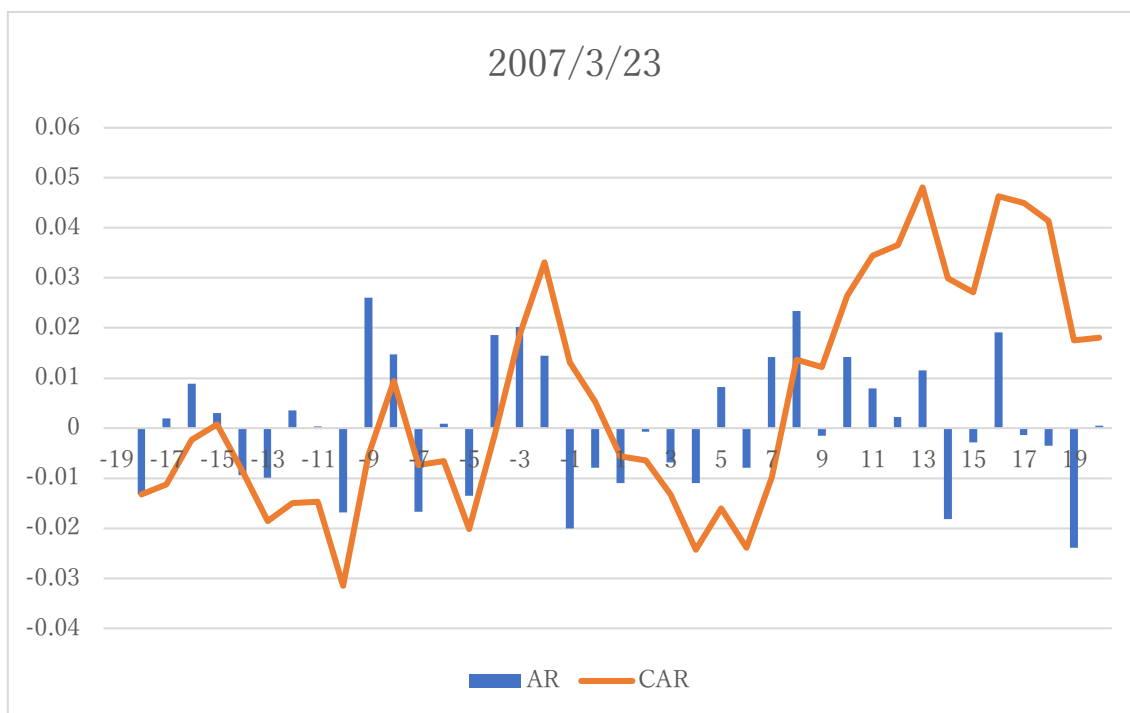


図 H 2007/03/23 のイベントスタディの結果

(A-9) 2007/03/30 の結果

以下の図 I に 2007/03/30 のイベントスタディを行った結果を示す。

2007/03/30 で観測されたトピックは Topic1、Topic2、Topic4、Topic9 である。イベントデー当日は-0.008572、前日は 0.00769276、翌日は 0.01470072 である。前日から当日にかけての異常リターンは減少しており、当日から翌日にかけての異常リターンは上昇している。つまりイベントデーを起因となって市場に影響を与えた可能性がある。このことから、Topic1、Topic2、Topic4、Topic9 のいずれかのトピックが市場に好ましい影響を与えた可能性が高いことがわかる。

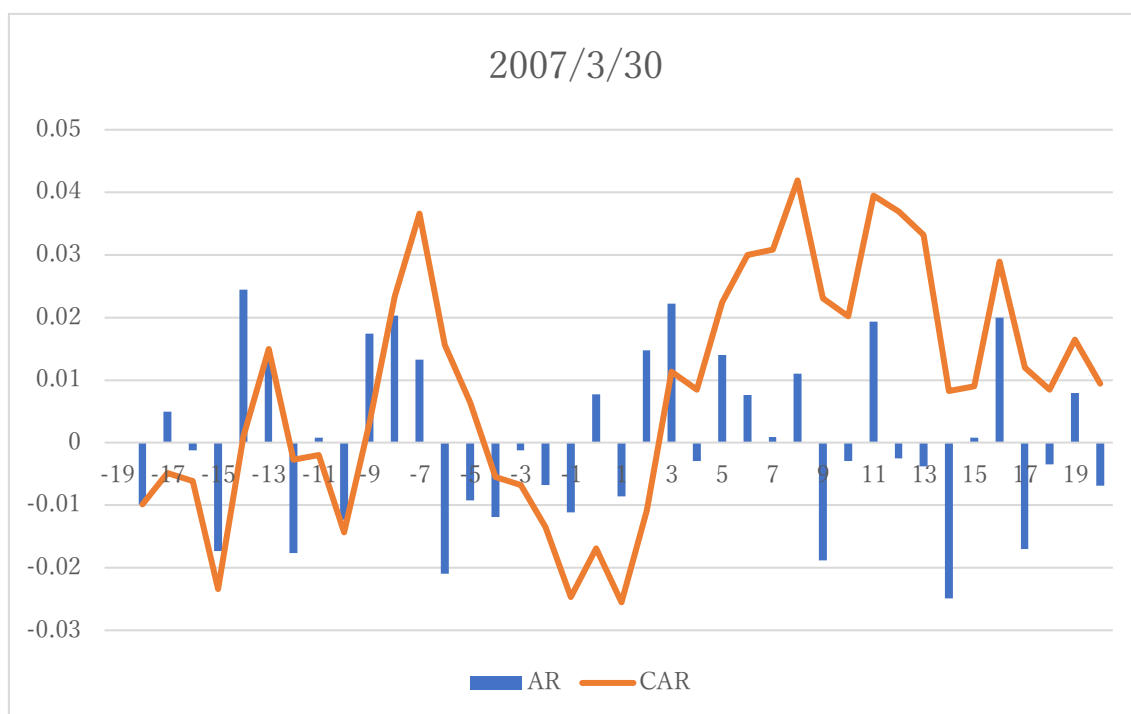


図 I 2007/03/30 のイベントスタディの結果