

有価証券報告書のテキスト分析： 経営者による将来見通しの開示と 将来業績

かとうだいすけ ごしまけいいち
加藤大輔／五島圭一

要 旨

本稿は、有価証券報告書の「企業の経営方針・経営戦略や経営者による経営成績の分析（Management's Discussion and Analysis: MD&A）」に含まれるテキスト情報の定量化を通じて、経営者によって開示された将来見通しが、将来の企業業績に対する予測力を有することを明らかにしている。これは、経営者が自主的に将来見通しを開示する媒体として、MD&A が一定の役割を果たしていることを示している。また、同予測力については、業種による差異が確認されたほか、売上高の小さい企業ほど高くなる傾向も見出されている。この結果は、ディスクロージャーの充実度が企業の属性によって異なりうることを示している。さらに、2018年に実施されたMD&Aの章立て変更に対応して、同予測力が、サービス業を営む企業において、他の業種に比べて高まったことも見出された。

キーワード： テキスト・マイニング、有価証券報告書、ディスクロージャー

.....
本稿の作成に当たっては、宮川大介准教授（一橋大学）、古澤知之氏（金融庁）、井上俊剛氏（金融庁）、原野浩氏（金融庁）、関根敏隆教授（一橋大学）、稲葉圭一郎氏（日本経済研究センター）、ならびに金融研究所スタッフから有益なコメントを頂いた。そして、本稿で使用した Loughran and McDonald Sentiment Word Lists の日本語訳においては、須藤直氏（日本銀行）に多大なご協力を頂いた。ただし、本稿に示されている意見は、筆者たち個人に属し、日本銀行の公式見解を示すものではない。また、ありうべき誤りはすべて筆者たち個人に属する。

加藤大輔 日本銀行金融研究所
（現みずほりサーチ&テクノロジーズ株式会社、
E-mail: daisuke1.kato@mizuho-ir.co.jp）

五島圭一 早稲田大学商学部講師、日本銀行金融研究所
（現東京大学エコノミックコンサルティング、早稲田大学産業経営
研究所招聘研究員、E-mail: keiichi@utecon.net）

1. はじめに

有価証券報告書の「経営者による財政状態及び経営成績の検討と分析」(Management's Discussion and Analysis: MD&A)は、上場企業の経営者が投資家と対話をする重要な場である。MD&Aは企業の経営成績や財務状態等の数値情報とともに、事業全体・セグメント単位での自社を取り巻く経営環境や経営成績に重要な影響を与えた要因、設定した重要業績評価指標の達成状況等について、経営者の視点による自社の経営成績に関する分析・検討内容が、非財務情報(記述情報)の形で記載されている。この情報は、開示企業の株主や債権者にとって重要である。なぜならば、当該企業の経営方針や経営戦略の意義や適切さの確認が可能になるほか、経営者自身の認識は当該企業の将来業績を予測するうえで有益な判断材料となるからである。例えば、MD&Aにおける資本調達やキャッシュ・フローに関する非財務情報は、株主や債権者にとっては、当該企業の企業金融、設備投資、および研究開発の現状の把握や、それらを踏まえての将来業績の予測に役立つものと考えられる。

2003年より、すべての上場企業が有価証券報告書にMD&Aの記載を行うことが義務付けられているが、その内容については一定のルールの範囲内で各企業の裁量に委ねられており、いわば半自発的なディスクロージャーという側面がある。現在、わが国において法令上記載が求められていることは、「経営者の視点による当該経営成績等の状況に関する分析・検討内容を具体的に、かつ、分かりやすく記載すること」と、「その際、事業全体及びセグメント情報に記載された区分ごとに、経営者の視点による認識及び分析・検討内容を、経営方針・経営戦略等の内容のほか、有価証券報告書に記載した他の項目の内容と関連付けて記載すること」である。したがって、財務情報と比べると、各企業にとって記載内容の自由度は高い。

企業が開示するMD&Aをはじめとする非財務情報の重要性は、金融・資本市場において増している。2015年6月に適用が始まったコーポレートガバナンス・コードの原則の1つに、「適切な情報開示と透明性の確保」が挙げられており、企業と株主との建設的な対話のために、非財務情報について、法令に基づく開示以外の情報提供にも主体的に取り組むべきとされている(東京証券取引所[2018])。そして、有価証券報告書におけるMD&Aの位置付けを拡充する施策が実行されてきた。これに伴って、MD&Aにおける定性的なディスクロージャー内容が果たしている役割を実証的に解明する必要性もまた高まっているといえる。この必要性に応えることは、金融市場インフラの重要な構成要素である有価証券報告書の意義に対する理解の向上に資するものと考えられる。

MD&Aにある定性的な非財務情報を実証的に分析するにあたっては、テキストの定量化が不可欠である。初期の研究においては、ごく限られた数のテキストを人

間が実際に読み込んだうえで、MD&A テキストに含まれるトピックの種類や、人間が付与したポジティブ・ネガティブ等を示すスコアが、アナリストによる将来の企業業績の予測や、将来において実現した企業業績といかに関連しているかが分析されていた (Pava and Epstein [1993]、Bryan [1997]、Barron, Kile, and O'Keefe [1999])。

近年の研究では、コンピュータや分析手法の発展によって人手による読解を介さずに分析が行われるようになったことから、分析に用いられるテキストのサンプルが大幅に増加している。こうした研究においては、テキストの定量化は、主として、記述量 (単語数、文字数) や読みやすさ、そしてトーンの算出を通じてなされており、それらと企業業績あるいは株式リターンとの関連が分析されている。ここでトーンとは、何らかの評価軸に対する記述内容の傾向を、ポジティブ・ネガティブの度合で示すものであり、企業が公表するテキストでは、企業業績あるいは株式リターンが重要な評価項目となる。

ファイナンスおよび会計研究の分野において、企業が公表するテキストの分析は、米国企業を対象としながら先行してきた。まず、記述量や読みやすさに関しては、Li [2008] は、1994 年から 2004 年に発行された米国の年次報告書に当たる Form 10-K を対象に、単語数および FOG 指数と呼ばれる読みやすさの指標を分析し、利益の低い企業ほど開示情報が読みにくいことや、読みやすい Form 10-K を開示している企業は利益の持続性が高いことを報告した¹。また、Loughran and McDonald [2014] は、1994 年から 2011 年に公開された米国の有価証券報告書に当たる Form 10-K のテキストを分析し、読みやすさの指標としては FOG 指数よりもファイルの容量のほうが優れていることを示した。さらに、Muslu *et al.* [2015] は、1993 年から 2009 年に公開された Form 10-K の MD&A に含まれるフォワード・ルッキングな記述の割合を分析し、株価に将来業績の情報があまり反映されていないときに MD&A のフォワード・ルッキングな記述の割合が増えることを示した。最後に、Lo, Ramos, and Rogo [2017] は、FOG 指数を用いて Form 10-K を分析し、前年度に利益調整を行った可能性の高い企業の MD&A の記述が複雑になりやすい傾向を見出した。

次に、米国企業のテキストのトーン分析については、Li [2010] が先駆的である。彼は、1994 年から 2007 年に発行された Form 10-K と米国の四半期報告書に当たる Form 10-Q の MD&A に記載されている非財務情報のトーンを、ナイーブ・ベイズ分類器を用いて算出したうえで、企業の財務情報や属性が企業の発行する MD&A のトーンに影響を与えていること、そして、トーンと将来業績の間に正の相関のあることを明らかにした。また、Loughran and McDonald [2011] は Form 10-K のテキストの分析を行い、Harvard Dictionary でネガティブに分類されている単語の約 4 分の 3 がファイナンス文書においてはネガティブな意味を持たないことを示した。そ

.....
1 FOG 指数の詳細は補論 2 に記した。

して、Form 10-K においてポジティブおよびネガティブを示す単語を含む極性辞書を独自に作成した。

わが国企業の有価証券報告書に関するテキスト分析は、記述量や読みやすさに関する研究が先行している。廣瀬・平井・新井 [2017] は Li [2008] と同様の手法を用いて、日本の有価証券報告書の MD&A 部分の読みやすさと総文字数を分析した。その結果、2003 年の MD&A 開示の制度化以降、文書の難易度の低下と総文字数の減少が進んでおり、MD&A の可読性が高まっていることを示した。また、時価総額が大きく特別損益を計上している企業で文章の難易度が高く、上場年数の長い企業で総文字数が少ないことや、総文字数が多いと将来業績が悪化することを示した。また、吉田 [2018] は、有価証券報告書の「【事業等のリスク】」の文章の長さを分析し、収益性の高い企業では多くのリスク情報を開示するほど高い収益性が継続し、収益性の低い企業では多くのリスク情報を開示するほど将来の収益性が小さくなることを報告した。

このように、わが国企業の有価証券報告書に関するテキスト分析については、米国における研究の流れを引き継ぐもとで、将来業績との関連性が分析されているものの、トーンの分析は筆者たちの知る限り、まだ存在しない。本稿の最大の貢献は、テキスト・マイニングの手法を用いて、このギャップを埋めることである。すなわち、2014 年から 2019 年にかけての延べ約 2 万本の有価証券報告書の MD&A テキストのトーンを算出し、その決定要因や、将来の企業業績に対する予測力の有無を分析している。この予測力が、MD&A におけるディスクロージャーの充実度を表すことを議論したうえで、同予測力には業種別あるいは売上規模別にみて差異があることや、MD&A に含まれるテキスト情報の拡充を目指す規制変更を受けて同予測力がいかに変化したかについて新たな知見を報告している。また、本稿が用いたテキスト分析手法の有用性もサポートされている。

本稿の構成は以下のとおりである。2 節では分析手法を、3 節ではデータを説明する。4 節では得られた結果を示し、5 節では得られた結果をまとめるとともに今後の課題を示す。

2. 分析手法

(1) トーンの計算

本稿は、MD&A に記載されている非財務情報のトーンを算出するために、極性辞書によるアプローチを採用する。このアプローチでは、特定の単語の極性情報（ポジティブあるいはネガティブ）を事前に定義し、文章に含まれるすべてのポジティブおよびネガティブな単語の数を数え上げることで、文章のトーンが計測される。具体的には、(1) 式によってトーン ($TONE_{i,t}$) を算出する。

$$TONE_{i,t} = \frac{N_{i,t}^{positive} - N_{i,t}^{negative}}{N_{i,t}^{positive} + N_{i,t}^{negative}}. \quad (1)$$

ここで、 $N_{i,t}^{positive}$ 、 $N_{i,t}^{negative}$ はそれぞれ、 t 年度における企業 i の MD&A の文章に含まれるポジティブ・ネガティブの単語数である。定義により、算出されるトーンの値は ± 1 の範囲に収まる。

本稿で用いる極性辞書は、Loughran and McDonald [2011] が作成した英語の辞書である Loughran and McDonald Sentiment Word Lists（以下、LM 辞書）を日本語に訳したものである。LM 辞書は Form 10-K のテキスト分析を行うために開発された辞書であり、ファイナンスや会計に関連するポジティブおよびネガティブを示す単語を含んでいるのが最大の特徴である。この辞書に含まれる単語数はポジティブ 354 語、ネガティブ 2,355 語の合計 2,709 語であるが、同じ単語の語形変化したものを含んでおり、日本語へ訳したときに重複するものが多い²。これらの重複を削除した結果、日本語訳した辞書に含まれる単語数はポジティブ 255 語、ネガティブ 1,374 語の合計 1,629 語となった。

(2) トーンの将来予測成分を用いた MD&A ディスクロージャーの分析

本稿は、Li [2010] にならって、MD&A において開示されている経営者による将来見通しのトーンに注目する。本稿では、MD&A の記述全体のトーンを計測し、そこから現状において入手可能な情報によって説明される成分を控除し、得られる

.....
2 例えば、LM 辞書には単数形と複数形がそれぞれ収録されているが、日本語へ訳したときには同一の単語となる。

残余部分を経営者による定性的な将来見通しに由来する成分とみなす（以下、補正済トーン）。

こうした定式化は、概念的には以下のとおりである。まず、ある企業が時点 t において公表する MD&A のトーンは、時点 t において入手可能な情報と、経営者による将来見通しによって規定されると考える。ここで、前者は時点 t における企業業績を含む財務情報や企業の属性等の定量的な情報であり、後者は経営者だけが把握できる非公開情報や主観に基づく定性的な情報である。後者に含まれる非公開情報は、新商品の開発や企業買収といった情報を含み、現時点の企業業績には表れてはいないものの、将来の企業業績を左右しうる。したがって、定性的な将来見通しの代理変数である補正済トーンは、時点 t において、翌時点 $t+1$ における企業業績の予測を手助けするはずである。なお、本稿では企業業績の指標として、株主のみならず債権者の視点も考慮する観点から、総資産利益率（Return On Assets: ROA）に注目している³。具体的には、時点 t における補正済トーンは、同時点におけるトーン（ $TONE_{i,t}$ ）を目的変数とし、時点 t に入手可能な定量的な指標（ $X_{i,t}^{Now}$ ）を説明変数として最小二乗法による回帰分析によって得られる残差のことであり、次の (2) 式における $TONE_{i,t}^{ADJ}$ である。

$$\begin{aligned} TONE_{i,t} &= \alpha + \beta X_{i,t}^{Now} + \epsilon_{i,t}, \\ TONE_{i,t}^{ADJ} &= TONE_{i,t} - T\hat{O}NE_{i,t}. \end{aligned} \quad (2)$$

ここで、 i は個別企業、 t は年度、 α は切片、 β は偏回帰係数ベクトル、 $\epsilon_{i,t}$ は誤差項を表す。また $T\hat{O}NE_{i,t}$ は (2) 式の回帰式で求められるトーンの推定値である。 $X_{i,t}^{Now}$ を構成するベクトル成分の変数の選択は、Li [2010] を参考にしており、表 1 に記載のとおりである。

なお、Li [2010] では、英文の読みやすさの指標を説明変数に加えており、本稿もこれに倣っている。具体的には、「日本語文章難易度判別システム（jReadability）」で提案されている読みやすさ指数を採用している（Sunakawa, Lee, and Takahara [2012]、李 [2016]）。なお、読みやすさ指数の詳細については、補論 2 を参照されたい。前述のとおり、補正済トーンは経営者による将来見通しの代理変数であるため、翌期の ROA（ $ROA_{i,t+1}$ ）の決定要因の 1 つとなるはずである。そこで、次の (3) 式の推定を通じて MD&A のトーンの将来の企業業績に対する予測力を検証する。

$$ROA_{i,t+1} = \alpha^* + \beta^* X_{i,t}^{Now} + \gamma^* TONE_{i,t}^{ADJ} + \epsilon_{i,t}^*. \quad (3)$$

3 ROA は経常利益総資産比率。国際会計基準（IFRS）または米国会計基準（USGAAP）を導入している企業については、「経常利益」の代わりに「営業利益」または Bloomberg が提供する「営業損益」を用いた。総観測数 20,800 のうち、IFRS、USGAAP が採用されている観測数はそれぞれ 733、120 で、それ以外（19,947 サンプル）ではすべて日本会計基準が採用されている。なお、Li [2010] において ROA は当期純利益総資産比率で計算されている。

表 1 変数一覧

変数	名称	定義
TONE	トーン	計算方法は 2 節 (1) を参照
TONE ^{ADJ}	補正済トーン	計算方法は 2 節 (2) を参照
ROA	ROA	(経常利益) / (総資産)
RET	株式リターン	過去 1 年の株式リターン
ACC	会計発生高	{ (経常利益) - (営業 CF) } / (総資産)
SIZE	時価総額	$\log_{10}\{ (\text{会計期末株価}) \times (\text{発行株式数}) \}$
MTB	時価簿価比	{ (株式時価総額) + (負債) } / (総資産)
RETVOL	株式リターン・ボラティリティ	過去 1 年の株式リターンの標準偏差
ROAVOL	ROA ボラティリティ	過去 5 年の ROA の標準偏差
NBSEG	事業セグメント数	$\ln\{1 + (\text{事業セグメント数})\}$
NGSEG	地域セグメント数	$\ln\{1 + (\text{地域セグメント数})\}$
AGE	上場年数	上場年数
SI	特別利益・損失	{ (特別利益) + (特別損失) } / (総資産)
READ	読みやすさ指数	計算方法は補論 2 を参照

ここで、 i は個別企業、 t は年度、 α^* は切片、 β^* は偏回帰係数ベクトル、 $\epsilon_{i,t}^*$ は誤差項を表す。推定の結果、 γ^* が統計的に有意に正であれば、当該企業の経営者が将来見通しを MD&A のテキスト情報を通じて開示していることを意味する。また、 γ^* の値が大きいほど、すなわち補正済トーンの限界的なインパクトが大きいほど、翌期 ROA 予測の手助けになるという意味で、その開示内容がより充実していることになる。したがって、本稿では係数 γ^* の有意性と値の大きさを MD&A におけるディスクロージャーの充実度を示す指標とする。

なお、Li [2010] では、MD&A から将来予測に関する表現 (Forward-Looking Statements) だけを抜き出してトーンを算出している。しかし、将来見通しを述べた文章だけが、経営者による将来見通しを反映しているとは限らない。例えば、過去の業績を分析・評価する記述についても、経営者が将来業績に対して楽観的である場合と悲観的である場合では、その記述のトーンは異なる可能性が高い。したがって、本稿では MD&A の全テキストから回帰分析によって、現状由来成分を差し引くことでトーンの将来見通しを抽出した。

(3) 企業属性や章立て変更による影響の検証

本稿は、さらに、経営者が MD&A を通じて行う、将来見通しに関する定性的なディスクロージャーの充実度と、他の要因との相互作用を分析する。すなわち、

(3) 式に、補正済トーンと交差する複数の変数（交差項）を追加し、重回帰分析を行う。

イ. トーンと企業の属性

1 節で述べたように、有価証券報告書を発行している企業にとって、MD&A における記載内容の自由度は財務情報と比べて高い。こうした条件下で、経営者が非公開情報や主観に基づく、定性的な将来見通しをより開示しようと思うのはどのような場合においてであろうか。このような情報のディスクロージャーの充実度は、企業経営者の裁量によるものであり、企業統治や企業金融の分野において重要なエージェンシー問題と直結するトピックであると思われる。そこで、同充実度への理解を深めるために、(3) 式に、企業属性を制御するダミー変数と補正済トーンの交差項を挿入する。ダミー変数は2つあり、それぞれ業種と売上に関連する。より大きな不確実性に晒されている業種に属する企業の経営者や、売上規模が小さい部類の企業の経営者は、先々において、資金繰りの問題に直面するリスクがより大きいものと考えられる。したがって、同充実度が高くなる、すなわちトーンの将来業績予測力が追加的に高まるかもしれない。また、エージェンシー問題とは別に、非競争的な市場において活動する企業、すなわち何らかの規制産業に属する企業の経営者にとっては、将来はある程度見通しが利くものと考えられる。このため、こうした業種の企業経営者は、将来見通しに関する定性的なディスクロージャーをより簡単に充実化できるかもしれない。

ロ. トーンと章立て変更

MD&A については、2016 年 4 月に公開された金融審議会「ディスクロージャーワーキング・グループ報告」において、有価証券報告書の非財務情報の開示充実を目的として、複数箇所に分散している MD&A の記述の整理・統合と、記述内容のさらなる拡充を求める提言がなされた（金融庁 [2016]）。これを受けて、2018 年 1 月に「企業内容等の開示に関する内閣府令」等の改正が行われ、2018 年度より発行される有価証券報告書について、記載内容の整理を目的とした章立ての変更に加えて、「事業全体及びセグメント別の経営成績等に重要な影響を与えた要因について経営者の視点による認識及び分析」、「経営者が経営方針・経営戦略等の中長期的な目標に照らして経営成績等をどのように分析・評価しているか」という2点を記載することが求められることとなった。2018 年の MD&A の章立て変更の政策効果はいかなるものだったのだろうか。同制度改革が経営者に対して定性的な将来見通しのディスクロージャーの充実化を促した、という政策効果は、当局者の期待するところであろう。そうした促進効果の有無を検証するために、(3) 式に、章立て変更の前後を制御するダミー変数と補正済トーンの交差項を挿入する。同促進効果があれば、章立て変更後、トーンの将来業績予測力は高まっていたはずである。

ハ. 章立て変更と企業の属性

章立て変更による影響は、企業の属性によって異なる可能性がある。例えば、同施策に対して、前述のとおり、さまざまなリスクや不確実性により晒されている業種に属する企業の経営者は、先々の資金調達の円滑化のためにより積極的な対応を行う可能性がある。また、規制産業に属する企業の経営者は、規制変更に対してより機敏に対応する可能性がある。そこで、本稿では、2節(3)ロ. で用いたモデルに企業の属性を示すダミー変数と章立て変更ダミーと補正済トーンの交差項を導入した重回帰分析を行う。もし、交差項の係数が有意であれば、章立て変更による影響が企業の属性によって異なることが示唆される。

3. データ

(1) データ

本稿では、2014年5月14日より2019年6月30日の間に発行された有価証券報告書を用いた。発行企業数は約3,400社、発行本数は約2万本である。有価証券報告書のデータは金融庁のEDINET（Electronic Disclosure for Investors' NETwork）より取得した。EDINETから有価証券報告書を取得する手順の詳細については、補論1を参照されたい。財務データについてもEDINETより取得した有価証券報告書より取得し、有価証券報告書に記載のない財務データおよびマーケット・データについてはBloombergより取得した。

(2) テキスト分析の手順

EDINETより取得した有価証券報告書のテキスト・データはhtm形式となっている。本稿では、htm形式のテキスト・ファイルからMD&A部分を抜き出し、htm形式のタグ除去等の処理を行って生のテキストを抽出した。処理後のテキストについて、MeCabを用いた形態素分析によって文章を単語に分割し、2節で説明した手法でトーンを算出した⁴。

本稿が目指すMD&A記載部分の章立て変更は、次のように2段階で行われた。すなわち、第1段階では2017年4月より、「第2【事業の状況】」の「3【対処すべ

4 「MeCab」：<https://taku910.github.io/mecab/>（最終閲覧日：2021年4月12日）

き課題】」の見出し名が、「3【経営方針、経営環境及び対処すべき課題等】」に変更された。そして、第2段階では2018年4月より、上記の改正による章立て変更が実施された。本稿では、章立て変更前後の同じ部分のテキストを分析対象とするため、(i)2017年度以前の章立て変更前については、「1【業績等の概要】」、「2【生産・仕入、受注及び販売の状況】」、「3【対処すべき課題】」、「7【財政状態、経営成績及びキャッシュ・フローの状況の分析】」の4つの章のテキストをMD&Aの記述とみなし、(ii)2017年度については上記(i)の「3【対処すべき課題】」の代わりに「3【経営方針、経営環境及び対処すべき課題等】」を分析対象に含め、(iii)2018年度以降の章立て変更後については、「1【経営方針、経営環境及び対処すべき課題等】」と「3【経営者による財政状態、経営成績及びキャッシュ・フローの状況の分析】」の2つの章のテキストをMD&Aの記述とみなして分析を行った。

(3) データの統計量

本稿の推計に用いる各変数の記述統計量を表2に、変数間の相関係数行列を表3にそれぞれ示す。このうち、本稿にとって関心が高いトーンとROAの関係を図1および図2に示す。図1が示すように、トーンがポジティブなほどROAも大きくなる傾向があり、特にROAが負の場合において、トーンはネガティブ側に多く分布している。また、図2にトーンと翌期のROAの分布を示す。図2が示すように、トーンと翌期のROAを比べても同じ期のROAと同じ傾向がみられる。

(4) 補正済トーンの計算

MD&Aのトーンを被説明変数とした重回帰分析の結果を表4に示す。Li [2010]と同様の推定方法である年ダミーのみを考慮したPooled OLS推定(2列目)と、年ダミーに加え、企業の固定効果を加味した最小二乗ダミー変数(Least Squares Dummy Variables: LSDV)推定の結果(3列目)を示す。両者を比較するとLSDVモデルのほうが自由度調整済み決定係数が大きく、またF検定でもPooled OLSモデルが棄却されることから、トーンからの補正済トーンの抽出は、LSDVモデルの推定を通じて行う。

表 2 変数の記述統計量

変数名	平均値	P5	P25	中央値	P75	P95	標準偏差
TONE	0.12	-0.34	-0.05	0.13	0.30	0.52	0.26
ROA	0.06	-0.02	0.03	0.05	0.09	0.17	0.10
RET	0.39	-0.44	-0.12	0.07	0.31	1.55	1.67
ACC	0.00	-0.09	-0.03	0.00	0.03	0.10	0.07
SIZE	10.35	9.31	9.80	10.25	10.83	11.83	0.88
MTB	1.54	0.61	0.84	1.00	1.41	3.60	3.35
RETVOL	0.53	0.17	0.26	0.35	0.51	1.67	0.69
ROAVOL	0.03	0.00	0.01	0.02	0.03	0.10	0.09
NBSEG	1.61	0.69	1.39	1.61	1.95	2.56	0.58
NGSEG	1.13	0.69	0.69	0.69	1.61	2.20	0.66
AGE	28	1	11	21	49	66	20.97
SI	0.02	0.00	0.00	0.00	0.01	0.06	0.07
READ	9.03	8.43	8.84	9.07	9.26	9.50	0.33

表 3 変数間の相関係数行列

	ROA	RET	ACC	SIZE	MTB	RETVOL	ROAVOL	NBSEG	NGSEG	AGE	SI	TONE	READ
ROA		-0.02	0.14	0.01	0.28	-0.03	0.11	-0.02	0.05	-0.11	0.27	0.11	0.02
RET	-0.04		-0.06	-0.03	-0.03	0.53	-0.05	0.00	0.00	0.17	0.00	0.05	0.01
ACC	0.33	-0.04		-0.02	-0.05	0.07	0.04	0.00	-0.05	-0.03	-0.07	0.04	-0.02
SIZE	0.03	-0.02	-0.03		0.09	-0.02	-0.02	0.18	0.13	0.14	0.00	-0.05	0.04
MTB	0.50	-0.03	0.15	0.04		0.09	0.14	0.00	-0.09	-0.11	0.02	-0.01	-0.01
RETVOL	0.02	0.51	0.05	-0.05	0.12		0.07	-0.01	-0.01	-0.06	0.01	-0.01	0.02
ROAVOL	0.17	-0.04	0.11	-0.07	0.34	0.17		-0.01	0.04	-0.09	-0.05	-0.02	0.02
NBSEG	-0.04	0.00	-0.03	0.21	-0.04	-0.03	-0.06		0.09	0.04	0.06	0.06	-0.03
NGSEG	0.01	0.01	-0.05	0.18	-0.07	-0.04	-0.01	0.14		0.15	0.04	0.02	0.04
AGE	-0.22	0.17	-0.13	0.17	-0.26	-0.04	-0.25	0.11	0.18		0.04	-0.01	-0.12
SI	-0.13	-0.01	-0.11	0.01	0.02	0.02	0.11	0.06	0.05	0.02		-0.06	-0.03
TONE	0.34	0.06	0.15	-0.03	0.15	0.02	0.01	0.03	0.01	-0.06	-0.13		-0.05
READ	0.06	0.01	0.01	0.02	0.06	0.03	0.07	-0.04	0.02	-0.13	-0.01	0.01	

備考：対角線の下側は Pearson の相関係数、対角線の上側は偏相関係数を示す。

图 1 $TONE_t$ 对 ROA_t

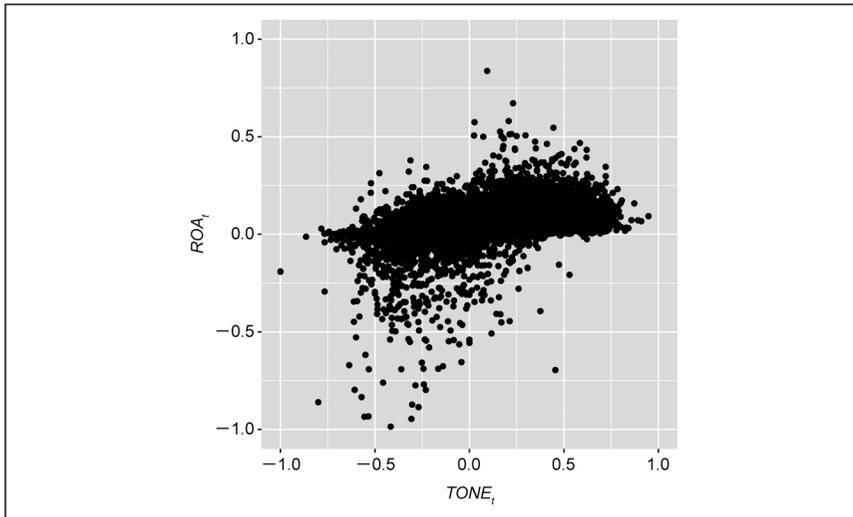


图 2 $TONE_t$ 对 ROA_{t+1}

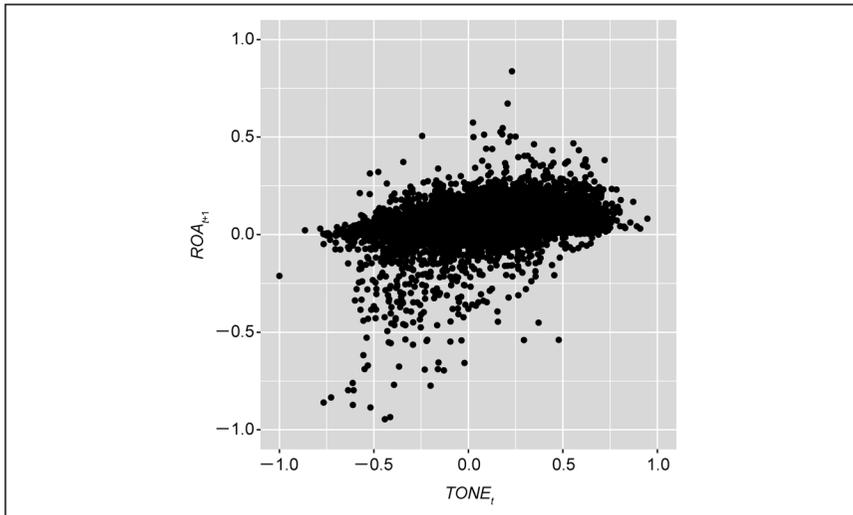


表4 トーンの決定要因

説明変数	Pooled OLS	LSDV	Li [2010]
<i>ROA</i>	1.334***	1.102***	0.212***
<i>RET</i>	0.005***	0.007***	0.038***
<i>ACC</i>	0.169***	0.086***	-0.139***
<i>SIZE</i>	-0.005**	-0.025***	-0.008***
<i>MTB</i>	-0.004**	-0.001	-0.010***
<i>RETVOL</i>	-0.009**	-0.007**	-0.418***
<i>ROAVOL</i>	0.075*	0.017	-0.006***
<i>NBSEG</i>	0.021***	—	0.018**
<i>NGSEG</i>	0.002	—	-0.015
<i>AGE</i>	-0.000**	0.015	0.004***
<i>SI</i>	-0.333***	-0.264***	-0.038***
<i>READ</i>	-0.048***	0.048***	-0.029***
観測数	18,322	18,322	105,846
自由度調整済み決定係数	0.23	0.28	0.21

備考：被説明変数はいずれも当期のトーンである。***、**、*はそれぞれ、両側確率1%、5%、10%で回帰係数が有意であることを示している。Pooled OLS は年のダミー変数を追加した Pooled OLS 推定、LSDV は企業と年のダミー変数を追加した LSDV 推定の結果である。*t* 値の算出には、企業と年についてクラスターロバスト標準誤差を利用している。LSDV 推定では、*NBSEG*、*NGSEG* は多重共線性により説明変数より除外された。Li [2010] は年のダミー変数を追加した Pooled OLS 推定で、*t* 値の算出には、企業と年についてクラスターロバスト標準誤差を利用している。

なお、本稿の Pooled OLS 推定で係数が有意に正となるのは、*ROA*、株式リターン (*RET*)、会計発生高 (*ACC*)、*ROA* ボラティリティ (*ROAVOL*)、事業セグメント数 (*NBSEG*) で、係数が有意に負となるのは、時価総額 (*SIZE*)、時価簿価比 (*MTB*)、株式リターン・ボラティリティ (*RETVOL*)、上場年数 (*AGE*)、特別損益 (*SI*)、読みやすさ指数 (*READ*) である。表4の第4列に示している Li [2010] による Pooled OLS 推定の結果と比較すると、係数の符号が一致する変数が多い。すなわち、*ROA*、株式リターン (*RET*)、時価総額 (*SIZE*)、時価簿価比 (*MTB*)、株式リターン・ボラティリティ (*RETVOL*)、事業セグメント数 (*NBSEG*)、特別損益 (*SI*) で符号が一致する⁵。

.....
5 読みやすさ指数 (*READ*) の係数の符号も一致しているが、FOG 指数は大きいほど難しい文章を指すのに対し、*jReadability* 指数は小さいほど難しい文章を指すため、実質的な符号は異なっている。

4. 分析結果

(1) トーンと将来業績

翌期の業績を被説明変数とした重回帰分析の結果を表5に示す。Pooled OLS 推定、LSDV 推定、一般化積率法 (Generalized Method of Moments: GMM) 推定の結果をそれぞれ第2、3、4列に示している⁶。有意性を判断するための t 値の計算では、企業と年についてクラスター・ロバスト標準誤差を利用している。また、比較のために、Li [2010] の結果を第5列に示している。

3種類の推定すべてについて、当期 ROA を含む企業の財務情報、属性、市場情報を調整した後においても補正済トーン ($TONE^{ADJ}$) 項の係数は有意に正であった。これらの結果は、MD&A のトーンが翌期 ROA の予測力を持っていることを示しており、企業が有価証券報告書の MD&A を通じて、経営者による将来見通しのディスクロージャーを行っている可能性を示唆する。そして、Li [2010] と整合的な結果である。

なお、3つの推定方法を比較すると、Pooled OLS モデルと LSDV モデルの間の F 検定では Pooled OLS モデルが棄却された。一方、この回帰分析では被説明変数である翌期の ROA (ROA_{t+1}) のラグ項 (ROA_t) が説明変数に含まれるために内生性の問題があり、LSDV モデルで得られた推定値では一致性が保たれないため、GMM モデルを用いることが望ましい。これ以降の交差項を用いた分析については、GMM 推定の結果に加えて、Li [2010] との比較のために Pooled OLS 推定の結果を示すが、解釈は GMM 推定の結果に依拠して行う。

.....
6 標準的な手法である二段階システム GMM 推定を Stata を用いて行ったところ、計算を収束させるためには説明変数の数を3つまで削減しなければならなかった。この弊害に加えて、そもそも有限標本下では二段階システム GMM 推定量の標準誤差は過小評価される傾向のあることが知られている (千木良・早川・山本 [2011])。これらを踏まえて、本稿では一段階システム GMM 推定量を採用している。この推定量については、効率性は保証されないものの、一致性は維持される。さらに、本稿では Bond, Leblebicioğlu, and Schiantarelli [2010] の手法によって、同推定量の信頼性を確認している。すなわち、動学パネル・データ分析において、ラグ項の係数について、Pooled OLS 推定量は高く見積もり、一方で LSDV 推定量は低く見積もることが知られている。したがって、もし GMM 推定量が両者の間となれば、信頼できるとおおまかに判断することができる、というものである。表5に示すように、GMM 推定によるラグ項の係数は Pooled OLS 推定量と LSDV 推定量による値の中間を取っており、Bond, Leblebicioğlu, and Schiantarelli [2010] の基準を採用すると、本稿の一段階システム GMM 推定の結果は信頼に足ると判断できる。

表5 将来業績とトーン

説明変数	Pooled OLS	LSDV	GMM	Li [2010]
<i>TONE^{ADJ}</i>	0.010***	0.010***	0.011***	0.006**
<i>ROA</i>	0.801***	0.116***	0.603***	0.679***
<i>RET</i>	0.001***	0.001***	0.001**	0.011***
<i>ACC</i>	-0.009	-0.002	-0.056***	-0.240***
<i>SIZE</i>	0.002***	0.005***	0.007***	0.002***
<i>MTB</i>	0.001**	0.001	-0.001	-0.002***
<i>RETVOL</i>	-0.003***	-0.001	-0.002**	-0.048***
<i>ROAVOL</i>	-0.036**	0.059	0.011	-0.016***
<i>NBSEG</i>	-0.001	—	-0.023	0.000
<i>NGSEG</i>	0.001	—	0.000	0.002***
<i>AGE</i>	0.000***	-0.003	0.000	0.000***
<i>SI</i>	0.087***	0.071***	0.136***	-0.422***
<i>READ</i>	0.001	0.004*	-0.006*	-0.001***
観測数		14,855		95,325

備考：被説明変数はいずれも翌期の ROA である。***、**、*はそれぞれ、両側確率 1%、5%、10% で帰帰係数が有意であることを示している。*TONE^{ADJ}* は補正済トーンを指す。Pooled OLS は年のダミー変数を追加した Pooled OLS 推定、LSDV は企業と年のダミー変数を追加した LSDV 推定、GMM は年のダミー変数を追加した一段階システム GMM 推定の結果である。*t* 値の算出には、企業と年についてクラスター・ロバスト標準誤差を利用している。LSDV 推定では、*NBSEG*、*NGSEG* は多重共線性により説明変数より除外された。Li [2010] は年のダミー変数を追加した Pooled OLS 推定で、*t* 値の算出には、企業と年についてクラスター・ロバスト標準誤差を利用している。

(2) トーンと企業の属性

MD&A における経営者による将来見通しのディスクロージャーの充実度が企業の属性によって異なる可能性を検証するため、4 節 (1) の重回帰分析の説明変数に以下のダミー変数と補正済トーンの交差項を追加した。

1. 業種ダミー：証券コード協議会による 10 業種（大分類）のうち、製造業を除く 9 業種についてダミーを設定する。例えば、商業ダミー (*DM_MER*) の場合、商業であれば 1、それ以外なら 0 を設定する。すべての業種ダミーの変数名は以下のとおりである。*DM_MER*：商業、*DM_TRA*：運輸・情報通信業、*DM_FIN*：金融業、*DM_SER*：サービス業、*DM_FAF*：農林・水産業、*DM_CON*：建設業、*DM_MIN*：鉱業、*DM_RES*：不動産業、*DM_ELG*：電気・ガス業。

表6 業種ダミーとトーン

説明変数	Pooled OLS	GMM
$TONE^{ADJ}$	0.002	-0.007
DM_MER	-0.002***	-0.020*
DM_TRA	0.002	-0.014
DM_FIN	-0.009***	-0.035**
DM_SER	0.004***	0.004
DM_FAF	-0.006	-0.011
DM_CON	0.004***	0.002
DM_MIN	0.000	-0.027***
DM_RES	-0.001	0.009
DM_ELG	-0.002	-0.011
$TONE^{ADJ} \times DM_MER$	0.008***	0.030***
$TONE^{ADJ} \times DM_TRA$	0.010***	0.059***
$TONE^{ADJ} \times DM_FIN$	0.004	0.004
$TONE^{ADJ} \times DM_SER$	0.012***	0.036***
$TONE^{ADJ} \times DM_FAF$	0.005	0.048
$TONE^{ADJ} \times DM_CON$	-0.003	-0.003
$TONE^{ADJ} \times DM_MIN$	-0.005	-0.008
$TONE^{ADJ} \times DM_RES$	0.015***	0.009
$TONE^{ADJ} \times DM_ELG$	0.006	0.038**

コントロール変数 = Yes

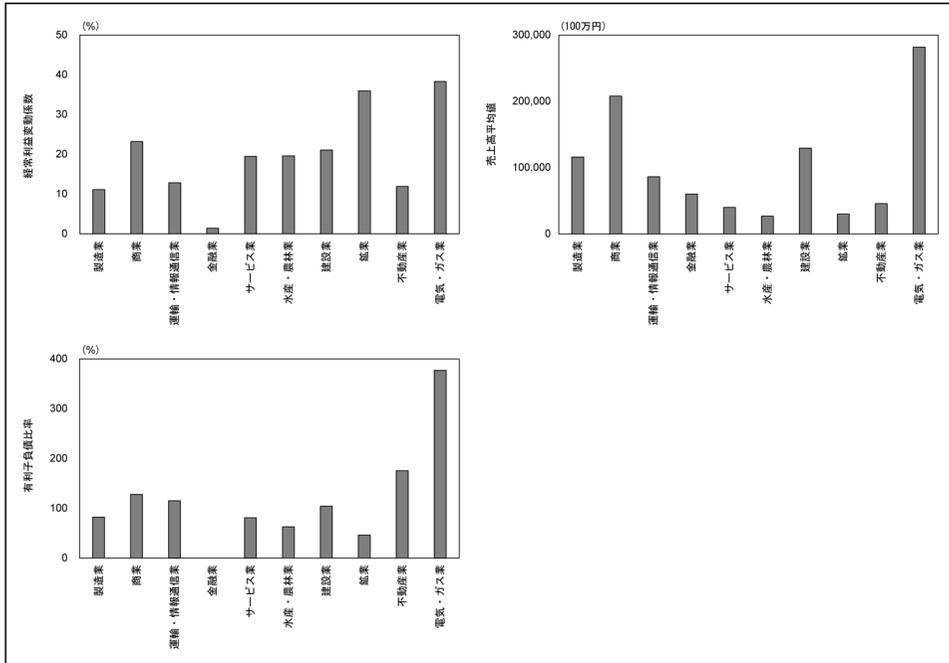
備考：被説明変数はいずれも翌期の ROA である。***、**、*はそれぞれ、両側確率 1%、5%、10% で回帰係数が有意であることを示している。 $TONE^{ADJ}$ は補正済トーンを指す。Pooled OLS は年のダミー変数を追加した Pooled OLS 推定、GMM は年のダミー変数を追加した一段階システム GMM 推定の結果である。“DM” で始まる項はそれぞれ以下の業種ダミー： DM_MER ：商業、 DM_TRA ：運輸・情報通信業、 DM_FIN ：金融業、 DM_SER ：サービス業、 DM_FAF ：農林・水産業、 DM_CON ：建設業、 DM_MIN ：鉱業、 DM_RES ：不動産業、 DM_ELG ：電気・ガス業。製造業をベンチマークとして推定した。観測数はいずれの結果でも 14,855 であった。コントロール変数の結果は省略している。なお、Pooled OLS モデルと LSDV モデルの間の F 検定では Pooled OLS モデルが棄却された。

2. 売上高ダミー ($DM_NETSALE$)：売上高が閾値を超える企業に 1、それ以外に 0 を設定する。閾値を 10 億円から 1 兆円まで 7 通りに変えて分析を行う。

イ. トーンと業種

業種ダミーと交差項を加えた重回帰分析の結果を表 6 に示す。補正済トーン項の係数は有意ではなく、交差項のうち、商業 (DM_MER)、運輸・情報通信業 (DM_TRA)、サービス業 (DM_SER)、電気・ガス業 (DM_ELG) の各ダミーの交差

図3 業種別の1社当たりの経常利益変動係数（左上）、売上高平均値（右上）、および有利子負債比率平均値（左下）



備考：売上高10億円以上の企業について、2014年度から2018年度の1社当たりの経常利益変動係数、売上高平均値、有利子負債比率平均値を業種別に集計。

資料：『法人企業統計調査』（財務省）

項の係数が GMM 推定で有意に正となった。10 業種のうち、製造業についてはダミーを追加していないため、補正済トーン ($TONE^{ADJ}$) 項の係数は製造業を営む企業の補正済トーン項の係数を示し、各業種ダミーの交差項の係数は、製造業を営む企業に対する補正済トーン項の係数の相対的な差を示している。よって、これらの結果は、交差項が有意に正である商業 (DM_MER)、運輸・情報通信業 (DM_TRA)、サービス業 (DM_SER)、電気・ガス業 (DM_ELG) の 4 業種を営む企業が公表する MD&A の補正済トーンが翌期 ROA の予測力を持っていることを示している。

この結果は、そうした企業が MD&A を通じて将来見通しのディスクロージャーを行っている可能性を示唆する。上記の 4 業種のうち、運輸・情報通信業と電気・ガス業は、規制産業と称される業種であり、競争的な製造業に比べて、将来見通しを立てやすいものと考えられる。一方で、図3をみると、上記の 4 業種すべてで製造業と比べて経常利益の変動係数が大きくなっており、相対的に不確実性が強い業種となっている。また、サービス業を除く 3 業種では、製造業と比べて有利子負債比率が高くなっており、相対的に他人資本依存が強い業種となっている。

表7 売上高ダミーとトーン

Pooled OLS							
説明変数	10 億円	30 億円	100 億円	300 億円	1,000 億円	3,000 億円	1 兆円
$TONE^{ADJ}$	0.005	0.013***	0.013***	0.012***	0.011***	0.010***	0.010***
$DM_NETSALE$	0.008***	0.005***	0.003***	0.000	-0.002***	-0.003***	-0.005***
$TONE^{ADJ} \times DM_NETSALE$	0.005	-0.004	-0.005**	-0.004***	-0.005***	-0.005***	-0.003
$N(DM_NETSALE=1)$	13,903	13,246	10,959	7,426	3,805	1,692	594
コントロール変数 = Yes							
GMM							
説明変数	10 億円	30 億円	100 億円	300 億円	1,000 億円	3,000 億円	1 兆円
$TONE^{ADJ}$	0.018	0.055***	0.032***	0.016***	0.013***	0.012***	0.011***
$DM_NETSALE$	0.005	-0.018*	-0.005	-0.001	-0.002	-0.003	-0.011*
$TONE^{ADJ} \times DM_NETSALE$	-0.009	-0.050***	-0.027***	-0.006	-0.007	-0.006	-0.006
$N(DM_NETSALE=1)$	13,903	13,246	10,959	7,426	3,805	1,692	594
コントロール変数 = Yes							

備考：被説明変数はいずれも翌期の ROA である。***、**、*はそれぞれ、両側確率1%、5%、10% で回帰係数が有意であることを示している。 $TONE^{ADJ}$ は補正済トーンを指す。Pooled OLS は年のダミー変数を追加した Pooled OLS 推定、GMM は年のダミー変数を追加した一段階システム GMM 推定の結果である。 $DM_NETSALE$ は売上高ダミーで閾値より売上げが大きい場合1、それ以外では0である。閾値を10億円から1兆円まで7通りに変えて分析を行った。 $N(DM_NETSALE=1)$ は売上高ダミーが1である観測数を示す。観測数はすべての結果で14,855であった。コントロール変数の結果は省略している。なお、Pooled OLS モデルと LSDV モデルの間の F 検定では Pooled OLS モデルが棄却された。

もっとも、建設業のように、製造業よりも経常利益の変動係数が大きく有利子負債比率が高いにもかかわらず、交差項の係数が有意ではない業種も存在しており、MD&A における将来見通しのディスクロージャーの充実度に関する業種間格差は、前述の競争度、不確実性、および資金繰りだけでもって説明されるわけではないように窺われる。そうした業種間格差の解明は今後の課題であるが、さまざまな要因の代理変数を補正済トーンと交差させることや、同トーンと経営者が開示する将来に関する数値情報の相互関係を制御することが考えられる。なお、本稿では各業種のダミーを含む回帰分析を行っており、ROA の業種によるオフセット成分は補正されている。

ロ. トーンと売上高

売上高ダミーと交差項を加えた重回帰分析の結果を表7に示す。GMM 推定の結果では閾値が30億円、100億円のモデルで交差項 ($TONE^{ADJ} \times DM_NETSALE$) の係数が有意に負であった。これらの結果は、売上高の小さい企業による MD&A のトーンの将来業績予測力、売上高の大きい企業に比べて高いことを示唆する。

売上高によってこのような差異が発生する理由は、前述のとおりかもしれない。すなわち、売上高を企業規模を示す指標と考えると、規模が大きいほど規模の経済

性や過去の規模拡大時になされた自己資本の蓄積を期待できるが、売上高の小さい企業については、先々において資金繰り難に陥るリスクが高くなるため、現時点においてディスクロージャーに熱心になると推論できる。

(3) トーンと章立て変更

2018年4月の章立て変更の影響を評価するため、公開日が2018年4月1日以降のデータを1、それ以外を0とする章立て変更ダミー (DM_HDCH) と補正済トーンとの交差項を追加して重回帰分析を行った。その結果を表8に示す。章立て変更ダミーの交差項 ($TONE^{ADJ} \times DM_HDCH$) は有意ではなく、企業の属性を考慮しない分析では章立て変更の有意な影響を確認できなかった。この結果は、2003年の米国SEC (U.S. Securities and Exchange Commission) ガイドライン更新による影響を検証しているLi [2010] と同様である。

(4) トーン、章立て変更と企業の属性

本稿では、章立て変更の与えた影響が企業の属性によって異なる可能性を評価するため、業種と売上高について、それぞれのダミーと章立て変更ダミーと補正済トーンの三重交差項を追加した重回帰分析を行った。

イ. 章立て変更と業種

章立て変更ダミーと業種ダミーの交差項を加えた重回帰分析の結果を表9に示す。補正済トーン ($TONE^{ADJ}$)、章立て変更ダミー (DM_HDCH)、業種ダミーの三重交差項のうち、サービス業 (DM_SER) の三重交差項がGMM推定で有意に正となった。また、補正済トーン項の係数は有意ではなく、補正済トーンと業種ダミーの交差項のうち、商業 (DM_MER)、運輸・情報通信業 (DM_TRA)、サービス業 (DM_SER)、不動産業 (DM_RES)、金融業 (DM_FIN)、建設業 (DM_CON)、電気・ガス業 (DM_ELG) の7つの各ダミーの交差項の係数がGMM推定で有意に正となった。これらの7つのダミーのうち、金融業 (DM_FIN) と建設業 (DM_CON) を除く5つのダミーについては、4節(2)で示した業種ダミーのみを含む分析でも交差項が有意に正となっていた。これらの結果より、サービス業を営む企業においては、MD&Aのディスクロージャー内容に将来見通しが他業種よりも強く反映されており、章立て変更によってその傾向がさらに強まったことが示唆される。また、運輸・情報通信業、サービス業、不動産業、電気・ガス業を営む企業については、MD&Aのディスクロージャー内容に将来見通しが他業種よりも強く反映され

表 8 章立て変更とトーン

説明変数	Pooled OLS	GMM
$TONE^{ADJ}$	0.010***	0.011***
DM_HDCH	0.002	0.004
$TONE^{ADJ} \times DM_HDCH$	0.000	0.000
ROA	0.801***	0.604***
RET	0.001***	0.001**
ACC	-0.009	-0.056***
$SIZE$	0.002***	0.007**
MTB	0.001**	-0.001
$RETVOL$	-0.003***	-0.003**
$ROAVOL$	-0.035***	0.009
$NBSEG$	-0.001	-0.017
$NGSEG$	0.001*	-0.001
AGE	0.000***	0.000
SI	0.087***	0.135***
$READ$	0.001	-0.006**

備考：被説明変数はいずれも翌期の ROA である。***、**、*はそれぞれ、両側確率 1%、5%、10%で回帰係数が有意であることを示している。 $TONE^{ADJ}$ は補正済トーンを指す。Pooled OLS は年のダミー変数を追加した Pooled OLS 推定、GMM は年のダミー変数を追加した一段階システム GMM 推定の結果である。 DM_HDCH は章立て変更ダミー。なお、Pooled OLS モデルと LSDV モデルの間の F 検定では Pooled OLS モデルが棄却された。

ているが、章立て変更による影響はなく、それ以外の業種を営む企業についても章立て変更による有意な影響はなかったことが示唆される。

ロ. 章立て変更と売上高

補正済トーンと章立て変更ダミーと売上高ダミーの三重交差項 ($TONE^{ADJ} \times DM_HDCH \times DM_NETSALE$) を加えた重回帰分析の結果を表 10 に示す。唯一、GMM 推定の閾値 30 億円のモデルにおいて、有意水準は 10% であるが、三重交差項の係数が有意に正となった。4 節 (2) の結果では、規模の小さい企業ほどディスクロージャーに熱心であることが示唆されたが、この結果からは、相対的にディスクロージャーに熱心ではないと考えられる規模の大きな企業において、章立て変更によってそのディスクロージャー内容に将来見通しが反映される傾向が強まったことが示唆される。

表9 章立て変更、業種とトーン

説明変数	Pooled OLS	GMM
$TONE^{ADJ}$	0.002	-0.004
$TONE^{ADJ} \times DM_{MER}$	0.008***	0.020***
$TONE^{ADJ} \times DM_{TRA}$	0.011***	0.025***
$TONE^{ADJ} \times DM_{FIN}$	0.003	0.017**
$TONE^{ADJ} \times DM_{SER}$	0.010**	0.025***
$TONE^{ADJ} \times DM_{FAF}$	0.012	0.019
$TONE^{ADJ} \times DM_{CON}$	-0.002	0.017**
$TONE^{ADJ} \times DM_{MIN}$	-0.001	0.006
$TONE^{ADJ} \times DM_{RES}$	0.016***	0.027***
$TONE^{ADJ} \times DM_{ELG}$	0.008	0.022**
$TONE^{ADJ} \times DM_{HDCH} \times DM_{MER}$	-0.001	0.002
$TONE^{ADJ} \times DM_{HDCH} \times DM_{TRA}$	-0.008	-0.007
$TONE^{ADJ} \times DM_{HDCH} \times DM_{FIN}$	0.002	0.001
$TONE^{ADJ} \times DM_{HDCH} \times DM_{SER}$	0.016*	0.018*
$TONE^{ADJ} \times DM_{HDCH} \times DM_{FAF}$	-0.029*	0.028
$TONE^{ADJ} \times DM_{HDCH} \times DM_{CON}$	-0.008	-0.009
$TONE^{ADJ} \times DM_{HDCH} \times DM_{MIN}$	-0.021	-0.027
$TONE^{ADJ} \times DM_{HDCH} \times DM_{RES}$	-0.004	-0.002
$TONE^{ADJ} \times DM_{HDCH} \times DM_{ELG}$	-0.004	-0.010
コントロール変数= Yes		

備考：被説明変数はいずれも翌期のROAである。***、**、*はそれぞれ、両側確率1%、5%、10%で回帰係数が有意であることを示している。 $TONE^{ADJ}$ は補正済トーンを指す。Pooled OLSは年のダミー変数を追加したPooled OLS推定、GMMは年のダミー変数を追加した一段階システムGMM推定の結果である。 DM_{HDCH} は章立て変更ダミー。それ以外の“DM”で始まる項はそれぞれ以下の業種ダミー。 DM_{MER} ：商業、 DM_{TRA} ：運輸・情報通信業、 DM_{FIN} ：金融業、 DM_{SER} ：サービス業、 DM_{FAF} ：農林・水産業、 DM_{CON} ：建設業、 DM_{MIN} ：鉱業、 DM_{RES} ：不動産業、 DM_{ELG} ：電気・ガス業。製造業をベンチマークとして推定した。コントロール変数およびダミー変数の結果は省略している。なお、Pooled OLSモデルとLSDVモデルの間のF検定ではPooled OLSモデルが棄却された。

表 10 章立て変更、売上高とトーン

Pooled OLS							
説明変数	10 億円	30 億円	100 億円	300 億円	1,000 億円	3,000 億円	1 兆円
<i>TONE^{ADJ}</i>	0.005	0.014***	0.013***	0.012***	0.011***	0.010***	0.010***
<i>DM_HDCH</i>	0.004	0.002	0.003	0.002	0.003	0.002	0.002
<i>TONE^{ADJ} × DM_HDCH</i>	0.000	-0.008	0.000	-0.002	0.000	0.000	0.000
<i>DM_NETSALE</i>	0.009***	0.005***	0.003***	0.000	-0.002**	-0.003**	-0.005***
<i>TONE^{ADJ} × DM_NETSALE</i>	0.005	-0.006	-0.005**	-0.005***	-0.005***	-0.005***	-0.003
<i>DM_HDCH × DM_NETSALE</i>	-0.002	0.000	-0.001	0.000	-0.001	-0.001	0.000
<i>TONE^{ADJ} × DM_HDCH × DM_NETSALE</i>	0.001	0.009	0.000	0.003	0.002	0.003	-0.001
<i>N(DM_NETSALE=1)</i>	13,903	13,246	10,959	7,426	3,805	1,692	594
コントロール変数 = Yes							
GMM							
説明変数	10 億円	30 億円	100 億円	300 億円	1,000 億円	3,000 億円	1 兆円
<i>TONE^{ADJ}</i>	-0.007	0.014	0.015***	0.016***	0.015***	0.013***	0.011***
<i>DM_HDCH</i>	0.001	-0.002	0.002	0.003	0.003	0.003	0.003
<i>TONE^{ADJ} × DM_HDCH</i>	-0.012	-0.020*	-0.002	-0.001	0.000	0.000	0.000
<i>DM_NETSALE</i>	0.015***	0.009**	0.006***	0.000	-0.004	-0.003	-0.014**
<i>TONE^{ADJ} × DM_NETSALE</i>	0.018*	-0.005	-0.005	-0.009***	-0.012***	-0.014***	-0.002
<i>DM_HDCH × DM_NETSALE</i>	0.003	0.006	0.001	0.001	-0.001	-0.001	0.001
<i>TONE^{ADJ} × DM_HDCH × DM_NETSALE</i>	0.013	0.022*	0.002	0.003	0.000	0.002	-0.001
<i>N(DM_NETSALE=1)</i>	13,903	13,246	10,959	7,426	3,805	1,692	594
コントロール変数 = Yes							

備考：被説明変数はいずれも翌期の ROA である。***、**、*はそれぞれ、両側確率 1%、5%、10% で帰帰係数が有意であることを示している。*TONE^{ADJ}* は補正済トーンを指す。Pooled OLS は年のダミー変数を追加した Pooled OLS 推定、GMM は年のダミー変数を追加した一段階システム GMM 推定の結果である。*DM_HDCH* は章立て変更ダミー、*DM_NETSALE* は売上高ダミーである。閾値を 10 億円から 1 兆円まで 7 通りに変えて分析を行った。*N(DM_NETSALE=1)* は売上高ダミーが 1 である観測数を示す。観測数はすべての結果で 14,855 であった。コントロール変数の結果は省略している。なお、Pooled OLS モデルと LSDV モデルの間の F 検定では Pooled OLS モデルが棄却された。

5. おわりに

本稿は有価証券報告書の MD&A の記述についてトーンを用いた分析を行い、以下の結果を示した。

1. MD&A のトーンが将来業績の予測力を持っており、この結果は企業が MD&A を通じて、経営者による将来見通しのディスクロージャーを行って

いる可能性を示唆する。

2. 業種によって MD&A の将来見通しのディスクロージャーの充実度が異なる。
3. 売上高の小さな会社のほうが MD&A のディスクロージャー内容に将来見通しがより強く反映されている。
4. 2018 年の章立て変更により、サービス業を営む企業で MD&A における将来見通しのディスクロージャーの充実度の向上が示唆される。

文章のトーンを指標として用いてわが国の有価証券報告書の大規模データを分析した研究は本稿が初めてであり、MD&A におけるディスクロージャーの内容や規制変更の効果に関する有用な知見を得られたことは本稿の貢献である。特に、本稿のテキスト・マイニングの手法を用いた統計的な分析により、業種や売上高によって MD&A のディスクロージャーの充実度が異なることや、2018 年の章立て変更によって特定の業種において MD&A のディスクロージャー内容が向上していることが明らかになった。本稿により、本稿で用いたテキスト分析の手法が、この分野における実証研究にとって有用であることが示された。

最後に、今後の課題を述べる。第 1 に、使用データの拡張である。本稿で用いた有価証券報告書のデータは 2014 年から 2019 年に発行されたものに限られている。もし過去データを追加できれば、景気後退局面を含むことになる。これは、本稿の分析結果の頑健性チェックに資するものと考えられる。逆に、将来データの追加は、章立て変更からより長い時間が経過する中においてなされる企業のディスクロージャー面での何らかの対応を捕捉する可能性がある。第 2 に、企業の属性とディスクロージャーの関係については、本稿をきっかけにさらなる検証を期待したい。具体的には、トーンの将来予測成分が将来業績に与える影響の決定要因の分析は、興味深い実証課題になるものと思われる。同影響については、個社レベルでの多種多様な性質はもとより、業種特有の不確実性、資金繰り、そして競争度合といった特性からも説明されるかもしれない。加えて、企業には有価証券報告書以外にも、四半期報告書やカンファレンス・コール、アニュアル・レポート等のさまざまなディスクロージャーがあり、これらのデータをも含めた分析をすることでより詳細な分析ができる可能性がある。最後に、トーン以外のテキストの特徴量の活用である。補論 2 で示すように、一般的な読みやすさの指標は有価証券報告書の評価には適していないが、他にも、本稿では取り扱わなかったさまざまな特徴量がある。例えば、言語学におけるモダリティを利用すれば、情報発信者の主観を定量化できる可能性がある。こうした、トーン以外の特徴量を用いることによって、企業によるディスクロージャーの内容に対するさらなる理解向上が期待される。

参考文献

- 金融庁、「金融審議会 ディスクロージャーワーキング・グループ報告—建設的な対話の促進に向けて—」、金融庁、2016年 (https://www.fsa.go.jp/singi/singi_kinyu/tosin/20160418-1/01.pdf、2021年4月12日)
- 東京証券取引所、「コーポレートガバナンス・コード～会社の持続的な成長と中長期的な企業価値の向上のために～」、東京証券取引所、2018年 (<https://www.jpx.co.jp/news/1020/nlsgeu000000xbfx-att/nlsgeu0000034qt1.pdf>、2021年4月12日)
- 東山昌彦・乾 健太郎・松本裕治、「述語の選択選好性に着目した名詞評価極性の獲得」、『言語処理学会第14回年次大会論文集』、2008年、584～587頁
- 廣瀬喜貴・平井裕久・新井康平、「MD&A情報の可読性が将来業績に及ぼす影響：テキストマイニングによる分析」、『経営分析研究』第33巻、2017年、87～101頁
- 吉田政之、「リスク情報の可読性と将来業績に関する実証分析」、『神戸大学大学院経営学研究科大学院生ワーキング・ペーパー』第06a巻、2018年、1～16頁
- 千木良弘朗・早川和彦・山本 拓、『動学的パネルデータ分析』、知泉書館、2011年
- 李 在鎬、「日本語教育のための文章難易度に関する研究」、『早稲田日本語教育学』第21巻、2016年、1～16頁
- Barron, Ori E., Charles O. Kile, and Terrence B. O’Keefe, “MD&A Quality as Measured by the SEC and Analysts’ Earnings Forecasts,” *Contemporary Accounting Research*, 16(1), 1999, pp. 75–109.
- Bond, Steve, Asli Leblebicioğlu, and Fabio Schiantarelli, “Capital Accumulation and Growth: A New Look at the Empirical Evidence,” *Journal of Applied Econometrics*, 25(7), 2010, pp. 1073–1099.
- Bryan, Stephen H., “Incremental Information Content of Required Disclosures Contained in Management Discussion and Analysis,” *The Accounting Review*, 72(2), 1997, pp. 285–301.
- Li, Feng, “Annual Report Readability, Current Earnings, and Earnings Persistence,” *Journal of Accounting and Economics*, 45(2–3), 2008, pp. 221–247.
- , “The Information Content of Forward-Looking Statements in Corporate Filings: A Naïve Bayesian Machine Learning Approach,” *Journal of Accounting Research*, 48(5), 2010, pp. 1049–1102.
- Lo, Kin, Felipe Ramos, and Rafael Rogo, “Earnings Management and Annual Report Readability,” *Journal of Accounting and Economics*, 63(1), 2017, pp. 1–25.
- Loughran, Tim, and Bill McDonald, “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks,” *Journal of Finance*, 66(1), 2011, pp. 35–65.
- , and ———, “Measuring Readability in Financial Disclosures,” *Journal of Finance*, 69(4), 2014, pp. 1643–1671.

- Muslu, Volkan, Suresh Radhakrishnan, K. R. Subramanyam, and Dongkuk Lim, “Forward-Looking MD&A Disclosures and the Information Environment,” *Management Science*, 61(5), 2015, pp. 931–948.
- Pava, Moses L., and Marc J. Epstein, “How Good is MD&A as an Investment Tool?” *Journal of Accountancy*, 175(3), 1993, pp. 51–53.
- Sunakawa, Yuriko, Jae-ho Lee, and Mari Takahara, “The Construction of a Database to Support the Compilation of Japanese Learners Dictionaries,” *Acta Linguistica Asiatica*, 2(2), 2012, pp. 97–115.

補論 1. 有価証券報告書データの取得方法

本稿に利用した有価証券報告書は金融庁の EDINET (Electronic Disclosure for Investors' NETwork) ホームページより取得した⁷。以下、取得方法を詳述する⁸。2019年3月に有価証券報告書データを取得するための API (Application Programming Interface) が公開され、プログラムを用いた機械的なデータ取得をネットワークに過度な負荷をかけることなく実行可能となった。API を用いたデータ取得の手順は以下のとおりである。

1. 日付を設定した URL にアクセスし、その日に公開された文書一覧を取得する⁹
2. 取得した文書一覧から取得したい文書の文書 ID を読み込む
3. 文書 ID を設定した URL にアクセスし、文書データを取得する¹⁰

手順 1 で取得した文書一覧は JSON (JavaScript Object Notation) と呼ばれるデータ形式になっており、さまざまなプログラミング言語で容易に扱うことができる。例えば python の場合、json ライブラリを用いて JSON 形式のファイルを取り込むことができる。

なお、EDINET より取得可能なデータは過去 5 年間に公開されたものに限られており、本稿で用いたデータの総量を制限している¹¹。データ取得可能な期間が延長されれば、有価証券報告書を通じた情報開示に関するより頑健な計量分析が期待できる。

.....
7 「EDINET」：<https://disclosure.edinet-fsa.go.jp> (最終閲覧日：2021年4月12日)

8 EDINET では有価証券報告書以外にも、四半期報告書や公開買付届出書等の文書も取得可能である。

9 <https://disclosure.edinet-fsa.go.jp/api/v1/documents.json?date='YYYY-MM-DD'&type=2>。ただし“YYYY-MM-DD”は日付である。

10 <https://disclosure.edinet-fsa.go.jp/api/v1/documents/'DOCID'?type=1>
ただし“DOCID”は文書 ID である。末尾の type 以下の数字はデータの種類を示し、1：本文 (テキスト)、2：本文 (PDF)、3：代替書面、添付文書、4：英文である。本稿では1のテキストデータを用いた。

11 米国の有価証券報告書に当たる Form 10-K は、SEC の EDGAR (the Electronic Data Gathering, Analysis, and Retrieval system) において、少なくとも過去数十年分のデータを無料で取得できる。「EDGAR」：<https://www.sec.gov/edgar.shtml> (最終閲覧日：2021年4月12日)

補論 2. 読みやすさの指標

FOG 指数は英文の読みやすさの指標であり、(A-1) 式で計算される。

$$FOG = \{ (\text{文当たり単語数}) + (\text{難単語率}) \} \times 0.4. \quad (\text{A-1})$$

ここで、難単語率は全単語に占める 3 音節以上の単語の割合を意味する。FOG 指数はその文章を理解するのに必要な公的教育を受けた年数を示しており、FOG 指数が 12 では高校卒業程度、17 では大学院卒業程度の読解力が要求される読みやすさであることを示す。

FOG 指数は英文の指数であり、和文には適用できない。そこで、本稿では読みやすさの指標として、日本語文章難易度判別システム (jReadability) で用いられている読みやすさ指数を採用した¹²。jReadability の読みやすさ指数は (A-2) 式で計算される (Sunakawa, Lee, and Takahara [2012]、李 [2016])。

$$READ = 11.724 - 0.056 \times (\text{平均文長}) - 0.126 \times (\text{漢語率}) \\ - 0.042 \times (\text{和語率}) - 0.145 \times (\text{動詞率}) - 0.044 \times (\text{助詞率}). \quad (\text{A-2})$$

ここで、READ は読みやすさ指数で、漢語率、和語率、動詞率、助詞率はそれぞれ、全単語に占める漢語、和語、動詞、および助詞の割合を示す。本稿では、この式を用いて各テキストの読みやすさ指数を計算した。指数の算出に必要な漢語、和語、動詞、助詞の同定に当たっては、日本語学習辞書支援グループによる「日本語教育語彙表 Ver 1.0」を用いた¹³。

本稿では、将来業績と読みやすさ指数の間に有意な関係はみられなかった。これは廣瀬・平井・新井 [2017] の結果と一致する。Loughran and McDonald [2014] では、FOG 指数を読みやすさの指標として有価証券報告書を分析することの限界を指摘している。jReadability は FOG 指数同様、教育の分野で提唱された指標である。日本語のファイナンス文書の分析に当たっては、それに特化した読みやすさの指標の開発が重要となる。

.....
12 「jReadability Portal」: <https://jreadability.net/> (最終閲覧日: 2021 年 4 月 12 日)

13 「日本語教育語彙表」: <http://jhlee.sakura.ne.jp/JEV/> (最終閲覧日: 2021 年 4 月 12 日)

補論 3. 極性辞書の信頼性

本稿では、LM 辞書を用いて MD&A のテキスト分析を行った。ここでは、この辞書の信頼性を評価するため、日本語評価極性辞書（名詞句編）を用いた場合との簡単な比較を行う（東山・乾・松本 [2008]）。この辞書には LM 辞書と同様に、単語とその極性情報が収録されている。

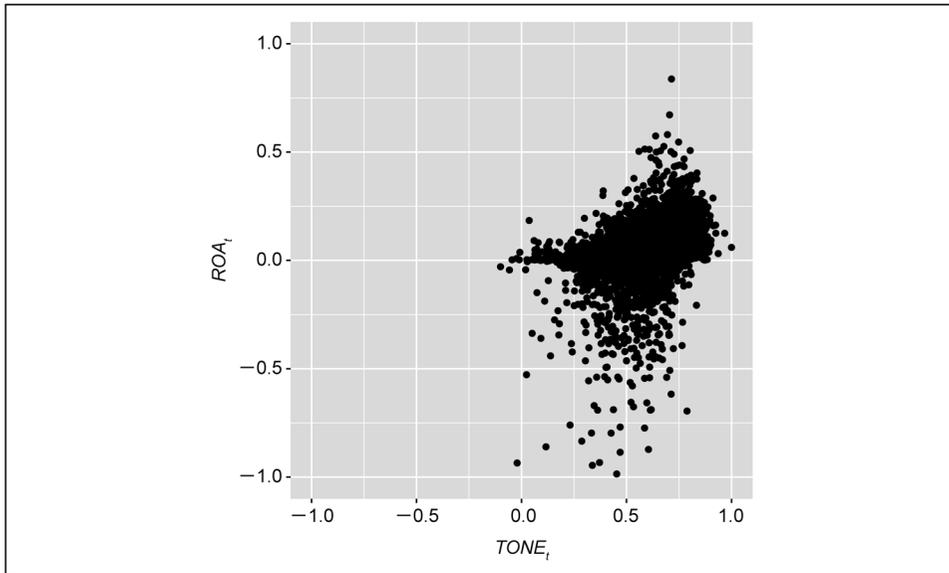
LM 辞書と日本語評価極性辞書に含まれる単語の中で、有価証券報告書に頻出した上位 30 単語を表 A-1 と表 A-2 にそれぞれ示す。LM 辞書については、企業の業績に対して極性を持つと思われる単語がリストアップされている。また、単語の出現総回数もポジティブ単語が約 94 万回に対してネガティブ単語が 77 万回で、その差は 20% 以下である。一方、日本語評価極性辞書に含まれる単語の中には、企業の業績に対しては極性を持たないと思われる単語が多く含まれている。例えば、ポジティブの上位 10 単語のうち、「売上」、「資産」、「活動」、「ため」、「資金」、「サービス」の 6 単語は、企業の業績に対して極性を持たないと考えられる。また、単語の出現総回数は、ポジティブ単語が約 480 万回に対してネガティブ単語が約 110 万回と 4 倍以上の差がある。特に、先ほど挙げた 6 つのポジティブとされている単語の出現回数の合計は約 107 万回となり、これだけでネガティブ単語の出現総回数に匹敵する。

同様に (1) 式によって算出したトーンを比較すると、これらの企業の業績に対しては極性を持たない単語をカウントしたことの影響が顕著にみられる。日本語評価極性辞書を用いて算出したトーンと ROA の関係を図 A-1 に示す。図から明らかのように日本語評価極性辞書を用いて算出したトーンはほとんどが正の側に分布しており、企業の業績に対しては極性を持たない単語をポジティブとして数え上げた結果、トーンが正の側に偏っていることが示唆される。一方、LM 辞書を用いて算出したトーンと ROA の関係を示す図 1 をみると、LM 辞書を用いて算出したトーンは正負の両側に広く分布している¹⁴。また、ROA が負である企業を比較すると、LM 辞書を用いて算出したトーンではトーンが負の側に多く分布しているが、図 A-1 では ROA が負である企業のトーンが小さい側へ偏る傾向はみられない。

以上に示したように、LM 辞書に含まれる単語は企業業績に対する極性を持っており、LM 辞書を用いて算出したトーンも正負の両側に広く分布し、ROA との間にも正の相関を持っていることから、本稿で使用した LM 辞書と、その辞書を用いて算出したトーンの信頼性は高いことが示唆される。

なお、2 節で述べたように、LM 辞書に含まれる単語数は、重複を除くとポジティブ 255 語、ネガティブ 1,374 語の合計 1,629 語である。一方、日本語評価極性辞書

14 LM 辞書を用いて算出したトーンの平均は 0.12、標準偏差は 0.26 で、日本語評価極性辞書を用いて算出したトーンの平均は 0.62、標準偏差は 0.12 である。

図 A-1 日本語評価極性辞書を用いて算出した $TONE_t$ 対 ROA_t 

に含まれる単語数は、重複を除くとポジティブ 3,486 語、ネガティブ 7,594 語の合計 11,080 語である。含まれる単語数が日本語評価極性辞書と比べて 1 桁近く少ないにもかかわらず、LM 辞書を用いたほうがもっともらしいトーンが算出されている。これは、LM 辞書に含まれる単語が企業業績に対する極性を持つ単語にうまく絞られていることを示唆している。Loughran and McDonald [2011] は一般的な辞書でネガティブとされる単語の約 4 分の 3 は有価証券報告書ではネガティブな意味を持たないことを示した。このように、日本の有価証券報告書においても同様に、一般的な辞書の極性をそのまま適用することはできず、ファイナンス文書のトーン算出に当たっては分析するテキストに対応した辞書や手法を用いる必要があると考えられる。

表 A-1 頻出単語上位 30 語：LM 辞書

順位	ポジティブ単語				ネガティブ単語			
	単語	出現回数	比率 [%]	累積 [%]	単語	出現回数	比率 [%]	累積 [%]
1	利益	372,325	39.60%	39.60%	減少	233,390	30.50%	30.50%
2	強化	120,269	12.79%	52.39%	損失	70,682	9.24%	39.74%
3	実績	61,636	6.56%	58.95%	目的	27,428	3.58%	43.33%
4	改善	53,169	5.65%	64.60%	研究	23,070	3.02%	46.34%
5	効率	37,071	3.94%	68.54%	発生	20,791	2.72%	49.06%
6	共同の	32,330	3.44%	71.98%	減損	19,669	2.57%	51.63%
7	安定	26,010	2.77%	74.75%	利用	18,329	2.40%	54.02%
8	積極的に	22,541	2.40%	77.14%	削減	18,183	2.38%	56.40%
9	達成	17,461	1.86%	79.00%	減価償却	17,693	2.31%	58.71%
10	最大	13,267	1.41%	80.41%	変動	17,547	2.29%	61.01%
11	有効	12,428	1.32%	81.73%	問題	15,046	1.97%	62.97%
12	収益性	11,610	1.23%	82.97%	厳しい	14,902	1.95%	64.92%
13	満足	11,327	1.20%	84.17%	未払	14,815	1.94%	66.86%
14	高める	8,097	0.86%	85.04%	利息	14,639	1.91%	68.77%
15	強み	7,625	0.81%	85.85%	懸念	11,588	1.51%	70.28%
16	安定した	6,608	0.70%	86.55%	減速	10,097	1.32%	71.60%
17	機会	6,512	0.69%	87.24%	不足	7,651	1.00%	72.60%
18	最高	6,250	0.66%	87.91%	悪化	7,560	0.99%	73.59%
19	成果	5,790	0.62%	88.52%	転換	6,179	0.81%	74.40%
20	革新	5,668	0.60%	89.12%	緩和	5,808	0.76%	75.16%
21	進捗	4,607	0.49%	89.61%	縮小	5,520	0.72%	75.88%
22	得る	4,513	0.48%	90.09%	調査	5,397	0.71%	76.59%
23	優秀な	4,320	0.46%	90.55%	不適切な	5,328	0.70%	77.28%
24	報酬	3,983	0.42%	90.98%	損なう	5,208	0.68%	77.96%
25	強化する	3,926	0.42%	91.40%	反動	5,155	0.67%	78.64%
26	十分に	3,778	0.40%	91.80%	災害	4,948	0.65%	79.28%
27	透明性	3,731	0.40%	92.19%	できない	4,181	0.55%	79.83%
28	強い	3,552	0.38%	92.57%	省略	3,854	0.50%	80.33%
29	イノベーション	3,424	0.36%	92.94%	終了	3,580	0.47%	80.80%
30	魅力	3,247	0.35%	93.28%	毀損	3,450	0.45%	81.25%
合計		940,249	100.00%			765,140	100.00%	

備考：比率はポジティブ（ネガティブ）単語の出現回数をすべてのポジティブ（ネガティブ）単語の出現総回数で割って計算した。

表 A-2 頻出単語上位 30 語：日本語評価極性辞書

順位	ポジティブ単語				ネガティブ単語			
	単語	出現回数	比率 [%]	累積 [%]	単語	出現回数	比率 [%]	累積 [%]
1	利益	346,651	7.11%	7.11%	負債	88,497	7.90%	7.90%
2	売上	259,222	5.32%	12.43%	支出	84,781	7.57%	15.47%
3	資産	212,385	4.36%	16.79%	損失	70,484	6.29%	21.77%
4	活動	185,736	3.81%	20.60%	費用	57,656	5.15%	26.91%
5	ため	170,684	3.50%	24.11%	取引	55,746	4.98%	31.89%
6	資金	147,192	3.02%	27.13%	課題	47,655	4.26%	36.15%
7	向上	107,232	2.20%	29.33%	価格	38,197	3.41%	39.56%
8	価値	96,316	1.98%	31.31%	債権	33,984	3.03%	42.59%
9	サービス	93,956	1.93%	33.23%	金額	32,732	2.92%	45.51%
10	収益	93,034	1.91%	35.14%	コスト	30,996	2.77%	48.28%
11	成績	89,570	1.84%	36.98%	リスク	30,802	2.75%	51.03%
12	業績	79,669	1.64%	38.62%	債務	28,978	2.59%	53.62%
13	現金	69,570	1.43%	40.04%	支配	21,007	1.88%	55.49%
14	取締役	69,428	1.42%	41.47%	発生	20,813	1.86%	57.35%
15	成長	69,333	1.42%	42.89%	対抗	20,210	1.80%	59.16%
16	情報	67,361	1.38%	44.27%	変動	17,832	1.59%	60.75%
17	規模	66,255	1.36%	45.63%	減価	17,709	1.58%	62.33%
18	収入	63,046	1.29%	46.93%	過去	14,794	1.32%	63.65%
19	提供	62,616	1.29%	48.21%	部品	13,880	1.24%	64.89%
20	顧客	61,854	1.27%	49.48%	低下	13,461	1.20%	66.09%
21	実績	61,631	1.26%	50.75%	減収	13,186	1.18%	67.27%
22	確保	60,517	1.24%	51.99%	問題	12,897	1.15%	68.42%
23	改善	55,114	1.13%	53.12%	懸念	11,589	1.03%	69.46%
24	資本	53,294	1.09%	54.21%	下落	9,345	0.83%	70.29%
25	技術	51,943	1.07%	55.28%	経費	9,274	0.83%	71.12%
26	設備	49,453	1.01%	56.30%	廃止	7,791	0.70%	71.81%
27	需要	48,879	1.00%	57.30%	低迷	7,444	0.66%	72.48%
28	将来	45,518	0.93%	58.23%	増税	6,965	0.62%	73.10%
29	実現	40,862	0.84%	59.07%	差別	6,259	0.56%	73.66%
30	積極的	37,616	0.77%	59.84%	手数料	5,885	0.53%	74.19%
合計		4,872,619	100.00%			1,119,965	100.00%	

備考：比率はポジティブ（ネガティブ）単語の出現回数をすべてのポジティブ（ネガティブ）単語の出現総回数で割って計算した。

