

証券経済学会第98回全国大会

統合報告書のテキスト分析と企業評価 —開示情報の定量化と市場評価との関係—

2026年9月14日

報告者 小方信幸

青山学院大学 客員教授

1. 研究の背景・動機・目的

- 投資家が企業の中長期的な価値創造能力を評価するためには、財務情報と非財務情報（特にサステナビリティ情報）の活用が不可欠となっている。
- 統合報告書は、財務情報とサステナビリティ情報を結び付けながら企業の価値創造プロセスを説明する開示媒体として普及している。
- 投資家が最も重視する企業の開示情報は統合報告書であるとする調査報告がある（生命保険協会 2026）。
- しかし、統合報告書の内容を客観的かつ定量的に評価する手法は十分に確立されていない。

- 本研究では、統合報告書の記述内容を定量化することにより、企業が投資家に提供するサステナビリティ情報が資本市場において、どのように評価されているのかを明らかにすることを目的とする。
- 具体的には、2021年から2025年に上場企業によって発行されたすべての統合報告書4,767企業年を分析対象とした。
- 収集したすべての統合報告書から、日経統合報告書アワード2025第1次審査10項目と各項目のキーワードに関連する特定用語を抽出して「辞書」を作成した。
- 企業年ごと10項目ごとに辞書の特定用語が出現する割合（出現率）を算出し、統合報告書発行後の株主資本コストとの関係を分析した。
- 今回の「辞書」は試作版の位置づけで、今後改修を行う方針である。

2. 先行研究

2-1 テキストデータと資本コストに関する研究

2-2 一般辞書・専門辞書によるテキストデータと財務・株価の分析

2-3 独自辞書によるテキストデータと財務・株価の分析

2-1 テキストデータと資本コストに関する研究

- 開示情報と資本コストの関係については、2000年頃から海外、国内ともにテキストデータを用いた多くの学術研究がみられる。
- 企業の年次報告書、決算説明書等の開示水準と資本コストの関係を分析した研究としては、Hail(2002), Botosan and Plumlee(2002), Francis and Olsson(2008)がある。
- Hail(2002)は、スイス企業73社の開示水準と残余価値モデルで推定した株主資本コストとの間に有意に負の関係があると報告している。

- 日本においては、村宮(2005)は利益予想の精度に、内野(2005)は開示行動に着目し、株主資本コストとの関係进行分析している。
- 最近の研究では、後藤(2026)が2023年3月期から有価証券報告書に新設された「サステナビリティに関する考え方及び取組」の開示義務化と株主資本コストとの関係を検証している。

後藤(2026)は、2023年6月開示開始の3月期決算企業を処置群、2024年2月開示開始の12月期決算企業を対照群とする差の差分析を採用した。2022年6月～2024年2月の2,092社、43,932企業月を分析した結果、義務化後の開示と株主資本コストの間に有意な低下が確認された。特に、統合報告書未発行、TCFD未賛同の企業の株主資本コストに開示の効果が表れた。

2-2 一般辞書・専門辞書による テキストデータと財務・株価の分析

- ・ テキストデータと企業業績、株式リターン、株式リスクなどとの関係を扱っている海外の先行研究として、Li (2008), Tetlock, Saar-Tsechansky and Macskassy (2008)を上げることができる。

Loughran et al. (2011)は新聞記事と企業の開示資料をテキストデータ化し、一般辞書または専門辞書を用いて文章の長さや株式リターン、および否定語と株式リターンの関係を分析している。

Li (2008)は米国の年次報告書(Annual Report)を対象に文章の長さで可読性を測定し、利益が低い企業や赤字企業ほど年次報告書が長く読みにくいとしている。一方、読みやすい年次報告書を作成する企業ほど正の利益の持続性が高いと結論付けている。

Tetlock et al.(2008)は、Dow Jones News とWall Street Journalの企業別ニュースを分析し、Harvard-IV辞書による否定語比率を説明変数とし、将来利益と翌日株式リターンを被説明変数として分析した。結果、否定語比率が高いニュースは、低い将来利益を予測すると報告している。

日本においては、有価証券報告書、決算短信、決算説明会資料からテキストデータを作成して業績・株価との関係进行分析したものとして、野田(2016)、山本・川代・栗田(2020)、黒木・真鍋・中川(2024)などがある。

野田(2016)は、日本企業の有価証券報告書を対象に、文字数と特定のキーワードを用いて開示内容を定量化し、企業評価との関係进行分析した。その結果、「経営成績」に関するテキスト情報を積極的に開示する企業は市場から相対的に高く評価されることを示した。

2-3 独自辞書によるテキストデータと財務・株価の分析

- ・ 米国では2010年代になると、年次報告書10kに基づく専門語辞書をつくり、辞書語の出現率と株価、リスク指標との関係进行分析するLoughran and McDonald (2011)、Campbell, Chen, Dhaliwal, Lu and Steele (2014)などの研究が行われた。

Loughran and McDonald(2011)は、1994～2008年の10-Kを基に、「否定」、「肯定」、「不確実性」、「訴訟」、「強いモダリティ」、「弱いモダリティ」の6種類の専門辞書を独自に構築した。辞書語の出現率を10-Kの提出日前後の超過収益率などに関連づけて検証している。

Campbell, Chen, Dhaliwal, Lu and Steele (2014) は、米国企業の 10-K における「Risk Factors」を対象として、リスク関連語を「システマティック・リスク」、「企業固有リスク」、「財務リスク」、「訴訟リスク」、「税務リスク」に分類する目的特化型辞書を構築した。辞書語の出現率と株式リスクの関係を分析している。

- ・日本においては、独自辞書を作成し辞書語の出現率と株価、株式リスクとの関係进行分析した研究として、五島・高橋(2017)と佐藤・池田・井上(2021)を上げることができる。

五島・高橋(2017)は、ニュース記事が言及した企業の株価変動を教師情報として機械学習を行い、学習器から各単語の重み付き極性値を抽出することにより、金融分野に特化したポジティブ・ネガティブ辞書を構築している。作成した辞書によってニュース記事を分類した結果、記事配信日の株式リターンについて分析している。

- 佐藤・池田・井上(2021)は、2014～2017年の東証一部上場企業 7,141企業年を対象に、有価証券報告書の「事業等のリスク」の記述と株式リスクとの関連を検証した。
- 2つの辞書を作成しテキスト分析を行い、リスクに関する用語の出現率が株式のリスク指標に正に相関することを確認した。
- 一方、過去のリスク指標を考慮すると、辞書指標は翌期のリスク変動に有意な説明力をもたないことから、年度ごとのリスク構造の変化を示す新規情報としては限定的であると結論づけている。

佐藤・池田・井上(2021)における2つの辞書

- 辞書① Campbell et al.(2014)のリスク辞書を翻訳し、日本固有語と頻出語を追加して、市場、財務、法規制訴訟、税、個別リスクの5分類、計1,111語に再構成した。
- 辞書② 辞書①の基準語から、Wikipediaで学習したword2vecによるコサイン類似度を用いて関連語を拡張した。

3-2 先行研究を踏まえた研究課題

- 大規模言語モデルの発展により、テキストデータと財務・株価指標との関係进行分析する研究が見られる。
- 米国では10k、日本では有価証券報告書という法定開示書類を基に独自辞書を作成し、辞書出現率と財務・株式指標との関係进行分析する先行研究がみらる。
- 一方、企業の任意開示書類である統合報告書を基に独自辞書を作成し、辞書語の出現率と財務・株価指標等と関係进行分析した研究は多くは行なわれていない。

本研究における取り組み

- 日本企業が発行した統合報告書を基に独自辞書を作成する。
- 統合報告書における辞書語の出現率を算出する。
- 積極的開示が資本コストを低下させるという先行研究の結果を踏まえ、辞書語の出現率と株主資本コストの関係を分析する。

仮説： 統合報告書に基づく辞書語の出現率と株主資本コストは負に相関する

3. 分析の対象・データ・方法

3-1 対象

- ・ 株式会社エッジ・インターナショナルに提供いただいた「統合レポート発行経年リスト(2021年版-2025年版)」に記載された、上場企業が2021年から2025年に発行した4,767企業年の統合報告書を分析対象とした。
- ・ しかし、辞書作成においては、4,767企業年のうち95件を除いた 4,672企業年の統合報告書のPDFで「辞書」を作成した。

辞書作成に使用しなかった95企業年の内訳は、ChatGPTまたはCodexがPDFを読み取ることができなかったものが69企業年、企業のウェブサイトから取得できなかったPDFが25企業年、企業のサイトに投資家および報道関係のみPDF取得可能としていたのが1企業年であった。

3-2 データ

- ・ テキストデータはChatGPTおよびCODEXを用いて、2021年から2025年に上場企業が発行した統合報告書のPDFが基となっている。
- ・ 分析に用いた株主資本コストは、QUICK社のAstra Manager から入手した

3-2-2 辞書構築

形態素解析

- ・ 形態素解析の主解析器としてSudachiPyを採用した。SudachiPyは複数の分割粒度(A/B/C)を選択でき、辞書形・正規化形等も取得できるため、統合報告書に多い複合表現や表記揺れを扱う辞書構築と相性がよい。
- ・ 一方、MeCabは長い研究利用実績を持ち、先行研究においても使用されている。本研究ではSudachiPyとMeCabの併用を行い、形態素解析器の違いに対する結果の頑健性を検証した。

* 形態素解析は、文章を言語の最小単位である形態素に分割し、それぞれの品詞や文法的な属性を識別する技術。

辞書構築方法

- ・ 日経統合報告書アワード2025の1次審査の10項目(Q1～Q10)と各項目のキーポイントから107のSeed Wordsを形成した。
- ・ Seed Wordsを起点として、形態素解析したテキストをPythonのルールベース処理により共起語、関連語および複合語を抽出した。
- ・ Human Reviewと条件付きルールの整備を経て、Q1～Q10別の辞書を確定した。

出現率

- ・ 確定した辞書をPythonで各企業年の統合報告書全文に適用し、企業年ごとにQ1～Q10別の辞書語出現回数を計算した。

辞書出現率

企業 i 、発行年 t 、Q項目 k の辞書出現率は、次のように定義した。

$$Rate_{kit} = \frac{Count_{kit}}{TotalWords_{it}}$$

$Count_{kit}$: Qkの辞書語出現回数

$TotalWords_{it}$: 統合報告書全文の有効総語数

回帰分析では、1万語当たりの出現回数を使用した。

$$Rate10K_{kit} = \frac{Count_{kit}}{TotalWords_{it}} \times 10,000$$

文書の長さが企業や年度によって異なるため、単純出現回数ではなく、文書長で標準化した出現回数を採用した。

《参考》日経統合報告書アワード2025

審査10項目(Q1～Q10)

- Q1 トップマネジメントのメッセージ
- Q2 企業価値創造を実現するための企業理念(パーパス・ミッション・カルチャー・バリュー・ビジョン)の記述
- Q3 価値創造の実現に向けた自社固有のマテリアリティの抽出のプロセスの提示
- Q4 自社の経営資源(人的資本、知的資本、社会・関係資本、自然資本など無形の資本を含む)の冷静な分析と中長期経営目標(KPI)・戦略(ビジネスモデルの变革含む)に関する記述
- Q5 投資家の分析に必要な十分な財務情報・財務関連情報が記述されているか

- Q6 中期的業容の展開にあたり説得力のある資本配分政策・財務政策・事業ポートフォリオ管理の記述があるか
- Q7 「環境関連」情報のマルチステークホルダーへの説明とソーシャルインパクトに関する記述
- Q8 「社会関連」情報のマルチステークホルダーへの説明とソーシャルインパクトに関する記述
- Q9 コーポレートガバナンス・システムの高度化が窺える記述があるか
- Q10 取締役の責任や役割、取締役会での議論の中身に関して有益な記述があるか

3-3 分析方法

- 辞書語の出現率と株主資本コストの関係を分析する。
- 分析対象の統合報告書は金融・証券・保険・その他金融453企業年を除いた。
- 説明変数 : 4,219企業年ごとに算出したQごとの出現率
- 被説明変数 : Quick社の株主資本コスト

統合報告書発行日から1ヶ月後、3ヶ月後、6ヶ月後の月初の株主資本コストを被説明変数としている。

固定効果モデルによる推計

$$COE_{ith} = \alpha_i + \lambda_t + \beta_k QRate_{kit} + \gamma \ln(Asset_{it}) + \varepsilon_{ith} \quad \text{--- (1)}$$

$$COE_{ith} = \alpha_i + \lambda_t + \beta_k QRATE_{Kit} + \gamma_1 \ln(Asset_{it}) + \gamma_2 ROE_{preit} + \varepsilon_{it} \quad \text{--- (2)}$$

COE_{it} : 企業*i*・発行年*t*における統合報告書発行後*h*の株主資本コスト。*h*は1M、3M、6M

Qk_RATE_{kit} : 企業*i*・発行年*t*の統合報告書における*Qk*対応辞書語の1万語当たり出現回数。

$\ln(Asset_{it})$: 総資産を自然対数化した企業規模(コントロール変数)

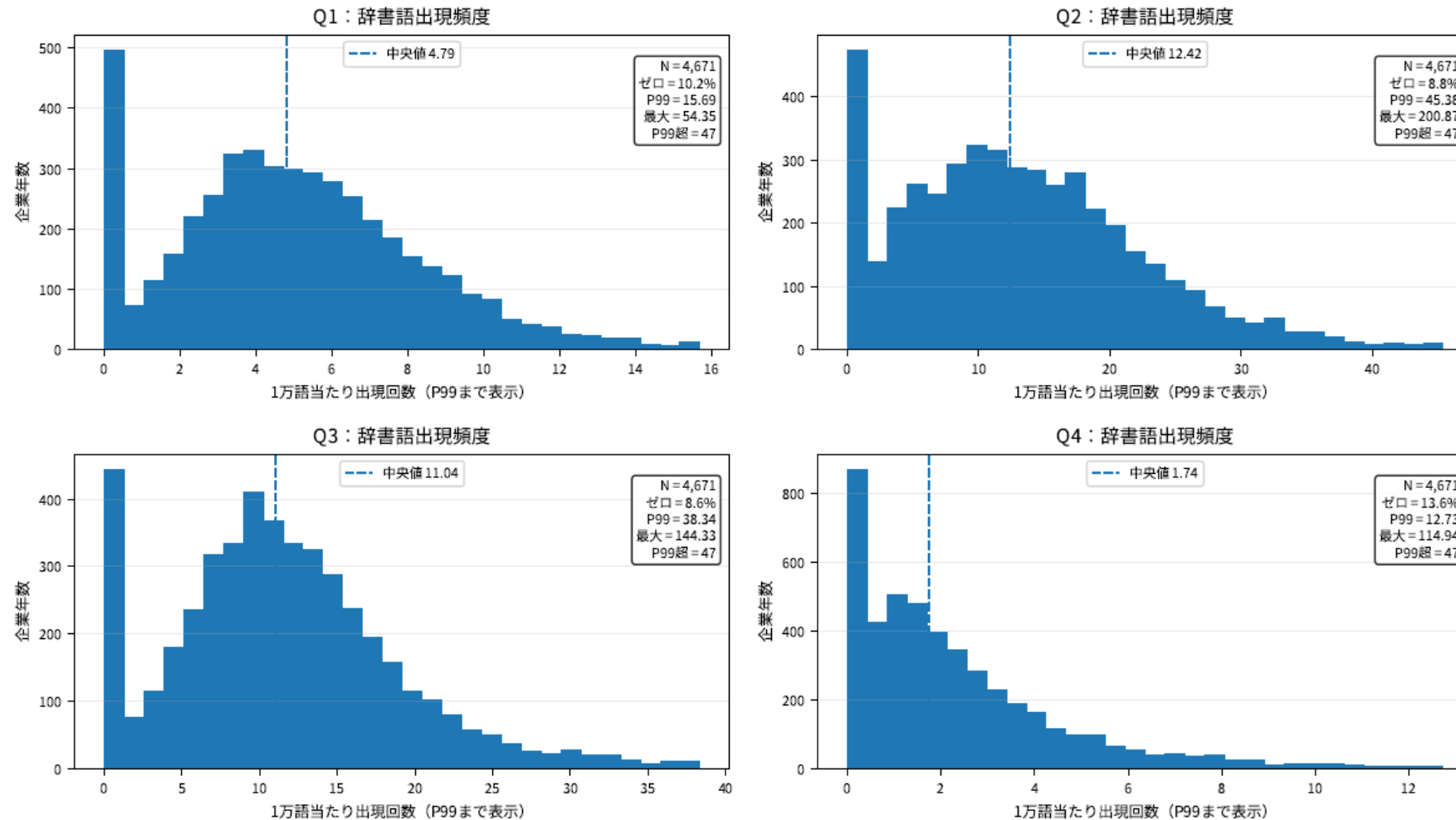
ROE_{preit} : 統合報告書発行日直前の直近本決算におけるROE。単位は%。

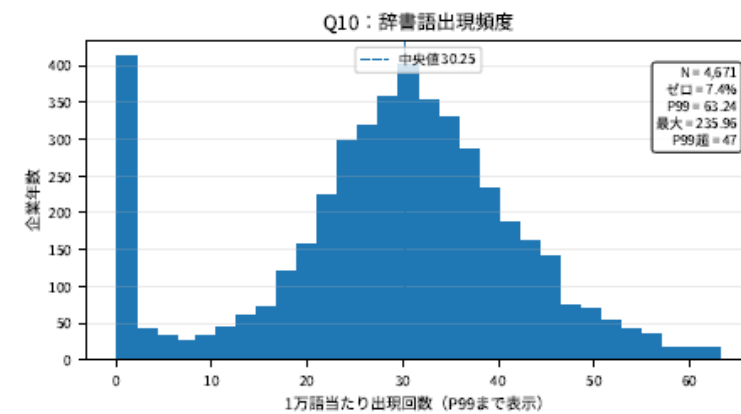
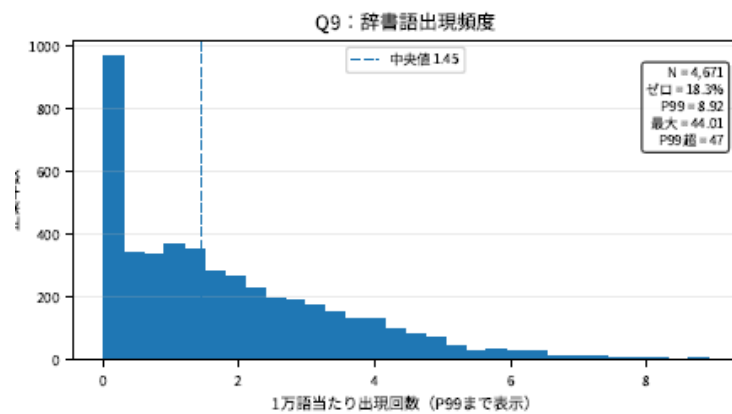
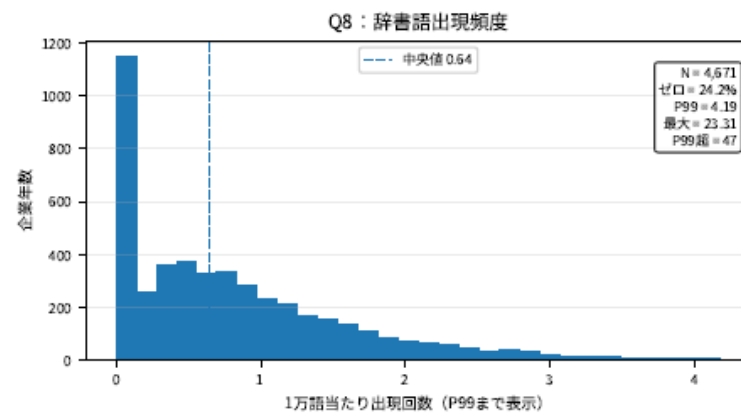
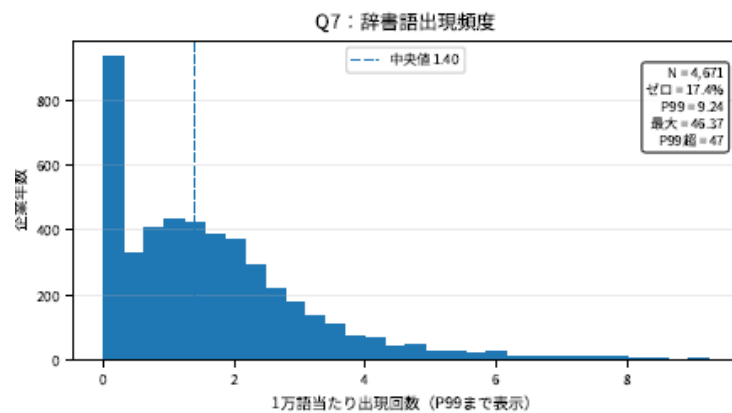
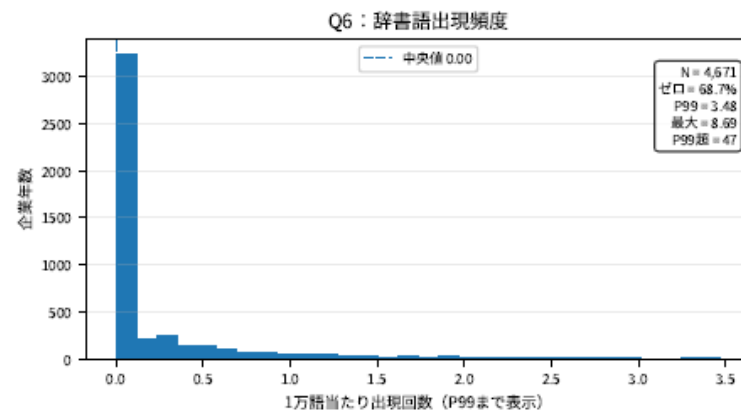
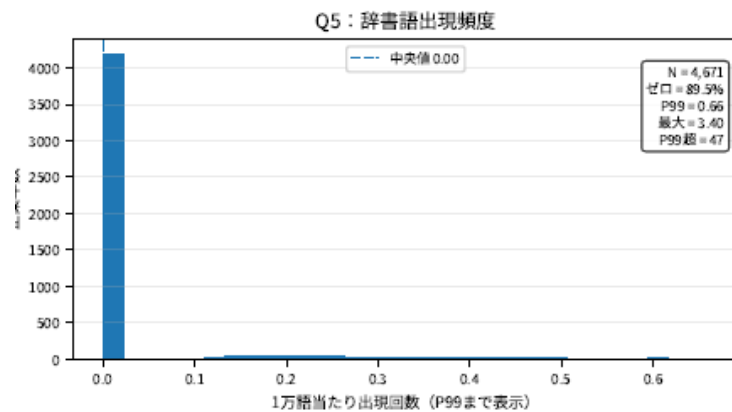
α_i : 企業固定効果

λ_t : 年固定効果

4. 分析結果 固定効果モデル(1) & (2)

図表1 Q1～Q10 辞書後出現頻度の分布





図表2 株主資本コストと辞書出現率の記述統計量（発行直前ROEあり）

変数	サンプル数 (N)	平均	標準偏差	最小	P10	中央値	P90	最大
株主資本コスト（1M）	3,865	6.836	2.342	3.291	4.058	6.548	10.105	13.340
株主資本コスト（3M）	3,865	6.831	2.328	3.313	4.083	6.526	10.108	13.469
株主資本コスト（6M）	3,865	6.848	2.314	3.339	4.152	6.508	10.145	13.469
Q1出現率	3,865	5.199	3.653	0.000	0.444	4.802	9.449	54.348
Q2出現率	3,865	14.177	10.938	0.000	1.654	13.090	26.114	200.873
Q3出現率	3,865	12.305	8.492	0.000	2.195	11.377	21.890	144.330
Q4出現率	3,865	2.529	2.734	0.000	0.000	1.785	5.678	31.813
Q5出現率	3,865	0.042	0.177	0.000	0.000	0.000	0.153	3.397
Q6出現率	3,865	0.309	0.734	0.000	0.000	0.000	0.970	8.693
Q7出現率	3,865	1.815	2.035	0.000	0.000	1.448	3.698	46.365
Q8出現率	3,865	0.891	0.969	0.000	0.000	0.673	2.024	11.962
Q9出現率	3,865	1.980	2.187	0.000	0.000	1.468	4.317	44.014
Q10出現率	3,865	29.353	14.013	0.000	7.921	30.395	44.849	163.517
総資産の自然対数	3,865	12.673	1.569	6.633	10.844	12.532	14.841	19.532
発行直前ROE	3,865	8.624	11.040	-117.093	1.582	8.129	17.156	188.952

（注）株主資本コストは、QUICKが提供するデータである。1M、3M、6Mは、それぞれ統合報告書発行後1か月、3か月、6か月の値を示す。サンプル数は発行前の直近本決算における自己資本利益率であり、単位は%ポイントである。P10およびP90は、それぞれ第10パーセンタイルおよび第90パーセンタイルを表す。括弧内は企業単位でクラスタリングした頑健標準誤差である。すべてのモデルに企業固定効果および年固定効果を含む。

（出所）筆者作成。

図表3 株主資本コストと辞書出現率の固定効果モデルによる推計結果(発行後3M、発行直前ROEあり)

被説明変数	株主資本コスト(%ポイント)										
変数	Q1個別	Q2個別	Q3個別	Q4個別	Q5個別	Q6個別	Q7個別	Q8個別	Q9個別	Q10個別	Q1～Q10
Q1出現率	0.00334 (0.00721)										-0.00370 (0.00938)
Q2出現率		0.00325 (0.00296)									0.00252 (0.00339)
Q3出現率			0.00090 (0.00333)								-0.00101 (0.00365)
Q4出現率				-0.01329 (0.01250)							-0.01519 (0.01276)
Q5出現率					0.06056 (0.20311)						0.06137 (0.20715)
Q6出現率						0.03004 (0.04435)					0.02273 (0.04480)
Q7出現率							0.00246 (0.01122)				0.00358 (0.01148)
Q8出現率								0.04747* (0.02722)			0.04156 (0.02828)
Q9出現率									0.01166 (0.01287)		0.00904 (0.01454)
Q10出現率										0.00135 (0.00201)	0.00021 (0.00240)
総資産の自然対数	1.62542*** (0.39294)	1.64433*** (0.39216)	1.62828*** (0.39326)	1.62465*** (0.39214)	1.62655*** (0.39296)	1.63522*** (0.39212)	1.62782*** (0.39302)	1.63380*** (0.39285)	1.62797*** (0.39298)	1.62386*** (0.39367)	1.65042*** (0.39206)
発行直前ROE(%)	0.01015** (0.00411)	0.01008** (0.00410)	0.01017** (0.00411)	0.01020** (0.00413)	0.01018** (0.00412)	0.01015** (0.00412)	0.01019** (0.00413)	0.01007** (0.00411)	0.01015** (0.00412)	0.01019** (0.00412)	0.01005** (0.00413)
サンプル数(N)	3865	3865	3865	3865	3865	3865	3865	3865	3865	3865	3865
企業数	1088	1088	1088	1088	1088	1088	1088	1088	1088	1088	1088
Within R ²	0.05141	0.05191	0.05137	0.05193	0.05140	0.05157	0.05135	0.05243	0.05166	0.05151	0.05383
企業固定効果	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
年固定効果	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(注)各列は固定効果モデルによる推計結果である。Q1～Q10個別列は各Qを1変数ずつ投入し、右端列はQ1～Q10を投入した。総資産の自然対数および発行直前ROEをコントロールした。括弧内は標準誤差。***、**、*はそれぞれ1%、5%、10%水準で統計的に有意であることを示す。サンプル数は1M・3M・6Mに共通する3,865企業年(1,088社)。

(出所)筆者作成。

分析結果 固定効果モデル(1) & (2)

(1) ROEなし

Q1～Q10, Q1-10は1M, 3M, 6M いずれの時点でも有意ではない。

(2) ROEあり

Q1～Q10はQ1-10は1M, 3M, 6M いずれの時点でも有意ではない。

仮説「辞書語の出現率と株主資本コストは負に相関する」は支持されない。

辞書指標が専門家評価を捉えていないのか、市場評価に反映されていないのか
この段階では結論付けできない。

説明変数の自然対数化

$$\log_Q_{kit} = \ln(1 + \text{QRATE}_{kit}), \quad k = 1, \dots, 10 \quad (3)$$

QRATE_{kit} は、企業*i*・発行年*t*の統合報告書における各*Q*の辞書語の出現回数を統合報告書全文の有効総語数で標準化した「1万語当たりの辞書語出現回数」である。

$Q1 \sim Q10$ のすべての説明変数に同じ変換を適用する。

1を加えることにより、 QRATE が0の企業年も $\log_Q=0$ として標本に保持する。

説明変数の自然対数化

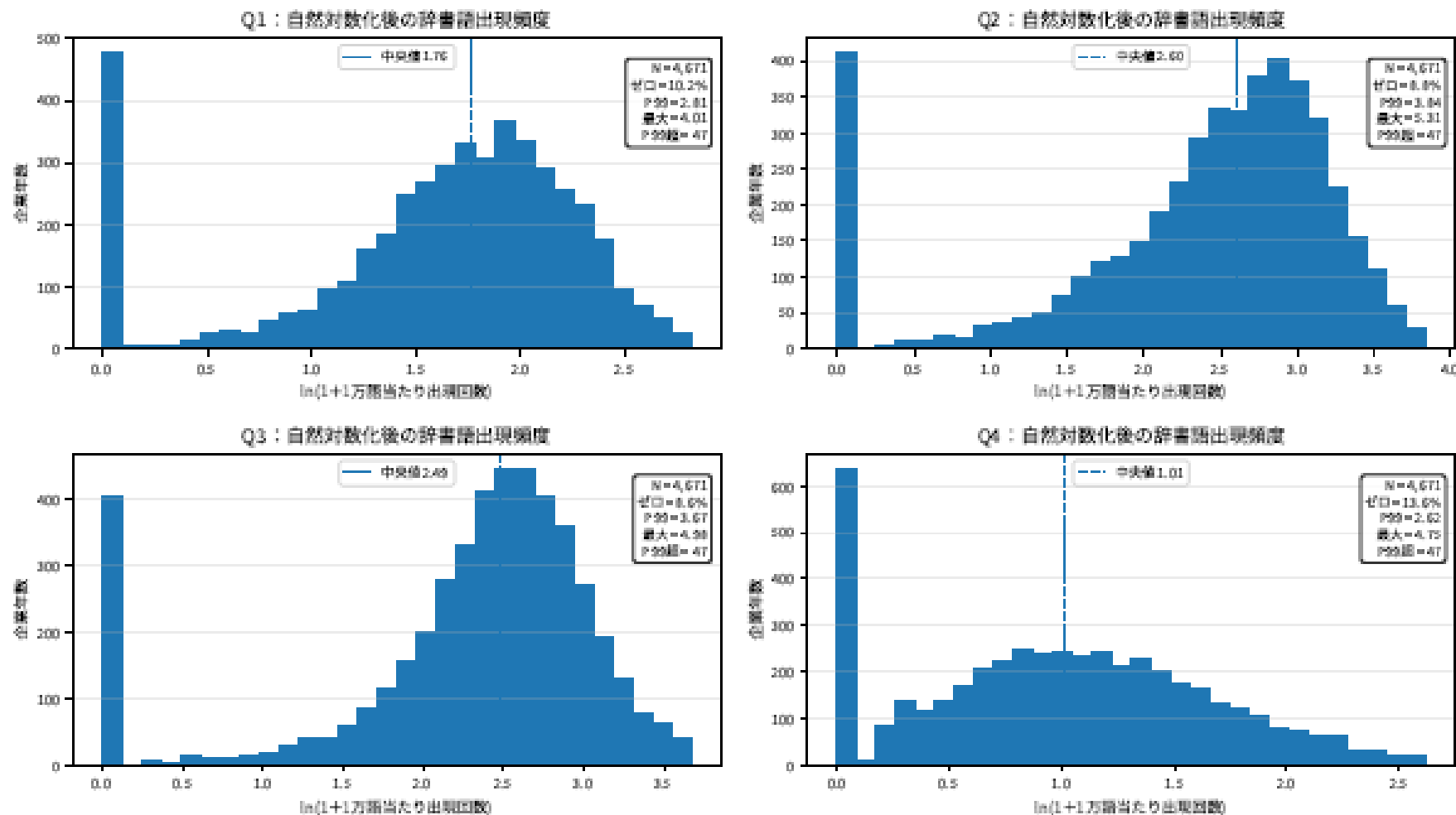
$$\log_Q_{kit} = \ln(1 + \text{QRATE}_{kit}), \quad k = 1, \dots, 10 \quad (3)$$

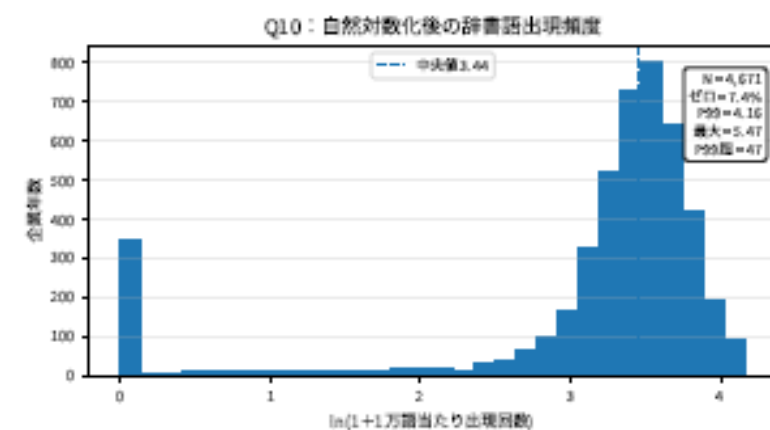
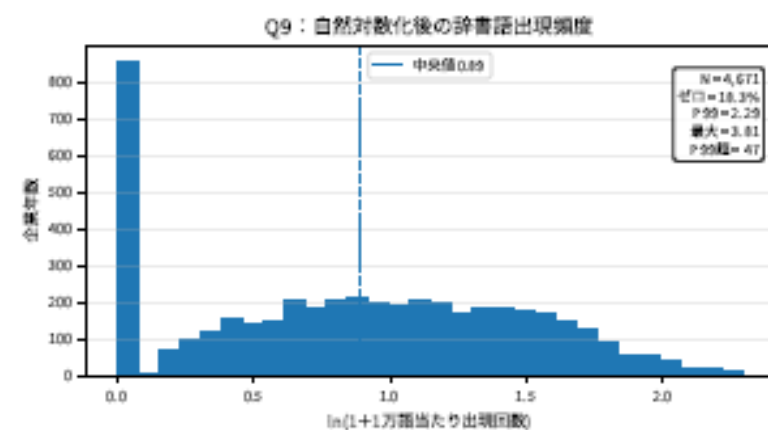
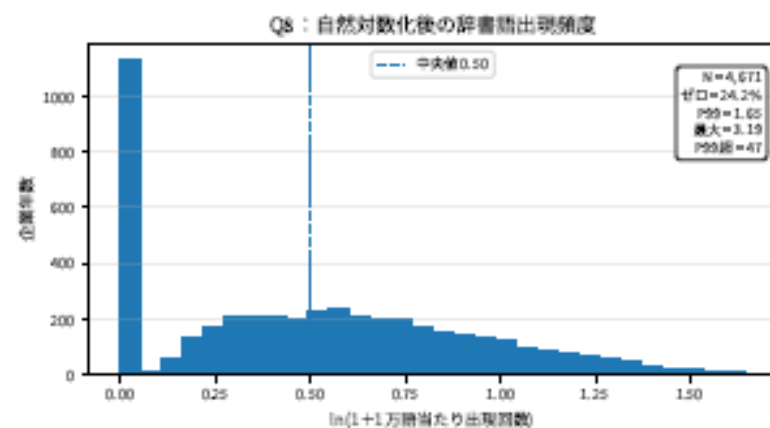
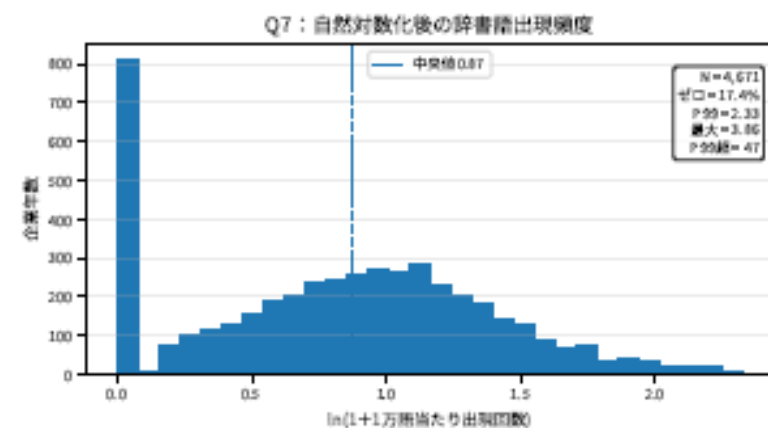
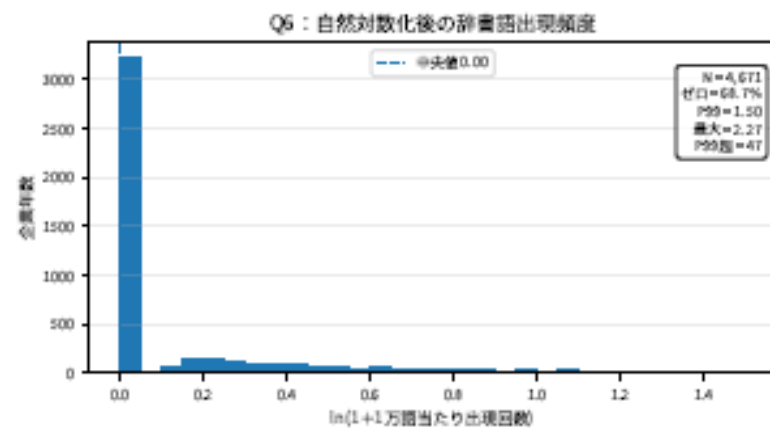
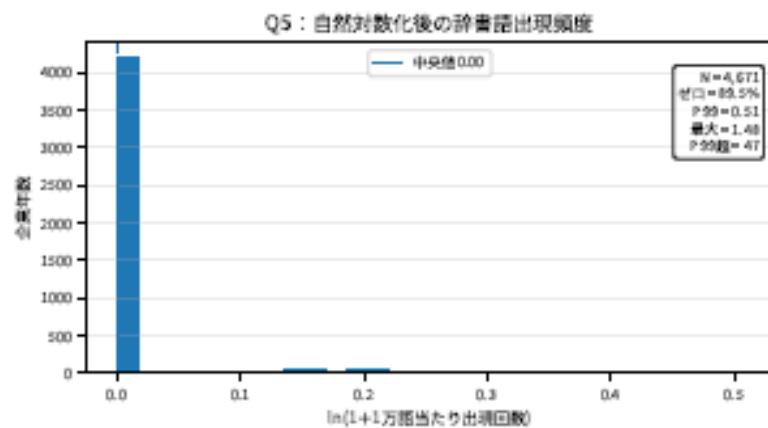
QRATE_{kit} は、企業*i*・発行年*t*の統合報告書における各*Q*の辞書語の出現回数を統合報告書全文の有効総語数で標準化した「1万語当たりの辞書語出現回数」である。

Q1～Q10とQ1-10のすべての説明変数に同じ変換を適用する
1を加えることにより、QRATEが0の企業年も $\log_Q=0$ として保持する。

分析結果 固定効果モデル(3)&(4)

図表4 自然対数化したQ1～Q10 出現頻度の分布





固定効果モデルによる推計

説明変数・株主資本コストの自然対数化

ROEなし

$$\text{COE}_{it_h} = \alpha_i + \lambda_t + \beta_k \log_Q_{kit} + \gamma_1 \ln(\text{Asset}_{it}) + \varepsilon_{it} \quad (3)$$

ROEあり

$$\text{COE}_{it_h} = \alpha_i + \lambda_t + \beta_k \log_Q_{kit} + \gamma_1 \ln(\text{Asset}_{it}) + \gamma_2 \text{ROE_pre}_{it} + \varepsilon_{it} \quad (4)$$

図表5 株主資本コストと自然対数化した辞書語出現頻度の記述統計量(発行直前ROEあり)

変数	サンプルサイズ (N)	平均	標準偏差	最小	P10	中央値	P90	最大
株主資本コスト(1M)	3,865	6.836	2.342	3.291	4.058	6.548	10.105	13.340
株主資本コスト(3M)	3,865	6.831	2.328	3.313	4.083	6.526	10.108	13.469
株主資本コスト(6M)	3,865	6.848	2.314	3.339	4.152	6.508	10.145	13.469
log_Q1 = ln(1+Q1出現頻度)	3,865	1.624	0.701	0.000	0.368	1.758	2.347	4.014
log_Q2 = ln(1+Q2出現頻度)	3,865	2.403	0.947	0.000	0.976	2.645	3.300	5.308
log_Q3 = ln(1+Q3出現頻度)	3,865	2.322	0.870	0.000	1.161	2.516	3.131	4.979
log_Q4 = ln(1+Q4出現頻度)	3,865	1.036	0.659	0.000	0.000	1.024	1.899	3.491
log_Q5 = ln(1+Q5出現頻度)	3,865	0.033	0.116	0.000	0.000	0.000	0.143	1.481
log_Q6 = ln(1+Q6出現頻度)	3,865	0.188	0.351	0.000	0.000	0.000	0.678	2.271
log_Q7 = ln(1+Q7出現頻度)	3,865	0.865	0.566	0.000	0.000	0.895	1.547	3.858
log_Q8 = ln(1+Q8出現頻度)	3,865	0.538	0.429	0.000	0.000	0.515	1.106	2.562
log_Q9 = ln(1+Q9出現頻度)	3,865	0.900	0.609	0.000	0.000	0.903	1.671	3.807
log_Q10 = ln(1+Q10出現頻度)	3,865	3.151	1.000	0.000	2.188	3.447	3.825	5.103
総資産の自然対数	3,865	12.673	1.569	6.633	10.844	12.532	14.841	19.532
発行直前ROE	3,865	8.624	11.040	-117.093	1.582	8.129	17.156	188.952

(注)1M、3M、6Mは、それぞれ統合報告書発行後1か月、3か月、6か月の値を示す。サンプルサイズは、3時点すべての株主資本コストおよび発行直前ROEが利用可能な共通標本3,865企業年(1,088社)である。Q1～Q10は、統合報告書全文1万語当たりの辞書語出現回数(Qk_RATE_10K)に対し、自然対数変換 $\ln(1+Q\text{RATE}k)$ を施した値である。発行直前ROEは統合報告書発行日前の直近本決算におけるROE(%)である。P10およびP90は、それぞれ第10パーセンタイルおよび第90パーセンタイルを表す。

(出所)筆者作成。

図表6 株主資本コストと自然対数化した辞書出現頻度の固定効果モデルによる推計結果(発行後3M、発行直前ROEあり)

被説明変数	株主資本コスト(%ポイント)										
変数	Q1個別	Q2個別	Q3個別	Q4個別	Q5個別	Q6個別	Q7個別	Q8個別	Q9個別	Q10個別	Q1~Q10
log_Q1 = ln(1+Q1出現頻度)	0.00112 (0.04063)										-0.05182 (0.07143)
log_Q2 = ln(1+Q2出現頻度)		0.00251 (0.03186)									-0.01916 (0.05752)
log_Q3 = ln(1+Q3出現頻度)			0.01099 (0.03420)								0.01772 (0.05842)
log_Q4 = ln(1+Q4出現頻度)				-0.11195** (0.05545)							-0.13342** (0.05689)
log_Q5 = ln(1+Q5出現頻度)					0.05556 (0.28558)						0.09531 (0.29199)
log_Q6 = ln(1+Q6出現頻度)						-0.00763 (0.09508)					-0.00080 (0.09624)
log_Q7 = ln(1+Q7出現頻度)							-0.00518 (0.05408)				0.01642 (0.05505)
log_Q8 = ln(1+Q8出現頻度)								0.06447 (0.06824)			0.08069 (0.07442)
log_Q9 = ln(1+Q9出現頻度)									-0.01680 (0.05432)		-0.01832 (0.07039)
log_Q10 = ln(1+Q10出現頻度)										0.01470 (0.02649)	0.04649 (0.05511)
総資産の自然対数	1.62701*** (0.39289)	1.62736** (0.39308)	1.62732** (0.39290)	1.62193*** (0.39171)	1.62636** (0.39304)	1.62656** (0.39337)	1.62663** (0.39291)	1.62939** (0.39284)	1.62659** (0.39288)	1.62439** (0.39352)	1.62290*** (0.39417)
発行直前ROE(%)	0.01018** (0.00412)	0.01017** (0.00412)	0.01015** (0.00411)	0.01026** (0.00414)	0.01018** (0.00412)	0.01018** (0.00412)	0.01018** (0.00412)	0.01009** (0.00411)	0.01020** (0.00413)	0.01017** (0.00412)	0.01025** (0.00416)
サンプル数(N)	3865	3865	3865	3865	3865	3865	3865	3865	3865	3865	3865
企業数	1088	1088	1088	1088	1088	1088	1088	1088	1088	1088	1088
Within R ²	0.05134	0.05134	0.05139	0.05350	0.05136	0.05134	0.05134	0.05172	0.05138	0.05146	0.05492
企業固定効果	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
年固定効果	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(注) 各列は固定効果モデルによる推計結果である。Q1~Q10個別列は自然対数化した各Qを1変数ずつ投入し、右端列は自然対数化したQ1~Q10を投入した。総資産の自然対数および発行直前ROEをコントロールした。括弧内は標準誤差。***、**、*はそれぞれ1%、5%、10%水準で統計的に有意であることを示す。サンプル数は1M・3M・6Mに共通する3,865企業年(1,088社)。自然対数化は $\log_Qk = \ln(1+Qk_RATE_10K)$ により行い、ゼロ観測を保持している。

(出所)筆者作成。

分析結果 自然対数化・固定効果モデル(3) & (4)

(1) ROEなし

Q4は1M 5%, 3M 5%, 6M 10% でいずれも負で有意

Q1-10のQ4は1M, 3M, 6M いずれの時点でも負で5%で有意

他のQは1M, 3M, 6M いずれの時点でも有意ではない。

(2) ROEあり

Q4は1M 5%, 3M 5%, 6M 10% でいずれも負で有意

Q1-10のQ4は1M, 3M, 6M いずれの時点でも負で5%で有意

他のQは1M, 3M, 6M いずれの時点でも有意ではない。

仮説 「辞書語の出現率と株主資本コストは負に相関する」は
支持されたとは言えない。

《参考分析》

辞書指標と専門家評価との基準関連妥当性

1. 目的

日経統合報告書アワード2025の参加企業を母集団とし、日経統合報告書アワードの第1次審査項目Q1からQ10を起点として構築した辞書が、実際のアワード受賞結果と統計的に関連しているかを検証する。

2. 対象

日経統合報告書アワード2025参加企業504企業・団体のうち、本研究では対象外の大学などの23組織・団体を除き、さらに銀行・証券・保険・その他金融など57社を除いた424社を分析対象とした。

分析方法

- ロジット分析

分析対象424社を受賞企業(1)と非受賞企業(0)として、これらを被説明変数とするロジット分析(logistic analysis, logit)を行う。

- 説明変数の自然対数化

説明変数は1万語当たりの出現回数 Q に1を加えた自然対数 $\ln(1+Q)$ 。

- コントロール変数と簡易業種ダミーの組合せで、3通りの分析を行う。

QUICK株主資本コスト：3つの分析モデル

被説明変数：統合報告書発行後 1、3、6 か月の 株主資本コスト

モデル 1：1M

$$COE_{it}^{(1M)} = \alpha_i + \lambda_t + \sum_{k=1}^{10} \beta_k \ln(1 + Q_{kit}) + \gamma_1 \ln Asset_{it} + \gamma_2 ROEpre_{it} + \varepsilon_{it}$$

企業固定効果 + 年固定効果

モデル 2：3M

$$COE_{it}^{(3M)} = \alpha_i + \lambda_t + \sum_{k=1}^{10} \beta_k \ln(1 + Q_{kit}) + \gamma_1 \ln Asset_{it} + \gamma_2 ROEpre_{it} + \varepsilon_{it}$$

企業固定効果 + 年固定効果

モデル 3：6M

$$COE_{it}^{(6M)} = \alpha_i + \lambda_t + \sum_{k=1}^{10} \beta_k \ln(1 + Q_{kit}) + \gamma_1 \ln Asset_{it} + \gamma_2 ROEpre_{it} + \varepsilon_{it}$$

企業固定効果 + 年固定効果

変数

変数	内容・測定方法
$COE_{it}^{(h)}$	株主資本コスト。統合報告書発行後1M・3M・6Mの各時点の値。
$\ln(1+Q_{kit})$	Q1～Q10対応辞書語の1万語当たり出現数 Qk_RATE_10K を $\ln(1+Q)$ に変換。 ゼロ観測は削除しない。
$\ln Asset_{it}$	統合報告書発行前の直近本決算に対応する総資産の自然対数。 企業規模のコントロール変数。
$ROE_{pre_{it}}$	統合報告書発行日前の直近本決算におけるROE(%)。 発行直前の収益性をコントロール。
α_i	企業固定効果。企業ごとに時間を通じて不変な観察不能要因を統制。
λ_t	年固定効果。各年に共通するマクロ環境・市場環境等を統制。
ε_{it}	誤差項

図表7 2025年日経統合報告書アワード参加企業(除く金融)の記述統計量

変数	サンプル数 (N)	平均	標準偏差	最小	P10	中央値	P90	最大
受賞ダミー(受賞=1、非受賞=0)	437	0.124	0.329	0.000	0.000	0.000	1.000	1.000
log_Q1 = $\ln(1+Q1_RATE_10K)$	437	1.766	0.687	0.000	0.889	1.913	2.458	3.257
log_Q2 = $\ln(1+Q2_RATE_10K)$	437	2.426	0.908	0.000	1.200	2.661	3.308	3.938
log_Q3 = $\ln(1+Q3_RATE_10K)$	437	2.489	0.831	0.000	1.791	2.642	3.197	4.134
log_Q4 = $\ln(1+Q4_RATE_10K)$	437	1.261	0.624	0.000	0.395	1.267	2.053	3.087
log_Q5 = $\ln(1+Q5_RATE_10K)$	437	0.054	0.131	0.000	0.000	0.000	0.188	0.871
log_Q6 = $\ln(1+Q6_RATE_10K)$	437	0.202	0.363	0.000	0.000	0.000	0.726	2.218
log_Q7 = $\ln(1+Q7_RATE_10K)$	437	0.893	0.514	0.000	0.199	0.888	1.588	2.654
log_Q8 = $\ln(1+Q8_RATE_10K)$	437	0.534	0.411	0.000	0.000	0.497	1.074	2.204
log_Q9 = $\ln(1+Q9_RATE_10K)$	437	1.155	0.614	0.000	0.000	1.267	1.868	2.882
log_Q10 = $\ln(1+Q10_RATE_10K)$	437	3.176	0.882	0.000	2.488	3.412	3.803	4.121
企業規模($\ln_asset$)	437	13.229	1.504	9.573	11.471	13.062	15.213	19.510

(注)2025年日経統合報告書アワード参加企業のうち銀行・証券・保険・その他金融を除外した共通サンプル437社を対象とする。受賞企業54社、非受賞企業383社。Q1～Q10は統合報告書全文1万語当たりの辞書語出現回数(Qk_RATE_10K)を $\ln(1+Qk_RATE_10K)$ により自然対数化した値である。企業規模は総資産の自然対数($\ln_asset$)。P10およびP90は、それぞれ第10パーセンタイルおよび第90パーセンタイルを表す。

(出所)筆者作成

図表8 Logitモデル1による推計結果(企業規模・業種ダミーなし)

被説明変数:日経統合報告書アワード受賞ダミー(受賞=1、非受賞=0)

変数	Q1個別	Q2個別	Q3個別	Q4個別	Q5個別	Q6個別	Q7個別	Q8個別	Q9個別	Q10個別	Q1～Q10
log_Q1 = ln(1+Q1出現頻度)	-0.04572 (0.23413)										-0.18923 (0.44964)
log_Q2 = ln(1+Q2出現頻度)		-0.00765 (0.17158)									-0.03932 (0.28730)
log_Q3 = ln(1+Q3出現頻度)			0.16668 (0.23867)								0.38632 (0.31258)
log_Q4 = ln(1+Q4出現頻度)				0.84034*** (0.24979)							0.88751*** (0.23644)
log_Q5 = ln(1+Q5出現頻度)					0.25012 (1.04506)						0.14991 (1.09420)
log_Q6 = ln(1+Q6出現頻度)						0.37628 (0.38419)					0.25152 (0.37740)
log_Q7 = ln(1+Q7出現頻度)							0.11419 (0.27881)				-0.00670 (0.32607)
log_Q8 = ln(1+Q8出現頻度)								0.17847 (0.34111)			0.28758 (0.34698)
log_Q9 = ln(1+Q9出現頻度)									0.49973* (0.28690)		0.78015** (0.32132)
log_Q10 = ln(1+Q10出現頻度)										-0.11902 (0.12916)	-0.81346** (0.35894)
サンプル数(N)	424	424	424	424	424	424	424	424	424	424	424
McFadden擬似R ²	0.0001	0.0000	0.0025	0.0370	0.0002	0.0032	0.0005	0.0008	0.0129	0.0019	0.0807
企業規模(ln_asset)	No	No	No	No	No	No	No	No	No	No	No
業種ダミー	No	No	No	No	No	No	No	No	No	No	No

(注)各列はLogitモデルによる推計結果である。Q1～Q10個別列は自然対数化した各Qを1変数ずつ投入し、右端列は自然対数化したQ1～Q10を投入した。括弧内は標準誤差。***、**、*はそれぞれ1%、5%、10%水準で統計的に有意であることを示す。金融4業種55社を除外し424社を用いた。

(出所)筆者作成。

図表9 Logitモデル2による推計結果(企業規模(ln_asset)統制)

被説明変数:日経統合報告書アワード受賞ダミー(受賞=1、非受賞=0)

変数	Q1個別	Q2個別	Q3個別	Q4個別	Q5個別	Q6個別	Q7個別	Q8個別	Q9個別	Q10個別	Q1～Q10
log_Q1 = ln(1+Q1出現頻度)	▼ -0.07152 (0.22970)										▼ -0.39381 (0.43565)
log_Q2 = ln(1+Q2出現頻度)		▼ 0.08698 (0.17944)									▼ 0.18124 (0.26977)
log_Q3 = ln(1+Q3出現頻度)			▼ 0.13736 (0.23519)								▼ 0.29932 (0.29338)
log_Q4 = ln(1+Q4出現頻度)				▼ 0.80381*** (0.25845)							▼ 0.88809*** (0.24178)
log_Q5 = ln(1+Q5出現頻度)					▼ 0.31256 (1.15064)						▼ 0.54217 (1.16249)
log_Q6 = ln(1+Q6出現頻度)						▼ 0.49985 (0.41008)					▼ 0.35073 (0.38442)
log_Q7 = ln(1+Q7出現頻度)							▼ 0.09061 (0.27725)				▼ -0.01114 (0.31643)
log_Q8 = ln(1+Q8出現頻度)								▼ 0.24710 (0.34934)			▼ 0.36336 (0.35976)
log_Q9 = ln(1+Q9出現頻度)									▼ 0.54294* (0.29296)		▼ 0.85635*** (0.32725)
log_Q10 = ln(1+Q10出現頻度)										▼ -0.10839 (0.14009)	▼ -0.83491** (0.34246)
総資産の自然対数	▼ 0.41946*** (0.09479)	▼ 0.42610*** (0.09468)	▼ 0.41605*** (0.09438)	▼ 0.41145*** (0.09774)	▼ 0.41880*** (0.09413)	▼ 0.42660*** (0.09584)	▼ 0.41863*** (0.09488)	▼ 0.42259*** (0.09440)	▼ 0.42831*** (0.09767)	▼ 0.41690*** (0.09445)	▼ 0.46148*** (0.10024)
サンプル数(N)	424	424	424	424	424	424	424	424	424	424	424
McFadden擬似R ²	0.0582	0.0586	0.0595	0.0894	0.0581	0.0632	0.0581	0.0593	0.0725	0.0592	0.1385
企業規模(ln_asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
業種ダミー	No	No	No	No	No	No	No	No	No	No	No

(注)各列はLogitモデルによる推計結果である。Q1～Q10個別列は自然対数化した各Qを1変数ずつ投入し、右端列は自然対数化したQ1～Q10を入した。すべての列で総資産の自然対数(ln_asset)をコントロールした。括弧内は標準誤差。***、**、*はそれぞれ1%、5%、10%水準で統計的に有意であることを示す。金融4業種55社を除外した共通標本424社を用いた。

(出所)筆者作成。

図表10 Logitモデル3による推計結果（企業規模・5業種分類統制）

被説明変数：日経総合報告書アワード受賞ダミー（受賞=1、非受賞=0）

変数	Q1個別	Q2個別	Q3個別	Q4個別	Q5個別	Q6個別	Q7個別	Q8個別	Q9個別	Q10個別	Q1～Q10
log_Q1 = ln(1+Q1出現頻度)	-0.08908 (0.22269)										-0.46676 (0.44408)
log_Q2 = ln(1+Q2出現頻度)		0.10443 (0.17871)									0.19207 (0.27153)
log_Q3 = ln(1+Q3出現頻度)			0.17625 (0.24897)								0.35069 (0.28426)
log_Q4 = ln(1+Q4出現頻度)				0.77267*** (0.27053)							0.85794*** (0.24883)
log_Q5 = ln(1+Q5出現頻度)					0.56512 (1.19305)						0.71945 (1.21877)
log_Q6 = ln(1+Q6出現頻度)						0.51317 (0.40503)					0.40822 (0.39039)
log_Q7 = ln(1+Q7出現頻度)							0.02408 (0.27434)				-0.06958 (0.32117)
log_Q8 = ln(1+Q8出現頻度)								0.14733 (0.37439)			0.22785 (0.36491)
log_Q9 = ln(1+Q9出現頻度)									0.52208* (0.28981)		0.81203** (0.32890)
log_Q10 = ln(1+Q10出現頻度)										-0.08141 (0.14217)	-0.74925** (0.35371)
総資産の自然対数	0.49036*** (0.10581)	0.49913*** (0.10649)	0.48811*** (0.10583)	0.48149*** (0.11033)	0.49193*** (0.10467)	0.49765*** (0.10603)	0.48863*** (0.10568)	0.48961*** (0.10514)	0.49730*** (0.10744)	0.48600*** (0.10522)	0.53540*** (0.11486)
消費財ダミー	0.83659 (0.56492)	0.83595 (0.56505)	0.90730 (0.57502)	1.02743* (0.58928)	0.83924 (0.56553)	0.89403 (0.54531)	0.83738 (0.56201)	0.84244 (0.56697)	0.83019 (0.57595)	0.82935 (0.56476)	1.04926* (0.59304)
投資財ダミー	0.37043 (0.54998)	0.36917 (0.55138)	0.41607 (0.55604)	0.58806 (0.57293)	0.37253 (0.55324)	0.39682 (0.53505)	0.36839 (0.54994)	0.36223 (0.55123)	0.35791 (0.56217)	0.36647 (0.55186)	0.68593 (0.60423)
サービスダミー	-0.07022 (0.63600)	-0.06799 (0.63944)	-0.00877 (0.64092)	0.29138 (0.66603)	-0.05619 (0.64052)	-0.05074 (0.63254)	-0.06593 (0.63785)	-0.06238 (0.63911)	-0.13417 (0.66334)	-0.07188 (0.63676)	0.24521 (0.72099)
公益ダミー	—	—	—	—	—	—	—	—	—	—	—
サンプル数(N)	424	424	424	424	424	424	424	424	424	424	424
McFadden擬似R ²	0.0927	0.0934	0.0948	0.1198	0.0929	0.0978	0.0922	0.0927	0.1056	0.0929	0.1656
企業規模(ln asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
業種ダミー	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(注) 各列はLogitモデルによる推計結果である。Q1～Q10個別列は自然対数化した各Qを1変数ずつ投入し、右端列は自然対数化したQ1～Q10を投入した。5業種分類の業種ダミーをコントロールした。括弧内は標準誤差。
***、**、*はそれぞれ1%、5%、10%水準で統計的に有意であることを示す。

(出所) 筆者作成。

辞書指標と専門家評価との基準関連妥当性に関する分析の結果

図表11 アワード受賞企業分析結果 3 Modelの比較

Model 1				Model 2				Model 3			
Q個別分析		Q1-10 分析		Q個別分析		Q1-10 分析		Q個別分析		Q1-10 分析	
変数	推定値	変数	推定値	変数	推定値	変数	推定値	変数	推定値	変数	推定値
	標準誤差		標準誤差		標準誤差		標準誤差		標準誤差		標準誤差
Q4	0.84034***	Q4	0.88751***	Q4	0.80381***	Q4	0.88809***	Q4	0.77267***	Q4	0.85794***
	(0.24979)		(0.23644)		(0.25845)		(0.24178)		(0.27053)		(0.24883)
Q9	0.49973*	Q9	0.78015**	Q9	0.54294*	Q9	0.85635***	Q9	0.52208*	Q9	0.81203**
	(0.28690)		(0.32132)		(0.29296)		(0.32725)		(0.28981)		(0.32890)
Q10	-0.11902	Q10	-0.81346**	Q10	-0.10839	Q10	-0.83491**	Q10	-0.08141	Q10	-0.74925**
	(0.12916)		(0.35894)		(0.14009)		(0.34246)		(0.14217)		(0.35371)

辞書指標と専門家評価との基準関連妥当性に関する分析の結果

- Q4に関しては3つモデルともに、個別QおよびQ1-10の各分析において、有意水準1%において正の相関を示している。
- この結果は、Q4が捉える経営資源・無形資産と中長期戦略に関する記述が、専門家による統合報告書評価と関連している可能性を示唆する。

Q4 自社の経営資源(人的資本、知的資本、社会・関係資本、自然資本など無形の資本を含む)の冷静な分析と中長期経営目標(KPI)・戦略(ビジネスモデルの変革含む)に関する記述

5. 本研究から得られた示唆と今後の課題

本研究から得られた示唆

- ・ 2021～25年に発行された統合報告書を対象として、日経統合報告書アワード2025の第1次審査10項目を起点とする独自辞書を構築し、統合報告書の記述内容を定量化した。
- ・ 辞書語の出現率と株主資本コストとの関係を固定効果モデルにより検証した結果、自然対数化後のQ4では一部に負の有意な関係が確認されたものの、Q1～Q10全体として「統合報告書に基づく辞書語の出現率と株主資本コストは負に相関する」という仮説が支持されたとは言えない。

- 一方、日経統合報告書アワード2025を用いた参考分析では、Q4を中心に辞書指標と専門家評価との統計的な関連が確認された。
- 本研究の結果は、統合報告書の記述内容について、専門家が評価する特徴と、資本市場における株主資本コストとして評価される特徴が必ずしも一致しない可能性を示唆する。
- ただし、本研究は因果関係を示すものではなく、現段階の辞書は試作版であるため、結果の解釈には慎重さが必要である。

今後の課題

1. 分析方法の再検討

(1) 各Qについて記述の「あり・なし」で分け、記述のあるQと株主資本コストの関係を分析する。

(2) 総合指標の作成

① 各Qと平均 $\bar{Q}$ との差分を標準偏差で除して標準化する(ZQ_{kit})。

ZQ_{kit} の平均を総合指標とし、株主資本コストとの関係を分析する。

$$ZQ_{kit} = \frac{Q_{kit} - \bar{Q}}{SD(Q_k)} \quad \text{Composite } Z_{kit} = \frac{1}{10} \sum_{k=1}^{10} ZQ_{kit}$$

② 自然対数化したQを用いた総合指標と株主資本コストとの関係を分析する。

$$\log Q_{kit} = \ln(1 + Q_{kit}) \quad \text{Composite } Z_{kit} = \frac{1}{10} \sum_{k=1}^{10} Z(\log Q_{kit})$$

2. 辞書構築の再検討

- 辞書の精緻化

現行辞書をVer.1として固定した上で、Q間の語彙重複、低頻度語、文脈依存語等を再検討し、辞書Ver.2を構築する。

- 辞書構築方法の拡張

Word2Vec等を用いた関連語拡張や、人手レビューとの組合せにより、各Qの概念をより適切に捉える辞書構築方法を検討する。

- テキスト指標の改善

単純な出現率だけでなく、開示の有無(Coverage)、記述の強度(Intensity)等を区別した指標を検討する。

- 市場評価との関係の追加検証

株主資本コストの測定方法、企業特性等を考慮した追加分析を行うとともに、専門家評価と市場評価の差異が生じる要因を検討する。

- Seed Words の再検討

業種別辞書というハードルは上がるものの、日経統合報告書アワード第1次審査項目に代えてSASB Standard による独自辞書の構築を検討する。

参考文献

- 内野里美(2005)「自発的な情報開示が自己資本コストに与える影響」『現代ディスクロージャー研究』第6号 pp.15-25.
- 黒木裕鷹・真鍋友則・中川慧(2024)「テキスト分析を用いた本邦決算説明会の情報価値」『人工知能学会論文誌』39巻4号 pp.1-8.
- 後藤嘉孝(2026)「有価証券報告書におけるサステナビリティ情報開示と株主資本コストに関する実証分析」『日本経営倫理学会誌』第33号 pp.253-267.
- 野田健太郎(2016)「有価証券報告書における定性情報の分析と活用 ―リスクの多様化にともなう望ましい対話のあり方―」『経済経営研究』Vol.35 No.1, pp.1-51.
- 村宮克彦(2005)「経営者が公表する予想利益の精度と資本コスト」『証券アナリストジャーナル』43(9), pp.82-96.

- 山本零、川代 尚哉、栗田昌孝(2020)「決算短信と四季報テキスト情報の投資戦略への利用可能性検証」『ジャフィー・ジャーナル』18巻 pp.46–62.
- Botosan, C. A. and M. A. Plumlee (2002), “A Re-examination of Disclosure Level and the Expected Cost of Equity Capital,” *Journal of Accounting Research*, 40(1), pp.21–40.
- Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M. and Steele, L. B. (2014), “The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings,” *Review of Accounting Studies*, 19(1), PP.396–455.
- Francis, J., D. Nanda and P. Olsson (2008), “Voluntary Disclosure, Earnings Quality, and Cost of Capital,” *Journal of Accounting Research*, 46(1), pp.53–99.
- Hail, L. (2002), “The Impact of Voluntary Corporate Disclosures on the Ex-ante Cost of Capital for Swiss Firms,” *European Accounting Review*, 11(4), pp.741–773.

- Li, Feng (2008) “Annual Report Readability, Current Earnings, and Earnings Persistence,” *Journal of Accounting and Economics*, 45, pp.221–247.
- Loughran, T. and McDonald, B. (2011), “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks,” *The Journal of Finance*, 66(1), PP.35–65.
- Tetlock, Saar-Tsechansky and Macskassy (2008) “More Than Words: Quantifying Language to Measure Firms’ Fundamentals,” *Journal of Finance*, 63, pp.1437–1467.