人工知能学会第二種研究資料
ビジネス・インフォマティクス研究会
SIG-BI BI-019-02

データマイニング法を用いた不正会計検知手法の検討

Study of accounting fraud detection using a data-mining approach

小畑崇弘 [1] 細尾忠敬 [1] 倉橋節也 [1]
Takahiro Obata[1] Tadataka Hosoo[1] Setsuya Kurahashi[1]

[1] 筑波大学大学院ビジネス科学研究科
[1] Graduate School of Business Sciences, University of Tsukuba

Abstract: 本研究では，データマイニング法（DM 法）を活用した不正会計検知手法の分析結果について報告する．DM 法とは，株式ファクター研究で近年提案された手法で，計算可能なファクターを網羅的に分析することで従来研究では注目されてこなかったが有効なファクターを見出そうとするものである．DM 法を用いたファクターの検証から不正会計検知モデル構築，予測結果までを報告する．

1 はじめに

上場企業による不正会計開示件数は年々増加傾向にあり，大きな社会問題となっている．不正会計の発覚は株価の急落や経営破綻といった重大な企業価値の棄損を引き起こすことが知られており，不正会計の恐れがある企業を検知しようとする研究が盛んに行われている．

本研究では，不正会計検知モデルのインプットデータとして用いる財務シグナルの検証に，データマイニング法を用いる．データマイニング法は，近年，株式ファクター研究の手法として Yan/Zheng[1] が提案したものであり，特定の財務シグナルに絞って分析するのではなく，分析可能な財務シグナルを網羅的に分析することで研究者の思い込みを排除し、新たな知見を得ようとする狙いがである．不正会計検知研究の代表的な先行研究でも、研究者が一部の財務データ等から作成したシグナルを活用しており，活用されているシグナルや財務データ以外にも有効なデータが潜んでいる可能性は拭えない．そこで，本研究ではまずデータマイニング法により不正会計検知に有効と考えられた財務シグナルを網羅的に検証する．さらに有効な財務シグナルのアウトプットを活用した不正会計検知モデル構築し，不正会計検知精度の向上を目指す．

本稿の構成は次の通りである．Section 2 で，関連研究として不正会計検知の代表的モデルやデータマイニング法について概略する．Section 3 は本研究の分析データについての説明である．Section 4 はデータマイニング法による財務シグナルの検証，Section 5 は不正会計検知モデルによる予測結果について報告する，Section 6 はまとめである．

2 関連研究

2.1 不正会計検知

不正会計検知の代表的手法として Beneish[2] が提唱した M-Score や Dechow et al[3] による F-Score がある．どちらの手法も米国企業に対する分析を基に構築されたもので，過去の実証研究などを参考に不正会計検知に有効と思われる財務シグナルを複数準備し、それらの財務シグナルの線形和を用いて不正の確率を推計するアプローチを取っている．下記は F-score の算出方法である．

$$Fscore = \frac{Probability}{0.004} \tag{1}$$

$$Probability = \frac{\exp(PredictedValue)}{1+\exp(PredictedValue)} \tag{2}$$

$$\begin{aligned} PredictedValue = {} & -7.893 + 0.790 rsst_{qcc} \\ & +2.518 ch_{rec} + 1.191 ch_{inv} \\ & +1.979 soft_{asset} + 0.171 ch_{cs} \\ & -0.932 ch_{roa} + 1.029 issue \end{aligned} \tag{3}$$

ここで (1) 式の分母の 0.004 は全サンプルにおける不正会計データの割合であり，(3) 式の各変数は，$rsst_{qcc}$: アクルアルズ[1]/総資産期中平均，ch_{rec}:売掛債権の増減額/総資産期中平均，ch_{inv}:棚卸資産の増減額/総資産期中平均，$soft_{asset}$:（総資産-償却対象有形固定資産-現預金）/総資産，ch_{cs}:売上高-売掛債権の増減額，ch_{roa}: ROA の増減，$issue$:新規資金調達の有無に関す

[1]アクルアルズは現金収入を伴う質の高い利益かどうかを見極めるために用いられる指標で，特別損益を除いた税引き後利益から営業キャッシュフローを引いて算出する．アクルアルズの値が小さい，またはマイナスの企業の利益は現金収入に裏付けられた質の高い利益といえる．

るダミー変数，である．国内においても不正会計検知研究は盛んに行われており，近年の研究としては東海林 [4] がある．東海林の研究では，不正会計データには財務データ間の歪みが存在するという監査実務の知見を反映するため，財務データ間の共分散とマハラノビス距離を活用して加工したインプットデータを用いて不正検知予測を行っている．

2.2 データマイニング法

データマイニング法（DM 法）は，前述の Yan/Zheng らが株式ファクター研究の中で用いた方法で，研究者の予見をできるだけ排除して客観的･網羅的に有意な財務シグナルを探索することを目的としている．国内株式市場において山田/後藤 [5] が DM 法を用いて分析を行い，これまであまり注目されてこなかった現金や人件費関連、支払い金利の変化といった財務ファクターの有効性を確認したと報告している．DM 法についてはその効用の一方で DM 法で発見された成果が，過適合や偽発見といった見せかけのパターン発見に過ぎない可能性などの懸念点も指摘されている．しかし，過適合や偽発見の多くは，研究者側が恣意的に研究デザインを設定できることに起因する面があると考えられることから，分析結果の一部のみを提示していることを懸念されるよりも網羅的に検証を行い結果を提示する DM 法の意義は小さくないと考えらえる．

3 分析データ

本研究では国内上場企業の 2008 年 3 月期から 2018 年 3 月期までの決算期間にあたる本決算および四半期決算を分析対象とした．但し，先行研究に倣い，一般的な企業の資金調達手法，財務構造や収益構造が大きく異なる銀行業など一部の業種を除外している．

研究のためには不正のない決算データ（以後，正常データと呼ぶ）と不正が行われている決算データ（以後，不正データと呼ぶ）が必要になる．正常データは不正を行っていない企業の決算データと不正を行ったが後に訂正報告がなされた決算データで構成される．正常データは日経 Needs 財務データ DVD 版により取得し，データ抜けなどがみられた場合は日経バリューサーチにより補完した．不正データについては，証券取引等監視委員会（Securities and Exchange Surveillance Commission，SESC）のウェブサイトで公表されている事例を収集した．さらに上場企業で不正会計が行われた場合にほとんどの企業が調査委員会や第三者委員会を設置して調査を行うことから，日経バリューサーチのキーワード検索機能を利用して’調査委員会’，’第三者委員会’，’社内調査’，’内部調査’，’外部調査’，’外部の専門家’，’事実が判明’，’調査報告書’ のキーワード抽出を行い，開示内容により経営層および経営に近い部門長レベルが関与していると確認できた事例を不正データとして用いた．表 1 は上記方法により収集した分析対象データの企業数および延べ決算期数をまとめたものである．なお，各決算期データにおける財務データ項目数は約 1,200 だが、全ての財務データ項目に数値が入っているわけではなく空欄になっているものが多い．

表 1: 企業数と延べ決算期数

	正常データ	不正データ
企業数	3,655	141
延べ決算期数	132,540	1,404

4 DM 法による不正会計検知に有効な財務シグナルの検証

前述の DM 法を用いた先行研究では，株価の将来リターンに対する各財務シグナルの単回帰係数の有意性検証をベースに分析を行っているが，本研究では不正の有無を教師データ，各財務シグナルをインプットデータとした決定木モデルを構築し、その正解率をみることで各財務シグナルの有効性を測ることとした．決定木モデルを採用したのは，教師データが不正の有無のため単回帰分析がなじまないことに加え，不正を行った企業の財務シグナル値が両端に偏っているケース（またはその逆）にも対応できるようにするためである．さらに，最終的な不正会計予測モデルではモデルの説明力が求められる可能性もあり，その点で予測モデルのアウトプットに至る判断過程をたどり易くしておく点も考慮した．具体的な検証モデルとしては最大分岐数が 1〜4 までの 4 種類の決定木モデルと，対比のためにロジスティック回帰モデルについても予測正解率を計測した．

インプットデータとしては各財務項目について 11 通りの財務シグナルを計算した．財務シグナルの計算方法については表 2 にまとめた．

次に各財務シグナルに対応する決定木モデルの学習および評価データだが，全ての決算期データを用いると正常データ数と不正データ数が大きく偏ってしまい，そうしたデータでの学習の結果，予測が不正なしのみになる恐れがある．この点を回避するため，本研究では不正あり企業の総数 141 社と同数の不正なし企業をランダムに選択するアンダーサンプリングを行い，不正あり企業およびアンダーサンプリングにより得られた不正なし企業の全ての決算データを用いて決定木モ

表 2: 財務項目 x に関する財務シグナルの計算方法

項目	内容
分母 (y)	総資産, 自己資本, 売上高
水準型シグナル	x_t/y_t
前期比変化幅型シグナル	$(x_t - x_{t-1})/y_{t-1}$
前年比変化幅型シグナル	$(x_t - x_{t-4})/y_{t-4}$
前期比成長型シグナル	$(x_t - x_{t-1})/x_{t-1}$
前年比成長型シグナル	$(x_t - x_{t-4})/x_{t-4}$

（注）添え字 t はどの期のデータかを表す．t-1 は 1 四半期前，t-4 は 1 年前を表す．

図 1: 学習データとテストデータの分割

デルの学習および評価をすることとした．不正あり企業の全ての決算データの中には不正がなかった期間のデータも含むことになる．不正あり企業の延べ決算期数は 5,069 期分あり，そのうちの 3,665 期分は不正なしの正常データ，残りの 1,404 期分が不正データである．これ以降で掲示する財務シグナルの検証結果は，ある一度のアンダーサンプリングにより得られた結果を示しており，その際の不正なし企業の決算データは企業数が 141 社，延べ決算期数は 5,034 期分で全て正常データである．

こうして取得した決算データについて決算年をもとに 2008 年分，2009 年分，…，2016 年分，2017〜2018 年分の 10 期間に分割し，1 期間分をテストデータ，それ以外を学習データとする 10th 交差検証法により各財務シグナルの有効性を検証した．図 1 は学習データとテストデータの分け方のイメージを示したものである．

各財務シグナルの有効性はテストデータにおける不正データの予測正解率を基にした．表 3〜5 はそれぞれ最大分岐数 2 の決定木モデル，最大分岐数 1 の決定木モデル（切り株モデル），およびロジスティック回帰モデルによる不正予測の正解率上位 10 シグナルを示したものである．最大分岐数 3 および 4 の決定木モデルは最大分岐数 2 の決定木モデルと非常に似通った結果のため省略した．なお，不正予測正解率を測定したのは，学習データ，テストデータの双方において財務シグナル値が計算できた割合がそれぞれ 3 分の 1 以上あった

表 3: 決定木（最大分岐数 2）による財務シグナル正解率上位 10

NO	signal 型	分子の財務項目	正解率（学習）		正解率（test）	
			不正無	不正有	不正無	不正有
1	ydiff_eq	受取配当金	98.8%	9.7%	98.5%	8.6%
2	ta	繰延税金資産	98.7%	7.5%	98.7%	5.6%
3	ta	有形固定資産処分損	99.5%	5.3%	99.3%	4.5%
4	sl	その他資産処分損	99.4%	4.4%	99.3%	3.4%
5	ygrowth	特別損失	100.0%	3.6%	100.0%	3.3%
6	qgrowth	その他資産処分損	99.3%	5.5%	99.3%	2.9%
7	qgrowth	特別損失	100.0%	3.4%	100.0%	2.9%
8	ydiff_sl	機械装置及び運搬具	100.0%	0.5%	100.0%	2.9%
9	ygrowth	その他資産処分損	99.6%	4.3%	99.3%	2.8%
10	sl	貸倒引当金	100.0%	2.6%	100.0%	2.7%

（注）signal 型の'ta' は水準型で分母が総資産，'eq' は水準型で分母が自己資本，'sl' は水準型で分母が売上高の財務シグナルを表す．qdiff_** は前期比変化幅型，ydiff_**は前年比変化幅型を表し，**の部分は分母を示す（水準型と同じ記号）．qgrowth は前期比成長型，ygrowth は前年比成長型を表す．

表 4: 切り株モデルによる財務シグナル正解率上位 10

NO	signal 型	分子の財務項目	正解率（学習）		正解率（test）	
			不正無	不正有	不正無	不正有
1	ydiff_eq	受取配当金	98.9%	9.0%	98.7%	8.8%
2	sl	その他資産処分損	99.0%	6.6%	99.0%	6.3%
3	ygrowth	特別損失	100.0%	3.6%	100.0%	3.3%
4	qgrowth	特別損失	100.0%	3.4%	100.0%	2.9%
5	sl	貸倒引当金	100.0%	2.6%	99.9%	2.7%
6	ygrowth	貸倒引当金	99.9%	2.5%	99.9%	2.5%
7	qgrowth	貸倒引当金	100.0%	2.5%	99.9%	2.4%
8	ta	退職給付に係る負債	99.8%	2.5%	99.8%	2.4%
9	sl	有形固定資産処分損	99.5%	3.4%	99.3%	2.3%
10	ygrowth	その他資産処分損	99.2%	4.8%	98.9%	2.3%

（注）signal 型の'ta' は水準型で分母が総資産，'eq' は水準型で分母が自己資本，'sl' は水準型で分母が売上高の財務シグナルを表す．qdiff_** は前期比変化幅型，ydiff_**は前年比変化幅型を表し，**の部分は分母を示す（水準型と同じ記号）．qgrowth は前期比成長型，ygrowth は前年比成長型を表す．

ものに限定している．

表 3 と表 4 の正解率 1 位はどちらも受取配当金の前年比変化幅型で分母は自己資本の財務シグナルで，正解率は近い値になっている．分割領域の予測値をみるとこの財務シグナルは値が-0.01 以下かどうかで不正ありかなしかを判断しており，財務シグナル値の大小が不正の有無と結びついているとみられる．値の大小と不正の有無の相関が強い財務シグナルであれば領域を二分割しさえすれば良いため，最大分岐数 2 以上の決定木モデルは不要と言えよう．特別損失や貸倒引当金を用いた財務シグナルも同類である．一方，表 3 の正解率 3 位の有形固定資産処分損の水準型で分母を総資産とした財務シグナルをみると，切り株モデルでの不正ありの正解率は学習データ，テストデータともに低迷しており，正解率の順位は 60 位にとどまっている．最大分岐数 2 の決定木モデルによるこの財務シグナル

表 5: logstic 回帰モデルによる財務シグナル正解率上位 10

NO	signal 型	分子の財務項目	正解率（学習）		正解率（test）	
			不正無	不正有	不正無	不正有
1	qgrowth	自己資本	99.6%	0.7%	99.6%	0.5%
2	ydiff_sl	引当金合計	100.0%	0.7%	100.0%	0.5%
3	ygrowth	為替換算調整勘定	99.9%	0.7%	100.0%	0.5%
4	ygrowth	累積その他の包括利益	100.0%	0.5%	100.0%	0.4%
5	qgrowth	引当金合計	100.0%	0.6%	99.9%	0.4%
6	qgrowth	半製品・仕掛品	100.0%	0.3%	100.0%	0.4%
7	qgrowth	累積その他の包括利益	100.0%	0.6%	100.0%	0.4%
8	ta	半製品・仕掛品	99.9%	0.4%	99.9%	0.4%
9	sl	累積その他の包括利益	100.0%	0.4%	100.0%	0.3%
10	ydiff_ta	貸倒引当金	100.0%	0.4%	100.0%	0.3%

（注）signal 型の'ta' は水準型で分母が総資産，'eq' は水準型で分母が自己資本，'sl' は水準型で分母が売上高の財務シグナルを表す．qdiff_** は前期比変化幅型，ydiff_**は前年比変化幅型を表し，**の部分は分母を示す（水準型と同じ記号）．qgrowth は前期比成長型，ygrowth は前年比成長型を表す．

の分割領域をみると財務シグナル値が-0.01 以下または 0.03 より大きい領域が不正ありを予測する領域となっている一方，切り株モデルは全ての領域で不正なしと予測する結果となっており，分岐数 2 の決定木モデルを活用することでこの財務シグナルを有効活用できるようになったと言える．

ロジスティック回帰モデルを用いた場合は，表 5 から分かるように不正なしの正解率が 100 パーセント近い値になっている一方で不正ありの正解率は 1 パーセントに満たない．ロジスティック回帰で不正ありと不正なしに領域を分割する場合，不正ありと予測する領域では，学習データ上，不正ありの教師データが半分以上になっていると考えられるが，そういった領域が非常に少ないことからほとんどの領域で不正なしと予測するモデルになっている可能性がある．この点を踏まえると，正常データであっても財務シグナルの値が極端な値になっているモノが相当数存在し，単一のシンプルな財務シグナルだけで不正の有無を捉えるのは難しいデータ分布状況にあると考えられる．この点の確認は今後の課題としたい．

5　不正会計検知モデルの構築

前セクションでみた通り，個別の財務シグナルでも不正会計検知にある程度有効とみられるものが散見されるが，その精度は不十分な水準であった．そこで複数の財務シグナルのアウトプットを説明変数とするロジスティック回帰モデルを構築し，これを最終的な不正会計検知モデルとして予測精度を確認した．後述するが，最終モデルの構築段階ではインプットデータは財務シグナルごとに予測値（不正/正常）と欠損値をダミー変数化したものになっている．この段階ではさらなる領域分割を行う決定木モデルよりも，係数等により財務シグナル間の重要度が把握しやすいロジスティック回帰モデルが望ましいと考え，最終モデルではロジスティック回帰モデルを採用した．

不正会計検知モデル構築から正解率の計測までの流れは次のようになる．

1. 前セクションの結果に基づき，正解率上位 n 個の財務シグナルを抽出する．
2. 正解率を計測するテストデータとして全ての正常データと不正データを統合したデータセットを準備し，前セクションと同様に決算年によってデータセットを 10 分割する．
3. テストデータに用いる決算年以外のデータを用いて財務シグナルごとに決定木モデルを学習する．なお，決定木モデルの学習はアンダーサンプリングした正常データと不正企業データを組み合わせたデータセットを用いる．
4. 学習した財務シグナルごとの決定木モデルから統合データセットに含まれる全てのサンプルに対する不正あり/なしの予測値を出力する．
5. 財務シグナルは銘柄や決算期によっては計算できない場合があることから，財務シグナルごとに不正あり，不正なし，欠損値の 3 つのダミー変数を準備する．
6. テストに用いる決算年のデータを除いた残り全ての統合データセットを用いて，ロジスティック回帰モデルのパラメータを推計する．
7. テストデータに対する不正あり/なしの予測値を出力し，正解率を計測する．

表 6 は正解率上位 20 の財務シグナルの予測結果を用いた不正会計検知モデルの予測精度をまとめたもので，表 7〜表 9 は同様に正解率上位 40，上位 60，上位 80 までを用いたモデルでの予測精度をまとめたものである．これらの表から不正会計検知モデルに用いる財務シグナルが 60 を超えると，正解率の改善は頭打ちになることがみてとれる．しかしながら，60 個の財務シグナルを用いた不正会計検知モデルでも不正なしデータの平均正解率は 20 パーセント程度にとどまっている．先行研究の手法は不正データに対する正解率が低くても 60 パーセントを超える水準にあり，本研究の正解率はまだまだ比較できる水準には達していない．一方，先行研究では正常データに対する正解率が軒並み 70 パーセント程度であることと比較すれば，本研究の不正会計検知モデルが正常データの正解率 99 %を維持していることは評価できる点でもある．正常データに対する正

表 6: 正解率上位 20 の財務シグナルを用いた不正会計検知モデルの正解率

test	学習 data 数		正解率（学習）		testdata 数		正解率（test）	
期間	不正無	不正有	不正無	不正有	不正無	不正有	不正無	不正有
2009	7917	1233	99.7%	13.6%	12261	171	99.4%	9.4%
2010	7901	1249	99.2%	15.1%	12349	155	99.1%	20.0%
2011	7885	1261	99.7%	11.7%	12490	143	99.5%	16.1%
2012	7891	1261	99.8%	11.3%	12632	143	99.7%	11.9%
2013	7876	1265	99.6%	13.7%	12888	139	99.2%	13.7%
2014	7882	1237	99.7%	13.1%	13276	167	99.4%	9.0%
2015	7862	1247	99.4%	15.2%	13660	157	99.4%	10.2%
2016	7824	1284	99.5%	14.1%	14086	120	99.3%	10.0%
17-18	7260	1305	99.1%	16.0%	22453	99	99.4%	14.1%
平均	7,810	1,260	99.5%	13.7%	14,011	144	99.4%	12.7%

表 7: 正解率上位 40 の財務シグナルを用いた不正会計検知モデルの正解率

test	学習 data 数		正解率（学習）		testdata 数		正解率（test）	
期間	不正無	不正有	不正無	不正有	不正無	不正有	不正無	不正有
2009	7917	1233	99.60%	18.17%	12261	171	98.92%	12.28%
2010	7901	1249	99.22%	19.14%	12349	155	98.91%	24.52%
2011	7885	1261	99.68%	15.62%	12490	143	99.06%	22.38%
2012	7891	1261	99.70%	14.67%	12632	143	99.52%	17.48%
2013	7876	1265	99.52%	17.63%	12888	139	98.84%	20.14%
2014	7882	1237	99.52%	17.87%	13276	167	99.13%	8.38%
2015	7862	1247	99.41%	19.09%	13660	157	99.20%	12.74%
2016	7824	1284	99.48%	18.69%	14086	120	98.89%	10.00%
17-18	7260	1305	99.15%	21.00%	22453	99	98.93%	11.11%
平均	7,811	1,260	99.47%	17.98%	14,011	144	99.04%	15.45%

表 8: 正解率上位 60 の財務シグナルを用いた不正会計検知モデルの正解率

test	学習 data 数		正解率（学習）		testdata 数		正解率（test）	
期間	不正無	不正有	不正無	不正有	不正無	不正有	不正無	不正有
2009	7917	1233	99.57%	21.01%	12261	171	98.74%	12.87%
2010	7901	1249	99.19%	21.86%	12349	155	98.73%	25.16%
2011	7885	1261	99.62%	19.19%	12490	143	98.74%	24.48%
2012	7891	1261	99.66%	18.24%	12632	143	99.19%	17.48%
2013	7876	1265	99.43%	21.66%	12888	139	98.43%	24.46%
2014	7882	1237	99.51%	20.45%	13276	167	98.91%	12.57%
2015	7862	1247	99.33%	21.49%	13660	157	99.09%	9.55%
2016	7824	1284	99.42%	20.87%	14086	120	98.92%	10.83%
17-18	7260	1305	99.01%	23.07%	22453	99	98.96%	19.19%
平均	7,811	1,260	99.41%	20.87%	14,011	144	98.86%	17.40%

表 9: 正解率上位 80 の財務シグナルを用いた不正会計検知モデルの正解率

test	学習 data 数		正解率（学習）		testdata 数		正解率（test）	
期間	不正無	不正有	不正無	不正有	不正無	不正有	不正無	不正有
2009	7917	1233	99.61%	21.41%	12261	171	98.82%	14.04%
2010	7901	1249	99.10%	22.82%	12349	155	98.77%	24.52%
2011	7885	1261	99.61%	20.54%	12490	143	98.57%	21.68%
2012	7891	1261	99.56%	19.51%	12632	143	98.97%	18.18%
2013	7876	1265	99.39%	22.37%	12888	139	98.30%	25.18%
2014	7882	1237	99.42%	20.70%	13276	167	98.94%	13.17%
2015	7862	1247	99.29%	23.10%	13660	157	99.03%	10.83%
2016	7824	1284	99.51%	21.73%	14086	120	98.85%	10.83%
17-18	7260	1305	98.95%	24.67%	22453	99	98.64%	21.21%
平均	7,811	1,260	99.38%	21.87%	14,011	144	98.76%	17.74%

解率の高さを維持したまま，不正データの検知精度を高めることが当面の大きな課題である．

6 まとめ

本研究では不正会計検知に有効な財務シグナルを検証するため，データマイニング法に則って分析可能な財務シグナルを網羅的に分析した．そして，そこから得られた不正データに対する予測精度の高いシグナル群を活用して最終的な不正会計検知モデルを構築した．最終モデルの不正データに対する正解率は高いモデルでも 20 パーセント弱とまだまだ低い水準にとどまっており先行研究の予測精度に対して大きく見劣りするが，一方で正常データに対する予測精度は 99 パーセント程度と既存研究にはみられない高い水準を保っている．現実的には不正データの数に比べて不正のなされていない正常データの比率は 1 : 100 程度であり，圧倒的に正常データの方が多い．そのため，正常データに対する予測精度を高い水準に保っているのは当研究のモデルの大きな利点と言える．正常データに対する高い予測精度を維持しつつ，不正データに対する予測精度を改善することが今後の大きな課題である．その他，データマイニング法で網羅的に検証した各財務シグナルの中身を詳細に分析することも手付かずとなっており，今後の課題と認識している．既存研究の多くは本決算のみを対象としているのに比べ，本研究では四半期データを分析対象に加えたことで分析対象データ数を増やすことが出来たが，その利点を活かせていない．利点を活用するために，特定時期の財務データ単独の分析だけでなく前後のつながりを活用した分析を行うことも今後の課題と考えている．

参考文献

[1] X. Yan and L. Zheng: *Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach*, The Review of Financial Studies 30(4), pp. 1382-1423 (2017)

[2] Messod D. Beneish: *The Detection of Earnings Manipulation*, Financial Analysts Journal, pp. 24-36 (1999)

[3] Patricia M. Dechow, Weili Ge, Chad R. Larson and Richard G. Sloan: *Predicting Material Accounting Misstatements*, Contemporary Accounting Research 28(1) pp. 17-82, (2011)

[4] 東海林和雄, 中村亮介, 尾崎幸謙: 日本の上場企業における売上過大計上による不正会計の検知–マハラノビス距離を用いた機械学習による方法–, 行動計量学 47(2), pp. 123-140 (2020)

[5] 山田徹, 後藤晋吾: 日本株ファクターモデルに足りないもの–データマイニング法を用いた探訪–, 現代ファイナンス 42, pp. 37-69 (2020)