

機械学習アルゴリズムによる 日本企業の四半期利益予測

早稲田大学大学院商学研究科 商学専攻
35201064 王思涵

2021 年 12 月 20 日

目次

第 I 部	はじめに	2
第 II 部	過去の研究	3
第 III 部	データ	6
1	サンプルデータ	6
2	目的変数	7
3	説明変数	8
第 IV 部	予測モデル	8
4	予測のデザイン	9
5	伝統的時系列モデル	10
5.1	ランダムウォーク	10
5.2	ARIMA モデル	10
5.3	多変量線形モデル	11
6	機械学習モデル	12
6.1	罰則回帰モデル	13
6.2	ランダムフォレスト回帰	14
6.3	ニューラルネットワークモデル	14

7	ハイパーパラメータの選択	17
8	予測精度指標	17
9	Diebold-Mariano(DM) 検定	18
第Ⅴ部 予測結果		19
10	時系列モデル間の予測精度比較	19
11	時系列モデル間の予測精度差の検定	22
12	アナリスト予測との予測精度比較	24
13	アナリスト予測との予測精度差の検定	27
第Ⅵ部 予測の投資指標としての有用性		27
第Ⅶ部 終わりに		29

第Ⅰ部

はじめに

1 株当たり利益 (Earnings Per Share: EPS) は、企業の当期純利益を発行株式数で割ったものであり、規模に依存しない企業の収益性を捉えた指標である。企業外部のステークホルダーである投資家は、企業の将来の EPS の予測をもとに、収益性の高いポートフォリオを構築したり、将来の株価収益率を算出して投資判断を行う。また、企業内部の経営者は将来の EPS の予測を用いて、営業予算の作成や設備投資の判断などの重要な意思決定を行う。したがって、EPS を正しく予測することは企業内外の幅広いステークホルダーにとって重要である。

企業の将来利益の予測は、人的な予測と、統計的・機械的な予測の 2 つに大別できる (桜井, 1990)。人的な予測の代表として、アナリスト予測がある。アナリスト予測とは、株式市場分析の専門家である証券アナリストが公表する利益予測である。他方、統計的・機械的な予測とは、過去の実績データをもとに何らかの時系列モデルを用いて将来の利益を予測するものである。モデルに基づく予測は、予測値を導出するまでの過程を全て自動化できるため、人的な予測に比べてコストが低いという特徴を有している (桜井, 1990)。

統計的・機械的な予測に関する研究で用いられているモデルは、1)「単変量」か「多変量」か、2)「線形」か「非線形」か、の 2 つの観点で分類できる (Zhang et al., 2004)。従来の、モデルに基づく EPS 予測の研究の多くは、自己回帰和分移動平均 (AutoRegressive Integrated Moving Average: ARIMA) モデル (Box, Jenkins, Reinsel and Ljung, 2015) などの単変量で線形な統計的時系列モデルを用いている (Foster, 1977; Griffin, 1977; Brown and Rozeff, 1979)。一方、近年の研究では、「ファンダメンタル会計変数 (Fundamental accounting variables)」を用いた多変量モデルによる EPS 予測の研究が増えている。ファンダメンタル会計

変数とは、売掛金、棚卸資産、資本的支出といったいくつかの会計変数のことであり、これらの変数には将来の EPS を予測する情報があるとされている (Lev and Thiagarajan, 1993; Abarbanell and Bushee, 1997) . また、四半期 EPS データには非線形性があるとされており (Hopwood and McKeown, 1986) , これを考慮するために、予測モデルに非線形性を反映させている研究も盛んである . 中でも従来の統計的時系列モデルに代替されるモデルとして、機械学習アルゴリズムが注目されている . 機械学習アルゴリズムには、従来の統計的時系列モデルに比べて、高次元データの処理に長けたモデルや非線形性を捉えられるモデルが豊富にある (Cao and You, 2020) . そのため機械学習アルゴリズムを用いることで、多変量なファンダメンタル会計変数から予測情報を抽出し、企業利益データの非線形性を捉え、精度の高い EPS 予測が得られると考えられる .

機械学習アルゴリズムによる EPS 予測の研究は多くの国の企業を対象として行われている . Zhang et al. (2004) はアメリカのニューヨーク証券取引所の上場企業、Cao and Gan (2009) は中国の上海証券取引所と深セン証券取引所の上場企業、Etemadi et al. (2015) はイランのテヘラン証券取引所の上場企業を対象に、機械学習アルゴリズムの 1 つである NN による四半期 EPS の予測を行っている . また、Cao and You (2020) はアメリカの上場企業を対象に、いくつかの代表的な機械学習アルゴリズムを用いた年次 EPS 予測を行っている . いずれの研究においても機械学習アルゴリズムが精度の高い予測を与える結果となっている .

このように近年、統計的・機械的な手法による EPS 予測の研究分野では、機械学習アルゴリズムを用いることで予測の精度が向上することが様々な国のサンプルで確認されている . しかし、太田 (2006) によると日本における人的な予測と従来の時系列モデル予測の精度比較研究については、明確な結論が出ていないにも拘らず、モデルによる予測よりも人的な予測の方が適切であると暗黙裡に見なされており、時系列モデルによる予測の研究は、現在では衰退している . 現に日本における統計的・機械的な EPS 予測の研究は限定的で従来の線形時系列モデルの適用に留まっており、高い予測精度が期待される機械学習アルゴリズムを用いた日本企業を対象とする EPS 予測の研究は、筆者の知る限り存在しない . また、日本では時系列モデルによる利益予測が衰退したあとの 2008 年度に四半期報告制度が創設されているため、四半期データを用いた EPS 予測の研究もない .

そこで、日本企業データにおいて、機械学習アルゴリズムは、従来の統計的時系列モデルに比べ EPS 予測の精度を向上させるのか、人的な予測よりも高い精度の予測を与えるのかを検証することを本稿の目的とする . 具体的には東京証券取引所一部上場企業を対象に機械学習アルゴリズムによる四半期 EPS 予測を行い、得られた予測値を用いて従来の線形時系列モデルによる予測やアナリストによる予測と精度比較をする . 追加的な検証として、各手法による予測値それぞれに基づいた株価予想利益率からロング・ショートポートフォリオを構築し、それを日本の証券市場に適用させて、機械学習アルゴリズムによる予測の投資指標としての有用性を考察する .

本稿の構成は次のとおりである . まず第 II 部では過去の EPS 予測に関する研究や機械学習アルゴリズムを用いた時系列予測に関する研究について紹介する . 次に第 III, IV 部で本稿で用いるデータ及び予測モデルについて説明し、第 V 部で本稿の予測結果について報告する . 第 VI 部では追加的な検証を行い、最後に 第 VII 部で本稿の結論を述べる .

第 II 部

過去の研究

これまで多くの企業利益予測の研究で、統計的・機械的な予測手法間の精度比較、統計的・機械的な予測手法と人的な予測手法の精度比較、人的な予測手法間の精度比較が行われている。桜井 (1990) や太田 (2006) のレビュー論文は、そのような数々の利益予測の研究に関する文献についてまとめている。アメリカの利益予測の研究に関して、桜井 (1990) は以下のようにまとめている。――まず EPS を予測する伝統的な時系列モデルについて、年次 EPS はランダムウォークによってうまく描写され、四半期 EPS は Brown and Rozeff (1979), Griffin (1977), Foster (1977) の 3 つの ARIMA モデルによってうまく描写されると述べている。一方で、アナリスト予測と伝統的な時系列モデルによる予測の比較については、アナリストによる年次および四半期の EPS 予測が伝統的な時系列モデルによる予測よりも正確であると述べている。その理由として、アナリストは広く最新の情報集合を予測に用いるからである。――>――太田 (2006) は、日本の利益予測に関する研究はそれほど盛んではないが、比較結果はアメリカのものと同様であることを指摘している。――>――そして、日本の時系列モデルによる利益予測については、「アナリスト予測との精度比較の明確な結論が出ていないにも拘わらず、最近の研究では時系列モデルによる予測よりもアナリスト予測を用いる方が市場の期待利益として適切であると暗黙裡に見なされており、時系列モデルによる利益予測の研究は、現在では衰退している」と述べている。――>

ここで注意すべき点として、上記で言及されている利益予測精度の比較の研究で扱われている時系列モデルの多くは、あくまでも当時研究が盛んであった伝統的な単変量で線形の時系列モデルのことであり、代替的な多変量モデルや非線形モデルについては議論されていない。太田 (2006) は「ナイーブな時系列を用いた場合には人的な予測の方が精度が高いが、高度な時系列モデルを用いた場合には時系列モデルの予測の方が人的な予測よりも精度が高いといえる」と述べており、モデルを発展させることで統計的・機械的な EPS 予測の精度は向上する余地があることを示唆している。

近年、多くの学術分野で応用され成果を上げている予測モデルとして、機械学習アルゴリズムの 1 つであるニューラルネットワーク (Neural Network: NN) モデルが挙げられる。NN は生物の神経細胞 (ニューロン) の構造と機能をもとに考案された数学モデルであり、従来の統計モデルに比べて非線形で不定形な問題にうまく対処できることから、多数の分野での利用が増加している (Tkáč and Verner, 2016)。特に、Hill et al. (1996) は、月次や四半期の時系列データにおいて NN の方が伝統的な統計モデルよりも正確な予測を与えるとし、時系列予測における NN の有用性を示している。その理由として、NN はあらゆる関数形を近似できる普遍性定理 (Universal Approximation Theorem) (Hornik et al., 1989) により、伝統的な線形時系列モデルでは捉えられない時系列データの非線形性を捉え、予測モデルの関数形の誤特定を回避できるからであると述べている。また、Hill et al. (1994) は、予測の対象である時系列データが、i) 財務的、ii) 季節的、iii) 非線形な特徴を有するとき、NN による予測は従来の線形時系列モデルよりも精度が高い傾向があると述べている。特に四半期 EPS は、上記の 3 つの特徴を有することが確認されている (Hopwood and McKeown, 1986)。

そこで、Callen et al. (1996) はニューヨーク株式市場の企業を対象に、単変量の四半期 EPS 予測における NN と従来の統計的手法である ARIMA モデルとの精度比較を行ったが、結果は予想に反して NN の予測精度が ARIMA モデルに比べて低いものとなっている。そして、この研究では NN の予測精度は文脈依存であると結論付けている。

一方、会計学の研究では、企業利益の予測について、将来の利益を説明する変数を特定することに注目して

きた。Lev and Thiagarajan (1993) は、アナリストが有価証券の価値評価において有用であるとされる会計変数を調査し、それらの会計変数が将来の企業利益と関連があると述べている。言及された会計変数は以下のとおりである。

- 棚卸資産 (Inventories)
売上原価の増加に対して過度な棚卸資産の増加は、売上高を増加させることが困難であることを示唆するため、ネガティブなシグナルとみなされる。さらに、棚卸資産が増加すると経営者は在庫を減らそうとするため、将来の利益が減少するシグナルになり得る。
- 売掛金 (Accounts receivable)
売上高に対して過度な売掛金の増加は、企業の製品販売が困難な状態にあることや、貸倒引当金の増加などを意味し、将来の利益が減少するシグナルになり得る。
- 資本的支出 (Capital expenditures)
資本的支出の過度な減少は、以前の投資水準を維持するための現在および将来のキャッシュフローが十分でないという経営者の懸念を意味する。このように一般的に資本的支出の減少は、短期的な経営志向とみなされ、将来の利益が減少するシグナルになり得る。
- 売上総利益 (Gross margin)
売上総利益は企業の競争の激しさや営業レバレッジなどの要因を捉え、これが企業の長期的なパフォーマンスに影響を与える。したがって、売上総利益は企業の利益の持続性や企業価値に関して有益な情報をもち、売上高に対して過度な売上総利益の減少は、将来の利益が減少するシグナルになり得る。
- 販売費および一般管理費 (Selling and administrative expenses)
ほとんどの場合、販売費および一般管理費は一定であるため、売上高に対して過度な販売費および一般管理費の増加は、コストコントロールの喪失や異常な販売努力を示唆し、将来の利益が減少するシグナルになり得る。
- 実効税率 (Effective tax rate)
法定税率の変更に依らない企業の実効税率の大幅な変化は、一般的には一時的なものとして捉えられる。したがって、過度な実効税率の低下は将来の利益が減少するシグナルになり得る。
- 労働力 (Labor force)
一般的に労働力の削減の発表に対してアナリストは好意的な反応を示す。したがって、過度な労働力の増加は将来の利益が減少するシグナルになり得る。

Abarbanell and Bushee (1997) では、上記の会計変数をファンダメンタル会計変数とし、ファンダメンタル会計変数が将来の EPS の変化を説明するかどうかについて検証している。分析の結果、ファンダメンタル会計変数と将来の EPS の変化の間に強い関係が確認され、ファンダメンタル会計変数に含まれる会計情報は、将来の EPS に対して予測能力を持つと述べている。

これを受け Zhang et al. (2004) は、Callen et al. (1996) の単変量 NN による四半期 EPS 予測の研究を発展させる形で、NN にファンダメンタル会計変数を反映させた多変量 NN による四半期 EPS 予測を行った。この研究では、単変量線形モデル、多変量線形モデル、単変量 NN、多変量 NN の 4 つのカテゴリー間で精度比較を行っており、結果として、NN は線形モデルよりうまくファンダメンタル会計変数の情報を反映し、より精度の高い EPS の予測値を与えた。その後、様々な国のサンプルで EPS 予測における NN の有用性を示した研究が行われている。Cao and Gan (2009) は中国の上海証券取引所と深セン証券取引所の上場企業、Etemadi et al. (2015) はイランのテヘラン証券取引所の上場企業を対象に、ファンダメンタル会計変数を反

映させた NN による EPS 予測を行っており、いずれも高い精度の予測を与える結果を得ている。

このように近年の四半期 EPS 予測の研究では多変量 NN が盛んに用いられているが、機械学習アルゴリズムの 1 つである NN だけが注目されており、他の機械学習アルゴリズムを用いた研究は限られている。これは EPS 予測の文脈に限らず、他の時系列予測研究でも NN のみに着目している傾向がある (Ahmed et al., 2010)。しかし、NN 以外の機械学習アルゴリズムが四半期 EPS 予測において成果を上げられる可能性も十分にある。代表的な機械学習アルゴリズムとして、罰則回帰モデルやランダムフォレスト回帰が挙げられる (Hastie et al., 2009)。罰則回帰モデルは、高次元データに対して変数選択・係数縮小を行うことで多数の説明変数から目的変数を予測する情報を抽出し、精度の高い予測を行うことができるため、四半期 EPS の多変量予測において有用であると考えられる。また、ノンパラメトリックで非線形なランダムフォレスト回帰 (Breiman, 2001) も、NN とは異なったアルゴリズムで四半期 EPS データの非線形性を捉えることができると考えられる。―― Cao and You (2020) は年次であるものの、これらの代表的な機械学習アルゴリズムを用いた EPS 予測を行っており、ベンチマークであるランダムウォークよりも精度の高い予測を与える結果を報告している。――

以上の過去の研究を踏まえ、現状の日本の EPS 予測における「時系列モデルが古典的な手法にとどまっていること」と「時系列モデルと人的な予測の優劣の曖昧さ」の 2 つの課題に対し、本稿は機械学習アルゴリズム (NN, 罰則回帰モデル, ランダムフォレスト回帰) にファンダメンタル会計変数を含めて日本企業の四半期 EPS 予測を行い、さらに時系列モデルとアナリストの EPS 予測精度を比較し、予測手法間の精度の優劣を明らかにする。

第 III 部 データ

1 サンプルデータ

日本の四半期報告制度は比較的新しく、2007 年の企業会計基準委員会 (ASBJ) による企業会計基準第 12 号「四半期財務諸表に関する会計基準」及び企業会計基準適用指針第 14 号「四半期財務諸表に関する会計基準の適用指針」の公表以降、上場企業の四半期報告書の提出が義務付けられた。^{*1}

これにより、四半期データを分析の単位とする本稿では、サンプル期間を四半期報告制度が適用された 2008 年度 (2008 年 4 月 1 日) から現在 2020 年度 (2021 年 3 月 31 日) までの計 52 四半期とする。また、本稿の分析対象企業 (サンプル企業) を、サンプル期間において以下の基準で選択する。

- 東京証券取引所一部上場企業
- 3 月決算企業
- 一般事業者 (「銀行業」, 「証券」, 商品先物取引業, 「保険業」, 「その他金融業」を除く)

^{*1} 2011 年に企業会計基準委員会 (ASBJ) は改正企業会計基準第 12 号「四半期財務諸表に関する会計基準」及び改正企業会計基準適用指針第 14 号「四半期財務諸表に関する会計基準の適用指針」等の公表をしている。この四半期報告制度の改正では財務諸表作成者の作成負担を考慮し、いくつかの四半期情報の開示を義務ではなく任意とされている。それに伴い、四半期報告制度の改正以降、資本的支出と期末従業員数の観測される頻度が減り、欠損が多く見られる。Zhang et al. (2004) は四半期単位で欠損している資本的支出と期末従業員数について、資本的支出は毎四半期均等であるとし、期末従業員数は前期の値を維持すると仮定して、欠損値を補填している。本稿でもこの方法に倣って資本的支出と期末従業員数の欠損処理を行うこととする。

表 1: 業種 (証券コード協議会中分類 33 種)

業種コード	業種名	企業数
3650	電気機器	108
6050	卸売業	101
3200	化学	92
3600	機械	83
2050	建設業	70
5250	情報・通信業	66
9050	サービス業	59
6100	小売業	54
3050	食料品	44
3700	輸送用機器	41
5050	陸運業	34
3450	鉄鋼	28
3800	その他製品	27
3100	繊維製品	24
3400	ガラス・土石製品	23
3250	医薬品	23
3550	金属製品	22
3750	精密機器	20
8050	不動産業	19
3500	非鉄金属	14
5200	倉庫・運輸関連業	12
3150	パルプ・紙	8
5100	海運業	7
4050	電気・ガス業	7
3300	石油・石炭製品	5
3350	ゴム製品	4
0050	水産・農林業	3
1050	鉱業	3
5150	空運業	2
合計		1,003

- 各四半期の中間月末 (5 月末, 8 月末, 11 月末, 2 月末) までに前四半期決算を発表している企業^{*2}

四半期 EPS(3 ヶ月) およびファンダメンタル会計変数の算出に必要な四半期会計データは「日経 NEEDS-FinancialQUEST」から収集している。結果, 1,003 社, 延べ 1,003 社 × 52 四半期 = 52,156 個の企業四半期が最終サンプルである。表 1 はサンプル企業の業種別の内訳を示している。なお, 所属業種は証券コード協議会の中分類 (全 33 種) に基づいている。

2 目的変数

予測の対象となる四半期 1 株あたり利益 (四半期 EPS) は以下のように定義される。

^{*2} IBES 予測及びポートフォリオ構築のタイミングをそろえるため。

表 2: 変数の記述統計量

	Y	INV	AR	CAPX	GM	SA	ETR	LF
count	52156	52156	52156	52156	52156	52156	52156	52156
mean	67.468	808.786	1272.915	173.152	513.363	385.510	0.339	16.205
std	707.529	7234.778	8087.579	2387.784	3577.896	2699.862	5.835	0.793
min	-37300.500	0.011	0.026	0.001	-30368.482	-482.938	-494.716	11.628
25%	3.490	91.615	181.897	9.296	53.156	39.603	0.241	15.662
50%	15.230	193.908	357.978	24.741	115.768	83.523	0.331	16.154
75%	37.540	415.649	717.761	64.527	222.790	166.317	0.414	16.693
max	31202.830	404482.462	293004.896	143224.390	104813.930	87333.561	817.750	22.531

$$\begin{aligned}
Y &= \text{四半期 1 株あたり利益} \\
&= \frac{\text{普通株式に係る四半期当期純利益}}{\text{普通株式の期中加重平均株式数}} \\
&= \frac{\text{損益計算書上の四半期当期純利益} - \text{普通株式に帰属しない金額}}{\text{普通株式の期中加重平均発行済株式数} - \text{普通株式の期中加重平均自己株式数}}
\end{aligned} \tag{1}$$

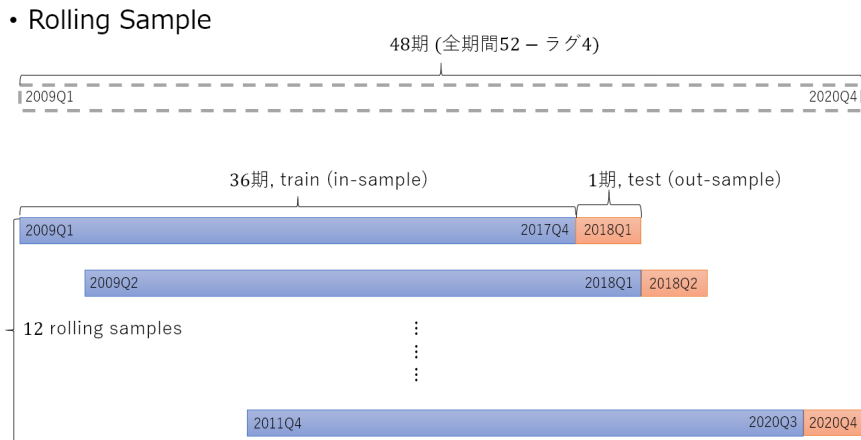
3 説明変数

単変量モデルによる四半期 EPS 予測は，過去の四半期 EPS の実績値のみを説明変数として用いる．一方，多変量モデルによって四半期 EPS を予測する場合は，過去の四半期 EPS に加えて，Lev and Thiagarajan (1993) や Abarbanell and Bushee (1997) が企業の将来の EPS に対して予測能力を持つと言及しているファンダメンタル会計変数も説明変数として用いる．本稿では，ファンダメンタル会計変数を実際に四半期 EPS 予測に用いた Zhang et al. (2004) に倣って以下のように 7 つのファンダメンタル会計変数を定義する．

$$\begin{aligned}
INV &= \frac{\text{棚卸資産 (円)}}{\text{普通株式の期中加重平均株式数}} \\
AR &= \frac{\text{売掛金 (円)}}{\text{普通株式の期中加重平均株式数}} \\
CAPX &= \frac{\text{資本的支出 (円)}}{\text{普通株式の期中加重平均株式数}} \\
GM &= \frac{\text{売上総利益 (円)}}{\text{普通株式の期中加重平均株式数}} \\
SA &= \frac{\text{販売費及び一般管理費 (円)}}{\text{普通株式の期中加重平均株式数}} \\
ETR &= \frac{\text{法人税 (円)}}{\text{税引前利益 (円)}} \\
LF &= \log \left(\frac{\text{売上高 (円)}}{\text{従業員数 (人)}} \right)
\end{aligned} \tag{2}$$

表 2 は予測に用いる変数の記述統計を示している．

図 1: ローリングサンプルのイメージ



第 IV 部

予測モデル

4 予測のデザイン

本稿では、サンプル企業 1,003 社について、それぞれ 1 社ごとに 1 四半期先予測を 12 期 (3 年) 分行う。まず、ある企業の全 52 四半期サンプル期間^{*3}をもとに、1 つあたり長さ 41 期のローリングサンプル (rolling sample) を 12 個形成する。次に、1 つのローリングサンプルを、過去 40 期の訓練データ (training data, in-sample) と最新 1 期のテストデータ (test data, out-of-sample) に分ける。そして訓練データを用いてモデルを推定し、テストデータと同じ期にあたる 1 期先の予測を行う。これを 12 個のローリングサンプルで繰り返すことによって、ある企業の 1 四半期先予測が 12 期分得られる。

つまり具体的には、ある企業の全 52 四半期サンプル期間 $\{Data_{2008Q1}, Data_{2008Q2}, \dots, Data_{2020Q4}\}$ から 12 個のローリングサンプルを作り、1 つ目の長さ 41 期のローリングサンプル $\{Data_{2008Q1}, Data_{2008Q2}, \dots, Data_{2017Q4}, Data_{2018Q1}\}$ を長さ 40 期の訓練データ $\{Data_{2008Q1}, Data_{2008Q2}, \dots, Data_{2017Q4}\}$ と長さ 1 期のテストデータ $\{Data_{2018Q1}\}$ に分ける。そして訓練データを用いてモデルを推定したあと、テストデータの説明変数をモデルに代入して訓練データの 1 期先の四半期 EPS, Y_{2018Q1} の予測値 \hat{Y}_{2018Q1} を求める。2 つ目のローリングサンプル $\{Data_{2008Q2}, Data_{2008Q3}, \dots, Data_{2018Q1}, Data_{2018Q2}\}$ も同様に 訓練データ $\{Data_{2008Q2}, Data_{2008Q3}, \dots, Data_{2018Q1}\}$ を用いて 1 期先の四半期 EPS の予測値 \hat{Y}_{2018Q2} を求める。同様に他のローリングサンプルに対しても、それぞれ同じモデルを異なる期間の訓練データで推定することで、1 企業につき予測期間は 2018 年度第 1 四半期 ~ 2020 年度第 4 四半期の 12 期となり、1 四半期先予測値系列 $\{\hat{Y}_{2018Q1}, \hat{Y}_{2018Q2}, \dots, \hat{Y}_{2020Q4}\}$ が得られる。この予測プロセスを全サンプル企業 1,003 社に対して行うことで、最終的に $12 \text{ 期} \times 1,003 \text{ 社} = 12,036 \text{ 個}$ の四半期 EPS 予測値が各予測手法から算出される。図 1 は本稿のローリングサンプルによる 1 期先予測の概要を示している。

^{*3} 実際のモデル推定では説明変数にラグ変数を含める。したがって利用できるサンプル期間は 52 - ラグの次数 である。モデルのデザインによってラグの次数は異なるため、利用できる訓練データの長さもモデルによって異なる。

5 伝統的時系列モデル

5.1 ランダムウォーク

Y_t をある企業における t 期の四半期 EPS とすると、もっとも単純な単変量モデルであるランダムウォークは次のとおりである。

$$\begin{aligned} Y_t &= Y_{t-1} + \epsilon_t \\ \hat{Y}_t^{RW} &= Y_{t-1} \end{aligned} \quad (3)$$

ただし、 ϵ_t は全ての時点 t において

$$\begin{aligned} E[\epsilon_t] &= 0 \\ E[\epsilon_t \epsilon_{t-k}] &= \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \end{aligned} \quad (4)$$

を満たすホワイトノイズである。

一方、四半期 EPS 時系列には 1 年単位での季節的変動パターンがあると考えられる。そこでランダムウォーク過程を 1 四半期単位ではなく、1 会計年度単位 (4 四半期) であると考慮した季節ランダムウォークが次のとおりである。

$$\begin{aligned} Y_t &= Y_{t-4} + \epsilon_t \\ \hat{Y}_t^{SRW} &= Y_{t-4} \end{aligned} \quad (5)$$

ランダムウォークは前四半期の実績値を、季節ランダムウォークは前年同四半期の実績値を予測値とするモデルであり、どちらも実際の予測においてパラメータを推定する工程がない単純なモデルである。そこで、本稿ではランダムウォーク及び季節ランダムウォークを他の予測に対するベンチマークとして用いる。

5.2 ARIMA モデル

従来の四半期 EPS を予測する単変量線形時系列モデルとして、自己回帰和分移動平均 (AutoRegressive Integrated Moving Average: ARIMA) モデル (Box et al., 2015)^{*4}、特にデータの季節性にも対応できるように一般化した季節自己回帰和分移動平均 (Seasonal AutoRegressive Integrated Moving Average: SARIMA) モデルを用いる。 $B^n Y_t = y_{t-n}$ と定義されるような B (backshift operator) を導入すると、SARIMA モデルの一般形は以下のとおりである。^{*5}

$$\phi_p(B)\Phi_P(B)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \quad (6)$$

^{*4} ARIMA モデルは提唱者の名前から Box-Jenkins モデルとも呼ばれる。

^{*5} 定数項 $\theta_\mu = (1 - \phi_1 - \dots - \phi_p)\mu$ は階差をとれば 0 となるため、通常取り扱いに従って 0 とする。

ただし ,

$$\begin{aligned}
p &: \text{トレンドの自己回帰過程の階数} \\
d &: \text{トレンド階差の次数} \\
q &: \text{トレンドの移動平均過程の階数} \\
P &: \text{季節変動の自己回帰過程の階数} \\
D &: \text{季節階差の次数} \\
Q &: \text{季節変動の移動平均過程の階数} \\
s &: \text{季節変動の周期}
\end{aligned} \tag{7}$$

$$\begin{aligned}
\phi_p(B) &= (1 - \phi_1 B - \dots - \phi_p B^p) \\
\theta_q(B) &= (1 - \theta_1 B - \dots - \theta_q B^q) \\
\Phi_P(B^s) &= (1 - \Phi_1 B^s - \dots - \Phi_P B^{sP}) \\
\Theta_Q(B^s) &= (1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ})
\end{aligned}$$

通常 , SARIMA モデルは $(p, d, q) \times (P, D, Q)$ の値の組み合わせを Box-Jenkins 法によってデータごとに選定し , モデルの構築を行う . 一方 , これまでの四半期 EPS の時系列予測の分野では全ての企業に適合する SARIMA モデルが探求され , 次の 3 つの SARIMA モデル (Foster, 1977; Griffin, 1977; Brown and Rozeff, 1979) があらゆる企業の四半期 EPS の時系列特性を描写するモデルであるとされている .

$$\text{Foster (1977)} : (p, d, q) \times (P, D, Q)_s = (1, 0, 0) \times (0, 1, 0)_4$$

$$\begin{aligned}
Y_t &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) + \epsilon_t \\
\hat{Y}_t^{ARIMA-F} &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) + \delta
\end{aligned} \tag{8}$$

$$\text{Griffin (1977)} : (p, d, q) \times (P, D, Q)_s = (0, 1, 1) \times (0, 1, 1)_4$$

$$\begin{aligned}
Y_t &= Y_{t-4} + (Y_{t-1} - Y_{t-5}) - \theta_1 a_{t-1} - \Theta_1 a_{t-4} + \theta_1 \Theta_1 a_{t-5} + \epsilon_t \\
\hat{Y}_t^{ARIMA-G} &= Y_{t-4} + (Y_{t-1} - Y_{t-5}) - \theta_1 a_{t-1} - \Theta_1 a_{t-4} + \theta_1 \Theta_1 a_{t-5} + \delta
\end{aligned} \tag{9}$$

$$\text{Brown and Rozeff (1979)} : (p, d, q) \times (P, D, Q)_s = (1, 0, 0) \times (0, 1, 1)_4$$

$$\begin{aligned}
Y_t &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) - \Theta_1 a_{t-4} + \epsilon_t \\
\hat{Y}_t^{ARIMA-BR} &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) - \Theta_1 a_{t-4} + \delta
\end{aligned} \tag{10}$$

ただし , δ は SARIMA モデルの定数項である .

本稿では先行研究に倣い , 特に支持されてきたこの 3 つのモデルを単変量線形時系列モデルとして用いることとする .

5.3 多変量線形モデル

ファンダメンタル会計変数が将来の四半期 EPS を線形に説明するかどうかについて確かめるため , Lev and Thiagarajan (1993) , Abarbanell and Bushee (1997) , Lorek and Willinger の研究で用いられているモデルをもとに , 以下のような多変量線形回帰モデルを考える .

$$\hat{Y}_t^{OLS1} = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-4} + \beta_3 INV_{t-1} + \beta_4 AR_{t-1} + \beta_5 CAPX_{t-1} + \beta_6 GM_{t-1} + \beta_7 SA_{t-1} + \beta_8 ETR_{t-1} + \beta_9 LF_{t-1} \quad (11)$$

$$\hat{Y}_t^{OLS2} = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-4} + \beta_3 INV_{t-4} + \beta_4 AR_{t-4} + \beta_5 CAPX_{t-4} + \beta_6 GM_{t-4} + \beta_7 SA_{t-4} + \beta_8 ETR_{t-4} + \beta_9 LF_{t-4} \quad (12)$$

2つのモデルはどちらも、四半期 EPS の 1 四半期前ラグ Y_{t-1} と 4 四半期前ラグ Y_{t-4} が自己回帰的な説明変数として含まれている。両者の違いとしてはファンダメンタル会計変数のラグの次数にあり、(11) のモデルではファンダメンタル会計変数の 1 四半期前ラグを、(12) のモデルでは季節性を考慮してファンダメンタル会計変数の 4 四半期前ラグをモデルの説明変数として含めている。一方、Cao and Parry (2009) は後述する機械学習アルゴリズムのモデルと利用できる情報の公平性に保つため、以下のように四半期 EPS とファンダメンタル会計変数のラグ変数を 1 四半期前、2 四半期前、3 四半期前、4 四半期前すべて含めた多変量線形回帰モデルを推定している。

$$\hat{Y}_t^{OLS3} = \beta_0 + \sum_{\tau=1}^4 (\beta_{1\tau} Y_{t-\tau} + \beta_{2\tau} INV_{t-\tau} + \beta_{3\tau} AR_{t-\tau} + \beta_{4\tau} CAPX_{t-\tau} + \beta_{5\tau} GM_{t-\tau} + \beta_{6\tau} SA_{t-\tau} + \beta_{7\tau} ETR_{t-\tau} + \beta_{8\tau} LF_{t-\tau}) \quad (13)$$

以上のファンダメンタル会計変数を説明変数として用いた 3 つの多変量線形回帰モデルを多変量モデルのベンチマークとして、他の手法による多変量予測と比較する。なお、これらのモデルのパラメータは最小二乗法を用いて推定する。

6 機械学習モデル

ここでは、本稿で用いる機械学習の手法について紹介する。なお、以下より共通して時点 t の説明変数ベクトル X_t について、単変量モデルでは目的変数である四半期 EPS のラグのみを用いてモデル構築を行う。

$$\begin{aligned} X_t &= (X_{t1}, \dots, X_{tk}) \\ &= (Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}) \end{aligned} \quad (14)$$

また、多変量モデルでは四半期 EPS のラグに加え、ファンダメンタル会計変数のラグも含めてモデルを構築する。

$$\begin{aligned} X_t &= (X_{t1}, \dots, X_{tk}) \\ &= (Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, \\ &\quad INV_{t-1}, INV_{t-2}, INV_{t-3}, INV_{t-4}, \\ &\quad AR_{t-1}, AR_{t-2}, AR_{t-3}, AR_{t-4}, \\ &\quad CAPX_{t-1}, CAPX_{t-2}, CAPX_{t-3}, CAPX_{t-4}, \\ &\quad GM_{t-1}, GM_{t-2}, GM_{t-3}, GM_{t-4}, \\ &\quad SA_{t-1}, SA_{t-2}, SA_{t-3}, SA_{t-4}, \\ &\quad ETR_{t-1}, ETR_{t-2}, ETR_{t-3}, ETR_{t-4}, \\ &\quad LF_{t-1}, LF_{t-2}, LF_{t-3}, LF_{t-4}) \end{aligned} \quad (15)$$

6.1 罰則回帰モデル

説明変数間に複数の強い相関が存在する場合，線形回帰モデルの最小二乗推定量の分散は大きくなる．そして高分散 (high variance) により係数の推定値が極端な値をとると，訓練データ内での予測精度は高くても，テストデータにおける予測精度が低くなる恐れがある．このように予測モデルが訓練データの規則性のみを捉えてしまい，テストデータを含めたデータ全体に対して汎化できていないことを過学習 (overfitting) という．

線形回帰モデルの係数の推定値が極端な値をとることに起因する過学習を防ぐために，係数の大きさに罰則を与えて推定する手法が罰則回帰モデル (penalized regression model) である^{*6}．罰則回帰モデルは，その罰則の与え方によって推定で最小化する目的関数が異なる．本稿では罰則回帰モデルのうち，代表的な Ridge 回帰，LASSO 回帰，Elastic Net 回帰の 3 つを用いる．

Ridge 回帰の係数は以下のように求められる．

$$\hat{\beta}^{Ridge} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{j=1}^k \beta_j X_{tj} \right)^2 + \lambda \sum_{j=1}^k \beta_j^2 \right\} \quad (16)$$

ここで， $\lambda \geq 0$ は罰則の強さを調節するパラメータであり，モデル推定の枠組みの中で決定されないパラメータ (ハイパーパラメータ) である．ハイパーパラメータの値はデータに応じて指定する必要があるが，本稿におけるハイパーパラメータの選択方法の詳細については 7 節で説明する．(16) 式で示されるように，Ridge 回帰では目的関数に係数の二乗和が含まれており，係数が過大な値をとることを防ぐ構造となっている．したがって Ridge 回帰は予測情報を持たない説明変数に対して係数縮小 (parameter shrinkage) し，モデルの過学習を防ぐ．

LASSO (Least Absolute Shrinkage and Selection Operator) 回帰モデルの係数は以下のように求められる．

$$\hat{\beta}^{LASSO} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{j=1}^k \beta_j X_{tj} \right)^2 + \lambda \sum_{j=1}^k |\beta_j| \right\} \quad (17)$$

LASSO 回帰は，Ridge 回帰と同様，目的関数にパラメータのサイズに関する罰則項が含まれているが，罰則として係数の絶対値の和を用いている点が Ridge 回帰との違いである．この罰則の与え方により，予測情報を持たない説明変数の係数が丁度 0 となるスパース推定が行われる．つまり，LASSO 回帰を推定することで変数選択が行われる．

Elastic Net 回帰の係数は以下のように求められる．

$$\hat{\beta}^{EN} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{j=1}^k \beta_j X_{tj} \right)^2 + \lambda \left\{ (1 - \alpha) \sum_{j=1}^k \beta_j^2 + \alpha \sum_{j=1}^k |\beta_j| \right\} \right\} \quad (18)$$

^{*6} 罰則回帰モデルの詳細については，Hoerl and Kennard (1970); Tibshirani (1996); Zou and Hastie (2005) などを参照されたい．

Elastic Net 回帰は、Ridge 回帰と LASSO 回帰両方の罰則項を利用する構造となっている。α は LASSO 回帰の罰則の影響の割合を調節するハイパーパラメータであり、罰則の強さを調節するハイパーパラメータ λ と併せて指定する必要がある。

なお、それぞれの罰則回帰モデルの予測値は次のとおりである。

$$\begin{aligned}\hat{Y}_t^{Ridge} &= \beta_0^{Ridge} + \sum_{j=1}^k \beta_j^{Ridge} X_{tj} \\ \hat{Y}_t^{LASSO} &= \beta_0^{LASSO} + \sum_{j=1}^k \beta_j^{LASSO} X_{tj} \\ \hat{Y}_t^{EN} &= \beta_0^{EN} + \sum_{j=1}^k \beta_j^{EN} X_{tj}\end{aligned}\tag{19}$$

6.2 ランダムフォレスト回帰

ランダムフォレスト (Breiman, 2001) は、決定木を弱学習器とするアンサンブル学習である。弱学習器とはそれ単独で用いると精度が低いような予測モデルであり、アンサンブル学習はその弱学習器を複数推定し、組み合わせることで予測精度の向上を図る手法である。弱学習器となる決定木は、ある変数のある閾値でデータを 2 分割し、この分割を繰り返すことで作成される。単独の決定木による予測値は、予測サンプルの説明変数をもとに作成された木の分岐を辿り、到達した末端ノードに属するサンプルの目的変数の平均値である。ランダムフォレスト回帰では、元のデータセットから標本及び変数無作為に復元抽出した複数のサブデータセットを構築し (ブートストラップサンプリング)、それをもとに弱学習器である決定木を複数推定する。そして各決定木の予測値の平均値が最終的なランダムフォレスト回帰の予測値となる。

決定木の数を M とし、 $m = 1, \dots, M$ 番目の決定木の t 時点における予測値を $Tree_m(\mathbf{X}_t)$ とすると、ランダムフォレスト回帰の予測値は以下のように表される。

$$\hat{Y}_t^{RF} = \frac{1}{M} \sum_{m=1}^M Tree_m(\mathbf{X}_t)\tag{20}$$

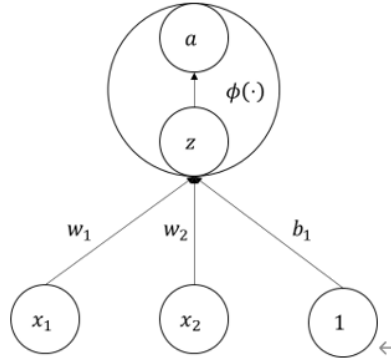
なお、ランダムフォレスト回帰の複雑さを調節する決定木の数や決定木の分岐の回数 (木の深さ) はハイパーパラメータであり、分析者が指定する必要がある。

ランダムフォレスト回帰は、ノンパラメトリックで非線形なモデルであり、説明変数と目的変数間に非線形な関係がある場合でも優れた予測が得られると考えられる。また、ブートストラップをもとに相関のない複数の決定木を推定して各決定木の予測値の平均を用いて予測を行うため、最終的な予測値の分散は小さく、予測の汎化性能が高まる性質を持つ。

6.3 ニューラルネットワークモデル

ニューラルネットワーク (Neural Network: NN) について述べる。まず NN の構造の最小単位であるパーセプトロンについて考える。 $\mathbf{X}_t = (X_{t1}, \dots, X_{tk})$ を入力 (説明変数) ベクトルとする。 (w_1, \dots, w_k) は入力ベクトルの要素それぞれに対応する重み (パラメータ) であり、スカラー値 b はバイアス項 (定数項) という。

図 2: パーセプトロンのイメージ



パーセプトロンではまず始めに $X_t = (X_{t1}, \dots, X_{tk})$ と (w_1, \dots, w_k) の内積と b の和であるブリアクティベーション z を計算する .

$$z = w_1 X_{t1} + \dots + w_k X_{tk} + b \quad (21)$$

次に z は非線形な関数である活性化関数 $\phi(\cdot)$ に渡され , アクティベーション a として出力される .

$$a = \phi(z) \quad (22)$$

以上が NN のパーセプトロンの構造であり , まとめると以下の式で表される .

$$a = \phi(w_1 X_{t1} + \dots + w_k X_{tk} + b) \quad (23)$$

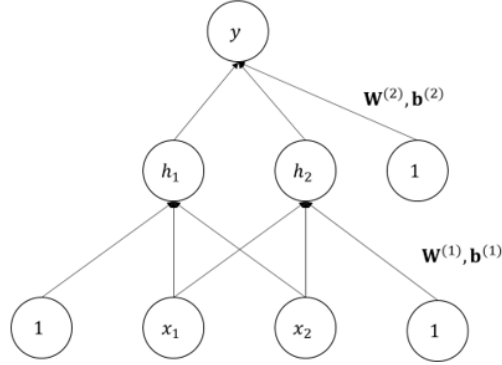
このパーセプトロンを複数組み合わせることで , NN が構成される .

NN と一口に言ってもその構造によって様々な種類のものがある . NN の 1 種である順伝播型ニューラルネットワーク (Feed-forward Neural Network: FNN) は NN のなかでも基本的な構造を持ち , 全てのパーセプトロンが結合した無閉路グラフで表現されすべての計算が逐次的に行われる . FNN の最も代表的なネットワークは多層パーセプトロン (multilayer perceptron: MLP) であり , 図 2 のように表される .

MLP の特徴として , 横のつながりのパーセプトロンをそれぞれ層としてまとめると少なくとも 3 つ以上の層から構成される . また MLP はネットワークの全てのパーセプトロンが次の層に含まれる全てのパーセプトロンとそれぞれ結合している . ネットワークの最初の層は入力層といい入力される特徴量がここにあたる . 最後の層は出力層といい , ネットワークが出力する値であり , 出力層のアクティベーションは予測値を表す . この 2 つを除いた間の層は全て隠れ層といい , MLP の多層構造を作る .

例えば , 入力層と 1 層の隠れ層 , 出力層から成る 3 層パーセプトロンを定式化してみる . まず , 隠れ層のアクティベーションをベクトル (H_{t1}, \dots, H_{tL}) とする . ただし , L は隠れ層に含まれるパーセプトロンの数である . また , 隠れ層の活性化関数を $\phi^{(1)}(\cdot)$ とし , 隠れ層のバイアス項を $(b_1^{(1)}, \dots, b_L^{(1)})$ とすると , 隠れ層のアクティベーションは以下のように表される .

図 3: MLP のイメージ



$$\begin{aligned}
 H_{t1} &= \phi^{(1)} \left(w_{11}^{(1)} X_{t1} + w_{21}^{(1)} X_{t2} + \cdots + w_{k1}^{(1)} X_{tk} + b_1^{(1)} \right) \\
 H_{t2} &= \phi^{(1)} \left(w_{12}^{(1)} X_{t1} + w_{22}^{(1)} X_{t2} + \cdots + w_{k2}^{(1)} X_{tk} + b_2^{(1)} \right) \\
 &\vdots \\
 H_{tL} &= \phi^{(1)} \left(w_{1L}^{(1)} X_{t1} + w_{2L}^{(1)} X_{t2} + \cdots + w_{kL}^{(1)} X_{tk} + b_L^{(1)} \right)
 \end{aligned} \tag{24}$$

次に，出力層について考える．予測する変数は本稿では 1 次元であるため，出力層のアクティベーションはスカラー Y となる．出力層の活性化関数を $\phi^{(2)}(\cdot)$ とし，バイアス項を $b^{(2)}$ とすると，出力層のアクティベーション（または予測値）は以下のように表される．図 3 は MLP のイメージを表している．

$$\hat{Y}_t = \phi^{(2)} \left(w_1^{(2)} H_{t1} + w_2^{(2)} H_{t2} + \cdots + w_L^{(2)} H_{tL} + b^{(2)} \right) \tag{25}$$

このように，MLP は線形結合モデルを非線形変換したものの繰り返しであることがわかる．この MLP の隠れ層の数を増やすことで，ネットワークをより深くすることができ，深層ニューラルネットワークが構築できる．一方，Qi (1999) は，隠れ層のパーセプトロン数が十分であれば，隠れ層が 1 つで，隠れ層の活性化関数にロジスティック関数，出力層の活性化関数に恒等関数を用いた 3 層パーセプトロンは，あらゆる連続関数を近似できると述べている．そこで本稿は，Callen et al. (1996) や Zhang et al. (2004) と同様に以下のような 3 層パーセプトロンを用いる．

$$\begin{aligned}
 \hat{Y}_t^{NN} &= f(\mathbf{X}_t \boldsymbol{\alpha}, \boldsymbol{\beta}) \\
 &= \sum_{l=1}^L (\alpha_0 + \alpha_l H_{tl}) \\
 &= \sum_{l=1}^L \left\{ \alpha_0 + \alpha_l \text{logistic} \left(\sum_{j=1}^k (\beta_{0l} + \beta_{jl} X_{tj}) \right) \right\}
 \end{aligned} \tag{26}$$

ただし，

X_t : 時点 t の入力ベクトル
 X_{tk} : k 番目の入力
 L : 隠れ層のパーセプトロン数
 α_l : 隠れ層の l 番目のパーセプトロンと目的変数の重み
 α_0 : 出力層のバイアス項
 β_{jl} : j 番目の入力と l 番目の隠れ層のパーセプトロンの重み
 β_{0l} : 隠れ層の l 番目のパーセプトロンのバイアス項

(27)

$$\text{logistic}(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)} : \text{ロジスティック関数}$$

NN の推定は誤差逆伝播法 (Backward propagation algorithm) を用いて重みの勾配を計算し、勾配降下によって行う。なお、NN の複雑さを調節する隠れ層のパーセプトロン数と、重みの更新量の幅を調節する学習率はハイパーパラメータであり、分析者が指定する必要がある。

7 ハイパーパラメータの選択

多くの機械学習アルゴリズムには、モデル推定の枠組みでは決定されない、分析者が指定する必要のあるハイパーパラメータが存在する。ハイパーパラメータの値が正しく設定できていないと、それをもとに推定されたモデルは過学習もしくは学習不足^{*7}となる恐れがあり、未知のデータに対して汎化性能を持たず、高い予測精度が期待できない。そこで本稿では適切なハイパーパラメータの値を設定するため、企業ごとにグリッドサーチを行う。

具体的には、ある 1 企業の 12 個のローリングサンプルそれぞれの訓練データのうち、最新時点のデータをあらたに検証データ (validation data) として分割する。例えば 1 つ目のローリングサンプル $\{Data_{2008Q1}, Data_{2008Q2}, \dots, Data_{2017Q3}, Data_{2017Q4}, Data_{2018Q1}\}$ では、訓練データを $\{Data_{2008Q1}, Data_{2008Q2}, \dots, Data_{2017Q3}\}$ 、検証データを $\{Data_{2017Q4}\}$ 、テストデータを $\{Data_{2018Q1}\}$ として 3 つに分割する。この分割を 12 個のローリングサンプルに対して行い、12 個の検証データ $\{Data_{2017Q4}, \dots, Data_{2020Q3}\}$ を得る。次に、候補となるハイパーパラメータの組を用意し、全通りのハイパーパラメータをそれぞれ用いて、訓練データのみをもとに 12 個のモデルを推定し、検証データの目的変数を予測する。そして、検証データに対する予測能力が最も高いハイパーパラメータの組合せを最適なモデル設定であるとする。最終的には、最適なハイパーパラメータをもとに訓練データと検証データの両方を用いてモデルを再推定し、テストデータの目的変数を予測する。図 4 は本稿におけるハイパーパラメータ選択のイメージを表している。

8 予測精度指標

各モデルの四半期 EPS 予測の精度を測るために、以下の予測精度指標を用いる。

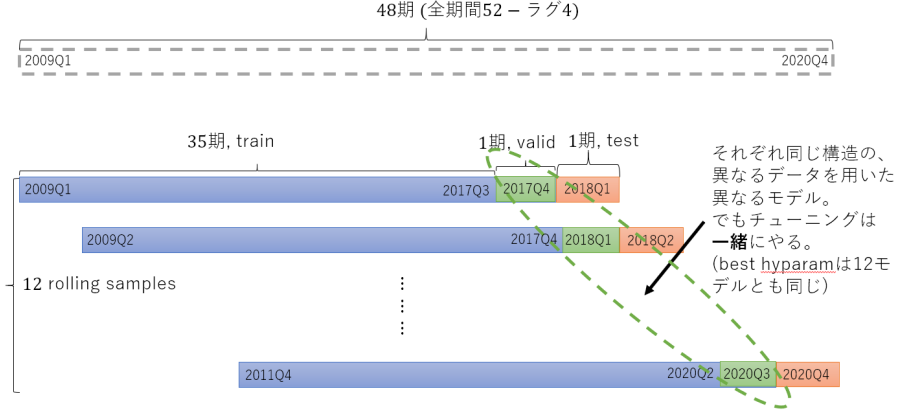
平均絶対誤差 (Mean Absolute Error: MAE)

^{*7} 過学習の反対語で、訓練データの規則性を十分とらえきれていないことを意味する。例えば、罰則付き回帰の罰則の強さを表す λ の値がデータに対して大きすぎる場合、係数の大きさに対する罰則が必要以上に強く、モデルがデータの規則性を十分に学習できない可能性が生じる。

図 4: ハイパーパラメータ選択のイメージ

チューニングの実装

・チューニング法 その2



$$MAE = \frac{1}{T} \sum_{t=1}^T |Y_t - \hat{Y}_t| \quad (28)$$

平均絶対誤差率 (Mean Absolute Percentage Error: MAPE)

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (29)$$

平均二乗誤差率 (Mean Squared Percentage Error: MSPE)

$$MSPE = \frac{1}{T} \sum_{t=1}^T \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)^2 \quad (30)$$

ただし、 T は予測期間であり、本稿では 2018 年度第 1 四半期～2020 年度第 4 四半期の 12 期間である。なお、絶対予測誤差率 $\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$ が 1 を超える予測サンプルについては、Brown and Rozeff (1979), Lorek and Willinger, Zhang et al. (2004) に倣い絶対予測誤差率の上界を 1 とし、この制約に基づいて算出した精度指標を報告することとする。また、そのような予測サンプルは Large Forecast Error サンプルとし、予測精度指標の値と併せて Large Forecast Error の割合も報告する。

9 Diebold-Mariano(DM) 検定

予測精度指標は異なる予測モデル間で予測精度を比較するために用いられる。一方、精度指標の値は本稿で用いているデータセットのみに由来しているため、精度指標の大小関係を比較するだけではモデルの精度の優劣を一般化することはできない。そこで、本稿では異なる 2 つの予測モデル間に統計的に有意な予測精度の差が存在するかどうかを検定するために、Diebold-Mariano(DM) 検定 (Diebold and Mariano, 2002) を行う。

具体的には、ある企業について、予測期間の実績値（真の値）を $\{Y_t; t = 1, \dots, T\}$ ，異なる 2 つの予測モデル（モデル 1 とモデル 2）に由来する予測値系列をそれぞれ $\{\hat{Y}_{1t}; t = 1, \dots, T\}$ ， $\{\hat{Y}_{2t}; t = 1, \dots, T\}$ とする．このとき予測誤差系列をそれぞれ，

$$\begin{aligned} \{e_{1t} = \hat{Y}_{1t} - Y_t; t = 1, \dots, T\} \\ \{e_{2t} = \hat{Y}_{2t} - Y_t; t = 1, \dots, T\} \end{aligned} \quad (31)$$

とし、予測誤差の損失 $L(e_t)$ ^{*8}をもとに 2 つの予測誤差系列の損失差 d_{12t} を以下のように定義する．

$$d_{12t} = L(e_{1t}) - L(e_{2t}) \quad (32)$$

モデル 1 とモデル 2 の予測誤差に差がないとき、損失差の期待値は 0 となる．つまり、DM 検定の帰無仮説と対立仮説は次のとおりである．

$$\begin{aligned} H_0 : E(d_{12t}) &= 0, \\ H_1 : E(d_{12t}) &\neq 0 \end{aligned} \quad (33)$$

この仮説検定を行うために以下のような検定統計量、

$$\begin{aligned} DM_{12} &= \frac{\bar{d}_{12}}{s.e.(\bar{d}_{12})} \\ \text{ただし、} \\ \bar{d}_{12} &= \frac{1}{T} \sum_{t=1}^T d_{12t} \\ s.e.(\bar{d}_{12}) &\text{は } \bar{d}_{12} \text{ の標準誤差。} \end{aligned} \quad (34)$$

を導入し、 d_{12t} について緩い仮定の下で DM_{12} が標準正規分布に漸近的分布することを用いる^{*9}．両側 DM 検定で統計的に有意な検定結果が得られなければ、モデル 1 とモデル 2 の間に予測パフォーマンスの統計的に有意な差がない結果となる．DM 検定量が正の値で統計的に有意であれば、モデル 1 の予測誤差はモデル 2 よりも大きく、モデル 2 の方が優れた予測を与えるという結果となる．反対に、DM 検定量が負の値で統計的に有意であれば、モデル 1 の予測誤差はモデル 2 よりも小さく、モデル 1 の方が優れた予測を与えるという結果となる．

第 V 部

予測結果

10 時系列モデル間の予測精度比較

^{*8} 損失関数 $L(e_t)$ は用いる予測精度指標によって異なる．例えば、MAE の場合は $L(e_t) = |e_t|$ ，MAPE の場合は $L(e_t) = \left| \frac{e_t}{Y_t} \right|$ ，

MSPE の場合は $L(e_t) = \left(\frac{e_t}{Y_t} \right)^2$ となる．

^{*9} DM 検定の詳細については Diebold and Mariano (2002) を参照されたい．

表 3: 時系列モデルによる 1 期先四半期 EPS 予測の精度 (1,003 社平均)

Model ^a	Q1				Q2				Q3				Q4				Overall			
	MAE ^b	MAPE ^c	MSPE ^d	Large Forecast Error(%) ^e	MAE	MAPE	MSPE	Large Forecast Error(%)	MAE	MAPE	MSPE	Large Forecast Error(%)	MAE	MAPE	MSPE	Large Forecast Error(%)	MAE	MAPE	MSPE	Large Forecast Error(%)
U-RW	44.2	0.668	0.572	42.6	27.6	0.529	0.408	25.7	29.7	0.499	0.372	22.3	47.9	0.621	0.514	36.1	37.3	0.579	0.467	31.7
U-SRW	27.1	0.547	0.436	27.8	27.4	0.505	0.390	23.8	26.3	0.461	0.338	20.4	47.2	0.573	0.464	32.2	32.0	0.521	0.407	26.0
U-ARIMA-F	32.6	0.549	0.437	29.5	25.7	0.480	0.362	22.4	28.3	0.477	0.355	21.4	46.7	0.573	0.463	32.3	33.3	0.520	0.404	26.4
U-ARIMA-G	37.7	0.567	0.460	32.4	30.9	0.481	0.366	23.9	33.2	0.488	0.370	23.5	50.2	0.570	0.461	31.8	38.0	0.526	0.414	27.9
U-ARIMA-BR	36.4	0.559	0.445	29.3	29.2	0.482	0.362	22.5	30.1	0.471	0.348	21.0	47.7	0.564	0.452	30.1	35.8	0.519	0.402	25.7
M-OLS1	64.1	0.626	0.527	38.8	50.6	0.556	0.447	32.0	40.4	0.528	0.414	27.4	60.9	0.603	0.499	35.4	54.0	0.578	0.472	33.4
M-OLS2	62.5	0.613	0.514	38.0	44.0	0.549	0.439	30.8	77.5	0.569	0.461	32.4	84.4	0.625	0.526	38.6	67.1	0.589	0.485	35.0
M-OLS3	333.8	0.821	0.769	67.9	278.6	0.798	0.738	63.8	263.9	0.790	0.727	62.0	225.4	0.797	0.733	62.7	275.4	0.801	0.742	64.1
U-RIDGE	41.9	0.576	0.469	33.5	35.4	0.506	0.391	26.2	37.0	0.499	0.373	22.2	49.7	0.575	0.464	31.5	41.0	0.539	0.424	28.3
U-LASSO	35.2	0.591	0.474	28.9	31.0	0.530	0.401	21.6	33.9	0.521	0.385	18.4	45.5	0.592	0.473	27.9	36.4	0.559	0.433	24.2
U-EN	34.0	0.581	0.462	28.4	30.7	0.519	0.390	21.4	32.7	0.510	0.372	17.7	45.6	0.591	0.471	27.6	35.7	0.550	0.423	23.8
U-RF	27.1	0.548	0.435	28.8	25.0	0.480	0.358	22.2	26.6	0.464	0.338	19.2	40.8	0.561	0.445	29.1	29.9	0.513	0.394	24.8
U-NN	50.7	0.624	0.515	34.4	49.4	0.580	0.460	28.2	49.4	0.554	0.430	25.7	60.5	0.626	0.516	34.0	52.5	0.596	0.480	30.6
M-RIDGE	38.1	0.527	0.419	29.3	33.9	0.483	0.374	25.1	36.6	0.468	0.355	24.2	45.7	0.534	0.429	30.3	38.6	0.503	0.394	27.2
M-LASSO	23.6	0.490	0.374	23.1	23.5	0.438	0.321	19.1	24.3	0.422	0.302	17.2	38.1	0.495	0.382	24.4	27.4	0.461	0.345	21.0
M-EN	22.9	0.479	0.362	21.8	22.7	0.425	0.308	18.2	23.7	0.417	0.296	16.8	37.7	0.492	0.377	23.7	26.8	0.453	0.336	20.1
M-RF	24.7	0.495	0.382	25.7	21.2	0.413	0.295	17.9	22.7	0.408	0.287	17.2	36.8	0.513	0.398	25.8	26.3	0.457	0.340	21.7
M-NN	35.8	0.588	0.480	33.7	36.8	0.534	0.416	27.1	33.0	0.517	0.392	23.4	46.4	0.595	0.485	33.4	38.0	0.558	0.443	29.4

^a 各モデルの先頭文字列 U-は単変量モデル, M-は多変量モデルを表す。また, RW はランダムウォークモデル, SRV は季節ランダムウォークモデル, ARIMA-F, ARIMA-G, ARIMA-BR はそれぞれ Foster (1977); Griffin (1977); Brown and Rozeff (1979) の ARIMA モデル, OLS1, OLS2, OLS3 はそれぞれ式 (11), 式 (12), 式 (13) の線形回帰モデル, RIDGE は Ridge 回帰モデル, LASSO は LASSO 回帰モデル, EN は Elastic Net 回帰モデル, RF はランダムフォレスト回帰モデル, NN はニューラルネットワークモデルを表す。

^b 平均絶対誤差 (Mean Absolute Error: MAE)

^c 平均絶対誤差率 (Mean Absolute Percentage Error: MAPE)

^d 平均二乗誤差率 (Mean Squared Percentage Error: MSPE)

^e 絶対予測誤差率 $\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$ が 1 を超える Large Forecast Error サンプルについては絶対予測誤差率を 1 としており, 表に全予測サンプルに占める Large Forecast Error サンプルの割合 (%) を示している。

表 3 は各予測指標の全企業平均を四半期別と全予測期間で集計したものを示している。まず、単変量モデル間での予測精度を比較すると、全予測期間平均で MAE, MAPE, MSPE が最も小さいモデルはランダムフォレスト回帰 (それぞれ 29.9, 0.513, 0.394), Large Forecast Error が最小のモデルは Elastic Net 回帰 (23.8%) であり、いずれの指標においても機械学習アルゴリズムの予測精度が伝統的な単変量時系列モデルの予測精度を上回る結果となった。各四半期別平均についても、第 1 四半期や第 3 四半期で季節ランダムウォークに劣る場合があるものの、機械学習アルゴリズム (特に Elastic Net 回帰やランダムフォレスト回帰) が最も精度の高い予測を与えている。

次に、多変量モデル間での予測精度を比較してみる。全予測期間平均で MAE が最小のモデルはランダムフォレスト回帰 (26.3), MAPE, MSPE, Large Forecast Error が最小のモデルは Elastic Net 回帰 (それぞれ 0.453, 0.336, 20.1%) であり、多変量モデルについても全ての指標で機械学習アルゴリズムの予測精度が伝統的な多重線形回帰モデルの予測精度を上回る結果となった。各四半期別平均についても全ての精度指標において機械学習アルゴリズムが伝統的な多重線形回帰モデルよりも精度の高い予測を与えている。この結果は、機械学習アルゴリズムが日本の四半期 EPS 予測において、1 期先四半期 EPS と過去の四半期 EPS の実績値やファンダメンタル会計変数との関係をうまく捉えている可能性を示している。

さらに単変量モデルと多変量モデルの予測精度を比較すると、ファンダメンタル会計変数を予測に用いることで、多変量線形回帰モデルはいずれも単変量線形時系列モデルの精度を下回る一方、多変量機械学習アルゴリズムによる予測は単変量機械学習アルゴリズムによる予測よりも精度が向上する傾向があった。この結果は Zhang et al. (2004) と整合的であり、ファンダメンタル会計変数には将来の四半期 EPS を予測する情報があり、その情報は伝統的な線形モデルでは捉えられないことを示唆する。

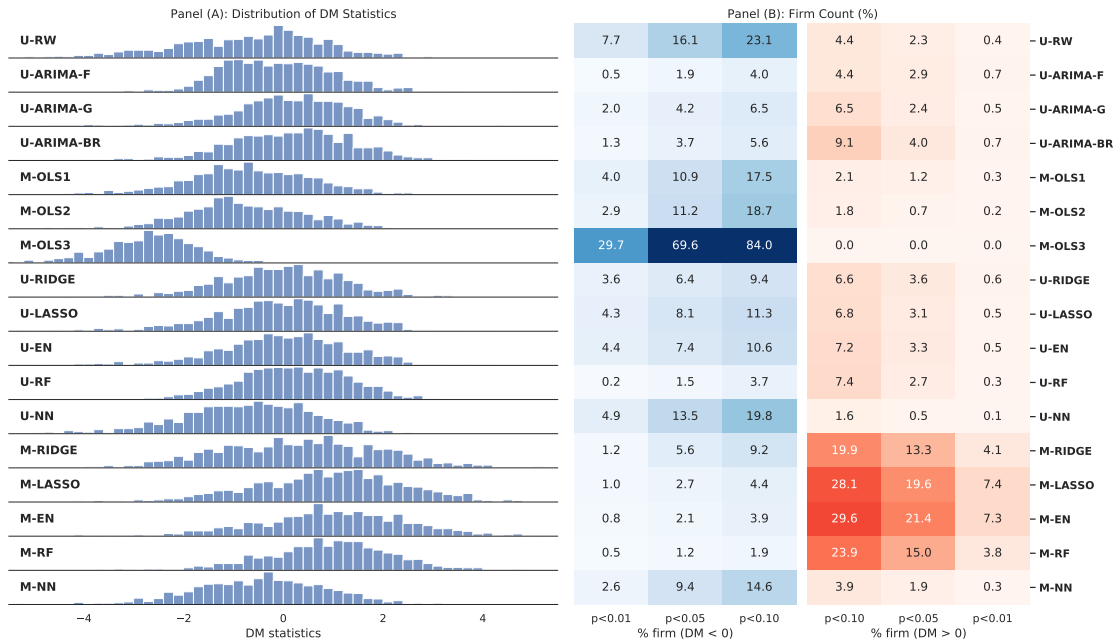
推定した全ての時系列モデルの中で、多変量機械学習モデル (特に Elastic Net 回帰, ランダムフォレスト回帰) がいずれの精度指標や四半期別平均においても最も精度の高い予測を与えている。＜――一方で、NN については過去の研究から期待されるような高い精度の予測結果は得られなかった。――＞ NN の精度が低い理由として、NN は本稿で用いた他の機械学習アルゴリズムよりも推定するパラメータが桁違いに多く、よりサンプルサイズが大きいデータが必要となるため、本稿の予測のデザインでは NN による予測の精度を高めるために十分なサンプルを用意できなかったことが挙げられる。また、日本企業の四半期 EPS データは隠れ層 1 層で近似できない関数形である可能性も考えられ、モデルの設定やハイパーパラメータ選択が適切に行われていない可能性もある。＜―― Chatfield et al. (1993), Hill et al. (1994), Callen et al. (1996) の研究は、NN は必ずしもあらゆる時系列予測の文脈でパフォーマンスが高いモデルではないとし、NN を用いることが適切な分野とそうでない分野があると指摘している。――＞本稿としては、日本企業の四半期 EPS 予測は NN にとってパフォーマンスを発揮しにくい文脈であるという結果を報告する。

どの時系列モデルに関しても、第 4 四半期の予測が他の四半期よりも精度が低い結果となっている。過去の四半期 EPS 研究でも同様な傾向が確認されており、これに関して桜井 (1990) は第 4 四半期には年次利益との一致のための利益調整が含まれているためであると解釈している。

なお、一般的に異なる予測手法の予測値系列を組み合わせ^{*10}得る組合せ予測 (Bates and Granger, 1969) は個々の予測よりも精度が高い傾向があるとされる。本稿では、単変量線形時系列モデル、多変量線形モデル、単変量機械学習アルゴリズム、多変量機械学習アルゴリズム、全時系列モデル、それぞれの組合せ予測を求めたが、いずれも目立った予測精度の向上は確認できなかった。

*10 単純平均など

図 5: 全企業 (1,003 社) の DM 検定結果のまとめ (モデル 1: SRW, loss: MAE)



(注) 各 DM 検定では U-SRW をモデル 1 としており、損失関数は MAE に対応するもの ($L(e_t) = |e_t|$) を用いている。Panel(A) はモデル 2 別の全 1,003 社の DM 統計量の分布をヒストグラムで表している。Panel(B) は全 1,003 社の DM 検定のうち、10%、5%、1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。色が濃いほど統計的に有意な企業の割合が大きいことを示している。

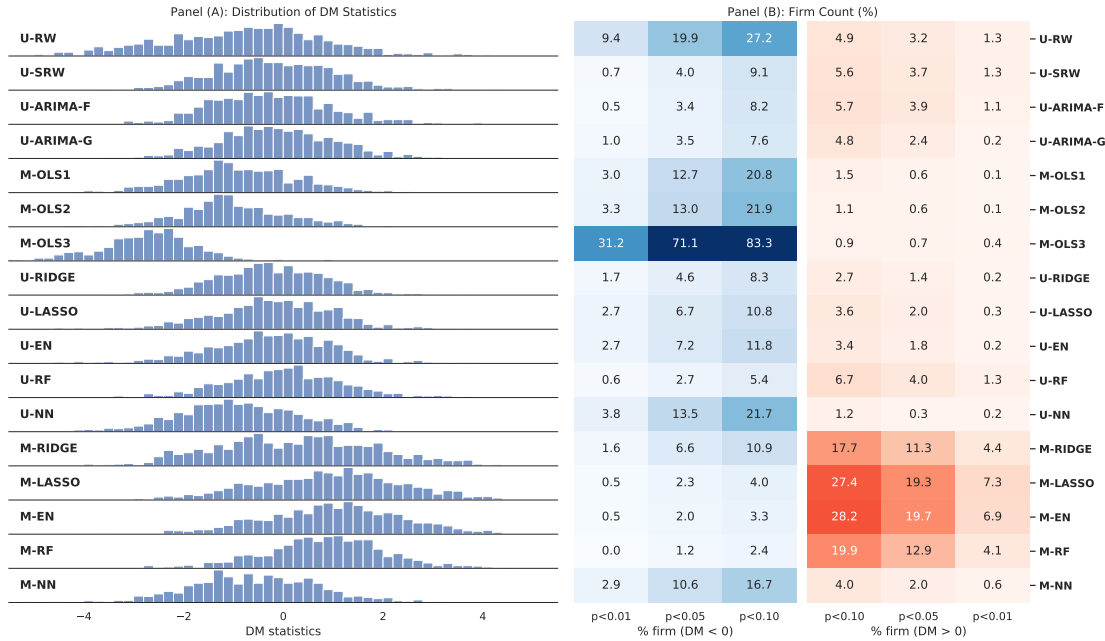
11 時系列モデル間の予測精度差の検定

表 3 において予測パフォーマンスが最も良いとされる多変量機械学習アルゴリズムが、他の時系列モデルに比べて統計的に有意に予測精度が高いかどうかを検定するために DM 検定を行う。まず、ベンチマークモデルと多変量機械学習アルゴリズムの予測精度の差の統計量有意性を検定するために、季節ランダムウォークをモデル 1 として、その他のモデル全通りをそれぞれモデル 2 とした場合の両側 DM 検定を各企業 (全 1,003 社) の予測値系列に対して行う。図 5.Panel(A) は全 1,003 社の DM 統計量の分布をモデル 2 別に示したヒストグラムであり、図 5.Panel(B) は全 1,003 社の両側 DM 検定のうち、10%、5%、1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。なお、図 5 の DM 検定で用いた損失関数は MAE に対応したものである^{*11}。

図 5.Panel(A) より、モデル 2 が < ー多変量 LASSO 回帰、Ridge 回帰、Elastic Net 回帰、ランダムフォレスト回帰の場合、DM 統計量の分布は正の値に集中している。ー > また、図 5.Panel(B) から、先述の 4 つの多変量機械学習アルゴリズムにおいて、< ー DM 統計量が正の値で統計的に有意な企業の割合は、10% 水準の場合約 20% ~ 30%、5% 水準の場合約 13% ~ 21%、1% 水準の場合約 4% ~ 7% となっている。ー > つまり、多変量 Elastic Net 回帰と多変量ランダムフォレスト回帰に加え、多変量 Ridge 回帰と多変量 LASSO 回帰も、季節ランダムウォークに比べて多くの企業で統計的に有意に予測誤差が小さい結果となっており、こ

^{*11} MAE 以外の精度指標、MAPE、MSPE に対応する損失関数を用いた DM 検定も同様な結果であった。

図 6: 全企業 (1,003 社) の DM 検定結果のまとめ (モデル 1: ARIMA-BR, loss: MAE)



(注) 各 DM 検定では U-ARIMA-BR をモデル 1 としており、損失関数は MAE に対応するもの ($L(e_t) = |e_t|$) を用いている。Panel(A) はモデル 2 別の全 1,003 社の DM 統計量の分布をヒストグラムで表している。Panel(B) は全 1,003 社の DM 検定のうち、10%, 5%, 1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。色が濃いほど統計的に有意な企業の割合が大きいことを示している。

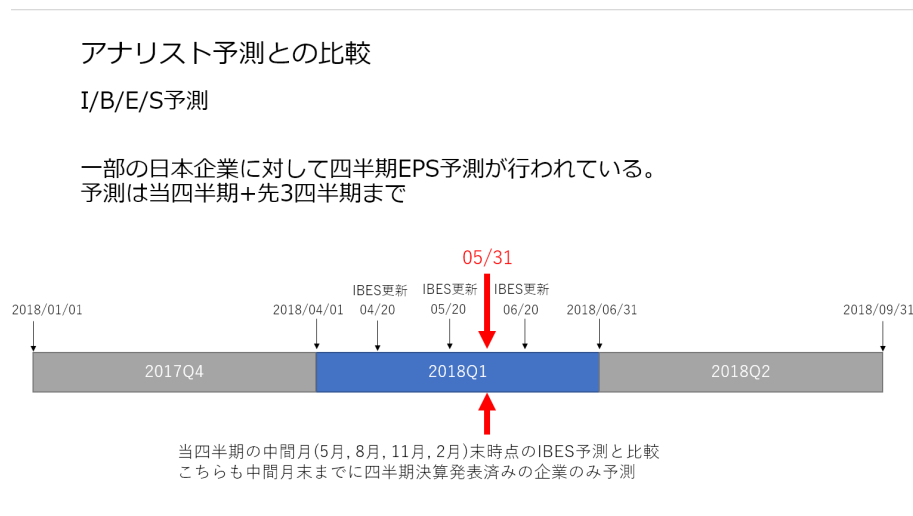
これは多変量機械学習アルゴリズムが季節ランダムウォークよりも優れた予測を与えていることを意味する。一方、多変量線形回帰モデルについては図 5.Panel(A) から <— DM 統計量の分布は負の値に集中している —>, 図 5.Panel(B) でも DM 統計量が負の値で統計的に有意な企業の割合が多いことが確認できる。つまり、線形回帰モデルはファンダメンタル会計変数を含めることでかえって予測精度が下がり、予測に何の情報も用いていない季節ランダムウォークよりも劣ってしまうことを意味する。単変量機械学習アルゴリズムについては、図 5.Panel(A) と図 5.Panel(B) より、統計的に有意に予測精度の差がある企業が少なく、季節ランダムウォークと同程度の予測パフォーマンスであることがわかった。

次に、伝統的線形時系列モデルと多変量機械学習アルゴリズムの予測精度の差の統計的有意性を検定するために、Brown and Rozeff (1979) の ARIMA モデルをモデル 1 として、その他のモデル全通りをそれぞれモデル 2 とした場合の両側 DM 検定を行う。図 6.Panel(A) は全 1,003 社の DM 統計量の分布をモデル 2 別に示したヒストグラムであり、図 6.Panel(B) は全 1,003 社の両側 DM 検定のうち、10%, 5%, 1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。なお、図 6 の DM 検定で用いた損失関数は MAE に対応したものである^{*12}。

Brown and Rozeff (1979) の ARIMA モデルをモデル 1 とした DM 検定の結果は、季節ランダムウォークをモデル 1 とした DM 検定の結果と同様であった。図 6.Panel(A) より、モデル 2 が多変量 LASSO 回帰、Ridge 回帰、Elastic Net 回帰、ランダムフォレスト回帰の場合、<— DM 統計量の分布は正の値に集中し

^{*12} MAE 以外の精度指標、MAPE, MSPE に対応する損失関数を用いた DM 検定も同様な結果であった。

図 7: I/B/E/S 業績予想の更新タイミング



ておりー> , 図 6.Panel(B) から DM 統計量が正の値で統計的に有意な企業の割合が多いことが確認できる (10%, 5%, 1% 水準で統計的に有意な企業の割合は, それぞれ約 18%~28%, 11%~20%, 4%~7%)。つまり, 多変量機械学習アルゴリズムは伝統的な線形時系列モデルである Brown and Rozeff (1979) の ARIMA モデルよりも多くの企業において統計的に有意に予測誤差が小さく, 優れた予測を与えていることを意味する^{*13}

12 アナリスト予測との予測精度比較

本稿では, 時系列モデルによる予測と人的な予測の精度を比較するために, 人的な予測として I/B/E/S(Institutional Brokers Estimate System) 業績予想を用いる。I/B/E/S 業績予想は, 複数のアナリストの利益予想の平均値であるコンセンサス予想であり, 日本においても I/B/E/S の四半期 EPS 予測がカバーされている企業がある^{*14}。I/B/E/S 業績予想は通常, 毎月第 3 金曜日に更新されており, 本稿の時系列モデルの予測タイミングと揃えるために各四半期の中間月末 (5 月末, 8 月末, 11 月末, 2 月末) 時点の I/B/E/S 業績予想の 1 期先四半期 EPS 予測を比較対象とする。図 7 は本稿及び I/B/E/S の四半期 EPS 予測のタイミングの概要を表している。

本稿で用いる I/B/E/S 業績予想のデータは「I/B/E/S on Datastream」から収集している。I/B/E/S 業績予想のサンプルについては, 第 III 部 1 節と同様な基準のもとで選択し, 予測期間についても時系列モデルと同様に 2018 年度第 1 四半期~2020 年度第 4 四半期とする。結果として 88 社, 延べ 88 社 × 12 四半期 = 1,056 個の企業四半期予測が得られた。なお, 時系列モデルと I/B/E/S 業績予想の予測精度を比較するために, 時系列モデルのサンプル企業については, 計算可能であった 1,003 社から I/B/E/S 業績予想が調査対象としている 88 社に減らして, 分析対象企業を I/B/E/S 業績予想と揃えることとする。表 4 は時系列モデルと I/B/E/S 業績予想の, 各予測指標の全 88 社平均を四半期別と全予測期間で集計したものを示している。

^{*13} Foster (1977), Griffin (1977) の ARIMA モデルをそれぞれモデル 1 とした DM 検定も同様な結果であった。

^{*14} 全ての上場企業はカバーされていない。

表 4: 時系列モデルによる予測と I/B/E/S 業績予想の精度比較 (88 社平均)

Model ^a	Q1					Q2					Q3					Q4					Overall				
	MAE ^b	MAPE ^c	MSPE ^d	Large Forecast Error(%) ^e		MAE	MAPE	MSPE	Large Forecast Error(%)		MAE	MAPE	MSPE	Large Forecast Error(%)		MAE	MAPE	MSPE	Large Forecast Error(%)		MAE	MAPE	MSPE	Large Forecast Error(%)	
U-RW	46.0	0.586	0.465	29.2		31.9	0.463	0.340	21.6		39.7	0.443	0.317	18.2		58.9	0.625	0.528	37.9		44.1	0.529	0.413	26.7	
U-SRW	35.9	0.491	0.372	23.9		29.4	0.443	0.330	22.3		41.8	0.454	0.331	20.5		57.4	0.602	0.496	36.0		41.1	0.498	0.382	25.7	
U-ARIMA-F	36.6	0.487	0.374	24.2		25.8	0.414	0.301	19.7		41.6	0.457	0.337	20.8		55.6	0.602	0.490	33.3		39.9	0.490	0.376	24.5	
U-ARIMA-G	37.2	0.522	0.405	27.3		29.8	0.427	0.321	21.6		41.4	0.467	0.344	21.6		56.5	0.607	0.498	36.4		41.2	0.506	0.392	26.7	
U-ARIMA-BR	35.8	0.498	0.380	22.7		27.1	0.420	0.303	18.2		36.9	0.434	0.309	18.2		51.3	0.583	0.471	31.8		37.8	0.484	0.365	22.7	
M-OLS1	49.7	0.574	0.458	29.9		40.7	0.481	0.368	26.5		46.5	0.487	0.368	22.7		69.0	0.597	0.486	36.0		51.5	0.535	0.420	28.8	
M-OLS2	46.9	0.550	0.438	29.2		39.1	0.480	0.366	24.6		85.5	0.507	0.385	25.4		68.1	0.641	0.548	43.6		59.9	0.544	0.434	30.7	
M-OLS3	261.9	0.777	0.712	60.6		197.9	0.780	0.712	59.8		207.2	0.747	0.672	54.9		219.0	0.780	0.714	60.2		221.5	0.771	0.703	58.9	
U-RIDGE	38.1	0.516	0.399	26.5		31.7	0.435	0.311	20.8		41.7	0.463	0.334	19.3		49.0	0.592	0.479	34.1		40.1	0.501	0.381	25.2	
U-LASSO	35.2	0.529	0.403	24.2		28.1	0.431	0.296	16.7		38.5	0.471	0.335	17.0		44.9	0.587	0.468	31.8		36.7	0.504	0.376	22.4	
U-EN	34.9	0.522	0.395	23.9		27.7	0.424	0.292	16.3		38.2	0.466	0.327	15.9		44.6	0.584	0.465	31.1		36.3	0.499	0.370	21.8	
U-RF	34.2	0.510	0.389	26.1		28.7	0.413	0.283	17.4		38.6	0.460	0.322	17.0		46.9	0.568	0.451	30.7		37.1	0.488	0.361	22.8	
U-NN	43.3	0.581	0.458	27.3		35.8	0.507	0.375	22.0		43.5	0.508	0.373	20.8		54.0	0.635	0.525	35.2		44.2	0.558	0.433	26.3	
M-RIDGE	48.2	0.509	0.402	28.0		41.9	0.490	0.379	25.4		44.9	0.461	0.346	22.3		54.3	0.551	0.457	34.1		47.3	0.503	0.396	27.5	
M-LASSO	25.0	0.432	0.315	18.6		22.4	0.389	0.263	13.6		30.7	0.391	0.276	17.4		36.6	0.488	0.378	26.1		28.7	0.425	0.308	18.9	
M-EN	25.8	0.426	0.310	18.2		24.0	0.404	0.277	14.8		31.3	0.407	0.291	17.8		39.4	0.501	0.392	27.7		30.1	0.435	0.318	19.6	
M-RF	26.3	0.439	0.320	19.7		22.6	0.375	0.262	17.4		34.1	0.396	0.275	18.2		40.2	0.526	0.410	28.8		30.8	0.434	0.317	21.0	
M-NN	41.0	0.536	0.414	28.0		36.1	0.505	0.381	25.0		41.3	0.483	0.355	20.5		54.1	0.606	0.493	33.0		43.1	0.532	0.411	26.6	
IBES	29.1	0.426	0.298	15.9		26.8	0.390	0.262	13.6		33.1	0.385	0.253	12.9		37.1	0.479	0.364	25.0		31.5	0.420	0.294	16.9	

^a 各モデルの先頭文字列 U-は単変量モデル, M-は多変量モデルを表す。また, RW はランダムウォークモデル, SRW は季節ランダムウォークモデル, ARIMA-F, ARIMA-G, ARIMA-BR はそれぞれ Foster (1977); Griffin (1977); Brown and Rozeff (1979) の設定の ARIMA モデル, OLS1, OLS2, OLS3 はそれぞれ式 (11), 式 (12), 式 (13) の線形回帰モデル, RIDGE は Ridge 回帰モデル, LASSO は LASSO 回帰モデル, EN は Elastic Net 回帰モデル, RF はランダムフォレスト回帰モデル, NN はニューラルネットワークモデルを表す。IBES は I/B/E/S 業績予想を表す。

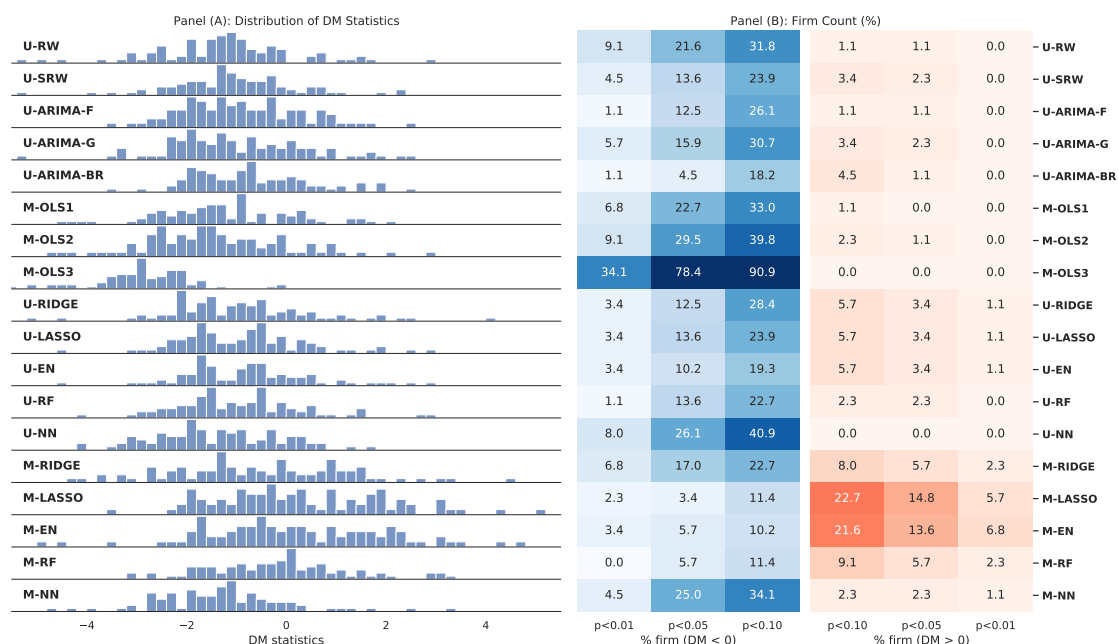
^b 平均絶対誤差 (Mean Absolute Error: MAE)

^c 平均絶対誤差率 (Mean Absolute Percentage Error: MAPE)

^d 平均二乗誤差率 (Mean Squared Percentage Error: MSPE)

^e 絶対予測誤差率 $\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$ が 1 を超える Large Forecast Error サンプルについては絶対予測誤差率を 1 としており, 表に全予測サンプルに占める Large Forecast Error サンプルの割合 (%) を示している。

図 8: 全企業 (88 社) の DM 検定結果のまとめ (モデル 1: IBES, loss: MAE)



(注) 各 DM 検定では IBES をモデル 1 としており、損失関数は MAE に対応するもの ($L(e_t) = |e_t|$) を用いている。Panel(A) はモデル 2 別の全 88 社の DM 統計量の分布をヒストグラムで表している。Panel(B) は全 88 社の DM 検定のうち、10%、5%、1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。色が濃いほど統計的に有意な企業の割合が大きいことを示している。

全予測期間平均で MAE が最小のモデルは多変量 LASSO 回帰 (28.7)、次いで多変量 Elastic Net 回帰 (30.1)、ランダムフォレスト回帰 (30.8) となり、多変量機械学習アルゴリズムがアナリスト予測である I/B/E/S 業績予想の MAE(31.5) を下回った。一方で、他の予測精度指標ではいずれも I/B/E/S 業績予想が最も小さい値となり (MAPE, MSPE, Large Forecast Error がそれぞれ 0.420, 0.294, 16.9%)、多変量機械学習アルゴリズム (特に LASSO 回帰, Elastic Net 回帰, ランダムフォレスト回帰) は I/B/E/S 業績予想と比較するとわずかに予測精度が低い結果となった。各四半期別平均についても同様な傾向が見られ、MAE においては機械学習アルゴリズムの方が I/B/E/S 業績予想よりも予測精度が高いものの、他の予測精度指標では I/B/E/S 業績予想よりも予測精度が低い結果となった。

なお、表 3 と表 4 を比較すると、時系列モデルでは、もともと計算可能な 1,003 社平均よりも I/B/E/S 業績予想が調査対象としている 88 社平均のほうが全体的に精度が高い傾向がみられる。――そもそもアナリストは全ての上場企業をカバーしているわけではなく、限られた企業のみを調査対象としているのが日本の現状である。これについて中井 (2006) は、アナリストは情報が入手しやすい企業をカバーする傾向にあると指摘しており、I/B/E/S 業績予想のアナリストの調査対象企業は比較的に予測が容易である可能性が示唆される。――

13 アナリスト予測との予測精度差の検定

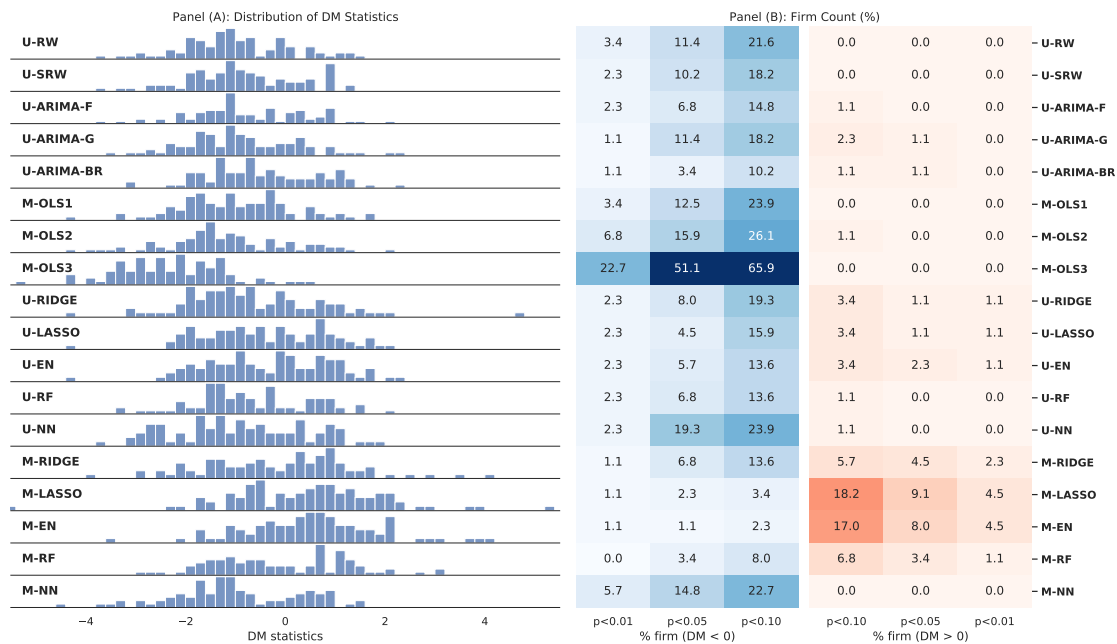
まず、表 4 で MAE が I/B/E/S 業績予想よりも小さい多変量 LASSO 回帰、Elastic Net 回帰、ランダムフォレスト回帰が、I/B/E/S 業績予想と比べて統計的に有意に予測誤差が小さいかどうかを DM 検定する。具体的には I/B/E/S 業績予想をモデル 1 として、その他のモデル全通りをそれぞれモデル 2 とした場合の、損失関数を MAE に対応するものとした両側 DM 検定を、I/B/E/S 業績予想が調査の対象としている企業（全 88 社）の予測値系列に対して行う。図 8.Panel(A) は全 88 社の DM 統計量の分布をモデル 2 別に示したヒストグラムであり、図 8.Panel(B) は全 88 社の両側 DM 検定のうち、10%、5%、1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。

図 8.Panel(A) より、多くの時系列モデル（特に単変量線形時系列モデル、多変量線形回帰モデル、単変量機械学習アルゴリズム）で $< -$ DM 統計量の分布が負の値に集中している。 $>$ 図 8.Panel(B) でも多くの時系列モデルにおいて DM 統計量が負の値で統計的に有意な企業の割合が多いことが確認でき、アナリスト予測は単変量線形時系列モデル、多変量線形回帰モデル、単変量機械学習アルゴリズムによる予測よりも精度が高いことを示している。一方、I/B/E/S 業績予想と多変量機械学習アルゴリズムの検定結果では、図 8.Panel(A) より、モデル 2 が多変量 LASSO 回帰と Elastic Net 回帰のとき $< -$ DM 検定量の分布が正の値に集中しており $>$ 、図 8.Panel(B) から、DM 統計量が正の値で統計的に有意な企業の割合が多いことが確認できる（10%、5%、1% 水準で統計的に有意な企業の割合は、それぞれ約 22~23%、14~15%、6~7%）。つまり、MAE において多変量機械学習アルゴリズムはアナリスト予測である I/B/E/S 業績予想よりも多くの企業において統計的に有意に予測誤差が小さく、優れた予測を与えていることを意味する。なお、モデル 2 が多変量ランダムフォレスト回帰の場合は有意水準と統計的に異なった企業の割合が同程度であり、I/B/E/S 業績予想と多変量ランダムフォレスト回帰の予測精度に差はない結果となった。

次に、損失関数を MAPE に対応するものとした両側 DM 検定を行い、図 9 にその結果をまとめている。表 4 では I/B/E/S 業績予想と多変量機械学習アルゴリズムを MAPE で比較すると、I/B/E/S 業績予想の方が小さい値であったが、図 9.Panel(A) より、モデル 2 が多変量 LASSO 回帰と Elastic Net 回帰のとき $< -$ DM 統計量が正の値に集中しており $>$ 、図 9.Panel(B) から、DM 統計量が正の値で統計的に有意な企業の割合が多いことも確認できる（10%、5%、1% 水準で統計的に有意な企業の割合は、それぞれ約 17~18%、8.9%、5%）。また、モデル 2 が多変量ランダムフォレスト回帰の場合は I/B/E/S 業績予想と同程度の精度の予測を与えている。なお、損失関数を MSPE に対応するものとした両側 DM 検定でも同様な結果が得られ、多変量機械学習アルゴリズムはアナリスト予測と比べると統計的に有意に精度が高い、もしくは精度の差がない予測を与える企業が多い結果となった。

通常、アナリストは本稿の時系列モデルが用いる情報（過去の四半期 EPS の実績値や過去のファンダメンタル会計変数）以外にも広範囲の情報を反映した予測を行うと考えられるため（桜井, 1990）、アナリスト予測は（多変量機械学習アルゴリズムも含めた）時系列モデルによる予測よりも精度が高いとされる。にもかかわらず本稿の DM 検定の結果から、多変量機械学習アルゴリズムはアナリスト予測と同等もしくはそれ以上に高い精度の予測を与えることが示された。

図 9: 全企業 (88 社) の DM 検定結果のまとめ (モデル 1: IBES, loss: MAPE)



(注) 各 DM 検定では IBES をモデル 1 としており、損失関数は MAE に対応するもの ($L(e_t) = \left| \frac{e_t}{\hat{Y}_t} \right|$) を用いている。Panel(A) はモデル 2 別の全 88 社の DM 統計量の分布をヒストグラムで表している。Panel(B) は全 88 社の DM 検定のうち、10%、5%、1% の水準で統計的に有意な企業の割合 (%) を DM 統計量の正負別に示したヒートマップである。色が濃いほど統計的に有意な企業の割合が大きいことを示している。

第 VI 部

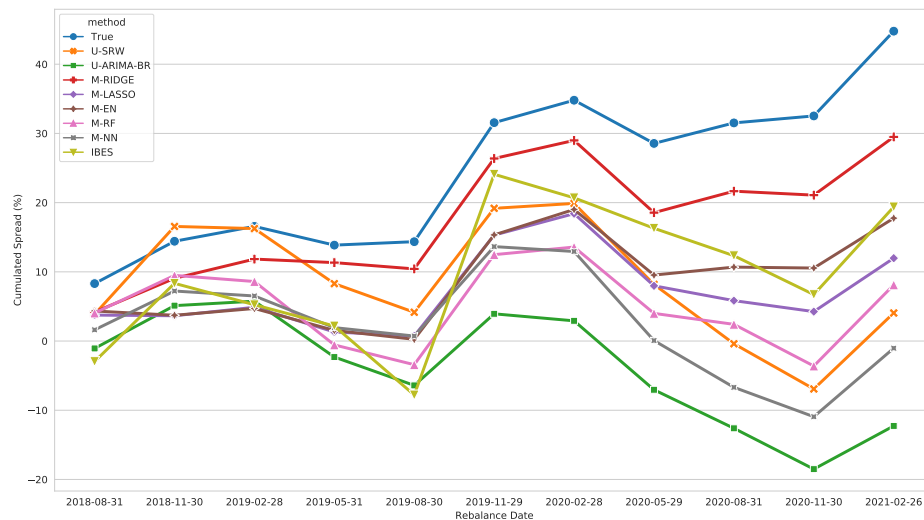
予測の投資指標としての有用性

最後に、本稿で推定した機械学習アルゴリズムによって得られた 1 期先四半期 EPS 予測の投資指標としての有用性について検証する。検証方法としては、代表的な投資指標である株価収益率を 1 期先四半期 EPS 予測値をもとに算出し、得られた株価予測収益率の高低によってロング・ショート・ポートフォリオを構築してどれくらいの超過リターンが得られるかを調べる。具体的には次のように行う。

1. 時点 $t-1$ (各四半期の中間月末) において、株価 P_{t-1} と 1 期先四半期 EPS 予測値 \hat{Y}_t を用いて株価予測収益率 $\frac{P_{t-1}}{\hat{Y}_t}$ を全企業に対して計算する。
2. 全サンプル企業 (時系列モデルは全 1,003 社、I/B/E/S 業績予想は全 88 社) の株価予測収益率を降順にソートし、倍率順に 1~5 までの 5 つのポートフォリオを作成する。
3. 高倍率企業で構成されるポートフォリオ 1 を空売りし、低倍率企業で構成されるポートフォリオ 5 を同じ額だけ買う。
4. 1 四半期 (3 ヶ月) 保有し、ポートフォリオ 1 とポートフォリオ 5 のスプレッドを観測する。

以上のような操作を毎四半期行い、ポートフォリオを 1 四半期ごとにリバランスする。図 10 は各予測手法に基づくロング・ショート・ポートフォリオ戦略の累積超過リターンを 2018 年 5 月末~2021 年 2 月末まで示

図 10: 各予測手法に基づいたロング・ショート・ポートフォリオ戦略による累積超過リターン



(注) リターン算出に用いた株価データは「日経 NEEDS-FinancialQUEST」から収集している。True は株価予測収益率の算出に 1 期先四半期 EPS の真の値を用いたときの超過リターンを示している。季節ランダムウォークモデル (U-SRW) は 1 年前の実績四半期 EPS を用いた株価予測収益率を算出しているとみなすことができる。なお、株価予測収益率が負である企業は除外してポートフォリオを構築している。

している。

まず、1 期先四半期 EPS の真の値をロング・ショート・ポートフォリオ戦略に用いると、プラスのリターンを獲得している。つまり、仮に 1 期前時点で次期四半期 EPS を完璧に予測できるのであれば高い水準の超過リターンを得ることができることを意味する。各予測手法のうち、最も高いリターンを獲得したのは多変量 Ridge 回帰であり、最終時点の累積超過リターンは約 30% となった。次点で I/B/E/S 業績予想が高いリターンを得ており、多変量 Elastic Net 回帰がそれをわずかに下回る結果となった。ここで、I/B/E/S 業績予想の超過リターンの標準偏差は 12.2 であり、他の時系列モデルの超過リターンの標準偏差 (平均 7.1) と比べると変動が大きいことがわかる。これは時系列モデルのサンプル企業が 1,003 社であるのに対し、I/B/E/S 業績予想のカバレッジが 88 社と少数であるためリターンの変動が大きくなっていると考えられる。なお、伝統的な線形時系列モデル (季節ランダムウォークモデル, Brown and Rozeff (1979) の ARIMA モデル) の超過リターンはいずれも低い水準の超過リターンであった。このように、機械学習アルゴリズム (特に多変量 Ridge 回帰) による予測には投資指標における有用性があることが確認された。

第 VII 部

終わりに

本稿では、機械学習アルゴリズムによる日本企業の四半期 EPS 予測を行い、伝統的な線形時系列モデルとの精度比較と、アナリスト予測との精度比較をそれぞれ行った。まず、伝統的な線形時系列モデルとの比較について、単変量予測ではあまり予測精度の差は見られなかったものの、多変量予測では機械学習アルゴリズムが明らかに伝統的な線形時系列モデルよりも精度の高い予測を与えた。この結果から、機械学習アルゴリズムは伝統的な線形時系列モデルでは捉えられない将来の四半期 EPS とファンダメンタル会計変数の関係性を捉えることができ、機械学習アルゴリズムを用いることで統計的・機械的な手法による日本企業の EPS 予測のパフォーマンスを向上させる可能性が示唆される。次に、アナリスト予測との比較について、多変量機械学習アルゴリズムはアナリスト予測と同等もしくはそれ以上に高い精度の予測を与えた。企業利益予測の研究分野では、長らく時系列モデルによる予測よりもアナリスト予測の方が適切であると暗黙裡に見なされてきたが、本稿の結果はその認識を覆すものとなった。さらに本稿の追加的な検証から、機械学習アルゴリズムによる四半期 EPS 予測値は投資指標としても優れており、多くの企業を予測対象とすることでより分散の効いた投資機会を得ることができると考えられる。

機械学習アルゴリズムがアナリストと同等な精度の予測を与えることに加え、統計的・機械的な手法である機械学習アルゴリズムは予測プロセスが自動化されているため低コストであり、かつ決算情報を公開する企業であれば全て予測対象とすることができる。米山 (2010) によると日本企業のアナリストカバレッジは減少している傾向にあり、特に新興市場のカバー率は全体で 10% に満たないものとなっている。日本証券業協会はこうした事態に対処するため、「アナリストの中立性に配慮しつつ、新興市場に上場する企業のアナリストカバレッジを高める施策を検討する」という提言を行った^{*15}。しかし現実の問題としてアナリストは、重要顧客である機関投資家の保有比率が高く、株式売買が盛んで、売買手数料が大きい企業のみをカバーする傾向にあり (中井, 2006)、市場で株式売買が少ない企業にアナリストの関心を引きつけることは困難で、費用対効果の問題も伴う。このような日本のアナリスト予測のカバレッジが限定的である現状に対し、機械学習アルゴリズムを用いることで、アナリストの関心を引かないような企業に対してもアナリストと同等もしくはそれ以上の精度の利益予測を低コストで提供できるであろう。

本稿の課題として、まず予測をより早い時点で実行できるようにすることが考えられる。本稿の予測のデザインは、前期四半期決算情報の開示後、その情報をもとに 1 期先四半期 EPS を予測するものとなっている。一方で、2 期先、3 期先のようにより早い時点で予測を行うことも十分に考えらる。その場合、予測に用いることができる情報集合を予測タイミング時点で利用可能なものに限定したうえで、複数期先予測の精度が保たれるのかを検証することが必要となるであろう。次に、過去の四半期 EPS の実績値やファンダメンタル会計変数以外にも予測に用いることができる情報があるのかどうかを吟味することも挙げられる。本稿の予測で用いた説明変数は、公開された対象企業自身の会計データに限定しており、1 企業内で完結しているものとなっている。しかし、企業の将来の利益を説明するものとして、マクロ要因や産業別要因、利害関係がある企業の情報なども考えられる。そのような対象企業外の情報の中から予測精度を向上させるものを模索することも有意義であろう。

機械学習アルゴリズムはハイパーパラメータ選択のプロセスやモデルの設定によって予測のパフォーマンス

^{*15} 日本証券業協会「新興市場のあり方を考える委員会報告書」参照。

が大きく変動するため、本稿で得られた結果は必ずしも最良なものではないかもしれない。それでも本稿の結果はアナリストと比べて劣らない、もしくは優れた予測を与えるものとなった。このことから、現在日本で衰退してしまっている時系列モデルによる利益予測の研究が、機械学習アルゴリズムを用いることで再興する余地が十分にあることが示唆される。今後の研究で、統計的・機械的な手法による企業利益予測を更に発展させ、あらゆるステークホルダーに利用可能な幅広い企業の予測情報が提供されることを期待したい。

参考文献

- Abarbanell, Jeffrey S and Brian J Bushee (1997) “Fundamental analysis, future earnings, and stock prices,” *Journal of accounting research*, Vol. 35, No. 1, pp. 1–24.
- Ahmed, Nesreen K, Amir F Atiya, Neamat El Gayar, and Hisham El-Shishiny (2010) “An empirical comparison of machine learning models for time series forecasting,” *Econometric Reviews*, Vol. 29, No. 5-6, pp. 594–621.
- Bates, John M and Clive WJ Granger (1969) “The combination of forecasts,” *Journal of the Operational Research Society*, Vol. 20, No. 4, pp. 451–468.
- Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung (2015) *Time series analysis: forecasting and control*.
- Breiman, Leo (2001) “Random forests,” *Machine learning*, Vol. 45, No. 1, pp. 5–32.
- Brown, Lawrence D and Michael S Rozeff (1979) “Univariate time-series models of quarterly accounting earnings per share: A proposed model,” *Journal of Accounting Research*, Vol. 17, No. 1, pp. 179–189.
- Callen, Jeffrey L, Clarence CY Kwan, Patrick CY Yip, and Yufei Yuan (1996) “Neural network forecasting of quarterly accounting earnings,” *International Journal of Forecasting*, Vol. 12, No. 4, pp. 475–482.
- Cao, Kai and Haifeng You (2020) “Fundamental Analysis Via Machine Learning,” Technical report, Working Paper.
- Cao, Qing and Qiwei Gan (2009) “Forecasting EPS of Chinese Listed Companies Using Neural Network with Genetic Algorithm,” *15th Americas Conference on Information Systems 2009, AMCIS 2009*, Vol. 5, pp. 2971–2981.
- Chatfield, Chris et al. (1993) “Neural networks: Forecasting breakthrough or passing fad?” *International Journal of Forecasting*, Vol. 9, No. 1, pp. 1–3.
- Diebold, Francis X and Robert S Mariano (2002) “Comparing predictive accuracy,” *Journal of Business & economic statistics*, Vol. 20, No. 1, pp. 134–144.
- Etemadi, Hossein, Ahmad Ahmadpour, and Seyed Mohammad Moshashaei (2015) “Earnings per share forecast using extracted rules from trained neural network by genetic algorithm,” *Computational Economics*, Vol. 46, No. 1, pp. 55–63.
- Foster, George (1977) “Quarterly accounting data: Time-series properties and predictive-ability results,” *Accounting Review*, Vol. 52, No. 1, pp. 1–21.
- Griffin, Paul A (1977) “The time-series behavior of quarterly earnings: preliminary evidence,” *Journal of Accounting Research*, pp. 71–83.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009) *The Elements of Statistical Learning*:

- Data Mining, Inference, and Prediction Second Edition*, New York, NY.
- Hill, Tim, Leorey Marquez, Marcus O'Connor, and William Remus (1994) "Artificial neural network models for forecasting and decision making," *International journal of forecasting*, Vol. 10, No. 1, pp. 5–15.
- Hill, Tim, Marcus O'Connor, and William Remus (1996) "Neural network models for time series forecasts," *Management science*, Vol. 42, No. 7, pp. 1082–1092.
- Hoerl, Arthur E and Robert W Kennard (1970) "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, Vol. 12, No. 1, pp. 55–67.
- Hopwood, William S and James C McKeown (1986) *Univariate time-series analysis of quarterly earnings: Some unresolved issues*, Sarasota, FL.
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989) "Multilayer feedforward networks are universal approximators," *Neural networks*, Vol. 2, No. 5, pp. 359–366.
- Lev, Baruch and S Ramu Thiagarajan (1993) "Fundamental information analysis," *Journal of Accounting research*, Vol. 31, No. 2, pp. 190–215.
- Lorek, Kenneth S and G Lee Willinger "A multivariate time-series prediction model for cash-flow data," *Accounting Review*, Vol. 71.
- Qi, Min (1999) "Nonlinear predictability of stock returns using financial and economic variables," *Journal of Business & Economic Statistics*, Vol. 17, No. 4, pp. 419–429.
- Tibshirani, Robert (1996) "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 1, pp. 267–288.
- Tkáč, Michal and Robert Verner (2016) "Artificial neural networks in business: Two decades of research," *Applied Soft Computing*, Vol. 38, pp. 788–804.
- Zhang, Wei, Qing Cao, and Marc J Schniederjans (2004) "Neural network earnings per share forecasting models: A comparative analysis of alternative methods," *Decision Sciences*, Vol. 35, No. 2, pp. 205–237.
- Zou, Hui and Trevor Hastie (2005) "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, Vol. 67, No. 2, pp. 301–320.
- 桜井久勝 (1990) 「会計利益の時系列特性と利益予測」, 『経営学・会計学・商学研究年報 / 神戸大学大学院経営学研究科編』, 第 36 号, 45–98 頁.
- 太田浩司 (2006) 「経営者予想に関する日米の研究: 文献サーベイ」, 『武蔵大学論集』, 第 54 巻, 第 1 号, 53–94 頁.
- 中井誠司 (2006) 「アナリストカバレッジと機関投資家保有」, 『証券アナリストジャーナル』, 第 44 巻, 第 6 号, 84–97 頁, 6 月.
- 米山徹幸 (2010) 「アナリストカバレッジの現状と課題」, 『政策科学学会年報』, 第 1 号, 40–47 頁.