

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

2020 年月日

目次

第 1 章	はじめに	2
第 2 章	過去の研究	5
第 3 章	データ	9
3.1	サンプルデータ	9
3.2	目的変数	10
3.3	説明変数	10
第 4 章	予測モデル	12
4.1	予測のデザイン	12
4.2	伝統的時系列モデル	13
4.3	機械学習モデル	16
4.4	ハイパーパラメータの選択	21
第 5 章	予測精度指標	22
第 6 章	予測の精度と結果	23
第 7 章	予測の投資指標としての有用性	24
第 8 章	終わりに	25

第 1 章

はじめに

1 株当たり利益 (Earnings Per Share: EPS) は株式 1 つ当たりに対する企業の当期純利益を表した指標である。EPS は企業の当期純利益を発行株式数で割ることで求められており、規模に依存しない企業の収益性を示す。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ことである。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機械化できるため、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 & 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 (e.g., 英語のもっと古い論文)。一方、近年の研究では、「ファンダメンタル会計変数」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 & Thiagarajan 1993, Abarbanell & 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 & 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は, 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 & 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の非線形を捉え, 高い精度の 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企業の EPS 予測の研究の前例として, 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 & 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 ANN モデルによる四半期 EPS の予測を行っている. また, Cao & 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もたらす結果となっている.

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て, モデルに多変量と非線形な特性を反映させ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とされている. 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による日本企業データを用いた 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できていたため, 四半期データを用いた EPS 予測の研究もない.

そこで本稿は, 日本企業データにおいて代替的な時系列予測モデルである機械学習アルゴリズムが, 統計的・機械的な手法による 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もたらすのかを検証する. 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. さらに, 機械学習アルゴリズムによる予測と各手法による予測の組み合わせ予測 (Bates & Granger 1969) を求めて, 機械学習アルゴリズムによる予測に他の予測手法の精度を向上させる情報を含んでいるかどうかを調べる. 追加的な検証として, 機械学習アルゴリズム

ムによる予測値と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と各手法による予測の投資指標としての有用性を考察する。

本稿の構成は以下のとおりである。第 2 章では、、

第2章

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モデルによってうまく描写され、四半期 EPS は Brown & 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たとしている。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としている。太田 (2006) では、日本の利益予測に関する研究の数が非常に少ないとしつつも、アメリカと同様な比較結果を得ているとしている。そして、時系列モデルによる予測に関しては、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 EPS 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経営者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予測よりも精度が高いといえる」と述べ、モデルの改善により統計的・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つである人工ニューラルネットワーク (Artificial Neural Network: ANN) モデルが挙げられる。ANN モデルは生物の神経細胞 (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 & Verner 2016)。特に、Hill et al. (1996) は、月次や四半期の時系列予測において ANN モデルの方が伝統的な統計モデルよりも正確な予測を与え、時系列予測における ANN モデルの有用性を示している。その理由として、ANN 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、伝統的な線形統計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、ANN モデル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も確認されている (Hopwood & 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 ANN モデルと従来の統計的手法である ARIMA モデルとの精度を比較して ANN の予測精度を検証したが、結果は予想に反して ANN モデルによる予測が ARIMA モデルによる予測よりも精度が低かった。そして、この研究では ANN の予測精度は文脈依存であると結論付けている。(Hill のディスポイント)

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 & 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してい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持ち、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 & 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 ANN モデルによる四半期 EPS 予測の研究を発展させる形で、ANN モデルにファンダメンタル会計変数を反映させた多変量 ANN モデルによる四半期 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 ANN モデル、多変量 ANN モデルの 4 つの精度比較を行い、結果として、ANN モデルは線形モデルよりも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

ルで EPS 予測における ANN モデルの有用性を示した研究が行われている。Cao & 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 ANN モデルによる 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 予測の研究では機械学習アルゴリズムの 1 つとして ANN モデルが盛んに用いられているが、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 ANN モデルのみに着目している傾向がある (Ahmed et al. 2010)。一方、ANN モデル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・フォレスト回帰モデルが挙げられる (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・フォレスト回帰モデル (Breiman 2001) も、ANN モデル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 & You (2020) は年次ではあるものの、これらの代表的な機械学習アルゴリズムを用いた EPS 予測をしており、ベンチマークであるランダムウォーク (Random Walk: RW) モデルよりも精度の高い予測を与える結果となっている。

(Hill et al の ANN ディスも少し入れる)

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 (ANN モデル、罰則回帰モデル、ランダム・フォレスト回帰モデル) 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 の予測を行い、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 3 章

データ

3.1 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)
- 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

^{*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している。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。

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、1,020 社、延べ 1,020 社 × 52 四半期 = 53,040 個の企業四半期が最終サンプルである。図表はサンプル企業の所属業種の割合を示している。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。

3.2 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。

$$\begin{aligned} Y &= \text{四半期 1 株あたり利益} \\ &= \frac{\text{普通株式に係る四半期当期純利益}}{\text{普通株式の期中加重平均株式数}} \\ &= \frac{\text{損益計算書上の四半期当期純利益} - \text{普通株式に帰属しない金額}}{\text{普通株式の期中加重平均発行済株式数} - \text{普通株式の期中加重平均自己株式数}} \end{aligned} \quad (3.1)$$

3.3 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 & Thiagarajan (1993) や Abarbanell & 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を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。

^{*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。

$$\begin{aligned}
INV &= \frac{\text{棚卸資産 (円)}}{\text{普通株式の期中加重平均株式数}} \\
AR &= \frac{\text{売掛金 (円)}}{\text{普通株式の期中加重平均株式数}} \\
CAPX &= \frac{\text{資本的支出 (円)}}{\text{普通株式の期中加重平均株式数}} \\
GM &= \frac{\text{売上総利益 (円)}}{\text{普通株式の期中加重平均株式数}} \\
SA &= \frac{\text{販売費及び一般管理費 (円)}}{\text{普通株式の期中加重平均株式数}} \\
ETR &= \frac{\text{法人税 (円)}}{\text{税引前利益 (円)}} \\
LF &= \log \left(\frac{\text{売上高 (円)}}{\text{従業員数 (人)}} \right)
\end{aligned} \tag{3.2}$$

多変量モデルによる予測では四半期 EPS とファンダメンタル会計変数の 4 期までのラグを説明変数として用いる。図表は予測に用いた変数の記述統計、及び相関行列を示している。

第 4 章

予測モデル

4.1 予測のデザイン

本稿では、サンプル企業 1,020 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^{*1}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、訓練データを用いてモデルを推定し、テストデータの説明変数を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し、1 企業につき 12 期分の 1 四半期先予測値系列

^{*1} 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 $52 - \text{ラグの次数}$ 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。

$(\hat{y}_{2018Q1}, \hat{y}_{2018Q2}, \dots, \hat{y}_{2018Q2})$ が得られる。この予測プロセスを全サンプル企業 1,020 社に対して行うことで、最終的に $12 \text{ 期} \times 1,020 \text{ 社} = 12,240 \text{ 個}$ の四半期 EPS 予測値が各予測手法から算出される。

図表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。

4.2 伝統的時系列モデル

4.2.1 ランダムウォークモデル

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モデルは次のとおりである。

$$\begin{aligned} Y_t &= Y_{t-1} + \epsilon_t \\ E[Y_t] &= Y_{t-1} \end{aligned} \quad (4.1)$$

ただし、 ϵ_t は全ての時点 t において

$$\begin{aligned} E[\epsilon_t] &= 0 \\ E[\epsilon_t \epsilon_{t-k}] &= \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \end{aligned} \quad (4.2)$$

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モデルが次のとおりである。

$$\begin{aligned} Y_t &= Y_{t-4} + \epsilon_t \\ E[Y_t] &= Y_{t-4} \end{aligned} \quad (4.3)$$

ランダムウォークモデルは前四半期の実績値を、季節ランダムウォークモデル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モデル及び季節ランダムウォークモデルを他の予測に対するベンチマークとして用いる。

4.2.2 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)^{*2}、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。^{*3}

$$\phi_p(B)\Phi_P(B)(1-B)^d(1-B^s)^DY_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \quad (4.4)$$

ただし、

$$\begin{aligned} p &: \text{トレンドの自己回帰過程の階数} \\ d &: \text{トレンド階差の次数} \\ q &: \text{トレンドの移動平均過程の階数} \\ P &: \text{季節変動の自己回帰過程の階数} \\ D &: \text{季節階差の次数} \\ Q &: \text{季節変動の移動平均過程の階数} \\ s &: \text{季節変動の周期} \end{aligned} \quad (4.5)$$

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

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 & 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。^{*4}

^{*2} ARIMA モデルは提唱者の名前から Box-Jenkins モデルとも呼ばれる。

^{*3} 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、通常に取り扱いに従って 0 としている。

^{*4} 先行研究では、企業ごとにモデルを構築するよりも、企業で共通の構築で予測したほうが全体として予測のパフォーマンスが良いと示されている。

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

$$\begin{aligned} Y_t &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) + \epsilon_t \\ E[Y_t] &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) + \delta \end{aligned} \quad (4.6)$$

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

$$\begin{aligned} Y_t &= Y_{t-4} + (Y_{t-1} - Y_{t-5}) - \theta_1 a_{t-1} - \Theta_1 a_{t-4} + \theta_1 \Theta_1 a_{t-5} + \epsilon_t \\ E[Y_t] &= Y_{t-4} + (Y_{t-1} - Y_{t-5}) - \theta_1 a_{t-1} - \Theta_1 a_{t-4} + \theta_1 \Theta_1 a_{t-5} + \delta \end{aligned} \quad (4.7)$$

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

$$\begin{aligned} Y_t &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) - \Theta_1 a_{t-4} + \epsilon_t \\ E[Y_t] &= Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) - \Theta_1 a_{t-4} + \delta \end{aligned} \quad (4.8)$$

ただし、 δ は SARIMA モデルの定数項。

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。

4.2.3 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 & Thiagarajan (1993)、Abarbanell & Bushee (1997)、Lorek & Willinger (n.d.) の研究で用いられているモデルを基に、以下のような多変量線形回帰モデルを考える。

$$\begin{aligned} E[Y_t] &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-4} + \beta_3 INV_{t-1} + \beta_4 AR_{t-1} + \beta_5 CAPX_{t-1} \\ &\quad + \beta_6 GM_{t-1} + \beta_7 SA_{t-1} + \beta_8 ETR_{t-1} + \beta_9 LF_{t-1} \end{aligned} \quad (4.9)$$

$$\begin{aligned} E[Y_t] &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-4} + \beta_3 INV_{t-4} + \beta_4 AR_{t-4} + \beta_5 CAPX_{t-4} \\ &\quad + \beta_6 GM_{t-4} + \beta_7 SA_{t-4} + \beta_8 ETR_{t-4} + \beta_9 LF_{t-4} \end{aligned} \quad (4.10)$$

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、(4.9) 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、(4.10) 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。

$$E[Y_t] = \beta_0 + \sum_{\tau=1}^4 (\beta_{1\tau}Y_{t-\tau} + \beta_{2\tau}INV_{t-\tau} + \beta_{3\tau}AR_{t-\tau} + \beta_{4\tau}CAPX_{t-\tau} + \beta_{5\tau}GM_{t-\tau} + \beta_{6\tau}SA_{t-\tau} + \beta_{7\tau}ETR_{t-\tau} + \beta_{8\tau}LF_{t-\tau}) \quad (4.11)$$

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。

4.3 機械学習モデル

ここでは、本稿で用いる機械学習的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 \mathbf{X}_t について、単変量モデルでは、

$$\begin{aligned} \mathbf{X}_t &= (X_{t1}, \dots, X_{t4}) \\ &= (Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}) \end{aligned} \quad (4.12)$$

多変量モデルでは、

$$\begin{aligned} \mathbf{X}_t &= (X_{t1}, \dots, X_{t32}) \\ &= (Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, \\ &\quad INV_{t-1}, INV_{t-2}, INV_{t-3}, INV_{t-4}, \\ &\quad AR_{t-1}, AR_{t-2}, AR_{t-3}, AR_{t-4}, \\ &\quad CAPX_{t-1}, CAPX_{t-2}, CAPX_{t-3}, CAPX_{t-4}, \\ &\quad GM_{t-1}, GM_{t-2}, GM_{t-3}, GM_{t-4}, \\ &\quad SA_{t-1}, SA_{t-2}, SA_{t-3}, SA_{t-4}, \\ &\quad ETR_{t-1}, ETR_{t-2}, ETR_{t-3}, ETR_{t-4}, \\ &\quad LF_{t-1}, LF_{t-2}, LF_{t-3}, LF_{t-4}) \end{aligned} \quad (4.13)$$

としてモデルを構築する。

4.3.1 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、in-sample での予測精度は高くても、out-of-sample 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 in-sample データの規則性のみを捉えてしまい、out-of-sample 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。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 回帰モデル、LASSO 回帰モデル、Elastic Net 回帰モデルの 3 つを用いる。

Ridge 回帰モデルの係数の推定式は次のとおりである。

$$\hat{\beta}^{Ridge} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{j=1}^k \beta_j X_{tj} \right)^2 + \lambda \sum_{j=1}^k \beta_j^2 \right\} \quad (4.14)$$

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り、ハイパーパラメータの選択方法の詳細については 4.4 節で説明する。(4.14) 式で示されるように、Ridge 回帰モデル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 回帰モデル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) 回帰モデルの係数の推定式は次のとおりである。

$$\hat{\beta}^{LASSO} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{j=1}^k \beta_j X_{tj} \right)^2 + \lambda \sum_{j=1}^k |\beta_j| \right\} \quad (4.15)$$

LASSO 回帰モデルは、Ridge 回帰モデルと同様、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 回帰モデルとの違いである。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 回帰モデルを推定することで変数選択が行われる。

Elastic Net 回帰モデルの係数の推定式は次のとおりである。

$$\hat{\beta}^{EN} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{j=1}^k \beta_j X_{tj} \right)^2 + \lambda \left\{ (1 - \alpha) \sum_{j=1}^k \beta_j^2 + \alpha \sum_{j=1}^k |\beta_j| \right\} \right\} \quad (4.16)$$

Elastic Net 回帰モデルは、Ridge 回帰モデルと LASSO 回帰モデル両方の罰則項を利用する構造となっている。 α は LASSO 回帰モデル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、罰則の強さの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。

$$\begin{aligned} E[Y_t]^{Ridge} &= \hat{y}_t^{Ridge} = \beta_0^{Ridge} + \sum_{j=1}^k \beta_j^{Ridge} x_{tj} \\ E[Y_t]^{LASSO} &= \hat{y}_t^{LASSO} = \beta_0^{LASSO} + \sum_{j=1}^k \beta_j^{LASSO} x_{tj} \\ E[Y_t]^{EN} &= \hat{y}_t^{EN} = \beta_0^{EN} + \sum_{j=1}^k \beta_j^{EN} x_{tj} \end{aligned} \quad (4.17)$$

4.3.2 ランダム・フォレスト回帰モデル

$$(4.18)$$

4.3.3 人工ニューラルネットワークモデル

人工ニューラルネットワーク (Artificial Neural Network: ANN) モデルについて述べる。まず ANN の構造の最小単位であるユニット (unit) またはニューロン (neuron) につ

いて考える。 $X = (X_1, \dots, X_k)$ を入力ベクトルとする。 (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 (weight) であり、スカラー値 b はバイアス (bias) という。ユニットではまず始めに $X = (X_1, \dots, X_k)$ と (w_1, \dots, w_k) の内積と b の和であるプレアクティベーション (pre-activation) z を計算する。

$$z = w_1 X_1 + \dots + w_k X_k + b \quad (4.19)$$

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

$$a = \phi(z) \quad (4.20)$$

以上が ANN のユニットの構造であり、まとめると以下の式で表される。

$$a = \phi(w_1 X_1 + \dots + w_k X_k + b) \quad (4.21)$$

このユニットを複数組み合わせることで、ANN が構成される。

ANN と一口に言っても様々な構造をもつ多くの種類がある。ANN 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: FFNN) は ANN のなかでも基本的な構造を持ち、全てのユニットが結合した無閉路グラフ (acyclic graph) で表現されすべての計算が逐次的に行われる。FFNN 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、図表のように表される。

MLP の特徴として、横のつながりのユニット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ユニットが次の層に含まれる全てのユニットとそれぞれ結合している (fully connected)。ネットワークの最初の層は入力層 (input layer) 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 (output layer) 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 (hidden layer) 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_1, \dots, H_L) とする (activation vector)。ただし、 L は隠れ層に含まれるユニットの数である。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。

$$\begin{aligned} H_1 &= \phi^{(1)}(w_{11}^{(1)} X_1 + w_{21}^{(1)} X_2 + \dots + w_{K1}^{(1)} X_K + b_1^{(1)}) \\ H_2 &= \phi^{(1)}(w_{12}^{(1)} X_1 + w_{22}^{(1)} X_2 + \dots + w_{K2}^{(1)} X_K + b_2^{(1)}) \\ &\vdots \\ H_L &= \phi^{(1)}(w_{1L}^{(1)} X_1 + w_{2L}^{(1)} X_2 + \dots + w_{KL}^{(1)} X_K + b_L^{(1)}) \end{aligned} \quad (4.22)$$

次に、出力層のについて考える。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。

$$Y = \phi^{(2)}(w_1^{(2)} H_1 + w_2^{(2)} H_2 + \dots + w_L^{(2)} H_L + b^{(2)}) \quad (4.23)$$

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) は隠れ層のユニット数が十分であれば、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 ANN モデルを用いる。

$$Y_t = f(\mathbf{X}, \alpha, \beta) = \sum_{l=1}^L (\alpha_0 + \alpha_l H_l) = \sum_{l=1}^L \alpha_l \text{logsig} \left(\sum_{k=1}^K (\beta_{0l} + \beta_{kl} X_k) \right) \quad (4.24)$$

ただし、

$$\begin{aligned}
& Y_t : \text{出力 (目的変数)} \\
& \mathbf{X} = (X_1, \dots, X_K) : \text{入力 (説明変数) ベクトル} \\
& X_k : k \text{ 番目の入力 (説明変数)} \\
& \alpha_l : \text{隠れ層の } l \text{ 番目のユニットと目的変数のウェイト (パラメータ)} \\
& L : \text{隠れ層のユニット数} \\
& \alpha_0 : \text{出力層のバイアス (定数項)} \\
& \beta_{kl} : k \text{ 番目の入力 (説明変数) と隠れ層の } l \text{ 番目のユニットのウェイト (パラメータ)} \\
& \beta_{0l} : \text{隠れ層の } l \text{ 番目のユニットのバイアス (定数項)} \\
& \text{logsig}(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)} : \text{ロジスティック関数}
\end{aligned}
\tag{4.25}$$

ANN モデルの推定は誤差逆伝播法 (Backward propagation algorithm) を用いてウェイトの勾配を計算し、勾配降下によって行う。なお、モデルの複雑さを調節する隠れ層のユニット数とウェイトの更新量の幅を調節する学習率をハイパーパラメータとして、検証データによって決定する。

4.4 ハイパーパラメータの選択

Cao & You (2020), 予測がうまい代わりに複雑さの調節が必要。

本稿ではグリッドサーチによって機械学習モデルのハイパーパラメータを選択する。

第 5 章

予測精度指標

第 6 章

予測の精度と結果

フリードマン? DM 検定

桜井 1990 の第 4 四半期が最も不正確なのと整合的

第 7 章

予測の投資指標としての有用性

第 8 章

終わりに

参考文献

- Abarbanell, J. S. & Bushee, B. J. (1997), ‘Fundamental analysis, future earnings, and stock prices’, *Journal of accounting research* **35**(1), 1–24.
- Ahmed, N. K., Atiya, A. F., Gayar, N. E. & El-Shishiny, H. (2010), ‘An empirical comparison of machine learning models for time series forecasting’, *Econometric Reviews* **29**(5-6), 594–621.
- Bates, J. M. & Granger, C. W. (1969), ‘The combination of forecasts’, *Journal of the Operational Research Society* **20**(4), 451–468.
- Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015), *Time series analysis: forecasting and control*.
- Breiman, L. (2001), ‘Random forests’, *Machine learning* **45**(1), 5–32.
- Brown, L. D. & Rozeff, M. S. (1979), ‘Univariate time-series models of quarterly accounting earnings per share: A proposed model’, *Journal of Accounting Research* **17**(1), 179–189.
- Callen, J. L., Kwan, C. C., Yip, P. C. & Yuan, Y. (1996), ‘Neural network forecasting of quarterly accounting earnings’, *International Journal of Forecasting* **12**(4), 475–482.
- Cao, K. & You, H. (2020), Fundamental analysis via machine learning, Technical report, Working Paper.
- Cao, Q. & Gan, Q. (2009), ‘Forecasting eps of chinese listed companies using neural network with genetic algorithm’, *15th Americas Conference on Information Systems 2009, AMCIS 2009* **5**, 2971–2981.
- Etemadi, H., Ahmadpour, A. & Moshashaei, S. M. (2015), ‘Earnings per share forecast using extracted rules from trained neural network by genetic algorithm’, *Computational Economics* **46**(1), 55–63.

- Foster, G. (1977), 'Quarterly accounting data: Time-series properties and predictive-ability results', *Accounting Review* **52**(1), 1–21.
- Griffin, P. A. (1977), 'The time-series behavior of quarterly earnings: preliminary evidence', *Journal of Accounting Research* pp. 71–83.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction Second Edition*, New York, NY.
- Hill, T., Marquez, L., O'Connor, M. & Remus, W. (1994), 'Artificial neural network models for forecasting and decision making', *International journal of forecasting* **10**(1), 5–15.
- Hill, T., O'Connor, M. & Remus, W. (1996), 'Neural network models for time series forecasts', *Management science* **42**(7), 1082–1092.
- Hopwood, W. S. & McKeown, J. C. (1986), *Univariate time-series analysis of quarterly earnings: Some unresolved issues*, Sarasota, FL.
- Hornik, K., Stinchcombe, M. & White, H. (1989), 'Multilayer feedforward networks are universal approximators', *Neural networks* **2**(5), 359–366.
- Lev, B. & Thiagarajan, S. R. (1993), 'Fundamental information analysis', *Journal of Accounting research* **31**(2), 190–215.
- Lorek, K. S. & Willinger, G. L. (n.d.), 'A multivariate time-series prediction model for cash-flow data', *Accounting Review* **71**.
- Qi, M. (1999), 'Nonlinear predictability of stock returns using financial and economic variables', *Journal of Business & Economic Statistics* **17**(4), 419–429.
- Tkáč, M. & Verner, R. (2016), 'Artificial neural networks in business: Two decades of research', *Applied Soft Computing* **38**, 788–804.
- Zhang, W., Cao, Q. & Schniederjans, M. J. (2004), 'Neural network earnings per share forecasting models: A comparative analysis of alternative methods', *Decision Sciences* **35**(2), 205–237.
- 桜井, 久. (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 (36), 45–98.
- 太田, 浩. (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 **54**(1), 53–94.