



HUB-FS Working Paper Series

FS-2025-J-001

POS データによる企業業績予測可能性の検証

大城 真太郎

一橋大学大学院経営管理研究科

横内 大介

一橋大学大学院経営管理研究科

First version: 2025 年 7 月 29 日

All the papers in this Discussion Paper Series are presented in the draft form. The papers are not intended to circulate to many and unspecified persons. For that reason any paper can not be reproduced or redistributed without the authors' written consent.

POS データによる企業業績予測可能性の検証

大城真太郎* 横内大介†

2025年7月29日

概要

本研究では、POS データを用いてメーカーの売上高の予測可能性を検証する。POS データは店頭での製品売上を記録したものであるため、POS データからメーカーの売上高予測が可能だと考えられる。

POS データはこれまで研究や実務の様々な場面で活用されてきた。在庫管理をはじめとした小売店での業務効率化、販促や価格戦略の効果検証、消費や物価等のマクロ経済指標の推計はよく知られたユースケースである。前述の通り POS データは、個社の売上高予測にも大いに活用できそうだが、実務面での活用例は伺えるものの、我々の調べる範囲において研究例は見当たらない。

そこで本研究では、POS データを用いて個社の売上高予測を試みる。関連する先行研究として Ishikawa et al. (2016) は、米国日用品メーカーの四半期実績売上高の成長率は、米国スーパーマーケットで観測される POS データの成長率と高く相関することを報告した。同報告では売上高予測の可能性の示唆に留まっているため、本研究では実際に POS データを用いて実績売上高の予測を行う。また同報告では POS データを全て使用しているが、POS データには様々な誤差要因が存在し、サンプリングバイアスも大きいため、本研究では POS データの選別を行った後に売上高予測を行うことで、その効果も併せて検証する。さらに株式投資への応用可能性も検証する。

* 一橋大学 大学院経営管理研究科, Email:bd21f001@g.hit-u.ac.jp

† 一橋大学 大学院経営管理研究科, Email:yokouchi@hub.hit-u.ac.jp

1 はじめに

Post-Earnings Announcement Drift(PEAD) は多くの株式市場で見られるアノマリーであり、企業業績が市場予想を上回った銘柄は決算発表前後において長期に正の超過リターンを示し、下回った銘柄は負の超過リターンを示す現象である (Ball and Brown, 1968, Bernard and Thomas, 1989). そのため、市場予想を上回る精度の企業業績の予測は、アクティブ戦略による運用者が超過リターンを得るための源泉となっており、より精度の高い業績予想手法が求められている。

企業の決算発表は四半期ごとに行われており、通常は終了期の企業業績と将来業績の会社予想値が公表される。Williams (1996) や太田 (2007) によれば、証券アナリストや投資家は主にこれらの決算発表情報に基づいて将来の業績を予想している。

一方で、Choi and Ziebart (2004) は、公表される会社予想値には企業属性等を背景とした構造的なバイアスが存在すると指摘している。具体的には、企業規模が大きいほど経営陣の計画達成への意識が高まる傾向が強いと考え、その結果、悲観的な予想値を出す傾向にあると述べている。また、高成長企業でも同様に、計画未達を回避すべく悲観的な予想値を出す傾向にあると述べている。

決算発表日は決算期末日から1ヵ月前後遅れることから、会社予想値の構造的バイアスに加えて、企業決算発表以外の情報を組み合わせた業績予想も試みられている。一般的には、独自に実施する企業取材で経営者との議論を通じて獲得する情報や、独自のリサーチから得られた情報を組み合わせて業績予想を作成することが知られている (Hassell and Jennings, 1986)。企業取材以外の情報源としては、企業が決算発表以外で公表する業績数値、例えば月次売上や出退店数、契約社数等が知られている。また、証券アナリストも投資家と同様のプロセスを通じて業績予想を作成・公表しているので、投資家は証券アナリストレポートや証券アナリストとの議論を通じて業績に関する情報を取得し、独自情報と併せて業績予想や投資判断を行っている。この一連の運用手法は、総称してボトムアップ運用と呼ばれており、アナリストによって個別企業ごとに行われている。

アクティブ運用のうち、もう一つの代表的な運用手法としてクオンツ運用が挙げられる。運用者は独自の株価予測モデルを構築し、それに基づいて投資判断を行うことが一般的であり、とりわけファクター投資は古くから広く行われている。銘柄ごとに PER や PBR、株価騰落率といったファクター値を付与し、過去データをもとに株価予測に有効なファクターや合成ファクターを見出して、それをもとに運用者が求める特性を持ったポートフォリオを構築する。このように人手やコストをかけて行われるアクティブ運用だが、Carhart (1997) は株式リターンの持続性は共通ファクターと運用コストでほぼ説明可能であると述べている。また、国内外のアクティブファンドは全体としてインデックス対比でパフォーマンスが劣後しているとも報告されており (SP グローバル, 2023), 継続的に超過収益を得ることは難しいと考えられている。加えて、金融庁は、2015 年から「フェア・ディスクロージャー規制」を敷いて決算発表前の企業取材を禁止し、また未発表の決算内容についての質問も禁止した。それにより取材に基づく将来業績の予測が従前より困難となった。

欧州では *MiFID* の施行に伴い、投資家は証券アナリストによるリサーチ情報自体に対価を払う必要があるため、多くの投資家はセルサイドからの情報取得を制限することとなった (FSA,

2006, 菅野, 2019). 将来, 日本でも *MiFID* が施行されれば投資家の情報が大きく制限されることとなり, アナリストによる取材を基本としたボトムアップ運用にとって大きな逆風となる. また, アナリスト取材が制限されればアナリスト予想指標の有効性が薄れることとなり, それを利用しているクオンツ運用も影響は免れない. そのため近年では, 上記のような従来型の調査や情報に加えて, オルタナティブデータを活用した企業業績予測や資産価格予測, 株式投資が試みられており, その規模は年々拡大している (Alternative Data.org). 例えばクオンツ領域では, オンラインで収集したテキスト情報を評価し, 将来株価の予測可能性が検証されている. Tetlock (2007) は, ウォールストリートジャーナルのコラムから悲観度を表す指標を作成し, 悲観度が強い場合には短期的に株式市場に下方圧力がかかり, その後株価はファンダメンタルズに回帰すること, また悲観度が極端に高いもしくは低い値を示すと市場の取引量が増加することを報告した. 日本語テキストでは, 竹中・浦野 (2021) は, 予測対象企業のネットニュースを自然言語処理によりポジティブネガティブ判定し, その企業の株価・財務指標を組み合わせた指標は, ニューラルネットワークモデルを用いると翌日の始値予測に対して有効であることを報告した. このオンラインテキスト情報は, 資産運用の実務においても広く応用され始めている. BlackRock 社は, 宿泊や交通, レジャー等の予約件数をオンラインテキストの中から収集し, 決算発表前に個別企業の業績予測の情報源として活用したり, GoldmanSachsAssetManagement 社は, 同じくオンラインテキストである特許情報から技術開発への取り組みを推測して同じく業績予想に活用している (日本経済新聞社, 2019). 他のオルタナティブデータでは, クレジットカードの決済データや POS データのような消費関連データに注目が集まっている. これらのデータは, センチメントではなく実消費額という数値データであり, そのデータ形式も構造化されていることから, 比較的扱いやすいデータとなっており, 我が国においても研究されはじめている (鈴木・森, 2023). しかし次節でも述べるが, 研究領域において, 消費関連データはマクロ経済指標予測への活用が大半であり, 個別企業の業績予測やそれに基づく個別株式投資戦略については, 実務では活用例が見られるが, 我々の調べた限りは研究例はほぼ見当たらない. そこで本研究では, POS データを用いたメーカーの売上高の予測可能性を検証する. POS データは, 店頭での製品売上を記録したものであるから, その店頭の製品売上の合計はメーカーの実績売上高と連動する可能性が高いと予想する. 特に, 国内小売店舗における売上が業績の大部分を占めるメーカーは, その傾向が強いことが予想されることから, 我々は POS データを用いた国内メーカーの売上高予測可能性は十分高いと考えた.

本論文の構成は以下の通りである. 2 節では, 学術面や実務面から, POS データ利活用のレビューを行う. 3 節では, 本研究で使用する POS データの概要説明と, バイアスを確認する. 4 節は, POS データによるメーカーの売上高の予測可能性を検証する. メーカーごとの予測可能性や, 最適な予測手法の比較検討を行う. また, 業績予測の検証過程で得られた POS データ取扱上の留意点も併せて解説する. 5 節にて, まとめと今後の展望を述べる.

2 POS データと利活用

POS データとは、スキャナーを用いて商品バーコードを会計時に読み取ることで記録した、購買日時・購買小売店・購買商品・購買個数・購買金額等の商品販売情報のことである。商品は JAN コードにより識別され、JAN コードは 13 桁もしくは 8 桁の数字で構成されており、主に事業者コードと商品コードから成る。また購入者情報も含むデータは ID-POS と呼ばれ、購入者に紐づくポイントカード等から記録される。スーパーマーケットやコンビニエンスストア等のほぼ全ての小売店舗には、POS システムが導入されている。POS データは、店舗売上や販売商品に関して網羅的かつリアルタイムな情報を含むため、多方面でその活用法が検討され、実際に活用されてきた。ここでは、先行研究や事例を紹介する。

流通現場では、購買商品のバーコードを読み取ることで出庫数量を記録し、商品入荷の際には商品や段ボールのバーコードを読み取ることで入庫数量を記録すれば、在庫推移がリアルタイムで把握可能となる。在庫推移から適切な発注タイミングや発注量が決まるため、店舗や卸売業者は適正在庫の確保や効率的な物流が可能となり、メーカーやサプライヤーは的確な生産計画を立てることができる。在庫関連以外にも、売れ筋品を素早く発見したり、返品交換等の商品追跡が容易となるなど、流通現場での活用範囲は多岐に渡る。日本では、1982 年にセブンイレブン社が初めて POS システムを導入して効率的な受発注管理を行ったことを皮切りに、小売業での導入が進み、広く普及している(阿部, 2005, 流通システム開発センター, 2012)。Weber and Kantamneni (2002) によると、POS データを活用した在庫管理やコスト削減、顧客の商品需要への柔軟な対応方法等、数多く活用例が報告されている。マーケティング分野においては、被説明変数を特定商品やブランド等の販売量、説明変数を他商品の広告量や販促量、過去の販売量等としてその関係を明らかにしている。また、特定商品等の販売量を最大化するために販売店やメーカーが講じるプロモーションを考察している。Blattberg and Wisniewski (1989) は、特定ブランドの販売個数と、通常単価や販売単価、広告タミー等を回帰することでそれぞれの販促効果を測定した。また高価格商品の値引は低価格商品の販売個数に負の影響を与えるが、逆は起きないことを確認した。Condon (2012) は、購買者の属性や販促といった外的要因が購買確率に与える影響を記述するために、購買者の購買履歴に応じて回帰係数を変化させるロジスティック回帰モデルを提案した。Hruschka (2017) は、特定商品群のための販促が他商品群の購買に与える影響を加味したモデルを導入し、その影響の存在を示した。そのほか、リアルタイムで詳細な市場動向の把握や、リアルタイムな自社品販売量の追跡が可能となっている。このように POS データは、流通やメーカーの課題を解決すべく活用されてきた。POS が普及する前は、協力者による手書きの購買情報をもとに上記分析を行っていたことを考えると、POS データはマーケティング分野での購買データ分析を飛躍的に発展させてきた。

一方近年では、マクロ経済分野において活用例が多数みられる。POS データを用いた例として、POS 売上を集計することで、消費や物価等のマクロ経済動向を捉えている。Imai and Watanabe (2015) は、POS データを用いて消費者物価指数を再現する手法を提唱し、両者の差異の要因を分析している。Abe and Tonogi (2010) は、大規模な日次 POS データセットから消費者物価を推計した

ところ、月次データから推計される物価と比較して、価格変更頻度が激しいことを報告した。また、ポイントカードの購買情報から物価指数を作成したり (CCC マーケティング株式会社)、クレジットカードの購買情報から消費動向統計を算出する (NOWCAST) など、POS データのみならず消費関連データは実務でも幅広く活用されている。

マクロ経済への応用に加え、POS 売上をメーカーごとに集計することで、メーカーの業績予測にも活用されている。実務上においては、POS データに加えてクレジットカードデータやスクレイピングデータのような他のオルタナティブデータを組み合わせて、企業業績を予測するサービスを提供する会社 (NOWCAST, hands) が多数存在したり、そのようなサービスや自社でのデータ分析を資産運用に活用する運用会社 (ゴールドマン・サックス・アセット・マネジメント株式会社、ブラックロック・ジャパン株式会社) が存在し、年々広がりを見せている。吉野・山本 (2023) は、個別企業のファンダメンタルズ分析の補助として、POS データから単価や数量、市場シェア、多角化度合いを算出し、その活用法を提唱している。

Ishikawa et al. (2016) は、米国日用品メーカーの四半期実績売上高の成長率は米国スーパーマーケットで観測される POS データの成長率と高く相関しており、また実績売上高のうちスーパーマーケットでの売上割合が高いメーカーほどその相関は高い傾向があるため、POS データを活用した企業売上高のナウキャスト可能性を示唆した。同報告では売上高予測の可能性の示唆に留まっているため、本研究では実際に POS データを用いて実績売上高の予測を試みた。また同報告では POS データを全て使用しているが、POS データには様々な誤差要因が存在し、サンプリングバイアスも大きいいため、本研究では POS データの選別を行った後に売上高予測を行うことで、その効果も併せて検証した。データ選別手法として、インデックストラッキングの先行研究 (Nakayama and Yokouchi, 2018) を参考に、過去の四半期実績売上高を良くトラックする店舗のみの POS データを使用することで予測精度は向上したが、経営者予想の方が遥かに高い精度であったため、さらなる工夫が必要であると考えられる。

3 本研究で使用する POS データ

本研究では、株式会社 TrueData が提供する POS データセット (以下、本 POS データ) を用いる。このデータセットでは、日次で収集された日本国内のドラッグストアチェーン 4 社とスーパーマーケットチェーン 5 社の POS データを、チェーンごとに規格の異なる箇所の標準化や分析者が分析可能になるようにデータを加工し、データベースへ収録している。

本 POS データは国内の全小売上の一部を集めたサンプルデータであり、日本全国で発生する全購買のうち、本 POS データが捕捉できている割合をカバー率とよぶこととする。なお、全購買については経済産業省発表の商業動態統計¹⁾に記録された売上高や店舗数のデータを用いる。図 1 は、2019 年の POS 総額 (橙) と商業動態統計販売額 (青) を示している (単位: 億円)。スーパーマーケットの売上高カバー率は約 9%、ドラッグストアは約 22%となっている。

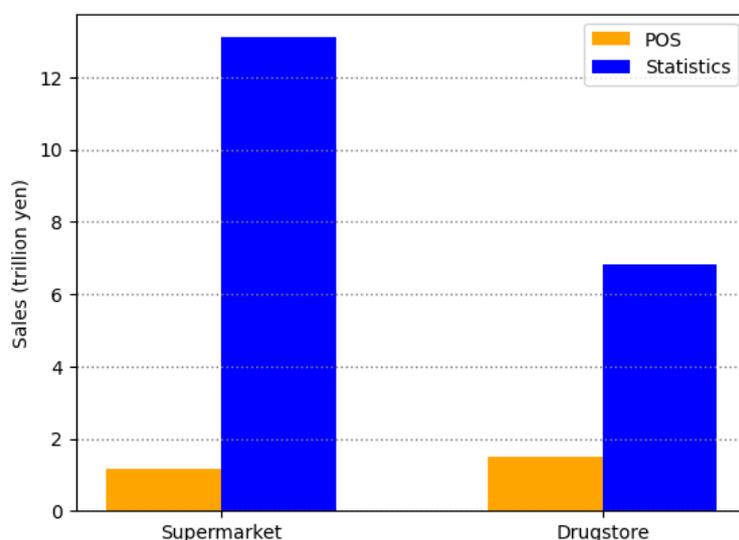


図 1 業態別、商業動態統計販売額と本 POS データ総売上高の比較

図 2 には 2019 年の業態別売上比率を示す。商業動態統計の約 40%を占める家電大型専門店・百貨店・ホームセンターの購買情報は、本 POS データには収録されていないことに留意する必要がある。

図 3 では、2019 年商業動態統計と本 POS データの地域別シェアを示すが、関東のシェアが大きく近畿地方が続くといったような大まかな傾向は一致しているが、九州地区で大きく差がある等、POS データベースには地域バイアスが存在する。さらに POS データでは海外売上は全く捕捉していないので、国内売上にバイアスが掛かっている。

以上で示したサンプリングバイアスを軽減すべく、本研究ではウェイトバック集計 (Battaglia et al., 2009) を行う。ウェイトバック集計とは、サンプルを母集団の構成比に合わせる方法である。本

1) <https://www.meti.go.jp/statistics/tyo/syoudou/index.html>

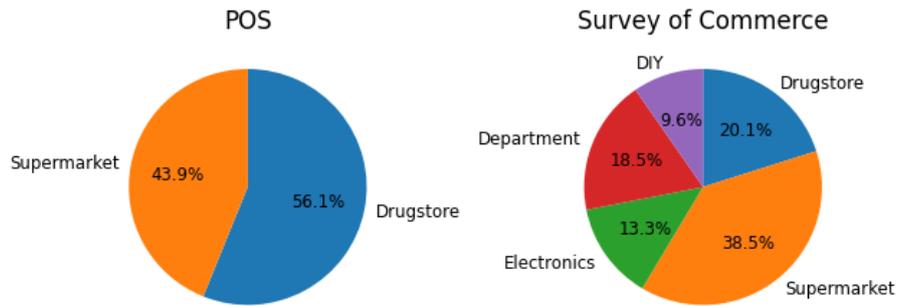


図 2 業態別、商業動態統計店舗数と本 POS データ店舗数の比較

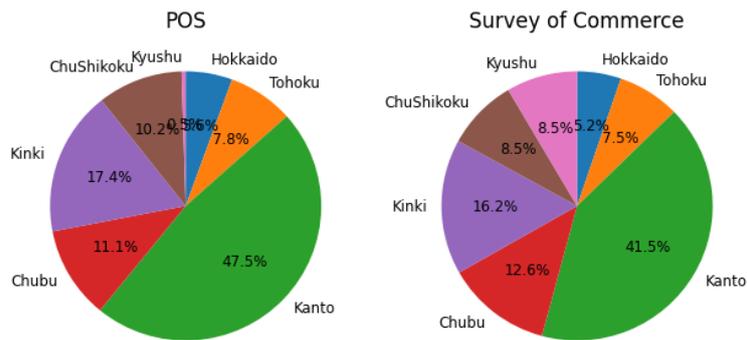


図 3 地域別 POS 売上比率, 商業動態統計小売販売額比率

研究では、地域別 POS 売上構成比を、商業動態統計の地域別売上構成比に合わせることをとする (図 3 参照)。簡易的に、2019 年商業動態統計の地域別構成比を全期間に適用する。

なお、商業動態統計から地域別スーパーマーケット・ドラッグストア構成比を算出可能 (表 1, 表 2) であるため、地域構成比に加えて、スーパーマーケット・ドラッグストア構成比も合わせることも可能ではあるが、ウェイトバックのメッシュを細かくするとサンプルが少なくなること、また結果に大きな差が出ないことから、本研究では地域構成比 (図 3 参照) のみをウェイトバックする。

地域	スーパー	ドラッグストア
北海道	4.74%	0.05%
東北	8.45%	0.49%
関東	18.90%	30.97%
中部	3.95%	0.28%
近畿	1.56%	22.10%
中四国	1.88%	6.23%
九州	0.00%	0.67%

地域	スーパー	ドラッグストア
北海道	3.83%	1.40%
東北	5.14%	2.30%
関東	26.49%	14.95%
中部	8.36%	4.20%
近畿	11.15%	5.07%
中四国	5.69%	2.81%
九州	5.05%	3.55%

収集対象の拡大や小売店舗の開廃業、休業に伴い、本 POS データの収集対象店舗は一定ではなく、図 4 で示すように増加傾向にある。対象企業の製品売上を全ての小売店舗で集計すると実績の企業売上に対してバイアスがかかるため、本研究では特定の時点で収録されている店舗に調査対象を固定する必要がある。

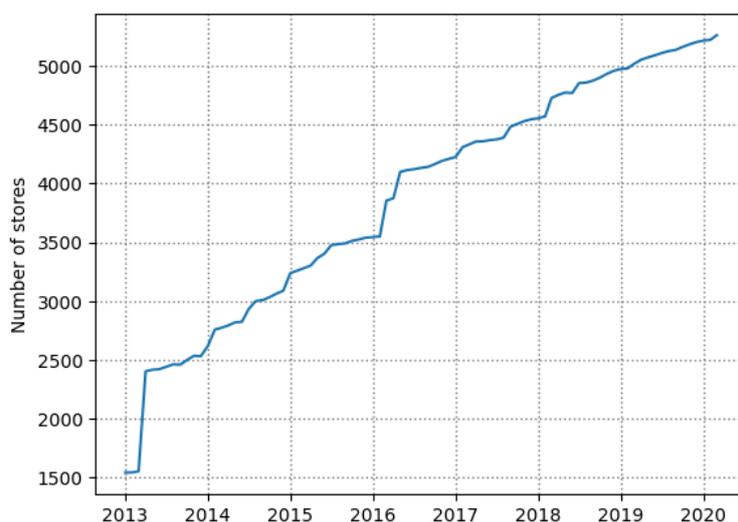


図 4 本 POS データにおける店舗数推移

4 売上高予測手法と結果

4.1 全 POS データを使用した売上高予測

本 POS データの収集対象店舗数は増加傾向にあるため (図 4), 全 POS データを合計して売上予測に用いると上方バイアスがかかる。そのため, 予測モデル構築にあたっては, データ期間の最初の時点から存在する店舗に限定してデータを組織化し直す必要がある。本研究の目的からすれば, できるだけ多くの店舗数が揃う時点からのデータを使いたいものの, 実店舗消費が大きく変化するコロナ禍は除外する必要があることから, 2015 年 4 月 1 日以前に売上が記録された店舗を分析対象とし, 2015 年 4 月 1 日から 2019 年 3 月 31 日を分析期間とした。

本研究では POS データを用いて実績売上高を予測するが, 上記の理由から使えるデータが 8 四半期のみとなるため, モデル構築に使えるデータは限定的となっている。そのため, 過度なモデル化による過学習の弊害を避けるべく, 本研究では, 四半期実績売上高を被説明変数, 四半期 POS 売上高を説明変数としたシンプルな線形回帰モデルを使い, POS データによる企業の実績売上予測の可能性について検証することとした。

また, 実績売上高予測を行うにあたっては, POS データと実績売上高とが高い相関を持つ, いわゆる小売業に属する企業を選んでおく必要がある。本研究では, 2015 年 4 月 1 日～2017 年 3 月 31 日の 8 四半期分の実績売上高 y を同期間の POS 売上高 x で単回帰

$$y = \beta_0 + \beta_1 x + \epsilon,$$

を行い, R^2 が大きい上位 50 銘柄を本検証の対象とした。

まずは, 予測期の直前 8 四半期売上高の単回帰モデルのあてはめを行った。そして, 得られた回帰式に予測期の四半期 POS 売上高を代入して四半期実績売上高の予測値を得た (図 5)。なお予測期は, 2017 年 4 月 1 日～2019 年 3 月 31 日の計 8 四半期とした。Ishikawa et al. (2016) では POS データの選別は行っていないため, 本項ではそのセッティングに準じて, 2015 年 4 月 1 日時点でデータ上に存在する店舗の POS データ全てを使用した。

さらに, 得られた四半期実績売上高予測値と, 公表済みの四半期実績売上高を用いて, 通期の実績売上高成長率予測を行った。ここで, 図 6 は, 第 3 四半期決算発表前の通期売上高成長率予測値の定義である。通期の成長率予測は, 4.4 項の経営者予想との比較で用いる。

また, 売上高予測の精度は, 実績売上高との平均絶対誤差率 MAPE

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|,$$

で評価した。

表 3 は江崎グリコとブルボンの回帰係数, 図 7 と図 8 は各四半期売上高の散布図と回帰直線を示している。データ期間は 2017 年 10 月 1 日から 2018 年 12 月 31 日までの 8 四半期である。係数 p 値

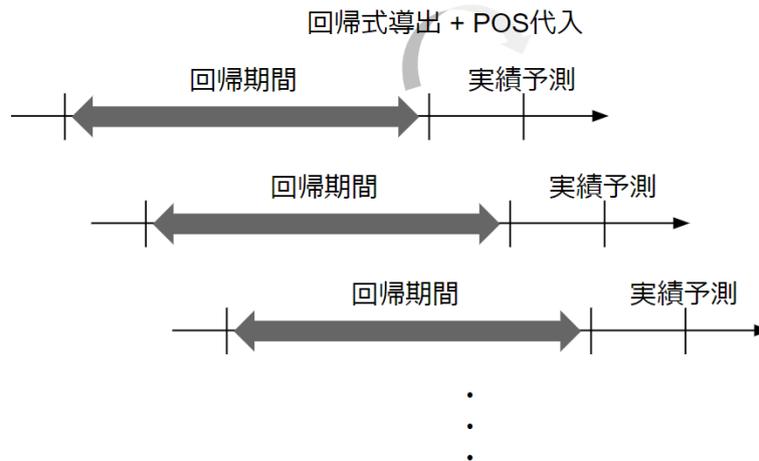


図5 回帰による四半期実績売上高の予測

$$\text{通期売上高成長率予測値} = \frac{\text{3Q予測売上高} + \text{2Q実績売上高} + \text{1Q実績売上高}}{\text{前年3Q実績} + \text{前年2Q実績} + \text{前年1Q実績}} - 1.$$

図6 四半期売上高予測に基づく通期実績売上高成長率予測

は 0.01 を大きく下回っているが、時間が経過するにつれてブルボンの残差が大きくなっている。

	江崎グリコ	ブルボン
回帰係数	43.1	16.4
標準誤差	5.3	4.6
t 値	8.0	3.5
p 値	0.0002	0.0116

また、図9は四半期売上高予測値のMAPEの分布を示しているが、やや大きなMAPEを示す銘柄がいくつかあることがわかる。

このようなあてはめの悪さを示す要因について、POSデータに立ち返って調べたところ、個別の店舗の特性や商品特性によって生じるケースがあることが分かった。

図10は、とある店舗の日次販売商品点数だが、21日間連続で売上点数が0となっている。これは長期の改装閉店等を行っていたと考えられるが、この店舗のPOSデータはそのまま予測に用いるのではなく、削除や補完といった対処が必要なことは明らかである。

長期間在庫が可能な商品は、メーカーの販売時点と小売店舗での販売時点が大きく異なる可能性があり、売上高予測値にラグが生じる。そのため予測誤差が大きくなる傾向にある。図11はフマキラー社の店舗及び実績の四半期売上高成長率であるが、POSのピークがQ2に集中している一方で、

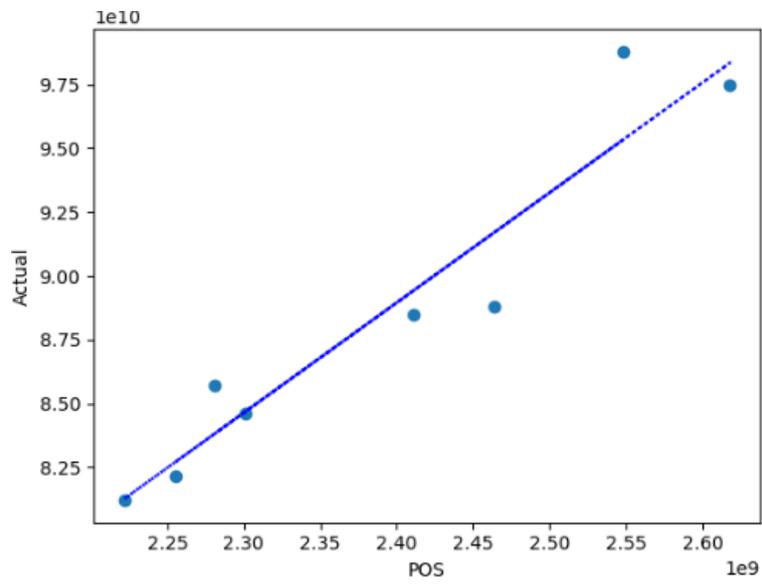


図7 江崎グリコの各売上高と回帰直線

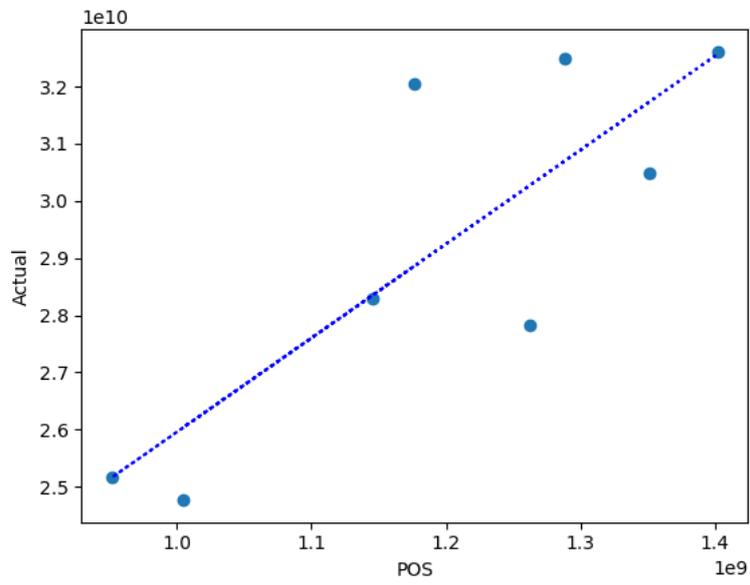


図8 ブルボンの各売上高と回帰直線

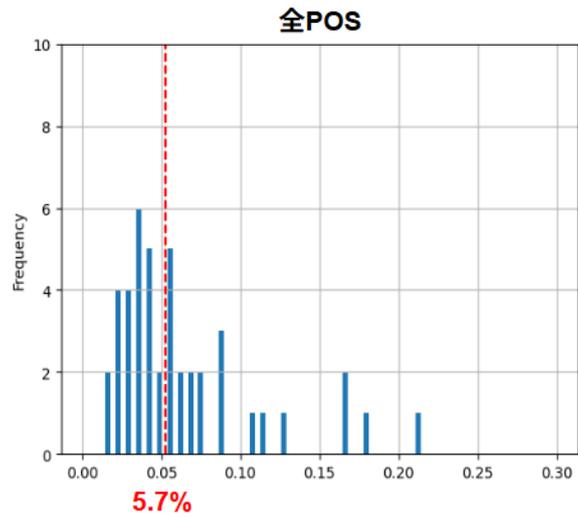


図 9 全 POS データを使用した四半期売上高予測値の MAPE

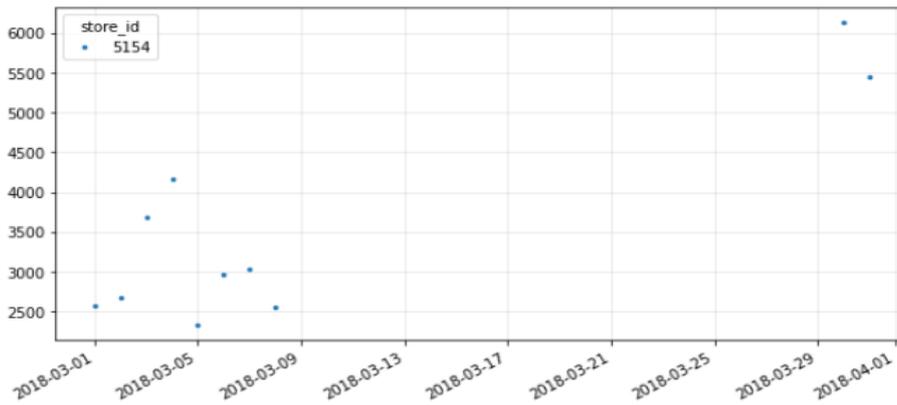


図 10 店舗 ID5154 番の日次販売商品数

実績では1Qがピークとなっている。これは、同社製品がよく売れる2Qを前に、1Qで小売業者へ販売しているためだと考えられる。

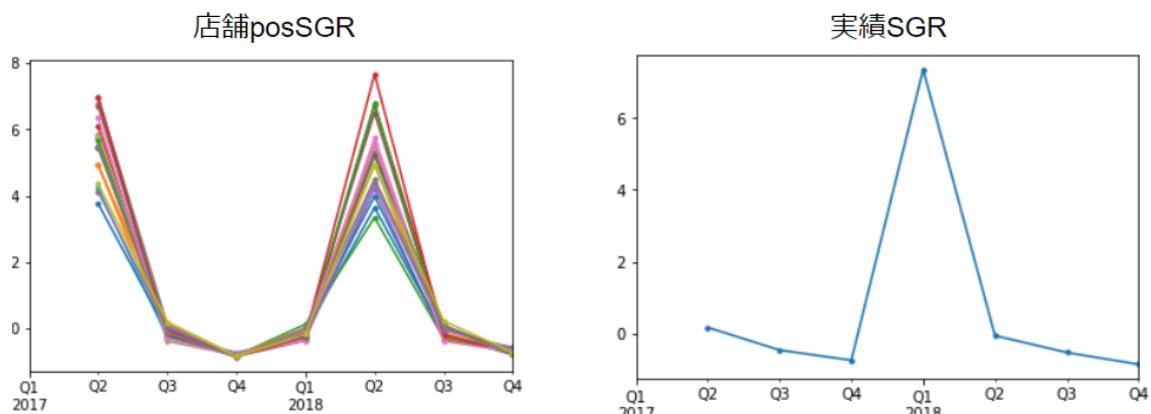


図 11 フマキラー (4998) の POS および実績売上高成長率

このほかにも、特定店舗や店舗群が大幅割引セールを行うことで POS 売上高が局所的に著しく伸びるケースは頻発していることが分かった。

4.2 類似度により店舗選別した売上高予測

前節で述べたように POS 売上高は実績売上高に沿った動きをすることは限らず、例えば店舗ごとの商品構成や販促活動といった店舗特性によって大きく左右されるため、POS データから適切に店舗を選択して、それらのみでデータ分析を実施する必要がある。分野は異なるが、Dose and Cincotti (2005) は、少数銘柄で株式ベンチマークをトラックするポートフォリオを構築するために、ベンチマーク構成銘柄間で階層的クラスタリングを行い、各クラスタからポートフォリオの構成銘柄を選択した。また、似たような手法として、Nakayama and Yokouchi (2018) も、ベンチマークとベンチマーク構成銘柄との Euclid 距離を計算してそれを非類似度とすることでポートフォリオの銘柄選択を行っている。本研究では、これらの手法を参考にして、ベンチマークを企業の実績売上高、構成銘柄を各店舗の POS 売上高、Euclid 距離を非類似度と見なし、企業の実績売上高と非類似度が低い売上高を示す上位 100 店舗、また幅広く非類似度を分散させた 100 店舗群 (以下、類似度バランス) を抽出し、その店舗群の POS 売上高のみを用いて企業の売上高予測を行った。なお、POS 売上高と実績売上高はスケールが異なるため、図 12 のように、それぞれ集計期初の売上高で全期間の売上高を除することで標準化を行い、時系列同士の非類似度を求める。

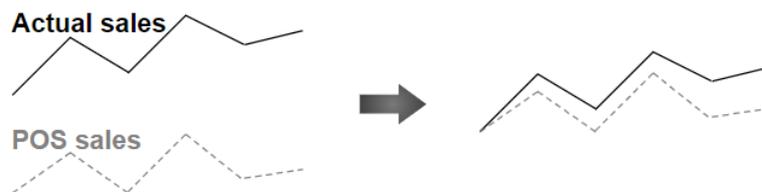


図 12 売上高推移の標準化イメージ

本研究においては、Nakayama and Yokouchi (2018) にならった Euclid 距離 $D_E(\mathbf{x}, \mathbf{y})$ 、直近期によりウェイトを付した Euclid 距離 $D_{WE}(\mathbf{x}, \mathbf{y})$ (Weighted Euclid 距離と呼ぶ)、時点のずれを許容する Dynamic Time Warping 距離 $D_{DTW}(\mathbf{x}, \mathbf{y})$ (DTW 距離と呼ぶ) を時系列の非類似度として用い、比較検討する。それぞれの距離は、 $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ 、 $\mathbf{y}_n = (y_1, y_2, \dots, y_n)$ を用いて以下のように記述される。

$$D_E(\mathbf{x}_n, \mathbf{y}_n) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2},$$

$$D_{WE}(\mathbf{x}_n, \mathbf{y}_n) = \sqrt{\sum_{i=1}^n \frac{1}{n-i+1} (y_i - x_i)^2},$$

$$D_{DTW}(\mathbf{x}_n, \mathbf{y}_n) = D_E(x_n, y_n) + \min(D_{DTW}(\mathbf{x}_n, \mathbf{y}_{n-1}), D_{DTW}(\mathbf{x}_{n-1}, \mathbf{y}_n), D_{DTW}(\mathbf{x}_{n-1}, \mathbf{y}_{n-1})).$$

図 13 に、高類似度 100 店舗群を対象に、各距離に基づいた四半期売上高予測値の MAPE の分布を示す。Euclid 距離と Weighted Euclid 距離、Dynamic Time Warping 距離では誤差に大きな差はない。

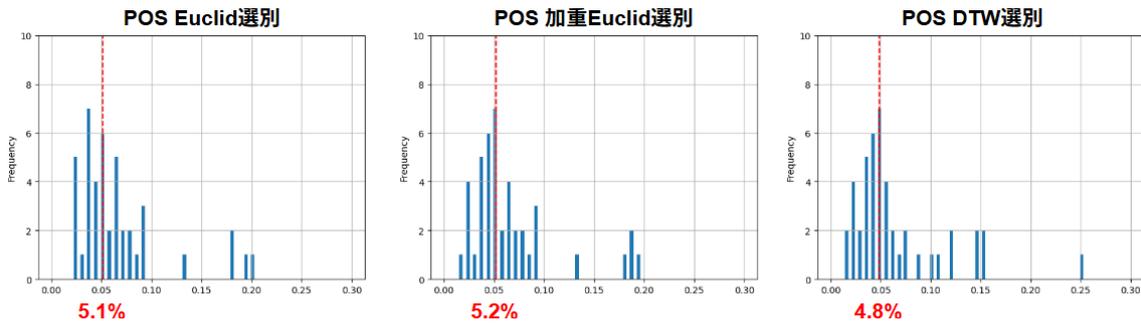


図 13 選別した POS データに基づく四半期売上高予測値の MAPE

図 14 は、全 POS データを用いた予測と POS データ選別後の予測値の MAPE である。後者は、予測誤差が小さい高類似度の MAPE を示す。

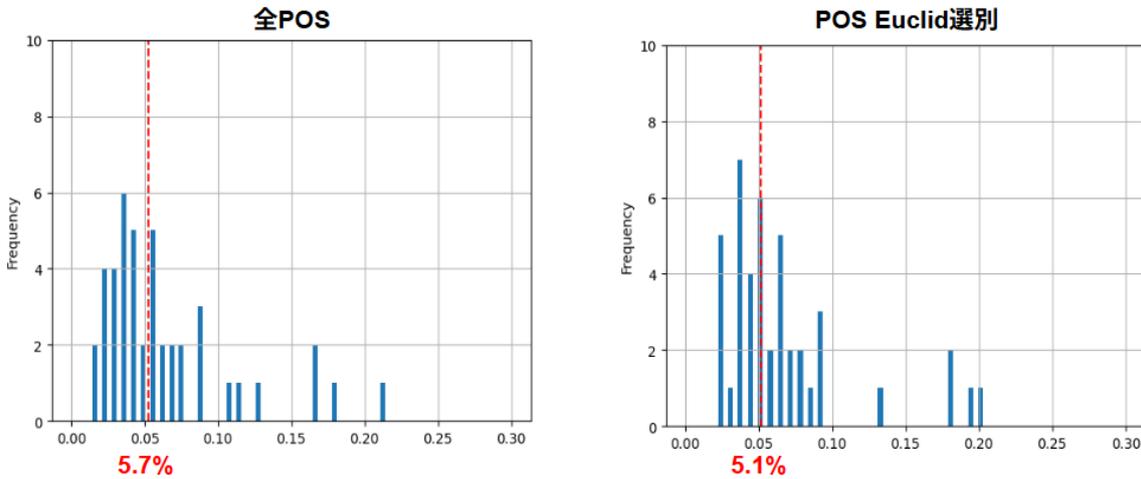


図 14 各手法の MAPE

図 14 より、全 POS データを使う予測に比べて、店舗を軸にデータを選別した予測の方が、予測誤差の中央値 (赤点線)、平均値ともに小さくなっている。

以上のように、サンプリングバイアスの大きい POS データにおいて、POS データにおける店舗選択の手法は、企業の実績売上予測という目的においては有効に機能することが明らかとなった。

4.3 株式投資への応用可能性

本項では、POS データを用いた売上高予測の株式投資への応用可能性を検証する。これまでと同様に POS データと実績売上高の R^2 が大きい上位 50 銘柄に関して、対前年度の実績売上高成長率 (actual_SGR)、経営者予想²⁾ (guidance_SGR_2018, guidance_SGR_2019)、POS データによる予測値 (POS_SGR)、対 TOPIX の株価超過リターン (ExcessReturn) を時系列図として並べ、それぞれの関係を調べた。なお経営者予想は通期で公表されることが一般的であるため、上記の各売上高成長率は通期である。

株価は日次で変動するため、POS による予測値も日次で算出するべく、未経過日の POS 売上は前年同日で補完し、前年同四半期の POS 売上で割ることで四半期売上高成長率予測値とした。用いる POS データは、高類似度 100 店舗群を対象とする。図 15 は、2018 年 10 月 3 日時点の四半期売上高成長率予測値である。未経過の 10 月 4 日以降は前年同日の POS 売上で補完している。このようにして求めた四半期売上高成長率予測値から四半期売上高予測値を求め、図 6 を用いて通期の売上高成長率予測値を求める。

$$\text{四半期売上高成長率予測値} = \frac{\begin{array}{ccccccc} \text{2018/10/01} & \text{2018/10/02} & \text{2018/10/03} & \text{2017/10/04} & \text{2017/10/05} & \dots & \text{2017/12/31} \\ \text{POS売上} & \text{POS売上} & \text{POS売上} & \text{POS売上} & \text{POS売上} & & \text{POS売上} \end{array}}{\begin{array}{ccccccc} \text{2018/07/01} & \text{2018/07/02} & \text{2018/07/03} & \dots & & & \text{2018/09/30} \\ \text{POS売上} & \text{POS売上} & \text{POS売上} & & & & \text{POS売上} \end{array}} - 1.$$

図 15 日次算出の POS による四半期売上高成長率予測

この調査からわかった時系列図の特徴は大きく分けると 2 つある。1 つ目は、POS 予測が実績売上高の推移を予見し、その後の株価変動を予見しているケースである。図 16 は江崎グリコ (2206) の例で、POS 予測と経営者予想が乖離し、その後の経営者予想修正を予見しており、その後の株価変動も予見している。2018 年 3 月期・2019 年 3 月期の第二四半期決算発表時に通期業績予想を下方修正し、直後から株価は急落しているが、POS データによる売上高予測は期初から売上高の低成長を示しており、実績売上高もそれに沿った推移となっている。ホクト (1379) など他銘柄でも同様の現象が確認できる。

図 17 はブルボン (2208) の例では、経営者予想修正は伴わないが、POS 予測が実績売上高の推移を予見し、その後の株価変動を予見している。2018 年 3 月期・2019 年 3 月期ともに経営者予想の修正はなかったが、POS 予測に沿って実績売上高が推移しており、株価リターンも概ねそれに沿った推移となっている。株価が売上高に連動する直接的な因果関係があるとは言えないが、例えば利益率がほぼ一定で売上高と利益が概ね連動しており、その結果売上高と株価が連動する傾向にある銘柄なら、POS 予測は株式投資に有用といえる。

2 つ目は、売上高推移に関係なく株価変動が起こるケースである。コストの増減により利益が変動し、好決算や経営者予想修正を伴って株価が変動したり、株主還元の増減で株価が変動したり、

2) 株価は業績予想値をベースに形成されることが考えられるが、アナリストコンセンサスは比較的規模の大きい企業にのみ付与されるため、ここでは上場企業のほぼ全てで公表されている経営者予想を参照した。



図 16 江崎グリコ 株価超過リターンと売上高, 売上高予測



図 17 ブルボン 株価超過リターンと売上高, 売上高予測

M&A 等の他要因で株価が変動するケースが考えられる。図 18 はハウス食品 G 本社 (2810) の例で、2018 年 3 月期の第一四半期決算では売上は概ね横ばいも大幅増益となり株価は大きく上昇、続く 2019 年 3 月期の第三四半期決算では資産売却益に伴う特別配当を発表し株価は大きく上昇したが、両者とも POS データからの予見は不可能である。実務で用いることを考えるのであれば、コストに連動する指数のような市場データや、ニュース・コーポレートアクションデータを参照することで、本予測を補完する必要があるだろう。

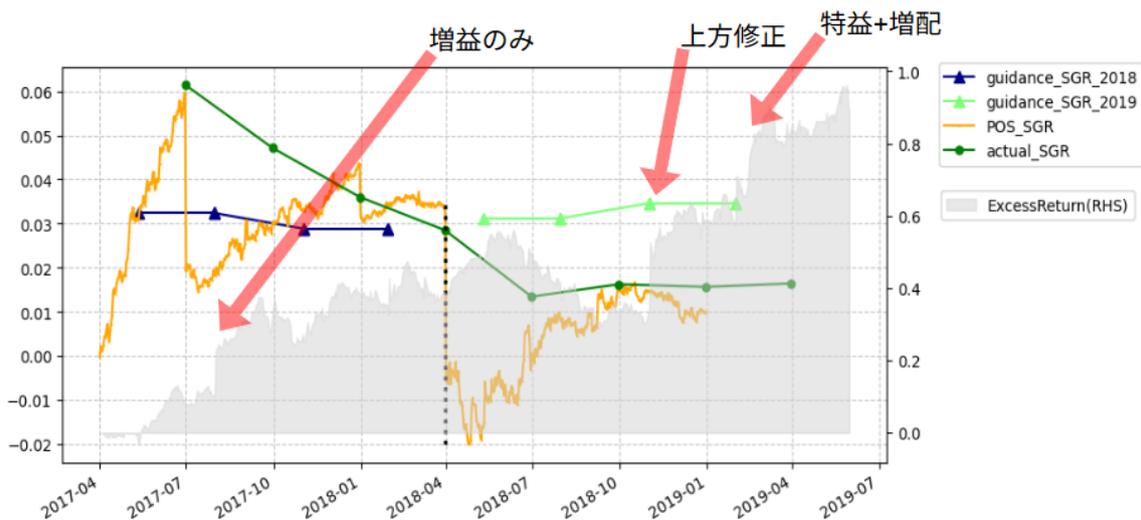


図 18 ハウス食品 株価超過リターンと売上高, 売上高予測

4.4 経営者予想との比較

本手法の通期売上高予測値と経営者予想の MAPE を図 19 に示す。経営者予想の誤差と比較すると、高類似度、類似度バランスともに、POS データによる予測値の誤差は大きくなった。

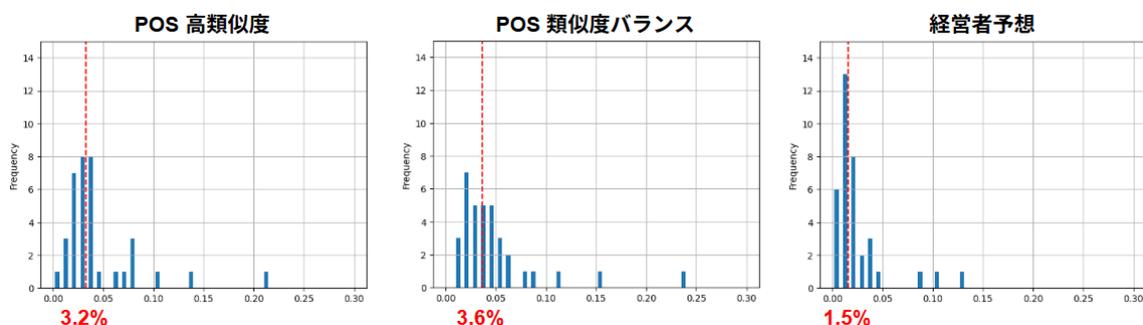


図 19 POS 予測と経営者予想の誤差比較

表 4 には、経営者予想よりも本研究手法による予測値のほうが正確であった企業を列挙したが、これらの企業については「本 POS データによる売上捕捉率が高い」、「国内売上比率が高い」、「小売比率が高い」ことがその特徴として挙げられる。

表 4 POS 予測の精度が経営者予想を上回る企業

企業名	POS 捕捉率	国内比率
江崎グリコ	3.9%	75%
かどや製油	2.7%	80%
中村屋	2.9%	非開示
フジッコ	8.7%	99%超

そもそも経営者予想は、需要予測や受注や、生産計画等の内部情報を多角的に用いて作成しており、また乖離があれば足元の数値を踏まえてすぐに見直すことができるため、POS データのみを使用した本研究手法と比較してきわめて有利な条件にある。そのため、現時点では経営者予想を補完する道具として POS データの予測を使うことが望ましいと思われる。もし、経営者予想以上の精緻な予測を実現するならば、POS データに加えて受注等の他財務情報やニュースなどの非財務情報と組み合わせる必要があると考える。

5 おわりに

5.1 結論

本研究では、POS データを用いて、企業の実績売上高予測を試みた。保有する POS データをすべて用いる場合と、過去の類似度をもとに店舗を選別する場合で、予測精度の比較を行った。その結果、以下の知見を得た。

- POS データ分析に際し、店舗由来の攪乱要因や商品由来の攪乱要因の一部を特定した。
- 過去の売上実績推移と類似度が高い店舗の POS データのみを用いることで、実績売上高の予測精度は向上させた。これは類似度による選別を行うことで、特異的な売り上げを示す傾向にある店舗を排除できたためと考えられる。
- 先行研究にならい類似度を分散させて店舗を選別しても、予測精度の向上は見られなかった。トレンドが突如変わりがちな先行研究の例と違い、店舗の売上挙動は安定しているためだと考えられる。
- 類似度ごとの予測精度には大きな差はなかった。その結果、計算が簡便な Euclid 距離で十分であることがわかった。
- 類似度をもとに POS データを選別して予測精度を上げても、経営者予想の精度には及ばなかった。経営者予想は企業の内部情報に直結した様々なデータをもとに作られるため、POS データのみを使用した本研究に比べて情報量が多いことが原因だと考えられる。
- 一部の相場銘柄においては、POS データを用いた売上高予想が実績売上高や経営者予想修正を予見しているケースがあった。経営者予想と組み合わせるなどすれば、POS データによる売上予測が株式投資にも有用である可能性が示唆された。

5.2 今後の課題

本研究では POS データと過去売上高のみを用いたが、その他の財務指標や非財務指標などを用いることで予測誤差を小さくできる可能性がある。また、売上高予測のモデルとして単純な線形回帰モデルを用いたが、フィッティングに使うモデルを工夫することで予測精度を向上させる余地はあるかもしれない。そもそも、本研究で使用した POS データは蓄積が少なく、サンプリングバイアスも大きいいため、単純に POS データのカバー範囲を広げることで予測精度の向上が図れるのではないと思われる。

本研究では個別株式投資を想定して POS データから四半期実績売上高そのものを予測したが、製品の販売数量や単価、市場シェア変動といった指標は、実績売上高や株価変動に対して予見性を持つと考えられる。このような仮説の検証も今後の課題としたい。

参考文献

- Abe, N. and Tonogi, A. (2010): Micro and macro price dynamics in daily data. *Journal of Monetary Economics*, **57**, 716-728.
- Ball, R. and Brown, P. (1968): An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, **6**(3), 159-178.
- Battaglia, M. P., Hoaglin, D. C. and Frankel, M. R. (2009): Practical considerations in raking survey data. *Survey Practice*, **2**(5).
- Bernard, V. L. and Thomas, J. K. (1989): Post-earnings-announcement drift: Delayed price response or risk premium?. *Journal of Accounting Research*, **27**, 1-36.
- Blattberg, R. C. and Winsniewski, K. J. (1989): Price-induced patterns of competition. *Marketing Science*, **8**(4), 291-309.
- Carhart, M. M. (1997): On persistence in mutual fund performance. *The Journal of Finance*, **52**(1), 57-82.
- Condon, E. M. (2012): Varying Coefficients in Logistic Regression with Applications to Marketing Research. *doctoral dissertation of Statistics in the Graduate College of the University of Illinois at Urbana-Champaign*, **11**(2), 167-192.
- Dose, C. and Cincotti, S. (2005): Clustering of financial time series with application to index and enhanced index tracking portfolio. *Physica A*, **355**(1), 145-151.
- Financial Services Authority (2016): The overall impact of mifid.
- Hassell, J. M. and Jennings, R. H. (1986): Relative forecast accuracy and the timing of earnings forecast announcements. *The Accounting Review*, **61**(1), 58-75.
- Hruschka, H. (2017): Multi-category purchase incidences with marketing cross effects. *Review of Managerial Science*, **11**, 443-469.
- Imai, S. and Watanabe, T. (2015): Replicating japan' s cpi using scanner data. *CARF Working Paper Series*, F-364.
- Ishikawa, A., Fujimoto, S. and Mizuno, T. (2016): Nowcast of firm sales using pos data toward stock market stability. *IEEE International Conference on Big Data*, 2495-2499.
- Nakayama, J. and Yokouchi, D. (2018): Applying time series decomposition to construct index-tracking portfolio. *Asia-Pacific Financial Markets*, **25**, 341-352.
- Tetlock, P. C. (2007): Giving content to investor sentiment:the role of media in the stock market. *The Journal of Finance*, **62**(3), 1139-1168.
- Weber, M. M. and Kantamneni, S. P. (2007): Pos and edi in retailing: an examination of underlying benefits and barriers. *Supply Chain Management*, **7**(5), 311-317.
- Williams, P. A. (1996): The relation between a prior earnings forecast bymanagement and analyst response to a current management forecast. *The Accounting Review*, **71**(1), 103-115.

- Ziebart, D. A. and Choi, J. H. (2004): Management earnings forecasts and the market's reaction to predicted bias in the forecast. *Asia-Pacific Journal of Accounting and Economics*, **11**(2), 167-192.
- 阿部誠・近藤文代 (2005): マーケティングの科学. 朝倉書店.
- 太田浩司 (2007): 業績予想における経営者予想とアナリスト予想の役割. 証券アナリストジャーナル, **45**(8), 54-66.
- 株式会社ナウキャスト : JCB 消費 NOW. <https://www.jcbconsumptionnow.com/>.
- 株式会社 hands : Peragaru. <https://peragaru.net/>.
- ゴールドマン・サックス・アセット・マネジメント株式会社 : GS グローバル・ビッグデータ投資戦略ファンド. [urlhttps://www.gsam.com/content/gsam/jpn/ja/gsitm/fund-center/fund-finder/gs-global-big-data-strategy-b-course-\(unhedged\).html](https://www.gsam.com/content/gsam/jpn/ja/gsitm/fund-center/fund-finder/gs-global-big-data-strategy-b-course-(unhedged).html).
- CCC マーケティング株式会社 : T ポイント物価指数. <https://www.cccmkhd.co.jp/>.
- 菅野泰夫 (2019): 欧州におけるリサーチ費用のアンバンドリングの実態. 証券アナリストジャーナル, **57**(1), 18-27.
- 鈴木源一郎・森成弥 (2023): クレジットカードデータを用いた個人消費動向把握の精度向上の取組. 経済財政分析ディスカッション・ペーパー, <https://www5.cao.go.jp/keizai3/discussion-paper/dp231.pdf>.
- 竹中幹・浦野昌一 (2021): ニュースの見出しを活用した株価予測. 人工知能学会第 35 回全国大会論文集, 1G4-GS-2c-05.
- 日本経済新聞社 (2019): 投資に生きる独自データ ホテル予約・特許…異変読む. <https://www.nikkei.com/article/DGXMZ039871650Q9A110C1920M00/>.
- ブラックロック・ジャパン株式会社 : ブラックロック・米国小型株式ビッグデータ戦略ファンド. <https://www.blackrock.com/jp/individual/ja/products/287540/blackrock-us-small-cap-big-data-strategy-fund-unhedged>.
- 流通システム開発センター (2012): 流通情報システム化実態調査報告書.
- 吉野貴晶・山本裕大 (2023): POS データ分析と運用実務への応用可能性について. 証券アナリストジャーナル, **61**(11), 7-19.

付録 本研究で使用する POS データの詳細

本 POS データセットには、POS データテーブル、商品マスタテーブル、店舗マスタテーブルという 3 つのテーブルが収録されている。POS データテーブルは、文字通り POS データを格納するためのテーブルである。各行にはある店舗の 1 日における 1 商品の購買数量と売上金額、そして製造者コードが記録されている。製造者コードには、当該企業の証券コードないし親会社の上場企業証券コードが紐づけられている。また、本テーブルにおいて商品は JAN コードで記録されているが、JAN コードとは商品のバーコードに付された 8~13 桁の数字列で、一般財団法人流通システム開発センターに登録された SKU (Stock Keeping Unit) 単位の商品コードであり、SKU とは受発注や商品在庫管理で用いる最小の管理単位のことである。我々が用いるデータベースでは、JAN コードを介して商品マスタテーブルを参照することで、具体的な商品名等の属性が分かり、店舗コードを介して店舗マスタテーブルを参照することで、チェーンや地域等の販売店舗属性が分かる。そして、2020 年時点で、このデータベースに掲載された購買記録数は約 100 億レコードになっている。

商品マスタテーブルとは、商品の属性を格納するテーブルである。商品ごとに、JAN コード、商品名、商品カテゴリー、所属するブランドコードが記録されている。商品カテゴリーは、例えば「食品」→「菓子類」→「菓子」→「キャンディ・キャラメル」のように、大分類から小分類の全 4 階層から構成される。ブランドコードとは、1 単位の数量や容量等の違いから、SKU 上は異なると見做された同ブランドの商品を纏めるために付与されている。2020 年時点で、商品数は約 650,000 である。

店舗マスタテーブルとは、小売店舗の属性を格納するテーブルである。店舗ごとに、店舗コード、所在地域 (日本全国を 7 地域に分類)、東京か否か、所属するチェーンが記録されている。2020 年時点で、店舗数は約 7,000 である。

本 POS データの店舗カバー率について、図 20 では 2019 年の POS データ集計対象店舗数 (橙) と商業動態統計事業所数 (青) を比較している。スーパーマーケットの店舗カバー率は約 20%、ドラッグストアは約 33%となっている。

図 21 には本 POS データの業態別売上高推移を示すが、データの収集開始当初はドラッグストア比率は 50%未満であったが、近年のドラッグストア業態の急成長により、直近でドラッグストアが約 60%を占めている。一方で前述の商業動態統計 (小売業態別) によると、ドラッグストアの販売シェアは 30%程度であり、本 POS データは当初よりドラッグストアに偏っており、偏りは増していることがわかる。

図 22 には直近の業態別商品カテゴリー比率を示すが、業態別でのカテゴリーバイアスは非常に強いことが分かる。業績予測対象企業の主力商品カテゴリーによっては、注意が必要である。

図 23 には 2019 年の地域別商品カテゴリー比率を示すが、業態別同様、地域別にもカテゴリーバイアスは非常に強いことが分かる。

図 24 では、地域バイアスを確認するために、1: 北海道, 2: 東北, 3: 関東, 4: 中部, 5: 近畿, 6: 中国四国, 7: 九州沖縄、と分類して時系列に売上を集計した。業態別と同じく、地域比率は時系列で大きく変化しており、関東や中部の伸びにバイアスが大きい。

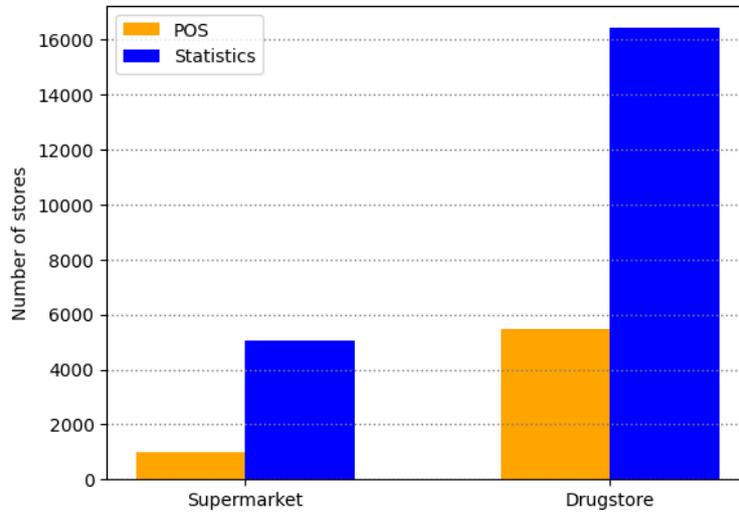


図 20 業態別、商業動態統計店舗数と本 POS データ店舗数の比較

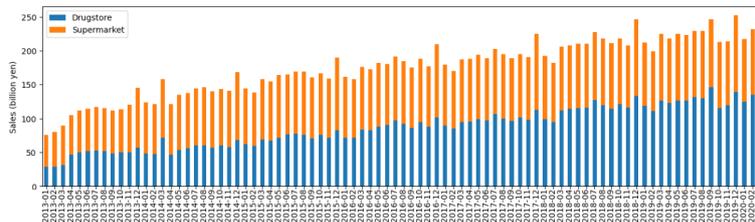


図 21 本 POS データにおける業態別売上高推移

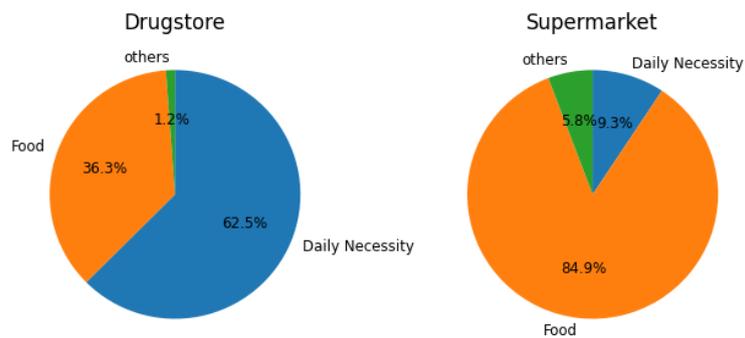


図 22 本 POS データにおける業態別商品カテゴリー比率

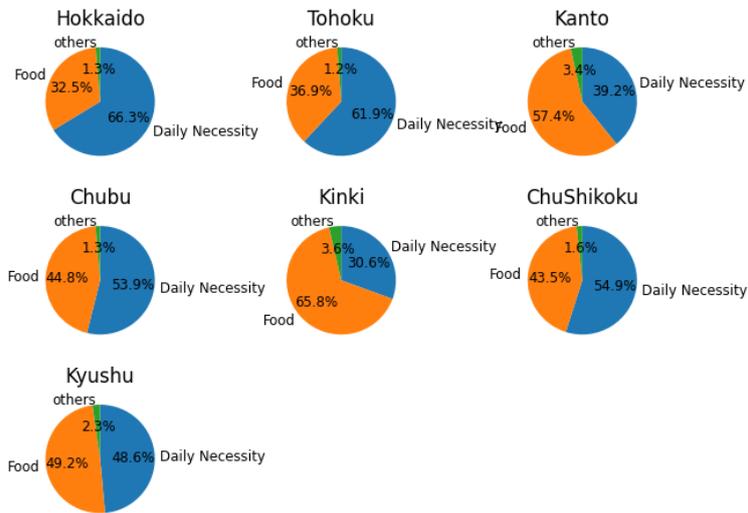


図 23 POS データベースにおける地域別商品カテゴリー比率

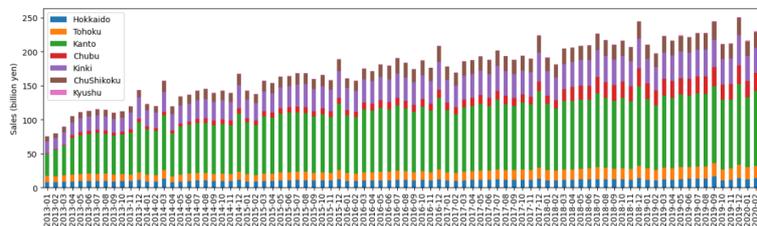


図 24 POS データベースにおける地域別売上高推移

付録 実績売上高予測における本 POS データの有用性

POS データベースに収録された企業のうち、食料品関連企業の POS 売上は、会社全体の売り上げに占める小売店での売上が大きい傾向にあるため、実績売上高に対する説明力が高いと考えられる。以下は、東証 33 業種のうち、食料品並びに水産農林に該当する 114 銘柄についての結果である。2014 年から 2019 年の四半期実績売上高と四半期 POS 売上高の相関係数 (縦軸) と、実績売上高に占める POS 売上高の割合 (横軸) との関係を以下に示す。実績売上高を POS で捕捉した割合が高いほど (横軸)、実績売上高と POS 売上高の相関が高いことが分かる。ゆえに実績売上高予測において POS データは有用であると考えられる。

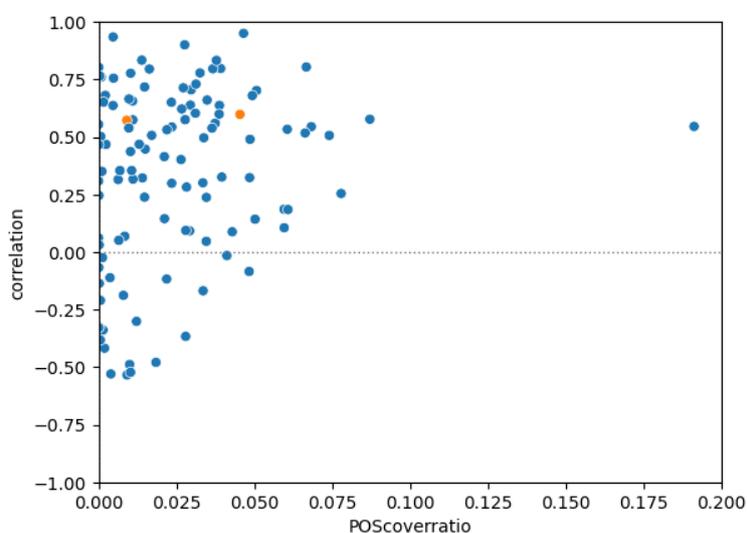


図 25 POS カバー率と POS-実績売上高相関