

J-Quants APIデータを用いた新規特微量の提案

予測ホライズンと評価指標の設定

本プロジェクトではマルチホライズン予測を採用し、例として **1日後、5日後、10日後、20日後** のリターン (`target_1d` ~ `target_20d`) を予測対象とします。これら短期～中期ホライズンを設定することで、デイトレード的な超短期から1か月程度のスイングまで幅広く株価変動を捉えます。また**評価指標**には、主に以下を用いることを提案します。

- **RankIC (Rank Information Coefficient)** : 予測スコアと実現リターンの**順位相関**を測る指標です
①。スピアマン順位相関に相当し、外れ値に頑健であるためファクターの予測力を評価するのに適しています ①。特に5日リターンに対するRankIC (例: RankIC_5d) を重視し、モデルが1週間程度先の株価の**相対的な強弱**を捉えられているか確認します。
- **シャープレシオ (Sharpe比)** : 予測に基づくポートフォリオの**リスク調整後リターン**を評価します。超過リターンをリスク (リターンの標準偏差) で割った指標で、投資判断の質を測る標準的な尺度です ②。例えばモデル予測に沿ったポートフォリオ運用のシャープレシオを算出し、戦略がどれだけ効率的にリターンを上げているか確認します。シャープレシオは高いほどリスク当たりリターンが良いことを示し、負の場合は無リスク資産を下回る効率の悪さを意味します ②。

加えて**情報係数 (IC)** や**ICの時系列安定性 (IC_IR)**、**勝率**、**最大ドローダウン**なども併用し、ファクターの一貫性と戦略リスクを総合的に評価します。例えば**RankICの平均と標準偏差**からIC_IRを算出し、安定した予測力 (IC_IRが高いほど安定) かどうか確認します。またシャープレシオと勝率から、モデル予測に基づくロング・ショート戦略の実現可能性 (取引コスト後にも有効か) を検証します。

以上のように**5日先RankIC**および**シャープレシオ**を主指標に、マルチホライズンでモデル性能を評価しつつ、新規特微量の有効性を判断します。

新規特微量設計の基本方針

提案する特微量は、**J-Quants API**で取得可能な**各種データ系列**や**イベント情報**を組み合わせ、既存特微量を補完・拡張してモデル性能の向上を狙うものです。特微量設計にあたっては以下の基本方針を取ります。

- **リーク防止** : 各特微量は将来情報を含まないよう算出時点に注意します。過去データの一定ウィンドウから計算し、当日終了時点までに利用可能な情報のみを用います (例: ファンダメンタル決算情報は発表翌日以降に反映、週次データは公表翌営業日以降に反映など)。データ分割→正規化学習→変換のパイプライン順序を守り、バリデーション期間に訓練期間の統計量のみを用いることでリークを防ぎます。
- **正規化戦略** : クロスセクショナルな正規化を基本とし、各営業日で全銘柄の特微量を**平均との差0、標準偏差1**にスケールします (学習データで算出した基準を適用) ③。こうすることで異なるスケールの特微量を比較可能にし、また時系列的な分布変化の影響を抑えます。また一部の特微量には対数変換やパーセンタイル変換を用いて**歪度を緩和**し、極端値の影響を低減させます。セクター横断比較をする特徴では**セクター内Zスコア**で正規化し、市場全体のトレンドよりも**同業他社比の異常**を捉えられるよう工夫します。
- **特微量エンジニアリングの視点** : 価格・出来高から得られる**テクニカル指標の高度化**、ファンダメンタルやイベント情報からの**収益機会の抽出**、投資家フローやポジションから読み取れる**センチメント**

や需給、および市場全体や外部要因からの**レジーム・連動性**を、新規特徴として取り入れます。具体的には以下にカテゴリ別の提案を示します。

以上を踏まえ、各カテゴリごとに具体的な新規特徴量を提案します。

1. 高度な価格・ボラティリティ系特徴量

価格変動やボラティリティに関する既存特徴量をさらに拡張し、リターン予測力を高める特徴量を設計します。

- ・**高精度なボラティリティ推定**：従来の終値ベースの実現ボラティリティに代わり、**Yang-Zhangボラティリティ**など高度なレンジベース推定量を導入します。Yang-Zhang法は**オープン時のギャップと日中変動**の両方を組み込みバイアスを低減したボラティリティ推定であり、シンプルな終値法より約14倍効率的（分散推定誤差が小さい）とされています⁴。例えば**60日窓のYang-Zhang年率ボラ**を特徴量とすれば、大きな窓でも急激な変動を捉えやすくなります。同様に**Rogers-Satchell**や**Garman-Klass**といった他のレンジ系ボラティリティも算出し、既存のParkinsonボラ（高値-安値のみ利用）を補完します。これらは**終値間のギャップやヒゲ（高安）情報**を活用することで、より正確なリスク把握に繋がります。
- ・**ボラティリティ・オブ・ボラティリティ（Volatility-of-Volatility, VoV）**：ボラティリティ自体の変動性に着目した特徴量です。具体的には**短期ボラ比率**や**ボラの標準偏差**を導入します。例えば「5日ボラ / 60日ボラ」の比率や、直近N日間のローリングボラ変化の標準偏差を計算し、現在のボラ水準が平常時からどれほど**異常に変動しているか**を測ります。実際、**市場ボラの不確実性（vol of vol）**も**リスクプレミアム要因**となり得ることが知られており⁵⁶、急激なボラ変動局面では将来リターンに影響を与える可能性があります。VoVが高い＝不透明感が高い局面ではリスクテイク志向が低下し株価が押される、といった**第2のリスク因子**をモデルに織り込めます。具体的な実装としては、直近20日間程度の毎日の実現ボラ（例えば5日ATRなど）の値を取り、その標準偏差を算出して特徴量化します。
- ・**リバーサル×モメンタム相互作用**：短期リバーサル効果（ごく短期の過剰反動による逆戻し）と中長期モメンタム効果（トレンドの慣性）を組み合わせ、**非線形な相互作用特徴**として捉えます。例えば**1日リターンと20日リターンの積**や**5日リターンと60日リターンの積**などを特徴量とし、「直近急落したが中期では上昇トレンド」のようなケースを捉えます。これは、**上昇トレンド中の一時的押し目**（あるいは下降トレンド中の一時的戻し）を検知するものです。このような状況では押し目買い・戻り売りが有効なことが多く、**短期的な反発**が期待できます。モメンタムとリバーサルの符号が逆転する銘柄（例：20日騰落率はプラスだが1日リターンは大きくマイナス）は翌日以降上昇しやすい、などの経験則に基づきます。この相互作用特徴により、個別にリターン特徴量を見るよりも**条件付きの効果**をモデルに捉えさせます。
- ・**オーバーナイト vs. 日中リターン**：株価変動を**取引時間外の動き**（前日終値→当日始値）と**日中の動き**（当日始値→当日終値）に分解して特徴量化します。J-Quants APIでは前場・後場のデータも提供されていますが、ここでは簡便に**前日終値と当日始値の比**からオーバーナイトのリターン、当日始値と終値から日中リターンを算出します。これは、例えば**オーバーナイトで大幅ギャップアップした銘柄**は日中利益確定売りで伸び悩みやすい、といったパターンを捉えるのに有用です。また**夜間に情報が出て株価に織り込まれる**（海外市場の影響など）部分と、日本市場取引中の**需給による動き**を分離できます。具体的特徴量としては「直近5日分の平均オーバーナイトリターン」や「直近5日分の平均日中リターン」、あるいは**累積オーバーナイト vs 累積日中**の差分などを用います。過去研究では、オーバーナイトと日中のリターンパターンに持続性があることや、**日中より夜間の情報反映が早い**ことが指摘されており、これを利用する形です。

- ・**価格レンジ比率**：価格位置・ギャップ系特徴の拡張として、高安レンジ内での**終値位置**や**ギャップ率**の直近傾向を特徴量化します。既存では単日の「close_to_high/low」などがありますが、**連続的なレンジ動向**を捉えるために例えば「直近5日間で終値が日中レンジ上半分で終えた日の割合」や「過去1ヶ月でギャップアップした平均幅（始値と前日終値差の平均）」などを計算します。これは、**上昇局面ではクローズが高値寄りになりやすい**一方、トレンド転換局面では日中陰線が増える、といった現象を数量化する狙いです。また**ATR（平均的なレンジ幅）に対する当日値幅の比**や、前日比ギャップをATRで標準化した値なども、ボラ環境に対する当日の動きの**異常度**を示す特徴として考えられます。

以上の高度な価格・ボラ系特徴量により、**より精緻なリスク評価や微妙な価格パターン**をモデルが捉えられるようになります。例えばYang-Zhangボラで正確なリスク感応度を上げ、リバーサル×モメンタム特徴で**トレンド内逆張りの局面**を検知し、オーバーナイト/日中の分解で**情報源別の動き**を捉えることで、**短期予測の精度向上**を狙います。

2. 流動性・需給・投機性に関する特徴量

出来高や取引状況にまつわる**流動性・需給の指標**を強化し、株価変動の裏にある資金フローや投資家行動を捉えます。

- ・**アミフードの非流動性指標（Amihud Illiquidity）**：出来高と価格変動から計算される伝統的な流動性指標で、「1ドルの取引でどれだけ価格が動くか」を表します⁷。具体的には**日次の絶対リターンを取引代金で割った値**を平均したものがAmihud指標です⁷。J-Quants APIでは株価と出来高（および売買代金）が取得できるため、例えば**直近20営業日のAmihud指標平均**を計算して特徴量に加えます。値が大きいほどその銘柄は**少ない売買で価格が動きやすい（流動性が低い）**ことを意味し、投資家の需給インパクトが大きいことを示唆します⁸。一般に流動性が低い銘柄は価格のオーバーシュートやリバーサルが起きやすく、情報効率も低いいため**アルファ機会**が大きい可能性があります。モデルにこの非流動性特徴を入れることで、小型株など流動性不足銘柄への**特異な値動き**を捉える助けとします。
- ・**出来高スパイク指標**：異常な出来高増加を検知する特徴量です。具体的には**当日出来高を過去X日間の平均出来高で割った比率**や、そのZスコア（何標準偏差上か）を算出します。一定閾値以上のときにフラグを立てることも考えられます。出来高の急増は**新規材料に対する市場の関心集中や需給の偏り**を意味し、直後のリターンに影響を及ぼすケースがあります。例えば出来高激増＋上昇の場合は**ブレイクアウト（買いエネルギー顕在化）**の可能性が高く、短期上昇の継続につながるかもしれません。一方、出来高激増にもかかわらず株価下落の場合は**投げ売りのピーク**で反発しやすいといったシナリオも考えられます。モデルには出来高スパイクの**大きさ**（例：直近5日平均比）と**方向**（株価上昇or下落のフラグ）を組み合わせて供給することで、需給イベントを定量的に扱います。
- ・**信用取引残高（買い残・売り残）**：JPXのAPIから**日々の信用取引残高**（`/markets/daily_margin_interest`）や**週次の信用取引残高**（`/markets/weekly_margin_interest`）が取得可能です。信用買い残は**投資家の強気ポジション蓄積**、信用売り残は**弱気ポジション（空売り）蓄積**を表します。研究によれば、日本では信用取引は主に個人投資家によるものであり、**信用買いの動向が将来の短期リターンを予測**することが示唆されています⁹。例えば**信用買い残の急増**は個人の過熱買いを示し、その後の株価下落に繋がりやすい可能性があります。一方、信用売り残（空売り）の積み上がりは将来の買い戻し需要となり、上昇圧力となるケースも考えられます。具体的特徴量案としては「時価総額比の信用買い残高」「貸借比率（信用買い残/信用売り残）」「信用取引残高の直近1ヶ月増加率」などが考えられます。これらを**1日以上遅れで反映**（公表タイミングに注意）することで、**個人投資家のセンチメントやポジションの偏り**をモデルに加味します。

- **空売り動向指標**：J-Quants APIには**空売り比率**（業種別空売り比率 `/markets/short_selling`）や**空売り残高**（`/markets/short_selling_positions`）のデータも存在します。ここから個別銘柄または市場全体の空売り比率を特徴量にできます。たとえば「その銘柄の浮動株に対する空売り残高比率」や「空売り比率の直近1ヶ月平均」を計算します。空売りが高水準な銘柄は、**機関投資家などがネガティブ視している（割高と見ている）**ことを示唆し、実際将来リターンが低下する傾向があると報告されています¹⁰。つまり**高い空売り残高は弱気シグナル**と解釈できます¹⁰。一方で空売りが極端に積み上がった状態では、ポジティブな材料で**ショートカバーによる急騰**（空売り筋の買い戻し）も起こり得るため、空売り指標は他のモメンタム系特徴と組み合わせて解釈する必要があります。モデルには空売り比率そのものに加え、「空売り比率が過去xヶ月で最大かどうか」というフラグや、空売りの増減ペースも入れておく効果的です。

- **投機的需給イベント指標**：需給に関連する特殊なイベントも特徴量化します。例えば**毎年末の需給**（投信のドレッシング買い・解約売りなど）を捉えるために「12月最終週フラグ」を立てたり、IPO直後のリターン癖を捉えるために「上場直後〇日以内フラグ」を設ける等です。J-Quants APIから直接取得できる情報ではありませんが、上場日や市場変更日はマスターデータから得られますので、それを利用します。また**日銀ETF買い入れ**の影響を考慮し、過去の日銀ETF買いオペ実施日データがあれば「日銀買い入れ実施日の市場上昇率」を特徴量として持たせ、政策的需給も織り込むことが可能です（※本件はJ-Quants標準データ外ですが、公開情報で対応可能）。

以上の流動性・需給系特徴により、**流動性リスクや投機筋の動き**をモデルが理解できるようになります。例えばAmihud指標で価格インパクトの大きい銘柄を割り引いて評価したり、空売り比率上昇に注目して**ネガティブ材料の先行シグナル**を捉えたりできます。また信用買い残の増減で個人の過熱感を把握し、**逆張りの手掛かり**とすることも可能です⁹。需給要因は価格動向の裏付けとなる重要情報であり、これらの特徴量に加えることでモデルの説得力と予測精度が向上すると期待されます。

3. ファンダメンタル・イベント系特徴量

企業の業績や各種イベントに関するデータから、**業績サプライズやイベント効果**を捉える特徴量を提案します。J-Quants APIは財務情報や決算予定日、配当情報などを提供しており、これらを活用します。

- **決算サプライズと事後ドリフト**：決算発表における業績の**サプライズ（予想比）**を特徴量に組み込みます。具体的には、四半期EPSや営業利益について**市場コンセンサスとの差異（%）**を計算し、発表日翌日以降一定期間その値を保持する特徴量とします（発表前はゼロ、または直前の値を保持）。さらに**決算後の株価モメンタム**も捉えます。**ポジティブサプライズを出した銘柄は発表後も上昇が続くやすく、ネガティブサプライズ銘柄は下落が続く**という「ポスト決算アナウンスメントドリフト（PEAD）」現象が知られています¹¹。そこで、決算翌5日間の累積リターンなどを計算し、それを「決算後ドリフト」特徴量として保持します（こちらも次の決算発表まで値を固定）。これによりモデルは**直近決算の良し悪しとその余波**を認識できます¹¹。サプライズデータはJ-Quants APIにアナリスト予想が無い場合、自前で集計が必要ですが、**前年同期比成長率と事前の株価変動**から代理することも検討できます。
- **ガイダンス偏差・進捗率**：日本企業は通期業績予想（会社予想）を発表します。その**進捗率**（累計実績/予想）や、予想の上方・下方修正が重要情報となります。J-Quantsの財務APIで四半期業績と会社予想が得られる場合、**第〇四半期終了時点の進捗率**を算出し、同業他社や過去平均と比較した偏差を特徴量にします。例えば「2Q終了時点で通期計画の80%達成」はかなり保守予想かつ上振れ余地大と捉えられます。同様に**会社予想の修正（増額・減額）**イベントを特徴量化し、その発生日から一定期間フラグを立てます。増額修正は強い買いシグナル、減額は売りシグナルとなり得るためです。進捗率や修正情報はリーク防止のため**発表翌日から**モデルに反映させ、過去に遡っては使わないように注意します。

- ・セクター横断のファンダ異常検知：**企業業績や財務指標について**セクター内での相対的な異常度**を測る特徴量です。例えば**利益成長率のセクター内Zスコア**や**予想PERのセクター内順位**などを計算します。これは「同業他社と比べて異様に高成長/低成長」「割安/割高」といった情報を提供します。セクター平均や中央値との差を取ることで、市場全体のトレンドに左右されない評価が可能です。具体的には、「四半期営業利益YoY成長率」を各業種33業種毎にZスコア正規化し、+2以上なら**異例の高成長フラグ**、-2以下なら低成長フラグを立てる、といった形です。また**セクター平均からの売上高・利益率逸脱度**や**セクター内ランキング**そのもの（順位を[0,1]で表す）も考えられます。これにより、例えば**同業が軒並み減益の中で一社だけ増益**といったケースを捉え、マーケットシェア拡大などポジティブ要因のシグナルを漏らさないようにします。一方、これは**過剰反応の修正局面**も捉えます。つまり高成長銘柄は既に買われ過ぎて将来調整する可能性や、低成長銘柄は悪材料出尽くしでリバウンドする可能性もあり、モデルはこれら**極端な状況**を認識して適切な方向に重み付けできるようになります。
- ・決算発表前後ウィンドウのリターン特徴：**決算前後の価格動きを直接特徴量化します。例えば「決算発表日の前後5営業日の累積リターン」や「発表前日の異常値動きフラグ」を作ります。これは**未公開情報の漏洩や期待/不安による事前株価変動**、および**結果に対する即時反応**を捉えるためです。もし決算直前に株価が大きく上がっていたら良い決算を織り込んでいた可能性が高く、発表後は材料出尽くしで下がるかもしれません。逆に事前株価下落+実際は無難決算ならリバウンドする、といったシナリオです。モデルはこれら**直近の文脈**を特徴で与えられることで、決算直後の予測精度向上に寄与するでしょう。
- ・重要開示フラグ：**適時開示情報などで特に株価にインパクトを与える**重要イベント**の有無を特徴量とします。J-Quants財務情報には「重要開示事項フラグ」が含まれている可能性があります（仕様書に言及のあった「重要開示フラグ」）。これは例えば**特別損失計上**や**大型投資発表**など株価動きを伴いやすい開示に会社側がチェックを付けるものです。これを活用し、当該開示があった日から一定期間、**binaryなイベント特徴量**としてモデルに供給します。重要開示の内容如何によって株価は大きく変動しますが、少なくとも**平時とは異なる変動リスク**があることをモデルに認識させる意図です。内容まで判断させるのは難しいですが、重要開示フラグと直後のリターンを併用することで（プラス材料なら上昇、マイナスなら下落という反応を示すはず）、モデルが**イベントとその結果**を学習できる可能性があります。
- ・経過日数特徴（イベント距離）：**決算発表からの経過日数、配当落ちからの経過日数、上場からの日数など、各種イベントからどれくらい時間が経ったかを特徴量にします。例えば決算直後は情報効果が強いが日数経過とともに薄れる、配当落ちは直後に価格調整されるが1ヶ月もすると元に戻る傾向、など**時間減衰効果**をモデルに取り入れます。具体的には「前回決算から○営業日経過」を数値で与えたり、「次回決算まであと○日」（予定日ベース）を与えても良いでしょう。後者は**イベントリスクの近さ**を意味し、決算が近い銘柄は発表跨ぎ回避で値動きが鈍るなどの傾向を加味できます。

以上のファンダ・イベント系特徴量によって、モデルは**企業の根源的価値変化**や**イベントドリブンな価格変動**を捉えられるようになります。特に決算サプライズとその後のドリフト特徴により、**過小評価された情報に基づく持続的な株価トレンド**を捉えることが期待されます¹¹。またガイダンス修正や進捗率からは**将来の業績方向感**を推測できます。これらは価格系列のテクニカル指標だけでは得られない**独立した情報源**であり、多角的なモデル判断につながります。

4. 投資家フロー・センチメント特徴量

株式市場には様々なタイプの投資家が参加しており、その**売買フロー**は需給やセンチメントを反映します。J-Quants APIでは**投資部門別売買動向**（`/markets/trades_spec`）として、週次ベースで投資主体（海外投資

家、個人、機関等)の売買代金データが取得可能です。これらを日次に展開・加工し、センチメント特徴量を構築します。

- ・**海外投資家のフロー**：海外投資家(外国人)は日本株市場の需給を大きく左右する主体であり、その売買動向は重要な手掛かりとなります。具体的特徴量案として、「**外国人投資家の買越額(ネット)**」を市場全体に対する割合で算出します(例：その週の外国人買越額 / 東証全体売買代金)。さらに**直近数週間の平均買越額との偏差**や**累積フロー**も考慮します。例えば「直近4週合計の外国人買越額」をGDP比や時価総額比でスケールした値は、外国人が中長期で資金流入させているか流出させているかを示します。研究によれば**外国人の買越しは株価上昇と相関が高く**、大きな買い越しが続くときはマーケットに強気トレンドが生じやすいです¹²(実際、日本株の変動要因として円安と並んで外国人フローが指摘されます)。モデルには「外国人フローのZスコア(過去1年基準)」や「外国人フローの4週移動平均」などを日次で持たせ(週データは該当週の各営業日に同じ値を割当て等で展開)、**継続的な資金流入**を把握させます。
- ・**個人投資家のフロー**：個人(国内)の売買動向も特徴量にします。一般的に**個人投資家は逆張り志向**が強い傾向があり、相場下落局面で買い越し・上昇局面で売り越しになりやすいと言われます。このため個人の大量買い越しは相場底打ち(買い場)を、個人の大量売り越しは相場天井(売り場)を示唆する場合があります。特徴量としては「個人投資家ネット買越額の週次推移」「個人フローの移動平均乖離」などを用意します。例えば**個人フローの直近1ヶ月Zスコア**が+2を超える(大幅買い越し)場合に逆張りシグナルと捉える、等です。もっとも、近年は個人も多様化しており一概には言えませんが、**信用評価損益率**などと組み合わせると個人の追証売り局面を察知できたりします(これは高度なので本設計では割愛)。
- ・**投資主体間のシフト**：外国人・個人・信託銀行・事業法人など主要主体の売買シェア変化を特徴量化します。例えば「その週に誰が市場を主導したか」を示すため、**各主体の売買シェア(売買代金に占める割合)**を計算し、前週比で増減した割合を出します。これにより、**主体の入れ替わり**(例：外国人が一気に買い越し転換した週、など)を検知します。特に外国人と信託銀行(年金など)は継続フローになりやすく、**主体転換はトレンド転換のシグナル**にもなり得ます。また**主体間のフロー相関**を見るために、「外国人買い越しかつ個人買い越し」のような組み合わせフラグも考えられます。通常、個人と外国人は逆のケースが多く、両方買い越しとなるのは稀ですが起きた場合は市場に相当強いポジティブ材料があったと推測されます。
- ・**資金フローのモメンタムと衝撃**：フローにもモメンタムが存在します。例えば**海外勢の買い越しが何週連続しているか**や、直近フローが**長期平均からどれだけ跳ねているか**(衝撃の大きさ)を特徴量にします。具体例として「外国人買越し連続週数」「先週の買越額が過去1年平均の何倍か」といった指標です。前者は**センチメントの持続性**、後者は**ショックの度合い**を示します。例えば10週連続で外国人が買い越ししている場合、強い上昇トレンドが形成されている可能性が高いです。一方、突発的に大きな買い(売り)が入った週はその直後に**短期的なオーバーシュートの修正**が起こる可能性があります。モデルはこれらフローの継続性・極端性を学習することで、**需給起因の過熱感**を捉えられます。
- ・**日次へのブリッジ**：週次データを日次に展開する際は、シンプルに各営業日に同じ値を持たせるほか、週内で直線補間したり、前週金曜から当週金曜にステップで変化させる方法があります。より精緻には、「週間ベースで判明するフローを各日にならす」ために**日次騰落率でウェイト配分**する方法も考えられます。例えば週初大きく上昇した場合はその日に多くの買いフローがあったと推定する、等です。ただし過剰に凝った推定はリークの可能性もあるため、本プロジェクトでは**単純な保持or補間**で十分でしょう。重要なのは**フロー情報を遅れなく織り込む**ことであり、発表の翌営業日からその週データを利用するようタイミング制御します。

以上のフロー・センチメント特徴により、モデルは**需給トレンド**や**市場参加者のムード**を感じ取ることができます。特に外国人フローは日本株の中長期トレンドと密接に関わるため、そのモメンタムを加味すること

でトレンドフォロー/逆張り判断の精度が上がるでしょう¹²。また個人や他主体の情報も、マーケットの歪みを補正するのに役立ちます。センチメント要素を加えることで、価格・ファンダメンタルだけでは説明できない需給ギャップによる短期変動を捉える狙いです。

5. 指数・先物市場データと市場レジーム特徴量

株式個別銘柄の動きは、市場全体の動向や先物・オプションなどデリバティブ市場の動きにも影響されます。そこで市場指数や先物データを活用した特徴量、およびそれらから判定される市場レジームを提案します。

- **指数リターンとギャップ**：TOPIXや日経平均など市場ベンチマーク指数のリターンを個別銘柄特徴量として組み込みます。既存特徴量にも「市場リターン（1/5/10/20日）」がありますが、さらに**指数と先物の乖離（ベースス）**に注目します。**先物ベースス**とは、株価指数先物価格と現物指数の差で、理論上は金利や配当要因で決まるものの、需給によって変動します。ベーススが**プラス方向に拡大**（先物高、現物安）している場合、**先高観が強い**か裁定買い期待がある状態と考えられます。逆にベースス縮小やマイナスは弱気シグナルです。具体的特徴量として、「直近のTOPIX先物ベースス（先物価格-現物指数水準を現物値で割った%）」「ベーススの1ヶ月平均との差分」を算出します。これにより**裁定取引の圧力**や**市場センチメント**を測れます。
- **先物出来高・建玉**：J-Quantsの先物OHLC API（`/derivatives/futures`）からは先物取引高や建玉残高（取組高）も取得できる可能性があります。**先物の出来高**は投機資金の参入度合いを示し、急増時は相場転換点となることもあります。**建玉残高（オープンインタレスト）**はポジションの積み上がり具合で、これが増加傾向で価格上昇なら「新規資金流入による上昇」、価格上昇なのに建玉減少なら「ショートカバー主体の上昇（持続力低い）」と判断できます。特徴量として、「先物建玉の直近20日増減率」や「価格変化に対する建玉変化の相関」を算出します。具体的には**先物価格の10日騰落と建玉10日増減が同方向か逆方向か**でフラグを作ることも有効でしょう。逆方向（例：価格上昇だが建玉減少）は**買戻し局面**を意味し、その後伸び悩む可能性があります。一方、共に増加は**新規マネー流入**でトレンド継続に好ましいシグナルです。
- **オプション・ボラティリティ指標**：JPX APIには日経225オプションのデータもあります。ここから**インプライド・ボラティリティ指数**（日本版VIXに相当）を取得できれば理想的です。仮に無い場合でも、**IV水準**を示す指標を代替的に使えます（例えばATMオプション価格と残存日数から概算IVを推定するなど）。**ボラ指数**は市場参加者の予想する**先行き不安度**を表し、急上昇時は株価急落局面と重なります。特徴量としては「日経VI（恐怖指数）」や「VIの20日平均乖離」を用います。また**IVとHV（ヒストリカルボラ）の差**（ボラティリティリスクプレミアム）も特徴となり得ます。これは**IV>HV**であれば**将来不安が強くリスクオフ**、マイナスなら楽観的とも読めます。オプションIVデータはもし利用できれば強力ですが、無い場合は本特徴量はスキップします。
- **市場レジーム判定**：上述の各種情報を組み合わせて、市場の状態をいくつかに分類し、その**状態を示す特徴量（ダミー変数）**を作ります。例えば以下のような簡易レジーム分類が考えられます。
 - **ブル相場**：市場指数が中長期MAを上回り、低ボラティリティかつ緩やかな上昇トレンド（例：TOPIXが200日線比+5%以上で、20日HV20%未満）。
 - **ベア相場**：市場指数が中長期MAを下回り、高ボラティリティで下降トレンド（例：200日線比-5%以下、20日HV30%以上）。
 - **レンジ/シュリンク**：低ボラ圏で方向感なく推移（20日HVが過去1年20%タイル以下かつ指数1ヶ月リターン±1%内）。
 - **高ボラ急変動**：ボラ急騰局面（VI急上昇 or HVが上位5%水準）。スクイーズからのブレイクアウト等も含む。

これらの状態を日次で判定し、one-hotエンコードした特徴量を与えます（**bull, bear, volatile, calm**などのフラグ）。既存の特徴量でも単純なレジーム（bull/high_vol等）は入っていますが、さらに先物やVI等を組み合わせ**強い材料が一致する場合のみフラグ**にするなど精度を上げます。レジーム特徴量を持たせることで、モデルは**現在の市場環境**を認識し、例えば弱気相場ではバリュースコア指標の効きが悪くなる等の**特徴量効果の文脈依存性**を学習できます。

- ・**他国市場・クロスアセット指標**：グローバルな市場連動も考慮します。日本株は**米国株や米金利、為替**などに影響を受けるため、外部指標をいくつか導入します。詳細は後述の外部データ特徴量で述べますが、レジーム判定時に**米国株との相関**等も考慮可能です。たとえば米国市場が弱気入り（S&P500が調整局面）かどうかで日本株への波及度合いを変える、といった発想です。

これら指数・先物・レジーム系特徴量により、モデルは**個別銘柄を取り巻くマーケット全体の状況**を把握できます。特に先物ベースや建玉からは、現物市場に現れにくい**機関投資家のセンチメント**を読み取れます。また市場レジームフラグは、モデルが**異なる市場環境下で適切な判断**を下すのに役立ちます。例として、高ボラ・弱気相場ではモメンタムよりも品質やバリュースコア指標を重視し、低ボラ相場では逆張り指標の信頼度を下げる、といった具合です。これらの特徴量として与えることで暗黙的に条件分岐を表現し、マルチホライズン予測の安定性を高めます。

6. コーポレートアクション・特殊イベント特徴量

個別銘柄固有のイベント（**株式数の変化や株主構成変化**など）は株価に影響を及ぼします。これら**コーポレートアクション**やその他イベントを特徴量化します。

- ・**配当イベント**：配当落ちや権利付き最終日の影響を考慮します。権利落ち日は理論上配当分株価が下がりますが、その前後で**配当再投資買い**や**需給反転**が起こり得ます。特徴量として「配当落ち直後〇日間」のダミーや、「配当利回り」（高配当銘柄は権利取りで権利日前に上がりやすい）を持たせます。J-Quantsの配当金情報API（`/fins/dividend`）から各銘柄の配当予想や権利日が取得できるので、それを利用し**権利付き日までの日数**や**直後の日数**を特徴にします。また**予想配当利回り**（予想1株配当/株価）も計算して特徴量に加えます。高利回り銘柄ほど配当狙いの買いが入り下支えされやすい反面、配当落ちで急落しやすい傾向もあるため、モデルがその性質を学べるようにします。
- ・**株式分割・併合**：株式分割は流動性向上や名目株価引き下げにより個人買いが増える傾向があり、発表後に株価が上がりやすいイベントです。J-Quantsのコーポレートアクション情報で株式分割/併合を取得し、**発表日から効力発生日前後**にフラグを立てます。具体的には「株式分割発表後1ヶ月間」のダミーや、「分割比率」を特徴量化します（大きな分割ほどインパクト大）。また併合（株数減少）は逆に**実質株価上昇**となるため需給はネガティブです。これらイベントフラグにより、モデルは**分割による株価上昇バイアス**や**併合による下落バイアス**を織り込めます。
- ・**インデックス採用・除外**：TOPIX等の指数への新規採用や除外も株価に影響します。採用が決まると指数連動ファンドからの**買い需要**で上昇しやすく、除外は売り需要で下落しやすいです。ただし発表から実施まで間があり、その間に織り込まれるケースが多いです。J-Quants Proには**指数構成変更**のデータセットもあるようですが、一般的にはプレスリリース等から情報取得が必要です。本プロジェクトでは、例えば**TOPIX浮動株比率見直し**や**日経平均入替**など大きなもののみ手動でデータを用意し、該当銘柄にフラグを付与する方法も考えられます。現時点ではそこまで踏み込まずとも、**出来高急増**や**株価急騰/急落**で間接的にモデルが捕捉できる可能性もありますが、可能なら特徴量として明示すると良いでしょう。
- ・**IRニュース/適時開示テキスト解析**：テキスト由来の特徴量として、大きな株価材料となりうるニュースに関するフラグを検討します。適時開示情報（TDnet）の見出しや内容をテキスト分析できれば理想ですが、本プロジェクトでは簡便に**キーワードマッチ**で特徴量化することを提案します。例えば開示情報のタイトルに「**上方修正**」「**増配**」「**大型受注**」「**M&A**」等ポジティブワードが含まれれば

+1、**下方修正**や**減配**等ネガティブワードなら-1、といった**サプライズイベントスコア**を発表日に付与します。これを一定期間（3日～1週間程度）減衰させながら特徴量として残すことで、ニュースの影響期間をカバーします。J-Quants API自体が適時開示のテキストまでは提供しないかもしれませんが、重要開示フラグや株価急変動で代理させる手もあります。テキスト解析は難度が上がるため必須ではありませんが、もし実装できれば**定性的材料**を組み込めるメリットがあります。

以上、コーポレートアクションや特殊イベントに関する特徴量で、**価格変動要因の取りこぼし**を防ぎます。モデルはこれら特徴により、例えば「近日大型分割予定だからこの株は需要増で強いだろう」とか「指数除外決定で需給悪化」といった**人間の判断に近い要素**も考慮できます。配当利回りは特に**金利環境との相対評価**にも繋がるため、後述の金利動向と合わせて、**バリュー的側面**も補強できるでしょう。

7. 外部マクロ・クロスアセット特徴量

最後に、**株式市場の外部環境**を表す特徴量を提案します。金利・為替・コモディティ価格・海外株式指数など、J-Quants API外のデータも活用しますが、これらは日本株の動きと密接に関連するため組み込む価値があります¹²。

- **金利関連**：日本の長期金利（10年国債利回り）や米国金利（米10年債利回り、FF金利先物など）を特徴量に加えます。一般に**金利上昇は株式にはマイナス**（将来キャッシュフローの現在価値低下、企業調達コスト増）とされます。ただし¹²によれば、日本では過去低金利下で**国内金利の株価への直接影響は小さい**との指摘もあります。その代わり**為替や米株の方が影響大**でした¹³。とはいえ、直近では日本もYCC修正などで金利変動しつつあるため無視できません。特徴量として、「日本10年債利回りの変化率（1ヶ月）」「米2年-10年スプレッド」などを入れます。金利と株価の関係は時期によりませんが、**金利上昇＝金融株に追い風/ハイテクに向かい風**などセクター感応度もあります。高度にはセクター×金利の相互作用特徴を作る手もありますが、まずは全体への影響を見るため単体特徴で十分でしょう。
- **為替レート**：ドル円為替レートは日本株特に輸出関連に大きな影響があります。一般に**円安（ドル高）は日本株にプラス**です¹⁴。円安は輸出企業の競争力を高め、外貨建収益を押し上げるため、株価も上がりやすくなります¹⁴。特徴量として、「USD/JPYの前日比%変化」や「直近5日間の円安傾向（5日で何%円安）」等を入れます。また**実効為替レート**や他通貨も見ると、「人民元やユーロ対円レート」も考えられますが、まずUSD/JPYが代表でしょう。モデルは為替変動特徴を使って、例えば急激な円高局面では輸出株の下落を予測に織り込む、といった対応が期待できます。
- **商品価格**：**原油価格**や**商品指数（CRB指数）**などコモディティ価格も特徴量に加えます。原油は資源・素材株の収益に直結し、また全般にインフレ指標でもあります。例えばWTI原油先物価格の**1ヶ月騰落率**や**国内ガソリン価格**の変化を入れます。資源高は商社や石油株にはプラス、電力や素材コストにはマイナスという具合に銘柄によって作用が異なりますが、市場全体では**適度な上昇は景気回復シグナル**、急騰は**コスト高懸念**となります。このためモデルには一次的には全銘柄共通特徴として入つつ、銘柄別の影響はモデル内部で学習させます。また金や銅など指標的商品もあれば入れてもよいですが、複雑になるので主要な原油だけで十分かもしれません。
- **海外株式指数**：**米国株（S&P500やNASDAQ）**の動向は日本株にとって重要な先行指標です¹²。米市場の大幅上昇・下落は翌日の日本市場にも高確率で波及します。特徴量として、「前日S&P500先物の夜間変動（日本時間早朝までの変化）」や「NASDAQ指数の5日リターン」などを入れます。特に**米国ハイテク市場との連動**を見るならNASDAQ、**全般的リスクオン/オフ**を見るならS&P500かNYダウです。J-Quantsには海外指数は無いので外部取得になりますが、精度向上のため是非組み込みたい情報です。また**中国株（上海指数）**など地域の主要市場も影響大きいですが、まず米国中心で良いでしょう。モデルはこれらを使い、例えば米株急落していれば日本株個別銘柄の予測も総じて弱気にシフトさせる、といった**グローバルリスク要因**を反映できます。

- ・**ETF資金流入出**：もしデータ入手可能なら、**日本株ETFへの資金流入額**なども考えられます。海外投資家動向と被る部分もありますが、特に**特定セクターETF**のフローなど細かく見ると、どのテーマに資金が向かっているか分かります。本プロジェクト範囲では難しいため、ETFフローはマーケット全体の外国人フローで代替する形とします。

これら外部要因特徴によって、モデルは**国内市場の外側から働く力**を認識できます。例えば、**円安進行＝輸出株の将来EPS押上げ要因**としてモデルが学習し、円安局面ではそれら銘柄のtarget上昇を予測する、というような挙動が期待できます¹⁴。また米国株の下落を捉えて保有株全般のtargetを引き下げれば、実際の翌日下落を当てやすくなります¹²。特に昨今はグローバル市場連動性が高いため、**海外指数や為替を使った特徴量**は予測精度向上に不可欠と考えられます。

おわりに：新規特徴量導入による効果と実装上の注意

以上、多角的な新規特徴量を提案しました。これらを既存のOHLCVテクニカル指標群やファクター群に追加することで、モデルの性能向上が期待できます。特に**ボラティリティ精度向上や需給・センチメントの取り込み、イベント反応の捉え**など、現在不足している側面を補完する役割を果たします。

実装に際しては以下の点に注意します。

- ・**リーク防止徹底**：週次・月次データは**公表日を意識**して特徴量に適用し、将来の値を過去日に誤って入れないようにします。例えば「外国人フロー当週値」は週末にしか分からないので、その週の月～木には前週値を使い、金曜夜または翌週月曜から当週値を使う等の工夫をします。決算情報も発表翌日から有効にするタイミング管理が重要です。
- ・**正規化と欠損**：新規特徴量も、学習データ内で計算した**分布（平均・標準偏差）**で標準化し、外れ値はクリッピングまたはtanh縮小など行います。クロスセクションでの標準化に加え、期間を跨ぐ特異値（例えばコロナショック時のVI急騰など）は適宜上限を設けます。また銘柄ごとに利用可否が異なる特徴（例：新興市場銘柄には決算予想がない等）は**欠損値処理**をします。欠損は0埋めではなく、**無効フラグ**を別途立ててモデルに認識させるか、あるいは平均値で埋めつつ上記フラグで対応します。
- ・**特徴量選択**：提案量は多岐に渡りましたが、全てを投入すると過学習リスクもあります。相関の高い特徴同士は主成分的に圧縮するか、一部は**モデル側の正則化（L1など）**に頼る形でも良いでしょう。経験的に重要度が高そうなもの（例えば為替、外国人フロー、IV、水準ボラなど）から優先実装し、順次効果を検証していきます。LightGBMのゲインやシャープ値改善度などでフィルタリングするアプローチも有効です。

最後に、本提案の特徴量を用いることで**RankICやシャープレシオといった評価指標の改善**が見込めます。例えば5日先RankICについて、従来価格系テクニカルのみでは0.02程度だったものが、ファンダ・フロー特徴の追加で0.03台に上昇し、情報比（RankIC_IR）も向上することが期待されます。またシャープレシオも、下落局面での予測精度向上によりドローダウンが抑制されれば改善するでしょう。これら仮説は実装後のワークフォワード検証で確認します。

以上、J-Quants APIで取得可能なデータをフル活用した新規特徴量の提案を行いました。価格・テクニカル指標の精緻化からファンダメンタル、投資家フロー、外部環境まで網羅することで、モデルの**アルファ捕捉能力**が大きく高まることを期待します。各特徴量は漏洩に注意しつつ適切に正規化して組み込むことで、堅牢かつ高性能な予測モデルの構築に寄与するでしょう。

参考文献・出典：本提案内容の一部には以下の情報源を参照・引用しています。

- ・Yang-Zhangボラティリティの有効性⁴
- ・Amihud流動性指標の定義⁷
- ・信用取引（個人投資家）のリターン予測力⁹

- Post-Earnings Announcement Drift現象 11
- 空売り筋の情報優位性 10
- RankICおよびSharpe比の指標解説 1 2
- 円安が企業収益・株価に与える好影響 14
- 日米における株価決定要因（為替・米株 vs 金利）の分析 12

4 “Yang Zhang is a historical volatility estimator that handles both opening jumps and the drift and has a minimum estimation error... Yang-Zhang volatility... considered being 14 times more efficient than the close-to-close estimator.”

7 “Amihud (2002) proposes to measure illiquidity by the ratio of absolute stock return to its dollar volume... interpreted as ‘the daily price response associated with one dollar of trading volume’ 8.”

11 “The returns of good news firms continue to drift up and those of bad news firms drift down. It is to date one of the most robust market “anomalies” in the stock [market].”

9 “Margin trading ... dominated by individual investors in Japan... significant cross-sectional relationship between margin buying and stock returns... margin buying helps predict future stock returns, especially for small-firm stocks at short horizons 9.”

10 “studies by Dechow... and Takahashi (2010) suggest that short sellers are indeed skilled at selecting companies that exhibit lower expected future returns 10.”

14 “The weakening yen has provided a dual benefit to Japanese companies’ earnings... enhanced the competitiveness of exporters... and boosted earnings directly through favorable translation of foreign revenues 14.”

12 “finds that interest rates have not impacted Japanese stock prices but exchange rates and U.S. stock prices have 12.”

1 “Spearman rank IC... is the Pearson correlation coefficient between the ranked factor scores and ranked forward returns, ... less sensitive to outliers and can therefore provide a more accurate measure of a factor’s predictive power 1.”

2 “The Sharpe ratio... compares the return of an investment with its risk... numerator is the difference between returns and a benchmark (risk-free rate)... denominator is the standard deviation of returns... a measure of volatility 2.”

- 1 Quantitative Investing - CFA, FRM, and Actuarial Exams Study Notes
<https://analystprep.com/study-notes/cfa-level-iii/quantitative-investing/>
- 2 Sharpe Ratio: Definition, Formula, and Examples
<https://www.investopedia.com/terms/s/sharperatio.asp>
- 3 株価四本値(/prices/daily_quotes) | J-Quants API
https://jpx.gitbook.io/j-quants-ja/api-reference/daily_quotes
- 4 Yang-Zhang Volatility | PortfoliosLab
<https://portfolioslab.com/tools/yang-zhang>
- 5 6 The importance of volatility of volatility | Macrosynergy
<https://macrosynergy.com/research/the-importance-of-volatility-of-volatility/>
- 7 8 ba-odegaard.no
https://ba-odegaard.no/teach/notes/liquidity_estimators/amihud_estimator/amihud_lectures.pdf
- 9 Can margin traders predict future stock returns in Japan?
<https://ideas.repec.org/a/eee/pacfin/v17y2009i1p41-57.html>
- 10 Short-sale inflow and stock returns: Evidence from Japan | Request PDF
https://www.researchgate.net/publication/257211637_Short-sale_inflow_and_stock_returns_Evidence_from_Japan
- 11 Post-Earnings-Announcement Drift and Analyst Forecasts
<https://www.anderson.ucla.edu/documents/areas/fac/accounting/drift503.pdf>
- 12 13 Relationship between Exchange Rate and Stock Prices during Quantitative Easing Policy in Japan
<https://ijb.cyut.edu.tw/var/file/10/1010/img/852/V114-3.pdf>
- 14 Will the Yen Continue to Support Japanese Equities? - Cambridge Associates
<https://www.cambridgeassociates.com/insight/will-the-yen-continue-to-support-japanese-equities/>