

データ駆動科学的手法を用いた企業群株価予測特性の定量化

日大生産工(学部) 白石 健也 日大生産工 佐々木 真

1. 背景

株価の動向を正確に予測することは金融取引において重要な役割を持つ。株価の動向を正確に把握・予想することで投資利益率を高くする。そのためには株価の変動に対して安定した予測精度が必要である。しかし、株価変動は多数の要因により複雑に変動するため予測は難しい。また、株価予測には常に暴落などのリスクを考慮する必要がある。これらの株価解析・予測は様々な方法で行われてきた。例えば線形時系列モデルに対し株価の短期予測を可能とした自己回帰和分移動平均モデル(AutoRegressive Integrated Moving Average, ARIMA)¹⁾や非線形や長期依存性を捉え予測精度を向上させた長短期記憶(Long Short-Term Memory, LSTM)²⁾などと株価時系列データの解析・予測は様々な方法で進んでいる。

これらの手法は単一の銘柄に対しての予測であるため、他企業や市場全体の動向といった外的要因による急な変動を十分に捉えられないという課題がある。そこで特定の企業群に属している株価はある程度相関して変動すると期待される。企業群内の株価変動を仮想空間における場の変動ととらえデータ駆動科学的手法を用いて複数の株価に対して予測する。これにより特定の株価の暴落などのリスクを分散できると期待する。

本研究では、複数の株価の時系列予測を可能とする手法の予測精度の定量化を行う。

2. ハンケルスパース動的モード分解

株価変動を仮想空間における場の変動ととらえ、時系列データとして株価予測を行う。そこで流体などの時間の進みにより変化する複雑なデータを解析する手法で使われる動的モード分解(Dynamic Mode Decomposition, DMD)を使い株価予測を行う。動的モード分解の支配方程式は以下になる。

$$\frac{dX}{dt} = F(X, t) \approx AX \quad (1)$$

動的モード分解は時空間における複雑な非線形の観測データ $F(X, t)$ から短い時間スケールで線形近似し時間発展算子 A と観測点 X との積により予測が可能となる³⁾。しかし、非線形に対して線形近似するため長期的な予測が困難である。そのためハンケル行列とスパース化を併用する。ハンケル行列 H の式を以下に示す。

$$H = \begin{bmatrix} \varphi(t) \\ \varphi(t + \Delta t) \\ \varphi(t + 2\Delta t) \\ \vdots \\ \varphi(t + (N-1)\Delta t) \end{bmatrix} \quad (2)$$

時系列データ $\varphi(t)$ を時間方向にスライド幅 Δt ずつずらしデータ層として $N-1$ 個積み重ね、行列の列数を増やす。そうすることで Δt 分先の変動をある地点の時系列データとして見かけ上増し、長期的な特徴を捉える。スパース化の式を以下に示す。

$$\min f(x) + \lambda \sum_{i=1}^n |w_i| \quad (3)$$

左の項 $f(x)$ は株価データと予測データの一一致度を評価し、右の項で L1 正規化によって重要なモードに対応する係数 w_i のみを残し、それ以外は 0 に近づける。左右の項の和を最小化することで最適なモードのみを残す。これらよって得られたモードをさらに最適なモードのみにフィルタする。これにより短期成長するモードやノイズを除去でき、過学習による予測精度の低下を抑制する⁴⁾。これらを併用し株価予測を行う。

3. 予測精度の定量化

ハンケルスパース動的モード分解は学習期間の長さ、ハンケル行列の大きさにより予測結果が大きく変わる。そこで株価予測の予測精度を定量化しパラメータの選定を容易にした。評価方法はスピアマン順位相関関数と構造的類似性指数(SSIM)と平均二乗誤差(RMSE)を併用した。スピアマン順位相関関数は二つの時系列データが同じ方向に変動しているかを評価する。

$$\rho_s = \text{corr}(r_x, r_y) \quad (4)$$

時系列データを要素の大きさで順番に並べたデータ r_x と r_y の相関を $-1 \sim 1$ の間で表す⁵⁾。構造的類似性指数(SSIM)は見た目の構造がどれくらい似ているかを評価する。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

データの大きさの平均 μ とばらつき σ をもとに比較して $-1 \sim 1$ で評価する⁶⁾。(C は安定化定数である) 平均二乗誤差(RMSE)は予測値と実測値の誤差を表す。これら以下のように併用する。

$$\text{score} = \rho_s \times SSIM(x, y) - RMSE \quad (6)$$

これにより、score が 1 に近いほど高精度の予測という予測精度の定量化をした。

4. 予測結果

2025年2月1日から8月31日までの自動車・電気機器・半導体関連の44社を使い予測をした。株価データをFig.1に示す⁷⁾。

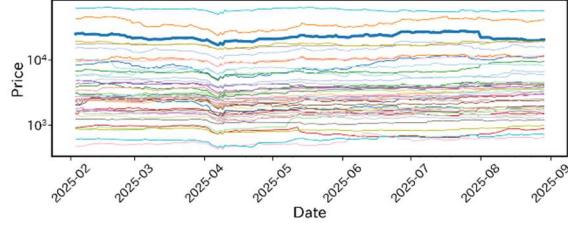


Fig. 1 自動車・電気機器・半導体関連 44 社の株価変動グラフ。縦軸は株価(日本円)の対数スケール、横軸は日付(2025年2月1日～8月31日)を示す。

予測に使用した全期間の株価グラフである。縦軸は株価価格(価格の差が大きいため対数化)、横軸は日付でプロットした。これらの取得したデータを使用し予測精度について評価した。ハングル行列のスライド幅 Δt を 2 で固定し、学習期間の長さとハングル行列の行数を変化させ、各 44 社で予測開始から 7 日間について score を計算し、予測精度を定量的に評価した。Fig.1 の青太線の株価の予測精度を評価した結果を Fig.2 に示す。

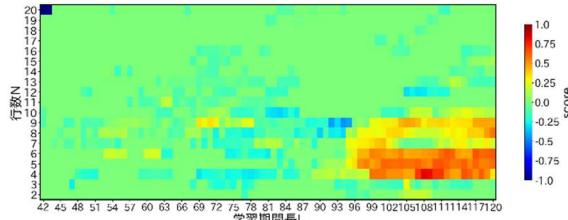


Fig. 2 予測開始から7日間における評価結果(score)。縦軸はハングル行列の行数 N、横軸は学習期間の長さ L を示す。

縦軸はハングル行列の行数N、横軸は学習期間の長さLである。scoreが1に近いほど赤色になり予測精度が良いことを示す。Fig.2の中で評価結果の良いパラメータをscore=0.815、学習期間Lを108、行数Nを4のときの予測結果をFig.3に示す。

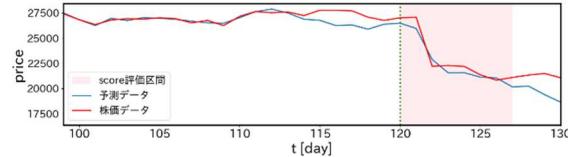


Fig. 3 評価結果(score)より選定したパラメータ ($L=108$, $N=4$, $score=0.815$) の株価予測と実際の株価データの比較。縦軸は株価価格(日本円)、横軸は2月1日から何日後を示す。

縦軸を株価価格、横軸を2月1日から何日後かを示し、実際の株価データ(赤のプロット)と予測データ(青のプロット)を比較する。120

日(緑の点線)から左が再現、右が未来予測である。120日から赤色の範囲が score で評価した範囲である。

Fig.2により score の評価を視覚的に比較でき、高精度のパラメータの選定を容易にすることができた。また、Fig.3の実際の株価データと未来予測の誤差は 0.0302 となり、下落の変動にも対応した予測可能であると言つたことがわかった。Fig.2 図上より行数 N が 4~6、学習期間長 L が 96~120 の範囲に score の結果に偏りが見られた。

5.まとめ

本研究では、企業群で株価を選定し、複数の株価を同時に学習することで、リスク分散ができると考え研究を行った。その結果、ある程度の株価の下落にも対応した予測が可能であることがわかった。また、score の結果による評価方法により予測精度の高いパラメータを定量的に評価した。これにより予測精度が数値的に得られパラメータの選定が容易になった。一方で、score の結果に偏りがみられた。この偏りに対して法則性を見つけることでより実用的なパラメータの選定が可能となると考える。

今後の課題として、今回の結果以上の株価の急変動に対応していくために使用する株価の企業群絞り込み方や銘柄の数を変えてより高精度の株価予測を実現する。また、今回のパラメータの選定は予測結果をもとに行っている。しかし、実運用となると予測データと比較する株価データがないため実用的ではない。そのためパラメータによる評価結果や予測を再現するところでの特徴や法則性見つけ出すことが必要である。それにより実用的なパラメータの選定が可能になると考える。

参考文献

- 1) Ayodele A. Adebiyi, Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 105–111, (2014)
- 2) Prashant Pilla, arXiv preprint, arXiv:2501.17336v1, pp. 1–15, (2025)
- 3) Schmid, J. Fluid Mech., Vol.656, pp. 5–28, (2010)
- 4) M. R. Jovanoviet al., Phys. Fluids 26, 024103 (2014).
- 5) Dominik Janzing, arXiv preprint, arXiv:2012.06287, (2020)
- 6) Zhou Wang, IEEE Transactions on Image Processing, Vol.13, pp. 600–612, (2004)
- 7) Yahoo! ファイナンス, https://finance.yahoo.co.jp/?utm_source=chatgpt.com