

株価データベースに対する周期変動の パターンマイニングとその評価

Periodic Fluctuation Pattern Mining in a Stock Price Database and its Evaluation

林 大祐[†]
Daisuke Hayashi

湊 真一[‡]
Shin-ichi Minato

羽室 行信[§]
Yukinobu Hamuro

岡田 克彦[¶]
Katsuhiko Okada

1. はじめに

近年ビッグデータを金融分野に活用する技術が注目されており、これまで明らかにされてこなかった個別銘柄レベルのマーケットアナリシーについても検証されるようになってきた [3]。本研究も大規模な株価データを扱いながら、パターンマイニングによる株価変動の特徴量抽出を目的としている。とりわけ「週」単位の変動に着目し、曜日によって発現する株価変動の特徴量をモデル化、収益の予測可能性の有無を検証する。ファイナンス領域の先行研究では、世界の株価指数の変動パターンには、週単位において特徴があることが報告されている [4]。株価指数は個別銘柄によって構成されるものであり、指数における週単位の変動の特徴量は個別銘柄により色濃く反映されているはずである。ここでは全上場銘柄の週単位の株価変動パターンを網羅的に抽出し、銘柄ごとに有利だと考えられる特徴を選び投資するモデルを提案する。いくつかの候補モデルの中から、学習する過去参照週数とシャープ比下限の組み合わせにより最適と考えられるモデルの結果について報告する。

2. 準備

2.1 使用データ

本研究では、約 15 年間 (2001 年 1 月 ~ 2015 年 5 月) の約 4000 銘柄についての時々刻々の売買取引の記録ではなく、各銘柄の日々の株価データ (始値、最高値、最安値、終値) を表形式で記録した物 (総行数:13,117,361 総バイト数:661,464,487) を使用した。

2.2 超過収益率 (AbnormalReturn)

超過収益率 (AbnormalReturn) とはファンド運用において期待水準 (ベンチマーク) を超えた収益割合のことを指す。本実験では TOPIX をベンチマークとした超過収益率を使用している。TOPIX は東証株価指数とも呼ばれ、1968 年の株式市場の時価総額を 100 とした場合の現在の時価総額を表す指標である。ある 1 銘柄の 1 日 (第 t 日) あたりの超過収益率を $aReturn(t)$ 、銘柄、TOPIX の終了価格をそれぞれ $close(t)$ 、 $topix(t)$ とすると (1) で 1 銘柄について第 t 日の超過収益率が算出される。

$$aReturn(t) = \frac{close(t)}{close(t-1)} - \frac{topix(t)}{topix(t-1)} \quad (1)$$

本研究では株式売買を日々の終了価格においてのみ行うという制約のもと、1 銘柄あたりでの前日比較による超過収益率の算出を行った。

2.3 シャープ比 (SharpeRatio)

シャープ比 (SharpeRatio) とは、ポートフォリオにおける超過収益率がそのリスクに占める比率を指し、本研究で言えば、提案手法がリスクに見合うだけの収益をあげているかどうかの指標である。利回りの同じポートフォリオでは、シャープ比が高いほど安定していると言える。ある 1 銘柄の 1 日 (第 t 日) あたりの超過収益率とその標準偏差、シャープ比をそれぞれ $aReturn(t)$ 、 $std(t)$ 、 $Sharpe(t)$ とすると $Sharpe(t)$ は (2) で表せる。

$$Sharpe(t) = \frac{aReturn(t)}{std(t)} \quad (2)$$

3. 提案手法

株式データに一連の処理を行い、投資モデルを構築するまでの基本手法について述べる。

3.1 各銘柄の売買パタンの全列挙

週単位でのマイニングを目論むため、使用する株価データ 15 年間 (2001 年 ~ 2015 年) に通して曜日番号と週番号を割り当てた。2001 年の第 1 週を 0 とし、2015 年に至るまで 750 週分の週番号を与えた。1 週間は 5 営業日で構成されている。本研究では週単位の株価の変動パターンを網羅的に学習するのであるが、ある銘柄を買ってから売りに出すまでの最大期間は 1 週間としている。つまり売りが次週にわたり行われることは許すが、その期間が 1 週間を超えることは許してはいない。従って 1 週間あたり、各銘柄の売買パターンは 25 通り存在する。約 4000 銘柄について 25 パタンの超過収益率を算出し全列挙した。(表 1) 例えば、ある銘柄の株式を $start$ 曜日に買い、 end 曜日に売るパターンを考える。このパターンにおけるシャープ比を $Sharpe(start, end)$ と定義する (3)。

$$Sharpe(start, end) = \sum_{k=start}^{end} aReturn(k) \quad (3)$$

3.2 提案する投資手法

全銘柄の売買パターン 25 通りが全列挙されているファイル群からモデルを構築する。本研究で提案している投資手法は非常にシンプルなものである。週 n において、過去 k 週間 (週 $t-k$ から週 $t-1$) において $Sharpe(start, end)$ の平均が最大となるような $start$ と end を求め、翌週 $t+1$ に $start$ 曜日で購入し end 曜日で精算するという方法である (表 2)。

4. 実験

本研究では大規模データ解析ツール MCMD[3] を用いてデータ処理し、モデルを構築した。モデルを実際のデータに適用し、シミュレーションするまでを実験セクションとする。

[†]北海道大学 情報科学研究科

[‡]北海道大学 情報科学研究科

[§]関西学院大学 経営戦略研究科

[¶]関西学院大学 経営戦略研究科

4.1 実験内容

直近の過去 k 週で最大のシャープ比を保有する銘柄毎の売買パターン (start 曜日に買い, end 曜日または end₁*曜日に売る) を抽出する。次週、銘柄毎に抽出したパターンでの実際の超過収益率を算出し集計することで週毎の超過収益率を算出し、モデルの有用性を週単位マイニングの過去参照週数 ($=k$) とシャープ比下限 ($=s$) の組み合わせによりシャープ比 s 以下のパターンは選択しない、最適と考えられる手法の評価を行った。

4.2 実験結果

実験結果を表 3 に示す。

5. 考察

実験で行った組み合わせの中では、 $(k, s)=(60, 0)$ の場合の超過収益率が最大値で 0.398 であった。提案手法で得られた結果から、全体としてマーケットに週単位の季節性変動が存在することは確認できた。ただ、ある時期モデルはうまく機能していない。株価がランダムウォークしている場合、いかなるモデルを用いたとしても超過収益率の獲得は望めないはずであるが、その理由として考えられるのがプロ投資家 (アービトラージャー) の存在である。仮に週単位の季節性が公知となれば、利益を最大化したいアービトラージャーは、先回りして同様に取引を行うことから、やがて期待していた超過収益率は獲得できなくなる。季節性のみならず、多くのマーケットアナリにはこうした傾向が常に存在し、実務的にはこうしたアービトラージャーの動向そのものを読み取る、新たな別の手法を提案する必要がある。

6. まとめと今後の課題

本研究では比較的単純な週単位のパターンマイニングでのモデル構築をし、どの程度の超過収益率が獲得できるのかについて実験し、実験結果を週単位マイニング過去参照週数とシャープ・レシオ下限比の組み合わせにより示した。今後の課題としては、実験結果では一見して好成績が出ているように見えるが、大型株中心に運用しても同様の傾向が出るのか、それとも、小型株特有の傾向が強くて好成績になっているのか、ということ詳しく調べることが必須である。そのため、アービトラージャーの動向を読み取り、超過収益率をより多くの (k, s) 組み合わせについて獲得できる手法を取り入れることが必要になる可能性がある。その他には、約 4000 銘柄すべてで行ったモデル構築を各銘柄毎に行い、銘柄ごとの収益率累積を勝ち負けのパターンで分類するといくつかのクラスに分けることができるかを確認してみる、各銘柄 25 パターンの中の最良パターンを必ず適応するのではなく、25 パターンのうち上位数パターンを選んで平均的に投資する戦略で実験するなど行っていきたいと考えている。

謝辞 本研究の一部は JSPS 科研費 15H05711 の助成による。

参照図表

表 1: ある銘柄の 1 週間の売買パターンと超過収益率収益率計算

月	火	水	木	金	月	火	水	木	金	収益率計算
買	売									aReturn(月, 火)
買		売								aReturn(月, 水)
買			売							aReturn(月, 木)
買				売						aReturn(月, 金)
買					売					aReturn(月, 月 ₊₁)*
:	:	:	:	:	:	:	:	:	:	:
				買		売				aReturn(金, 火 ₊₁)*
				買			売			aReturn(金, 水 ₊₁)*
				買				売		aReturn(金, 木 ₊₁)*
				買					売	aReturn(金, 金 ₊₁)*

* end 曜日は start 曜日の次週であることを意味する。

表 2: 直近 k 週でのパターンマイニングフロー

マイニングに用いる週**	抽出パターンを適応する週**
0,1,2,3,4..k-1	k
1,2,3,4,5..k	k+1
2,3,4,5,6..k+1	k+2
:	:
750-k+1..747,748,749, 750	751

** 各週は週番号で表記した。

表 3: 過去参照週数 k とシャープ比下限値 s による結果比較

k	s=0.0	s=0.5	s=1.0	s=1.5
5	0.161	0.161	0.151	0.121
10	0.200	0.211	0.148	0.044
15	0.242	0.255	0.107	0.044
20	0.271	0.290	0.151	0.098
40	0.335	0.376	0.197	0.178
60	0.398	0.369	0.269	0.243

参考文献

- [1] 岡田克彦, 羽室行信, 森田裕之, "周辺文脈アプローチを利用した新聞記事内容と株価に関する分析", 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集 2010, pp.128-129, [2010]
- [2] 前川浩基, 中原孝信, 岡田克彦, 羽室行信, "「言葉」にみる株価収益率の予測可能性について: 探索的アプローチによるモデル構築と out of sample の実証テスト", Business accounting review, 12, pp.117-130, [2013]
- [3] 岡田克彦, 羽室行信, Stephane, Cheung, "銘柄別期間全列挙に基づく季節性アナリのパターンマイニング" The 30th Annual Conference of the Japanese Society for Artificial Intelligence, [2016]
- [4] Jaffe, Jeffrey, Randolph, Westerfield, "The Week End-Effect in Common Stock Returns", The International Evidence, Journal of Finance, 40, 2, pp.433-454, [1985]
- [5] NYSOL:大規模表形式データ解析のためのオープンソフトウェアツール, <http://www.nysol.jp>