

日本株短期モメンタム投資における優位性領域の探索と予測モデル構築に向けた課題

分析の目的

株式市場において「長期的に機能する手法」は存在するのだろうか？

これについて、よく通説としてこんなことが言われる。

どんな手法であれ、それが本当に儲かるものであればすぐに市場参加者に広まり、やがて機能しなくなる。したがって一貫して機能する手法は市場には存在しない

しかし、これまでトレーディング実務を経験した中で、私はこの市場の前提について懐疑的である。その理由としては、1:優位性があると分かっているにもかかわらず心理的に実行困難なパターンがあること,2:実際に素晴らしい成績を上げ続けているトレーダーを生で見たこと、3:自身でも部分的ではあるが継続的な収益を手にした経験があることが挙げられる。こうした背景から、市場の前提を鵜呑みにせず、市場の可能性について自身の手で

探ってみることは有意義な試みだと考えた。

そこで、今回のレポートでは自身の経験の中で「長らく優位性を感じてきた手法（短期
モメンタム投資）」について分析を行い、その結果をまとめた。

ただしレポート内容は、あくまでもモデル構築前段階の「探索的データ分析（EDA）」
を行った経過、洞察をまとめたものである。レポート全体の概要としては「市場のどこ
に優位性が潜んでいて、それを収益に結びつけるためにどんな課題がありそうか」を可
視化・整理するような内容になっています。

分析の概要

日本株短期テクニカルトレードにおいて、東証上場銘柄（東証 1 部，東証 2 部，マザーズ、JASDAQ）の広い範囲で成立する汎用性の高い優位性領域が存在するか検証。（※
ただし運用資金 100 万円をスリッページ無しで売買するのに十分な流動性があること
を銘柄選択の前提とする）

分析方法

1:手法概要

2:売買ルール定義

3:データセット詳細

4:収益算出方法

5:データ前処理

6:分析に用いた説明変数一覧

1:手法概要

今回検証を行った手法は、株価の勢いを利用する「短期モメンタム投資」と呼ばれるものになります（イメージとしては上昇している銘柄を上がっている最中に買い、更に上がったところで売り抜ける）。この戦略の特性上、市場の売買が最も集中する「寄付」（日本時間 9:00）に目を向け、その中でも特に材料発表などファンダメンタルズの理由から当日の注文が一方方向に偏る「**特別気配スタート銘柄※1**」に絞って検証を行いました。

※1 特別気配：東証システムでは、材料や決算発表などにより注文が片方向へ殺到した場合、市場オープンと同時に価格決定させるのではなく、東証が定める「更新値幅ルール」に基づいて一定時間間隔（3分）ごとに現在価格を更新し、市場需給が一致したタイミングで価格を確定させる「板寄せ方式」が採用されている。

2:売買ルール定義

1:エントリー：特別気配スタートした銘柄に寄付きで順張りエントリー。（買い特別気配であれば買いポジション、売り特別気配であれば空売りポジションを持つ）。

2:決済：エントリーした当日の「大引け」で手仕舞い。（ストップ高でも当日の大引けで決済）。

3:損切り：「前日の終値」をロスカットポイント（損切りポイント）とする。前日の終値まで株価が逆行した場合、逆指値成行注文によりロスカット。

4:利益確定：利益確定の指値は設定しない。損益確定はロスカットポイントまで株価が逆行する or 当日の大引けに行う。

3:データセット詳細

5 年間(2015/1/1~2021/11/10)における個別株時系列データ（楽天証券から取得）を使用。これらを市場特性別に 4 つのデータセットに分類した（大型株、市場人気の乏しいボロ株、流動性条件を満たす銘柄群×2）。

1:data_market_capitalization_top100(サンプル数 n=100)

東証銘柄のうち時価総額 top100 銘柄。トヨタ、ソニー、キーエンスなど市場人気が高く、時価総額の大きい銘柄群で構成されている。株価指数（TOPIX, 日経平均）に組み入れられ、ファンド資金が豊富に入っていることが特徴。

2:borokabu_100_300(n=124)

いわゆるボロ株。株価 100 円-300 円未満を条件にスクリーニング(2021/11/10 時点)。市場人気の乏しく、時価総額が小さい銘柄群。ファンドや長期個人投資家の資金流入が少なく、個人投資家や短期トレーダーの資金が大きなウェイトを占めていると考えられ

る。

3:karioki2(n=54)

検証者が流動性条件を元に抽出した銘柄群。

4:karioki3(n=59)

検証者が抽出した銘柄群。

karioki2 と選定基準は同じだが、選定を行った日が1ヶ月程度異なるため karioki2 と区別した。

4:収益算出方法

東証には約 3700 銘柄が上場していますが、銘柄ごとに株価、時価総額、ボラティリティが大きく異なります。

今回のように「複数の銘柄に同一システムを適用する手法」の場合、純粋な値幅換算収益 (※1)を pl として用いる方法では各トレードで獲得した収益に対して妥当な評価を行うことができません(ボラティリティの大きな銘柄、株価水準の高い銘柄の収益を加

重に見積もってしまう)。

その為、pl 算出にあたって以下の標準化処理を施しました。

トレード銘柄ごとに「直近 20 日間における 1 日の平均ボラティリティ(ATR)」を算出し、純粋な獲得値幅 (pl) を ATR 幅で割った「ATR 換算収益 (pl_atr)」を計算(※2)。

今回の検証では、この標準化された収益指標である「**pl_atr**」を用いることにします。

※1: 売却価格から購入価格を差し引いた値 (空売りの場合はその逆)

※2:

pl_atr 計算例 1: 株価 200 円の銘柄を購入し、220 円で売却。直近 20 日間の平均ボラテ

ィリティ ATR=20 円の場合、 $pl_atr = (\text{売単価 } 220 \text{ 円} - \text{買単価 } 200 \text{ 円}) / 20 = 2$

pl_atr 計算例 2: 株価 15000 円の銘柄を購入し、14000 円で売却。直近 20 日間の平均

ボラティリティ ATR=500 円の場合、 $pl_atr = (\text{売単価 } 14000 \text{ 円} - \text{買単価 } 15000 \text{ 円}) / 500 \text{ 円}$

$= -2$

5:データ前処理

処理 1: 「特別気配スタート」が前提となるため、スクリーニングとして「始値」が前日比±2%未満（※1）の銘柄を除外。

処理 2 : 東証のシステム障害が発生し、終日取引が停止された 2020/10/1 のデータを delete 処理。

処理 3 : ストップ高（ストップ安）で終日値段が付かなかった銘柄をトレード結果から除外（実際にはエントリーできないため）

※1: 「2%」という数字は「特別気配スタート」が約束される最小値に最も近い整数値。

特別気配スタートでない銘柄は、寄付の瞬間までポジション方向を系統的に確定することはできない)

6:分析に用いた変数一覧

収益（目的変数: pl_atr ）との相関が考えられる 11 種類の説明変数を用意(主にテクニカル指標、出来高、マクロ動向)。※以下表記は買いポジションにおける解釈、括弧()は売りポジションの場合を示す。

Position: 売買種別（買いポジション:l, 売りポジション:s で表す）

X1:前日終値からのギャップの大きさ。計算式： $x1=|当日始値-前日終値|/ATR$ 。

ギャップが大きいほど、寄付の段階で既に買い込みが進み、高値掴みになるリスクが高いように見える。

X2:出来高規模。計算式： $x2=エントリー日の一日出来高/直近20日間における一日平均出来高$ 。出来高ボリュームが大きいほどその銘柄に対する需給の変化が大きいことから pl との相関が考えられる。ただし終値時点で初めて決定する数値データであるため、「寄付トレード」においてはリークが発生する。従ってこのまま実際の予測モデルに使用することは出来ないことに注意が必要。

X3:前日終値から当日始値にかけて、直近20日間の高値（安値）を上抜いた（下抜いた）日数。この日数が大きいほど、直近20営業日中に買った投資家の含み損が解消さ

れていることを示す

X4: エントリー価格時点において直近 60 日間の高値（安値）を超えていないローソク足の日数（ $0 \leq x4 \leq 60$ の範囲。買いポジションにおいて新高値に価格が位置しているなら最小値 0 を取る。新安値であれば最大値 60 を取る。売りポジションであれば新安値で最小値 0, 新高値で 60 を取る）。上値の抵抗の大きさ（既存投資家の含み損ポジションの大きさ）と収益との相関を調べたい。

X5: 前日終値から当日始値にかけて、直近 20 日間の高値（安値）を何日連続して上抜いたか（下抜いたか）。直近の値動きに対するインパクトの強さと収益との相関を調べる。

X6: 前日が ATR 幅以上の下落(上昇)であるかどうかのフラグ (True: $x6=1$, False: $x6=0$)。前日のローソク足は最も直近の需給を反映していると考えられる。前日大きく売られた銘柄が上がった場合、トレンド転換になりやすいのではないか？

X7: 前日が陰線（陽線）であるかどうかのフラグ (True: $x7=1$, False: $x7=0$)。X6 と似た指標。X6 は大きな投げがあったかどうかのフラグだが、こちらは大きさは問わない。

X8:当日の日経平均株価の始値時点における前日比 (%)。マクロの動きと pl の相関。

X9:当日のザラ場(市場取引期間。日本株は 9:00-15:00)中の値上がり幅。計算式： $x9=(前日終値-前日始値)/ATR$ 。X6, X7 フラグを定量化した指標。

X10:直近 3 日間における上昇率。計算式： $x10=(前日の終値-3 日前の始値)/ATR$

X11:日経平均のザラ場中の上昇率 (%)。マクロ動向に対する pl の抵抗性を可視化するための指標。ただしリークが発生するため、実際の予測モデルには使えない。

※なお、X1,X2,X3,X4,X5, X10 は変数に固有のパラメータ値を持つが、この値には検証者が適当だと考えられる値を設定した。値が小さすぎると変数としての意味が希薄となり、大きすぎるとサンプル数が減ってしまう影響に配慮した。

分析結果

1:各変数の相関

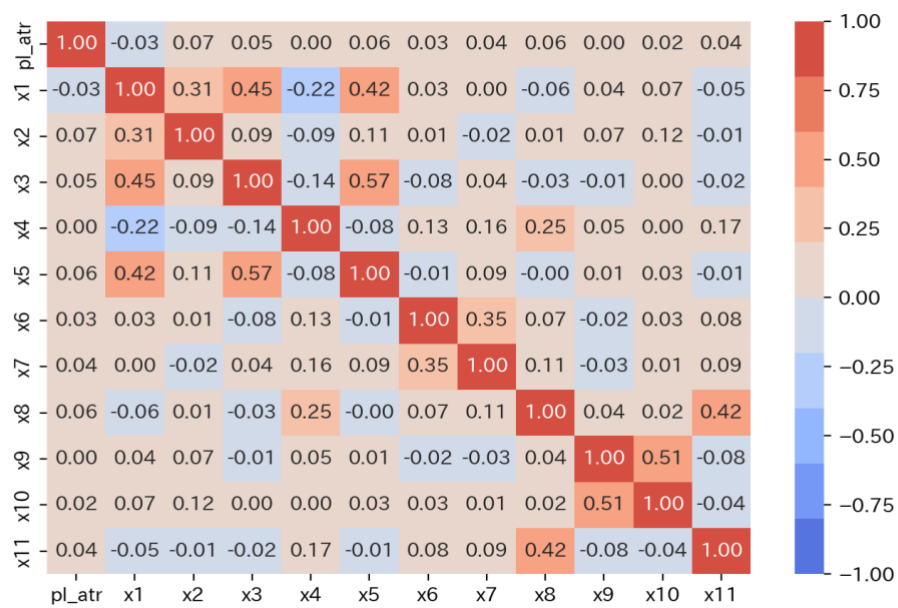
2: pl_atr との相関を検定

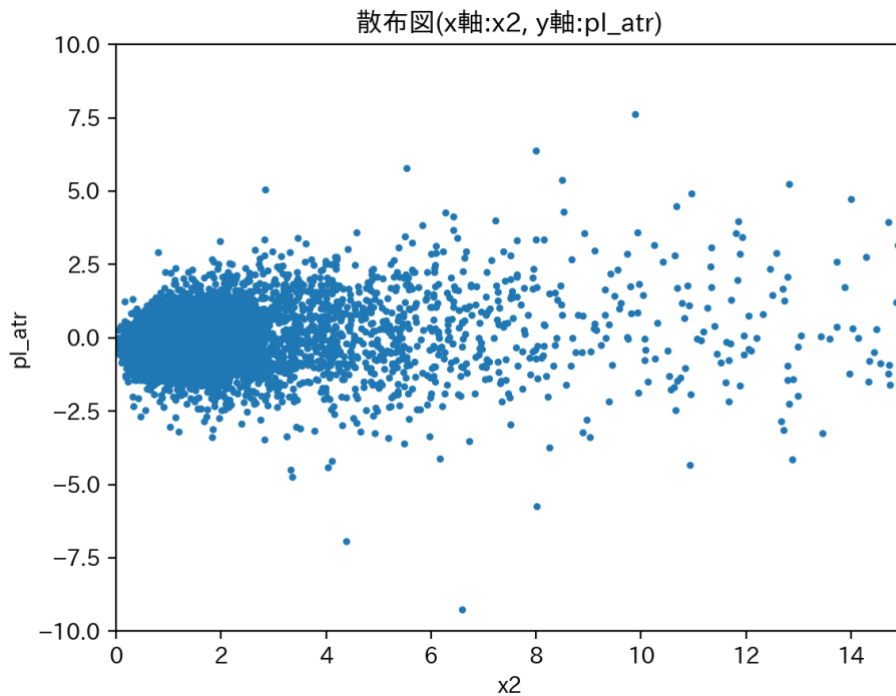
3:決定木による優位性領域の探索

4: 2次元ヒートマップ表による優位性領域の可視化。

1:各変数の相関

各変数間のスピアマン相関係数表、pl_atr との散布図を以下に示します(karioki2 群)。





相関表の結果から、pl_atr に対してどの変数も相関係数 0.1 未満という結果となった。

散布図からも pl_atr との相関がハッキリ見て取れる変数は確認されない(グラフが多いので散布図は x2 だけ添付)

しかし一般的なデータと比較して金融データは過分散の傾向があり、この点を考慮して優位性解釈を進める必要がある。実際に、日本証券取引所グループの考察によると相関係数 0.15 程度のスコアでも高い予測精度があるとみなされているようです。

(引用 <https://japanexchangegroup.github.io/J-Quants-Tutorial/>)

“

予測と真の値には、正の相関(0.144192)が見受けられるので、ある程度の相関関係が発生しています。一般的なデータで 0.144 という数字が出てもほぼ無相関に見えますが、金融データでは 0.144 というスコアは高い部類に入ります。このように視覚化すると、予測値と真の値の関係性を可視化できます。

(2章 10 節 予測結果の可視化方法)

“”

今回の結果においても相関係数 0.07 (他データセットにおいては 0.11) と、金融データという文脈においては気になる結果が得られています。しかし、この相関が誤差の範疇を超えたものであるかどうかについて確信が持てないため検定を行いました。

2:pl_atr との相関を検定

Karioki2 のトレード結果サンプル(n=7495)を $pl_atr > 0$, $pl_atr < 0$ の 2 群に分け、2 群間において各説明変数の平均値に有意差があるか両側検定(有意水準 $\alpha = 0.05$)。

(対象群は karioki2。質的データはカイ二乗検定、量的データについては正規分布が仮定できる場合は welch の t 検定,仮定できない場合はマンホイットニーの U 検定を適用)

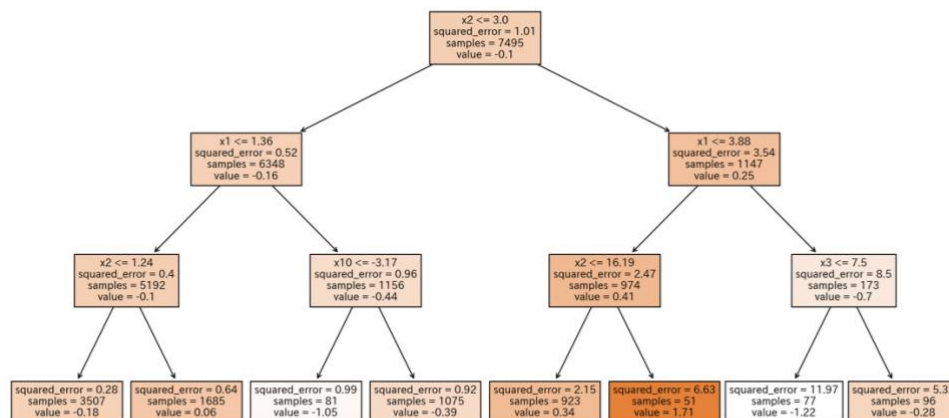
説明変数	P 値	T 値, [U 値]、(χ^2 乗値)
Position(買 or 売)	0.24	(1.32)
X1	0.13	[6962436]
X2	5.0×10^{-48}	[8166060]
X3	0.004	2.88
X4	0.86	-0.17
X5	0.0005	3.45
X6	0.01	(6.69)
X7	0.14	(2.15)
X8	0.0002	3.71
X9	0.11	-1.58
X10	0.42	-0.79

以上の結果から、説明変数{x2,x3, x5, x6, x8}と収益 (pl_atr) の相関は有意である結果が得られました。次にこれら相関の結果として、優位性領域 ($E(pl_atr)>0$) が発生して

いるか調べます。

3:決定木による優位性領域の探索

まずどのデータから探索を行うべきかの糸口を探る為、「回帰木モデル」を用いて可視化しました(value: pl_atr)。



この結果から以下の傾向が確認できます。

- 1: x2 が pl_atr に対して最も影響の大きい特徴量となっていそう
- 2: 第一分岐で期待値プラス/マイナスの領域を上手く分離できている

3: x1 が次に重要な特徴量となっていそう

※なお他3つのデータセットについても同様の可視化を行ったところ、これらにおいても x2 が第一分岐に選択される結果となりました。

決定木の結果を踏まえて、説明変数 {x1,x2} に的を絞り、目的変数 {pl_atr} との関係を2次元ヒートマップ図で可視化しました。

4: 2次元ヒートマップ表による優位性領域の可視化。

(**赤色**は期待値プラスの領域を示す、**青色**は期待値マイナスの領域を示す)

左→収益期待値 $EV(pl_atr)$ のヒートマップ、

右→サンプル数(N)のヒートマップ (クロス集計表)

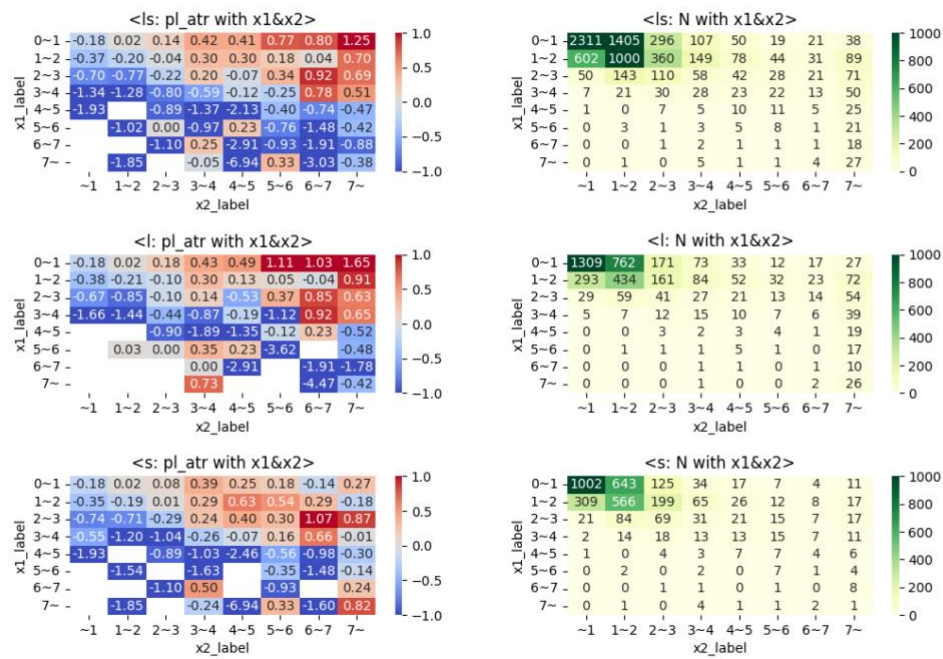
ls (上段) →買いポジション、売りエントリーの合算

l (中断) →買いエントリーのみ

s（下段）→売りエントリーのみ

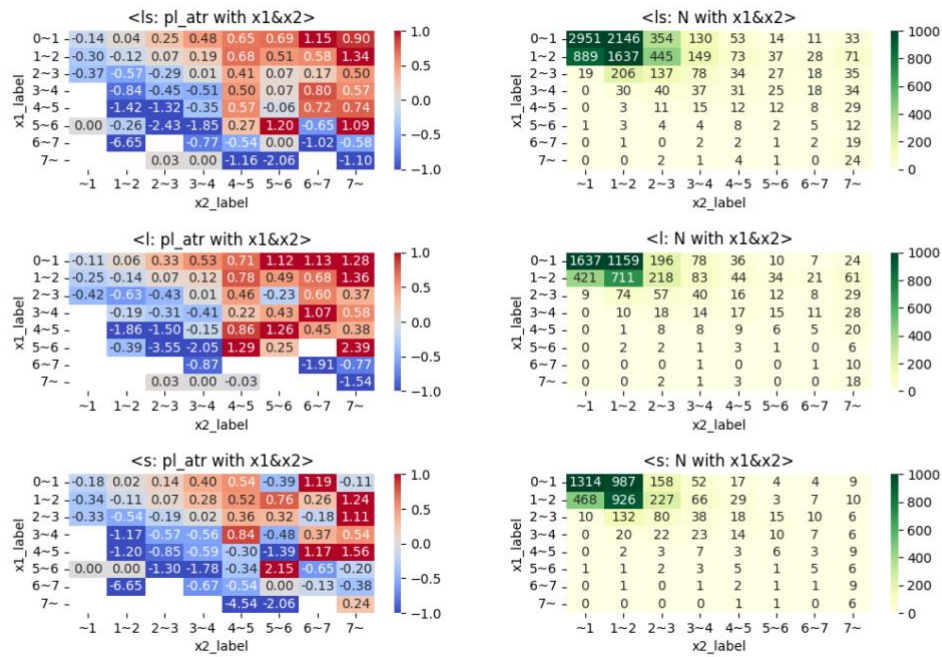
<karioki2>

figure1

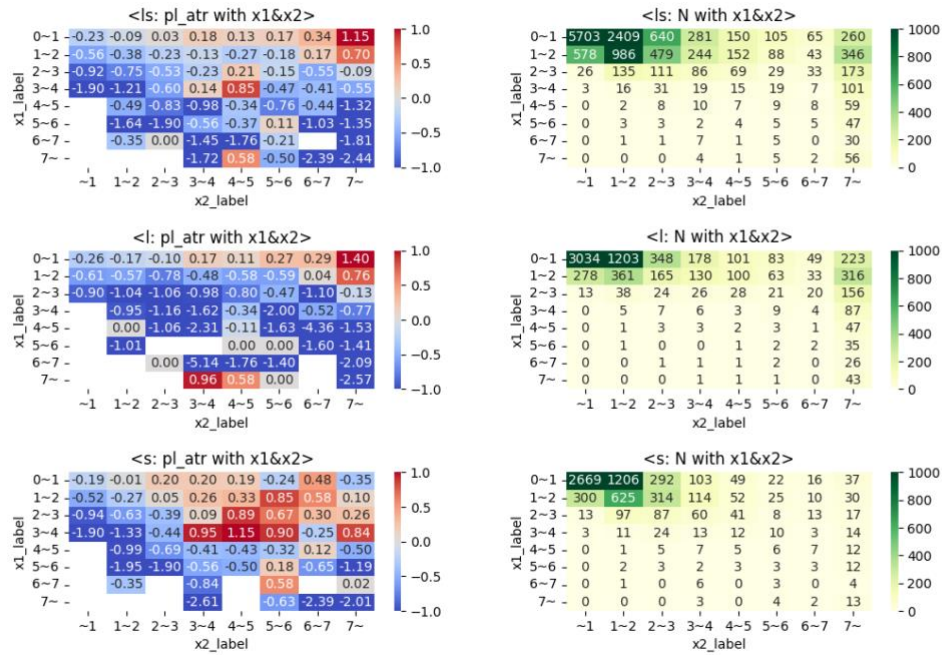


<karioki3>

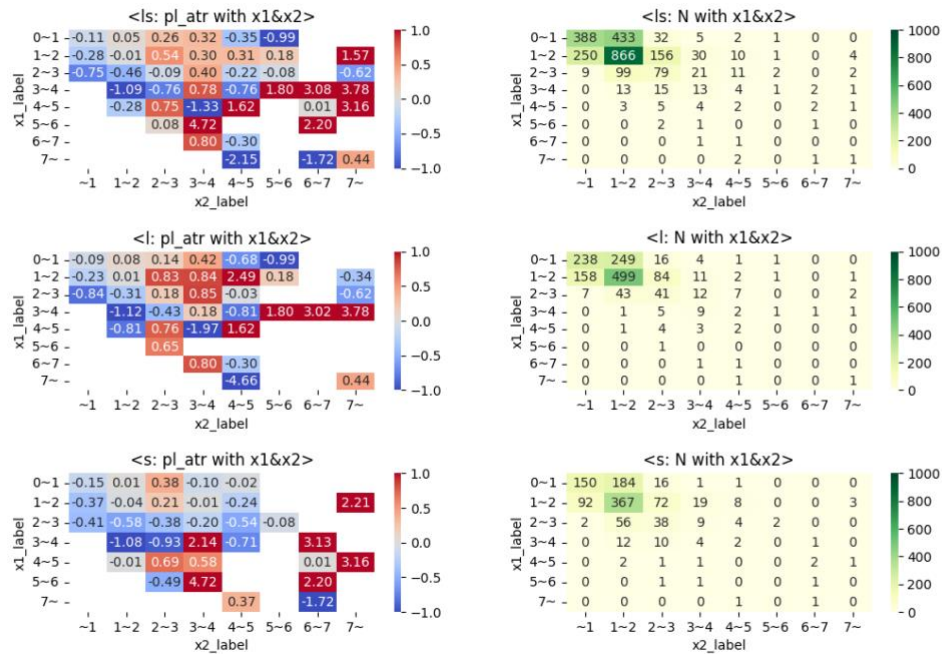
figure1



<borokabu_100_300>



<capital100>



ヒートマップ表の結果から、 pl_atr と説明変数 $\{x_1, x_2\}$ の関係には以下の傾向が確認される。

- 1: 全てのデータセットに共通して $\{pl_atr, x_2\}$ には単調相関の関係が垣間見える。また、全てのデータセットにおいて期待値プラス($pl_atr > 0$)の領域が発生している。
- 2: $\{pl_atr, x_1\}$ の関係について各データセット間で差異が見られる (borokabu 群の優位性領域は際立って狭い)

3: ポジション種別(lors)についても各データセット間で差異が見られている (borokabu 群だけ優位性領域が s のほうが広い)

考察

1: pl_atr に対して最も強い相関を持つ変数は「X2:出来高」であることが判明した。

2: データセットの区分によって {x1, ポジション種別} の有効範囲に差異が見られたことから、何らかの「市場区分」が優位性領域の広さに作用していると考えられる。

3: 短期モメンタム株式投資において、{x3~x11}で与えられたような「テクニカル指標」、
「マクロ動向」の影響は「出来高」に比べて小さい

予測モデル構築に向けて

今回の結果から洞察が引き出された一方で、今後の予測モデル構築に向けての課題は多くあることも判明した。具体的には以下のような問題を解決しなければならない。

1:収益予測に有効な変数として{x2}が得られたが、{x2}は事前に参照できるデータ量ではなく、実際の予測に使えない。従って {x2} と強い相関があり、かつ事前参照可能な変数を見つけ出す必要がある。その候補としては「寄付出来高」、「前日の材料の有無」、「材料の種別」などが挙げられる。

2: 各データセット間で変数 {x1, position} の優位性範囲にバラツキが見られたことから、「市場区分」は優位性領域を決定する欠かせない変数となっている可能性は高い。そのため「市場区分」を上手く表現する変数を見つけ出すことも予測モデル構築に必須となる。候補としては「上場区分」、「時価総額」、「業種・セクター」、「発行株数に占める浮動株の割合」などが挙げられる

3:今回のデータは直近5年間のデータを用いているが、さらに期間を遡っても優位性は維持されているか。※なお今回の検証期間内(2015/1/1~2021/11/10)における年間収益分布は確認済で、期間内においては年度によらず安定した成績を維持している。グラフは github リポジトリ (https://github.com/wabisachin/kabu_analysis/report) にて示す。