

## <日本株短期モメンタム投資における優位性領域の探索、得られた結果の背景的考察>

### ・はじめに

優位性がどこに存在し、その優位性はどこからやってくるのか。この謎を解明することは多くの投資家にとってのロマンであり、最大の目標の一つと言えます。

世の中には数多くの投資手法が存在し、それぞれの手法がどのような条件のもとで機能するのか。一人の投資家がその全貌を知ることは

今回はその一例として、「短期モメンタム投資※1」と呼ばれる投資手法に焦点を合わせ、テクニカル側面（ローソク足や出来高などの過去の時系列データから将来の株価変動を予想しようとする立場）から説明できる優位性がないか、探索していきます。

なお、得られた結果については単に数値の解釈にとどまらず、背景でどんな市場原理が働いた上でこのような結果が得られたのか、ファンダメンタル側面から解釈していきます。

※1 モメンタム投資とは相場の勢いを利用して収益を上げる投資手法。値動きと同じ方向ヘトレードすることから「順張り」とも呼ばれる。

### ・分析の目的

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します。）

### ・分析の概要

- 1:売買ルールの定義
- 2:使用するデータセット
- 3:収益の算出方法
- 4:データの前処理

- 4:分析に用いた各説明変数の概要
- 5:各説明変数の相関を可視化
- 6:有意差の検定
- 7:相関が見られた変数を中心に優位性領域が存在するか可視化
- 8:結果に対しての背景的理由（ファンダメンタルズの考察）
- 9:今回の検証で引き出されたこと
- 10:今後の結果から示唆された可能性

#### ・ 検証手法の定義

今回検証を行うのは「短期モメンタム投資」です。

具体的には当日の株式市場オープン時（9:00）に「特別買い気配(※1)」スタートした銘柄を寄付で購入し、当日の大引けで売却するデイトレード手法の検証を行います（買いエントリーの場合。売りエントリーについては※2）。

※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、いきなり価格決定するのではなく、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）ごとに現在値を更新していき、市場需給が一致したタイミングで初めて始値を決定する「板寄せ方式」が採用されている）。

※2 空売りの場合は「特別売り気配」スタートした銘柄を寄付と同時に空売りし、その日の大引けで買い戻す。

#### ・ 売買ルール詳細

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。（買い特別気配で始まれば買いポジション、売り特別気配で始ま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手仕舞う）。

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：利確指値は設定しない。手仕舞いはロスカットポイントまで株価が逆行するか、当日の大引けに行く。(ストップ高で大引けを迎えても当日に決済)

~~※手法検証の前提として、資金管理方式は「固定リスク(銘柄にかかわらず毎回同じ金額のリスクを取る)」を想定して期待値検証を行う(後述する `pl_atr` で収益評価を行う方法にも適用しやすい)。複数銘柄を同一システムに適用する手法の場合、資金管理方式には他にも、「固定資金方式」、「固定ユニット方式」などが存在する。一般にファンドのような大量資金を投入、長期投資が運用の方針となっている場合、固定資金方式に近い形で運用が行われるケースが多い。運用資金の乏しい個人投資家の場合は固定資金による運用方法は実際には現実的ではなく、信用取引を活用して複利的なリターンを目標にする固定リスク方式、固定ユニット方式での運用が適している。~~

#### ・事前処理

前処理 1: 「特別気配スタート」が今回の手法の前提となるため、スクリーニングとして「始値」が前日比±2%未満(※1)の銘柄はあらかじめ除外。

前処理 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 のデータは delete 処理。

前処理 3: ストップ高(ストップ安)で終日値段が付かない銘柄については、実際にはエントリーできないのでトレード結果から除外。

※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通常の銘柄については株式市場オープン(9:00)と同時に価格決定される。なおこの場合、寄付の瞬間まで前日より上がって始まるのか、下がって始まるのか事前に判別することはできない。したがって、ポジション方向を系統的に確定することはできない。)

#### ・収益(pl)の評価方法

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、ボラティリティは大きく異なります。

今回のように、複数の銘柄に同じシステムを適用することが前提となる手法の検証を行う場合、純粋な値幅（※1）を pl として用いる方法では、その手法によって得た収益につい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で生じた収益を加重に見積もってしまう)。

この影響を抑えるため、今回は各トレードの pl 算出に当たり以下の処理を施しました。

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)。

なお今回の検証では収益指標として、この修正された pl\_atr を用います。

※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、 $pl\_atr = (売単価 220 円 - 買単価 200 円) / 20 = 2$

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、 $pl\_atr = (売単価 14000 円 - 買単価 15000 円) / 500 円 = -2$

#### ・分析の流れ

- 1:収益性との相関が考えられる 10 種類の変数(ローソク足、テクニカル指標、出来高、日経平均)を説明変数として用意
- 2:各説明変数と pl\_atr との関係を分析
- 3:優位性が示唆された変数について、他変数との 2 次元ヒートマップ表を作成し、より詳細に優位性領域を特定。
- 4:優位性が示された領域について、ランダムな株価変動(市場仮説理論によるところのランダムウォークを想定)と比較して統計的な有意差があるかどうか検定。

#### ・分析に用いた変数一覧

目的変数 Y: pl\_atr

説明変数 X:x1-x10

※以下表記は買いポジションでの意味、括弧内()は売りポジション。

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

ギャップが大きいほど、既に関入が進んでしまっている可能性が高く、高値掴みになってしまう印象を受けるが、本当にそうか。

X2:出来高規模（計算式： $x2=エンタリー日の一日出来高/直近20日間における一日平均出来高$ ）。

出来高が大きいほど需給に与える影響は大きいことから pl との相関が考えられる

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:エンタ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）。

上値の抵抗の有無（他投資家の含み損玉の多さ）と収益との関係を調べたい。

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）。直近の動きに対するモメンタムの強さがどう収益に反映されるか調べる。X3と似た変数。

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ので変数に加えた。前日の動きと相反の動きをしたトレンド転換を捉えている可能性が高いのでは？とみた。

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と同じ理由。X6は明確な投げがあったかどうかのフラグだが、こちらはあくまで方向感だけ。

X8:当日の日経平均株価の始値時点における前日比(%)。マクロ動向と当日の pl の関係は相関していると考えられる。

X9:当日のザラ場中の値上がり幅（ローソク足における前日の陽線、陰線の長さ）計算式： $x9=(前日終値-前日始値)/ATR$ 。X7フラグを定量化したもの。

X10:直近3日間における上昇率 計算式： $x10=(\text{前日の終値}-3\text{ 日 前 の 始 値})/\text{ATR}$

X11:日経平均株価のザラ場中の上昇率 (%) (※ザラ場・・・取引時間中のこと。  
株式市場では寄付と大引けを除いた 9:00-15:00 を指す)。マクロ動向に対する抵抗性を  
把握するために使用。

※X1,X2,X3,X4,X5, X10 は説明変数が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。

#### →各変数の特徴と分類

~~今回用意した変数は大きく分けて3種類に分類されます。~~

~~前日 (エントリ 前) から値が確定している変数(セットアップとして活用できる)~~

~~当日のエントリと同時に値が確定する変数(トレーダーは発注方法により値のコント  
ロールが可能)~~

~~エントリ段階では値が確定できない変数(事後変数。トレーダーは事前にコントロ  
ールできない。)~~

#### ・データセット詳細

直近5年間(2015/1/1~2021/11/10)の個別株時系列データ(楽天証券から取得した csv  
ファイル)。

これらを4つのデータセットに分けた(時価総額 top100, 価格 100-300 円のボロ株セク  
ター、検証者自身が流動性の観点から有意抽出した銘柄2セット)

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。各株価指数に組み入れられて、ファンドの資金が豊富に  
入っていることが主な特徴。

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)。  
市場人気が乏しく、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、短期トレーダーの資金が相対的に大きなウエイトを占めていると

考えられる。

3:karioki2(n=54)

検証者が有意抽出した銘柄群。

市場区分、成長性、業績、時価総額などは考慮せず、流動性を判断基準として選定。

4:karioki3(n=59)

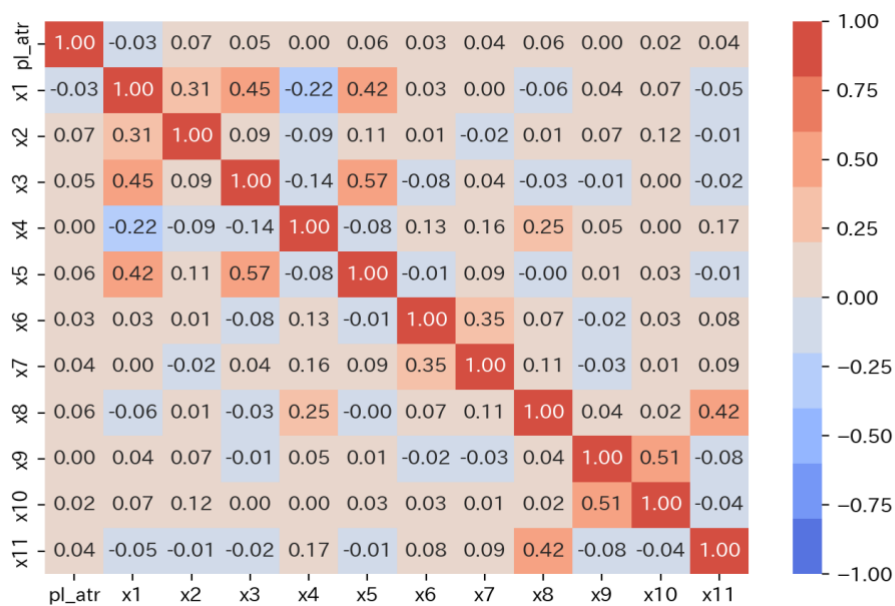
検証者が有意抽出した銘柄群。

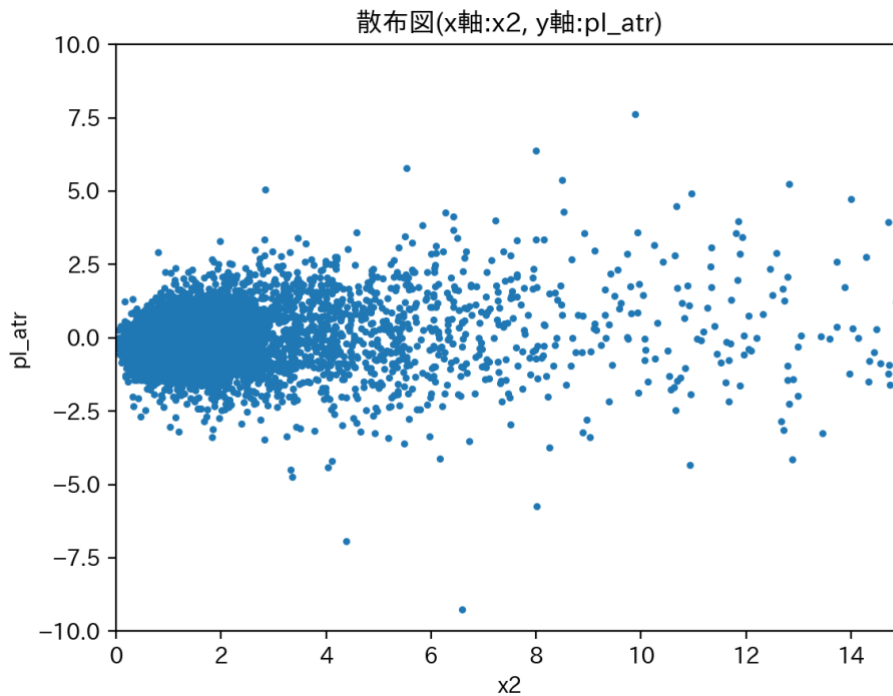
Karioki2 と選定基準は同じだが、選定を行った日が1 ヶ月程度異なるため karioki2 と区別した。

※なお検証にあたって、なぜこのような方法でデータセットを抽出したのか。その辺りの経緯と妥当性については備考欄にまとめました。

#### ・分析結果

まず概要として、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)。





まず pl\_atr との相関を見ると、どの説明変数も相関係数 0.1 未満、散布図からも、pl\_atr との相関がハッキリと確認できるような変数は見つかりませんでした。(他データセットについても同様)

ただし一般的なデータと比較して金融データはノイズが大きく、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、日本取引所グループ)

今回の結果からも相関係数 0.7 (他データセットにおいては 0.11) とそれに準じる相関を持つ変数が確認されていることから、この小さな相関が誤差の範疇であるか詳細に確かめるため検定を行いました。

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有意水準  $\alpha = 0.05$  で両側検定を行いました。(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変数は welch の t 検定,仮定できない変数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]、( $\chi^2$ 乗値)
------	-----	---------------------------



Position(買 or 売)	0.24	(1.32)
X1	0.13	[6962436]
X2	$5.0 \times 10^{-48}$	[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）との相関には有意差があると言えそうです（ただし、サンプル数が多く検出力が高いので参考程度ではあります）。

ただ、最終的に重要なのは、これらの相関によって、優位性領域（EV(pl\_atr)>0）が発現している程度に大きな相関があるかどうかです。

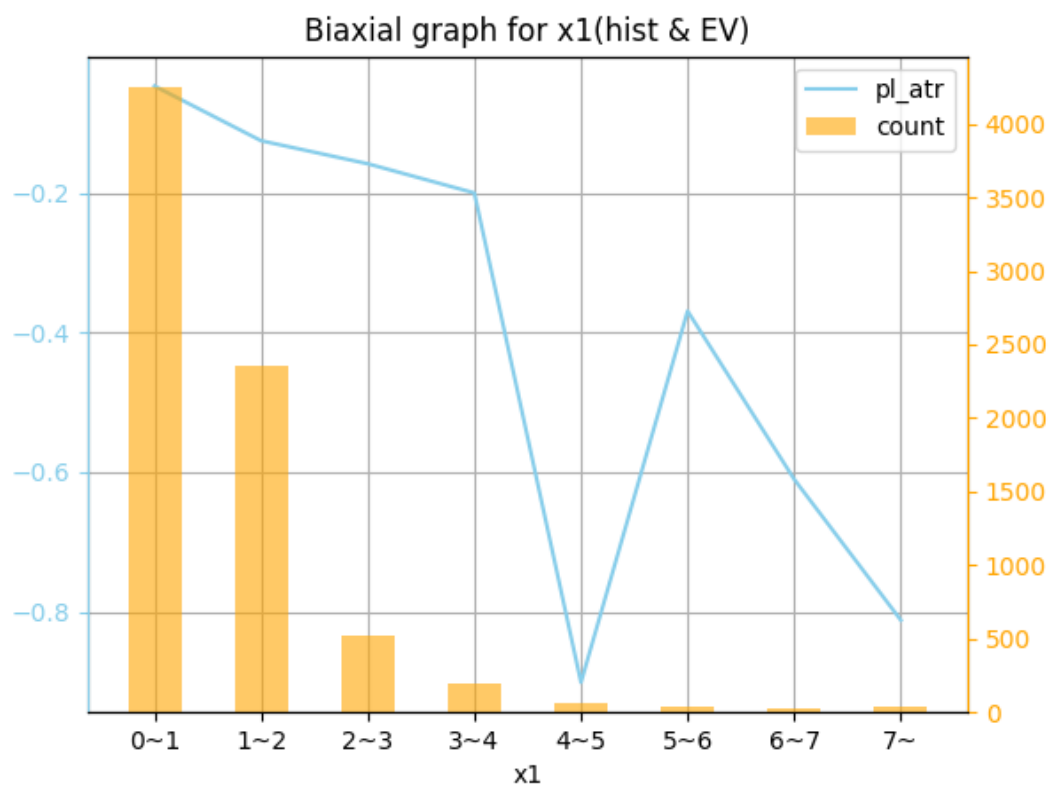
しかし、「トレーディングで収益を上げる」という文脈において1回1回のトレード結果を精度良く説明できるモデル構築は必ずしも必要ではなく、同じようなトレードを例えば100回行った時にトータルリターンが得られるような領域が見つかるとするならば、その領域には優位性があると判断することができる。このような観点で優位性領域が存在するか探索することが大切なのではないかと考えた。

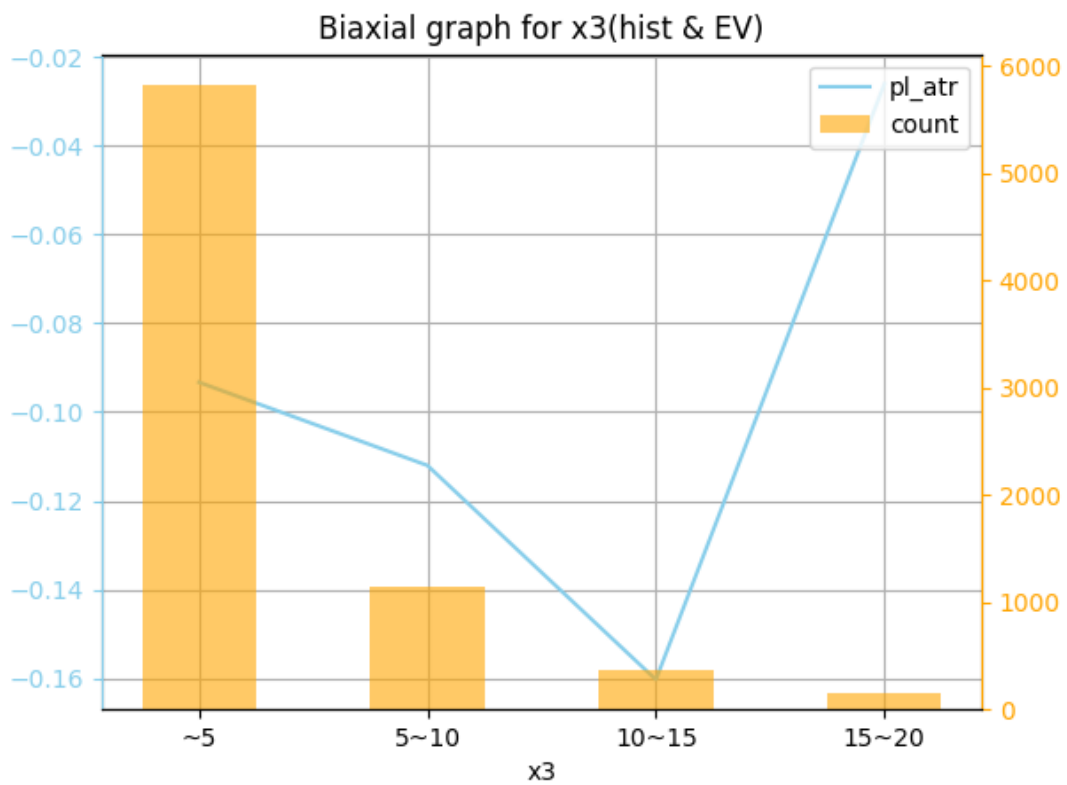
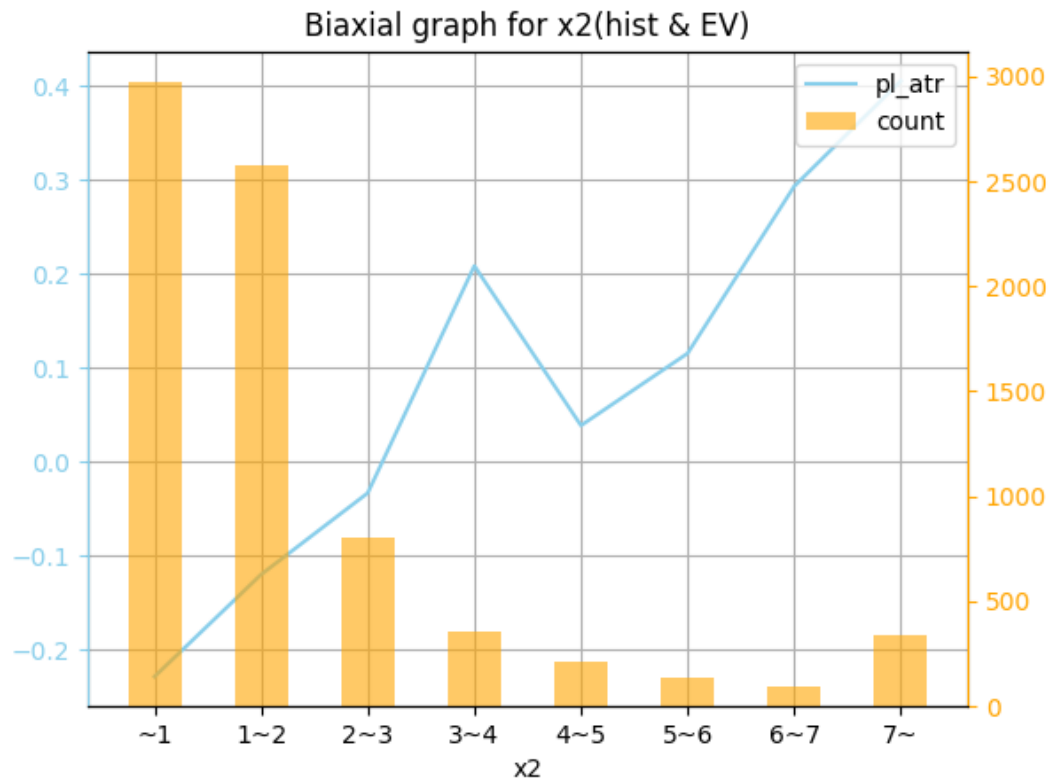
その様子を可視化するために今度は各説明変数をビンで区切り、各領域内での期待値EV(pl\_atrの平均値)を収益指標として優位性検証を行いました。

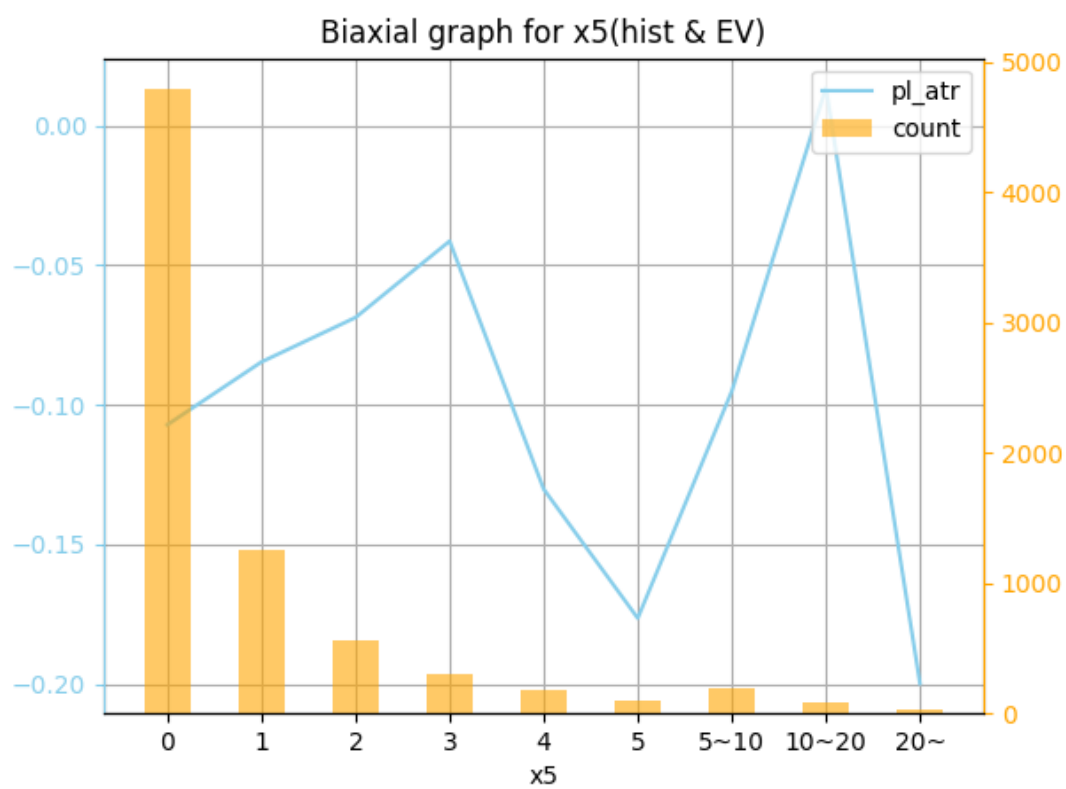
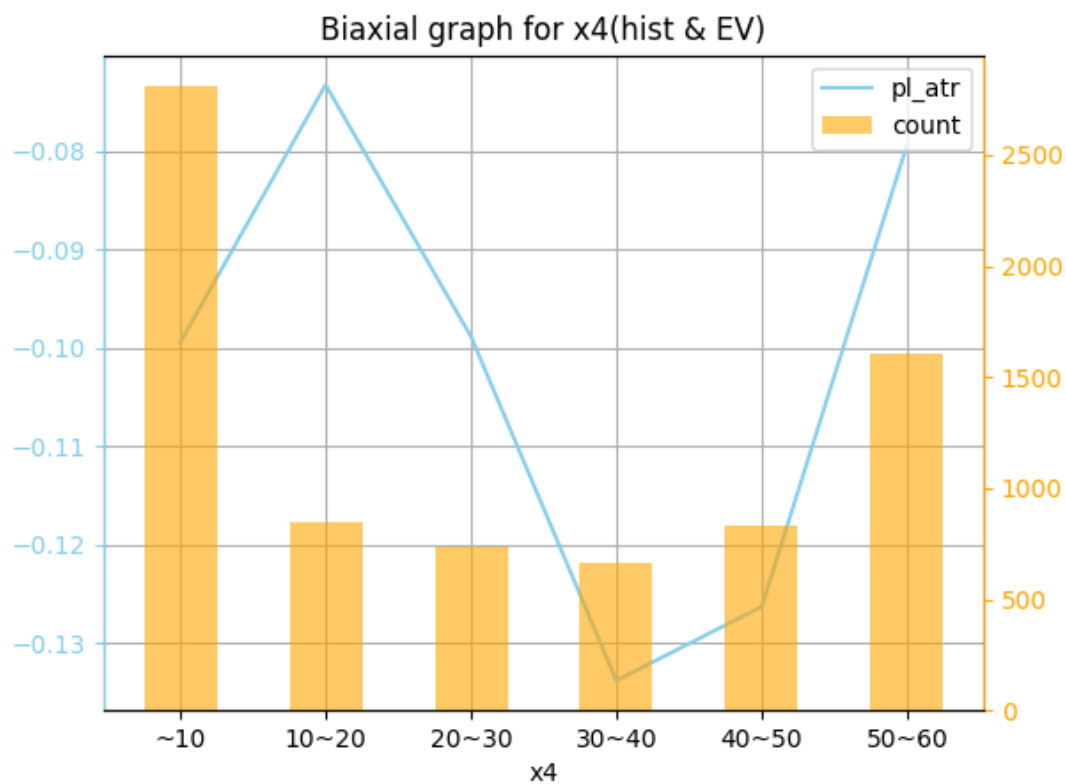
まず、単位トレード当たりの収益期待値EV（各ビン区間におけるpl\_atrの平均値）と各説明変数{x1~x10}の関係についての一次元関係を調べました。

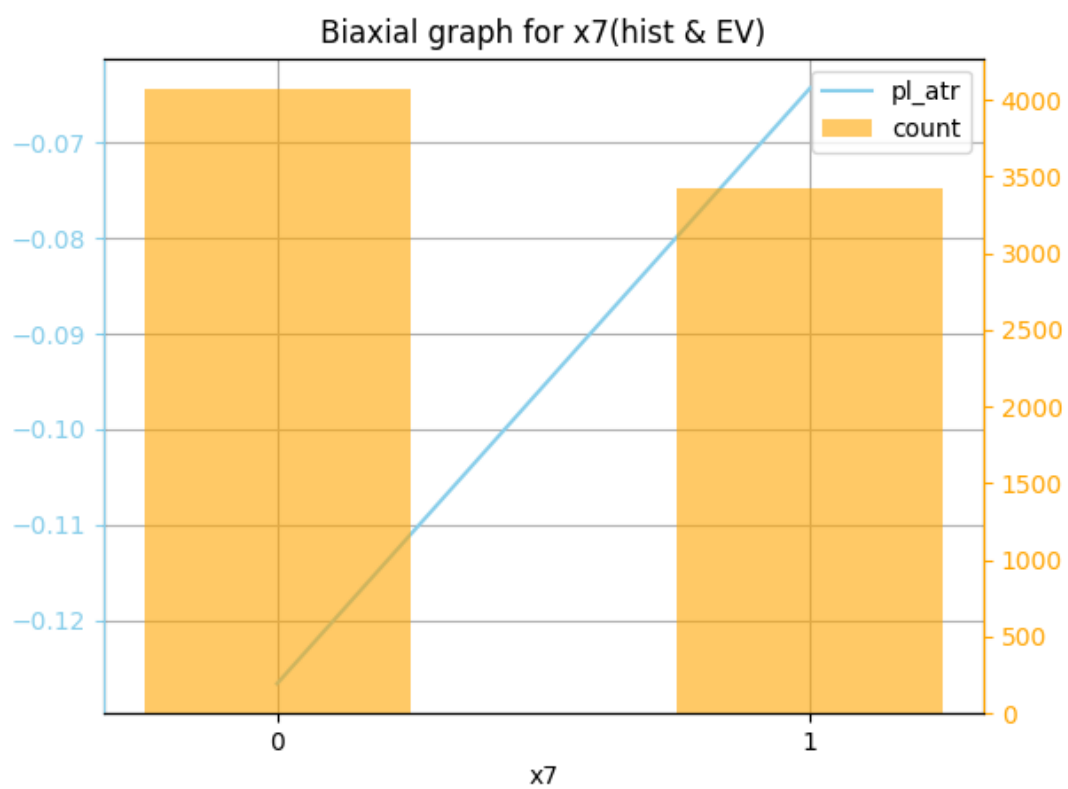
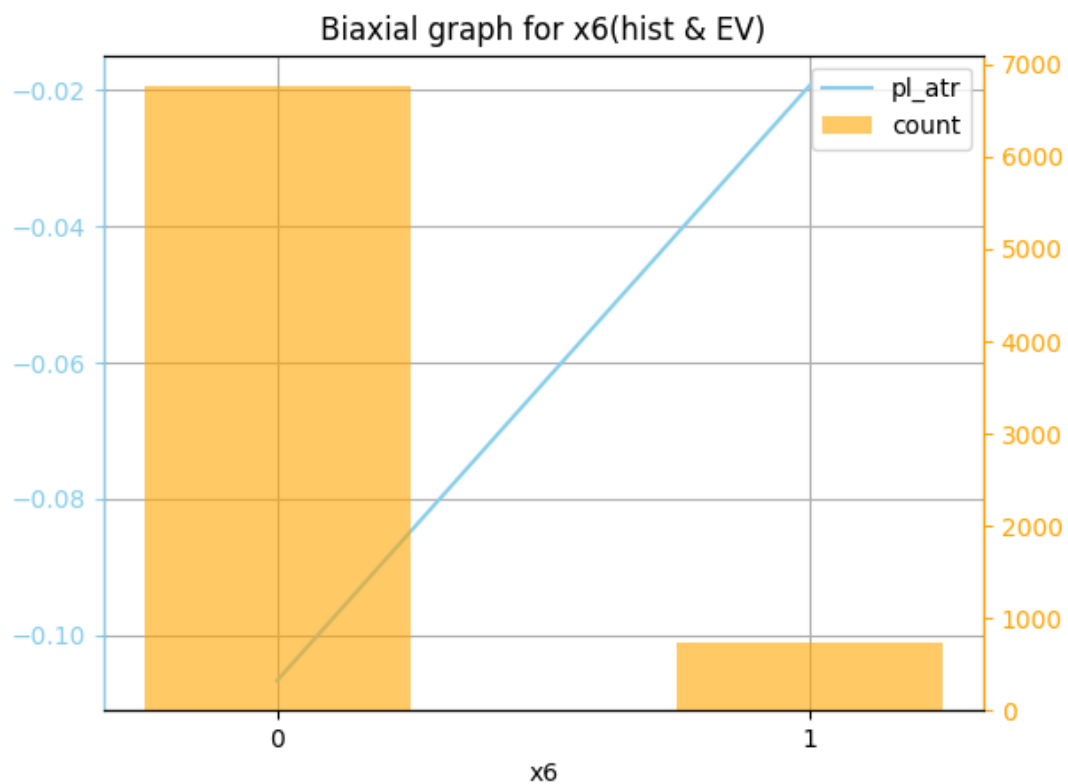
以下がその結果となります(karioki2)。

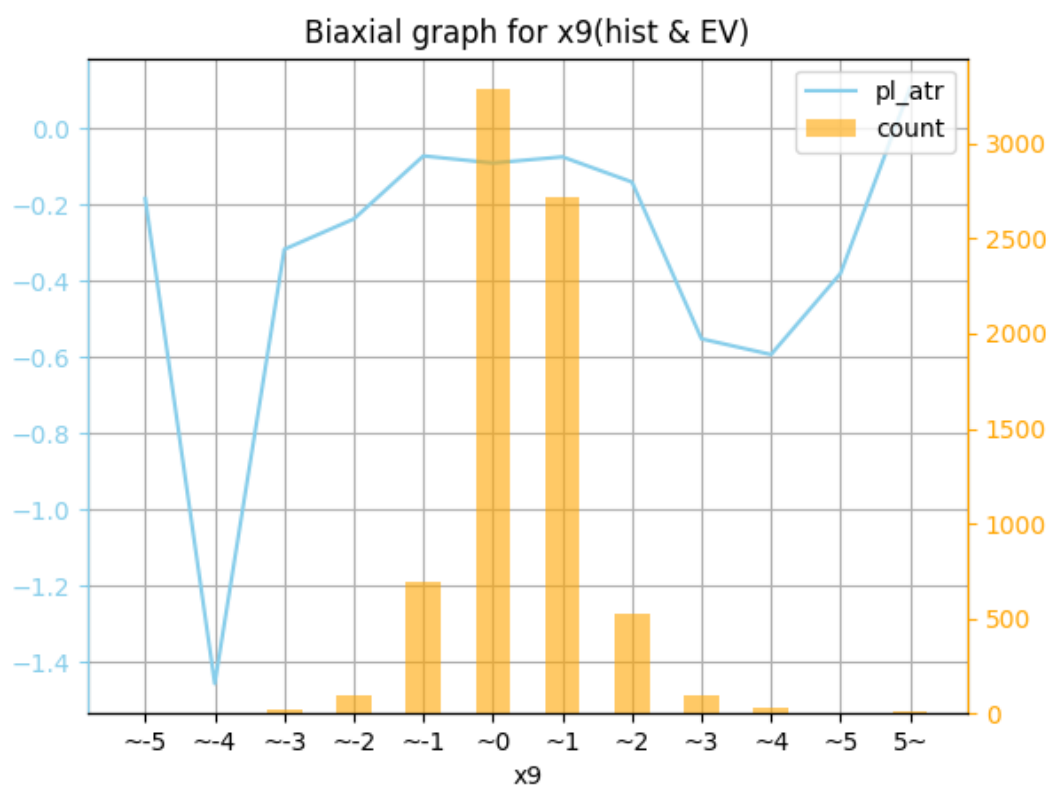
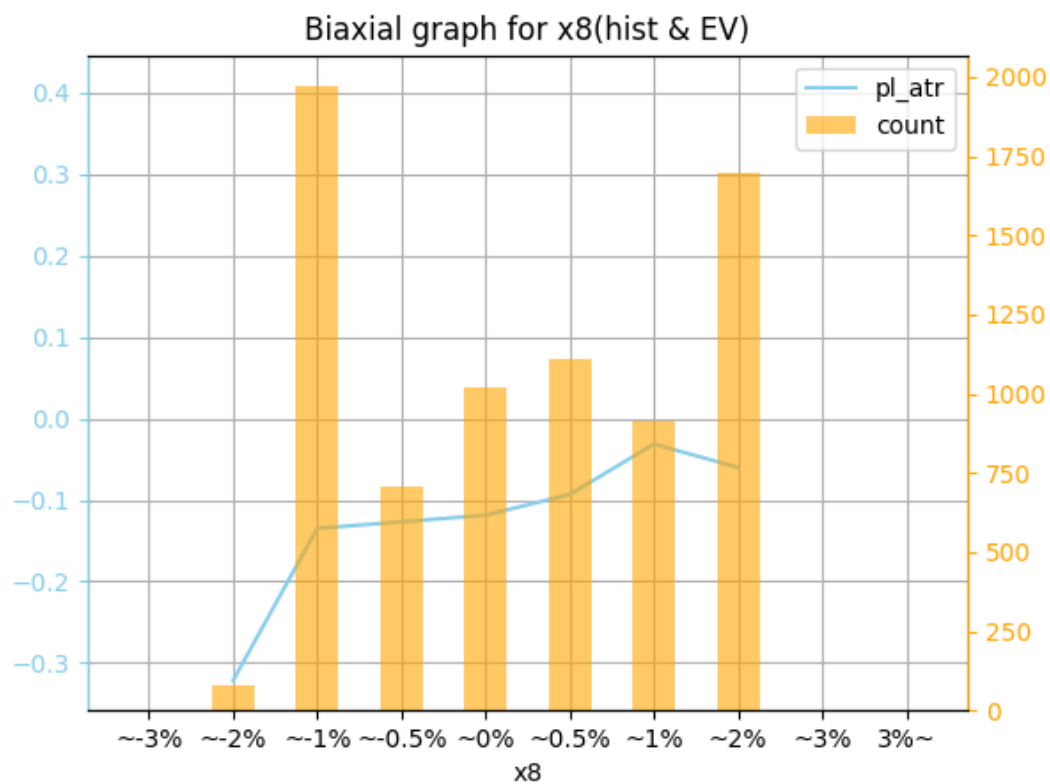
※グラフ枚数が多いので他データセットの結果については私の github リポジトリ ([https://github.com/wabisachin/kabu\\_analysis](https://github.com/wabisachin/kabu_analysis))を参照。analysis ディレクトリに画像ファイルを添付してます。

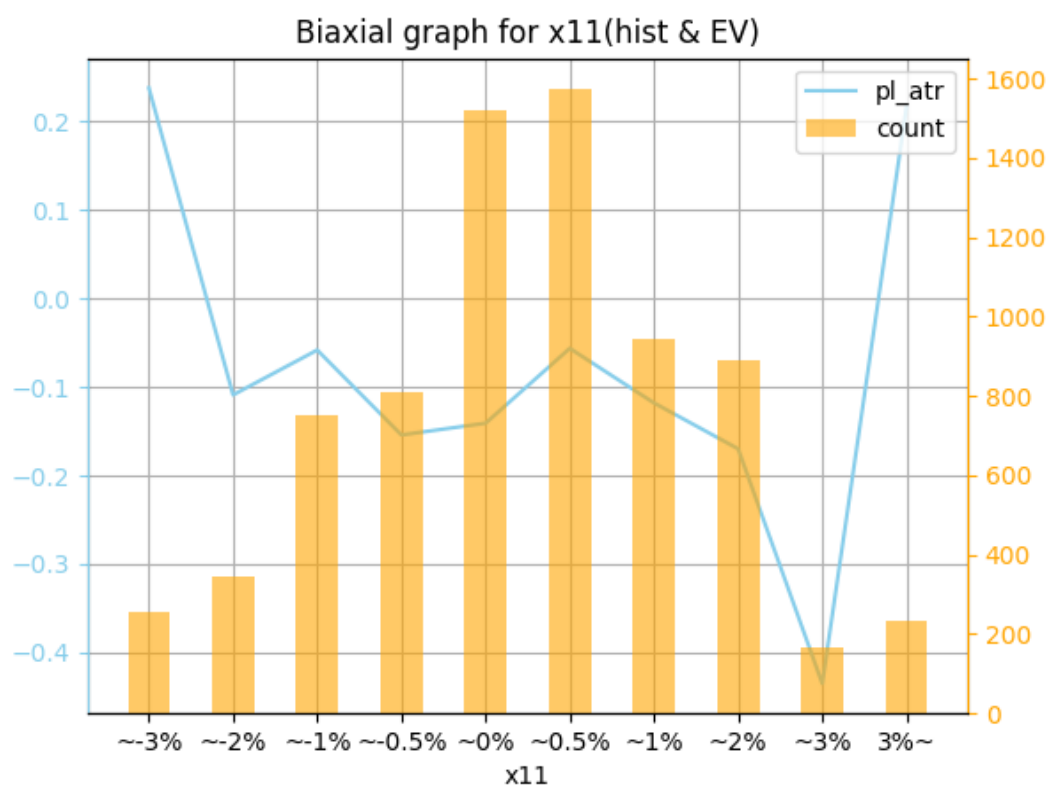
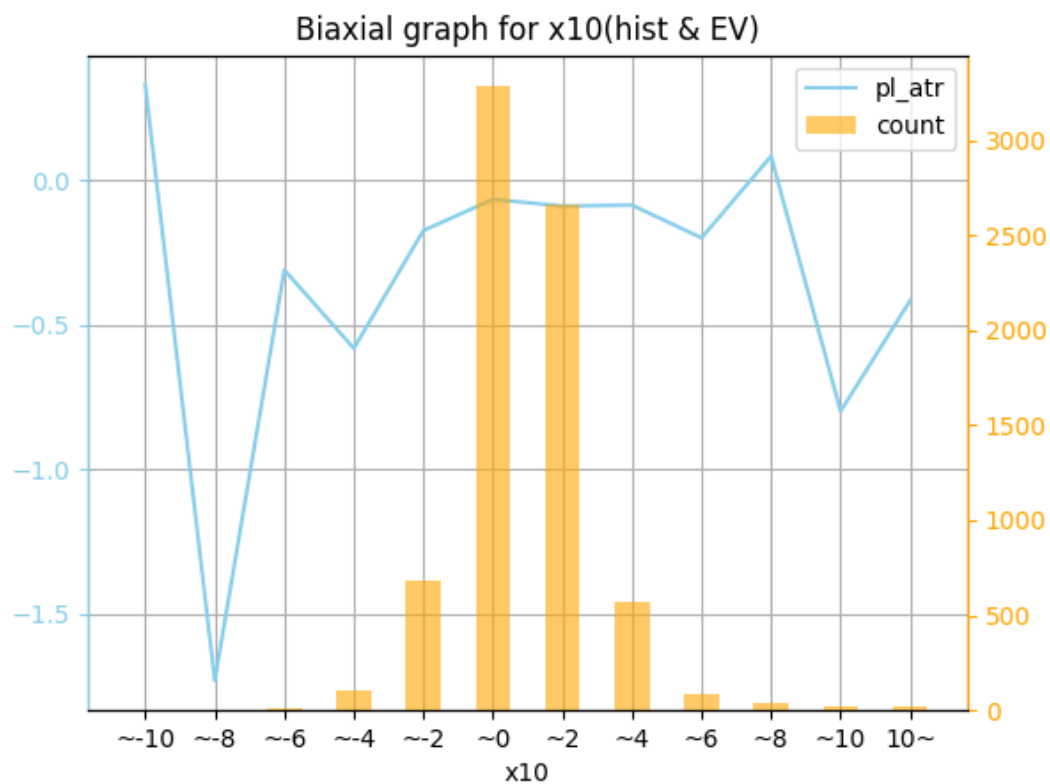










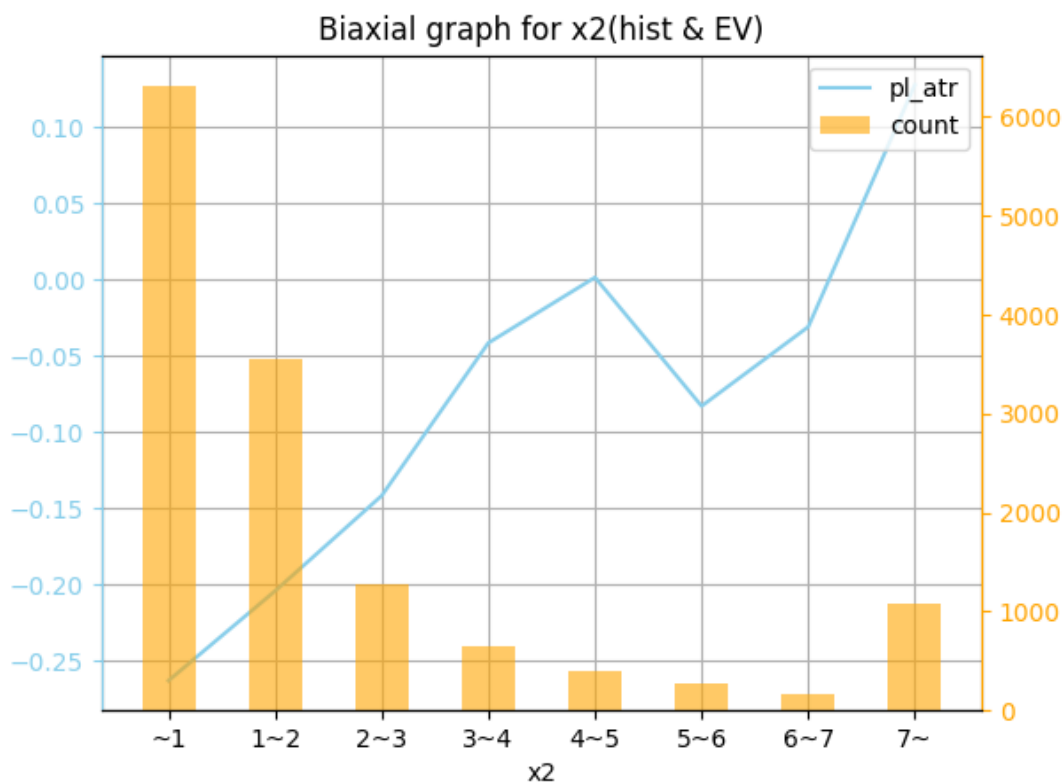


折れ線部分(青)が1トレード当たりの収益期待値  $EV(pl\_atr)$  と各説明変数  $x1 \sim x10$  との関係を示した1次元グラフ、棒グラフ(黄色)が各ビン区間におけるサンプル数(ヒストグラム)を示しています。(  $pl\_atr=1$  であれば、1トレード当たり平均して「平均ボラティリティ幅 (ATR)」と同程度の差益が見込めることを示している。)

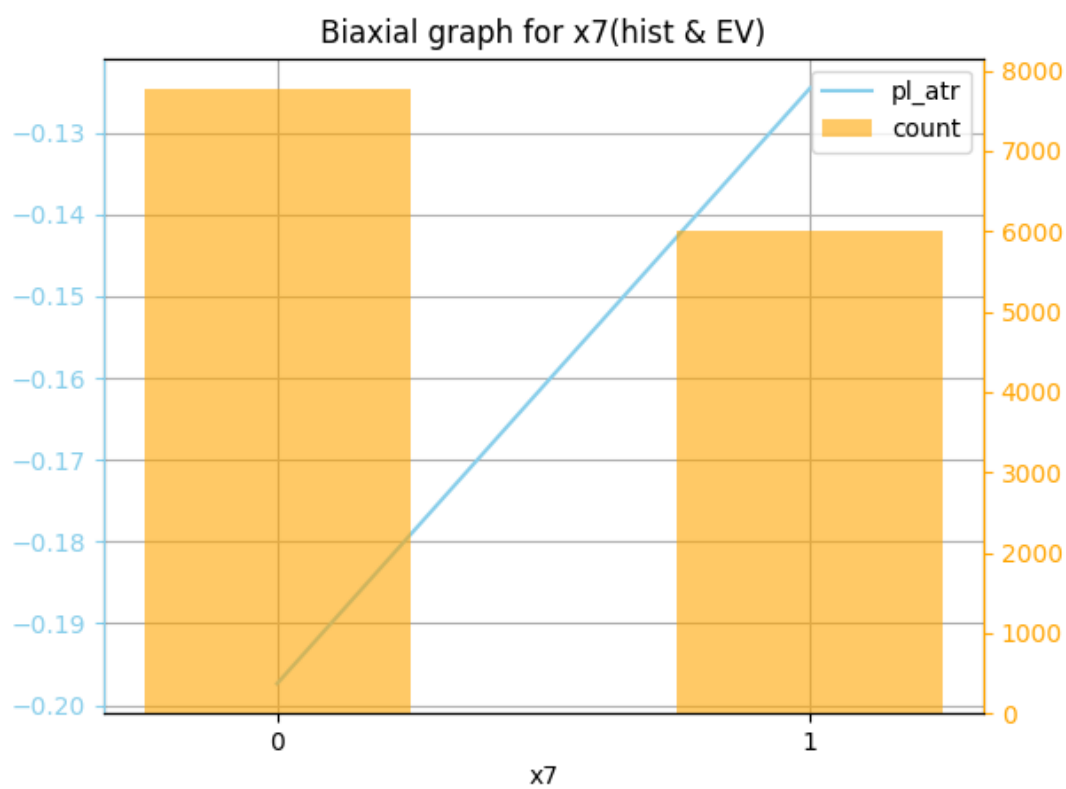
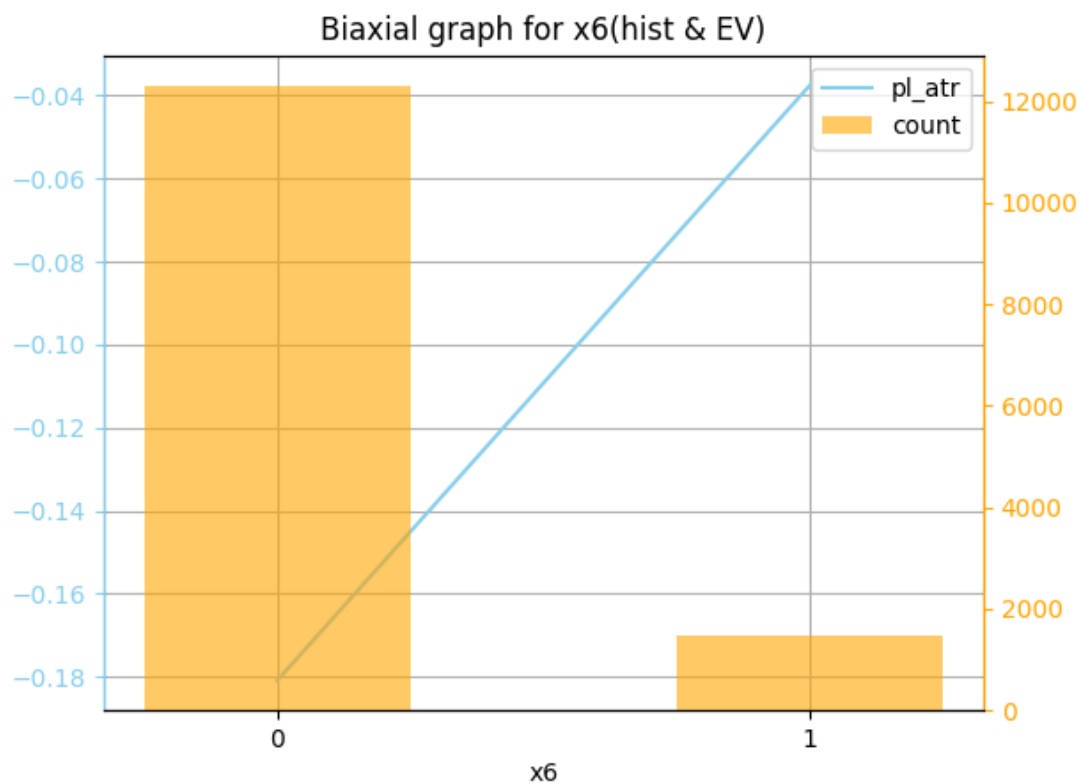
まず、  $pl\_atr > 0$  の領域が存在するかどうかに着目します。  $x2$  にその傾向が示唆されています(その他変数については一部相関が見られるものの、単一変数で際立った収益性の兆候(  $pl\_atr > 0$  )までは確認されない。)

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、  $pl\_atr$  と説明変数  $\{x2, x6, x7\}$  との間に似た相関の様子が確認されました。以下は他データセットにおける  $\{x2, x6, x7\}$  の結果です。

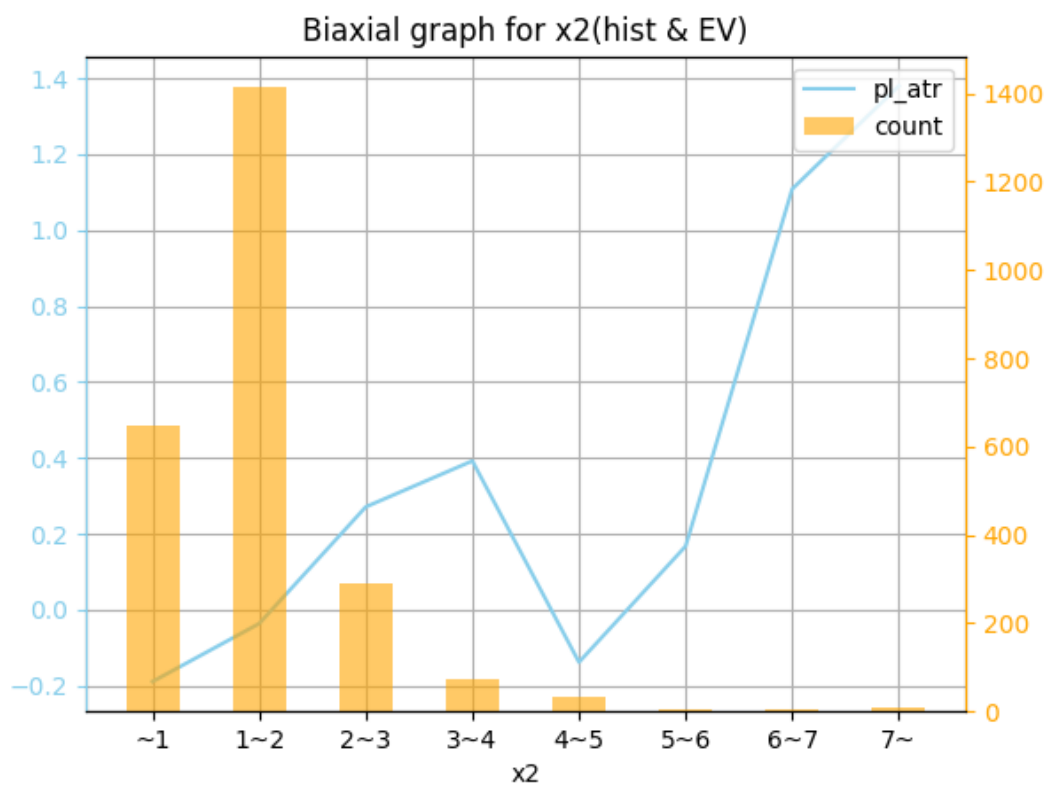
<borokabu\_100\_300>

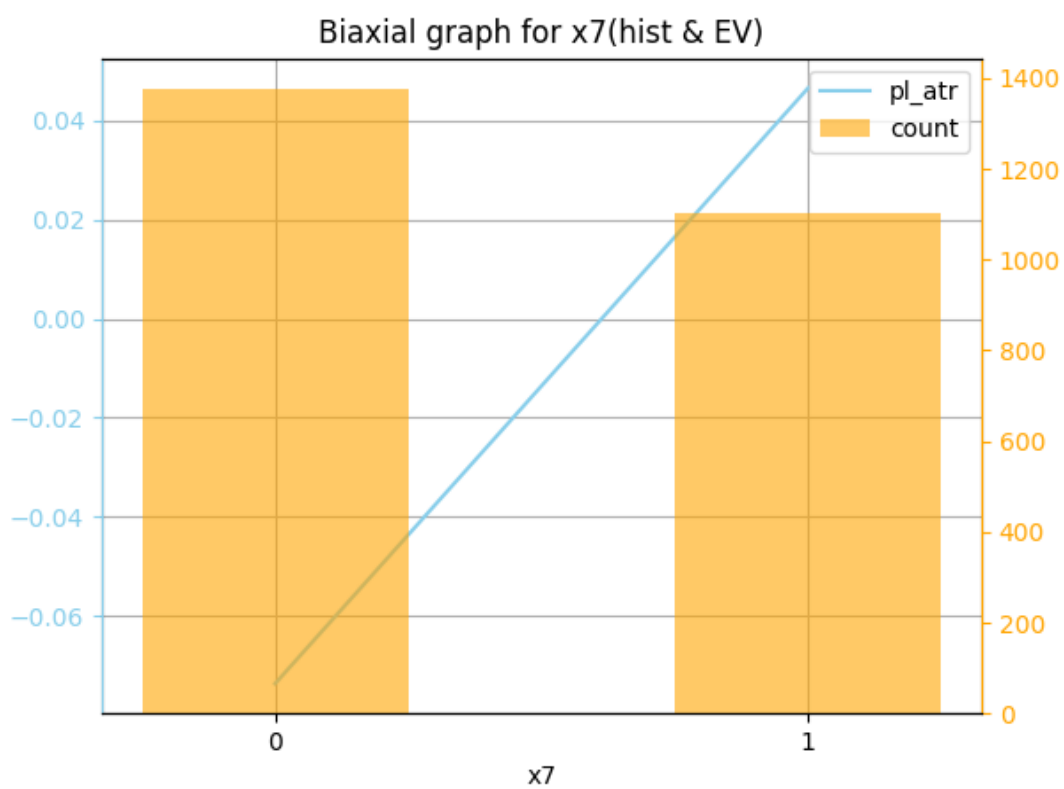
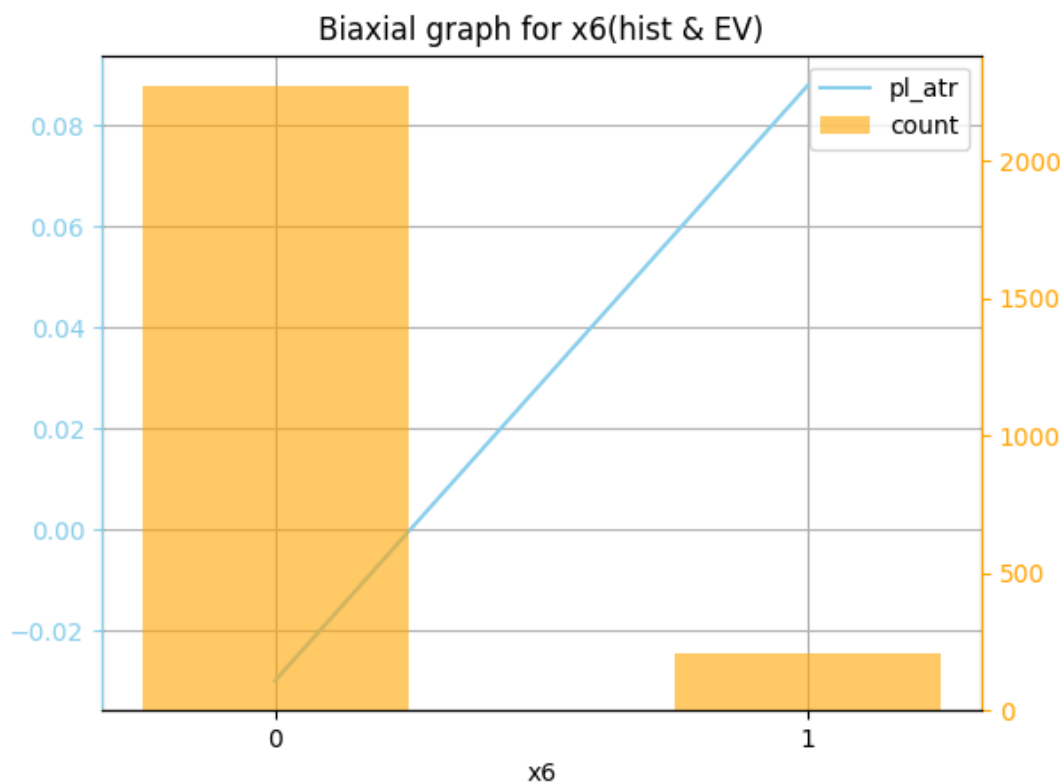




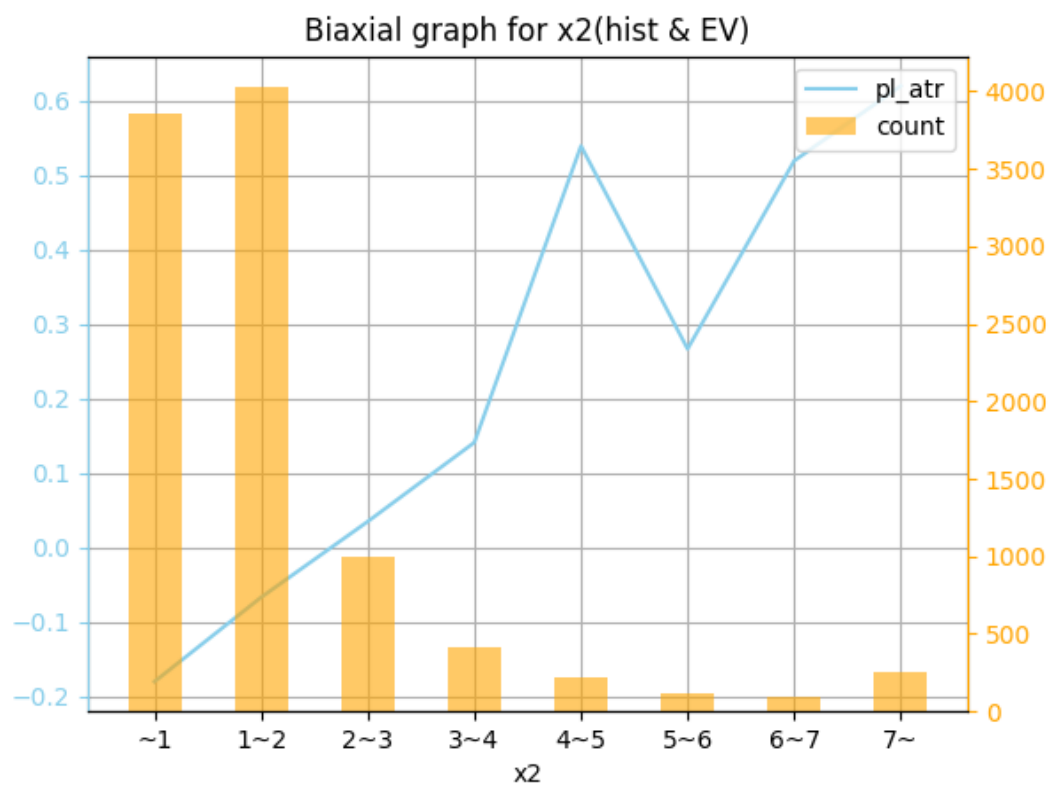


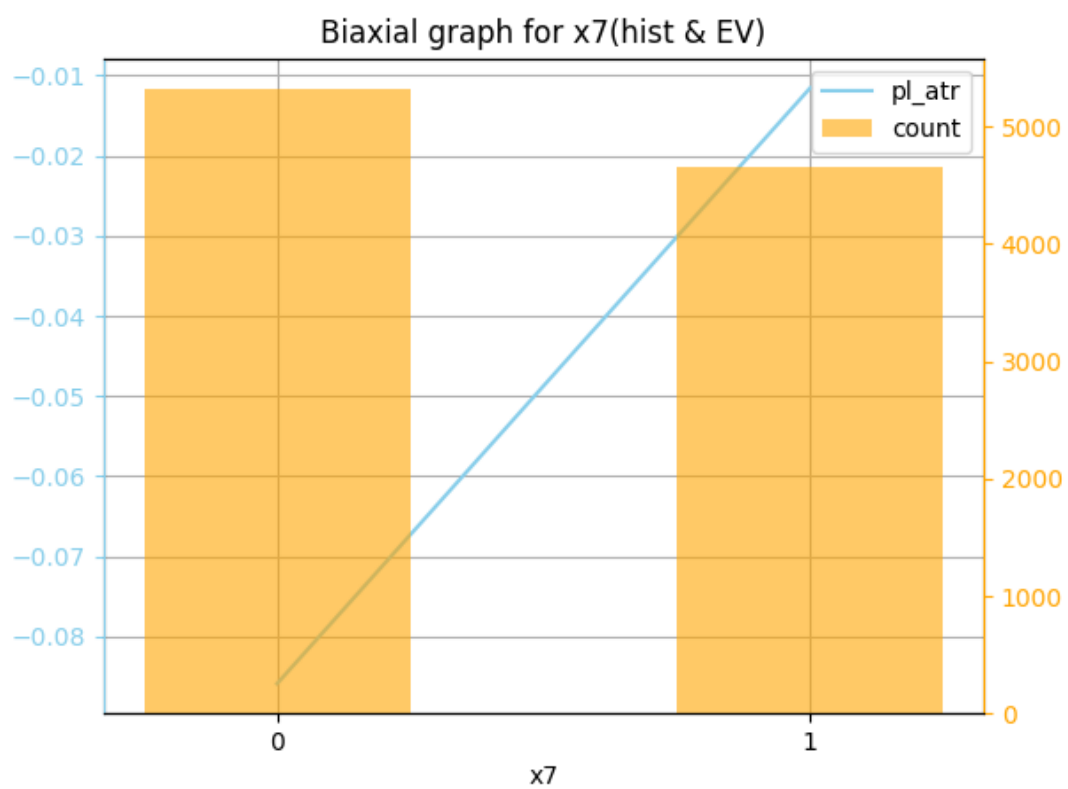
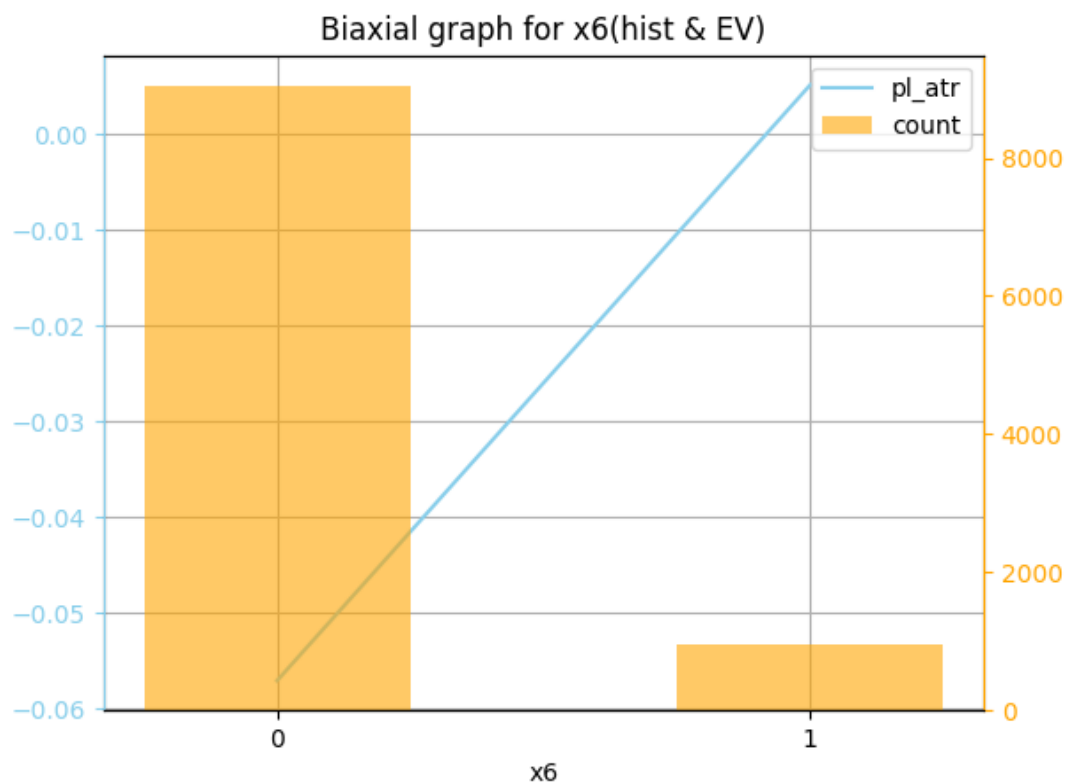
<data\_market\_catipalization\_top100>





<karioki3>





ここまでの検証をまとめると、 $x_2$  について以下の特徴が確認できます。

傾向 1: 損益分岐点( $pl\_atr=0$ )を大きく上回る水準の期待値領域が存在( $pl\_atr>0$ )

傾向 2: 比較的ハッキリとした線形性

これらの特徴が全てのデータセットにおいて共通して確認されたこと、また期待値プラスの領域が際立って確認された変数( $pl\_atr>0$ )が  $x_2$  のみであったことから、 $x_2$  に絞って優位性領域の探索を行います。

( $X_2$ ) × (その他説明変数)について  $pl\_atr$  との関係をそれぞれ 2 次元ヒートマップ図で可視化しました。

(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

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

figure1

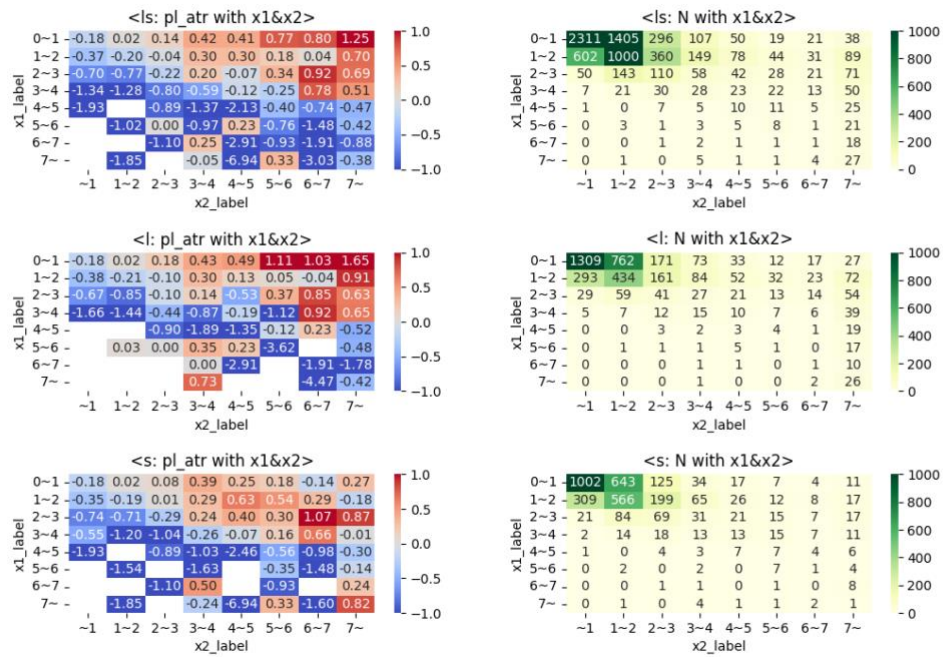


figure1

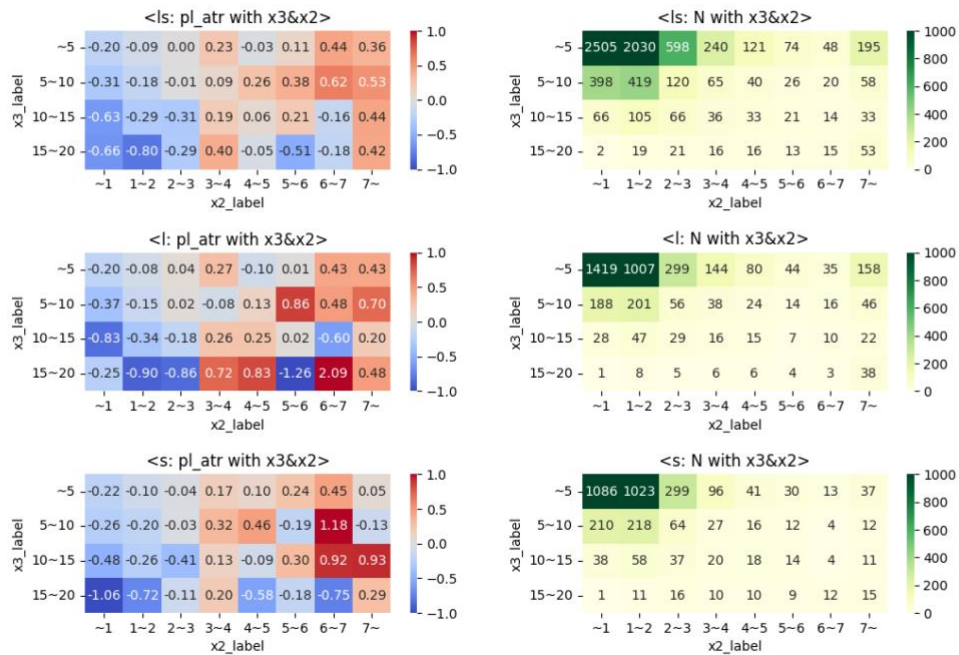


figure1

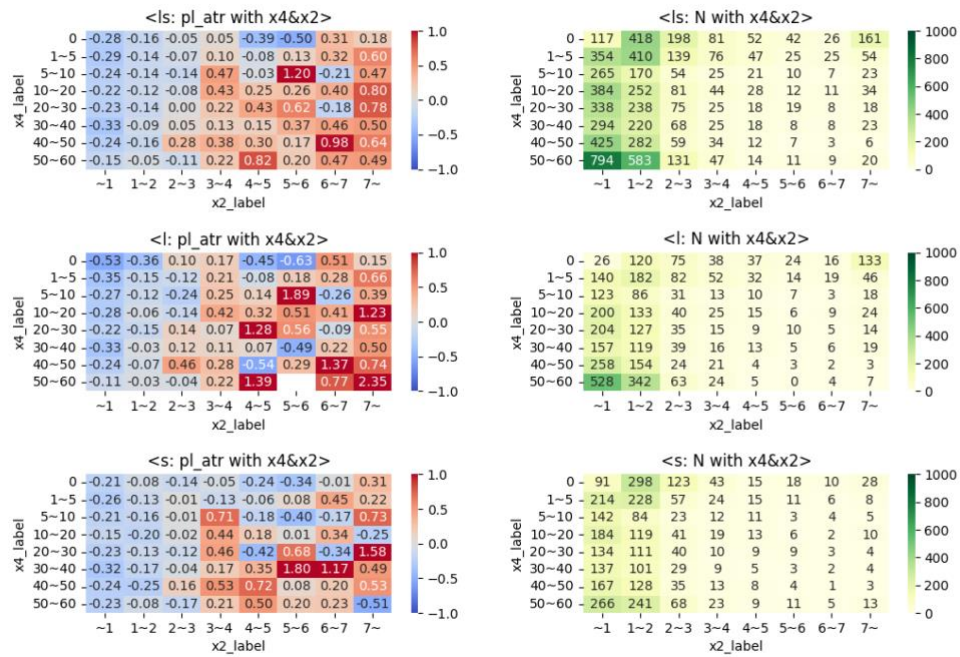


figure1

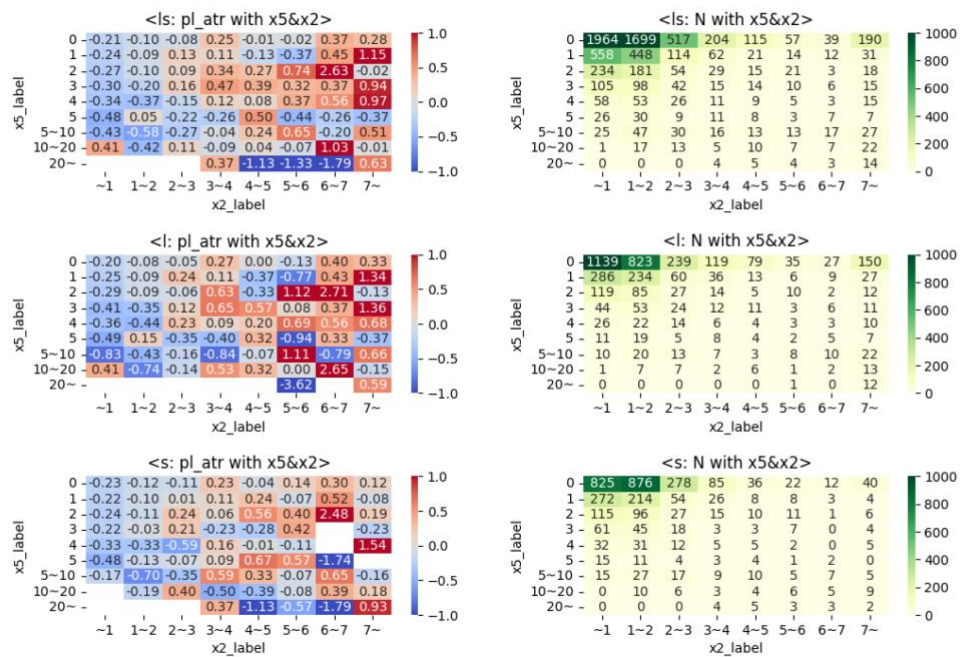




figure1

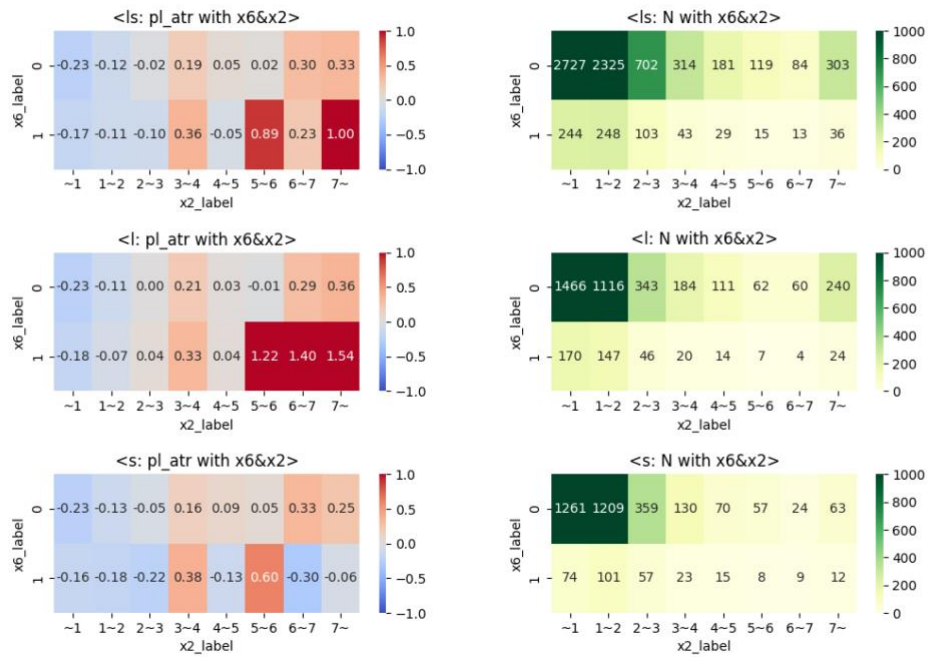


figure1

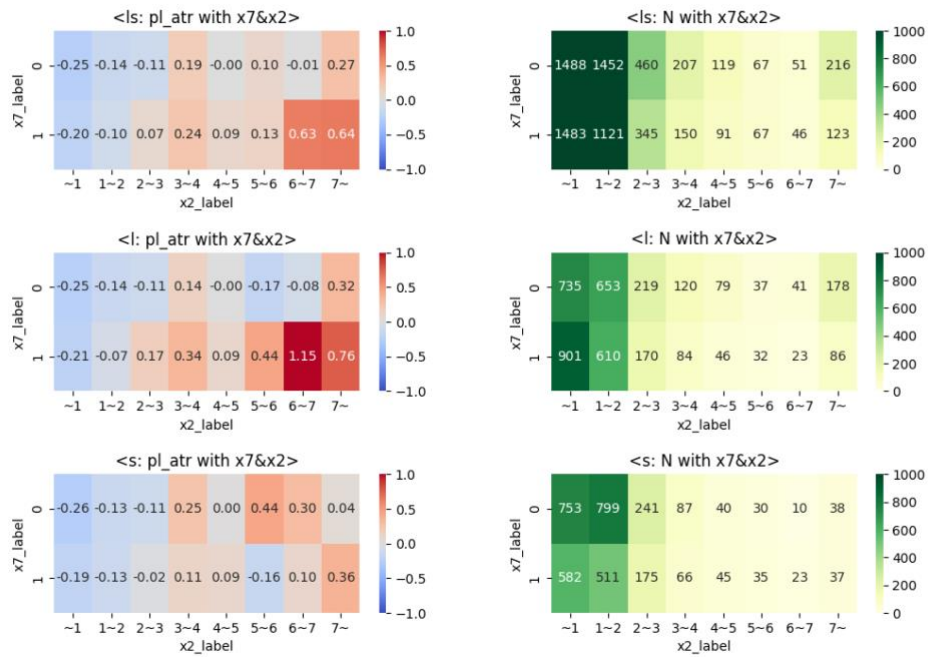


figure1

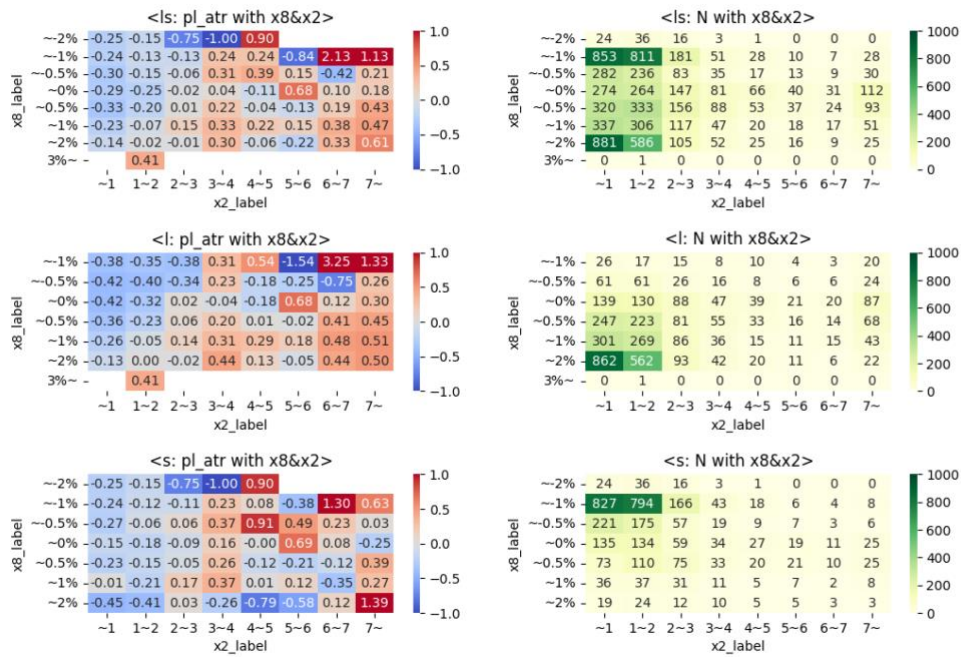


figure1

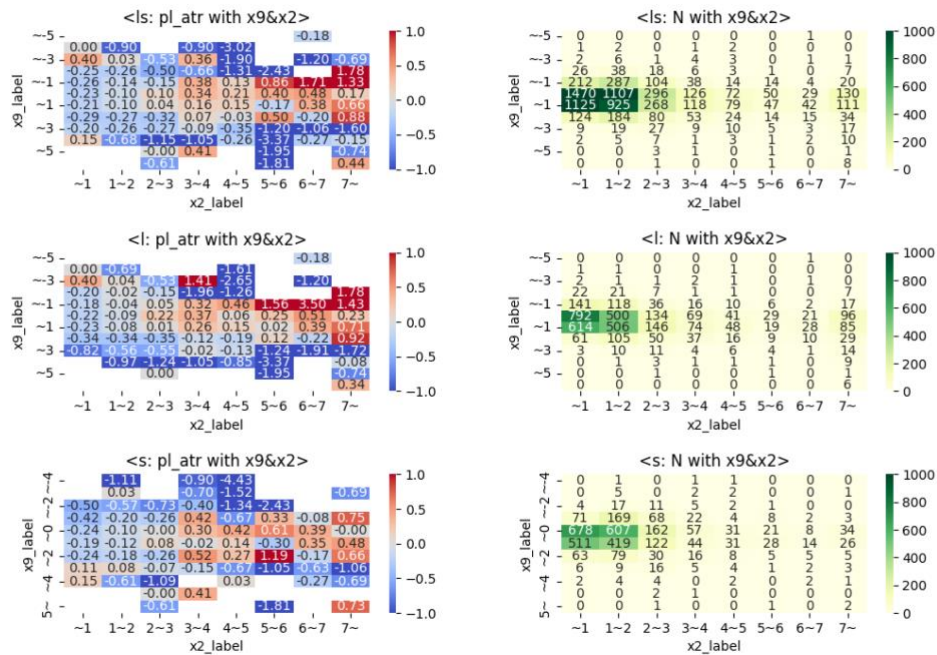


figure1

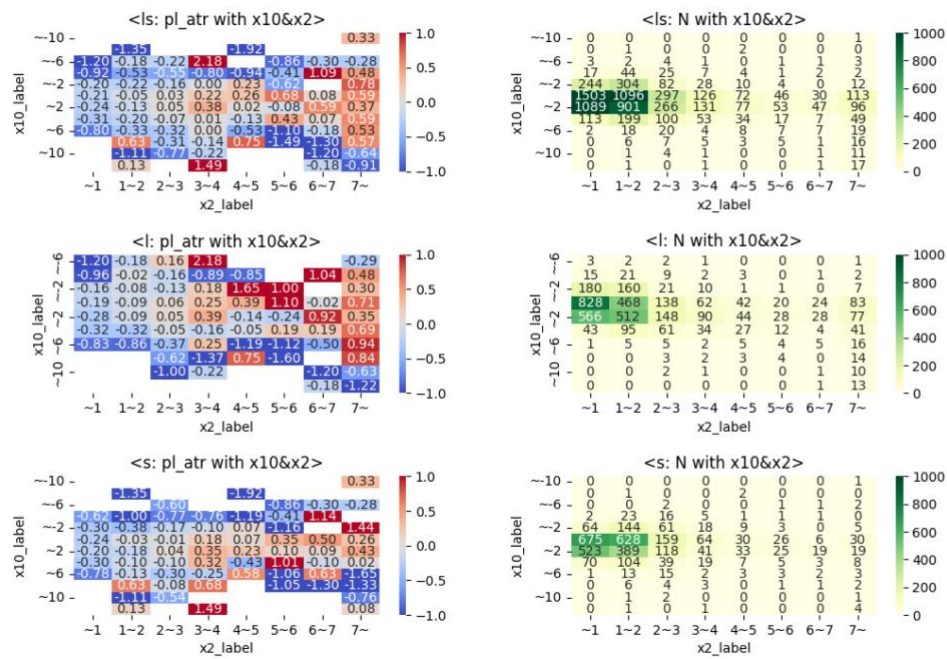
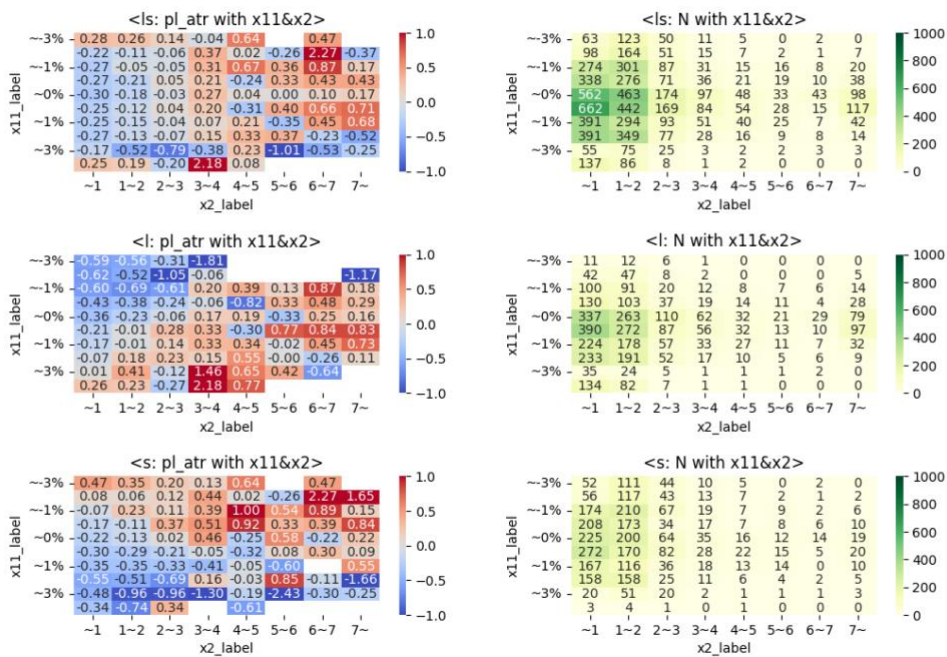


figure1



マップ左側が各ビン区間における pl\_atr の期待値 EV，マップ右側がサンプル数 N を示しています。

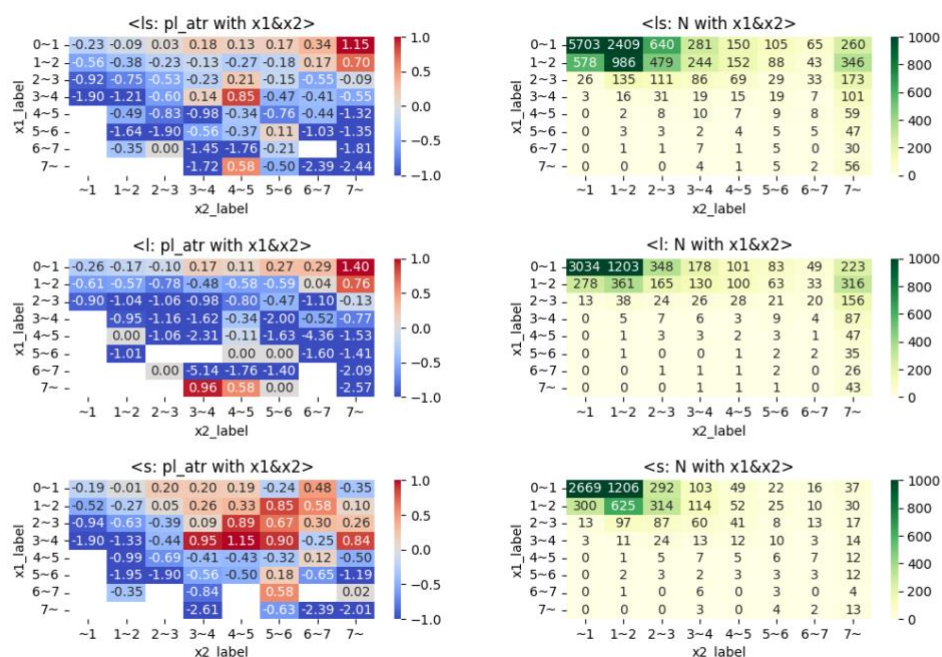


データ全体を眺めると、左ヒートマップの右側領域が赤に染まっている傾向が見られます。これは説明変数 x2 に pl\_atr との線形性が見られたことから、先程の 1 次元データの結果を反映したものといえます。

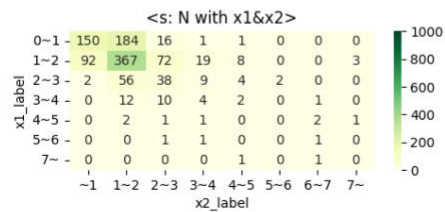
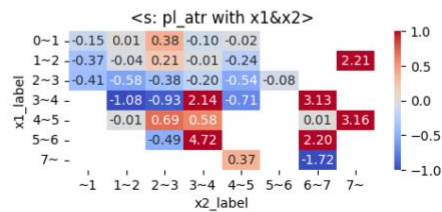
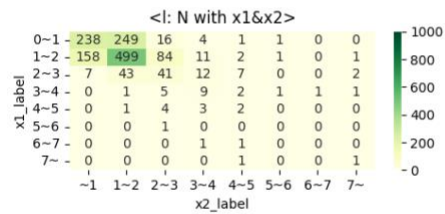
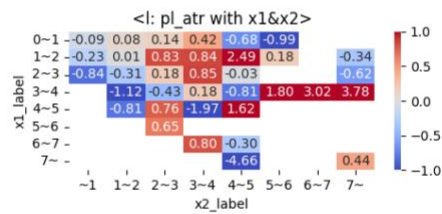
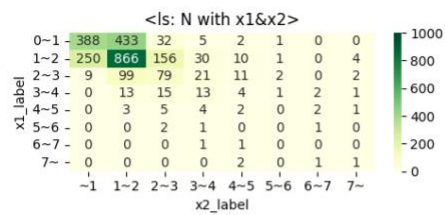
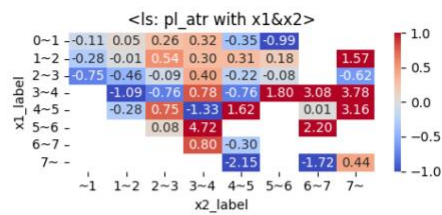
その上で、この領域を更に区分する説明変数が見つかるか？という尺度で全てのデータセットについて見ていきます。その結果、x1, x7 について特徴的な傾向が見られたので、その概要を順にまとめていきます。

- ・ x 1 について

<borokabu\_100\_300>

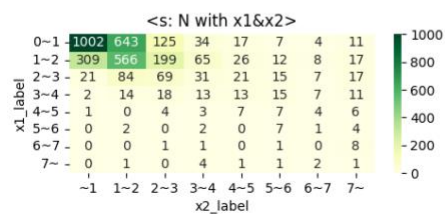
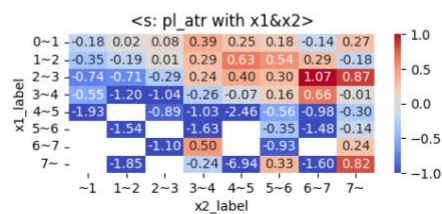
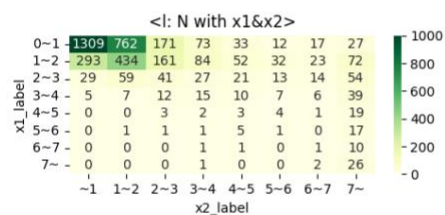
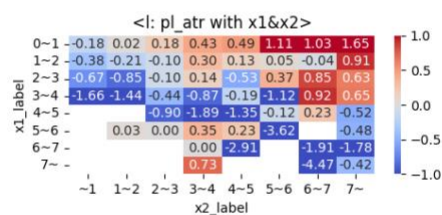
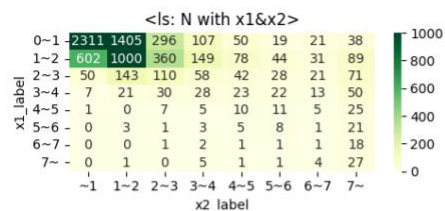
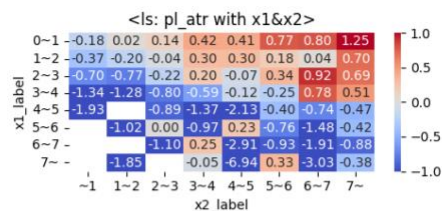


<capital100>



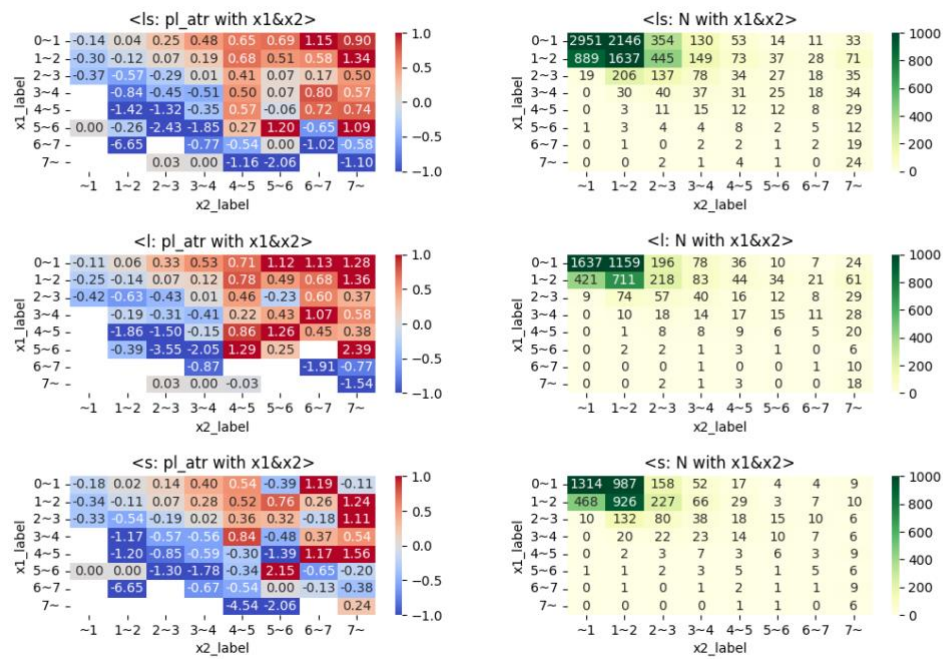
<karioki2>

figure1



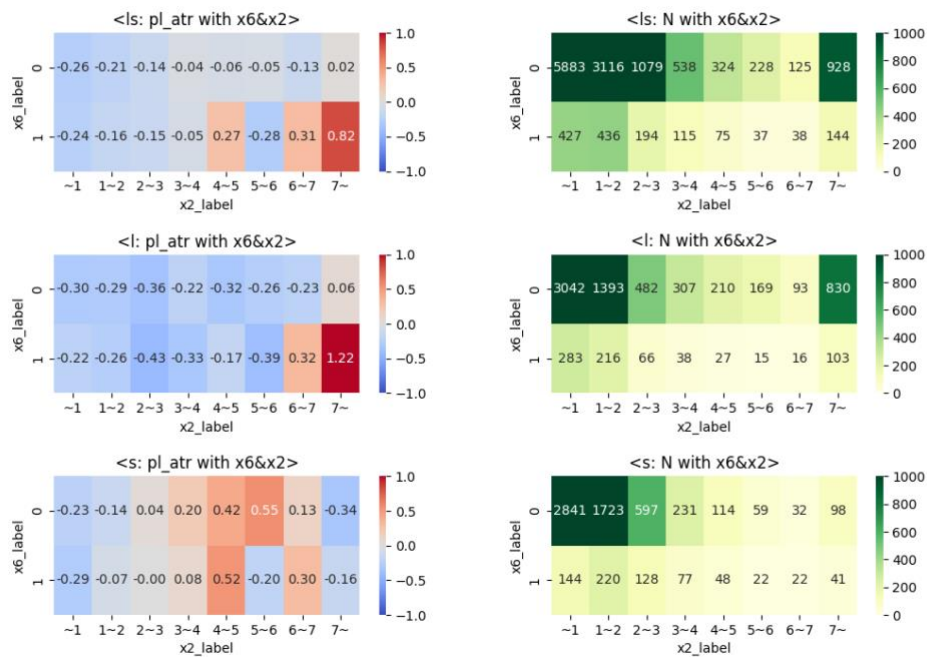
<karioki3>

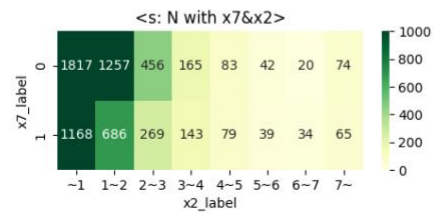
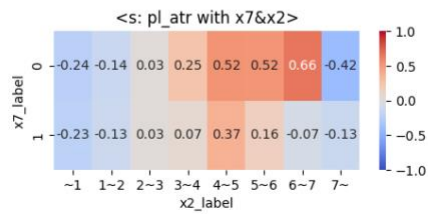
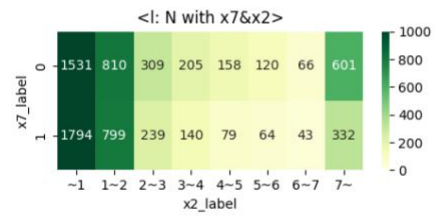
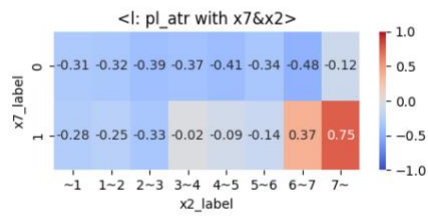
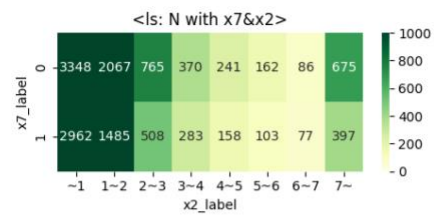
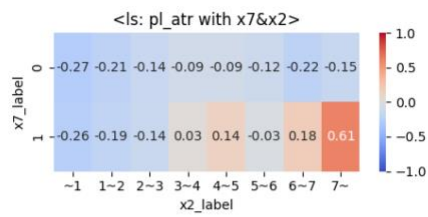
figure1



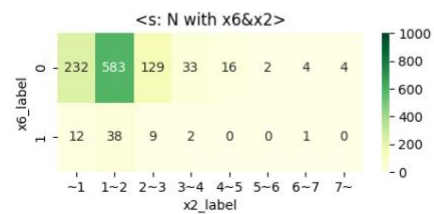
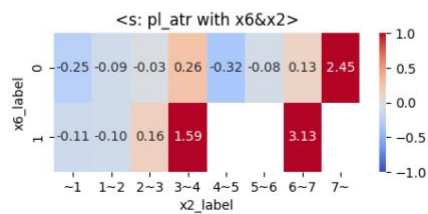
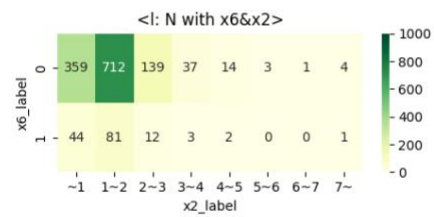
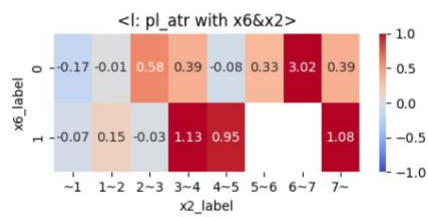
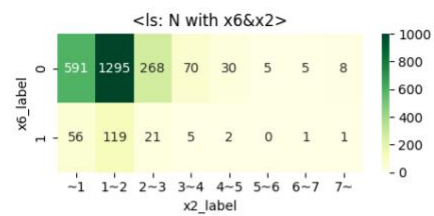
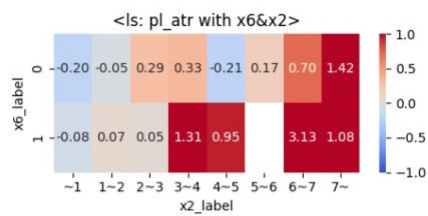
• x 7 について

$\langle borokabu100\_300 \rangle$

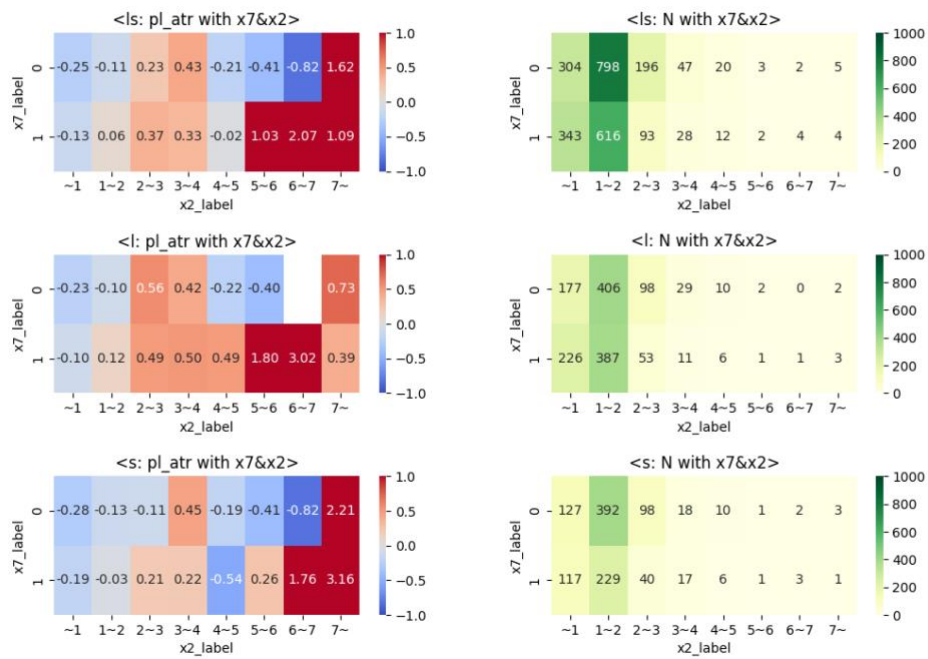




<data\_market\_catipalization\_top100>







<karioki3>

figure1

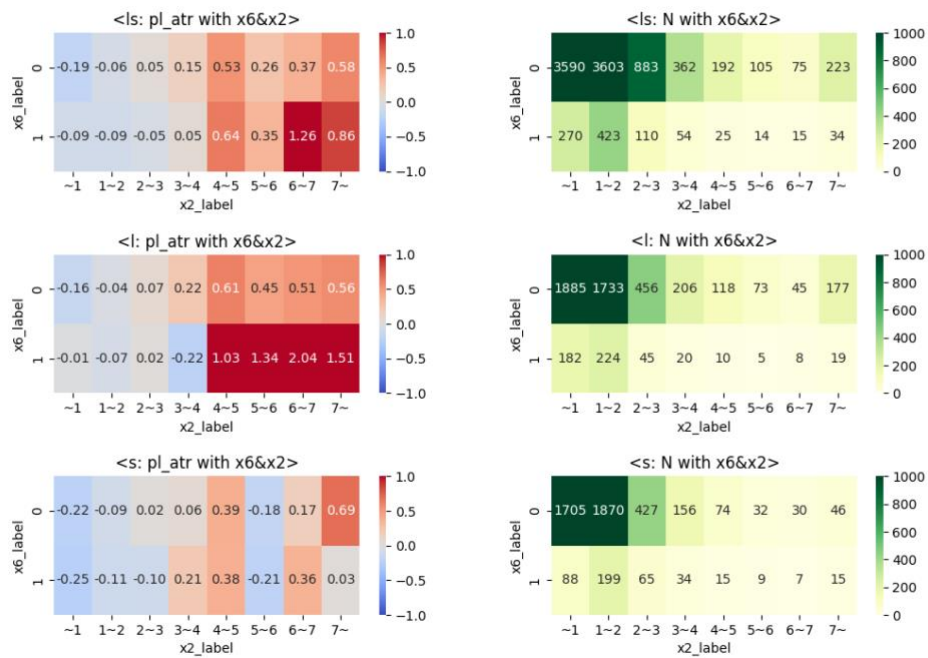
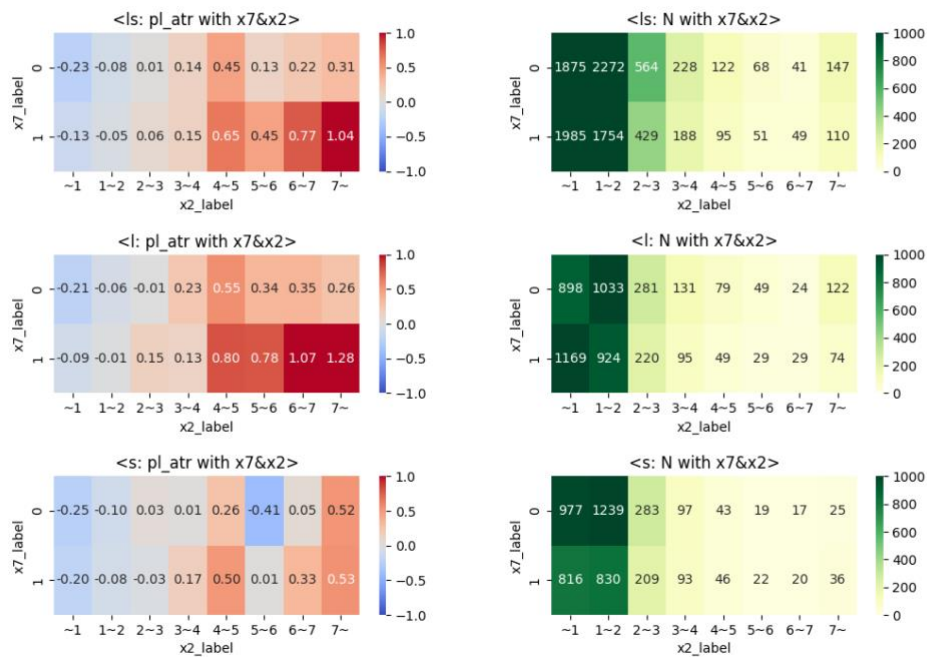




figure1



・上記の傾向を踏まえた優位性領域の選定

ここまでの結果をまとめると以下のようになった。

- 1:あらゆるデータセットにおいて  $x2$  が優位性を形作る最も大きなファクターとなっている
- 2:ただし、その優位性領域の水準（優位性が発現し始めるパラメータ値）については各データセット間で異なる（その背景的理由についての考察は後述）
- 3: $x7$  は  $x2$  に依存しない（独立）、相対パフォーマンスが高い
- 4: $x1$  は小さいほどいい。（その理由については後述）
- 5: $x1$  がどこまで許容されるかはデータセットによって異なる。このことは市場区分がパラメータ値の決定に影響を与えている可能性が示唆される。

これらについてそれぞれ考察していきます。

検定は2群比較くらいしかできそうにないな。恣意的に優位性領域を区切っているから。絶対リターンを追求するという意味 ( $pl\_atr > 0$ ) での統計的有意差があるかという問題と、その変数が優位性を切り分けているか(相対リターン)の問題の2つの問題がある。検定を行う際もそこを区分して考える必要がある。

期待値領域の区間推定。T 推定。

・検証結果についてのファンダメンタルズ面から見た考察

borokabu 群だけ優位性領域が狭く出た理由を長期保有の大口による買い(機関投資家、個人投資家を含む長期投資家による買い切りの買い)が乏しいことが根本的なドライバーだと考えられます。これらのセクターでは短期的な急騰を狙う短期投資家の資金ウエイトが大きく、長期保有主体の大口が長期的に保有する背景的理由が乏しいと思われます。他にも個人投資家のしこり玉(含み損ポジション)が多くあることも、その他の領域で優位性を維持できない理由になっていると考えられます。つまり、他の3つの群と比較して、長期的な資金が入りにくい市場環境に鳴っていること、資金が抜けていくスピードが最も早いと考えられることが、この群の優位性領域が狭い理由と考えられます。

~~大口や長期個人投資家と一緒に買い、大口が売るより先に売り抜けるイメージ。その際、短期トレーダーが占める割合が大きすぎると優位性を失っていく(主要な投資主体と同時に買って、早く売り抜けることが優位性の本質)。~~

~~優位性の本質はギャップアップの大きさは出来高急増(新規需要)に見合った水準になっているか?。価格が上がるほど売り需要が増えるため、X2 が大きいほど、上まで買いに行っていていいのは「グラフ上ではななめの領域に見えている」こと。背景はそうということだと見ている。~~

・データからは市場特有の背景が優位性領域の範囲について大きな影響を与えている可能性が示唆される。

・データから垣間見える可能性

順張り手法の検証結果から、断片的にはあるがギャップ幅が大きな部分(x1 が小さい)については逆張り手法が有効である可能性が示唆される。順張り手法において X2 と線形性が得られたことから、逆張り手法についても x2 と

・今回の結果から言える優位性の傾向と特徴

最も優位性に作用するのは出来高(X2)である。データを眺めると決算発表、上方修正がその背景として働いている傾向が見られているが、未検証。

X2 に対する背景要因。なぜセクター別でパラメータが変わったのか。

- ・実践に活用できる優位性領域の選定

今回の検証結果から、あらゆる市場において  $x_2$  に線形性が見いだされることが確認された。また、 $x_7=1$  フラグがあるかないかは、 $x_1$  については市場セクターによって上限が異なるが、どのデータセットにおいて  $0 < x_1 < 2$  は低リスク高リターンの領域であることが判明している。絶対リターンを追求するならば、この領域を掴む方法を考えることが必須となる。次に、相対リターンにおいては  $x_7=0, x_7=1$  で優位性を分ける働きがある。

しかも、 $x_7$  は前日から事前に参照できる変数であるため、実用が可能であり、かつ前日が陰線であるか陽線であるかの違いだけであるため、識別も容易い。

- ・現在の運用上の収益状況。楽天証券の口座。

~~→この手法が優位性を消失する可能性のある重大事項~~

~~日本株が24時間取引に移行する時、ただし、PTS 売買(夜間取引。jnx,chi-x)の状況を見ると、活発ではない。人間の生活サイクル等から馴染んでいない等の理由から普及までの道のりは険しいのではと感じています。~~

~~→企業の決算発表タイミングの主流が大引け後→ザラ場中に切り替わる時(アメリカではこちらが主流となってきた)→そもそも手法の前提となっているギャップアップ、ギャップダウンがなくなる。~~

~~→更新値幅ルールがなくなるとき。~~

~~ミスマイクが付きまくり、市場の混乱を招く恐れがあるのでわざわざこれを採用する合理性は薄いと思われます。~~

~~→検証されていないが、見落とせない重要点~~

~~→リーマンショック時のデータが含まれていないこと~~

~~デイトレードという手法の特性上、他の手法と比較してマクロ動向に対して抵抗性を持つことは伺える。今回の検証結果からも部分的ではあるが、少なくともマクロ動向(日経平均株価)に対して一定の抵抗性を持つことは伺えた。(x8,x11とpl\_atrの相関性は備考にまとめた)。年度別の収支グラフからも、年度によらず比較的安定した優位性を維持していることは確認された。しかし、異なるサンプル期間において本当の結果がど~~

~~うであるか確かめることは今後検証を精査していく中で最重要項目の一つとなる。~~  
~~→ATRを基準に pl を算出しているため、ATR の値が小さい場合 pl は大きな値（外れ値）となる。その影響を最小化する処理として、ATR がある一定値以内となる場合フィルターを設ける処理も考えられたが、ロジックが複雑になり、その基準にも一定の恣意性が含まれることからひとまず採用を見送った。念の為、収益の分布表を確認すると「優位性があると判断した領域」については特に目立った外れ値は確認されなかった（それ以外の領域については一部外れ値  $pl\_atr=33$  が観測された）ことから、少なくとも優位性領域については、このリスクはひとまず小さいと考えられる。~~  
~~→スリッページ、手数料が考慮されていない点~~  
~~板が厚く、流動性の高い銘柄群を扱っているため、全スクリーニング（3700 銘柄）の方法と比べてその影響は抑えられていると考えられるが、実際の売買ではスリッページは生じる。その影響を厳密に評価するためには、日足だけでなく日中足の情報が必要になったり、その分析に必要なデータを持ち合わせていないので現在のところは難しく、参考程度の評価しかできない。使用する証券会社の売買手数料ルール、~~

#### ・今回のデータの実践的活用方法

優位性は比較的広範囲に渡って見られるが、実践にあたっては、使える領域を信頼できる領域に慎重に絞る必要がある。

今回の検証結果の最も有用な使い道として考えられるのはヒートマップを期待値テーブルとして利用することである。特に x3, x7,x9,x12 については前日から値が確定していて、確定値として安心して利用できる変数である。これによりこの範囲においてはあらかじめ期待値テーブルを自分でコントロールすることが可能となる。X2 においては事後変数であり、あらかじめ値を特定できない。なので、x2 がどんな値になったとしても期待値プラス ( $pl\_atr>0$ ) を理論上キープできる領域期待値テーブルを利用するという方針が手堅い。運用イメージとしては、優位性領域をえらぶというよりは、無駄な領域を掴まないようにするという使い方がベターだろう。

特に x7 についてはカテゴリー変数であり、その差は目視でも活用しやすい。

#### ・データ活用にあたっての壁

唯一線形な関係が見られた肝心の説明変数 x2 が事後データで、事前に参照できないデータである。ただしアプローチする方法がないわけではなく、寄付出来高と 5 分足データとの比較から予測式を算出することにより x2 のボリュームそのものの予測が可能だと考えられる。直接事前に参照できる方法があればいいのですが、公開データの中では

取得できる媒体を見つけることができませんでしたので、得られたデータの中で。

- ・機械学習の可能性

現在のところ、ドライバーの背景的理解と可視化に重点を置いていて、古典的な記述統計的アプローチでの優位性検証を進めている段階です(まだ機械学習の学習は進んでおりません！)。ただ、~~今後のシステム実装を踏まえるにあたって、得られた優位性をモデル化して、システム運用がしやすい形へ落とし込んでいくという流れは生じてきます。現在では個人投資家向けにシステムトレードを実現するためのツールも充実してきており、(2021/6月に開始された楽天RSSではExcelマクロ機能を使って自動発注システムを構築することが可能となった)、自分で開発したロジックを実際の株式相場に投入できる機会が増えてきています(fxや商品先物市場では以前からMT4などのツールでシステムトレードすることが可能だった)。~~

各変数間の相関から、なにが擬似変数になっているかの確認。機械学習モデルに組み込む時に特徴量抽出に使う。

ヒートマップで確認すると優位性領域が斜めに分断されていたことから決定木モデルよりも、サポートベクターマシン、k近傍法を試してみたい。

- ・年度別の収益結果を算出してパフォーマンスのバラツキを確かめる

- ・運用の方針

現在のところ、事後変数が優位性の源泉となっていることから、優位性領域をエントリー時にドンピシャで狙うことはできません。なので、運用の大きな方針としては、「成功したら利益が大きい、失敗してもリスクは小さい」という優位性領域を見出し、活用する必要があります。

~~→今回の手法でもたらしえた優位性には実際には2つの優位性が混合している。損切りルールによる優位性と手法そのもの(エントリータイミングと決済タイミング)によってもたらしえた優位性が存在する。その分離方法をまとめておく。損切りルールによりもたらしえた優位性は今回の検証のメインではないので外した。検定したいのは後者。ただし、その検証方法は存在する。~~

~~正規分布を想定したときの  $p < 0$  部分を  $p = 0$  にすべて丸め込んでいるから、中心から左側部分で盛り上がっている。だとしたら、2つの面積を比べれば損切り効果がわかる。~~

~~手法そのものの優位性は平均値  $x$  のランダムウォーク  $\mu=0$  からのズレである。~~

#### ~~→今後の方向性~~

~~事前に参照可能な変数  $x_3, x_7, x_9, x_{12}$  を使って重回帰分析を実施して、適当な重み付けをする必要がある。その際に、各変数間にはある程度の相関が考えられるため、相関度合いを反映させて次元削減処理を行う必要がある（特に  $x_7, x_9, x_{13}$  は前日の動きを測定する似たような機能を持つパラメータであることから相関は強い）。~~

#### ・備考欄

備考1：データセットに用いた銘柄選択の恣意性の影響（karioki2, karioki3）について

サンプル期間(2015-2021)を十分長く設定すること(銘柄選定については楽天証券のスマホアプリから行っている。直近の株価の動きを反映して無意識に銘柄選択を行っている可能性があるが、楽天証券のスマホアプリからは直近1年間のデータしか参照できない為、それ以前のデータについては検証者自身のバイアスは関与していないと考えられる。このことから直近1年間の収益結果についてはバイアスの関与を疑う必要があるが、それ以前の結果についてはその影響は小さくなり、無作為抽出でサンプルを集めた場合に近づいていくと考えている)、サンプルサイズ  $n$  を十分に大きく取ること ( $n=54,59$ )、恣意性が混入しにくい、かつ選定した銘柄群とは異なる特徴(※)を持ったデータセット data\_market\_capitalization\_top100, borokabu\_100\_300) を併用して総合的な判断を行うことにより、その影響を抑えることに努めた。

※異なる特徴とは時価総額の大きさを観点としている。Karioki2,karioki3 のデータ群は300億-1000億程度の時価総額を持つ銘柄のウェイトが大きいため、時価総額が圧倒的に大きな銘柄群(時価1兆円以上の超大型株)、圧倒的に小さい銘柄群(100億未満の株)をデータセットに集めた。

#### ・このデータセットを用いた経緯と妥当性の考察

当初は上場銘柄全て(約3700社)をデータセットとして用いる、もしくはこれらのデータの中からいくつかをランダム抽出する方法を検討しましたが、主に以下の理由からその方法を回避しました。

1:優位性が示唆される領域がそもそも見つかるかどうか。その可能性そのものが未知で

あったので、まずは小さなサンプルを使ってアタリをつけたいと考えた。

2:流動性を考慮すると実務的にトレードできる銘柄は実際にはかなり限られている(上場銘柄の半数にも満たない)。このような中でランダム抽出を行うと母集団(実際に売買できる)から外れたサンプルが標本内に多く含まれてしまい、そのことが検証結果を歪めてしまう懸念があった。

こうした背景から、今回の検証ではその代替手段として 4 つのデータセットを標本として用いることにしました。その際、使用するデータセット数(標本数)を 1 つではなく 4 つ用いることにしたのは、それぞれのデータセットが固有に持つ特徴(時価総額、市場区分、ファンドの組み入れ、長期投資家の保有割合など)について、それを区分するのに適した変数を検証者自身が上手く表現できなかったからです。各データセットの特徴が結果に影響を与えている場合であっても、その違いを捉えられるようにしておくために、簡易的な措置として 4 つのデータセットを用いることにしました。

時価総額で区分したデータセットを 2 つ(価格形成の背景が異なる可能性が高い群)、検証者自身が流動性が十分にあると判断したデータセット(有意抽出) 2 つ、計 4 つのデータセットを用いて検証を行うことにしました。

ただし、選んだサンプルについては検証者の趣向性やバイアスを反映している可能性があるため、検証者自身が配慮した点について別途記載しました(※備考 1:データセットに用いた銘柄選択の恣意性の影響について)。

- ・ 検証に使用したツール

Python (Pandas, matplotlib, seaborn)

- ・ データ収集

楽天証券口座における時系列データを使用(株式分割後の株価修正済データ)。

~~相関係数からはどの変数も `pl_atr` に対して顕著な傾向は確認されませんでした。(散布図からも同様)。~~

~~ただ、ここで株式収益データが持つ特徴について考察しておく必要があります。~~

仮説：株式市場の収益データにおいて、長期的に目に見えるほど強い相関を示す変数を得ることはあまり期待できない（相関が強いほど他者の目から見ても収益性がわかりやすくなるので、超過リターンを奪い合うゲームの中では強い相関の存続は難しく、結果として収益結果データは長期的にはランダム性の強い収益分布に収束していくと考えられる（例外として、市場自体がまだ未成熟であったり、twitterのテキストマイニングなど他者から見えにくい、かつ参入障壁の高いロジックを駆使した手法の中には視覚的にも分かりやすい優位性が眠っている可能性がある」と個人的には考えています）。したがって、今回のような古典的なチャート分析、テクニカル分析を用いた手法検証の場合、ロジックそのものが他者にも再現しやすく、既にその優位性に対する折り込みが進んでいる可能性は十分に考えられる。結果として、今回のようなローソク足を題材にした手法検証にあたっては、どの変数も統計上の相関が人間の目では比較的に見えにくいことを前提に、一見ランダムに見える中から弱い相関を注意深く探っていく姿勢が重要になると思われる。