

FX 証拠金取引の投資パフォーマンスの決定要因

- 機械学習を用いた分析 -

岩壺 健太郎
(神戸大学 教授)

要旨

FX 投資家の個人取引データから算出された建玉の保有期間、レバレッジ、勝率、リスクリワード比と月間収益率を使ってクラスター分析をしたところ、スキャル投資家を中心としたクラスターが 2 つと、デイトレおよびスイング投資家を中心としたクラスター 1 つに分類される。前者は収益率が高く、レバレッジが低く、勝率が高いクラスターとその逆の収益率が低く、レバレッジが高く、勝率が低いクラスターに分類され、3 つ目の保有期間が長いクラスターの特徴は前者 2 つのクラスターの間位置する。この結果は保有期間で分類し、そこに属する平均的な投資家の取引を分析することでは得ることのできなかつた結果である。また、ランダムフォレストによって収益率の予測モデルを推計したところ、説明変数の重要度が高い順に勝率、リスクリワード比、レバレッジとなり、保有期間の重要性は非常に低いことが明らかになった。株式投資では長期投資が推奨されることが多いが、FX 証拠金取引では長期投資と短期投資のどちらのパフォーマンスがいいかという議論は必ずしも重要ではないことを示している。

I はじめに

本協会会報 No.114 に掲載した拙論「FX 証拠金取引の投資戦略とパフォーマンス」(2017 年 10 月)では、FX 証拠金取引を行う投資家の中で、ポジションの保有期間が長い投資家ほど投資パフォーマンスが高いことを示した。具体的には、保有期間が 1 週間以上の長期投資家の収益率が最も高く、スイング投資家(保有期間が 1 週間以内)、デイトレ投資家(1 日以内)、スキャル投資家(1 時間以内)の順に収益率が悪化することを示した。しかし、そこで行った分析は各投資家グループの平均的な収益率を比較しているにすぎない。グループ内の異質性(Heterogeneity)が高い場合、保有期間で類型化することは問題かもしれない。たとえば FX 投資家の約半数に上るスキャル投資家は、平均的には収益率が低いとはいえ、非常にパフォーマンスが優れた投資家が多いことも知られており、同グループ内での収益率の分散が大きいことが想定される。

そこで、本研究では保有期間が同じ投資家を 1 つのグループにするのではなく、クラスター分析によってデータの特性に基づいた分類を行う。その際、勝率やリスクリワード比、レバレッジといったパフォーマンスに影響する変数も同時に分析に加えて、どの変数が類型化に影響しているのかを検証する。また、ランダムフォレストという機械学習の手法を用

いて、パフォーマンス予測に対する説明変数の重要度を比較する。

機械学習は学習の方法から、「教師あり学習」「教師なし学習」「強化学習」の3つの枠組みに大別される。教師あり学習とは学習データに正解を与えた状態で学習させる手法であり、教師あり学習で解く問題として代表的なものに「回帰」と「分類」がある。回帰とは連続する数値を予測するもので、分類とはあるデータがどのクラスに属するかを予測するものである。ランダムフォレストは、データを条件分岐によって分割していくことで分類する「決定木」という手法を拡張させ、ランダムにデータを選んで決定木を複数作り、それぞれの決定木で予測された結果の多数決により、最終的なクラスを判定する手法である。他方、教師なし学習とは学習データに正解を与えない状態で学習させる手法であり、その中のアルゴリズムの1つとして、データ間の類似度にもとづいてデータをグループ分けするクラスタリングがある。本稿では、収益率を勝率やリスクリワード比、レバレッジと同様、説明変数の1つとして用いて、教師なし学習のクラスター分析を行う他、教師あり学習のランダムフォレストによって、収益率を予測する説明変数の分析を行う。

FX 証拠金取引において、利益を出せるかどうかは勝率とリスクリワード比の組み合わせによる。勝率が高いとリスクリワード比（1取引当たりの平均収益額/平均損失額）が低くても利益を出せるが、勝率が低いならばリスクリワード比が高くないと利益を出せない。このように、勝率とリスクリワード比は投資戦略上、重要な変数である。勝率やリスクリワード比は完全にコントロールすることはできないものの、これらの変数を念頭に置きながら投資戦略を構築することが多く、パフォーマンスに影響する。

レバレッジも新規ポジションを持つときにはコントロール可能な変数であるが、ポジションが悪化すれば実質レバレッジが高まるように、完全にコントロールすることはできない。これもパフォーマンスに影響する変数として考慮すべき重要な変数の1つである。

II データ

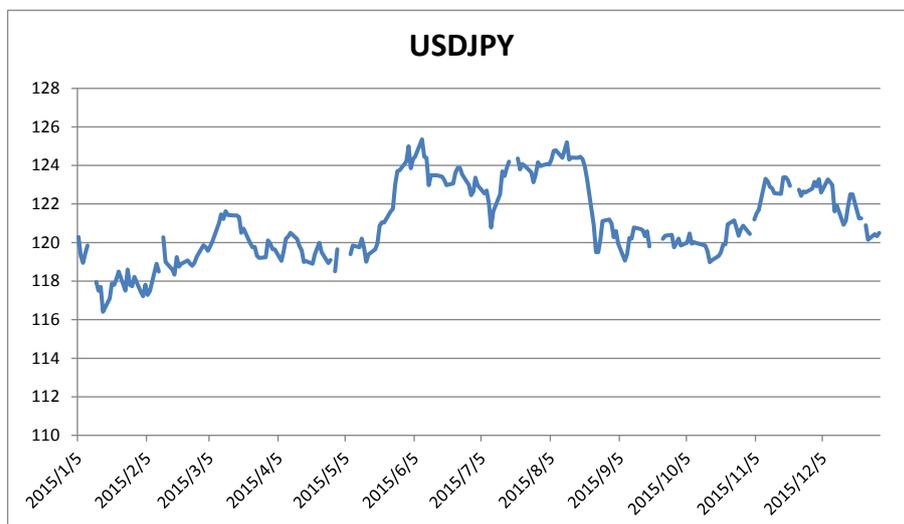
本稿では、金融先物取引業協会の会員 FX 会社から提供された顧客取引データを用いて分析を行う。サンプル期間は2015年1月から2月までの2カ月間である。図1は2015年のドル円レートの推移を描いたものである。2013年～2014年の急激な円安が一服し、トレンドが容易につかめないう時期をサンプル期間としている。

分析には、取引データと日次証拠金データの2種類のデータを用いる。取引データは、顧客ID（個人情報保護のために変換されたもの）、通貨ペア（22通貨）、約定日時、取引の種類（新規、仕切）、売買、約定価格、約定数量が記録されており、証拠金データは、ハッシュ処理済み顧客ID、日付、預託証拠金額、有効証拠金額、必要証拠金額、評価損益、実現損益、スワップ損益が含まれている。

同じ市場環境での投資家間の比較を行うため、1カ月ごとに1,000人の投資家をランダムにサンプリングし、同じ分析を2回行うことで、異なる経済環境下でも同じ分析結果がみられるかどうかを確かめる。ただし、以下の条件を満たす投資家の中からサンプリングして

いる。その条件とは①月に 1 回以上取引をしている投資家、②前月末時点でポジションを保有していない投資家、③月初に預託証拠金を差し入れている投資家である。

図 1 2015 年のドル円相場



(出典) 日経 FQ から筆者作成。

(1) 収益率

月ごとに 1,000 人の投資家の月間収益率を計算し、クロスセクション分析を行う。1 カ月を超える取引を分析対象から外すというのは分析上の欠点ではあるが、その割合は非常に小さく、また、長期の収益率の比較だと同じ経済環境という仮定から逸脱する可能性があるため、やむを得ない選択と考えている。長期だと保有期間や戦略を替える投資家もいるため、投資家間の比較が困難になるという理由もある。

含み損益を考慮に入れた月次の口座評価額収益率は以下の式で表される。

$$\text{日次口座評価額グロス収益率}_t = \frac{\text{有効証拠金}_t + \text{出金額}_t}{\text{有効証拠金}_{t-1} + \text{入金額}_t}$$

ここで、 t は t 日の終わりの時点 (NY クローズ) を指している。入出金額は

$$\text{入出金額}_t = \text{預託証拠金}_t - \text{預託証拠金}_{t-1} - (\text{確定損益}_t + \text{スワップ損益}_t)$$

によって算出する。この値が正のときは入金があったとみなし、負のときは出金があったとみなす。月次リターンは

$$\text{月間収益率}_t = \left(\prod \text{日次グロスリターン}_t - 1 \right) \times 100(\%)$$

ここで、注意すべきは、入出金は 1 日に 1 回しか行われないと仮定していることと、入金はその日の投資に使われ、出金はその日の取引の後に行われると仮定していることである。入出金の時点が分からないため、仕方のない仮定であるが、日次収益率が極端に高い、もしくは低い場合はサンプルから省くことで、その影響を軽減した。

(2) 保有期間

各投資家を保有期間に応じて、スキャル投資家、デイトレ投資家、スイング投資家の3つに大別する。スキャル投資家は保有期間が1時間以内、デイトレ投資家は1日以内、スイング投資家は1日超の保有期間を有していると仮定する。1カ月以上の取引は分析の対象外とする。

分類の仕方は以下の通りである。まず、各投資家のすべての取引を通貨ごとに新規約定から仕切約定までの保有期間をもとにスキャル、デイトレ、スイングの3つに分類する。その上で、ある分類がその投資家の仕切約定回数の5割を超えている場合、当該分類を当該投資家の保有期間と決定する。たとえば、スキャルが6割、デイトレが2割、スイングが2割の投資家はスキャル投資家となり、いずれの分類も5割未満だと分析対象から除くことにする。また、スキャルが5割、デイトレが5割の投資家は保有期間が長い方、つまりデイトレ投資家となる。

(3) レバレッジ

投資家が収益を確保するためにはリスク管理が欠かせない。ここでは、リスクの代理変数としてふさわしいレバレッジを分析対象とする。1日の終わり（NYクローズ時）のオーバーナイトのレバレッジを計測する。計算式は以下の通りである。

$$\text{月次レバレッジ} = \text{日次レバレッジ} \left(\frac{\text{必要証拠金}_t \times 25}{\text{有効証拠金}_t} \right) \text{の月間平均}$$

なお、レバレッジがゼロの日は平均から省いた。

(4) 勝率

各投資家の月別勝率を仕切注文ベースの取引数のうち、実現益を出した取引の比率を勝率とする。各取引の約定価格と約定数量をもとに実現損益を算出し、実現損か実現益かを判定する。

(5) リスクリワード比

リスクリワード比は1取引当たりの平均実現益を1取引当たりの平均実現損で割ったものを指す。ただし、分母である実現損が0の場合、つまり勝率が100%のときはリスクリワード比を算出することができない。そこで、リスクリワード比の算出にあたり、勝率100%と勝率0%の場合を省いている。また、クロス円以外の通貨ペアについては、その約定時の円レートで円換算しなければならない。22通貨ペアのうち、実現損益の計算が容易な10個のクロス円の通貨ペア(USD/JPY、EUR/JPY、GBP/JPY、AUD/JPY、NZD/JPY、CHF/JPY、CHF/JPY、HKD/JPY、CNH/JPY、ZAR/JPY)と、取引量が多いドルストレート3通貨ペア(EUR/USD、GBP/USD、AUD/USD)は円換算することで実現損益の算出を行ったが、

その他の通貨ペアについては取引量の少なさから分析から省いた。¹

Ⅲ 分析結果

(1) 収益率による四分位グループの比較

まず、月間収益率で全サンプルを4分割し、各四分位群の特徴を比較したのが表1である。2015年1月の各四分位群の月間収益率の平均値は、下位から順に-46.0%、-9.6%、0.8%、54.7%となっている。収益率が極端に高い(低い)投資家が平均値を引き上げ(引き下げ)ている可能性が高いので、中央値(メディアン)も比較すると、下位から順に-40.0%、-8.8%、0.6%、15.3%であり、第1四分位群(下位1/4)では平均値と中央値があまり変わらないのに対し、第4四分位群(上位1/4)では中央値が平均値よりもかなり低い。2015年2月の月間収益率も中央値を見る限り、1月とそれほど変わらない。

各4四分位群において保有期間の異なる3グループがどのくらいの割合で入っているかを比較したところ、第1、第2四分位群ではスキャル投資家が50%強、第3、第4四分位群では50%をわずかに下回る。反対に、デイトレ投資家は第1、第2四分位群に比べて第3、第4四分位群で比率を高めている。2015年2月ではこの傾向がより顕著に見られ、デイトレ投資家とスイング投資家の第3、第4四分位群の比率が大幅に高まり、合わせて50%を優に超えている。

レバレッジは第1四分位群と第4四分位群が高く、第2四分位群と第3四分位群が低い。これは1月と2月に共通してみられる特徴である。高い収益率を上げるためにレバレッジを高めている投資家がいる一方で、レバレッジが高いことでパフォーマンスが悪化している投資家がいることを示唆している。

勝率は収益率が高くなるにつれ、高くなっている。第4四分位群では7割を上回る勝率上げており、勝率が100%という投資家も第4四分位群の中の2割に達している。

リスクリワード比は月内に損切りと利確をとともに経験した人を対象としているため、勝率が0%や100%の投資家の割合が多い第3四分位群と第4四分位群ではサンプル数が少ないことに注意する必要があるが、第3四分位群と第4四分位群のリスクリワード比は勝率が高いにもかかわらず高い。

¹ リスクリワード比の算出において、計算を省略した通貨ペアには NZD/USD、EUR/AUD、EUR/CHF、GBP/CHF、USD/CHF、CAD/CHF、AUD/CHF、EUR/GBP、USD/HKD がある。

表 1 (A) 月間収益率での投資家分類 (2015 年 1 月)

	第 1 四分位 群(下 1/4)	第 2 四分位 群	第 3 四分位 群	第 4 四分位 群(上 1/4)
月間収益率(平均値)	-46.0%	-9.6%	0.8%	54.7%
月間収益率(中央値)	-40.0%	-8.8%	0.6%	15.3%
保有期間				
スキヤル割合	50.6%	50.4%	47.1%	45.2%
デイトレ割合	27.2%	31.1%	39.1%	34.7%
スイング割合	15.1%	20.2%	14.7%	14.6%
レバレッジ(倍)	16.00	10.03	5.40	12.08
勝率	53.6%	47.3%	63.7%	74.9%
リスクリワード比	0.68	8.51	3.27	5.79
勝率 100%割合	11.3%	14.7%	31.1%	23.4%
勝率 0%割合	15.1%	28.2%	14.7%	7.1%

(注) 保有期間 (スキヤル+デイトレ+スイング=100%)

勝率(100%、0%は各四分位に占める割合)

リスクリワード比 (1 取引当たりの平均収益額/平均損失額)

表 1 (B) 月間収益率での投資家分類 (2015 年 2 月)

	第 1 四分位 群(下 1/4)	第 2 四分位 群	第 3 四分位 群	第 4 四分位 群(上 1/4)
月間収益率(平均値)	-36.1%	-5.1%	2.0%	28.6%
月間収益率(中央値)	-30.0%	-4.3%	1.6%	15.6%
保有期間				
スキヤル割合	63.8%	65.4%	42.0%	40.2%
デイトレ割合	24.4%	28.9%	40.4%	42.3%
スイング割合	4.5%	5.3%	19.2%	12.6%
レバレッジ(倍)	16.17	9.18	6.41	13.62
勝率	52.0%	50.0%	71.5%	73.4%
リスクリワード比	1.03	0.66	5.25	4.19
勝率 100%割合	4.5%	4.1%	42.9%	24.8%
勝率 0%割合	6.9%	15.4%	12.7%	5.7%

(注) 保有期間 (スキヤル+デイトレ+スイング=100%)

勝率(100%、0%は各四分位に占める割合)

リスクリワード比 (1 取引当たりの平均収益額/平均損失額)

図 2 (A) 勝率とリスクリワード比の組み合わせの散布図 (2015 年 2 月)

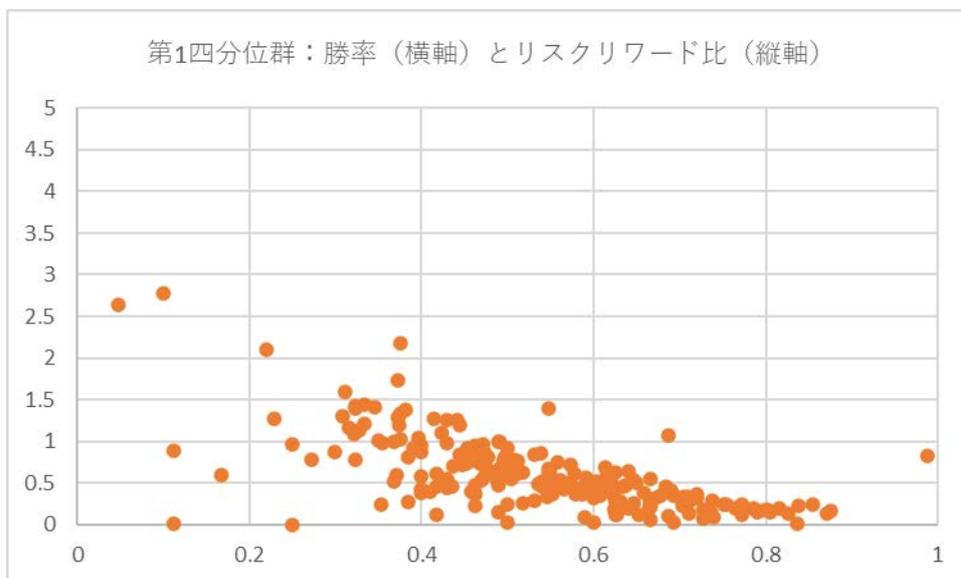
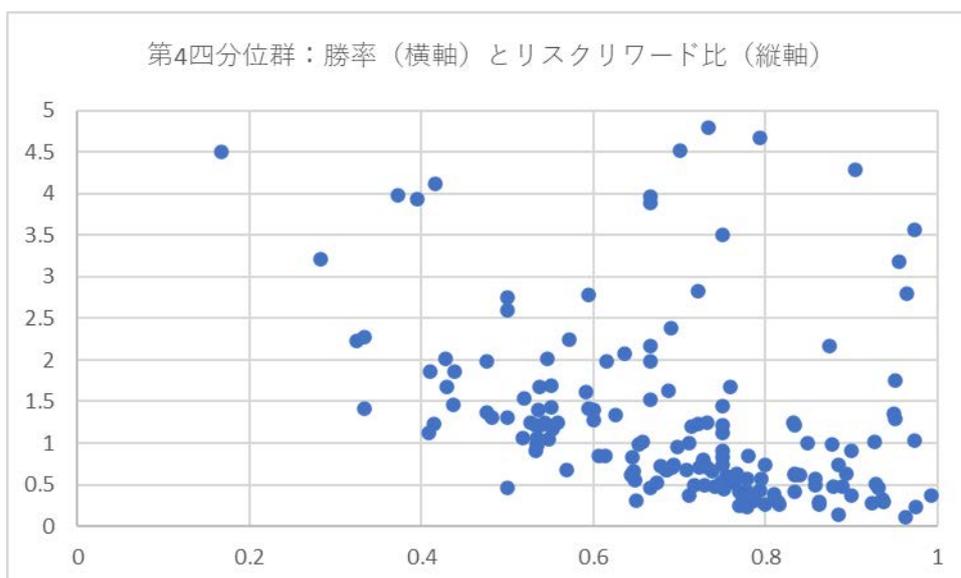


図 2 (B) 勝率とリスクリワード比の組み合わせの散布図 (2015 年 2 月)



(2) 勝率とリスクリワード比の組み合わせ

投資パフォーマンスを向上させるために損小利大を目指すと、損切り貧乏になって勝率が悪化したりすることがある。反対に、勝率を上げるようと利益確定を急ぐとリスクリワード比が悪化し、損大利小になりがちである。このように、勝率とリスクリワード比の間には負の相関があるように思われる。

図 2 では、2015 年 2 月の第 1 四分位群（下位 1/4）と第 4 四分位群（上位 1/4）のグルー

プの勝率とリスクリワード比の組み合わせの散布図を比べてみた²。リスクリワード比を算出するにあたって、勝率 100%と 0%の投資家を除いているため、サンプル数は減っているが、勝率とリスクリワード比にはゆるやかな負の相関がみられる。パフォーマンスが低い第 1 四分位群の投資家は勝率が高くても、多くの投資家のリスクリワード比が 1 を下回っており、パフォーマンスが高い第 4 四分位群とは対照的である。

(3) 保有期間別グループの比較

次に、表 2 で保有期間別グループの特徴を比較する。月間収益率は平均値、中央値ともにスイング投資家が高く、続いてデイトレ投資家、スキャル投資家の順となっている。レバレッジについては、投資家グループ間であまり違いは見られない。2015 年 1 月では、各投資家の勝率に違いはないが、2 月ではスイング投資家、デイトレ投資家、スキャル投資家の順で勝率が高い。リスクリワード比もスイング投資家が高く、スキャル投資家が低い。勝率 100%の比率も同様である。

表 2 (A) 保有期間別の投資家分類 (2015 年 1 月)

	スキャル 投資家	デイトレ 投資家	スイング 投資家
月間収益率(平均値)	-2.5%	0.8%	6.7%
月間収益率(中央値)	-3.0%	-0.4%	2.4%
レバレッジ(倍)	10.34	10.81	11.58
勝率	60.4%	60.8%	59.1%
リスクリワード比	1.60	4.73	5.64
勝率 100%割合	9.5%	26.7%	37.7%
勝率 0%割合	10.8%	16.8%	20.8%

(注) 勝率(100%、0%は各投資家グループに占める割合)

リスクリワード比 (1 取引当たりの平均収益額/平均損失額)

表 2 (B) 保有期間別の投資家分類 (2015 年 2 月)

	スキャル 投資家	デイトレ 投資家	スイング 投資家
月間収益率(平均値)	-7.0%	3.0%	4.8%
月間収益率(中央値)	-2.3%	1.0%	2.5%
レバレッジ(倍)	10.73	11.43	10.65
勝率	58.8%	62.7%	72.8%

² 2015 年 1 月のデータでも同じ結果になることを確認している。

リスクリワード比	2.60	3.34	5.61
勝率 100%割合	8.7%	25.7%	39.8%
勝率 0%割合	8.1%	11.4%	17.6%

(注) 勝率(100%、0%は各投資家グループに占める割合)
 リスクリワード比 (1 取引当たりの平均収益額/平均損失額)

(4) 階層型クラスタリングによるクラスタ分析

データ間の類似度にもとづいてデータをグループ分けするクラスタリングの手法は、階層的クラスタリングと非階層的クラスタリングに大別される。階層的クラスタリングはもっとも似ている組み合わせからまとめていくもので、距離が近いもの、または近いクラスタをつなげてより大きなクラスタを作っていく方法である。

収益率、保有期間、レバレッジ、勝率、リスクリワード比の5つの変数の単位が異なるので、平均が0、標準偏差が1となるように標準化することによってスケールを揃えた。階層的クラスタを形成する際には、ウォード法に基づいてクラスタを形成した。³

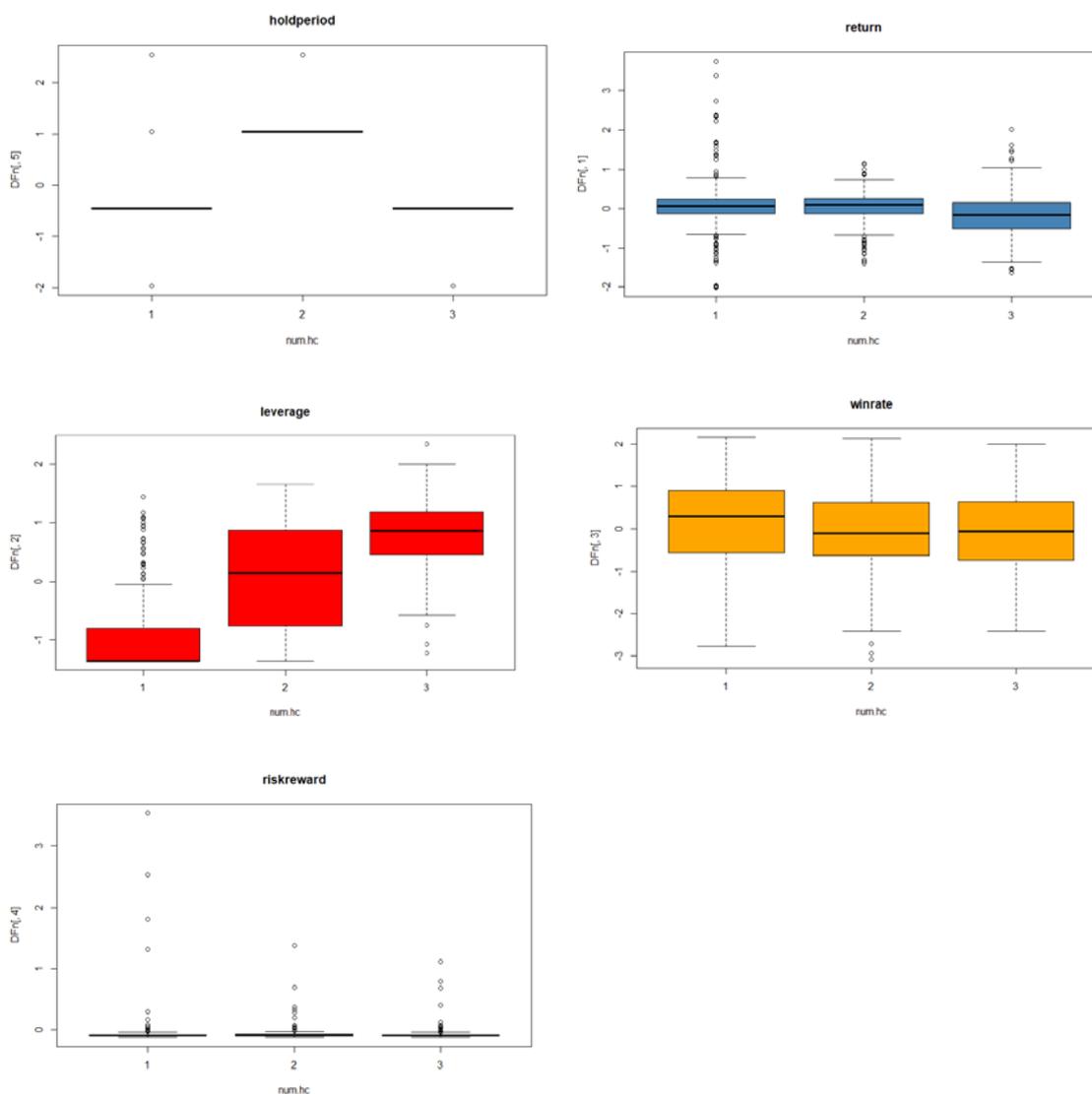
2015年2月のデータをもとに、すべてのサンプルの間の距離を計算して描いた樹形図(デンドログラム)から適切なクラスタの数は3と判明した⁴。第1クラスタには219人(31.6%)、第2クラスタには215人(31.1%)、第3クラスタには258人(37.3%)が含まれている。保有期間についてはスキャル投資家を0、デイトレ投資家を1、スイング投資家を2となるようにインデックス化したが、標準化によってスキャル投資家は0.016、デイトレ投資家は1.518、スイング投資家は3.021に変換されている。

図3に描かれたボックスプロットによると、クラスタ1と3はスキャル投資家を中心とした、2はデイトレ投資家とスイング投資家を中心としたクラスタであることが分かる。クラスタ1と3の違いを列挙すると、クラスタ1はクラスタ3に比べて収益率が高く、レバレッジが低く、勝率が高い。リスクリワード比の平均値はほぼ同じであるものの、クラスタ1にはリスクリワード比が高い投資家が含まれている。一方、デイトレ投資家とスイング投資家を中心としたクラスタ2は収益率がクラスタ1と同程度で、レバレッジは分散が大きいものの、クラスタ1と3の中間であり、勝率も中程度である。このように、スキャル投資家は特性の異なる2つのグループが混在しており、1つのグループにまとめることは適していないことがクラスタ分析によって明らかになった。

³ ウォード法とは2つのクラスタを結合したとき、それにより移動したクラスタの重心とクラスタ内の各サンプルとの距離の2乗和と、元々の2つのクラスタ内の重心とそれぞれのサンプルとの距離の2乗和の差が最小となるようなクラスタ同士を結合する手法である。

⁴ 2015年1月のデータでも適切なクラスタの数は3であり、各クラスタの特徴も同じであった。

図3 ボックスプロット（各クラスターの説明変数）



(5) ランダムフォレストによる説明変数の重要度

ランダムフォレスト (random forest) と呼ばれる手法を用い、「機械」が学習した結果を通じて説明変数の影響度合いを推定する。ランダムフォレストとは複数の決定木（データを条件分岐によって分割していくことで分類する方法）を組み合わせることで予測を行うアルゴリズムであり、特定の関数式を仮定しないため、従来の回帰モデルとは異なり説明変数の選択に制約が非常に少なく、過学習（オーバーフィッティング）の影響を軽減することが可能である。これは、ランダムフォレストが過学習を回避するため、1つのデータをリサンプリングして複数の回帰木 (regression tree) を学習するためである。この回帰木のサンプルを分割するたびに、全ての説明変数からランダムにいくつかの説明変数を選ぶことからランダムフォレストと呼ばれている（内閣府、2018）。

分析の目的は説明変数が被説明変数に対し、どの程度影響するかを探索することであるため、ランダムフォレストの予測値ではなく、変数重要度 (variable importance) を用いて影響を評価した。ランダムフォレストは従来の回帰モデルのように説明変数の係数を推定するわけではないため、説明変数がランダムで選択された際の予測誤差の大きさを計測した変数重要度が一般的に評価では用いられる。予測誤差が大きいほど変数への重要度が高いと評価できるため、変数重要度の高い説明変数ほど被説明変数への影響度が高いと考える。

表 3 では、2015 年 2 月の FX 投資家の収益率の予測モデルにおける各説明変数の変数重要度を比較している。%IncMSE とは、ある説明変数がランダムに置換されることによって MSE (平均二乗誤差) がどれだけ増加するかを示しており、IncNodePurity とは、ある説明変数がランダムに置換される前後の RSS (残差平方誤差) の差によって算出されるジニ係数 (Gini Index) から測定されるものである。いずれの指標を用いても、変数の重要度の順位は同じで、収益率の予測において勝率が最も重要で、次に、リスクリワード比、レバレッジの順で重要である。反対に、保有期間は%IncMSE と IncNodePurity のいずれにおいても大幅に低く、重要度の低さが示された。

表 3 ランダムフォレストによる変数の重要性

	%IncMSE	IncNodePurity
レバレッジ	14.87	47.21
勝率	36.58	66.01
リスクリワード比	33.12	58.22
保有期間	7.14	9.05

IV 結語

本稿では、クラスター分析とランダムフォレストを用いて、FX 投資家の分類と収益率予測における説明変数の重要度に関する分析を行った。クラスター分析からは、スキャル投資家には 2 つの特性が異なるグループが混在していることが分かり、ランダムフォレストからは、勝率、リスクリワード比、レバレッジと比べて、収益率予測において保有期間の重要性が非常に低いことが明らかになった。

勝率とリスクリワード比が投資パフォーマンスにとって重要であることは論を俟たないが、リスクリワード比を分析に採用するために勝率が 0%や 100%をとった投資家を分析から省くことになったことはやむを得ないとはいえ、分析結果に影響を与えている可能性がある。とりわけ、保有期間が長い投資家には勝率が 0%や 100%が多くなる傾向が見られ、追加的な検証が必要であろう。

参考文献

岩壺健太郎「FX 証拠金取引の投資戦略とパフォーマンス」金融先物取引業協会会報 No.114,
2017 年 10 月
内閣府 (2018) 「日本経済 2017-2018—成長力強化に向けた課題と展望—」(平成 30 年 1 月
18 日)