

ファンダメンタルファクターモデル（リターンモデル）における機械学習手法の応用可能性検証

Fundamental Factor Models using Machine Learning

杉友 盛佑*¹ 南 正太郎*²
Seisuke Sugitomo Shotaro Minami

*¹エピックパートナーズインベストメント株式会社
Epic Partners Investments Co., Ltd

*²あすかアセットマネジメント株式会社
Asuka Asset Management Co., Ltd

Fundamental factor models are one of the important methods for the quantitative active investors (Quants), so many investors and researchers use fundamental factor models in their work. But often we come up against the problem that highly effective factors do not aid in our portfolio performance. We think one of the reasons why is that the traditional method is based on multiple linear regression. Therefore in this paper, we tried to apply our machine learning methods to fundamental factor models as the return model. The results show that applying machine learning methods yield good portfolio performance and effectiveness more than the traditional methods.

1. はじめに

計量的アクティブ運用（クオンツ）の代表的なツールにマルチファクターモデルがある。これは、株式の投資収益率を複数のファクターで説明するものである。ファイナンス分野における一般的なマルチファクターモデルは、1976年にステファン・ロスにより提唱された裁定価格理論（APT: Arbitrage Pricing Theory）と同義に使用されることがあるが、運用実務においては、CAPMをベースとしたBARRA型アプローチやFama-French型アプローチのマルチファクターモデルも広く用いられている。マクロ経済変数を先験的に与えて、個別企業の株式収益率を求める方法や、過去の株式収益率から因子分析によって因子を引き出す方法などは、APT型のマルチファクターモデルに分類される。一方、PERやPBRに代表される投資指標のような個別企業の株式が持つ銘柄属性を用いて個別企業の株式収益率を求める方法は、BARRA型またはFama-French型のマルチファクターモデルに分類される。Barra型アプローチはBar Rosenbergにより紹介され、Grinold and Kahn (2000), Conner et al(2010)により拡張されたものである。ある1時点における株式銘柄群の株式収益率は共通のファクターによって説明されるとすることから、クロスセクション回帰分析によって計算される。またFama-French型は、Eugene Fama and Kenneth French (1992)によって初めて紹介された。本稿はFama-French型のマルチファクターモデルに分類されるものである。

もう一つ、マルチファクターモデルを用いるに当たって明確にしておかなければならない事項がある。

個別企業の株式収益率を銘柄属性で説明するマルチファクターモデルの利用方法として通常2種類考えられる。1つは、銘柄属性の寄与度から相場の傾向を判断して、将来の株式収益率を計算するために利用する方法（リターンモデル）。もう一つは、銘柄属性の寄与度をマーケット・リターンで時系列回帰してファンダメンタル・ベータを求め、ポートフォリオの属性を分解する方法（リスクモデル）である。本稿においては、リ

ターンモデルを想定している。

以上が分析するうえで押さえておかなければならない前提である。これらは、ファイナンス分野でも混同して議論されるケースがある。本稿の目的は、計量的アクティブ運用における機械学習手法の応用可能性を探ることであり、今後の計量アクティブ運用の分野への発展に寄与することであることから前提を明確にしておく。

本稿におけるマルチファクターモデル（以下ファンダメンタルファクターモデルと呼ぶ）を以下で定義する。

$$R_{it} = \sum_{j=1}^k X_{ijt} f_{jt} + \epsilon_{it} \quad (1)$$

ただし、 R_{it} はt期における企業iの株式収益率、 X_{ijt} はt期における企業iの第jファクターのファクター・エクスポージャー、 f_{jt} はt期における第jファクターのファクター・リターン、 ϵ_{it} はファクターでは説明できない誤差

個別企業の将来の株式収益率を複数の銘柄属性（ファクター）による重回帰分析により算出するモデルである。このモデルの下では株式収益率とファクターの間における関係は線形であるが、金融市場の複雑性を考えると非線形を仮定するほうがより適切に関係性を表現できると考えられる。そこで、本稿では、非線形な関係を表現できる機械学習の代表的な手法（サポートベクターマシン、勾配ブースティング、ニューラルネットワーク）を用いて、従来の線形モデルと比較し、非線形手法の実務運用における有効性と応用可能性を検証する。

2. 関連研究と基礎的概念

2.1 関連研究

[Joseph 14]は、PBRのファクター・リターンを予測するために、VIX、VIXの1カ月変化率、PBRの分布、PBRの分布の変化率、1カ月前のPBRのファクター・リターンを変数とし、パラメータの推定に縮小推定法（Shrinkage Method）を用いたロジスティック回帰分析を用いた検証を行っている。また、プライスモメンタムについても同様の分析を行っている。

連絡先: 杉友 盛佑 エピックパートナーズインベストメント株式会社 sugitomo@epicgroup.jp

ともに、翌期の予測精度が大幅に高くなっている結果を得ている。他の機械学習手法として、CART (Classification and Regression tree) を使った検証もしており、時期によってはロジスティック回帰分析よりも有効性が見られることを示した。

上記の分析は、ファクター・リターンの時系列予測に他の変数を用いて非線形の機械学習手法の適用を試みたものであるが、リターン・モデルについて適用したものではない。リターン・モデルでは、ファクター・リターンの予測値 (ファクター・ウェイトと呼ぶ) と直近のファクター・エクスポージャーから、将来の予測リターンが計算され、これを予測アルファと呼ぶが、上記の研究は予測モデルのマルチファクターに言及したものではない点で我々の研究とは異なる。

以降では、分析に用いた機械学習分析手法の基礎的概念について整理する。

2.2 サポートベクター回帰

サポートベクター回帰は非線形回帰の一手法であり、説明変数の特徴空間への非線形写像を考え、特徴空間において線形回帰を行うという手法である。本研究で用いる ϵ -SVR は、線形回帰の時、 ϵ -insensitive 損失関数 $L^\epsilon(x, y, f)$ を使って線形関数 $y = f(x) = \langle w \cdot x \rangle + b$ を推定する。

$$L^\epsilon(x, y, f) = \max(0, |y - f(x)| - \epsilon) \quad (2)$$

なお予測値が実測値を上回る場合は $L^\epsilon(x, y, f) = \xi$ とし、予測値が実測値を下回っている場合は $L^\epsilon(x, y, f) = \hat{\xi}$ で表現する。最終的には以下の主問題を解くことになる。

$$\min_{w, L^\epsilon, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l L^\epsilon(x_i, y_i, f), \quad (3)$$

$$s.t. (\langle w \cdot x_i \rangle + b) - y_i \leq \epsilon + \xi_i \quad (4)$$

$$y_i - (\langle w \cdot x_i \rangle + b) \leq \epsilon + \hat{\xi}_i \quad (5)$$

$$\xi_i, \hat{\xi}_i \geq 0, i = 1, 2, \dots, l \quad (6)$$

ここで、上記の式における元データ x を非線形関数 $\phi(x)$ によって高次元空間に写像することにより、非線形回帰へと拡張できる。上記の主問題は実際には双対定理を使い双対問題として解くことができる。この時双対問題中に $\langle \phi(x_i) \cdot \phi(x_j) \rangle$ と内積計算が現れる。一般的に高次元空間で内積計算を直接行うことはとても煩雑であり、この内積計算を行うことができるカーネル関数 $K(x_i, x_j)$ を適用する。本論文では、カーネル関数として、ガウシアンカーネル $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2})$ を用いる。

2.3 決定木、勾配ブースティング

取り得る選択肢や起こり得るシナリオを樹形図の形で洗い出し、それぞれの選択肢の期待値を比較した上で意思決定を行う手法が決定木であり、金融業界やコンサルティング業界においてよく用いられる手法である。本研究では、より精緻なモデルを構築するため、アンサンブル学習の一つである勾配ブースティングという手法と決定木を組み合わせる。勾配ブースティングとは、学習データから復元抽出を繰り返し複数のデータセットを作成し、それぞれに対して弱学習器を作り、全ての弱学習器の解で多数決を取り最終解を求める方法である。弱学習器を作成する際に、前回作成した弱学習器の結果を利用して、誤分類された値の重みを大きくするように更新す

る。この重みづけの際に勾配降下法を用いるのが、勾配ブースティングである。本研究では、この弱学習器に決定木を用いた、GBDT(Gradient Boosting Decision Tree) をモデルとして用いる。

2.4 ニューラルネットワーク

ニューラルネットワークには、階層型ニューラルネットワークと非階層型と呼ばれるものがあるが、今回用いる手法は、階層型ニューラルネットワークである。階層型ニューラルネットワークとは、入力層、中間層、出力層をもつネットワークである。入力層に説明変数を取り、中間層でそれらの説明変数に対してランダムに重みづけを与え、その結果が出力層である目的変数に近づくように、最適な重みを計算していく。あたかも人間の脳の中の神経細胞であるニューロンが、多数の他の神経細胞から信号を受け取り、意思決定を行っていることと類似していることからニューラルネットワークと呼ばれている。

2.5 実装

筆者らが構築した株式分析システムに基づいている。データは FactSet より取得、基本となるデータベース・計算システムおよびファンダメンタルファクターモデルについては Python、機械学習の計算については R と機械学習パッケージである 'nnet' を用いた。

3. 検証

3.1 検証における考え方

通常、ファンダメンタル・ファクターモデルにて算出した将来の株式収益率が実運用で活用できるかどうかについては、ファクター・リターンの有効性と安定性を検証しなければならない。しかしながら、機械学習手法 (サポートベクターマシン、勾配ブースティング、ニューラルネットワーク) についてはファクター・リターンに該当する係数を特定するのは困難である。そのため次のような検証を行う。まず、機械学習手法の応用の結果求められる将来の株式収益率を個別銘柄対比で大きい順にソートし 5 分位のグループに分ける。最も大きいグループをロングポートフォリオ、最も小さいグループをショートポートフォリオとし、それぞれ翌期のポートフォリオリターンを計測する。その後、ロングポートフォリオとショートポートフォリオの差分 (スプレッド・リターン) を計算。分析期間にわたって繰り返し計算を行い、従来の線形モデルと機械学習手法の結果を比較する。もし、将来の収益率の予測力が高いほど、実現した将来の株式収益率と予測した収益率の差は小さくなり、ロング・ショートポートフォリオの累積リターンは大きくなるはずである。なお、実証するにあたり、実現値と予測値を用いて RMSE および MAE も併せて計算している。

3.2 検証手順

ユニバースは、TOPIX500 構成銘柄とした。使用するファクターについては一般的によく使われる投資指標である PER、PBR、ROE、対数時価総額、3ヵ月 β の 5 つとする。ただし、PER と PBR については逆数に変換している。分析期間は 2000 年 1 月末から 2017 年 6 月までである。ポートフォリオについては毎月リバランスによりリターンを計測している。なお、個別銘柄の所属する業種によるファクター値の水準の違いによ

る影響を除くため、まず東証 33 業種内で基準化 (z-score) し、その後、再度 TOPIX500 全体で基準化した。

比較基準となる重回帰モデルについては、過去 1 年間 (t 期よりも前) の各銘柄の月末ファクター値を説明変数、翌月末 (t 期) リターンを目的変数として回帰係数 (ファクター・リターン) を算出。得られた回帰係数 (ファクター・リターン) をテスト期間における月末の各銘柄のファクター値 (ファクター・ウェイト) にかけることで、各銘柄の将来の株式収益率 (期待リターン) を算出している。先の検証についての考え方に従って、得られた期待リターンをもとに作った 5 分位ポートフォリオの 1 分位をロング、5 分位をショートし、翌月まで保有した時のロングポートフォリオとショートポートフォリオのリターン差 (スプレッド・リターン) を求めている。これを月次でローリングし得られたリターンを累積した。比較対象である機械学習手法 (サポートベクターマシン、勾配ブースティング、ニューラルネットワーク) についても分析のプロセスは同様である。期待リターンを計算する際に、機械学習手法を用いている。

有効性を検証するために、ポートフォリオのパフォーマンスの比較および予測精度の 2 つのパターンを示す。なお、ポートフォリオパフォーマンスについては、月次平均リターン、月次標準偏差、およびシャープレシオ。予測精度として、計算期間に渡り累積した RMSE と MAE である。

3.3 検証結果

	重回帰分析	GBDT	SVM	NN
平均リターン	0.10	0.09	0.12	0.18
標準偏差	2.83	2.86	2.89	2.72
シャープレシオ	0.037	0.034	0.042	0.068
累積 RMSE	1832.17	1831.61	1823.53	1832.92
累積 MAE	1428.71	1428.14	1418.28	1428.39

表 1: 実験結果

ポートフォリオのパフォーマンスの比較では、月次平均リターンはニューラルネットワークによるモデルが最も高い結果となった。ボラティリティを考慮したシャープレシオについても同様の結果である。予測精度の比較では、実績値と予測値から求めた RMSE と MAE の累積値はともに SVM が最も良い結果となった。GBDT についてはポートフォリオのパフォーマンスについては重回帰分析に劣るものの、累積 RMSE、累積 MAE については精度が高い結果となった。

4. 結論

本研究では、ファンダメンタルファクターモデルに着目しリターン予測の精度について、従来の重回帰分析に加え、GBDT、SVM、ニューラルネットワークを適用し比較した。結果として、当てはまりの精度である累積 MAE は全ての非線形モデルで改善が見られ、累積 RMSE、月次平均リターンや月次シャープレシオにおいても、多くのモデルで改善が見られた。このことは、金融市場における銘柄のリターンとファクター値間の関係は従来の線形関係ではなく、非線形関係が存在すると考えられ、そのような非線形関係を捉えられるモデルが、従来のモデルに対し優位性を持つと考えられる。

本研究の今後の展望として、非線形の分析は実運用においても重要な意味がある。例えば、ファクターにベットしている

大規模ファンドやスマートベータと呼ばれるファンドは、ティルトしているファクターの有効性は見られるもののファンドパフォーマンスがそぐわないことが良くある。これは、従来のマルチファクターモデルは線形モデルであるために、本来非線形で評価すべきものを線形評価することによるズレが起因している可能性がある。また、クロスセクション回帰分析への応用 (BARRA 型) も期待されるだろう。もしくはクロスセクションと時系列回帰分析の両方を考慮した多変量回帰分析への応用も興味深い結果が得られるかもしれない。計量アクティブ運用の分野はそれだけでも奥深い研究対象であることから、基本的な分析手法や考え方を押さえた上で、機械学習手法を応用するのは簡単ではない。しかしながら、実運用において広く使われる考え方でもあることに加え、計量アクティブ運用と機械学習手法の相性はよい側面もあることを考えれば、今後あらゆる角度からの応用が求められる。課題にしたい。

5. 留意事項

本論文は筆者らが所属するエピックパートナーズインベストメンツ株式会社及びあすかアセットマネジメント株式会社の公式見解を表すものではありません。すべては個人的見解であります。

参考文献

- [南 17] 南正太郎, 光定洋介: AI の資産運用への応用の可能性と限界, 証券アナリストジャーナル, Vol.55, 8月号 (2017).
- [吉野 03] 吉野貴晶: 株式投資のための定量分析入門, 日本経済新聞社, (2013).
- [T.DANIEL 95] T.DANIEL COGGIN AND FRANK J. FABOZZI: The Handbook of Equity Style Management, Addison-Wesley, 2nd edition (1998) (邦訳: 資産運用時代新時代の株式投資スタイル, 野村総合研究所 訳, 野村総合研究所発行, (1996)).
- [Eric 11] Eric Zivot: Factor Models for Asset Returns, University of Washington (2011).
- [Argimiro 14] Argimiro Arratia: Computational Finance An Intodutory Course with R, ATLANTIS PRESS (2014).
- [Joseph 14] Joseph Mezrich: Factor forecasting with machine learning, Nomura Equity Research Report (2014).
- [Ludwig 06] Ludwig B Chincarini, Daehwan Kim: Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management, McGraw-Hill Library of Investment and Finance (2006).
- [Richard 99] Richard C. Grinold, Ronald N. Kahn: Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk, McGraw-Hill Library of Investment and Finance (1999).