グローバル株式市場における深層学習を用いた マルチファクター運用の実証分析

Deep Learning for Multifactor Models in Global Stock Markets

阿部 真也 中川 慧 Masaya Abe Kei Nakagawa

野村アセットマネジメント株式会社 Nomura Asset Management Co.,Ltd.

Many studies have been undertaken by using deep learning techniques to predict stock returns in terms of timeseries prediction. However, from the viewpoint of the cross-sectional prediction, there are no examples that verify its effectiveness in the global stock market. This paper implements deep learning to predict stock returns in the cross-section in the global stock market and investigates the performance of the method. Our results are followings.

1. Deep learning is superior in terms of return / risk as compared with random forest and ridge regression. 2. Especially, in terms of risk, the deep learning model is outstanding.

3. If market efficiency declines, opportunities for return may increase.

1. はじめに

株価を予測するためのアプローチとして、(1)株価を時系列 データとみなして予測する方法、(2)あるクロスセクション(断 面)での銘柄間の相対的な順位(魅力度)に基づいて予測する 方法の2つがある。一般的な時系列データとは異なり、株価を 時系列データとしての予測、すなわち過去の株価を特徴量とし て利用し、将来の株価を予測するという前者のアプローチは短 期的には可能でも、長期的には難しいことが知られている。

一方で、後者の複数の株価をクロスセクションデータとし て見た中で、どういった銘柄群が相対的にリターンが高くな り、どういう銘柄群が低くなるかは、中長期的な期間である程 度明らかにされている。こういったクロスセクションで株価を 説明する要因を「ファクター」という。実際にはこれら複数の ファクターを組み合わせ、株価の相対的な魅力度をより詳細に、 多角的に明らかにするマルチファクターモデルが用いられる。 ファイナンスの学術的には、[Fama 92] による Fama-French の3ファクターモデルが一般的である。以降、[Harvey 16] に よれば、2012年までの間に316ものファクターが「発見」さ れた。しかも、これらファクターの大部分は過去10年間で検 証された。そのため近年、運用者が検討すべきファクターは 急速に増大しているものの、次元の呪いの問題から 316 もの ファクターを同時に検証することはできない。一方で、ファイ ナンスにおいては統計的な扱いや理論的な解析が容易である こと、結果の頑健性から線形回帰モデルを用いる。実務的に は、これらのうち幾つかのファクターと将来リターンの相対 的順位との回帰分析から、ある銘柄についての総合的な魅力 度を計算し、当該魅力度に基づいてポートフォリオ構築を行 う。しかしながら、これらファクターと将来リターンの関係は 複雑であることが古くから知られているため、単純な線形回 帰モデルでは予測精度に限界がある。金融市場の非線形な関 係をとらえるため、単純な線形回帰モデルではなく、機械学

 連絡先:連絡先:野村アセットマネジメント株式会社 〒103-0027 東京都中央区日本橋1丁目12-1 E-mail: m-abe@nomura-am.co.jp
本稿の内容は筆者らが所属する組織を代表するものではなく、本稿の全ての誤りは、筆者らの責に属するものである。 習の手法を用いて、時系列予測の観点から、将来リターンを 予測する問題に取り組んでいる研究が学術的には盛んである [Bahrammirzaee 10, Cavalcante 16, Nakagawa 18a]。

しかしながら、クロスセクションでのマルチファクターモ デルの観点から、機械学習を適用する試みは [Sugitomo 18, Abe 18, Nakagawa 18b] といった少数の例を除くとあまりな く、特に様々なファクターを用い、グローバルな株式市場にお いてその有効性を検証した例は筆者らの知る限り存在しない。

そこで本稿では、グローバルな株式市場において画像認識、 音声認識、自然言語処理など様々なタスクにおいて、優れた非 線形関数近似能力を示している深層学習を用いたマルチファク ターモデルに基づく相対的な魅力度の有効性を検証する。

2. 深層学習を用いたマルチファクターモデル

銘柄の相対的な魅力度を判断するマルチファクターモデルを 構築するためにはまず、データとして PER や ROE などの株 価を説明するであろう指標 (ディスクリプター)を幾つか用意 する。[Fama 92] などのファイナンスの学術研究とは違い、そ れらのディスクリプターを標準化したものを複数組み合わせて ファクターを作る。例えば、株価の割安感を表すバリューファ クターを、PBR、PER、配当利回りを 1:1:1 で組み合わせて作 る。そして複数のファクターから回帰モデルで相対的な魅力度 を予測する。最後に予測された魅力度を期待リターンとして、 銘柄間のリスクと組み合わせてポートフォリオを構築する。

従来はディスクリプターからファクターの作成、そしてどの ファクターをいくつ使うのかを人が選択していたため、モデル の見直し・チューニングには非常に多くの時間がかかった。ま たそれらの多くは運用者であるファンドマネージャーの経験に よって決定されていた。しかしながら、ディスクリプターから 株式リターンの予測を表現力に優れた深層学習を用いて置き換 えることで、長年の経験をパラメータチューニングに置き換え ることができる。

具体的には、ディスクリプターをインプットとして与え、ア ウトプットとして一定期間後の将来リターンの順序関係を与え る。すると隠れ層にはこれらディスクリプターが非線形に合成 されたファクターとファクターリターンが表現される (図 1)。



図 1: 深層学習を用いたマルチファクターモデル

3. 実証分析

3.1 データセット

ここではグローバルな株式市場を対象に分析を行う。分析 のユニバースとして、機関投資家の投資対象として使用され ることの多い MSCI 社の株式指数である、MSCI Americas、 MSCI Europe & Middle East、MSCI Pacific の3地域を使 用する。

特徴量として用いるディスクリプターは表 1 の通りであ る。これらのディスクリプターの計算のため、WorldScope、 Thomson Reuters、I/B/E/S および EXSHARE からデータ を取得した。実績財務データは、WorldScope と Reuters Fundamentals(WorldScope 優先)から、予想利益 (No.15-16) は Thomson Reuters Estimates、I/B/E/S Estimates(Thomson Reuters 優先) から取得している。これらのディスクリプター はすべて実務でよく使用されるものである。

3.2 分析方法

特徴量として、表1のディスクリプターを用い、教師データ には翌月の現地通貨建てのリターンとする。これらは月次系列 データである。また各データは時点毎に0から1のパーセン タイル化の前処理を行っている。訓練期間として直近120か 月(ただし、直近1か月を除く)のデータを使用し、モデルの 学習を行う。モデル学習後は、直近時点の特徴量をモデルに入 力し、各銘柄の予測スコア(相対魅力度)を算出する。この処 理を3地域別に毎月行っていく。具体的なモデルとして深層学 習とベンチマークとしてランダムフォレスト、リッジ回帰を用 いる。

1. 深層学習 (DNN): パッケージ TensorFlow を用いて実装 を行った。7層の全結合型の構造で、活性化関数は全て ReLU 関数を用いた。ネットワークにおける隠れ層の構 造 (ユニット数) は {150-150-100-100-50-50} で あり、各層に対応するドロップアウトは (50% - 50% -30%-30%-10%-10%)である。最適化アルゴリズムに 学習率 0.001 の Adam を用い、 ミニバッチサイズは 300 とした。初期値 (分析開始月) については、バイアスを 0、 ウェイトを TensorFlow の関数"tf.truncated_normal"の 平均 0、標準偏差を $\sqrt{2/M}$ として与えた。ただし、M は 前層のユニットのサイズである。また各層でバッチ正規 化を行っている。学習度合いについては、各時点の銘柄 数が異なり、訓練データサイズが一定ではないため、エ ポックス数ではなく、訓練データの特徴量から計算した 予測スコアと教師データの順位相関係数を用いた。具体 的には、訓練データ内の時点(年月)別に順位相関係数を

計算し、その平均値が 0.20 を超えたらネットワークの学 習を終了させている。また fine tuning として、分析開 始月の後の学習については、前月の訓練過程において順 位相関係数 0.18 を超えた時点のネットワークを呼び出し て、学習を開始させている。

- ランダムフォレスト (RF):パッケージ scikit-learn のク ラス"sklearn.ensemble.RandomForestRegressor"を用 いて実装を行った。パラメータについては、特徴量の数 (max_features)を20、木の数 (n_estimators)を1,000、 深さ (max_depth)を5とした。
- リッジ回帰 (RR):パッケージ scikit-learn のクラ ス"sklearn.linear_model.Ridge"を用いて実装を行った。 正則化パラメータ (alpha) は、1(デフォルト値) とした。

パフォーマンス評価のために構築するポートフォリオは、予 測スコアで5分位し、最上位の第5分位に属する銘柄群を等 ウェイトで買った「5分位ポートフォリオ」と、5分位ポート フォリオと最下位の第1分位に属する銘柄群を等ウェイトで 売った1分位ポートフォリオを合成した「5分位-1分位ポート フォリオ」の2パターンとする。5分位ポートフォリオは投資 対象銘柄の等ウェイトリターンを引いて、5分位-1分位ポート フォリオは、5分位ポートフォリオから1分位ポートフォリオ のリターンを引いて、ポートフォリオのリターンを計測する。 なお、検証期間は2005年1月から2017年12月までとし、毎 月予測スコアを計算し、ポートフォリオ内の構成銘柄を入れ替 えていく。

3.3 分析結果

5分位ポートフォリオ、5分位-1分位ポートフォリオの結果 のサマリーがそれぞれ表 2、3である。年率リターン、リスク、 リターン/リスク比 (R/R)、最大ドローダウン (最大 DD)*¹を 評価指標とする。なお、表中の太字は DNN、RF、RR それぞ れの手法内で最も良いものを表している。

表 2、3 ともに、DNN は R/R で見て全ての地域で他の手 法を大きく上回っている。また年率のリスクと最大ドローダ ウンに着目すると、DNN は他の手法と比べて小さく、リスク を回避した運用を行う場合に特に DNN に優位性がある。年 率リターンについては、Europe & Middle East を除いて、両 ポートフォリオ戦略において DNN が最も高く、低リスクであ りながらも、相対的に高いリターンを獲得できている。対し て、RFと RR は地域ごとに優劣のばらつきがあり、どちらが 優れているとは言い難い。なお、RR は最大ドローダウンが大 きくなりがちであることが確認できる。地域別に結果を確認す ると、使用するモデルに関わらず Pacific、Europe & Middle East、Americas の順でパフォーマンスが悪化していることが わかる。この結果は、一般にアジア、欧州、米国の順に株式市 場の流動性が高くなり、市場の効率性が向上することで、収益 獲得が困難となることと整合的であると考えられる。

4. おわりに

本稿では、深層学習を用いたマルチファクターモデルに基づ く相対魅力度の有効性の検証を、グローバル株式市場で行った。 各地域の MSCI Americas、MSCI Europe & Middle East、 MSCI Pacific を対象とした実証分析の結果、以下の知見が得 られた。

^{*1} リターンの最大収益時からの下落率をドローダウンという。その ドローダウンが最大のケースを最大ドローダウンといい、リスク指 標として用いられる。

| No | ディスクリプター | 項目説明 |
|----|------------------------|------------------------------------|
| 1 | B/P | 総資産/時価総額 |
| 2 | E/P | 当期純利益/時価総額 |
| 3 | D/P | 配当/時価総額 |
| 4 | S/P | 売上高/時価総額 |
| 5 | CF/P | 営業キャッシュフロー/時価総額 |
| 6 | ROE | 純利益/純資産 |
| 7 | ROA | 営業利益/総資産 |
| 8 | ROIC | 税引後営業利益/(有利子負債+株主資本) |
| 9 | アクルーアルズ | -(流動資産の変化+現金の変化-流動負債の変化-減価償却費)/総資産 |
| 10 | 総資産成長率 | 総資産の前期からの成長率 |
| 11 | 流動比率 | 流動資産/流動負債 |
| 12 | 自己資本比率 | 純資産/総資産 |
| 13 | 総資産回転率 | 売上高/総資産 |
| 14 | CAPEX 成長率 | 資本的支出の前期からの成長率 |
| 15 | EPS リビジョン (対 1 か月) | 利益予想変化 (対 1 か月) ※ローゼンバーグ方式 |
| 16 | EPS リビジョン (対 3 か月) | 利益予想変化 (対 3 か月) ※ローゼンバーグ方式 |
| 17 | 1 か月モメンタム | 過去1か月リターン |
| 18 | 12 か月モメンタム (直近 1 か月除く) | 過去 12 か月リターン (直近 1 か月除く) |
| 19 | ボラティリティ | 過去 60 カ月リターンの標本標準偏差 |
| 20 | 歪度 | 過去 60 カ月リターンの標本歪度 |

| 表 1 | 1: 夜 | も用し | った- | ディ | スク | IJ | プ | ター | の定義 |
|-----|------|-----|-----|----|----|----|---|----|-----|
|-----|------|-----|-----|----|----|----|---|----|-----|

表 2:5 分位ポートフォリオの結果 MSCI Americas

| Q5 | DNN | RF | RR |
|--------|---------|---------|---------|
| 年率リターン | 1.76% | 0.14% | 0.34% |
| 年率リスク | 3.51% | 5.05% | 5.88% |
| 年率 R/R | 0.50 | 0.03 | 0.06 |
| 最大 DD | -11.21% | -25.26% | -27.19% |

MSCI Europe & Middle East

| Q5 | DNN | RF | RR |
|--------|--------|---------|---------|
| 年率リターン | 2.55% | 1.73% | 2.81% |
| 年率リスク | 3.29% | 4.73% | 5.23% |
| 年率 R/R | 0.78 | 0.36 | 0.54 |
| 最大 DD | -6.47% | -15.94% | -16.61% |

MSCI Pacific

| Q5 | DNN | RF | RR |
|--------|--------|---------|---------|
| 年率リターン | 4.59% | 3.15% | 3.08% |
| 年率リスク | 3.59% | 4.90% | 4.81% |
| 年率 R/R | 1.28 | 0.64 | 0.64 |
| 最大 DD | -6.93% | -10.03% | -10.36% |

表 3: 5 分位-1 分位ポートフォリオの結果 MSCI Americas

| Q5-Q1 | DNN | RF | RR | | |
|--------|---------|---------|---------|--|--|
| 年率リターン | 4.16% | 2.23% | 2.04% | | |
| 年率リスク | 7.41% | 11.28% | 11.58% | | |
| 年率 R/R | 0.56 | 0.20 | 0.18 | | |
| 最大 DD | -28.40% | -38.68% | -42.40% | | |

MSCI Europe & Middle East

| Q5-Q1 | DNN | RF | RR |
|--------|---------|---------|---------|
| 年率リターン | 5.01% | 5.53% | 6.90% |
| 年率リスク | 7.58% | 12.90% | 12.14% |
| 年率 R/R | 0.66 | 0.43 | 0.57 |
| 最大 DD | -17.92% | -37.72% | -34.71% |

MSCI Pacific

| Q5-Q1 | DNN | RF | RR | | |
|--------|---------|---------|---------|--|--|
| 年率リターン | 9.48% | 8.26% | 7.18% | | |
| 年率リスク | 6.49% | 9.11% | 9.04% | | |
| 年率 R/R | 1.46 | 0.91 | 0.79 | | |
| 最大 DD | -13.11% | -18.73% | -19.01% | | |

- どの地域でも、深層学習による株価予測モデルはランダ ムフォレストやリッジ回帰に比べ、リターン/リスク比の 面で大きく優れている。
- 低リスクという観点で深層学習モデルは優れており、リ ターンについても MSCI Europe & Middle East を除き、 相対的に高い水準を獲得している。
- 手法に関わらず、市場の効率性が向上すると、収益獲得の機会が低下する可能性がある。

参考文献

- [Abe 18] Abe, M. and Nakayama, H.: Deep Learning for Forecasting Stock Returns in the Cross-Section, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 273–284, Springer (2018)
- [Bahrammirzaee 10] Bahrammirzaee, A.: A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems, *Neural Computing and Applications*, Vol. 19, No. 8, pp. 1165–1195 (2010)
- [Cavalcante 16] Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., and Oliveira, A. L.: Computational intelligence and financial markets: A survey and future directions, *Expert Systems with Applications*, Vol. 55, pp. 194–211 (2016)
- [Fama 92] Fama, E. F. and French, K. R.: The crosssection of expected stock returns, the Journal of Finance, Vol. 47, No. 2, pp. 427–465 (1992)
- [Harvey 16] Harvey, C. R., Liu, Y., and Zhu, H.: … and the cross-section of expected returns, *The Review of Fi*nancial Studies, Vol. 29, No. 1, pp. 5–68 (2016)
- [Nakagawa 18a] Nakagawa, K., Imamura, M., and Yoshida, K.: Stock Price Prediction with Fluctuation Patterns Using Indexing Dynamic Time Warping and k*-Nearest Neighbors, New Frontiers in Artificial Intelligence, pp. 97–111 (2018)
- [Nakagawa 18b] Nakagawa, K., Uchida, T., and Aoshima, T.: Deep Factor Model, in *MIDAS 2018 MIDAS @ECML-PKDD 2018 - 3rd Workshop on MIning DAta for financial applicationS* (2018)
- [Sugitomo 18] Sugitomo, S. and Minami, S.: Fundamental Factor Models Using Machine Learning, Journal of Mathematical Finance, Vol. 8, No. 01, p. 111 (2018)