

令和 5 年 6 月 5 日現在

機関番号：22604

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K01691

研究課題名（和文）データマイニングの危険性を考慮した資産価格のファクター・プレミアムに関する研究

研究課題名（英文）Study on factor premia of asset prices in consideration of the dangers of data-mining

研究代表者

内山 朋規（UCHIYAMA, Tomonori）

東京都立大学・経営学研究科・教授

研究者番号：50772125

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：資産のリターンが予測可能であることはファイナンス理論における標準的な見方であるが、統計的にこれを検出することは容易でない。より有意な実証的証拠を得ようとデータやモデルのマイニングを行うと、実際には無意味であるにもかかわらず有意に見えてしまうというオーバーフィッティングを引き起こす。本研究は、大規模データを用いた実証分析により、時系列とクロスセクションの双方のリターンの予測可能性に関して、機械学習を利用することの有効性と危険性を明らかにした。

研究成果の学術的意義や社会的意義

ファクター・プレミアムに関する研究はファイナンス研究における中心的テーマであり続けているとともに、直接応用可能なことから実務においても注目されてきた。本研究の新規性は、近年発展が著しい機械学習の技術がどのように有効なのかを批判的思考のもとで探求した点である。本研究の結果は、ファイナンス研究が人々の意思決定の改善に役立つことを通じて社会に貢献していくためには、データサイエンスの技術を活用することが有効であるとともに、データマイニングの危険性を考慮して活用するのが重要であることを示している。

研究成果の概要（英文）：It is a standard view in finance theory that the returns on assets are predictable, but detecting this statistically is not easy. Mining data and models to obtain more significant empirical evidence leads to overfitting, in which the results appear significant even though they are meaningless. This study uses large data sets to reveal the effectiveness and risks of using machine learning in empirical analysis to return predictability in the time series and the cross-section.

研究分野：ファイナンス

キーワード：資産価格 ファクター リスクプレミアム オーバーフィッティング 機械学習 データマイニング
予測可能性

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

資産価格は将来キャッシュフローの期待割引価値として定まり、その割引率(言い換えれば、資本コスト)はリスクフリーレートにプレミアムが加わったものである。また、割引率は資産に投資した際の期待リターンでもある。このことから、プレミアムの特徴を解明することは、資産価格理論における中心的テーマであり続けてきた。プレミアムはクロスセクションで異なり、かつ時系列に変動しているため、何らかの予測変数(ファクターや特性と呼ばれる)によって、クロスセクションと時系列の双方でリターンが予測可能であることは、現在の学術界における標準的な考え方である。株式市場を例にとると、純資産株価倍率や配当利回りはそれぞれクロスセクションと時系列におけるよく知られた予測変数である。

しかし、プレミアムの実証的な特徴を満足に記述できるモデルはいまだ存在しない。このため、多くの研究者がさまざまな資産(株式、債券、クレジット、デリバティブ、通貨、コモディティ、不動産)を対象に、クロスセクションあるいは時系列における予測変数(ファクター)の研究を競い合い、新たな予測変数を「発見した」とする報告が続いている(筆者も含まれる)。例えば Harvey, Liu and Zhu (2016, Rev Finance Stud)によれば、米国株式市場におけるクロスセクションの予測変数として、これまでトップジャーナルに「発見」が報告されたファクターは316個もあり、「発見」される頻度は年々高まっている。他の資産クラスや時系列の予測変数に関しても同様で、米国ファイナンス学会会長講演(Cochrane, 2011, J Finance)では、学術界のこうした状況が“zoo of new factors”と称されている。

さらに近年では、機会学習やビッグデータに注目が集まっている。データスヌーピングとも呼ばれ、ネガティブな意味の用語として使われてきたデータマイニングが、ファイナンスの分野でもポジティブな意味として使われる場面が見受けられる。予測対象のリターンデータは時間の経過分しか増えない一方で、情報入手コストの低減に伴い予測変数の候補として使用可能なデータは飛躍的に増大している。数値データのみならず、今日ではSNS投稿などのテキスト分析や、画像解析といった非構造化データも利用可能である。近年は工学的な側面から機械学習への注目度が増している。計算処理能力の向上も相まって、データマイニングはこれからますます容易になるだろう。しかし、データマイニングはオーバーフィッティング(過剰適合)という忌めしき事態を招く。

2. 研究の目的

ファクター・プレミアムの予測可能性に関して、合理的説明と行動論的説明の双方からさまざまな理論的根拠が示されているものの、統計的にこれを検出することは容易ではない。そのため新たな実証的証拠を発見するには、データを念入りに調べることになる。一方で、念入りに調べるほど、本来は意味がないにもかかわらず、興味深いパターンが偶然に出現してしまうというオーバーフィッティングの可能性が高まる。偽発見であればこうした成果に価値はなく、むしろ成果が応用されることによって弊害をもたらす。従来資産価格理論の実証では、この影響が軽視されてきた。

そこで、資産価格のプレミアム(期待リターン)を対象にオーバーフィッティングの影響を定量的に評価し、これを考慮した実証分析を実施した。ファイナンスの学術研究の発展に貢献するのはもちろんのこと、実務の意思決定にも応用可能な基礎研究を行った。

3. 研究の方法

本研究はファイナンスの研究であり、方法論は先人が築き上げたファイナンス研究に基づく。ただし、機械学習の技術を活用する。機械学習はこれまでファイナンスが頼ってきた伝統的な計量経済学とはカルチャーが大きく異なる。機械学習では予測変数となりうる潜在的なデータを積極的に活用して、さまざまな異なるアルゴリズムを比較利用し、アウトオブサンプルでの予測精度を高めることを主眼としている。特に変数間の相互作用を考慮した非線形性を扱うためのモデル(回帰木やニューラルネットワークなど)や、高次元の変数を扱うための各種の技術が発展してきた。こうした技術を利用して、大規模データを用いた実証分析を行った。

4. 研究成果

得られた主要な成果は以下のとおりである。

(1) オーバーフィッティングの危険性に関する既存研究の多くはクロスセクションのリターン予測に関するもので、時系列についてはこれまでほとんど探究されていない。そこで時系列予測に関して、オーバーフィッティングの影響はどの程度生じるのか、また、オーバーフィッティングを考慮した場合に有意水準をどの程度に設定するべきかについて、マイニングを考慮した多重検定の手法を考察した。さらに、実際のデータを用いて、変数選択によるマイニングと機械学習に伴うモデル選択によるマイニングの影響を分析した。この結果、みかけ上はアウトオブサンプルの予測精度が高く統計的に有意なものを容易に構築できるが、多重性を考慮した検定を行うと、ほとんどが偽発見として判定される。時系列はクロスセクションに比べて予測対象の標本数が少ないことから、オーバーフィッティングの影響はより深刻である。資産の実現リターンは事前の期待値と事後的な期待外リターンから構成される。モデルとデータによって近似する対象は、観測できない期待リターンの成分である一方、予測精度を高めようとするとうる偶然に期待外リターンにフィットしてしまう。このことは、工学的なデータサイエンスの技術を金融の予測に用いることには危険が伴うことを意味する。

(2) 大規模データと機械学習を用いた米国株式市場のクロスセクションでのリターン予測に関する代表的な研究に Gu, Kelly and Xiu (2020, Rev Finance Stud)がある。この研究では、900超の企業特性を予測変数の候補の中から、さまざまな機械学習を用いて予測力のある変数を探索し、線形モデルに比べて非線形モデルの方が予測力は高いことを示している。しかし、バリュエーション効果などこれまでにリターン予測力がある企業特性が知られている。こうした既知の予測力を持つ企業特性をどのように組み合わせるのかという問題に対して、機械学習は有効かもしれない。これを分析するため予測変数のマイニングをせずに、114種類の機械学習モデルを用いて実証した結果、モデルフリーの方法による予測力に及ばず、複雑なモデルほど予測力が悪化することを得た。これは真の資産価格モデルの決定係数は低いため、モデルエラーやパラメータ推定の誤差に起因することによる。またシミュレーションによる結果から、学習用データを大量に用いれば複雑な機械学習モデルの予測力は高まるが、現実的なヒストリカルデータの期間では学習に限界があることも得た。金融データを実験で増やすことはできない。こうした結果は、単純に機械学習に頼ることはできず、経済理論に基づく制約などを設けて利用した方が良いことを示唆する。

(3) 一方で、既存研究で予測力があると報告されてきた予測変数は、多くの変数の候補を試し、偶然に予測力が高かったものだけを報告した結果に過ぎない可能性がある。これを考察するには、多くの変数の候補を対象にした検定を行えばよい。そこで、123種類の企業特性の組み合わせをランダムに抽出することによって、各種の機械学習モデルによる株価リターンの予測の精度を検証した。この結果、機械学習モデルは簡素な線形モデルと比べて、平均的には予測精度が変わらないことを得た。ただし、なかには優れた予測力を持つ予測変数と機械学習アルゴリズムの組み合わせが存在する。これが偶然によるものか、あるいは真に予測力があるものかを識別するために、多重検定の枠組みで検定した結果、偶然とはいえ予測力を持つものが一定程度存在することが判明した。これは、予測変数と期待リターンとの間の非線形な関係や予測変数間の相互作用を機械学習によって捕捉できる可能性を示している。

(4) 前述の結果により、資産の期待超過リターンを推定する際に経済理論に基づく制約を課すべきであることが示唆される。資産価格理論に基づけば、資産の期待超過リターンは、システムティックなファクターに対するエクスポージャー(ベータ)とファクターのプレミアムによって定まる。ただし、理論的にファクターは複数存在するはずであるものの、ファクターが何かは自明ではなく実証的に特定される必要がある。そこで、Gu, Kelly and Xiu (2021, J Econom)の方法を利用して、日本株式市場を対象に機械学習により実証的なファクターモデルを構築した。この方法の利点は、ファクターモデルという経済学的な制約に加えて、時間変動するベータとファクターリターンを同時に推定できる点にある。構築したファクターモデルを用いて、理論的に説明するのが難しい実証的特徴であるわが国市場のバリュエーション効果の検証をしたところ、標準的に用いられている Fama-French の 5 ファクターモデルよりも説明力が高いことを得た。さらに、期待超過リターンを推定する上で、ファクターを特定することに加え、ベータの時間変動も考慮することが重要であることも明らかになった。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 内山朋規・菊川匡	4. 巻 11
2. 論文標題 債券運用におけるキャリア・ロールダウンのファクター投資	5. 発行年 2021年
3. 雑誌名 企業年金	6. 最初と最後の頁 16-19
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 内山朋規・瀧澤 秀明・菊川匡	4. 巻 65(7)
2. 論文標題 資産リターンの予測可能性と機械学習 - 危険なデータマイニング -	5. 発行年 2020年
3. 雑誌名 オペレーションズ・リサーチ	6. 最初と最後の頁 367-373
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 内藤誠・内山朋規・清水康弘・西内翔	4. 巻 59(3)
2. 論文標題 アルファカベータか - 機械学習アプローチによるバリュエーション効果の解明 -	5. 発行年 2020年
3. 雑誌名 証券アナリストジャーナル	6. 最初と最後の頁 29-40
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 水門 善之、内山 朋規	4. 巻 57
2. 論文標題 金融市場が織り込む消費税率引上げの実施確率	5. 発行年 2019年
3. 雑誌名 証券アナリストジャーナル	6. 最初と最後の頁 80-89
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 水門 善之、内山 朋規	4. 巻 12
2. 論文標題 消費増税の影響を除くインフレ期待の計測と期待形成の分布変化の検証	5. 発行年 2019年
3. 雑誌名 行動経済学	6. 最初と最後の頁 49-52
掲載論文のDOI (デジタルオブジェクト識別子) 10.11167/jbef.12.S49	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計16件 (うち招待講演 5件 / うち国際学会 2件)

1. 発表者名 内山朋規
2. 発表標題 機械学習と株式投資
3. 学会等名 第17回日本統計学会春季集会 (招待講演)
4. 発表年 2022年

1. 発表者名 内山朋規・上川知雄
2. 発表標題 年金運用におけるホームバイアスのコスト
3. 学会等名 日本ファイナンス学会第4回秋季研究大会
4. 発表年 2022年

1. 発表者名 Tomonori Uchiyama and Hiroshi Takahashi
2. 発表標題 Announcement Effects of Dividend Increases and Share Repurchases: Evidence from the Japanese Market
3. 学会等名 JFA-PBFJ Special Issue Conference (国際学会)
4. 発表年 2021年

1. 発表者名 内山朋規
2. 発表標題 機械学習と株式投資
3. 学会等名 第314回MPTフォーラム（招待講演）
4. 発表年 2021年

1. 発表者名 内山朋規
2. 発表標題 機械学習と株式投資
3. 学会等名 第26回東京ファイナンスフォーラム
4. 発表年 2021年

1. 発表者名 内藤誠・清水康弘・内山朋規
2. 発表標題 企業特性に基づく最適ポートフォリオ 機械学習アプローチ
3. 学会等名 日本ファイナンス学会第29回大会
4. 発表年 2021年

1. 発表者名 内藤誠・内山朋規・清水康弘・西内翔
2. 発表標題 アルファかベータか 機械学習アプローチによるバリュエーション効果の解明
3. 学会等名 第310回MPTフォーラム（招待講演）
4. 発表年 2021年

1. 発表者名 菊川匡・内山朋規・売野隆一
2. 発表標題 グローバル債券市場におけるキャリア・ロールダウンのファクター効果と適用法
3. 学会等名 日本ファイナンス学会第28回大会
4. 発表年 2020年

1. 発表者名 清水康弘・内山朋規
2. 発表標題 機械学習はファクターの合成方法を学習できるか リターン予測とオーバーフィッティング
3. 学会等名 日本ファイナンス学会第28回大会
4. 発表年 2020年

1. 発表者名 岩澤誠一郎・内山朋規
2. 発表標題 Who Does Better/Poorer Stock Market Forecasts? Evidence from Japanese Professional Survey
3. 学会等名 日本ファイナンス学会第28回大会
4. 発表年 2020年

1. 発表者名 鈴木誠・内山朋規
2. 発表標題 ライフサイクルにおける動的ポートフォリオ選択問題：外国投資とフォワードバイアスを用いた為替ヘッジ
3. 学会等名 日本金融・証券計量・工学学会2020年度冬季大会
4. 発表年 2020年

1. 発表者名 内山朋規
2. 発表標題 ESG/SDGsと債券投資
3. 学会等名 FTSE Russell・野村証券セミナー「気候リスクとキャリー・ファクター：ソブリン債への取り組み」（招待講演）
4. 発表年 2021年

1. 発表者名 岩澤誠一郎、内山朋規
2. 発表標題 Who does better/poorer stock market forecasts? Evidence from Japanese professional survey
3. 学会等名 行動経済学会第13回大会
4. 発表年 2019年

1. 発表者名 水門 善之、内山 朋規
2. 発表標題 消費増税の影響を除くインフレ期待の計測と期待形成の分布変化の検証
3. 学会等名 行動経済学会第13回大会
4. 発表年 2019年

1. 発表者名 Tomonori Uchiyama
2. 発表標題 Return Predictability and Machine Learning: Dangers of Data-Mining
3. 学会等名 International Workshop "Digital Innovation in Finance" (国際学会)
4. 発表年 2018年

1. 発表者名 内山朋規
2. 発表標題 リターンの予測可能性と機械学習
3. 学会等名 Group Discussion of the Dangers of Backtesting バックテストの危険性（招待講演）
4. 発表年 2019年

〔図書〕 計2件

1. 著者名 アンティ・イルマネン 著、加藤康之・内山朋規・角間和男・鹿子木亨紀・野村アセットマネジメント株式会社 訳	4. 発行年 2021年
2. 出版社 一般社団法人金融財政事情研究会	5. 総ページ数 636
3. 書名 期待リターン	

1. 著者名 ラッセ・ヘジェ・ペデルセン 著、内山朋規，角間和男，浦壁厚郎 訳	4. 発行年 2018年
2. 出版社 金融財政事情研究会	5. 総ページ数 485
3. 書名 ヘッジファンドのアクティブ投資戦略：効率的に非効率な市場	

〔産業財産権〕

〔その他〕

-

6. 研究組織	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計2件

国際研究集会 TMU Workshop on Finance 2019	開催年 2019年～2019年
国際研究集会 International Workshop "Digital Innovation in Finance"	開催年 2018年～2018年

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------