

# ビッグデータと AI の ファイナンスへの利用

おおはしかずひこ  
大橋和彦

## 1. はじめに

本日は「ビッグデータと AI のファイナンスへの利用」というテーマで基調講演を行う機会を賜りまして、誠にありがとうございます。このテーマは既に世間で注目を浴びているものです。例えば、最近、日本経済新聞が「投資 AI と人間」というテーマで 5 回の特集を組みました。内容をみると、人工知能（Artificial Intelligence: AI）を使い米国で活躍している日本人、侍クオンツがいたり、AI のみを使って投資判断をしているファンドがあったりという話がある一方、AI を利用しない人間が地道な銘柄選択を通じてあまり目立たない企業の株式を購入したところ、後に株価が大幅に上昇したという話もありました。別の例としては、ある経済学者が日本経済学会の講演で「将来的に経済学は機械学習の植民地になる」と指摘されました。これを受けて経済学者達は戦々恐々とされているかもしれません。

私の理解では、今回のファイナンス・ワークショップは、「ビッグデータと人工知能を用いたファイナンス研究の展開」をテーマとした前回からの流れを引き継いでいます。前回は、東京大学の和泉先生が基調講演をされ、AI がさまざまな分野において急速に応用されている理由を 2 つ指摘されました（和泉 [2019]）。第 1 の理由は、ビッグデータを利用できる環境になったことです。第 2 の理由は、AI、特に機械学習といった、ビッグデータを分析する技術が発展したことです。実際、ビッグデータとそれを処理する AI の利用がさまざまな方面に非常な勢いで拡大しています。あわせて和泉先生は、このことがどのような変化をもたらすのか、またその先にどのような世界が生まれるのか模索していく必要性を指摘されました。

私は、この基調講演において、今回のワークショップのテーマである「ビッグデータ・AI を活用したリスク計測・分析」の論点を整理する目的のもと、ビッグ

.....  
本稿は、2019 年 11 月 28 日に日本銀行金融研究所が開催したファイナンス・ワークショップ「ビッグデータ・AI を活用したリスク計測・分析」における基調講演である。ただし、本稿に示されている意見は、筆者個人に属し、日本銀行の公式見解を示すものではない。また、ありうべき誤りはすべて筆者個人に属する。

大橋和彦 一橋大学大学院経営管理研究科教授 (E-mail: kohashi@hub.hit-u.ac.jp)

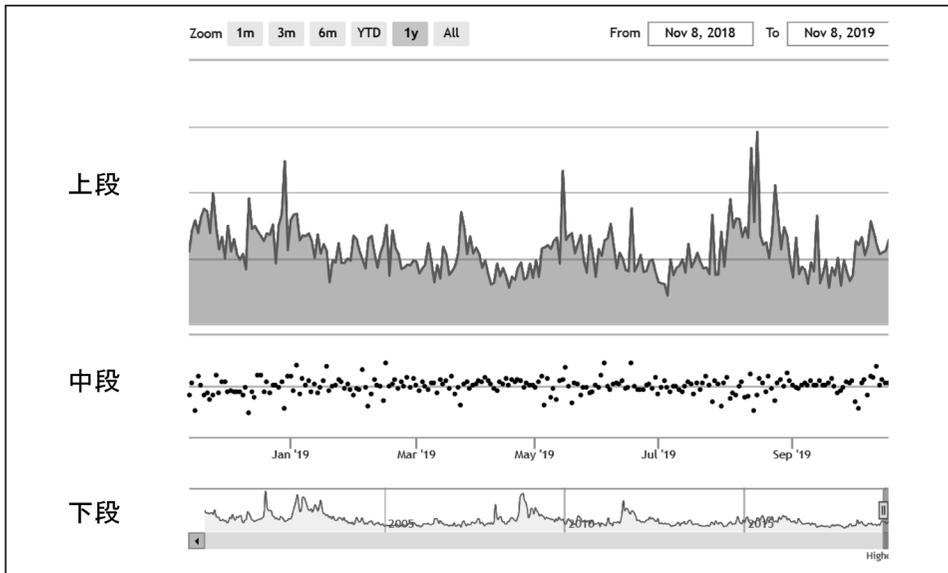
データと AI のファイナンスへの利用方法を 3 つに分類します。第 1 の分類は、ビッグデータ等の新しいデータを使って新しい側面をみるというものです。第 2 の分類は、AI や機械学習等の新しい手法を使って、新しい視点で分析するというものです。最後の分類は、新しいデータと新しい手法を用いれば新しい課題が生まれますので、それにどのように対応するかというものです。これら 3 つの分類はお互いに密接に関係しているため、別個に考えることは不適切かもしれませんが、ここでは「ビッグデータと AI のファイナンスへの利用」について論点を整理するため、あえてそうした取扱いを行います。

## 2. 新しいデータで新しい側面をみる

第 1 の分類に当たる試みは、ファイナンス研究において繰り返されてきたといっても過言ではありません。まず、価格データを使ったファイナンス研究であれば、データの高頻度化が綿々に行われてきました。年次が月次になり、月次が週次、週次が日次、日次が日中、さらには日中データの頻度を増やしていくといった具合です。そうした高頻度データは、マーケット・マイクロストラクチャーといわれる分野において最もよく使われてきました。同分野においては、取引の仕組みや制度、そして市場参加者の特性が、価格発見機能や流動性供給機能といった市場機能に与える影響が分析されます。近年では、高頻度取引 (high frequency trading) やアルゴリズム取引 (algorithmic trading) がそのプレゼンスを高める中、それらが市場機能に対して与える影響の解明に力が注がれており、そこで使われるデータは、取引ごとのティックデータです。私が学生だった 30 年ほど前、既にティックデータを用いたマイクロ・マーケットストラクチャーの分析が始まっていましたが、近年の分析において利用されるデータの頻度は、当時に比べてかなり高いものになっています。

高頻度データのもう 1 つの応用例はボラティリティ分析です。ボラティリティはファイナンス研究において非常に重要な指標であり、高頻度データの利用は、その理解の深化に役立っています。例えば、日中のデータを使うことによって、まさに今日実現しているボラティリティの算出が可能になりました。オックスフォード・マン定量ファイナンス研究所 (Oxford-Man Institute of Quantitative Finance) という英国の研究機関では、毎日、主要な株価指数に関する実現ボラティリティを計算して、研究者にも公開しています (図 1 参照)。日次実現ボラティリティの応用例として、その期待値とインプライド・ボラティリティの差をとることにより、分散リスク・プレミアム (Variance Risk Premium) を計算できます。この指標は、ボラティリティが時間を通じて変化することに対して投資家が要求するプレミアムを表

図1 ドイツ株価指数（Deutscher Aktienindex: DAX）の日次実現ボラティリティ



備考：上段は選択期間（2018年11月8日～2019年11月8日）における日次実現ボラティリティ（夜間は除く）、中段は始値と終値の変化率（Open-to-Close Return）、下段は日次実現ボラティリティの長期ヒストリカル・データを示している。

資料：Heber *et al.* [2009]

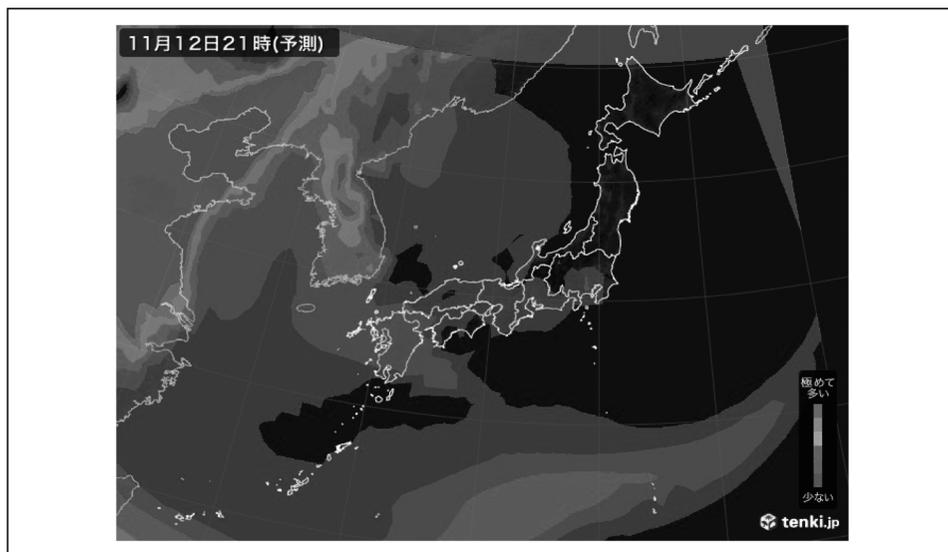
しており、アセット・プライシングの分野において、昨今注目を浴びています。

利用可能となった新しいデータには、高頻度データだけではなく、オルタナティブ・データと呼ばれるものもあります。これは、投資の意思決定に伝統的に用いられてきた一般的な公開情報（価格、取引量、財務・非財務情報等）以外のデータの総称です。具体例としては、衛星画像データがあります。これを分析すれば、世界中の石油タンクの備蓄量を知ることができます。備蓄量は石油の価格形成に大きな影響を与えるため、同データの利用は、石油価格の予測に役立つというわけです。また、衛星画像データを使って小売店等の駐車場の様子をみることにより、お店がどのくらい繁盛しているのか調べることもできます。

オルタナティブ・データのもう1つの例として、環境データが挙げられます。経済活動が活発になるとPM2.5（微小粒子状物質）の濃度が高くなりますので、それをみれば活発度合いがわかるというわけです（図2参照）。

さらなる例は、店頭レジで商品が販売されたときに記録されるデータ、すなわち販売時点情報管理（Point of Sales: POS）データです。これを使えば、どの商品がいつ、どこで、いくらで売れたかがわかります。この情報は、マーケティングに利用できるほか、学術的にも日次の物価指数を構築するために利用されています。

図2 PM2.5の分布予測



資料：日本気象協会 tenki.jp

このほか、ウェブページの閲覧履歴やテキストデータもオルタナティブ・データです。特に後者については、ニュースやレポート、ツイッターの文章をテキスト分析し、その結果を利用するというアプローチが、最近、広く採られています。

このように、伝統的なファイナンス研究では用いられてこなかったさまざまな新しいデータが利用可能となる中で、それらを分析することを通じ、新たな知見が得られているのです。

### 3. 新しい手法で新しい視点の分析をする

ビッグデータとAIのファイナンスへの利用に関する第2の分類は、新しい手法で新しい視点の分析をするというものです。この種の分析は、実務の中でも盛んに行われています。例えば、AIによってビッグデータを処理すれば、企業や個人の信用評価を行うことができます。また、こうした分析を通じて、企業の不正会計を検知・予測することも可能になるそうです。さらに、足許の国内総生産（Gross Domestic Product: GDP）・物価を予測するナウキャストリングや、投資家センチメントの定量化を通じた株価予測もそうした分析に当たります。

学術研究においても、ビッグデータとAIの利用は盛んです。日本銀行金融研究

所から近年公表された論文でもそうした分析が行われています(熊野・五島 [2018]、五島・高橋・山田 [2019]、風戸・黒崎・五島 [2019])。また、2019年に海外で発表された論文の中で、AIを使った分析アプローチを特徴付けるものとして、次の2本を紹介したいと思います。

まず1つは Kostovetsky and Warner [2020] です。テキスト分析の手法を用いて目論見書の内容を定量化し、それと投資信託の独自性・商品差別化の関係を分析しています。投資信託の商品情報は、多くの場合、リターンやリスクといった指標にくわえて、目論見書のテキストを通じて顧客に伝達されます。例えば、「愛国ファンド」(Patriot Fund) と称する投資信託では、米国の発展に貢献する企業のみに投資しているという特徴があります。こうした特徴は、財務指標からは計測できません。同研究では、こうした独自性をテキスト分析によって計測・定量化したうえで、①小規模で新しい投資信託グループの方が、そうでない投資信託グループよりも目新しく独自性の高い商品を開発し、高い手数料を顧客から徴収することができることや、②確立した (established) 投資信託では、独自性の高い方がそのパフォーマンスに対する資金の流出入の感応度が低くなることを報告しています。この研究は、新しいデータを新しい手法で加工したうえで、従来の方法で分析したものと評価できます。もう1つの論文は Gu, Kelly, and Xiu [2019] です。彼らは、資産価格の予測に当たり、機械学習を用いて有効な変数の絞込みや関数形の選択を行ったうえで、選択されたモデルが、既存のモデルよりも優れていることを示しています。この研究は、従来からあるデータを新しい手法で分析したものと評価できます。

#### 4. 新しい課題への対応

ビッグデータとAIのファイナンスへの利用に関する第3の分類は、新しい課題にどのように対応するかというものです。AI、特に機械学習はブラック・ボックスです。このため、AIによる分析結果は、何となくしっくりこないということもありません。どうしてそのような答えが出てくるのか説明できないにもかかわらず、AIで決めた人事や信用スコアに頼ってよいのでしょうか。また、金融機関は、顧客から預かった資金の運用について、なぜ得をしたのか、あるいは損をしたのか、顧客に対して説明する責任がありますが、AIによる運用においてこの責任を果たすことができるのでしょうか。機械学習のような手法をファイナンスの分野で使うためには、「なぜ」に答えられること、すなわち解釈可能性が求められます。

解釈可能性とは一体何なのでしょう。前述の海外論文2本を思い出して下さい。「新しいデータを新しい手法で加工し、線形回帰といった従来の方法で分析した」1本目 (Kostovetsky and Warner [2020]) と、「従来からあるデータを機械学習と

いった新しい手法で分析した」2本目（Gu, Kelly, and Xiu [2019]）のうち、分析結果に対する納得感が高いのはどちらでしょうか。多くの人が1本目を選ぶものと思われます。この理由は、説明変数がどのようなメカニズムで被説明変数を左右するか明らかな場合において、人間はその分析結果が解釈可能であると感じるからだと思えます。

もしそうだとするならば、こうした解釈可能性を部分的にでも確保する手法として、説明変数の変化が被説明変数に与える影響の度合いから変数の重要性を評価する手法（Janitza, Celik, and Boulesteix [2018]、宇宿ほか [2019]）や、説明変数やモデルの構造を先に定めてからパラメータの推定に機械学習を利用する手法が検討されてしかるべきでしょう。

ここで注意したいのは、①解釈可能性と因果関係は別物であること、②モデルを解釈するために課した制約が予測精度を減じてしまう、すなわち、モデルの解釈可能性と予測精度の間にトレードオフがあること、そして、③人間にはよくわからないものの、機械にはわかる何らかの構造が存在し、それが人間の理解を深める可能性があることです。AI、特に機械学習の解釈可能性は、近年、数学者をはじめとする研究者から注目を集めています。この分野の研究は、今後、大いに進展するものと期待されます。

## 5. おわりに

---

最後に、ビッグデータとAIのファイナンスへの利用に関する3つの分類を、本ワークショップにおける3本の報告論文と関連付けたいと思います。新しいデータで新しい側面をみるという第1分類は、Ogawa, Ubukata, and Watanabe [2019]に該当し、新しい手法で新しい視点の分析を行うという第2分類は、Miyakawa and Shintani [2019]に該当します。そして、新しい課題への対応という第3分類は、解釈可能性を求めた点で五島・三木 [2019]に該当します。こうした関連付けが一助となり、参加者の皆さまにとって本ワークショップが有意義なものとなることを期待しています。ご清聴、ありがとうございました。

## 参考文献

- 和泉 潔、「ビッグデータと人工知能を用いたファイナンス研究の潮流」、『金融研究』第38巻第1号、日本銀行金融研究所、2019年、15～28頁
- 宇宿公紀・加藤尚吾・加藤由樹・千田国広、「LINEグループにおいて返信ができないことで生じるネガティブ感情：ネガティブ感情が生じるまでの時間と性格特性及びLINEメール依存度との関係」、『日本認知心理学会発表論文集』、日本認知心理学会第17回大会、2019年、115頁
- 風戸正行・黒崎哲夫・五島圭一、「日本銀行による景気判断のトーン分析」、金融研究所ディスカッション・ペーパー No. 2019-J-16、日本銀行金融研究所、2019年
- 熊野雄介・五島圭一、「金融政策アナウンスメントとアルゴリズム取引：ウェブページへのアクセス情報を用いた検証」、金融研究所ディスカッション・ペーパー No. 2018-J-11、日本銀行金融研究所、2018年
- 五島圭一・高橋大志・山田哲也、「自然言語処理による景況感ニュース指数の構築とボラティリティ予測への応用」、『金融研究』第38巻第3号、日本銀行金融研究所、2019年、1～42頁
- ・三木翔太、「深層学習による株価収益率ボラティリティ予測：自己注意機構付きモデルの応用」、日本銀行金融研究所ファイナンス・ワークショップ報告論文、2019年
- Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu, “Empirical Asset Pricing via Machine Learning,” NBER Working Paper No. 25398, National Bureau of Economic Research, 2019.
- Heber, Gerd, Asger Lunde, Neil Shephard, and Kevin Sheppard, “Oxford-Man Institute’s Realized Library, Version 0.3,” Oxford-Man Institute, University of Oxford, 2009.
- Janitza, Silke, Ender Celik, and Anne-Laure Boulesteix, “A Computationally Fast Variable Importance Test for Random Forests for High-dimensional Data,” *Advances in Data Analysis and Classification*, 12(4), 2018, pp. 885–915.
- Kostovetsky, Leonard, and Jerold B. Warner, “Measuring Innovation and Product Differentiation: Evidence from Mutual Funds,” *Journal of Finance*, 75(2), 2020, pp. 779–823.
- Miyakawa, Daisuke and Kohei Shintani, “Systematic Disagreement between Human and Machine Predictions,” paper presented at the Financial Workshop held by the Institute for Monetary and Economic Studies, Bank of Japan, 2019.
- Ogawa, Toshiaki, Masato Ubukata, and Toshiaki Watanabe, “Stock Return Predictability and Variance Risk Premia around the ZLB,” paper presented at the Financial Workshop held by the Institute for Monetary and Economic Studies, Bank of Japan, 2019.

