

研究報告書

「関数空間上への機械学習理論の展開と高頻度金融データ解析」

研究タイプ: 通常型

研究期間: 平成27年10月～平成31年3月

研究者: 荻原 哲平

1. 研究のねらい

金融機関は保有する株式の株価変動による金融資産価値の変動をコントロールする必要があり、株価データや財務データから統計解析を用いて将来の株価の分散・共分散を予測してリスク・コントロールを行っている。特に年金基金の株式運用やインデックス・ファンドの運用においては株価変動リスクをより正確にコントロールすることが求められ、株価の分散・共分散の高い予測精度が必要となる。従来の株価変動分析では一日の終わりの株価を用いた日次データによる解析が中心であった。このような低頻度のデータにおいては株価分散にランダムな構造を取り入れた GARCH 型モデルや、株価変動を財務数値や市場全体との連動性などの幾つかのファクターで説明するマルチ・ファクター・モデルなどが提案され、これらを用いて株価変動リスクを定量化して株式運用に用いられている。一方で、近年各証券取引所における一日内の全ての取引の取引時刻・取引価格・売買高等の情報を記録したような「高頻度金融データ」の利用可能性が高まっており、その膨大な情報量から金融市場の分析の大きな進展が期待されている。しかし、高頻度金融データには日内の周期性や強い観測ノイズの存在、観測のランダム性、複数資産の観測時刻の不一致性、株価変動の強い非正規性などの複雑な構造が統計解析を困難にしているため、それらを包括的に考慮したモデルの構築や統計解析は困難を極めている。

機械学習の手法は複雑な構造をもつ事象に対し、データの構造をコンピュータに学習させることでコンピュータの計算力を活用した大規模な解析を可能にする。特にデータ量が豊富である時に効果を発揮する。高頻度金融データはその複雑なモデル構造の特定が困難である一方で、膨大なデータ量を含むため、機械学習を用いたアプローチが有効であると考えられる。高頻度データの複雑なデータ構造から従来の機械学習手法の適用は困難であるが、本研究では最新の確率解析・統計解析手法と機械学習を融合させることで、高頻度金融データの扱いを可能にすることを目指す。このようなアプローチにより、高頻度金融データの豊富な情報量を株価変動リスクの予測に効率的に活用することが可能になり、特に従来の低頻度のデータでは得られないような金融危機などの市場が急激に変化する局面に対応できるリスク管理手法の確立へ貢献していく。

2. 研究成果

(1) 概要

証券価格を時間の関数としてみた時に、関数空間の元とすることができる。この関数空間に確率構造を入れて、データを高頻度観測する時に推定関数を構築し、理論的な挙動を解析し

た。まず株価変動をパラメトリック・モデルで記述し、高頻度金融データ特有の問題として、異なる株式間で観測時刻が一致しない「非同期観測」の問題や、観測に仮想的なノイズが生じるという経験則を考慮した上で、近似的な対数尤度関数を用いてパラメータを最尤型推定する手法を開発し、その効率性等の結果も含めて、【論文1】【論文2】に出版した。

その後、この手法を発展させて、パラメータ推定にニューラル・ネットワークを取り入れた手法を開発した。ニューラル・ネットワーク・モデルでは株価変動を記述する真のモデルがパラメトリック・モデルに含まれない状況が自然であり、このようなモデルはミスペシファイド・モデルと呼ばれる。確率過程の高頻度観測モデルにおいてはミスペシファイド・モデルの先行研究がないが、上記の最尤型推定量の理論を発展させることにより、推定関数の収束値を特定した。さらに最尤型推定量に推定バイアスが生じることを示し、このバイアスを修正した推定量を提案することで、推定精度が理論的に最良のレートを達成していることを確認した。これらの提案手法のアルゴリズムに関する特許出願を行った。

このような提案手法に対して、確率過程の基本的なモデルのシミュレーションにより、推定関数がモデル構造を精度よく推定できることを確認した。また、日本株式市場の主要銘柄の高頻度金融データから株価構造、特に株価変動の分散・共分散に相当するボラティリティ・共変動をニューラル・ネットワークにより学習した。高頻度データ特有の観測の複雑さに対処し、データからボラティリティ・共変動の構造を学習するアプローチは今までにない画期的なものである。特に日経平均のボラティリティ・インデックスを株価変動の説明変数に加えることでボラティリティの予測精度が向上することを確認した。

関連する研究として、ベイズ型推定法の理論を発展させ、一般の損失関数に対するベイズ型推定法の一般理論を確立して【論文3】にまとめた。また、提案手法を代表的な機械学習手法の一つであるカーネル法へと拡張させる研究や、ミスペシファイド・モデルの理論を応用した株価変動のノンパラメトリック推定量の構築の研究などを行った。

(2) 詳細

研究テーマ1【関数空間上の確率モデルに対する最尤型推定手法の構築】

高頻度金融データ特有の問題である非同期・ノイズ付観測を考慮した統計モデルにおいて、最尤型推定量の研究を行った。このようなモデルでは最尤推定に必要な尤度関数を計算することは困難であるが、局所ガウス近似を用いた近似的な対数尤度関数(疑似対数尤度関数)を構築することで最尤型推定量を構築した。さらに観測が高頻度になる極限において最尤型推定量が一致性や漸近混合正規性等の推定量として望ましい性質を満たすことを確認した。

研究テーマ2【関数空間上の確率モデルに対するニューラル・ネットワーク手法の構築】

確率過程を高頻度観測する統計モデルにおいて、機械学習の代表的手法であるニューラル・ネットワークを構築し、観測が高頻度になる極限における推定関数の理論的な挙動を研究した。ニューラル・ネットワークを取り入れることによりあらかじめ株価モデルを特定することなく、その非線形構造をデータから効率よく学習することが可能となる。非同期・ノイズ付観測といった観測の複雑さの下では、ニューラル・ネットワークで通常用いられる最小二乗型のパラメータ推定は困難であり、研究テーマ1で構築した疑似対数尤度関数を損失関数として用い

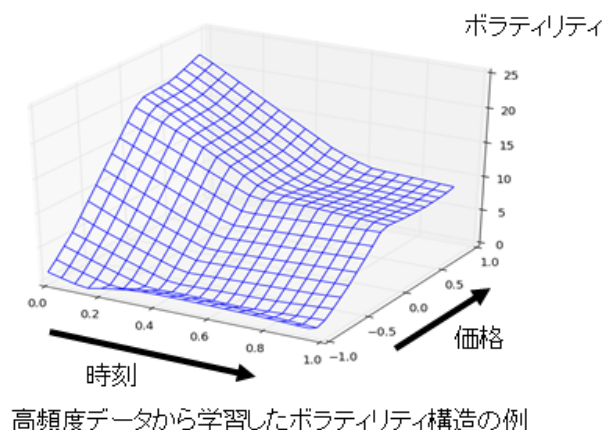
ることにより高頻度観測に適用した推定関数を構築した。研究テーマ1の理論を発展させることにより、このようなニューラル・ネットワークを含むミスペシファイド・モデルにおける推定関数の収束値を理論的に特定した。パラメトリック・モデルが真のモデルを含む状況では推定関数の極限への収束の速さがデータ数の $-1/4$ 乗となり、これはあらゆる推定手法の中で最良になっていることが示されている。ミスペシファイド・モデルでは上記の最尤型推定量は最速レートを達成せず推定バイアスが生じることを示し、このバイアスを修正した推定量を提案することで最速レートを達成することを確認した。

また、最尤型推定量を用いた推定関数は巨大な行列の逆行列計算を含み、多くの計算量が必要とされるが、逆行列計算を含まない近似的な推定量を提案することにより、計算時間を飛躍的に改善した。

研究テーマ3【シミュレーション・日本株式市場データに対する分析】

代表的な確率過程のモデルである幾何ブラウン運動、Cox-Ingersoll-Ross モデルに対して研究テーマ2の提案手法のシミュレーション分析を行った。プログラミング言語 Python の Chainer パッケージを用いて上記の疑似対数尤度関数を独自の損失関数として実装し、学習を行った。シミュレーションに用いられるモデルを未知としてデータから学習させた時に、データが高頻度になるにつれて推定関数が真のモデルを精度よく近似することを確認した。

また、研究テーマ2の提案手法により日本株式市場の主要 30 銘柄の高頻度金融データから株価構造、特に株価変動の分散・共分散に相当するボラティリティ・共変動をニューラル・ネットワークにより学習する実証分析を行った。ボラティリティを時間の関数と見た時に取引開始時や終了時にボラティリティが上昇するという、経験的に知られている結果を確認した。また、株価モデルで重要な「ボラティリティの長期記憶性」を表現するために、ボラティリティの説明変数として「日経平均ボラティリティ・インデックス」を導入して東証株式市場の高頻度データからの学習を行い、想定通りボラティリティ・インデックスの値が高い時に株価ボラティリティの推定値が高くなるという現象を確認した。さらに、過去データからボラティリティ構造を学習し、将来のボラティリティを予測する分析を行い、「日経平均ボラティリティ・インデックス」の予測における有用性を示した。



研究テーマ4【提案手法の発展と関連研究】

疑似対数尤度関数を用いたアプローチにより、最尤型推定量だけでなく、ベイズ型推定量も構築することが可能であり、Yoshida (AISM 2011)の多項式型第偏差不等式を用いた理論を応用することにより、ベイズの損失関数が一般の関数である時にその漸近理論を扱う手法を確立した。

また、ニューラル・ネットワーク以外の機械学習の展開として、代表的な機械学習手法の一つであるカーネル法を関数空間上の確率モデルに適用した。ニューラル・ネットワークと異なり、将来時刻の参照が起こる問題とパラメータの次元が無限大へ発散する問題が生じるが、過去データを用いて推定関数を作成する等によりこれらの問題を解決し、観測が高頻度になる極限において推定関数の収束先を特定した。

さらに、ニューラル・ネットワークの理論開発で用いたミスペシファイド・モデルの理論を応用して、一般の拡散過程をシンプルなブラウン運動とみなして最尤型推定量からボラティリティ・共変動推定量を導出し、一致性・漸近混合正規性等の理論的に望ましい性質が得られた。このようなアプローチにより最尤型推定量を計算できないような隠れ変数を含む確率過程モデルに対して効率的な共変動推定量が構築されると期待され、特に重要な株価モデルの一つである確率ボラティリティ・モデルへの適用が可能になる。確率ボラティリティ・モデルに対するシミュレーションにより、既存の推定量に比べてこの提案手法の推定精度が最も良いパフォーマンスとなることを確認した。

3. 今後の展開

高頻度金融データにパラメトリック・モデルを導入し、株式市場の株価間の分散共分散行列を過去データから学習させることで、リーマン・ショックや東日本大震災などの金融危機に伴う証券市場の構造変化に対して迅速に対応できるリスク・コントロールモデルの構築が期待される。現在主流である日次データを用いたマルチ・ファクター・モデルによるリスク・コントロールでは、このような急激な構造変化に弱く、株式資産のリスク・コントロールができなくなるため、代替モデルとしての利用が可能である。

また、現在株式売買注文情報の点過程によるモデリングと統計解析が活発に研究されている。このようなモデリングは株価形成プロセスをより自然に反映しており、機械学習をこのような点過程モデルで展開していくことで株式市場のミクロ構造をより正確にとらえた解析が可能になると期待される。特に株式の大量売買執行が株価に与える影響(マーケット・インパクト)は巨額の株式売買を行う必要のある年金基金などの投資家にとって重要な問題となる。

本研究は高頻度金融データの分析への応用を意識して行ってきたが、関数空間上の確率構造や観測ノイズ、非同期観測といったデータ観測上の問題は他分野の統計解析においても現れると考えられ、より広く応用できる可能性がある。

4. 自己評価

ニューラル・ネットワーク以外の機械学習の手法(特にカーネル法)に提案手法を拡張することをより研究する想定だったが、計算コストの面であまり相性が良くないため、理論研究にとどまった。一方でニューラル・ネットワークと提案手法の相性が良く、想定していた以上の高速計算が可能となった。ミスペシファイド・モデルの理論研究は確率過程では困難が多く、時間がかかってしまったが、推定バイアスの存在とその補正や、ランダム・パラメータに対する収束

定理などの先行研究では見られないような理論的成果が得られたのは数学的にも興味深いと感じた。また、ミススペシファイド・モデルを応用することで共変動を推定する新たな良い手法を発見することができた。

実証分析では、民間企業との共同を通して社会実装に向けて軌道に乗せていくことを考えていたが、この点に関してはうまく協力が得られず、想定していたより進めることができなかった。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

- | |
|--|
| 1. Ogiyara, T. Local asymptotic mixed normality property for nonsynchronously observed diffusion processes, Bernoulli, 2015, 21 2024–2072. |
| 2. Ogiyara, T. Parametric inference for nonsynchronously observed diffusion processes in the presence of market microstructure noise, Bernoulli, 2018, 24(4B) 3318–3383. |
| 3. Ogiyara, T. On the asymptotic properties of Bayes-type estimators with general loss functions, Journal of Statistical Planning and Inference, 2019, 199 136–150. |

(2) 特許出願

研究期間累積件数: 1 件(公開前の出願件名については件数のみ記載)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

口頭発表

- Parametric inference for diffusion processes with high-frequency financial data, The 4th Institute of Mathematical Statistics Asia Pacific Rim Meeting, Jun 2016, Hong Kong
- 非同期・ノイズ付観測された拡散過程に対する統計解析, 日本数学会 2016 年度秋季総合分科会特別講演, 2016 年 9 月, 大阪
- Parameter estimation for misspecified diffusion with market microstructure noise, Stochastic Processes and Risk Analysis, Oct 2018, Tokyo
- Local asymptotic mixed normality for diffusion processes with irregular observations, Asymptotic expansion and Malliavin calculus, Nov 2018, Paris

研究詳解

- 荻原哲平, 拡散過程による日内株価データのモデリングと統計推測理論, 統計数理 65-1, 5–20. (2017)