

令和 6 年 6 月 2 日現在

機関番号：12601

研究種目：若手研究

研究期間：2019～2023

課題番号：19K13668

研究課題名（和文）高次元高頻度データの統計解析

研究課題名（英文）Statistical analysis of high-dimensional high-frequency data

研究代表者

小池 祐太 (Koike, Yuta)

東京大学・大学院数理科学研究科・准教授

研究者番号：80745290

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）：本研究では、高次元高頻度金融データからその相関構造を統計推測するための手法について研究し、以下の研究成果を得た。

（1）精度行列と呼ばれる、金融資産の間の見せかけの相関を排した相関関係を計測する統計量を推定する方法を提案し、さらに推定誤差の分布を近似計算する方法を与えた。（2）非常に多数の仮説を統計的に検証する際に理論的に有用な役割を果たす、統計量の正規近似の相対誤差を系統的に評価するための理論を開発した。（3）データ間の相関関係を駆動する主因であるファクターがいくつか存在するか推定する方法を開発した。

研究成果の学術的意義や社会的意義

高次元高頻度データの統計学に関するこれまでの理論的研究は点推定が主流であり、特に推定量の一致性や収束レートに関するものがほとんどであった。すなわち、データ数を多くするにつれて推定誤差が0に近づいていくことは示されてきたが、具体的に推定誤差がどの程度の大きさか見積もる研究はこれまでほとんどなされてこなかった。本研究では、高次元高頻度データの相関構造に対するいくつかの推定量に対して推定誤差の確率分布の近似手法を与え、かつその理論的正当性のある程度一般的な枠組みで示したという点で意義がある。

研究成果の概要（英文）：This study has investigated statistical inference methods for the correlation structure of a large number of financial assets from their high-dimensional high-frequency data. Specifically, I have obtained the following results:

(1) I have proposed a method to estimate the precision matrix of a large number of assets from their high-dimensional high-frequency data. Besides, I have developed a method to approximately compute the distribution of the estimation error. (2) I have developed a theory to systematically estimate the relative errors of normal approximations for various statistics. This serves as justifying the validity of some multiple testing procedures. (3) I have proposed a method to estimate the number of relevant factors from high-frequency data.

研究分野：数理統計学、計量ファイナンス

キーワード：高頻度データ 高次元共分散推定 多重検定 ファクターモデル

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

金融市場における取引のデータを用いた、複数の金融資産の収益率間の共分散構造の推定は、金融機関におけるリスク管理や投資戦略の策定において基本的な役割を果たす。金融市場では多くの資産が高頻度で取引されており、このような高頻度データを用いることで1日単位といった非常に短いスパンでも一定以上のサンプル数を確保できる。一方で、データの次元は分析対象の資産数であるが、通常は投資対象の資産は非常に多数であるため、たとえ高頻度データを用いたとしてもデータの次元がサンプル数を超えてしまうケースが往々にしてある。そのため、このような高次元高頻度データを用いて金融資産の収益率間の共分散構造を推定するための統計理論の整備が、近年の高頻度データ解析における重要な研究課題として、2010年代ごろから活発に研究されてきた。

2. 研究の目的

従来の高次元高頻度データを用いた共分散構造推定の研究では、共分散行列やその逆行列(精度行列)の点推定の研究がメインであった。すなわち、具体的な推定量を構成し、その推定量の推定誤差がサンプル数を無限大とした極限で0に(確率)収束するという性質(一致性)を持つことを示すという研究がほとんどであった。本研究の目的は、推定誤差のより具体的な評価を通じて、区間推定や仮説検定といった統計推測の問題を解くことである。

3. 研究の方法

区間推定や仮説検定を実行するためには、推定誤差が従う確率分布を(近似)計算できればよい。このために、高次元高頻度データに基づく共分散行列や関連する統計量に対する推定量の誤差分布を近似計算するための統計理論を開発する。併せて、開発した統計理論を現実の問題に対して適用した際に十分な近似精度が得られるかどうかを検証するために、人工データを用いた数値実験を行う。さらに、実際の金融高頻度データに開発した統計手法を適用することで新たな知見を得ることを目指す。

4. 研究成果

(1) 高次元高頻度データを用いた精度行列に対する統計推測理論の整備

金融資産の収益率間の精度行列は、大域的分散ポートフォリオの投資比率の計算や、資産間の偏相関構造の決定などにおいて重要な役割を果たす。しかし、データの次元がサンプル数よりも大きいような高次元の設定では、標準的な共分散行列の推定量は逆行列を持たないため、精度行列の推定量の構成自体が非自明な問題となる。そのため、共分散行列もしくは精度行列になんらかの構造的制約を課して推定量を構成する必要がある。本研究では、独立観測の設定でよく用いられる、精度行列の非対角成分にスパース性を課して重み付き L^1 型罰則を付けた対数擬似尤度関数の最大化によって精度行列の推定量を構成する方法 (weighted graphical Lasso) を高次元高頻度データに適用した際の理論的性質を明らかにした。具体的には、上記の罰則付き対数尤度のインプットとして与える共分散行列の推定量が最大値ノルムに関する集中不等式を満たしていれば、得られた weighted graphical Lasso 推定量は作用素ノルムに関して一致性を持つことを示した。

上の結果は精度行列の点推定に関する研究成果であるが、本研究の主目的である統計推測を実行するための理論も構築した。すなわち、weighted graphical Lasso そのものは一次のバイアスを持つため、ワンステップ型のバイアス補正を適用した debiased version を考え、その推定誤差の確率分布を近似計算する手段を与えた。具体的には、インプットとして与える共分散行列の推定量がランダムなスパース多面体に含まれる確率を混合正規分布で近似できるのであれば、得られた debiased weighted graphical Lasso 推定量の誤差分布が矩形に含まれる確率を混合正規分布で近似できることを示した。

高頻度データではデータの観測頻度や目的に応じて共分散行列推定量の構成法を様々に変更する必要があるため、上で述べたような抽象的な条件下で精度行列の一致推定や統計推測を実行できる理論を構築したことは、今後の拡張を考える上でも意義がある。一方で、これらの抽象的な条件が実際に確認可能であることを例示するために、高頻度データにおける最も標準的な共分散推定量である実現共分散行列が実際にこれらの条件を満たすことを適当な仮定のもとで示した。また、金融データはしばしばファクター構造を持つため、全体の精度行列に対するスパース性の仮定は強すぎることが知られて

いる。そのため、既知のファクター構造を除いたあとの残差過程に対して理論を適用することの正当性も示した。

(2) 金融資産間のネットワーク構造の推定への応用

精度行列の非対角成分が0であることは、対応する変数間に他の変数を与えたもとの偏相関関係が存在しないことに対応している。このことを用いて、精度行列の0でない非対角成分を判別することで、変数間の偏相関関係を可視化したネットワーク（偏相関ネットワーク）を作成することができる。そこで、S&P500 指数構成銘柄間の偏相関ネットワークを、高頻度データを用いて(1)で開発した手法を用いて推定した。上述の通り、金融資産の収益率にはしばしばファクター構造が観察されるため、精度行列のスパース性をそのまま仮定するのは現実的でない。そのため、S&P 指数(厳密には SPDR S&P500 ETF)の収益率をファクターとみなして、そこから起因する変動を取り除いた残差系列の精度行列を推定してネットワークを作成した。点推定のみに基づく方法(weighted graphical Lasso)では、一般には精度行列の support recovery (ゼロ成分の特定)は完全には保証されておらず、しばしば冗長な推定結果を与える。今回の分析結果でもその現象が観察されたため(図1) de-biased version の漸近混合正規性を利用した仮説検定に基づく方法によってゼロ成分を特定する方法についても検証した(図2)。その結果、得られたネットワークには業種の構造がかなり反映されており、ある程度信憑性のある結果となった。

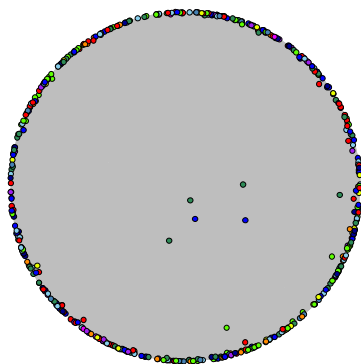


図1：S&P500 指数構成銘柄の残差系列の偏相関ネットワーク。Weighted graphical Lasso による推定結果

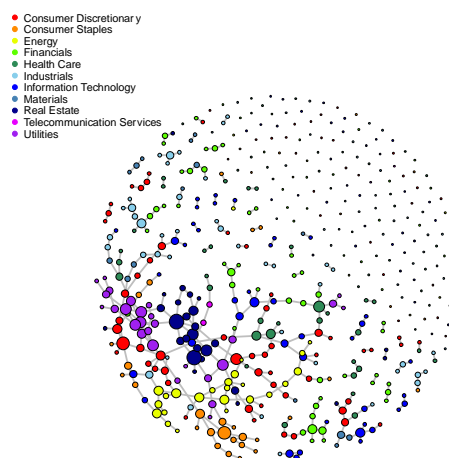


図2：S&P500 指数構成銘柄の残差系列の偏相関ネットワーク。De-biased weighted graphical Lasso による有意性検定に基づく推定結果

(3) 非線形な統計量に対する正規近似の相対誤差評価の導出

図2で有意な偏相関関係を検出する際に、多数の変数のペアについて同時に仮説検定を行う必要があった。このような場合、個々の検定のP値を補正しなければ過剰に帰無仮説を棄却してしまうという問題、いわゆる多重検定の問題が生じる。実際、図2の作成時にも、この問題に対処せずにネットワークを作成すると、辺が残りすぎて結果の解釈が難しくなったため、「Holmの方法」として知られるP値の補正法を用いて多重検定の問題に対処している。Holmの方法はFamily-Wise Error Rate (FWER)を有意水準以下に抑えるようにP値を補正する方法であるが、実行すべき検定の数が非常に多いと結果が保守的になりすぎるという問題点がある。そのため、より積極的に帰無仮説を棄却できるFalse Discovery Rate (FDR)を有意水準以下に抑える方法の利用が、近年高次元データ解析において広まっている。FDRをコントロールして多重検定を実行するための代表的な方法として、Benjamini-Hochberg法(BH法)がよく知られている。BH法の適用を高次元データにおいて理論的に正当化するには、正規近似の相対誤差に対するCramer型の評価が必要となるが、高頻度データ解析で利用されるようなセミマルチンゲールの汎函数で与えられる統計量に対してはそのような結果はほとんど知られていない。本研究では、Steinの方法と呼ばれる数学的技法が適用できるような統計量に対してそのような評価を導出するための系統的な理論を開発した。

(4) 「重要な」ファクター数の推定方法の開発

すでに述べたとおり、金融資産の収益率にはファクター構造がしばしば観察される。こ

のファクター構造を推定することは、すでに述べたような共分散行列や精度行列の推定をする際に重要であることに加えて、金融経済学の観点からも興味がある。上記の応用ではファクターは市場ファクターのみで既知と仮定していたが、実際にはその他にもファクターが存在することが過去の実証研究から示唆されているため、ファクター構造は未知としてそれを推定する方が妥当だと考えられる。さらに、分析対象すべての資産に影響を与えるような「強い」ファクターから、一部の資産にしか影響しない「弱い」ファクターまで存在し、それらを理論的に区別することが重要であることが先行研究で指摘されている。本研究では、そのようなファクターの強弱を反映したようなモデルを仮定したもとで、高次元高頻度データを用いて「重要な」ファクターの数を推定する方法を与えた。

5. 主な発表論文等

〔雑誌論文〕 計12件（うち査読付論文 12件 / うち国際共著 7件 / うちオープンアクセス 11件）

1. 著者名 Koike Yuta	4. 巻 36
2. 論文標題 High-Dimensional Central Limit Theorems for Homogeneous Sums	5. 発行年 2023年
3. 雑誌名 Journal of Theoretical Probability	6. 最初と最後の頁 1-45
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10959-022-01156-2	査読の有無 有
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 Chernozhukov Victor, Chetverikov Denis, Kato Kengo, Koike Yuta	4. 巻 10
2. 論文標題 High-Dimensional Data Bootstrap	5. 発行年 2023年
3. 雑誌名 Annual Review of Statistics and Its Application	6. 最初と最後の頁 427 ~ 449
掲載論文のDOI（デジタルオブジェクト識別子） 10.1146/annurev-statistics-040120-022239	査読の有無 有
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 該当する
1. 著者名 Chernozhukov Victor, Chetverikov Denis, Kato Kengo, Koike Yuta	4. 巻 50
2. 論文標題 Improved central limit theorem and bootstrap approximations in high dimensions	5. 発行年 2022年
3. 雑誌名 The Annals of Statistics	6. 最初と最後の頁 2562-2586
掲載論文のDOI（デジタルオブジェクト識別子） 10.1214/22-AOS2193	査読の有無 有
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 該当する
1. 著者名 Koike Yuta	4. 巻 4
2. 論文標題 Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles	5. 発行年 2021年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 257 ~ 297
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s42081-020-00096-7	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 Koike Yuta	4. 巻 4
2. 論文標題 Inference for time-varying lead-lag relationships from ultra-high-frequency data	5. 発行年 2021年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 643 ~ 696
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-021-00106-2	査読の有無 有
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 Fang Xiao, Koike Yuta	4. 巻 31
2. 論文標題 High-dimensional central limit theorems by Stein's method	5. 発行年 2021年
3. 雑誌名 The Annals of Applied Probability	6. 最初と最後の頁 602 ~ 631
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/20-AAP1629	査読の有無 有
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 該当する

1. 著者名 Fang Xiao, Koike Yuta	4. 巻 32
2. 論文標題 New error bounds in multivariate normal approximations via exchangeable pairs with applications to Wishart matrices and fourth moment theorems	5. 発行年 2022年
3. 雑誌名 The Annals of Applied Probability	6. 最初と最後の頁 602 ~ 631
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/21-AAP1690	査読の有無 有
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 該当する

1. 著者名 Yuta Koike	4. 巻 22
2. 論文標題 De-biased graphical Lasso for high-frequency data	5. 発行年 2020年
3. 雑誌名 Entropy	6. 最初と最後の頁 456 ~ 456
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e22040456	査読の有無 有
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 Fang Xiao, Koike Yuta	4. 巻 28
2. 論文標題 From p-Wasserstein bounds to moderate deviations	5. 発行年 2023年
3. 雑誌名 Electronic Journal of Probability	6. 最初と最後の頁 1 ~ 52
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/23-EJP976	査読の有無 有
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 該当する

1. 著者名 Chernozhukov Victor, Chetverikov Denis, Koike Yuta	4. 巻 33
2. 論文標題 Nearly optimal central limit theorem and bootstrap approximations in high dimensions	5. 発行年 2023年
3. 雑誌名 The Annals of Applied Probability	6. 最初と最後の頁 2374 ~ 2425
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/22-AAP1870	査読の有無 有
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 該当する

1. 著者名 Oga Akihiro, Koike Yuta	4. 巻 170
2. 論文標題 Drift estimation for a multi-dimensional diffusion process using deep neural networks	5. 発行年 2024年
3. 雑誌名 Stochastic Processes and their Applications	6. 最初と最後の頁 104240 ~ 104240
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.spa.2023.104240	査読の有無 有
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 Fang Xiao, Koike Yuta	4. 巻 34
2. 論文標題 Large-dimensional central limit theorem with fourth-moment error bounds on convex sets and balls	5. 発行年 2024年
3. 雑誌名 The Annals of Applied Probability	6. 最初と最後の頁 2065 ~ 2106
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/23-AAP2014	査読の有無 有
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 該当する

〔学会発表〕 計23件（うち招待講演 17件 / うち国際学会 11件）

1. 発表者名 Yuta Koike
2. 発表標題 Asymptotic mixed normality of realized covariance in high-dimensions
3. 学会等名 The 5th International Conference on Econometrics and Statistics (EcoSta 2022) (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 小池祐太
2. 発表標題 高次元中心極限定理の誤差評価に関する最近の進展
3. 学会等名 大阪大学確率論セミナー (招待講演)
4. 発表年 2022年

1. 発表者名 小池祐太
2. 発表標題 金融高頻度データにおける先行運行関係
3. 学会等名 統計数理研究所 リスク解析戦略研究センターシンポジウム (招待講演)
4. 発表年 2022年

1. 発表者名 小池祐太
2. 発表標題 ティックデータのフィルタリング: Daily TAQ データを例にして
3. 学会等名 探索的ビッグデータ解析と再現可能研究 (WS-EBDA-RR-2022) (招待講演)
4. 発表年 2022年

1. 発表者名 Yuta Koike
2. 発表標題 Central limit theorems in high-dimensions: Recent developments
3. 学会等名 Risk and Statistics, 3rd Tohoku-ISM-UUIm Joint Workshop (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 小池祐太
2. 発表標題 高次元データに対する正規近似理論
3. 学会等名 第25回情報論的学習理論ワークショップ (IBIS2022) (招待講演)
4. 発表年 2022年

1. 発表者名 Yuta Koike
2. 発表標題 Did the introduction of ETF Market Making Incentive Scheme affect lead-lag relationships in the Tokyo Stock Exchange?
3. 学会等名 33rd (EC)^2 Conference (国際学会)
4. 発表年 2022年

1. 発表者名 Yuta Koike
2. 発表標題 High-dimensional CLT with general covariance structure
3. 学会等名 15th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2022) (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 小池祐太
2. 発表標題 対数凹な独立同分布確率ベクトルの和に対する高次元中心極限定理
3. 学会等名 日本数学会2023年度年会
4. 発表年 2023年

1. 発表者名 Akihiro Oga, Yuta Koike
2. 発表標題 Drift estimation for a multi-dimensional diffusion process using deep neural networks
3. 学会等名 The 4th International Conference on Econometrics and Statistics (EcoSta 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 小池祐太
2. 発表標題 De-biased graphical Lasso for high-frequency data
3. 学会等名 統計数理研究所 リスク解析戦略研究センター 第8回金融シンポジウム「金融が直面する新環境への対応と方法論 III」(招待講演)
4. 発表年 2021年

1. 発表者名 Yuta Koike
2. 発表標題 Gaussian approximation for high-dimensional data: Recent progress
3. 学会等名 Maths & Stats Colloquium Series (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Yuta Koike
2. 発表標題 Central limit theorems in high-dimensions: Recent developments
3. 学会等名 15th International Conference on Computational and Financial Econometrics (CFE 2021) (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 小池祐太
2. 発表標題 高次元データに対する正規近似: 最近の進展
3. 学会等名 多様な高次元モデルの理論と方法論: 最前線の動向 (招待講演)
4. 発表年 2022年

1. 発表者名 Xiao Fang, Yuta Koike
2. 発表標題 Gaussian approximation to high-dimensional Wishart matrices under a moment assumption
3. 学会等名 日本数学会2021年度秋季総合分科会
4. 発表年 2021年

1. 発表者名 小池祐太
2. 発表標題 実現共分散行列の高次元漸近混合正規性
3. 学会等名 日本数学会2022年度年会
4. 発表年 2022年

1. 発表者名 小池祐太
2. 発表標題 高次元中心極限定理の最近の展開
3. 学会等名 日本数学会2020年度秋季総合分科会（招待講演）
4. 発表年 2020年

1. 発表者名 小池祐太
2. 発表標題 Homogeneous sumに対する高次元中心極限定理
3. 学会等名 日本数学会2021年度年会
4. 発表年 2021年

1. 発表者名 Yuta Koike
2. 発表標題 Asymptotic mixed normality of realized covariance in high-dimensions
3. 学会等名 The 3rd KAFE-JAFEE International Conference on Financial Engineering（招待講演）（国際学会）
4. 発表年 2019年

1. 発表者名 小池祐太
2. 発表標題 高次元共分散行列推定に基づく最小分散ポートフォリオのパフォーマンス比較
3. 学会等名 科研費 計量ファイナンス研究集会（招待講演）
4. 発表年 2019年

1. 発表者名 Yuta Koike
2. 発表標題 Asymptotic mixed normality of realized covariance in high-dimensions
3. 学会等名 The 62nd ISI World Statistics Congress 2019 (ISI-WSC 2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Yuta Koike
2. 発表標題 De-biased graphical Lasso for high-frequency data
3. 学会等名 CMStatistics 2019 (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Yuta Koike
2. 発表標題 Asymptotic mixed normality of realized covariance in high-dimensions
3. 学会等名 The 11th ICOSA International Conference (招待講演) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔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0件

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
---------	---------