

科学研究費助成事業 研究成果報告書

平成 30 年 5 月 21 日現在

機関番号：34504

研究種目：基盤研究(A) (一般)

研究期間：2013～2017

課題番号：25245035

研究課題名(和文)高次元データのためのベイズ計量分析に関する研究

研究課題名(英文)Bayesian econometric analysis of high-dimensional data

研究代表者

古澄 英男(KOZUMI, Hideo)

関西学院大学・経済学部・教授

研究者番号：10261273

交付決定額(研究期間全体)：(直接経費) 18,300,000円

研究成果の概要(和文)：本研究では、高次元の経済データを分析するために必要な新たなベイズ計量モデルの開発を行った。具体的には、共分散回帰モデルにもとづく計量モデルを構築し、さらに、確率的探索法や縮小型事前分布を用いた変数選択法の開発を行った。提案する計量モデルを効率的に推定するため、シミュレーションにもとづく推定方法の開発もあわせて行った。数値実験やファイナンス・マーケティングに関する実証分析を行い、提案する方法の有用性について検討した。

研究成果の概要(英文)：In this research project, we have developed new Bayesian econometric models for analyzing high-dimensional economic data. Specifically, we have developed Bayesian models based on covariance regression and proposed variable selection methods by applying the stochastic variable selection and shrinkage priors. For the proposed models, we have also developed efficient simulation-based methods. We have investigated the performance of the proposed approaches by conducting simulation studies and real data analysis.

研究分野：経済統計学

キーワード：ベイズ統計学 マルコフ連鎖モンテカルロ法 高次元データ

1. 研究開始当初の背景

(1) コンピュータの処理速度の向上および情報処理技術の進展によって、データの蓄積が急速に進んでいる。近年利用可能なデータの特徴として、データが持つ次元の大きさを挙げることができる。例えば、消費者行動分析で用いられるデータでは、消費者の属性や購入した商品の属性など数多くの項目について記録されている。こうしたデータの次元は、100以上であったり、ときとして1000や10000のオーダーであったりする。また、データの次元がデータ数を超えることも珍しくない。

(2) 複雑な経済現象を明らかにするために、こうした高次元のデータを解析することは一つの有効な接近方法である。しかし、従来の計量手法だけでは、高次元のデータを適切に分析し、そこから有益な情報を引き出すことは困難であった。また、高次元データ固有の問題も存在していた。例えば、100変数の共分散行列を考えたとき、パラメータ数は約5000と膨大になる。そのため、パラメータを正確に推定することは難しく、また共分散行列の構造を解釈することも容易ではない。こうした状況や問題を解決するために、高次元データに対する新たな計量モデルを構築し、さらにそれらに対して効率的な推定方法を開発することが必要とされていた。

(3) 高次元データに対する統計手法については、すでに経済学以外の分野(例えば、生物統計、遺伝子解析、画像解析など)において様々な取り組みが始められていた。しかし、標本選別、内生性の問題、打ち切りデータ、質的データといった計量経済学において重要な問題を扱うために必要とされる計量モデルや統計的分析方法についてはほとんどなかった。こうした制約のため、経済学分野では高次元データを用いた応用研究は多くなく、十分な実証結果の蓄積には至っていない状況であった。そこで、高次元データに対する新たな計量分析の枠組みを構築し、実際の経済問題への応用を行うことによって、高次元データの有用性を示すことも重要な課題の一つであった。

2. 研究の目的

本研究の主たる目的は、次の三点にまとめることができる。

(1) 高次元データに対するベイズ計量モデルの開発

本研究では、計量分析で最もよく用いられている回帰分析を念頭に置き、高次元データとして、(a) 被説明変数が高次元の場合、(b) 説明変数が高次元の場合を取り上げ、それら进行分析するためのベイズ計量モデルの開発を行う。

(2) 計量モデルに対する統計的分析方法の開発

本研究で提案する計量モデルでは潜在変数を利用するため、尤度関数などが複雑で解析的に解けないことが予想される。そこで、シミュレーションによる推定方法(主としてマルコフ連鎖モンテカルロ法)の開発を目指す。研究代表者および研究分担者は、これまでシミュレーションによる推定方法について研究を行ってきており、いくつかの研究成果を得ている。これらの研究成果を一層発展させることにより、既存の方法よりも効率かつ簡便な推定方法を提案する。

(3) 現実の経済問題への応用

本研究で提案する計量モデルとその統計的分析方法の有用性を示すため、実際の経済データを用いた分析を行う。主たる応用分野としては、ファイナンスと消費者行動分析を考えることにする。また、既存のモデルや推定方法との比較を行うことにより、本研究の接近方法の有用性を検証する。

3. 研究の方法

(1) 本研究の遂行に必要なデータの収集を行う。また、必要に応じて収集したデータを整理・加工する。

(2) 高次元データを分析するための新たな計量モデルならびにシミュレーションによる推定方法の開発を行うため、既存研究においてどのようなモデルや推定方法が提案されてきたかを精査し、その問題点を明らかにする。また関連する実証研究についても調べ、どのような研究成果が得られているのかを整理する。

(3) 高次元データに対する計量モデルを開発する。具体的には以下のとおりである。

被説明変数が高次元である場合、変数間の相関構造の理解が重要となる。しかし、高次元データの共分散行列はパラメータ数が多く、推定精度や解釈の難しさといった問題がある。本研究では、共分散回帰とよばれる方法に着目し、ファクターモデルの構造を取り入れることによって、柔軟でかつパラメータ数の少ないベイズ計量モデルの構築を行う。

説明変数が高次元である場合、それらの選択が重要な課題の一つである。この変数選択の問題に関しては、確率的探索法にもとづく新たなベイズ計量モデルを作成する。さらに、変数間の類似性を考慮したモデル、説明変数の一部が内生変数であるモデルへの拡張を行う。

状態空間モデルを利用することで、提案するモデルに時系列構造を取り入れ、時系列データを分析するためのモデルへと拡

張を行う。また、階層ベイズモデルを考えることによって、質的データやグループデータを分析できるように拡張する。

(4) 本研究で提案するモデルに対して、シミュレーションによる推定方法を開発する。具体的には、マルコフ連鎖モンテカルロ法の適用を考える。さらに、より効率的な推定のために粒子フィルター法についても検討を行う。

(5) 本研究で得られた計量モデルやその推定方法を数値実験や実データへの応用を通して検証する。

これらの研究を遂行するにあたり、研究代表者が研究全体を統括し、各課題に対して研究代表者および研究分担者が役割分担をして研究を進める。

4. 研究成果

本研究の主な研究成果は、以下の通りである。

(1) 被説明変数が高次元である場合については、「3. 研究の方法」で述べたように共分散行列のパラメータ数が多くなり、その解釈が難しいという問題があった。この問題を解決するため、本研究では変量効果を用いた定式化を採用し、共分散回帰モデルにファクターモデルの構造を導入することによって新たな計量モデルの開発を行った。また、このよう接近方法により、従来よりも柔軟かつパラメータ数の少ないベイズ計量モデルを構築することができることを示した。本研究で提案する計量モデルの特徴として、パラメータの解釈が容易である、時系列モデルへの拡張が容易である、すなわち、ファクターモデルを動学的ファクターモデルに変更するだけで、容易に時系列構造をモデルに取り入れることができるという点を挙げることができる。

次に、提案する計量モデルに対してシミュレーション法による推定方法の開発を行った。ここでは、マルコフ連鎖モンテカルロ法の一つであるギブス・サンプラーによって推定できることを明らかにした。ギブス・サンプラーは実装が簡単であり調整が不要なアルゴリズムであるため、本研究で提案する推定方法は実用上非常に有用であると考えられる。さらに、ここでの結果を応用することにより、EM アルゴリズムとよばれる方法でパラメータの最尤推定値を得ることができることも示した。

本研究で提案する推定方法の有用性を調べるためにいくつかの数値実験を行い、提案する推定方法が既存の方法よりも効率的であることを確認した。さらに、ファイナンスとマーケティングに関する実証分析などを行い、これまでの分析では分からない知見を

得ることができた。

(2) (1)では被説明変数の共分散行列を直接モデル化していた。本研究ではこれとは異なる接近方法として、共分散行列のコレスキー分解に対してモデリングを行う方法についても研究を進めた。この方法の利点は、推定すべきパラメータ数が少なくなる、共分散行列の正値定符号性が常に保たれる、相関係数に対応するパラメータのとりうる値に制約がないことである。特に ρ はシミュレーションによる推定方法の開発においては重要であり、本研究ではメトロポリス・ヘイスティング法とギブス・サンプリング法とを組み合わせた推定方法の開発を行った。また、より効率的な推定を行うため、利用可能な追加的情報を取り込むなどの工夫も行った。数値実験や株価収益率を使った実証分析を行い、モデルの当てはまりや予測の精度の観点から、提案するモデルが既存モデルよりも優れていることを示した。

(3) 説明変数が高次元である場合については、通常の線形回帰モデルを基本として、重要な説明変数を選択するための計量手法の開発を行った。具体的には、各回帰係数に対して重要な変数に対応する事前分布と重要ではない変数に対応する事前分布の二つを設定し、さらに説明変数の重要性を表す二値潜在変数を導入することによってベイズ階層モデルの構築を行った。本研究で提案するモデルでは、潜在変数が1(0)であれば対応する説明変数は重要である(重要ではない)ことを表しており、モデルや推定結果の解釈が容易である。また、実行が容易なギブス・サンプリングによってモデルのパラメータを推定することができることも示した。本研究で提案する方法がうまく機能するかどうかを確認するため、数値実験や実データによる実証分析を行った。その結果、説明変数がかかり多い場合でも、適切に説明変数を選択できることを示した。

(4) (3)の方法を用いることにより、必要な説明変数を適切に選択できることが分かったが、それと同時にギブス・サンプリングの収束が遅いことも明らかとなった。そこで、(3)に代わる方法として、縮小型事前分布を用いた変数選択について研究を行った。具体的には、回帰係数の事前分布としてラプラス事前分布と馬蹄事前分布を用いることを考えた。前者のラプラス事前分布はLASSOに対応した事前分布である。

これらの事前分布を直接利用したのでは、パラメータの推定が困難である。そこで本研究では、各事前分布を正規分布にもとづく混合分布の形に書き直すことによってこの問題の解決を試みた。この接近方法により、いずれの事前分布を用いても実行が容易なギブス・サンプリングによって推定することが

できることを示した。本研究で提案する方法の有用性を調べるため、ここでも数値実験や実データを用いた分析を行い、適切に説明変数が選択されること、またギブス・サンプリングの収束が改善されていることを確認した。

(5) 前述の(1)～(4)の研究では、被説明変数は正規分布にしたがうことが仮定されていた。「1. 研究開始当初の背景」で述べたように、経済学においては様々な種類のデータを分析する必要がある。そこで本研究では、(1)～(4)の研究成果を、分位点回帰モデル、空間モデル、質的データやグループデータを分析するためのモデルなどへ拡張を行った。これと同時に、本研究で提案するモデルを推定するために必要なシミュレーションによる推定方法の開発も行った。いずれの拡張においても、既存の方法よりも予測精度が改善されることなどを示した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計57件)

- [1] Miyawaki Koji, Omori Yasuhiro and Hibiki Akira (2018). "A discrete/continuous choice model on a nonconvex budget set," *Econometric Reviews*, Vol.37, 89-113, 査読有.
DOI:10.1080/07474938.2015.1032166
- [2] Sugawara Shinya and Omori Yasuhiro (2017). "An econometric analysis of insurance markets with separate identification for moral hazard and selection," *Computational Economics*, Vol.50, 473-502, 査読有.
DOI:10.1007/s10614-016-9594-z
- [3] Shirota Shinichiro, Omori Yasuhiro, Hedibert F. Lopes and Haixiang Piao (2017). "Cholesky realized stochastic volatility model," *Econometrics and Statistics*, Vol.3, 34-59, 査読有.
DOI:10.1016/j.ecosta.2016.08.003
- [4] Terui Nobuhiko and Li Yinxing (2017). "Measuring large scale market responses from aggregated sales regression model for high dimensional sparse data," *Discussion Paper of DSSR Graduate School of Economics and Management, Tohoku University*, Vol.63, 1-27, 査読無.
- [5] Usui Takehiro, Chikasada Mitsuko and Kakamu Kazuhiko (2017). "Does garbage pricing increase the immoral disposal of household waste?," *Applied Economics*, Vol.49, 3829-3840, 査読有.
DOI:10.1080/00036846.2016.1270414
- [6] Tsunehiro Ishihara, Yasuhiro Omori and Manabu Asai (2016). "Matrix exponential stochastic volatility with cross leverage," *Computational Statistics and Data Analysis*, Vol.100, 331-350, 査読有.
DOI:10.1016/j.csda.2014.10.012
- [7] Makoto Takahashi, Toshiaki Watanabe and Yasuhiro Omori (2016). "Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution," *International Journal of Forecasting*, Vol.32, 437-457, 査読有.
DOI:10.1016/j.ijforecast.2015.07.005
- [8] Kazuhiko Kakamu (2016). "Simulation studies comparing Dagum and Singh-Maddala income distributions," *Computational Economics*, Vol.48, 593-605, 査読有.
DOI:10.1007/s10614-015-9538-z
- [9] Miyawaki Koji, Omori Yasuhiro and Hibiki Akira (2016). "Exact estimation of demand functions under block rate pricing," *Econometric Reviews*, Vol.35, 311-343, 査読有.
DOI:10.1080/07474938.2013.806857
- [10] 照井伸彦 (2015). "マーケティング・アナリティクス-ビッグデータとスモールデータの統計モデリング-, " *応用統計学*, 44巻, 5-13, 査読有.
DOI:10.5023/jappstat.44.3
- [11] Haruhisa Nishino and Kazuhiko Kakamu (2015). "A random walk stochastic volatility model for income inequality," *Japan and the World Economy*, Vol.36, 21-28, 査読有.
DOI:10.1016/j.japwor.2015.06.003
- [12] Kobayashi Genya, Kakamu Kazuhiko, Sato Eisaku and Kozumi Hideo (2014). "An integrated purchase model using Gaussian copula," *Behaviormetrika*, Vol.41, 147-167, 査読有.
DOI:10.2333/bhmk.41.147
- [13] Terui Nobuhiko and Ban, M. (2014). "Multivariate structural time series models with hierarchical structure for over-dispersed discrete outcome," *Journal of Forecasting*, Vol.33, 376-390, 査読有.
DOI:10.1002/for.2301
- [14] Kakamu Kazuhiko, Yunoue Hideo and Kuramoto Takashi (2014). "Spatial patterns of flypaper effects for local expenditure by policy objective

in Japan: A Bayesian approach,"
Economic Modelling, Vol. 37, 500-506,
査読有.

DOI:10.1016/j.econmod.2013.11.028

- [15] Takahashi Makoto, Omori Yasuhiro and Watanabe Toshiaki (2013). "News impact curve for stochastic volatility models," Economic Letters, Vol.120, 130-134, 査読有.
DOI:10.1016/j.econlet.2013.03.001

〔学会発表〕(計94件)

- [1] Yamauchi Yuta, Omori Yasuhiro, "Multivariate realized stochastic volatility with dynamic pairwise correlations," 1st International Conference on Econometrics and Statistics (2017).
- [2] Terui Nobuhiko, "Measuring large-scale market responses from aggregated sales: Regression for high-dimensional sparse data," INFORMS Annual Conference 2017 (2017)
- [3] Terui Nobuhiko, "A threshold model for discontinuous preference change and satiation," 38th ISMS Marketing Science Conference (2016).
- [4] Genya Kobayashi and Kazuhiko Kakamu, "Approximate Bayesian computation for Lorenz curves from grouped data," ISBA 2016 World Meeting (2016).
- [5] Yamauchi Yuta and Omori Yasuhiro, "Multivariate realized stochastic volatility with dynamic pairwise correlations," 9th International Conference on Computational and Financial Econometrics (2015).
- [6] Kakamu Kazuhiko, "Direct and indirect effects on road productivity in Japan," 24th International Workshop on Matrices and Statistics (2015).
- [7] Ishihara Tsunehiro and Omori Yasuhiro, "A dynamic factor stochastic volatility model with leverage effect and its application," 7th International Conference on Computational and Financial Econometrics (2013).

〔図書〕(計5件)

- [1] 大森 裕浩, 新世社:サイエンス社, コア・テキスト計量経済学, 2017, 362.
- [2] 古澄 英男, 朝倉書店, ベイズ計算統計学, 2015, 208.

6. 研究組織

(1)研究代表者

古澄 英男 (KOZUMI, Hideo)
関西学院大学・経済学部・教授
研究者番号: 10261273

(2)研究分担者

大森 裕浩 (OMORI, Yasuhiro)
東京大学・大学院経済学研究科(経済学部)・教授
研究者番号: 60251188

照井 伸彦 (TERUI, Nobuhiko)
東北大学・経済学研究科・教授
研究者番号: 50207495

各務 和彦 (KAKAMU, Kazuhiko)
神戸大学・経営学研究科・准教授
研究者番号: 00456005

宮脇 幸治 (MIYAWAKI, Koji)
関西学院大学・経済学部・准教授
研究者番号: 40550249

石原 庸博 (ISHIHARA, Tsunehiro)
大阪大学・金融保険センター・講師
研究者番号: 60609072
(平成28年度より研究協力者)

(3)連携研究者

(4)研究協力者