

論文審査の要旨及び担当者

No.1

報告番号	甲 第 号	氏 名	マクリン 謙一郎
論文審査担当者	主 査	：星野 崇宏 (慶應義塾大学経済学部教授 博士(経済学))	
	副 査	：中妻 照雄 (慶應義塾大学経済学部教授 Ph.D.)	
		：廣瀬 康生 (慶應義塾大学経済学部教授 Ph.D.)	
	面接担当	：伊藤 幹夫 (慶應義塾大学経済学部教授 博士(経済学))	
		：櫻川 昌哉 (慶應義塾大学経済学部教授 博士(経済学))	
(論文審査の要旨)			
論文題名 : Essays on Bayesian Econometrics for Big Data			
1. 論文の概要			
<p>本論文の主たる目的は、ビッグデータのベイズ計量経済分析のためのモデリングと数値計算の新たな手法を提案することである。近年の数値計算でのスケーラビリティと言われるデータの大規模化に対する対応の必要性と大規模データの利用可能性の増大は、遺伝学、マクロ経済学、金融、マーケティングなどの数多くの分野における大規模データの応用を拡大するとともに、大規模データすなわち「ビッグデータ」の分析手法の研究への関心も高めつつある。ここでいう「ビッグデータ」とは、単に大量のデータという意味ではなく、従来の統計的および計量経済学的方法（古典的な最小二乗法など）では、推論、予測および意思決定に使用できる有益な情報を得られないような状況を指す。具体的には、(1) 推定すべきパラメータの数 p がデータセットの観測値の数 n を上回る、(2) p が n を下回る状況であっても p が十分 n に近い、という状況を想定している。このような場合、推定値そのものを計算できなかつたり、計算できても数値が不安定になったりすることが知られている。すると、実証分析における推論、予測および意思決定の結果が歪められるなどの問題が発生する危険性が出てくる。</p> <p>このような取り扱いの難しさにもかかわらずビッグデータが使われる1つの理由は、集計量では捉えられない（あるいは集計によって情報の損失が生じてしまう）場合には、可能であれば集計される前のデータを分析する方がより有用な情報を得ることができるからである（これはある意味で後述する第2章の実証結果に通じる話である）。経済学の実証研究でもマイクロデータの利用が普及して久しいが、近年のビッグデータ分析の隆盛はこの流れを汲むものである。もう1つの理由として、できるだけ多くの入手可能な変数を組み合わせて利用することで、予測の精度を高められるという期待があることも挙げられる（これは第5章の内容に関係するテーマでもある）。</p> <p>そのため機械学習や計量経済学以外の統計学の分野でも、この意味でのビッグデータの分析手法が盛んに研究されてきた。しかし、特に経済データを念頭に置いたモデリングの研究は十分である</p>			

とはいえない。この状況を鑑み、本論文は、経済のビッグデータを対象とするベイズ的アプローチによるモデリングの手法とそのため数値計算技法の提案を目指すものである。

本論文は6つの章からなる。第1章は論文の導入部分と全体のサマリーに過ぎないので、以下では第2章から第5章の内容について簡潔に紹介する。

第2章では、日米の個別銘柄（日経平均とS&P500の採用銘柄）の株式リターンという大規模データを利用して、ボラティリティにおけるレバレッジ効果の非線形性について検証を行っている。ここでいうところの「レバレッジ効果」とは、今日の株式リターンのボラティリティ（分散）と前日の株式リターンの間の負の相関を指す。この負の相関は、株価が下落することで株式の時価総額に対する負債の比率（レバレッジ）が増加し、結果として企業の破綻リスクが上昇することから生じると解釈される。株価指数のリターンについては、ほぼ全ての株式市場で両者の間に負の相関が確認されている。しかしながら個別銘柄については強くレバレッジ効果が確認されないことがわかってきた。この「本来の解釈上はレバレッジ効果ははっきりと出るべきである個別銘柄では見られない代わりに多くの銘柄からなるポートフォリオであるところの株価指数ではレバレッジ効果が見られる」という逆説的な現象を解明するため、本章では個別銘柄のレバレッジ効果は非線形であるという仮説を立て、非線形レバレッジ確率的ボラティリティモデルを粒子学習で推定することで検証を行った。その結果、日経平均とS&P500の採用銘柄の多くに対してレバレッジ効果の非線形性を確認することに成功した。

第3章では、第2章で使用した粒子学習をグラフィカル・プロセッシング・ユニット（GPU）の並列コンピューティング・アーキテクチャ上で効率的に並列実行するためのアルゴリズムの提案とその検証を行っている。粒子学習（および粒子フィルター）は、アルゴリズムの特性上、比較的並列化が容易であるとされてきた。しかし、唯一のボトルネックが粒子の再抽出の過程で、そこを並列処理できれば粒子学習の完全並列化が可能となる。本章では、GPU上で大規模な並列処理を行うことを前提にした斬新な並列計算アルゴリズムを開発し、これによって粒子学習を完全並列化することで従来の手法よりも飛躍的に計算時間を短縮できることを示した。

第4章では、大規模なマクロ経済パネルデータを分析するためのスパース動的因子モデルの効率的な推定法を開発している。動的因子モデルは、観測される変数の背後に観測されない複数の因子が存在し、それと連動することで個々の変数が変動するという構造になっている。動的因子モデルはマクロ経済変数間の関係性を分析するための強力なツールであるが、複数の観測されない因子を識別するためには何らかの制約を因子間に課す必要がある。本論文では、アドホックな制約の代わりに因子負荷量（因子にかかる係数）にスパース性を導入することで識別問題を回避するとともに、このスパース動的因子モデルをEMアルゴリズムで安定して推定する方法を提案した。さらに実証研究として本章の提案手法を米国のマクロ経済パネルデータに適用し、世界金融危機の前後でマクロ変数間関係性に構造変化が生じたことを発見した。

第5章では、McAlinn and West (2018) で提案された予測統合の方法を発展させて大規模予測モデルを構築するとともに予測を踏まえた意思決定を行うための新しいフレームワークを提案している。複数の予測モデルを統合する方法としてはベイズ的モデル平均(BMA)などが知られている。しかし、どうしても時期によって個々のモデルの予測精度は変化するため、予測統合の際の重みを各モデルの予測の実績に応じて動的に変化させた方が望ましいと考えられる。これを自動的に行う手法は McAlinn and West (2018) で提案されている。本章では、これを発展させ、意思決定に必要な解釈可能性を保ちながら、多数の共変量をうまくモデル内で統合するフレームワークを提案している。そして、米国のインフレ率や産業別株式超過リターンの予測に適用し、他の方法 (BMA や LASSO など) に対する提案手法の優位性を示した。

2. 論文の評価

近年の各方面における大規模データの利用の拡大は、「ビッグデータ分析」が単なるバズワードの領域を超えて広くデータサイエンス全体の中で定着してきたことを示唆するといえるだろう。しかし、工学の領域から入ってきた機械学習・深層学習などの手法の経済分析への応用には、「理論なき計測」とのそしりを受けかねない使われ方が散見されるのも事実である。この文脈において、本論文の研究成果は、経済データの特性と統計モデルの経済学的解釈の可能性を考慮してしっかりと実証分析を行うことを目指す計量経済学の系譜の中に位置付けられるものである。提案されている手法も最新の研究成果を反映したものばかりであり、先行研究にも見られない新規性を含んでいる。

しかしながら本論文に問題点がないわけではない。例えば第2章の結果そのものは興味深いが実用性に欠けるのではないかという指摘があった。また、タイトルは「ビッグデータのためのベイズ計量経済学」であるが、必ずしもビッグデータとは関係のない研究 (例えば第3章) も入っているのは一貫性にかけるのではないかという声も出た。但し、これらの指摘された課題は本論文の学術的な価値や貢献を損なうものではない。第2章は **fact finding** に重きを置く内容ではあるものの分析に使用しているモデルは予測やリスクヘッジへの利用が可能なものであるし、第3章のアルゴリズムはスケーラビリティの点でビッグデータへも十分応用できるポテンシャルを有している。

以上の所見から、本論文の内容はビッグデータの計量経済分析での利用の促進のための手法の提案という点において極めて新規性のある研究成果であり、審査委員会は全員一致で本論文が博士 (経済学) の学位を授与するにふさわしい内容のものであると判断した。