

令和 2 年 6 月 11 日現在

機関番号：32601

研究種目：若手研究(B)

研究期間：2016～2019

課題番号：16K17100

研究課題名(和文) 超高次元データにおける因子推定とDI予測

研究課題名(英文) Econometric Analysis on Ultra-high dimensional Factor models and DI forecasts

研究代表者

田中 晋矢 (Tanaka, Shinya)

青山学院大学・経済学部・准教授

研究者番号：80727149

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：本研究課題では「超高次元経済時系列データの適用を想定した超高次元線形回帰モデルおよび超高次元パネルベクトル自己回帰(PVAR)モデルにおける係数および予測値の罰則付回帰推定量の理論特性についての理論的考察」ならびに「罰則付回帰のマクロ経済予測への応用」に関する研究を行い、その結果として超高次元統計モデルは時間従属性の強い経済時系列データに適用した場合でもうまく機能すること、ならびに経済予測をはじめとするマクロ計量経済分析のための有用な手段となりうることを明らかにした。

研究成果の学術的意義や社会的意義

近年では膨大な情報・サイズを有する"超"高次元データ(所謂ビッグデータ)が工学、医学、情報通信等の分野において積極的に利用されていることは周知のとおりである。その一方で経済学分野での超高次元データの利用は特に本研究課題を開始した2016年当時において世界的に見ても非常に少ないという状況にあった。その大きな理由として標準的な高次元統計解析手法を経済データに適用した場合の推定量の"ふるまい"について未解明である部分が多かったことが挙げられよう。本研究課題の研究成果は当該問題の解決に大きく寄与し、さらに実際の経済データを用いて超高次元経済データを用いた計量経済分析が実証的にも有用であることを示した。

研究成果の概要(英文)：In this research project, we focused on both theoretical and empirical studies concerning ultra-high dimensional macroeconomic data: we investigated theoretical properties of the penalized regression estimators of the ultra-high dimensional linear regression model with time-dependent regressors and the panel VAR model, as well as empirical application of the penalized regression to macroeconomic forecasting. Through our contributions, we showed the ultra-high dimensional statistical models worked well both theoretically and empirically in macroeconomic time series data.

研究分野：計量経済学

キーワード：超高次元経済データ分析 罰則付回帰 Lasso 経済予測

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

- (1) 2000年代以降のITの著しい進歩を受け、特に近年においてはありとあらゆるデータの収集やアクセスが可能となってきた。同時にこのような超高次元データ(所謂ビッグデータ)を扱うことのできる高次元統計モデル・推定手法に関する理論研究が目まぐるしい勢いで発展してきており、近年では理系分野、とりわけ画像工学・医学・情報通信等の分野において超高次元データ分析が広く活用されている。
- (2) その一方で超高次元経済データの計量経済分析への活用は限定的である。その主たる要因として経済データに即した高次元統計モデル・推定手法に関する理論研究が発展途上であることが挙げられる。経済データの大きな特徴として時点(および横断面)に関する従属性を有することが挙げられるが、データが従属性を持つ場合に標準的な超高次元統計モデルおよび推定法がうまく機能するかという部分で未知数な部分が多いのが現状といえる。
- (3) しかしながら、例えば日経 NEEDS あるいはインテージ社全国小売店パネル調査のような超高次元マクロ経済データセットを用いて経済予測を行うなど、豊富な情報量を有する超高次元経済データを計量経済分析に活用することは多大なメリットがあることが期待される。

2. 研究の目的

本研究課題では1.における研究背景から生ずる問題意識のもと、

- (1) 高次元統計モデル、特に経済予測に親和性の高い高次元因子モデルに着目し、当該モデルへ従属性のある超高次元経済データを適用した場合の推定量の特性に関する理論的考察を行うこと
- (2) (1)で得られた理論研究の成果に基づき超高次元経済データを用いたマクロ経済予測、特に因子モデルに基づいたマクロ経済予測(DI予測)に関する実証研究を行い、超高次元経済データをマクロ経済予測に用いることでどれだけ予測精度の向上が見られるかを検証すること

を目的とする。

3. 研究の方法

2.における研究の目的を達成するため、本研究課題では次のような段階別に研究テーマを設定しつつ研究に着手していく：

- (1) 高次元統計モデルのうち、経済データ分析を行う際に最も標準的な統計モデルである線形回帰モデルの素直な拡張であるスパース性を持つ超高次元線形回帰モデル(罰則付線形回帰モデル)に着目し、データに従属性が存在する場合の回帰係数および予測値に関する罰則付回帰推定量の理論的性質を明らかにする。
- (2) マクロ経済時系列データ分析を行う際に頻りに用いられるベクトル自己回帰(VAR)モデルに着目し、VAR係数および予測値に関する罰則付回帰推定量の理論的性質を明らかにする。
- (3) (1)および(2)の研究成果を踏まえ、超高次元線形回帰モデル・超高次元ベクトル自己回帰モデルに基づくマクロ経済予測を主とした実証研究を行う。
- (4) (1)および(2)の研究成果を踏まえ、超高次元因子モデルにおける因子推定およびDI予測に関する理論研究に着手する。
- (5) (4)の結果を踏まえ、超高次元因子モデルに基づくマクロ経済予測を主とした実証研究を行う。

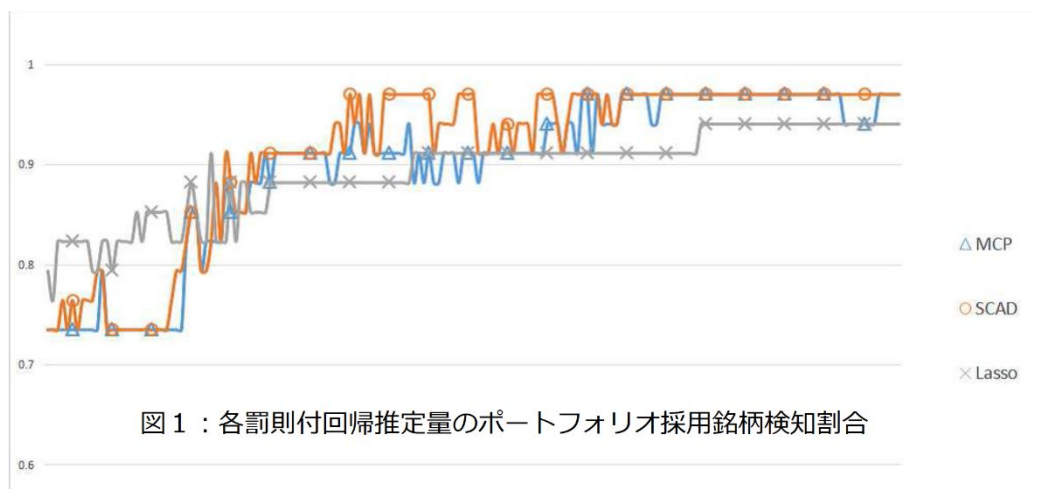
4. 研究成果

3.における研究の方法により、本研究課題の研究期間において得られた研究成果は次のとおりである：

- (1) データが従属性を持つ場合における超高次元線形回帰モデルの罰則付回帰推定量の理論特性を明らかにした。より具体的には罰則付回帰推定量のうち最も代表的と考えられるL1(Lasso)罰則ならびに非凸罰則(SCAD, MCP)により求められるものに焦点をあて、回帰係数および予測値の推定誤差に関する不等式(オラクル不等式)を導出した。その結果とし

て Lasso, MCP, SCAD に基づき超高次元経済データを用いた経済予測を行うことは妥当性を持つことを明らかにした。

- (2) (1)を根拠とし、罰則付線形回帰モデルに基づく米国の大規模マクロ経済月次データセット FRED-MD を用いた米国 GDP の超短期予測 (now-casting) を試みた。ここで GDP が四半期系列である一方で予測に用いる(説明)変数は月次系列であり、データ周期が異なるという問題が生じる。そのため MIDAS (Mixed Data Sampling) という手法を罰則付線形回帰モデルに応用し予測を行っているが、MIDAS を導入することにより説明変数の総数が 1000 を超え、超高次元経済データセットに対応できる罰則付線形回帰モデルの採用が望ましい状況となる。分析の結果、罰則付線形回帰モデルに基づく GDP 予測値はその他の代表的な予測モデルに基づく予測値と比較して良好な予測精度を有することを明らかにした。
- (3) (1)に関連する実証研究として(2)に加え罰則付線形回帰モデルのファイナンスへの応用も試みた。非凸罰則由来の罰則付回帰推定量である SCAD および MCP は特定の条件を満たす場合に変数選択に関する一貫性を持つことが知られているが、仮に経済(金融)時系列データにおいても同様に変数選択の一貫性が成立するのであれば各ファンドのポートフォリオの運用成績から非公開である各ファンドのポートフォリオ構成銘柄を検知できるのではないかと考え、米国のニューヨーク証券取引所に上場している 1853 銘柄から構成される週次株価データを用いて検証を行った。前述のとおり各ファンドのポートフォリオ構成銘柄は非公開であるため株価データからランダムに 34 銘柄を抽出して仮想ポートフォリオを構成し、これを各罰則付回帰推定量により検知できるかを検証した結果が図 1 である。図 1 は横軸がサンプルサイズの大きさ、縦軸が検知割合(最上部 1, 最下部 0.6)を表すが、非凸罰則由来の SCAD および MCP は期待通りサンプルサイズが大きい場合に極めて 1 に近い割合でポートフォリオ採用銘柄を検知できることを明らかにした。



- (4) (2)で述べた通り SCAD や MCP は特定の条件を満たす場合に変数選択に関する一貫性を持つが、これに加え非ゼロ係数に関して漸近正規性を持つことが知られている(オラクル性)。しかしながらこの特定の条件が満たされない場合には SCAD や MCP に基づいて検定を行うと誤った結論を導くことになる。当該問題を解決する方法として近年注目されているのが非スパース化 Lasso 推定量であり、漸近正規性の成立に当該条件は不要である。しかしながら経済時系列のように時点従属性を持つデータに適用した場合には非スパース化 Lasso が漸近正規性を持つかどうかは未知数である。そこでデータに時点従属性が存在する場合における非スパース化 Lasso 推定量に基づく検定統計量の有限標本特性をシミュレーション分析により考察した。その結果検定の実効サイズおよび実効検出力が良好であることを明らかにし、時点従属性がある場合においても非スパース化 Lasso が漸近正規性を有することを示唆する結論を導いた。
- (5) 複数財・複数国間の相互依存関係やショックの波及過程を記述できるパネルベクトル自己回帰 (Panel VAR) モデルへ罰則付回帰の考え方を応用し、超高次元経済データを Panel VAR モデルに適用した場合の VAR 係数および予測値に関する推定・検定方法の提案を行った。具体的には、はじめに VAR 係数および予測値の Lasso 推定量の推定誤差に関する不等式 (オラクル不等式) を導出し、次に VAR 係数に関する仮説検定を可能とするために VAR 係数に関する非スパース化 Lasso 推定量の漸近正規性を示した。この結果は(4)で得られた結論に理論的確証を得たことを意味する。さらに全ての個別 VAR 係数に関する(多重)検定を正しく行うために False discovery rate (FDR) をコントロールする方法を

提案し、この方法がうまく機能することをシミュレーション分析により示した。

(1), (2), (3)の研究成果はセミナー報告を経て Uematsu and Tanaka (2019)として纏められ、世界的に定評のある査読付欧文学術雑誌である *Econometrics Journal* に掲載された。なお、Uematsu and Tanaka (2019)は掲載巻号の Editor's choice 論文に選出されたことを付記しておく。(4)の研究成果は田中(2020)として邦文学術雑誌(紀要)である『青山経済論集』に掲載された。(5)の研究成果は Summer Workshop on Economic Theory 2019 にてセミナー報告がなされたが、理論部分の増強・追加ならびに実証部分の追加を図るため大幅な見直しを行った関係で残念ながら研究期間中に学術雑誌への掲載は叶わなかった。(5)については現在できるだけ早い時期にワーキングペーパーとして公開および学術雑誌への投稿を行うべく精力的に研究を継続中である。

以上が研究期間に得られた研究成果であるが、結局のところ超高次元線形回帰モデルおよびこの応用としての経済予測に関する理論・実証研究に終始することになり、研究当初に主目的として考えていた超高次元因子モデルに関する理論・実証研究に関するまとまった研究成果を挙げることは叶わなかった。これは当初の想定以上に超高次元線形回帰モデルに関する興味深い理論的問題が残されており、超高次元線形回帰モデルと超高次元因子モデルには共通する部分が多いことからこれらの問題を先に解決するほうが社会的に意義・インパクトのある超高次元因子モデルに関する理論・実証研究を行うことができると判断して超高次元線形回帰モデルに関する研究に専念したことが原因である。残された超高次元因子モデルに関する理論・実証研究は本研究課題の研究期間によって得られた研究成果を踏まえて責任をもって研究期間終了後も引き続き継続していく。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Uematsu Yoshimasa and Tanaka Shinya	4. 巻 22
2. 論文標題 High dimensional macroeconomic forecasting and variable selection via penalized regression	5. 発行年 2019年
3. 雑誌名 The Econometrics Journal	6. 最初と最後の頁 34-56
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/ectj.12117	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 田中晋矢	4. 巻 71
2. 論文標題 経済時系列データにおける非スパース化 Lasso にもとづく検定統計量の有限標本特性 シミュレーション分析による接近	5. 発行年 2020年
3. 雑誌名 青山経済論集	6. 最初と最後の頁 41-56
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
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 -

〔学会発表〕 計4件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 田中晋矢(報告者), 植松良公, 山形孝志, 新谷元嗣
2. 発表標題 Large-Scale Panel Vector Autoregressive Models
3. 学会等名 Summer Workshop on Economic Theory (SWET) 2019
4. 発表年 2019年

1. 発表者名 田中晋矢(報告者)
2. 発表標題 Statistical Inference on high-dimensional economic data through the desparsified Lasso: Monte Carlo evidence and its possible directions
3. 学会等名 Data Science Workshop (東北大学大学院経済学研究科サービス・データ科学研究センター)
4. 発表年 2019年

1. 発表者名 田中晋矢(報告者), 植松良公
2. 発表標題 High-dimensional Macroeconomic Forecasting and Variable Selection via Penalized Regression
3. 学会等名 ICS FSファカルティセミナー(一橋大学大学院国際企業戦略研究科)
4. 発表年 2018年

1. 発表者名 田中晋矢(報告者), 植松良公
2. 発表標題 Macroeconomic Forecasting and Variable Selection with a Very Large Number of Predictors: A Penalized Regression Approach
3. 学会等名 計量経済学セミナー(京都大学経済研究所)
4. 発表年 2016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
研究協力者	植松 良公 (Uematsu Yoshimasa)		
研究協力者	山形 孝志 (Yamagata Takashi)		
研究協力者	新谷 元嗣 (Shintani Mototsugu)		

6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 協力 者	小野寺 敬 (Onodera Takashi)		