

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 13 日現在

機関番号：11301

研究種目：基盤研究(B) (一般)

研究期間：2013～2016

課題番号：25280005

研究課題名(和文)大規模で非定常な時系列・時空間データのモデル化とその推定・検定・予測法の研究

研究課題名(英文)Modelling of non-stationary spatio-temporal data with estimation, testing and forecasting

研究代表者

松田 安昌(Matsuda, Yasumasa)

東北大学・経済学研究科(研究院)・教授

研究者番号：10301590

交付決定額(研究期間全体)：(直接経費) 8,600,000円

研究成果の概要(和文)：非定常な構造をもつ大規模な時空間系列を研究対象として、そのモデル化とその推定・検定・予測法を提案し、実データを使って実証研究を行った。本研究では時空間相関を時間相関と空間相関の積で与えられるセパラブルモデルを用い、空間相関にはContinuous Autoregressive Moving Average (CARMA) モデルを、時間相関にはARMAモデルをあてはめる時空間モデルを開発した。そしてアメリカ大陸6000地点で月次で観測される降雨量のデータに応用し、大規模データに対応する高速なパラメータ推定法、時空間相関のセパラビリティの検定、大規模データに対応する高速な予測法を提案した。

研究成果の概要(英文)：We have developed a model to analyze huge spatio-temporal data set with non-stationary structures and proposed a method of parameter estimation, testing of correlation structures and forecasting with applications to empirical data mainly from environmental studies. Specifically, We have considered a separable model whose spatio-temporal correlations are given by product of spatial and temporal correlations. In the empirical studies in this research, we employed continuous autoregressive moving average (CARMA) models for spatial behaviors and traditional ARMA models for temporal ones. We applied the separable models to the US rain fall data recorded monthly at 6000 points irregularly spaced inside US continent and found proposed an efficient estimation, testing and forecasting methods that can cope with huge data set composed of several million data points.

研究分野：時系列解析、時空間統計学

キーワード：spatio-temporal data Whittle estimate spectral density kriging forecasting separable correlations irregularly spaced data

1. 研究開始当初の背景

1900年代初頭に創始された統計学は、小・中標本を前提に一樣な構造をもつ定常なデータを対象として統計理論を構築してきた。そして2000年代に入るまで定常な中標本を前提とする流れが続き、主な応用先は独立同分布な系列の分析を基本とする医学・生物学や、高々数十年程度の年次または四半期の小規模なマクロデータを分析する回帰モデルや時系列解析の方法が主となる計量経済学であった。

ところが2000年代に入って以降、急速なIT技術の進歩により、標本数が数万・数十万にのぼる大規模なデータを観測することが可能となった。大規模なデータは種々の分野で、時系列および時空間系列の形で蓄積されており、現に更新されつつある。代表的な例として、国土交通省により公開される日本全国数万地点で公示地価の年次系列や、市区町村レベルで年次で集計された疾病率の数万次元の時空間系列などがある。データが得られる分野は計量経済学、医学生物学、環境、疫学や情報技術など多様な分野にわたる。大規模なデータは今までの小規模なマクロデータではわからない精密な学問的結論を導く可能性を秘めているが、小標本を前提とした既存の統計学的手法を単純に応用することはできない。

大規模な時系列・時空間系列に対して、既存の研究では主に二種の方法が考えられてきた。その二種の違いは、共分散行列が数万次の大規模行列になり逆行列の計算が不能になってしまうという難点への対処の違いによる。第一の方法は covariance tapering 法とよばれ、ある一定数以上のラグの相関を0とおくことで共分散行列を疎(スパース)行列に帰着させる方法である(Kaufman, et al.(2008))。第二の方法は predictive process 法とよばれ、ある一定数以下のラグの相関を無視する近似を行う方法である(Banerjee, et al. (2008))。いわば両者は正反対の近似法を使って大規模行列の演算を行う。Matsuda and Yajima (2009)は後者の立場による近似法の一種と捉えることができる。

しかし既存の大規模データ研究は定常性を仮定しており、非定常性を問題にしていない。大規模なデータは時間・空間にわたって得られるが、時間・空間にわたって確率構造が定常であると仮定するのは不自然である。よって非定常性をモデル化することはより精密な分析につながる可能性が期待される。既存の研究は大規模な次元の行列演算の処理に関心を集中させてきたため、非定常性に対する扱いが疎かになったのではないかと考えている。そこで本研究の対象は非定常な構造を持つ大規模なデータとし、そのモデル化とモデルの推定・検定・予測法を正当化する理論を構築し、計量経済学、医学生物学および環境において観測される大規模なデータへの実証研究を行うことを目的とする。非定

常性を考慮することで、定常性を仮定した分析に比べて精密な学問的成果が期待される。

2. 研究の目的

非定常な構造をもつ大規模な時系列・時空間系列を研究対象として、そのモデル化とその推定・検定・予測法を提案し、実データを使って実証研究を行う。具体的には、(1)提案する統計手法の良さを理論的に正当化すること、(2)計量経済学、(3)空間疫学、(4)環境関連問題への応用を通して、3分野において学問的な知見を導くことである。

(1)の理論的正当化において、頻度論的立場における漸近理論の構築およびベイズ統計学の立場における適当な事前分布の設定と事後分布の理論的導出という2点を明らかにしたいと考えている。近年の統計学は実用性を重視して理論を軽視する傾向があるが、本研究は理論的正当化を疎かにしない。

(2)において、国土交通省による公示地価および同省のマイクロデータ「物流センサス」を分析するためのモデルを提案する。前者では地価の短期変動を明らかにすることを目指して、地価の非定常な時空間構造をモデル化し、その推定法を考案して短期の将来予測量を構成する。特に大都市と地方の地価変動の差異をモデルから明らかにする。後者では地域間の貨物輸送費の決定モデルを推定し、更にその結果を用いて高速道路整備の社会的便益を計測することを目指す。具体的には、まず運輸企業の利潤最大化、荷主の時間費用を含めた輸送費用最小化を組み合わせた一般均衡モデルから貨物輸送の均衡を導出する。その輸送運賃モデルを距離、輸送所用時間、ロットサイズ等を説明変数に用いてヘドニックアプローチに基づいて推定する。更にその結果を用いて荷主の費用関数を推定し、各荷主ごとに時間費用及びその分布の推定値を得る。そこから得られた時間費用の分布を用いて、最終的に高速道路整備による社会的便益を推定する。

(3)において、市区町村ごとに集計された年次データをもとに、疾病地図の作成、医療費の地域差、疾病集積性の検討を行う。既存研究では、時間をとめた空間データを対象としてモデル分析がなされている。この分析法を時空間データへ拡張し、非定常性をモデリングすることで既存の方法の精度をさらに向上させる。

(4)において、二種の環境データ分析を行う。ひとつは地図・画像などの複雑なマルチメディアデータとして観測される世界各地の降雨量、森林面積の時空間データ、もうひとつは日本各地の海岸で観測される波高の時空間データである。前者では、砂漠化・温暖化に社会経済的要因がどのように影響しているかをモデル分析し、森林減少と人為的な活動の関連を明らかにする。後者では、Levy過程を用いて、巨大津波が起きるリスクを統計的に解析し、そのリスクの特徴やリスクのレベルに見合わせた防災対策を提言する。

3. 研究の方法

応用グループは「計量経済学」、「疫学」、「環境問題」において実証分析を行うためのデータを収集する。この3分野を選んだ理由は、非定常で大規模なデータの分析が最も有効だと考えられる学問分野であるからである。いずれの分野のデータにも共通することは、空間、時間または空間・時間にわたる数万点におよぶ大規模な標本数を持ち、かつ、時間・空間に渡って確率構造が連続的に変動する非定常性をもつ点にある。

「計量経済学」(西山・矢島)におけるテーマは、非定常な時空間相関構造のモデリングと時空間回帰モデルへの適用である。国土交通省から公表される公示地価およびマイクロデータ「物流センサス」を対象に、過去数年間にさかのぼってデータを収集する。まず、データのもつ非定常性を検定する統計量を考案し、次に非定常性をどのようにしてモデル化するかを明らかにしていく。

「疫学」(高橋・栗原)では、時間および空間疾病集積性を明らかにすることをテーマとする。市区町村別に集計された疾病数のデータを年単位で収集する。

「環境問題」(西井・陳)におけるテーマは2つある。降雨量・森林面積と都市化の関係、および津波発生頻度を分析することである。前者では、人工衛星・自動化測定機器で日々測定されているオゾン層濃度、土壌水分、地表面温度、土地被覆状況などテラバイトオーダーの情報を収集する。後者では、津波発生が危惧されている宮城県鮎川および四国土佐清水を中心に全国の主要海岸における波高の時系列データを収集する。

理論グループは、以上の3分野で収集した大規模なデータに対して統計分析法を構築する。ここで統計分析とはモデリングとそのモデルに基づく推定・検定・予測のための理論と方法を指す。理論グループは次の3つのアプローチによる統計分析法の確立を目指す。

第一は、頻度論の立場(柿沢・生川)からのアプローチである。頻度論とは、コルモゴロフによる確率論の創始以来の古典的な立場からのモデリングである。モデルパラメータを未知の定数とみなし、尤度最大化によって推定を行い、漸近理論を構築してパラメータ推定量の漸近分布を導出して検定を行う。大規模データへの適用の仕方に困難なポイントが3つある。1)非定常性のモデル化、2)大規模データを処理しうる高速な推定法、3)漸近理論の漸近を再定義すること、の3点である。3)ではMixed Asymptoticsを適用した漸近理論の構築を目指す。本研究は時・地点に渡って観測されるため、漸近理論を既存のものとは異なる枠組みで構築しなければならない点に理論的な興味がある(Stein(1999))。

第二はベイズ統計学の立場(丸山)からのアプローチである。ベイズアプローチでは、モデルパラメータを未知の定数と捉えず、適

当な事前分布を持つものとする。データを観測した後にパラメータの事後分布に基づいて推定・検定・予測を行う方法であるが、事後分布を解析的に導出することが一般には難しいため、通常はMCMC法等による近似解法を使って近似することが多い。本研究では、事後分布を解析的に導出できるタイプのモデルと事前分布を設定することで、近似法に寄らない事後分布によるベイズ統計分析法を提案することを目指す。

第三は経験ベイズ法(松田)からのアプローチである。このアプローチは頻度論とベイズ理論の中間と捉えることができる。パラメータを未知の定数と見做さず事前分布を設定するところはベイズ理論と同じであるが、事前分布に超パラメータとよばれるパラメータを設定し、超パラメータをその周辺尤度関数を最大化することで推定する点が頻度論的である。超パラメータを使ってモデリングする階層モデルは、通常の頻度論アプローチにくらべて柔軟なモデリングを可能とするため、様々な非定常性のモデル化が期待される。ただし、漸近理論の構築が頻度論に比べて難しくなってしまうというデメリットもある。

以上の研究計画の成果を発表し、研究計画をさらに進めることを目的として、平成25年11月に東北大学に共同研究者と会議(開催責任者 松田)を開くことを予定している。会議では応用3分野からのデータ分析の実際と理論の3アプローチからのモデリングの現状を報告する。

4. 研究成果

4年間にわたる本研究の結果、時間・空間にわたる非定常なデータを対象として、「モデルの理論構築」と「実証研究」を、統計家と応用研究者が協力して行うことができた。さらに統計学の立場からは、大規模・非定常データに対する汎用性のある統計モデル分析法を確立することができ、応用研究の立場からは、既存の定常性を仮定したモデル分析では得られない新たな学問的知見を獲得することができた。以下、具体的に研究成果を詳述する。

(1) Wavelet 時空間モデルの提案

不規則に位置する大規模な時空間データをモデル化する方法をHaar waveletをもとに構成した。不規則に位置する時空間データの最大の難点はいかにして直交系を構成するかという点にある。規則的に位置する通常の時系列では、フーリエ級数で直交関数系を容易に構成することができる。本研究では、フーリエ直交系をHaar waveletを使って不規則に位置するデータの直交系を構成した。この直交系により、百万のオーダーの標本点からなる大規模な時空間データに対しても高速に推定できる尤度計算法を提案することができた。この尤度は時系列のWhittle尤度に対応するものであり、時系列より不規則に位置する時空間データへ一般化された

Whittle 尤度とみなすことができる。

(2) Wavelet 時空間モデルの空間非正常性への拡張

次に で提案した wavelet 時空間モデルを空間相関構造に非正常性を持たせるように拡張に成功した。 で構成したモデルでは時空間相関構造は時間差，空間地点の差のみに依存する定常モデルである。そこで，時間方向には定常のまま，空間方向のみに非正常性を持たせるよう拡張を行った。ここでは Haar wavelet の係数を空間に依存させパラメトリックに依存性を表現して空間非正常な wavelet 時空間モデルを提案した。

(3) 空間 CARMA モデルの開発

Continuous Autoregressive Moving Average (CARMA) モデルは Peter Brockwell によって提案された連続時間時系列モデルである。本研究は時系列 CARMA モデルを空間モデルへと拡張した random field の一種である。CARMA random field の空間相関を明示的にパラメトリックな関数として表現することに成功した。不規則に位置する空間データへ CARMA モデルをあてはめるため，同定法，推定法，kriging(予測)法を提案した。ここで特に推定・予測法として MCMC 法を用いた Bayes 推定法を提案した。従来の尤度に基づく推定法と比べ時間はかかるが，予測と一体化した統計手法で推定と予測を分けて行う必要がない。シミュレーションおよび実証分析により推定精度の良さや優れた予測パフォーマンスを持つことを確かめることができた。

(4) 成果の公表・発表について

以上で得られた研究成果は下記の国際集会を主催し，共同研究者と共に公開・発表した。

Brockwell 教授(コロラド州立大学)，Huang 教授(アカデミアシニカ)を招待して開催。Recent Advances in Time Series, Spatial and Spatio Temporal Data Analysis. Dec. 5-7, 2013,

IMS 主催による国際研究集会に独立セッションを開いて出席。"Recent progress of spatial and spatio-temporal statistics", Third ims-aprm meeting in Taipei, 2014 年 6 月，台北

Tiao 教授(シカゴ大学)Fryzlewicz 教授(ロンドン大学政治経済学院)を招待して開催。International Conference on Statistical Analysis of Large Scale High Dimensional Socio-Economics Data 2014 年 11 月，東北大学

Robinson 教授(ロンドン大学政治経済学院)を招待して開催。Recent Progress in Time Series and Related Fields, 2015 年 12 月，東北大学

Brockwell 教授(コロラド州立大学)を招待して開催。Recent Progress in

Time Series Analysis in honor of professor Hosoya.

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 27 件)

Brockwell, P. and Matsuda, Y. Continuous auto-regressive moving average random fields on R^n . Journal of the Royal Statistical Society, Series B. 査読有, Vol. 79 (3), 2017 p. 833-857 DOI: 10.1111/rssb.12197

Qin, P. and Nishii, R. Statistical prediction of Dst index by solar wind data and t-distributions. IEEE transactions on Plasma Sciences. 査読有, Vol. 43, 2015, p. 3908-3915. DOI: 10.1109/TPS.2015.2485661

Kim, S., Hayashi, K. and Kurihara, K. Geostatistical data analysis with outlier detection. Journal of the Korean Data Analysis Society, 査読有, Vol. 16, 2014, 2285-2297.

Nishii, R. Regression analysis and its development. Mathematics for Industry, 査読有, Vol. 5, 2014, p. 249-262.

Wali, U., Tsukuda, Y. and Matsuda, Y. Term Structure Forecasting of Government Bond Yields with Latent and Macroeconomic Factors: Do Macroeconomic Factors Imply Better Out-of-Sample Forecasts? Journal of Forecasting, 査読有, Vol. 32(8), 2013, p. 702-723.

[学会発表](計 19 件)

松田安昌, CARMA random fields. The 10th ICSA international Conference. 2016.12/19-22, 上海(中国)

松田安昌, Levy driven CARMA random fields. 2015 年度統計関連学会連合大会. 2015.9/6-9. 岡山大学(岡山県岡山市)

松田安昌, Wavelet analysis of spatio-temporal data. The 3rd ISM Asia Pacific Rim meeting, 2014.6/29-7/3, Taipei(Taiwan)

松田安昌, 大規模で非正常な時空間データの wavelet 分析. 2013 年度統計関連学会連合大会. 2013.9/8-11. 大阪大学(大阪府豊中市)

[図書](計 2 件)

西井龍映, 実データの統計観測を題材とした学部教育，大学教育の数学的リテラシー，東信堂，2017，344 ページ。

矢島美寛, 時系列解析ハンドブック，編集翻訳，朝倉書店，2016，788 ページ。

〔産業財産権〕

出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

松田 安昌 (MATSUDA, Yasumasa)
東北大学・大学院経済学研究科・教授
研究者番号：10301590

(2) 研究分担者

矢島 美寛 (YAJIMA, Yoshihiro)
東北大学・大学院経済学研究科・客員教授
研究者番号：70134814

(3) 研究分担者

西井 龍映 (NISHII, Ryuei)
九州大学・マスフォインダストリ研究所・
教授
研究者番号：40127684

(4) 研究分担者

栗原 考次 (KURIHARA, Koji)
岡山大学・大学院環境学研究科・教授
研究者番号：20170087