

令和元年6月21日現在

機関番号：62603

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00067

研究課題名(和文) スパース正則化を利用した多変量時系列モデリングとその応用に関する研究

研究課題名(英文) Multivariate time series modeling with sparse regularization and its applications

研究代表者

川崎 能典 (Kawasaki, Yoshinori)

統計数理研究所・モデリング研究系・教授

研究者番号：70249910

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：統計モデルにおいて目的変数を説明する候補変数が高次元の状況下で、適切な変数選択・グルーピングを行う円滑閾値型推定方程式を、時系列解析の文脈に応用した。ボラティリティ予測モデルにおける変数選択を、経験類似度に基づくモデル化と、動的トピックモデルから抽出したトピック時系列の選択という2つの問題で議論し、その有効性を確認した。対数死亡率に対する要因分解モデルの残差に見られる時間軸効果を、先験的な仮定によらずスパース推定する方法を提案した。また、高次元モデルで周辺効果の強い変数に隠れて効果が見えにくい真の因果変数を拾い上げる方法論を提案した。

研究成果の学術的意義や社会的意義

IoTの推進により、学術・社会の両面でさまざまなセンサーデータが取得可能になっており、その多くは時間と共に観測される時系列データで、往々にして多変量である。従来の多変量時系列モデルは、比較的少数の変数間の相互共分散を通じてリード・ラグ関係を抽出するものであったが、ラグが深くなると高次元では推定が破綻する。本研究で試みたスパース推定との組合せは、今後の大容量の時系列解析につながる成果である。

研究成果の概要(英文)：We promoted the smooth-threshold estimation equations (STEE) to develop a prediction model with high accuracy even in high dimensional time series analysis. First, we worked with variable selection problem in volatility forecasting. We focused on empirical similarity-based models which turned out to produce better forecasting. We also compared topic score series which were extracted news text data using a dynamic topic model. Some topic score series are found to help forecasting. We also applied sparse regularization to vector autoregressive models, especially to the residual vector series from Lee-Carter model for log-mortality. Finally we proposed a variable selection method with which we can salvage true causal variables masked by other variables with strong marginal correlation.

研究分野：統計科学

キーワード：スパース正則化 円滑閾値型推定方程式 ボラティリティ 経験類似度 トピックモデル 多変量自己回帰モデル 対数死亡率 マスク効果

1. 研究開始当初の背景

背景には正則化法による変数選択理論の進展がある。90年代に入ると統計科学においては、モデルが含む未知母数にL1型のペナルティ(ゼロ方向への絶対値型縮小制約)を課して推定する方法、すなわちスパース正則化法が変数選択の手法として盛んに研究されるようになった。

年代順に記すと、Frank and Friedman (1993, Technometrics)によるbridge regression, Tibshirani (1996, JRSS(B))によるLeast Absolute Shrinkage and Selection Operator (LASSO), Fan and Li (2001, JASA)によるSmoothly Clipped Absolute Deviation penalty (SCAD), Zou and Hastie (2005, JRSS(B))によるelastic net, Zou (2006, JASA)によるAdaptive LASSOが挙げられる。

本研究課題の前身である基盤研究(C) (課題番号25330049)では、Ueki (2009, Biometrika)で提案された計算負荷の少ないスパース正則化法と、それを高次交互作用の検証を含む変数グルーピング問題に拡張したUeki and Kawasaki (2011, Electronic J. Statist.)に基づき、判別・分類・パターン認識における予測変数や因子の探索に利用できる汎用的方法に関する研究を推進した。

2. 研究の目的

上述の研究開始当初の背景に照らし、これまで取り組んできた円滑閾値型推定方程式(STEE)によるスパース正則化法の方法論を、時系列解析の諸問題に水平展開することである。事例として、高次元多変量自己回帰モデルにおける変数選択があり、もうひとつはボラティリティ予測モデルでの変数選択がある。

3. 研究の方法

本研究は円滑閾値型推定方程式(smooth-threshold estimating equation, STEE)に立脚し変数選択・自動グルーピングを適用あるいは拡張するので、まずSTEEの基本的な着想を説明する。

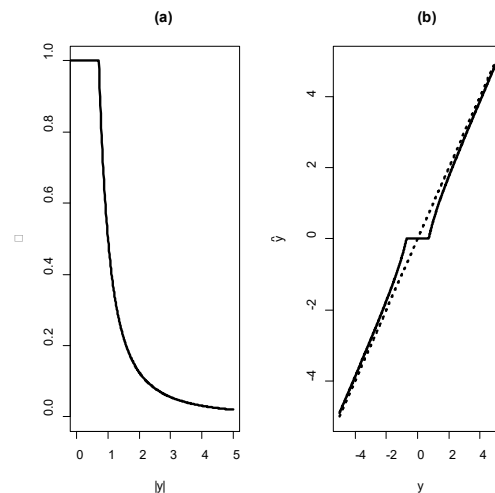
罰則付き損失関数として $L(\theta) + \sum_{j=1}^d \rho_j(\theta_j)$ を考える。ここで $\rho_j(\theta_j)$ はj番目のパラメータの非負罰則関数とする。従って設定としてはZou (2006)のadaptive lassoと同様である。

我々は罰則関数として、ある既知の重み $w_j \in [0, \infty]$ に基づき $\rho_j(\theta_j) = w_j \theta_j^2 / 2$ という定式化を採用する。もし $w_j = \infty$ であれば、罰則項が損失関数に優越するので、 θ_j はゼロとなる。パラメータがd次元であるとする、d本の推定方程式を解くことで解は得られる。

一見すると罰則関数は2次制約であり、スパース正則化というよりリッジタイプの制約と思われるかもしれない。しかし以下に述べるように、STEEでは閾値則が w_j の中に含まれている。

新たにパラメータ $\delta_j \in [0, 1]$ を導入し、 $w_j = \delta_j / (1 - \delta_j)$ という形でパラメータを取り直す。推定方程式はおのずと $(1 - \delta_j) \partial L(\theta) / \partial \theta_j + \delta_j \theta_j = 0, (j = 1, \dots, d)$ と書き換えられる。j番目の推定方程式で $\delta_j = 1$ ($w_j = \infty$ に対応)であれば $\theta_j = 0$ となり、スパース解に帰着する。Adaptive lassoの議論を借用することで、 δ_j は θ_j の推定に先立ってデータから決めておくことができる。我々の問題においては例えば $\hat{\theta}_j^{(0)}$ を \sqrt{n} 一致を持つような何らかの初期推定量として $\delta_j = \min(1, \lambda / |\hat{\theta}_j^{(0)}|)$ と決めることが自然である。チューニングパラメータ (λ, γ) はBIC型の規準で選択すればよい。

STEEの閾値関数の挙動とそのときの δ_j の推移を右図に示す。右パネルで横軸が大きいところから徐々に左に向かって(0に向かって)動いたとき、左パネルで徐々に δ_j の値が1に近づき、1に達したところでスパース解($\hat{y} = 0$)生成している様子がわかる。この図とFan and Li (2001, JASA)のFigure 2 (c)を見比べれば、STEEが本質的にSCADと同じような効果を持つことが理解できる。



4. 研究成果

(1) ボラティリティ予測モデルにおける変数選択

事例ベース意思決定理論に基礎を置いた経験類似度 (ES, Empirical Similarity) という概念を適用することにより、異なるモデルから生じるボラティリティ予測値を結合する方法を研究した。経験類似度の枠組みでは、意思決定者が予測モデルや予測値の尤もらしさに関する確率評価を行わずに、専ら類似性のみによって将来を予測することができる。具体的には、過去のモデル予測値と対応するボラティリティの実現値との距離を定量化することによって、予測の

組合せの重みを決定する。そして、決定された重みを用い将来のボラティリティを予測する。

本研究では、経験類似度モデルから得られたボラティリティの予測値とその他時系列モデルの予測値とを実証的に比較を行った。モデルの予測力比較については、誤差関数に基づくモデル信頼集合 (Model Confidence Set, MCS) を用いることにより、複数の銘柄と推定予測期間におけるモデルの予測力を順位付けし、最良モデルの累積頻度を分析し評価を行った。対応する論文としては②と⑤が、学会発表としては⑦、⑧、⑨、⑩、⑭、⑱などがある。

更に、テキストデータから抽出した時系列情報をボラティリティモデルに取り込み、予測性の向上を検証する研究を行った。具体的には動的トピックモデルから推定される各トピック時系列を Heterogeneous Autoregressive Model (HAR モデル) の右辺に取り込み、Patton の誤差関数に基づくモデル信頼集合 (Model Confidence Set, MCS) を用いることによりモデルの予測力比較を行った。特定のトピックスコアは実現ボラティリティの動きによく似ており、HAR モデルの予測力を改善するトピック系列を事後的に発見することに成功した。対応する論文としては①が、学会発表としては②、③、⑥がある。

(2) 高次元多変量自己回帰モデルにおける変数選択

対数死亡率に対する Lee-Carter モデルの残差に関し、高次元多変量時系列モデルをあてはめる予備的な研究を実施した。結果、生年の順序という構造を利用して、同時応答を含む形で単一方程式に基づく推計を行い、その際にパラメータに対してスパース正則化を課すアプローチからをとった。Igawa (2013, Asia Pacific Journal of Risk and Insurance) は、Lee-Carter モデルの当てはめ残差を観察して、(1) 1 期前の 1 歳下の年齢 (同生年コーホートの 1 年前) の残差と当年の残差の相関 (効果 1: 生年コーホート効果) と、(2) 1 期前の同年齢 (1 つ上の生年コーホートの 1 年前) の残差と当年の残差の相関 (効果 2: 補足的な年齢・期間効果) を想定し、先験的にスパースな係数構造を持つ多変量自己回帰項を持つ LC-VAR モデルを提案して分析している。上記の先験的なスパース制約は目視で確認できる妥当な経験則であり、多変量 AR モデルをスパース推定しても概ね選ばれる係数ではあるが、実際には他のラグ項も選択される。推定された係数の現実的含意等については、なお残された課題である。また、スケラム分布に基づく整数値自己回帰モデルに関する研究も行った。発表論文としては④がある。多変量自己回帰モデルと因果性分析に関する解説論文として図書の③がある。

(3) 線形重回帰における「マスクされた変数」の効率的探索

この研究成果は、時系列データを含む広範囲のデータに共線性が存在するときに、Elastic-Net に代表されるスパース正則化法は周辺効果のある変数だけを捨て、真の見落とすことを示し、その代替案を提示したものである。

多重共線性が深刻な説明変数群を使って推定された線形飽和重回帰モデルでは、有意な説明変数は偶発的に有意になっているに過ぎない可能性がある。一方、飽和重回帰モデルで有意にならず、かつ目的変数との周辺相関がないと思われる説明変数でも、特定の説明変数セットと共に推定されれば有意になることがある。本研究では、効果が他の変数にマスクされている説明変数を効率的に探索する方法を提案する。飽和回帰モデルが推定可能な状況であっても、部分回帰的全探索は計算負荷が高い。ここで提案する方法は、説明変数・被説明変数を合わせた変数グループ内で全てのペアの相関係数を計算し、特定の説明変数から目的変数までの最短経路をダイクストラ法で求めることで、効果を浮かび上がらせる共変量集合を特定する。この方法は、ケースの次元より説明変数の次元が大きい状況下で、かつ説明変数群に多重共線性が潜んでいる状況でも適用可能である。研究成果は、論文としては③が、学会発表としては⑩がある。

5. 主な発表論文等

[雑誌論文] (計 5 件)

- ① Morimoto, T. and Kawasaki, Y., (2017). Forecasting financial market volatility using a dynamic topic model, Asia-Pacific Financial Markets, Vol. 24(3), 149-167. (DOI:10.1007/s10690-017-9228-z) [査読有]
- ② 森本孝之, 川崎能典 (2017). 経験類似度に基づくボラティリティ予測, 統計数理, 第 65 巻 1 号, 155-180. [査読有]
- ③ Ueki, M., Kawasaki, Y. and Tamiya, G. (2017). Detecting genetic association through shortest paths in a bi-directed graph, Genetic Epidemiology, Vol. 41, 481-497. (DOI: 10.1002/gepi.22051) [査読有]
- ④ 中嶋雅彦, 酒折文武, 川崎能典 (2017). 整数値自己回帰モデルの最近の発展, 統計数理, 第 65 巻 2 号, 323-339. [査読有]
- ⑤ Morimoto, T. and Kawasaki, Y., (2017). Volatility Forecasting with Empirical Similarity: Japanese Stock Market Case. In JSM Proceedings, Business and Economics Statistics Section. Baltimore, MD: American Statistical Association. 2483-2510.

[査読無]

[学会発表] (計 19 件)

- ① 貝淵響, 川崎能典, “Value-at-Risk estimation: A novel GARCH-EVT approach dealing with bias and heteroscedasticity”, 第13回日本統計学会春季集会, 2019年3月10日, ポスター発表.
- ② 川崎能典, テキストデータからの情報抽出を利用した時系列予測, 科研費研究集会「経済統計・政府統計の理論と応用」, 2019年1月31日, 一般講演.
- ③ Kawasaki, Y., “Forecasting Financial Market Volatility Using a Dynamic Topic Model”, 2019 ISI-ISM-ISSAS Joint Conference, 2019年1月19日, 一般講演.
- ④ 川崎能典, 円滑閾値型推定方程式による与信スコアリング, リスク解析戦略研究センター第6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」, 2018年12月20日, 一般講演.
- ⑤ Kaibuchi, H. and Kawasaki, Y., “Comparison of EVT methods for GARCH-EVT approach applied to financial time series”, 11th International Conference of the ERCIM WG on Computational and Methodological Statistics, 2018年12月15日, 一般講演.
- ⑥ Kawasaki, Y., “Forecasting Financial Market Volatility Using a Dynamic Topic Model”, CEQURA Conference 2018 on Advances in Financial and Insurance Risk Management, 2018年10月5日, 一般講演.
- ⑦ Morimoto, T and Kawasaki, Y., “Volatility Forecasting with Empirical Similarity: Japanese Stock Market Case”, 11th International Conference on Computational and Financial Econometrics 2017, Senate House, University of London, London, U.K., 2017年12月18日, 一般講演.
- ⑧ 川崎能典, Quarticityを考慮した経験類似度に基づくボラティリティ予測, リスク解析戦略研究センター第5回金融シンポジウム「ファイナンスリスクのモデリングと制御IV」, フクラシア丸の内オアゾ, 東京, 2017年12月14日, 一般講演.
- ⑨ Kawasaki, Y. and Morimoto, T., “Volatility Forecasting with Empirical Similarity: Japanese Stock Market Case”, CEQURA Conference 2017 on Advances in Financial and Insurance Risk Management, Kardinal Wendel Haus, Munich, Germany, 2017年9月25日, 一般講演.
- ⑩ Kawasaki, Y. and Morimoto, T., “Volatility Forecasting with Empirical Similarity: Japanese Stock Market Case”, Joint Statistical Meeting 2017, Baltimore, MD, U.S.A., 2017年8月2日, ポスター発表.
- ⑪ Kawasaki, Y. and Kurisu, D., “Scale mixture of Skewed Kalman filter and its application”, ISI 61st World Statistics Congress, Palais de congrès, Marrakech, Morocco, 2017年7月21日, 招待講演.
- ⑫ Kawasaki, Y. and Ueki, M., Effective search for masked explanatory variables in linear regression, ANU-UC-ISM Joint Symposium on Environmental Statistics 2016, University of Canberra, Canberra, Australia, 2016年12月3日, 一般講演.
- ⑬ Kawasaki, Y. and Aoki, Y., Risk analysis of asymmetric price changes in Japanese commodity futures, CEQURA Conference 2016 on Advances in Financial and Insurance Risk Management, Kardinal Wendel Haus, Munich, Germany, 2016年9月26日, 一般講演.
- ⑭ 森本孝之, 川崎能典, 経験類似度に基づくボラティリティの推定と予測, 2017年度統計関連学会連合大会, 南山大学名古屋キャンパス, 名古屋市, 2017年9月6日, 一般講演(英語).
- ⑮ Kawasaki, Y. and Yoshida, Y., Intraday periodicity of high frequent commodity futures data, The 3rd ISM International Statistical Conference, University of Malaya, Kuala Lumpur, Malaysia, 2016年8月10日, 特別講演.
- ⑯ Kawasaki, Y. and Ueki, M., Sparse predictive modeling for bank telemarketing success using smooth-threshold estimating equations, Joint Statistical Meeting 2016, Chicago, U.S.A., 2016年8月2日, ポスター発表.
- ⑰ 川崎能典, 内閣支持率と株価収益率の因果関係, 応用経済時系列研究会第33回研究報告会, 立教大学, 2016年7月2日, 一般講演.
- ⑱ Morimoto, T. and Kawasaki, Y., Volatility forecasting with empirical similarity: Japanese stock market case, The 36th International Symposium on Forecasting, Santander, Spain, 2016年6月21日, 一般講演.
- ⑲ Kawasaki, Y. and Aoki, Y., “Asymmetric modeling of price change in commodity futures”, First Seoul-Tokyo-Stanford Workshop on Financial Statistics and Risk Management, Seoul National University, Seoul, Korea, 2016年4月1日, 招待講演.

[図書] (計 3 件)

- ① 川崎能典, 「ベイズ的セミパラメトリック回帰」, 日本統計学会編『統計科学百科事典』, 丸善出版, p. 1666-1667, 2018年12月25日. [分担翻訳]
- ② 川崎能典, 「時系列回帰」, 日本統計学会編『統計科学百科事典』, 丸善出版, p. 628-630, 2018年12月25日. [分担翻訳]
- ③ 川崎能典, 「VARモデルによる因果関係の推論—内閣支持率と株価を例に」, 岩波データサイエンス Vol. 6 (岩波データサイエンス刊行委員会編), 岩波書店, 2017年6月, pp. 68-81.

[産業財産権]

該当事項なし

[その他]

特記事項なし

6. 研究組織

研究代表者単独での研究プロジェクトである

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。