

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 4 日現在

機関番号：12601

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500364

研究課題名(和文) 金融経済データに対する予測マシンの開発

研究課題名(英文) Study of forecasting system for financial economic data

研究代表者

佐藤 整尚 (SATO, Seisho)

東京大学・経済学研究科(研究院)・准教授

研究者番号：60280525

交付決定額(研究期間全体)：(直接経費) 2,500,000円、(間接経費) 750,000円

研究成果の概要(和文)：本研究では、与えられたデータに基づいて、対象となる金融経済時系列の将来値を時系列的に予測するシステムを開発した。分離情報最尤法というロバストな方法で第一段階の変数探索を行い、次にAIC最適な多変量自己回帰モデルを推定することで、現代のビッグデータ社会に必要な、大規模データからの予測が可能となった。この方法を、様々な経済変数や企業の経営変数などに応用すれば、様々な知識発見が期待できる。

研究成果の概要(英文)：In this study, I developed a time series forecasting system for financial economic data from given data sets. Separating Information Maximum Likelihood method was used for robust estimating in the first step, then I utilized Vector Autoregressive models for the predicting models which are optimal in the sense of AIC. It is very important for "Big data" era that this system can be applied for large scaled data. A new knowledge discovery is expected when this method is applied for various economic data and business data.

研究分野：時系列解析

科研費の分科・細目：情報学・統計科学

キーワード：時系列予測 多変量自己回帰モデル 分離情報最尤法 季節調整法 逐次予測 X12Decomp AIC最適化
ビッグデータ

1. 研究開始当初の背景

(1) 近年 IT 技術の普及により、金融経済分野においても様々な情報やデータが蓄積されるようになった。一方、日本経済の置かれている環境はグローバル化の進展とともに一層複雑化し、従来の、単純化された経済モデル、金融モデルでは現実の現象を必ずしも正しく捉えきれなくなってきた。

(2) このような状況下でこれらの膨大なデータから意味ある情報を抽出すること、言い換えれば付加価値の高い情報の生産を行うことは益々困難になってきていた。しかしながら、高度成長期と異なり、将来に対する不確実性が増している中で、主に企業部門で、データに基づく客観的な予測方式の開発が期待されてきた。

(3) コンピュータ環境が激変しており、CPU のクロックの上昇が抑えられて、その代わりに、より多くのコアを内蔵する CPU が普及してきた。したがって、プログラムを高速に走らせるために、並列化が可能なアルゴリズムが求められる時代となった。

(4) 研究代表者はこれまでに、多変量自己回帰モデルを使った逐次予測の方法を開発してきた。この方法は与えられたデータセットに対して、逐次的に多変量自己回帰モデルを当てはめて、予測を行い、その予測値の変化から、足元の経済状況を読み解くというものであった。

2. 研究の目的

(1) 研究の目的はビッグデータ時代にふさわしい、大規模データを使った、新しい予測システムの開発である。具体的には、最新の統計学的手法を使った客観的な方法を用いて、高速に予測値を出せる方法を研究する。また、マクロ経済時系列には季節性や、リーマンショックなどの異常値を含むものが多く、これらに対する対策も検討する。

(2) 予測のベースには多変量自己回帰モデルを採用するが、問題は、変数の選び方にある。目的となる変数に対して、どのような変数を採用すれば良いかを決める必要がある。しかしながら、多変量自己回帰モデルは複数の変数の組み合わせによって予測値が決まるため、多数の時系列から最適な系列の組み合わせを探索するのは困難であり、新しい手法を開発する必要がある。また、多変量自己回帰モデルにおいては、定常な時系列を対象にしている事から、一般の非定常な時系列を定常化する手法の開発も必要である。

(3) 単に手法の開発のみならず、実際に動作するプログラムを開発することまでを目標とする。また、現在のコンピュータ環境が大規模データに対応するため並列化が可能

なアルゴリズムの開発も行う。

3. 研究の方法

(1) 大規模なデータセットから目的となる変数に関係の深い変数を抽出する方法として、研究代表者がかつて、研究を行い、提案した分離情報最尤法を考える。この方法はもとも、金融資産価格の高頻度観測データに対してロバストな分散を推定する方法であったので、一般のマクロ経済時系列にたいして適用したときにおこる問題等をシミュレーションや実際のデータを使って調べる。

(2) 一般のマクロ経済時系列は季節性を含み、かつ非定常であるので、季節調整や定常化の方法を考える必要がある。しかしながら、近年、リーマンショックや東日本大震災で時系列に大きな異常値が見られることが多く、これに対して、既存の季節調整法を改良し、実用に耐える手法を開発する。

(3) 分離情報最尤法で得られた情報をもとに最適な変数の組み合わせを探索するアルゴリズムを完成させる。また、シミュレーションにより、最適な組み合わせになっているかチェックを行う。

(4) 最後に実際のデータに適用して、そのパフォーマンスを調べる。開発に当たっては既存のソフトウェア資産を生かすため R と呼ばれる統計言語を用いて実装する。

4. 研究成果

(1) 分離情報最尤法で求めた変数間の相関について、マクロ経済時系列でもロバストなものになることが分かった。具体的には、以下の通り。

ノイズが存在する場合でもバイアスが少くない推定が可能である。通常的相关係数ではノイズが存在する場合には正しく推定できない。分離情報最尤法を用いると、ほぼ正しく推定できることがシミュレーションによって確認された。

季節性が存在する場合でも季節調整を行うことなく原系列のままでも適用が可能であることが分かった。原系列のままであると通常的相关係数ではバイアスが生じてしまう。また、正しく季節調整を行うには多くのコストがかかるため、対象となるすべての系列に対して行うのは難しい。したがって、この性質は計算スピードの面で大きなプラスである。

通常的相关係数では同時点の相関のみを検出するが、分離情報最尤法による相関は時点がずれている相関(相互相関)も同時に検出することができる。多変量自己回帰モデルでは主に相互相関をベースにして予測を行うので、この相互相関をいっぺんに推定できることは重要である。

マクロ経済時系列は通常、金融データに比

べてデータ長が短い。このため、分離情報最尤法を用いると効率性の意味で、やや、劣ることが分かった。しかしながら、ここで求める相関は変数選択のための情報であるので、それほど、精度は要しない。したがって、実用上はあまり問題とならないと思われる。

非定常な時系列同士のレベルでの相関を取ると、見せかけの相関をしめすことがあるが、分離情報最尤法は階差系列をベースにしており、このような現象は起こらないことがシミュレーションにて確認された。

なお、分離情報最尤法は並列可能なアルゴリズムである。

(2) 分離情報最尤法で目的の系列に相関があると判定された系列については、次のステップで定常化のプロセスを施す。

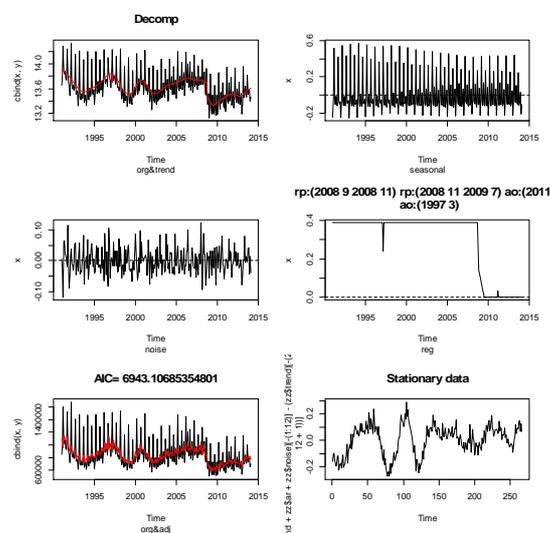
まず、状態空間モデルを使って季節調整を施す。具体的には Decomp というプログラムによって行う。

Decomp では異常値や構造変化を検出する機能が不足しているため、これを追加してものが、今回開発した X12Decomp である。ここでは、様々なダミーを用いて、処理を行うことが可能である。

全ての異常値や構造変化を推定すると、かなりの時間がかかるので、今回は、消費増税とリーマンショックの時期、東日本大震災の時期のみを対象に処理を行った。具体的には、リーマンショックの時期(2008年10月から2009年6月にかけて)に RAMPx2(傾向型構造変化)を、消費増税(1997年3月)と東日本大震災(2011年3月)に加法的異常値を仮定した。

X12Decomp にて季節調整を行うとトレンド成分、AR成分、季節成分、ノイズ成分、異常値・構造変化成分に分けられる。これを使って、k期前のトレンド成分を今期の調整系列から引いた系列を作り、これを定常系列とする。kはユーザーが選ぶものとする。

以下の図は機械受注統計を X12Decomp で分解し、定常化した例を示している。



(3)(2) で求めた定常系列の組み合わせを用いて多変量自己回帰モデルを当てはめて、予測モデルを構築する。

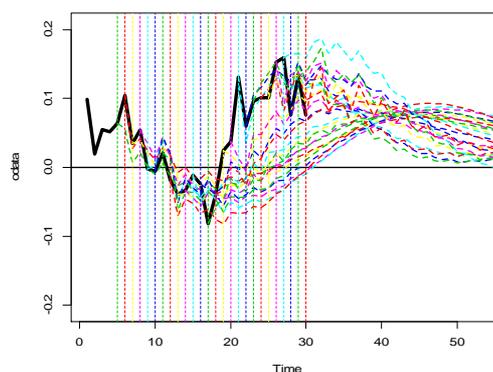
モデルに組み込む変数は(1)の段階で比較的少数に絞り込まれており、なおかつ、推定された相関の絶対値の大小で優先順位がふられている。これをもとに、探索を行う。

まず相関の大きい順に変数を加えていき、多変量自己回帰モデルを当てはめて、AIC(赤池情報量規準)を計算する。とりうる最大の変数の数はあらかじめ決めておく。なお、ラグの選択が重要であるが、ここでは、すべての変数が同じラグを持つモデルであると仮定して、AIC 最適なラグをその都度、選択した。最小の AIC を与える変数の組み合わせを暫定のモデルとする。このモデルから、変数を増減させて、さらに AIC 最適なモデルを探す。このような探索を何度か行って、ベストなモデルを探し出す。このような方式が取れるのは、(1)の段階で、候補となる変数が少数に絞り込まれているからである。

シミュレーションによっての方法でベストなモデルが探索できるかを調べた。その結果、ほぼ、ベストなモデルに近いモデルが選ばれることが分かった。全く同じではない場合でも、AIC としては大差のないモデルであり、実用上は問題ないと思われる。

(4)(3) でもとめた多変量自己回帰モデルを用いて、実際の予測を行う。ここでは、直近の値からの予測だけではなく、少し過去を遡った予測も同時している。これを、逐次予測と呼ぶが、たんに将来の予測だけではなく、あしもとの動きが、過去も予測と比べてどうであるかを示している点が特徴である。

以下は機械受注のデータに対して、約200系列のデータセット中から最適な組み合わせを選び、逐次予測を行った例である。



この図では、景気の転換点で予測の向きが大きく変わっていることが分かる。

(5) このように、多数のデータセットの中から、最適な変数を選び出し、逐次予測を行うことが可能となった。この方法を、様々な経済変数や企業の経営変数などに応用すれば、様々なことが発見することが可能となる。

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 2 件)

Kunitomo, N. and Sato, S., “Separating Information Maximum Likelihood Estimation of the Integrated Volatility and Covariance with Micro-Market Noise”, North American Journal of Economics and Finance, 査読有, 2013.

doi:10.1016/j.najef.2013.02.006

国友直人, 佐藤整尚, 「日本のマクロ経済統計の課題」, ファイナンス・景気循環の計量分析, 査読無, 第 10 章, 2011, 259-289.

[学会発表](計 10 件)

Chunhang Chen and Seisho Sato, “Statistical Analysis of Extremal Tsunamis Occurrence Risk with Levy Processes”, シンポジウム「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」, 2013 年 12 月 06 日, 東北大学(宮城県仙台市)

今喜典, 佐藤整尚, 「地域金融機関の業種別貸出と金融機関リスク」, 統計数理研究所第2回金融シンポジウム, 2013 年 11 月 05 日, 学術総合センター(東京都千代田区)

三崎 広海, 国友直人, 佐藤整尚, “The SIML estimation of volatility and covariance for irregularly spaced, unsynchronized and noisy high frequency data”, 統計関連学会連合大会, 2013 年 09 月 09 日, 大阪大学(大阪府豊中市)

佐藤整尚, 「状態空間モデルによる金融・経済データ解析」, 大内賞受賞記念シンポジウム, 2013 年 02 月 12 日, 国際文化会館(東京都港区)

国友直人, 佐藤整尚, 三崎広海, “A Robust Estimation of Integrated Volatility, Covariance and Hedging Coefficients under Non-linear Adjustments, Micro-market

Noises and Random Sampling”, 統計関連学会連合大会, 2012 年 09 月 11 日, 北海道大学(北海道札幌市)

佐藤整尚, 「電力需要データの季節調整について」, 応用経済時系列研究会, 2011 年 11 月 4 日, 同志社大学東京オフィス(東京都千代田区)

陳 春 航, 佐藤整尚, “Statistical Analysis of Tsunamis with Levy Processes”, 統計関連学会連合大会, 2011 年 9 月 7 日, 九州大学(福岡県福岡市)

佐藤整尚, 「Web 上での計算サービス」, 統計関連学会連合大会, 2011 年 9 月 6 日, 九州大学(福岡県福岡市)

Kunitomo, N. and Sato, S., “On Improving the Quality of Quarterly GDP in Japan”, ISI2011, 2011 年 8 月 24 日, コンベンション・センター・ダブリン(アイルランド・ダブリン)

佐藤整尚, 国友直人, 「景気判断と平滑化問題: GDP 公表値を巡って」, 応用経済時系列研究会, 2011 年 6 月 25 日, 立教大学(東京都豊島区)

[図書](計 0 件)

[産業財産権]
出願状況(計 0 件)

取得状況(計 0 件)

[その他]

6. 研究組織

(1) 研究代表者

佐藤 整尚 (SATO, Seisho)
東京大学・大学院経済学研究科・准教授
研究者番号: 60280525