

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 1 日現在

機関番号：34504

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K03406

研究課題名(和文)ビッグデータ解析を利用したボラティリティの推定に関する研究

研究課題名(英文)A study on estimation of volatility using big data analysis

研究代表者

森本 孝之 (Morimoto, Takayuki)

関西学院大学・理工学部・教授

研究者番号：80402543

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：1. データの整備：情報供給に関する代理変数としてオンライン上におけるロイター・ジャパンのニュース記事を用いた。そして、取得したオンラインニュース記事に対し形態素解析を行い、品詞ごとのデータに変換した。
2. モデルの定式化：文書の確率的生成モデルの一つである動的トピックモデルを用いて算出されたトピックスコアとボラティリティの各ラグ値を説明変数に持つ時系列モデルに実装し、各モデルのボラティリティ予測力を誤差関数を用い比較した。
3. 得られた結果：トピックスコアを含めたモデルを用いることにより、ボラティリティの予測力が向上することを実証的に示すことができた。

研究成果の概要(英文)：1. Data manipulation: We used Reuters Japan's online news article as a proxy variable for information supply. Then, we executed morphological analysis on the acquired online news articles and converted these into data for each word class.
2. Model formulation: We implemented time series models with lag values of topic score calculated using a dynamic topic model which is one of stochastic generating models of documents and volatility as explanatory variables.
3. Outcome: By using the model including the topic score, we could demonstrate empirically that the predictive power of volatility improves.

研究分野：経済統計学

キーワード：ビッグデータ オンラインニュース 動的トピックモデル トピックスコア 予測 実現ボラティリティ

1. 研究開始当初の背景

テキストデータから得られた情報を、ボラティリティ変動モデルと関係付け分析した国外の論文に Vlastakis and Markellos (2012) [2] がある。この論文では、米国市場の VIX (恐怖指数) の時系列モデルに関連するニュース記事の個数と Google 検索数指数を外生変数として加え、これらの変数が有意であるかどうかを調べている。その結果は、関連するニュース記事の個数と Google 検索数指数は共に有意に VIX に影響を与えるという事実を見出している。ただし、ここで外生変数として用いられた関連するニュース記事の個数については、以下のような問題点が挙げられる。

- A) どのような話題 (トピック) の記事が有効に働いていたのかが分からない。
- B) 個々の話題 (トピック) の盛衰の方がより効果的に時系列モデルに働くのではないか？

そこで申請者は上記の問題点を克服すべく研究を実行し、その結果は Morimoto, Yamauchi and Kawasaki (2014) [3] に纏められている。この論文の研究概要は以下の5点である。

- (ア) ニュース記事をトピック分析し、トピックごとにボラティリティとの関係を検証する。
- (イ) 用いるデータは、日次のニュース記事集合とする。
- (ウ) ニュース記事中にどのトピック (話題) がどの比率で含まれているかを分析する。
- (エ) 各トピックに関連する記事がどの程度あったかを表すスコアを日次で算出する。
- (オ) 算出したスコアを、ボラティリティの時系列モデルに外生変数として組み込み、モデルの予測力が向上するかどうかを調べる。

上記の背景及びこれまでの研究成果をもとに、本研究はビッグデータ解析を用いたボラティリティの推定手法の体系的な確立を目標に、研究期間内では以下の3点に着目する。

- A) Morimoto, Yamauchi and Kawasaki (2014) [3] で取り扱われたニュース記事以外にも、Twitter や ソーシャルネットワークといったビッグデータも分析の対象とし、各データがボラティリティに与える影響の強弱を明らかにする。
- B) 時系列文書集合における新しいトピックの出現や、既存のトピックの消滅をうまく捉えたモデルとして、時系列トピック、特に (階層) Dirichlet 過程などのノンパラメトリックベイズを用いる。これにより、適切なトピック数をデータ

から推定し、時間単位ごとにトピック数が増えるモデルを構築することが可能となる。

- C) Morimoto, Yamauchi and Kawasaki (2014) [3] では、ボラティリティの時系列モデルとして、単純な自己回帰モデルのみを用いたが、より実際の GARCH, SV といったボラティリティ変動モデルを用い、ボラティリティの推定及びその予測パフォーマンスの比較分析を行う。

Varian (2014) [4] は、ビッグデータを機械学習の手法により分析することが、これからの計量経済学にとっていかに重要であるかを豊富な経験と知識に基づいて説いている。そして、テキストデータを用い金融市場の分析に取り組んだ研究は、最近のビッグデータ・ブームに伴い増え始めている。しかし、これらの研究の多くは、原資産価格の上下変動にのみ着目しており、テキスト情報とボラティリティの関係に関する体系的な本研究は稀有である。予想される結果としては、あるトピック (話題) とボラティリティとの動きは連動しており、これらの間に何らかの因果性、相関性が見いだされることが挙げられる。さらに、こういったテキスト情報がボラティリティの予測力の向上に寄与する可能性も期待できる。また、ファイナンス理論的な側面から、テキスト情報の詳細な動きを分析することにより、金融市場への情報の流入がどのようにボラティリティに影響を与えるかを、可視的に捉えることも可能になる。

2. 研究の目的

本研究は、ビッグデータ、とりわけテキストデータから得られた情報を、GARCH, SV, といったボラティリティ変動モデルに組み入れることにより、ボラティリティ推定の精度が向上するかを理論的、実証的に検証することを目的としている。つまり、新聞の経済ニュース、電子掲示板の口コミ、Twitter, といった情報が、金融市場の原資産価格変動の先行指標として、ボラティリティの動きに何らかの影響を与える可能性があるかを研究する。これは、インターネット上の風邪や熱といった単語の検索数の増加が、実際のインフルエンザの流行の先行指標となり得る可能性を示した Ginsberg et al. (2009) [1] の金融市場への発展的な拡張となる。

参考文献

- [1] Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting flu epidemics using search engine query data," *Nature* 457, 1012-1014, 2009.
- [2] Vlastakis, N. and Markellos, R. N., "Information demand and stock market volatility," *Journal of Banking Finance*, 36,

1808–1821, 2012.

[3] Morimoto, T., Yamauchi, K. and Kawasaki, Y., “Forecasting Financial Market Volatility Using a Dynamic Topic Model,” preprint, 2014.

[4] Varian, H. R., “Big data: New tricks for econometrics,” *The Journal of Economic Perspectives*, 28, 3-27.

3. 研究の方法

ビッグデータ解析を利用したボラティリティの推定に関する研究では、以下の研究項目を実行した。

- (1) まず、本研究を進める上で最も大切となるビッグデータの入手、加工、そして分類を中心に行った。ビッグデータといっても多種多様であるが、本研究では新聞記事データ及びインターネット上の Twitter などの観測文書から得られるテキスト情報を主に用いた。こういった日本語のテキストデータを実際の分析に用いるためには、いわゆる形態素解析というデータの前処理が必要となる。これは、例えば $D = \text{“今日はいいい天気だ”}$ という文章は $D = \{ \text{“今日”, “は”, “いい”, “天気”, “だ”} \}$ と 5 つの品詞に分解できる。こういった処理を、Mecab という形態素解析ソフトを用い、すべての新聞記事や Twitter のテキストデータに対して行った。そして、これら文書がトピック及び単語を表す確率分布にしたがって生成されたと仮定し、観測した文書からトピック分布、単語分布を推論した。これらの作業は、元々のデータが非常に大きいため、実際に処理を行うためには膨大なメモリとハードディスクの容量と十分な計算機の処理能力が必要であった。
- (2) 新聞記事、Twitter などにおける文書がトピックと単語を表す確率分布に従って生成されたと仮定し、観測した文書からトピック分布、単語分布を推論しスコアを算出した。
- (3) 算出されたスコアを実現ボラティリティの自己回帰モデルに外生変数として含むことにより、ボラティリティの予測力が向上するかを比較分析した。

4. 研究成果

(1) 研究用データの整備

- ① 日経 NEEDS から購入した DVD データを 1 分間隔データに加工した上で、実現ボラティリティ (Realized Volatility, RV) を計算した。
- ② 分析に使用するデータとして株価指数 TOPIX を用いた。標本データ期間は 2008 年 1 月から 2012 年 12 月まで

であり、計 1223 日分の日次 RV を利用した。

- ③ 情報の供給に関する代理変数としてオンライン上におけるニュース記事を用いるが、本研究ではとりわけ経済ニュースに強いロイタージャパン (<http://jp.reuters.com/>) のテキストデータを利用した。
- ④ 取得したオンラインニュース記事を mecab により形態素解析を行い、品詞ごとのデータに変換した。

(2) モデルの定式化

- ① オンライン上におけるニュース記事のデータに対しては、文書の確率的生成モデルの一つである動的トピックモデル (Dynamic topic model) を用いた。
- ② 実現ボラティリティの自己回帰モデルにトピックスコアを含めることにより、ボラティリティ予測のパフォーマンスを改善することを本研究の目的としている。そこで、ボラティリティとトピックスコアのラグ値を説明変数に持つ AR, HAR といった基本的なモデルに加え、Bollerslev et al. (2016) による ARQ, HARQ モデルについても実装し、各モデルのボラティリティの予測力を MSE および QLIKE という 2 つの誤差関数を用い比較した。

(3) 得られた結果

すべてのモデルにおいてトピックスコアを含めた方がボラティリティ予測のパフォーマンスを改善することを示すことができた。

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 8 件)

1. T. Morimoto and Y. Kawasaki, Volatility Forecasting with Empirical Similarity: Japanese Stock Market Case, In JSM Proceedings, Business and Economics Statistics Section, Baltimore, MD: American Statistical Association, 2017 年, 2483-2510. 【査読無】
2. T. Morimoto and Y. Kawasaki, Forecasting financial market volatility using a dynamic topic model, *Asia-Pacific Financial Markets* 24(3), 2017 年, 149-167. 【査読有】
3. 森本孝之, 川崎能典, 経験類似度に基づくボラティリティ予測, *統計数理* 65(1), 2017 年, 155-180. 【査読有】
4. 森本孝之, GARCH-MIDAS モデルの本邦株

式市場への適用可能性について, オイコ
ノミカ(名古屋市立大学経済学会) 54(1),
2017年, 63-74. 【査読無】

5. T. Morimoto and S. Nagata, Robust estimation of a high-dimensional integrated covariance matrix, Communications in Statistics - Simulation and Computation 46(2), 2017年, 1102-1112. 【査読有】
6. T. Morimoto, European Option Pricing Under Fractional Brownian Motion with an Application to Realized Volatility, FORMA 31(Special Issue: Forms in Economics and Business Administration), 2016年, S29-S40. 【査読有】
7. D. B. Nugroho and T. Morimoto, Box-Cox realized asymmetric stochastic volatility models with generalized Student's t-distributions, Journal of Applied Statistics 43(10), 2016年, 1906-1927. 【査読有】
8. D. B. Nugroho and T. Morimoto, Estimation of realized stochastic volatility models using Hamiltonian Monte Carlo-based methods, Computational Statistics 30(2), 2015年, 491-516. 【査読有】

[学会発表] (計 10 件)

1. Takayuki Morimoto, Economic Policy Uncertainty and Financial Market Volatility: Evidence from Japan, Kagawa International Symposium "Recent Developments in Statistics and Econometrics", Kagawa University, March 2018
2. Takayuki Morimoto, Volatility forecasting with empirical similarity: Japanese stock market case, The 11th International Conference on Computational and Financial Econometrics (CFE 2017), University of London, December 2017
3. 森本孝之, 経験類似度に基づくボラティリティの推定と予測, 第34回応用経済時系列研究会・研究報告会, 株式会社リファレンス, 2017年11月
4. 森本孝之, 経済政策不確実性と金融市場ボラティリティ, 研究集会 大規模統計モデリングと計算統計 IV, 東京大学, 2017年9月
5. 森本孝之, 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企画セッション「計算統計学と従属データ解析手法・ソフトウェア開発の相乗発展」, 南山大学, 2017年9月

6. Takayuki Morimoto, Economic Policy Uncertainty and Financial Market Volatility: Evidence from Japan, The 2nd International Statistical Institute Regional Statistics Conference, ISI RSC 2017, Bali International Convention Center (BICC), March 2017
7. 森本孝之, Economic Policy Uncertainty and Financial Market Volatility: Evidence from Japan, 科学研究費補助金基盤研究 C (代表者:前川功一)「経済時系列モデルのパラメータ変化に関するモニタリング手法の研究開発」研究集会, 広島経済大学, 2017年2月
8. 森本孝之, 現在進めている研究について, 平成28年度数学協働プログラム「政治・社会事象の数的分析に関するスタディグループ」第1回会合, 明治大学, 2016年10月
9. 森本孝之, ボラティリティ・モデリング, 平成27年度数学協働プログラム研究集会「ウェアラブル機器によって得られた医療ビッグデータを利活用するための数理モデルの開発」キックオフセミナー, 福井大学, 2015年9月
10. 森本孝之, 非整数ブラウン運動とその日本の株式市場への応用, 平成27年度数学協働プログラム研究集会「統計科学の新展開と産業界・社会への応用」・2015年度統計関連学会連合大会企画セッション「統計的従属性モデリングの理論と応用」, 岡山大学, 2015年9月

[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]

ホームページ等

<http://sci-tech.ksc.kwansei.ac.jp/~mori/mot/>

6. 研究組織

(1) 研究代表者

森本 孝之 (MORIMOTO TAKAYUKI)

関西学院大学・理工学部・准教授

研究者番号: 80402543