

# Google 検索数指数と日本の株式市場

関西学院大学大学院理工学研究科

数理科学専攻 森本研究室 芦田莉奈

## 1 はじめに

近年、情報通信技術の発展に伴い、インターネット上の検索履歴データや、blog、twitter などを通して日々発信されるテキストデータ等、「ビッグデータ」と呼ばれる様々な種類の情報が利用可能となっている。その中でも、Google が無料で提供している Google トレンドにより取得できる、Google 検索数指数データを用いて様々なものを予測する研究は多く行われている。そこで今回、日々変動する株価と Google 検索指数の関係性に目を付け、日本の株価変動において Google 検索指数がどれほど影響を受けており、連動しているのかをいくつかの方法を用いて研究していく。

## 2 推定方法

VAR モデルを用いる。VAR を用いる目的として、1 つには、複数の変数で予測精度向上を図ることであり、もう 1 つには、変数間の動学的関係の分析を行うことである。

## 3 VAR モデル

ベクトル自己回帰 (VAR) モデルは、AR モデルをベクトルに一般化したものであり、VAR( $p$ ) モデルは  $y_t$  を定数と自身の  $p$  期の過去の値に回帰したモデルである。すなわち、

$$y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t, \epsilon_t \sim W.N.(\Sigma)$$

というモデルである。ここで、 $c$  は  $n \times 1$  定数ベクトルであり、 $\Phi_i$  は  $n \times n$  係数行列である。

## 4 実現ボラティリティ

$RV_{(m)}$  は実現ボラティリティ (realized volatility) と呼ばれており、時間と共に変動するランダムなボラティリティの実現値である。実際の  $RV_{(m)}$  の計算にあたっては、非均

等間隔に並んだティックデータをそのまま用いるのではなく、5 分間隔、30 分間隔と言った等間隔幅  $h = t_i - t_{i-1} = T/m$  のグリッド（格子）点を取り、各グリッド点において、直前に起こった取引値により補間をする“直近ティック補間（previous-tick interpolation）”か、そのグリッドをはさむ両側（直前・直後）の取引値の線形補間（linear interpolation）をするかが行われる。

## 5 Google トレンド

Google トレンドとは、あるトピックについての関連キーワードを入力すると、Google におけるユーザーにより検索されたキーワードの検索回数を時間経過に沿ってどのように変化したかを分析し、図 1 のように検索ボリュームが時系列に折れ線グラフで表示され、その時々の「検索ユーザーの関心の高まり」がグラフやランキングにより一目でわかるようになっている。

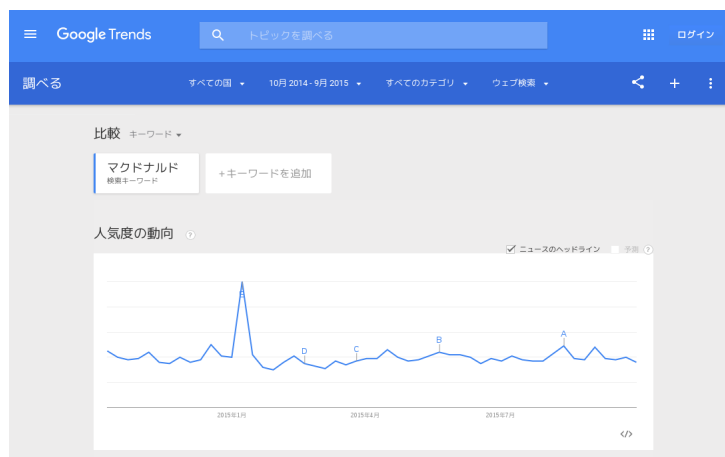


図 1 「マクドナルド」の 2014.9～2015.10 の検索数指数

## 参考文献

- [1] 沖本竜義 (2010), “経済・ファイナンスデータの計量時系列分析” 朝倉書店
- [2] 林高樹, 吉田朋広 (2008) “高頻度金融データと統計科学”, 21 世紀の統計科学 I, 10, 267-304
- [3] 銀座で働くデータサイエンティスト (<http://tjo.hatenablog.com/entry/2013/07/25/194546>)
- [4] VAR モデル 時系列分析 Logics of Blue (<http://logics-of-blue.com/var>)