



ソーシャルメディアと金融市場（〈特集〉計量分析の理論と応用）

辻, 隆司

(Citation)

国民経済雑誌, 209(1):13-25

(Issue Date)

2014-01

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCOI)

<https://doi.org/10.24546/81008949>

(URL)

<https://hdl.handle.net/20.500.14094/81008949>



ソーシャルメディアと金融市場

辻 隆 司

国民経済雑誌 第 209 巻 第 1 号 抜刷

平成 26 年 1 月

ソーシャルメディアと金融市場

辻 隆 司

ソーシャルメディアの重要性が高まる中で金融市場との関係性に対する関心が高まっている。そこで本稿では、ソーシャルメディアと金融市場の関係を探るべく、日経平均株価と株価に関するブログ情報、為替レートと為替に関するブログ情報を対象にグランジャーの因果性テストとインパルス応答関数による分析を行った。その結果、ブログ情報と株価市況に明確な因果関係がみられないこと、また、ブログ情報と為替市況については一定の因果関係がみとめられたものの、その動きは、為替市況の動きをもとに為替に関するコメントをブログに書き込んでいる可能性が高く、ブログ情報の動向が為替市況に影響を与えていないことが明らかになった。

キーワード ソーシャルメディア、ブログ、株価、為替、VEC

1 はじめに

金融市場におけるソーシャルメディア¹⁾の存在感が大きくなりつつある。ブログや twitter, Facebook などのソーシャルメディアで金融市場に関する情報のやり取りが活発になっているが、こうした動向を踏まえて、民間企業ではいわゆる“口コミ”情報を株価の分析や予測に活かす試みが始まっている²⁾。近年、ビッグデータの活用が注目されているが、こうした動きは、今後ますます進展することが予想される。

他方で、学術研究の動向に目を向けると、ソーシャルメディアと金融市場の関係に着目した研究が増えつつある。代表的な先行研究としては、Antweiler and Frank (2004) や丸山ほか (2009)、諏訪ほか (2012)、渡辺 (2012) などがある。まず、Antweiler and Frank (2004) では、ダウ・ジョーンズ工業株価指数とダウ・ジョーンズインターネット指数の組み入れ銘柄45社を対象に、株式に関するチャットのやり取りが株価に影響を及ぼすか否かを検証している。その結果、株価の上昇に対して有意な結果は見いだせなかったが、メッセージ数がボラティリティと有意な関係にあることを明らかにしている。また、丸山ほか (2009) では、自然言語処理と機械学習の手法を用いて投稿データを「弱気」、「中立」、「強気」の3種類に分類した上で、東証一部上場銘柄50社を対象にインターネット掲示板と株価指標の関係を検証している。その結果、弱気投稿数が多いと翌日の株価リターンはマイナスになる傾向があ

ることを明らかにしている。諏訪ほか（2012）では、Antweiler and Frank（2004）や丸山ほか（2009）が特定の銘柄のみを対象にしていたのに対して東証一部全体を分析対象にし、インターネット株式掲示板の投稿数と株価リターンの関係を分析した。その結果、株式掲示板の強気指数が部分的に株価リターンの説明変数になることを示唆している。さらに、渡辺（2012）では、ブログの為替に関するコメントと為替レートの関係について時差相関分析を用いたシンプルな分析を行っている。その結果、ブログのコメントデータと為替レートの間には、同時点において強い相関性がみられるものの、時差を伴った相関はほとんどみられないことを明らかにしている。

しかし、これらの先行研究では、テキストマイニング技術については多くの工夫がなされているものの、ソーシャルメディアと金融市場の関係性の時系列分析については簡単な相関分析にとどまっているなど、応用時系列分析の観点からみると厳密な分析がなされていたとは言い難い。ところが、ソーシャルメディアと金融市場を対象にした研究は、情報学やマーケティングなどの分野で比較的活発に行われているものの、応用時系列分析を主とする計量経済学の分野では先行研究がほとんどみられない。これには様々な理由が考えられるが、一つはデータ収集・加工の困難性があげられる。ソーシャルメディア情報を過去に遡って集めて時系列データを生成する場合、膨大な手間がかかる上に高度な情報処理技術が必要になる。近年では、民間企業等がソーシャルメディア分析を行うためのデータベースを開発し、そのサービスを有償で提供するケースが増えているが、多くは法人向けのサービスであるため、利用コストが高額になる。また、ソーシャルメディアに掲載されている文章の文脈に沿って、分析対象のキーワードを研究者が意図した意味合いで収集加工するためには高度なテキストマイニング技術が必要とされる。こうした技術は情報学やマーケティングなどの分野では多用されているものの、計量経済学分野での応用は進んでいないため研究上の障害になっている。

そこで本稿では、ソーシャルメディアと金融市場の関係を探るべく、シンプルな応用時系列分析を行う。具体的には、日経平均株価と為替レートを対象に次の手順で分析を進める。まず、使用データの特徴を捉えるべく、簡単な記述統計に加えて単位根検定と共和分検定を行う。そして、日経平均株価の市況動向と株価に関するブログ情報、為替レートの市況動向と為替に関するブログ情報のそれぞれの因果関係と影響度合いを分析するために、グランジャーの因果性テストとインパルス応答関数による分析を行う。

2 使用データの構築方法と基本特性

2.1 使用データの構築方法

本稿の使用データは、次の4種類の系列である。金融市場に関するデータとして、日経平

均株価（日経225）と為替レート（米ドル／円）の2系列，ソーシャルメディアデータに関するデータとして，ブログ上に記述された日経平均株価に関するコメントを集計したデータと，為替に関するコメントを集計したデータの2系列である。

まず，日経平均株価については，Yahoo! ファイナンス⁴⁾の株式・株価時系列から日次の終値を取得し，その対数値を分析に用いた。また，為替レート（米ドル／円）についても，同じく Yahoo! ファイナンスから入手した。外国為替・為替時系列から日次の終値を取得し，その対数値を利用した。

一方，ソーシャルメディアデータについては，富士通株式会社が提供する「FUJITSU DataPlaza ソーシャルメディア分析ツール」を用いてデータを入手した。同ツールは，ウェブで展開されるブログ，twitter，Facebook，電子掲示板，News などの様々なソーシャルメディアから，情報を収集することを可能とする。ソーシャルメディアデータの収集量は，現時点で国内最大級を誇り，ブログについてはブログ事業者を中心に41サイト以上を対象として1日あたり約100万記事を収集している。twitter に関しては，1日あたり約3,000万記事を収集しており，これは日本のツイートデータの約30%に相当する規模である。Facebook については，公開されているウォールの投稿記事を対象に1日あたり約10万記事，掲示板については，国内主要掲示板サイトの投稿記事から1日あたり約55,000記事を収集している。また，News に関しては，国内主要ニュースサイトの記事情報120サイト以上から，1日あたり約14,000記事を収集している。この分析ツールを使用すれば，指定した検索ワードの出現数をカウントし，検索期間の指定で前日から過去に遡って13か月間に限り，日次，週次，月次の時系列データを生成することができる。

ただし，既述のとおり，本稿ではブログのみを分析対象にした。twitter や Facebook は，個人の意見や意向がタイムリーに反映しやすいと考えられるものの，Facebook についてはウォールに一般公開されている部分に限られ，また，twitter についてはリツイート機能の安易な利用で重要な意味を持つことなく拡散するケースがあり，検索キーワードの出現数で分析する際は結果の意味づけが困難になることが予想された。そもそも，獲得したい真の検索キーワードの特定が難しい。twitter の場合，その機能の性質上，ユーザーの思いつきの言葉が掲載されることが多いなど，必ずしも厳密な言葉使いで利用されとは限らない。例えば，日経平均株価に関するコメントであっても，“株価”や“株”などの略語で表現されるケースが多いと想定され，単純なキーワードの出現数では真の実態に迫ることが難しいと考えられた。また，News と掲示板については同分析ツールによる情報収集数が少ないため，バイアスがかかりやすいと判断した。以上の点を考慮して，本稿ではブログのみを対象にデータの収集を行い，分析用の時系列データを構築した。

データの具体的な構築方法は次のとおりである。まず，日経平均株価に関するソーシャル

メディアデータ（以下、株価ブログデータと称す）は、単純に日次あたりの出現数を用いた。ただし、日経平均株価は、様々な略称で用いられることが多い。このため、本稿では一般的によく用いられる略称と考えられる“日経平均株価”，“日経平均”，“日経225”を検索キーワードとしてデータの収集を進めた。

一方、為替レート（米ドル／円）に関するソーシャルメディアデータについては、対米ドルの日本円の為替動向を示す場合によく使われる言葉である“円高”と“円安”を検索キーワードにしてデータの収集を行った。もっとも、円高、円安は、必ずしも対米ドルに対して使用されるとは限らないが、わが国の場合は一般的に対米ドルの日本円の動きに対して使用されることが多いと考えられるため、これを期待して検索キーワードとして採用した。ただし、“円高”及び“円安”の出現数をそのまま分析に使用せず、渡辺（2012）に従い、以下の計算式に従って算出された時系列データを用いた（以下、為替ブログデータと称す）。

$$W_t = \frac{L_t}{L_t + H_t} \quad (1)$$

上記（1）式の、 L_t は t 期における“円安”の出現数、 H_t は t 期における“円高”の出現数、右下の添字 t は日次を示す。したがって、この式で算出された W_t は、“円高”及び“円安”の出現数を母数とした場合の“円安”の出現率を示す。実際の分析には株価ブログデータ及び為替ブログデータともに対数値を用いた。

なお、データ収集期間は、2012年2月6日～2013年2月28日を対象にした。本稿の使用データは全て東京証券取引所の休業日を除く営業日ベースに統一して抽出したため、サンプル数は264日分の264件である。

2.2 使用データの基本特性

構築したデータの単純な時系列グラフと基本統計量は図1及び表1のとおりである。基本的な傾向としては、日経平均株価と株価ブログデータは、一定の相関性があるようにみえる。また、為替レート（米ドル／円）と為替ブログデータにも相関性があるようにみえる。また、基本統計量のJarque-Beraを確認すると、使用データは基本的に正規分布に従っている可能性は低いことがわかる。

次に、対象データの時系列的な性質を検証するために単位根検定を行う。検定方法は、ADF（Augmented Dicky-Fuller）検定に基づいて行った。拡張項の次数は、最大次数を15としてシュワルツバイズ情報基準（SBIC: Schwartz Bayesian information Criterion）を用いて同定した。検定結果は、表2のとおりである。ADF検定を2次階差まで実施したところ、1%有意水準において全てのデータが1次の和分過程（I(1)）に従っていることが明らかになった。

図1 使用データの時系列推移

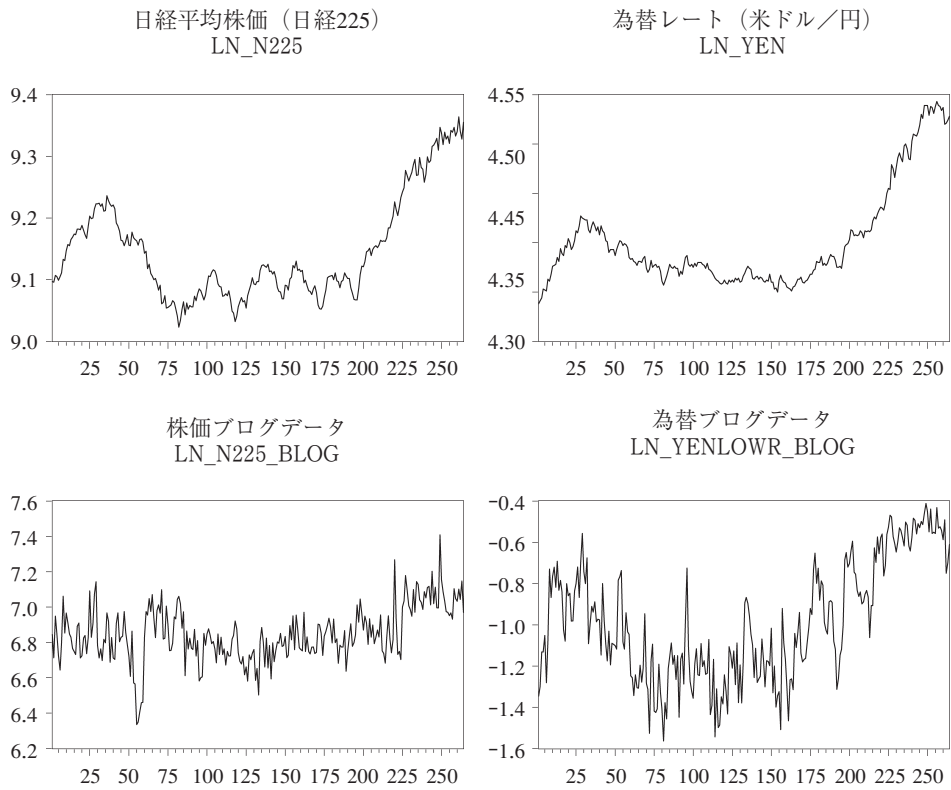


表1 基本統計量

	日経平均株価 (日経225) LN_N225	株価ブログ データ LN_N225_BLOG	為替レート (米ドル/円) LN_YEN	為替ブログ データ LN_YENLOWR_BLOG
Mean	9.1470	6.8466	4.4015	-0.9851
Median	9.1155	6.8290	4.3810	-1.0360
Maximum	9.3640	7.4080	4.5430	-0.4120
Minimum	9.0230	6.3350	4.3380	-1.5630
Std. Dev.	0.0831	0.1546	0.0512	0.2876
Skewness	0.9423	0.0247	1.5063	0.1915
Kurtosis	2.9289	3.8632	4.2176	1.9828
Jarque-Bera	39.123	8.224	116.137	12.995
Probability	0.000	0.016	0.000	0.002
Observations	264	264	264	264

表 2 ADF (Augmented Dicky-Fuller) 検定

① レベル

変数	Model I			Model II			Model III		
	ラグ次数	t 値	P-値	ラグ次数	t 値	P-値	ラグ次数	t 値	P-値
LN_N225	0	1.385	0.96	0	0.204	0.97	0	-0.428	0.99
LN_N225_BLOG	3	0.092	0.71	1	-5.255	0.00	1	-5.814	0.00
LN_YEN	0	2.160	0.99	0	0.476	0.99	0	-0.287	0.99
LN_YENLOWR_BLOG	1	-1.211	0.21	1	-3.117	0.03	1	-3.623	0.03

② 1 次階差

変数	Model I			Model II			Model III		
	ラグ次数	t 値	P-値	ラグ次数	t 値	P-値	ラグ次数	t 値	P-値
D(LN_N225)	0	-16.560	0.00	0	-16.655	0.00	0	-16.764	0.00
D(LN_N225_BLOG)	0	-23.449	0.00	0	-23.406	0.00	0	-23.360	0.00
D(LN_YEN)	0	-16.081	0.00	0	-16.328	0.00	0	-16.402	0.00
D(LN_YENLOWR_BLOG)	0	-22.314	0.00	0	-22.281	0.00	0	-22.238	0.00

③ 2 次階差

変数	Model I			Model II			Model III		
	ラグ次数	t 値	P-値	ラグ次数	t 値	P-値	ラグ次数	t 値	P-値
D2(LN_N225)	7	-11.182	0.00	7	-11.160	0.00	7	-11.148	0.00
D2(LN_N225_BLOG)	9	-11.223	0.00	9	-11.200	0.00	9	-11.175	0.00
D2(LN_YEN)	8	-10.354	0.00	8	-10.343	0.00	8	-10.324	0.00
D2(LN_YENLOWR_BLOG)	9	-10.927	0.00	9	-10.909	0.00	9	-10.890	0.00

(備考)

- 1) 変数欄の LN_N225 は日経平均株価 (日経225), LN_N225_BLOG は株価ログデータ, LN_YEN は為替レート (米ドル/円), LN_YENLOWR_BLOG は為替ログデータをそれぞれ示す。また, D (・) は 1 次階差, D2 (・) は 2 次階差を示す。
- 2) t 値は, Dicky-Fuller 分布に従う。
- 3) Model I は, ドリフト項及びトレンド項なし, Model II は, ドリフト項ありでトレンド項なし, Model III は, ドリフト項及びトレンド項ありを示す。具体的には, 以下のとおりである。ただし, y は検定対象の変数を示し, μ はドリフト項, t はトレンド項, u は誤差項, β, γ, δ はパラメーターを示す。

次に, 共和分検定を行う。本稿では, 次節において, 日経平均株価と株価ログデータ, 為替レート (米ドル/円) と為替ログデータのそれぞれの変数間の因果関係分析を行う。このため, ここでは日経平均株価と株価ログデータの共和分検定と, 為替レート (米ドル/円) と為替ログデータの共和分検定を行う。この分析によって各変数間の長期均衡関係の有無を確認することができる。検定方法は, Johansen (1988) が開発したトレース検定 (Trace 検定) 及び最大固有値検定 (L-max 検定) を採用した。

表 3 は, 共和分検定の結果である。ここでは, データ系列及び共和分ベクトルの特性に応じて 5 種類のモデルを想定して検定を行った。この結果をみると, 全てのモデル形式において 1% 有意水準で有意な結果が得られている。すなわち, 日経平均株価と株価ログデータ, 為替レート (米ドル/円) と為替ログデータのいずれの系列間においても共和分関係が検

出された。

$$\text{Model I : } \Delta y_t = \beta y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + u_t \quad H_0: \beta=0, H_A: \beta < 0 \quad (2)$$

$$\text{Model II : } \Delta y_t = \mu + \beta y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + u_t \quad H_0: \beta=0, H_A: \beta < 0 \quad (3)$$

$$\text{Model II : } \Delta y_t = \mu + \delta t + \beta y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + u_t \quad H_0: \beta=0, H_A: \beta < 0 \quad (4)$$

表3 Johansen の共和分検定：日経平均株価と株価ログデータ

Model I				
Eigenvalues	null	alternative	L-max	Trace
0.09528	$r=0$	$r=1$	25.833 **	27.317 **
0.00573	$r \leq 1$	$r=2$	1.484	1.484
Model II				
Eigenvalues	null	alternative	L-max	Trace
0.09768	$r=0$	$r=1$	26.519 **	28.006 **
0.00575	$r \leq 1$	$r=2$	1.487	1.487
Model III				
Eigenvalues	null	alternative	L-max	Trace
0.09766	$r=0$	$r=1$	26.512 **	26.514 **
0.00001	$r \leq 1$	$r=2$	0.002	0.002
Model IV				
Eigenvalues	null	alternative	L-max	Trace
0.10423	$r=0$	$r=1$	28.399 **	30.703 *
0.00889	$r \leq 1$	$r=2$	2.304	2.304
Model V				
Eigenvalues	null	alternative	L-max	Trace
0.10422	$r=0$	$r=1$	28.395 **	28.770 **
0.00145	$r \leq 1$	$r=2$	0.375	0.375

(備考)

- ・最大固有値検定統計量 (L-max) とトレース検定統計量 (Trace) の右付け印は、P 値に基づく検定結果であり、“***” は、1 % 有意、“**” は、5 % 有意を示す。
- ・モデルタイプの意味は以下のとおりである。
 Model I ; 共和分ベクトルに定数項及びトレンド項なし，系列はランダムウォーク
 Model II ; 共和分ベクトルに定数項あり，トレンド項なし，系列はランダムウォーク
 Model III ; 共和分ベクトルに定数項あり，トレンド項なし，系列はドリフト付ランダムウォーク
 Model IV ; 共和分ベクトルに定数項及びトレンド項あり，系列はドリフト付ランダムウォーク
 Model V ; 共和分ベクトルに定数項及びトレンド項あり，系列はドリフト及びトレンドあり

表 4 Johansen の共積分検定：為替レート（米ドル／円）と為替ブログデータ

Model I				
Eigenvalues	null	alternative	L-max	Trace
0.07939	$r=0$	$r=1$	21.341 **	23.246 **
0.00736	$r \leq 1$	$r=2$	1.905	1.905
Model II				
Eigenvalues	null	alternative	L-max	Trace
0.13108	$r=0$	$r=1$	36.251 **	38.164 **
0.00739	$r \leq 1$	$r=2$	1.913	1.913
Model III				
Eigenvalues	null	alternative	L-max	Trace
0.12533	$r=0$	$r=1$	34.548 **	34.576 **
0.00011	$r \leq 1$	$r=2$	0.028	0.028
Model IV				
Eigenvalues	null	alternative	L-max	Trace
0.12534	$r=0$	$r=1$	34.552 **	36.341 **
0.00691	$r \leq 1$	$r=2$	1.788	1.788
Model V				
Eigenvalues	null	alternative	L-max	Trace
0.12007	$r=0$	$r=1$	33.002 **	33.620 **
0.00239	$r \leq 1$	$r=2$	0.618	0.618

(備考)

- ・最大固有値検定統計量（L-max）とトレース検定統計量（Trace）の右付け印は、P 値に基づく検定結果であり，“**” は、1 % 有意，“*” は、5 % 有意を示す。
- ・モデルタイプの意味は以下のとおりである。
 Model I；共積分ベクトルに定数項及びトレンド項なし，系列はランダムウォーク
 Model II；共積分ベクトルに定数項あり，トレンド項なし，系列はランダムウォーク
 Model III；共積分ベクトルに定数項あり，トレンド項なし，系列はドリフト付ランダムウォーク
 Model IV；共積分ベクトルに定数項及びトレンド項あり，系列はドリフト付ランダムウォーク
 Model V；共積分ベクトルに定数項及びトレンド項あり，系列はドリフト及びトレンドあり

3 ソーシャルメディアデータと金融市場の因果関係分析

3.1 ベクトル誤差修正モデルの推定

本節では、ソーシャルメディアデータと金融市場の因果関係を分析する。具体的には、日経平均株価（日経225）と株価ブログデータの2変数間、為替レート（米ドル／円）と為替ブログデータの2変数間のそれぞれのグランジヤーの因果性テストとインパルス応答関数による分析を行う。既述のとおり、本稿で使用する4系列のデータは、単位根検定の結果、い

いずれも 1 次の和分過程 (I (1)) に従うことが明らかになった。また、共和分検定の結果、日経平均株価（日経225）と株価ログデータの 2 変数間、為替レート（米ドル／円）と為替ログデータの 2 変数間のいずれも共和分関係にあった。このため、ここでは、ベクトル誤差修正モデル（VECM: vector error-correction models）を採用し、関数の推定を行った。

なお、Johansen の共和分検定を行った際に、共和分ベクトル及びデータ系列の構造に応じて 5 種類のモデルタイプを想定して検定を行ったが、どのモデルタイプが最もあてはまりが良いかについて赤池の情報量基準（AIC: Akaike information Criterion）、シュワルツベイズ情報基準（SBIC: Schwartz Bayesian information Criterion）を用いて確認した。その結果、日経平均株価（日経225）と株価ログデータの 2 変数間では、共和分ベクトルに定数項及びトレンド項なし、データ系列はランダムウォークである Model I⁵⁾ が最もあてはまり具合が良好であった。このため、これら 2 変数を対象にした VECM の推定の際は、Model I タイプを仮定して推定した。為替レート（米ドル／円）と為替ログデータの 2 変数間については、共和分ベクトルに定数項あり、トレンド項なし、データ系列はランダムウォークである Model II⁶⁾ が最も良好であったため、このモデルタイプを仮定して推定した。

また、ラグ次数の同定については、最大次数を 5 として赤池の情報量基準（AIC: Akaike information Criterion）を用いて同定した。その結果、日経平均株価（日経225）と株価ログデータの 2 変数間の VECM のラグ次数は 1 次となり、為替レート（米ドル／円）と為替ログデータの 2 変数間の VECM のラグ次数は 2 次となった。

3.2 グランジャーの因果性

次に、推定した VECM を用いてグランジャーの因果性テストを行う。検定結果は、表 5 のとおりである。表中の数値は Wald 検定統計量であるが、これをみると、日経平均株価（日経225）と株価ログデータの 2 変数間では有意水準 10% において、グランジャーの意味での因果関係がみられなかった。他方、為替レート（米ドル／円）と為替ログデータの 2 変数間では、為替ログデータから為替レート（米ドル／円）に対してはグランジャーの意味での因果関係がみられなかったが、為替レート（米ドル／円）から為替ログデータに対しては有意水準 1 % においてグランジャーの意味での強い因果関係がみられた。

3.3 インパルス応答関数による分析

3.2 節の分析で変数間の因果関係が明らかになったが、グランジャーの因果性テストでは影響の期間や方向性はわからない。そこで、ここではインパルス応答関数の推定を通じて、変数間の影響の方向性と期間を明らかにする。分析結果は、図 2 のとおりである。

日経平均株価（日経225）と株価ログデータの 2 変数間では、グランジャーの因果性テ

表5 グランジャーの因果性

○日経平均株価における Granger の因果性テスト

		原因変数	
		日経平均株価（日経225） D(LN_N225)	株価ブログデータ D(LN_N225_BLOG)
結果変数	日経平均株価（日経225） D(LN_N225)	—	0.684
	株価ブログデータ D(LN_N225_BLOG)	0.462	—

○為替レートにおける Granger の因果性テスト

		原因変数	
		為替レート（米ドル／円） D(LN_YEN)	為替ブログデータ D(LN_YENLOWR_BLOG)
結果変数	為替レート（米ドル／円） D(LN_YEN)	—	0.068
	為替ブログデータ D(LN_YENLOWR_BLOG)	105.259**	—

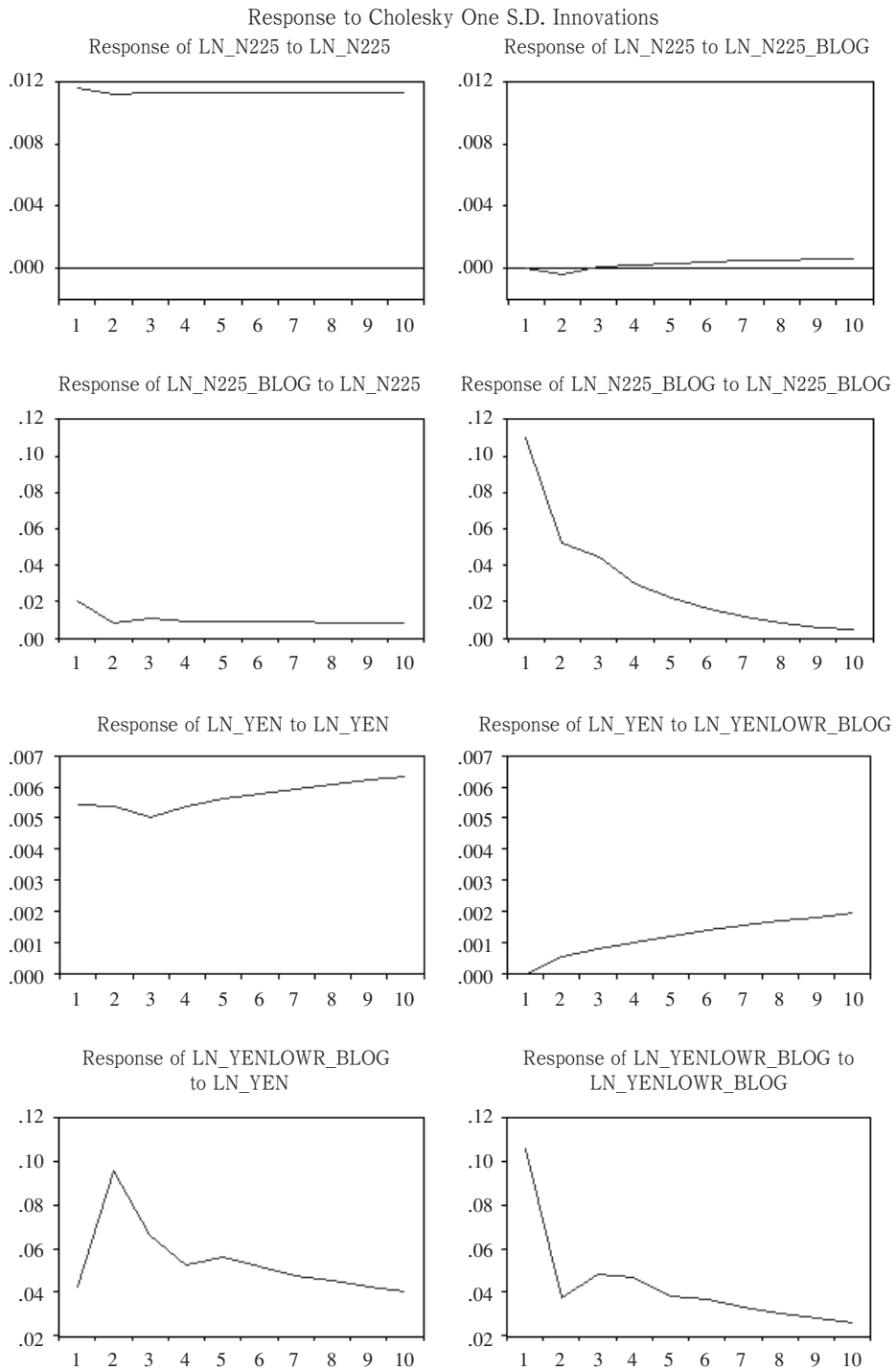
(備考)

- 1) 表中の数値は、Wald 検定統計量を示す。
- 2) 数値の右付け記号は，“***” が1%有意，“**” が5%有意，“+” が10%有意を示す。
- 3) D(LN_N225) は、日経平均株価（日経225）の対数値の1次階差、
D(LN_N225_BLOG) は、日経平均株価関連キーワードの日次出現数の対数値の一次階差を示す。
D(LN_YEN) は、為替レート（米ドル／円）の対数値の1次階差、
D(LN_YENLOWR_BLOG) は、“円安”の日次出現率（＝円安／（円安＋円高））の対数値の一次階差を示す。

ストの結果でも明らかになったように互いに有意な影響を及ぼさない。ただし、株価ブログデータには自己相関性がみられた。ショックから2～3日程度までは比較的強い影響を及ぼすが、日を迫うごとにその影響は減衰することが明らかになった。

一方、為替レート（米ドル／円）と為替ブログデータの2変数間では、グランジャーの因果性テストの結果、為替ブログデータは為替レート（米ドル／円）に対して影響を及ぼさず、為替レート（米ドル／円）の動きは為替ブログデータに強い影響を与えていることがわかった。図2をみると明らかなように、その影響の方向性は正相関性がみられる。また、為替レート（米ドル／円）に対するショックが与えられると、概ね2営業日目に最も強く影響し、その後は減衰している。そして、概ね10営業日目程度でその効果は消滅することが明らかになった。また、為替ブログデータにも自己相関性がみられた。ショックから1日程度は比較的強い影響を及ぼすが、2日目には大幅に低下する。その後は徐々に影響は減衰し、10営業日ほどでその影響がほぼ消滅することが明らかになった。

図2 インパルス応答



4 おわりに

本稿では、ソーシャルメディアデータと金融市場の関係性を探るべく、日経平均株価（日経225）と株価ブログデータ、為替レート（米ドル／円）と為替ブログデータを対象にグランジャーの因果性テストとインパルス応答関数による分析を行った。本稿で使用したデータは、単位根検定の結果、いずれも1次の和分過程（ $I(1)$ ）に従い、また、共和分検定を行ったところ、日経平均株価と株価ブログデータ、為替レートと為替ブログデータのそれぞれのデータ間について共和分関係が見いだされたため、ベクトル誤差修正モデル（VECM）を用いて分析を進めた。

その結果、ブログと株価市況に明確な因果関係がみられないことが明らかになった。また、ブログと為替市況の関係性については一定の因果関係がみとめられたものの、その因果の方向性から、為替市況の動きをもとに為替に関するコメントをブログに書き込んでいる可能性が高く、ブログ情報の動向が為替市況に影響を与えていないことが明らかになった。既述のとおり、ソーシャルメディアを株価の分析や予測に活用する試みが散見されるが、金融市場の分析に活用していく場合、本稿の分析で明らかになったようなデータ間の時系列特性に十分に配慮し、慎重なスタンスで分析を進める必要があろう。

ただし、本稿の分析では、極めてシンプルな方法でソーシャルメディアデータの生成を行っている。既述のとおりであるが、“日経平均株価”、“日経平均”、“日経225”と“円高”、“円安”を検索キーワードとして日次の出現数をカウントしたに過ぎない。為替レート（米ドル／円）の分析については、“円高”、“円安”というキーワードを使用することによって、ある程度市況の方向性を反映したブログデータを構築することができたと考えられるが、日経平均株価（日経225）の分析では、市況の方向性を示す言葉でキーワードの抽出を行っていないため、データ生成の段階で様々な問題が含まれていた可能性が高い。より厳密な分析を行うためにはテキストマイニングの技術を活用し、ソーシャルメディアの文脈から意味を見いだした上で必要なデータを抽出する工夫が必要であろう。情報学の分野では、金融市場におけるヘッドラインニュースなどを分析対象としてテキストマイニングを活用した研究が多数行われている。⁷⁾こうした先行研究を参考に、より発展的な研究を進める必要があろう。⁸⁾

注

- 1) ソーシャルメディアとは、ソーシャル・ネットワーキング・サービス（SNS: Social Networking Service）を通じて発信される多様な情報媒体のことを意味する。代表的な媒体としては、ブログ（Blog）や twitter, Facebook, 電子掲示板（BBS: Bulletin Board System）などがある。
- 2) 株式会社ホットリンクなどが株価予測システムの研究を進めている（内山ほか（2012）の第2章参照）。

- 3) チャットとは、インターネット上の対話型のリアルタイムコミュニケーションツールのことを意味する。
- 4) Yahoo! ファイナンス (URL: <http://finance.yahoo.co.jp/>)
- 5) 表3の備考欄を参照。
- 6) 表4の備考欄を参照。
- 7) Bloomberg やロイター, QUICK などの専用情報端末によってタイムリーに提供される情報を示す。
- 8) 高橋ほか (2007), 上瀧ほか (2009) など。

参 考 資 料

- Antweiler, W. and Frank, M. Z. (2004) "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards", *Journal of Finance*, Vol. 59, No. 3, pp. 1259-1294
- Dickey, D. A. and Fuller, W. A. (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, Vol. 49, No. 4, pp. 1057-1072
- Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, Vol. 12, Issue 2-4, pp. 231-254
- 浅井学 (2000) 「VAR モデルにおける共和分, ECM, 因果関係の分析」『立命館経済学』, 第48巻・第6号, pp. 41-59
- 上瀧弘晃, 高橋悟, 高橋大志 (2009) 「クレジット市場におけるヘッドラインニュースの効果」『人工知能学会全国大会論文集』, 23rd, ROMBUNNO.3G3-OS12-1
- 内山幸樹, 浅野弘輔, 榊剛史, セーヨー・サントイ, 田中勇樹 (2012) 「ブログ分析システム」, 高安美佐子編『ソーシャルメディアの経済物理学』第2章, 日本評論社, pp. 9-53
- 諏訪博彦, 梅原英一, 太田敏澄 (2012) 「ファクターモデルによるインターネット株式掲示板の投稿と株式リターンの分析」『情報処理学会論文誌』, Vol. 53, No. 1, pp. 117-125
- 高橋悟, 高橋大志, 津田和彦 (2007) 「株式市場におけるヘッドラインニュースの効果についての研究」, ファイナンス学会第15回大会論文, pp. 373-383
- 羽森茂之 (2000) 『計量経済学』中央経済社
- 丸山健, 梅原英一, 諏訪博彦, 太田敏澄 (2008) 「インターネット株式掲示板の投稿内容と株式市場の関係」『証券アナリストジャーナル』, Vol. 46, No. 11・12, pp. 110-127
- 丸山健, 梅原英一, 諏訪博彦, 太田敏澄 (2009) 「インターネット株式掲示板の投稿内容と株式市場の関連性」『人工知能学会研究会資料』, SIG-FIN-002-09
- 渡辺広太 (2012) 「ソーシャルメディアデータの応用分析」, 高安美佐子編『ソーシャルメディアの経済物理学』第5章, 日本評論社, pp. 144-171