

## 経済物理アプローチによる市場データの分析 Analysis of Market Data based on Econophysics Approach

我妻 勇貴<sup>†</sup>  
Yuki Wagatsuma

岩崎 清斗<sup>‡</sup>  
Kiyoto Iwasaki

大久保 誠也<sup>†</sup>  
Seiya Okubo

斉藤 和巳<sup>†</sup>  
Kazumi Saito

### 1. はじめに

現在、さまざまな時系列データを解析することにより、種々の事象の原因や関係を明らかにする研究が、さかんに行われている。そのような時系列データの例として、株価があげられる。株価は、政治・経済情勢、発行企業の業績や市場の需給等を反映し常に変動し、社会や経済の動きを表す鏡とされ、日経平均株価などの指標は景気の動向を表す1つの尺度として用いられている。その分析には、経済物理アプローチに基づく多くの研究成果が報告されている [1]。経済の動向を表す指標は、株価変動以外にも為替レートや物価、金利などさまざまである。これらの系列データは相互に影響しながら変動していることが知られている。一方の時系列が、他方の時系列にどれほど影響を与えるかを分析することは、経済動向の予測などのタスクにおいて重要な研究課題である。また、同様の手法を他のデータに適用することで、関係性を定量的に評価したり、従来は知られていなかった関係が明らかになることが期待される。

本稿では、農産物(野菜)の品目別取扱実績の金額、数量、平均価格のそれぞれを異種時系列として扱い分析する。本分析では、時系列間の類似尺度には、ピアソン相関係数とコサイン類似度の2つの定義を採用し、類似度定義の違いが、分析に与える影響についても調査する。評価実験では、15年間の月別市場データを用いる。特に、平均価格と数量の異種時系列ペアに焦点を絞り、類似度定義の違いによる特徴などを評価する。

### 2. 分析法

本稿では、次のようにして異種時系列の関係の強さを分析する。 $N$ 品目に対して、時系列ペア  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  と  $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$  が入力データとして与えられたとする。具体的には、品目  $n$  に対し、 $\mathbf{x}_n$  は価格の時系列、 $\mathbf{y}_n$  は数量の時系列などを表すとする。いま、時刻  $t$  での品目  $n$  の価格を  $x_{n,t}$  とし、時刻  $t$  は  $1, \dots, T$  をとるとすれば、Mantegna の研究 [1] と同様に、時系列ペア  $\mathbf{x}_n$  と  $\mathbf{y}_n$  の類似度を次式のピアソン相関係数で定義することができる。

$$p(n) = \frac{\sum_{t=1}^T (x_{n,t} - \bar{x}_n)(y_{n,t} - \bar{y}_n)}{\sqrt{\sum_{t=1}^T (x_{n,t} - \bar{x}_n)^2 \sum_{t=1}^T (y_{n,t} - \bar{y}_n)^2}} \quad (1)$$

ここで、 $\bar{x}_n$  と  $\bar{y}_n$  は、次式のように、 $\mathbf{x}_n$  と  $\mathbf{y}_n$  のそれぞれの平均値を表す。

$$\bar{x}_n = \frac{1}{T} \sum_{t=1}^T x_{n,t}, \quad \bar{y}_n = \frac{1}{T} \sum_{t=1}^T y_{n,t} \quad (2)$$

一方、機械学習やテキストマイニングの分野 [2] では、次式により定義される、コサイン類似度も頻繁に利用されている。

$$c(n) = \frac{\sum_{t=1}^T x_{n,t} y_{n,t}}{\sqrt{\sum_{t=1}^T x_{n,t}^2 \sum_{t=1}^T y_{n,t}^2}} \quad (3)$$

本稿では、異種時系列の関係の強さの分析を行う際に、式 (1) と (3) の類似度定義の違いが、実験結果にどのような影響を与えるかを実証評価する。その際、各異種時系列を次式のように正規化し、プロットすることにより比較評価する。

$$\bar{x}_{n,t} = \frac{x_{n,t} - \min_{1 \leq t \leq T} \{x_{n,t}\}}{\max_{1 \leq t \leq T} \{x_{n,t}\} - \min_{1 \leq t \leq T} \{x_{n,t}\}} \quad (4)$$

すなわち、 $0 \leq \bar{x}_{n,t} \leq 1$  を満たし、式 (1) と (3) のピアソン相関係数やコサイン類似度の定義に依存せず、異種時系列を比較するために式 (4) を採用する。

### 3. 実験による評価

本実験では、2002年1月から2016年12月までの期間における、品目別取扱実績(野菜)の数量、金額、平均価格の月別データを用いた<sup>§</sup>。ただし、この15年間のすべての月で欠損値のない品目に限定している。分析対象の野菜数は  $N = 137$  である。なお、この期間の月の総数は  $T = 180$  となる。

表1には、平均価格と数量の時系列に関するピアソン相関係数  $p$  とコサイン類似度  $c$  で上下5位となった野菜品目の比較を示す。ここで、 $r$  は順位を意味する。表1より、ピアソン相関係数  $p$  とコサイン類似度  $c$  のどちらを採用しても、上位については、「まいたけ」、「まめもやし」の同一野菜品目が現れるなど類似した傾向を示していることが分かる。これに対して、下位の品目については、同一品目は現れず、それぞれ異なる結果になっている。

図1に、ピアソン相関係数  $p$  とコサイン類似度  $c$  のそれぞれ最上位群と最下位群となった野菜品目に着目し、式 (4) で正規化した平均価格 (value) と数量 (quantity) の時系列の比較を示す。すなわち、それぞれの時系列は、最小値と最大値が0と1になるようにスケールしてプロットしている。図1(a)から(d)には、ピアソン相関係数  $p$  とコサイン類似度  $c$  で最上位群の品目となる「まいたけ」、「まめもやし」、「マッシュルーム」、及び「たらの芽」を示す。これらの平均価格と数量は比較的類似した時系列となっていることが分かる。これらに対して、図1(e)と(f)に示すピアソン相関係数  $p$  で最下位群の「いんげん」と「かぶ」では、平均

<sup>†</sup> 静岡県立大学 経営情報学部

<sup>‡</sup> 静岡県工業技術研究所 電子科

<sup>§</sup> <http://www.shijou-tokei.metro.tokyo.jp/>

表1: 平均価格と数量の時系列に関するピアソン相関係数とコサイン類似度での上下5位の比較

ピアソン相関係数		コサイン類似度			ピアソン相関係数		コサイン類似度		
$r$	$p$	野菜名	$c$	野菜名	$r$	$p$	野菜名	$c$	野菜名
1	0.704	まいたけ	0.997	まめもやし	133	-0.818	きゅうり	0.123	さんとうさい
2	0.551	まめもやし	0.987	マッシュルーム	134	-0.833	なす	0.122	うめ
3	0.518	たらの芽	0.985	まいたけ	135	-0.845	ぜんまい水煮	0.059	ほしだいこん
4	0.435	しめじ	0.971	なめこ	136	-0.845	いんげん	0.043	まつたけ
5	0.431	だいだい	0.970	セルリー	137	-0.846	かぶ	0.018	しそ

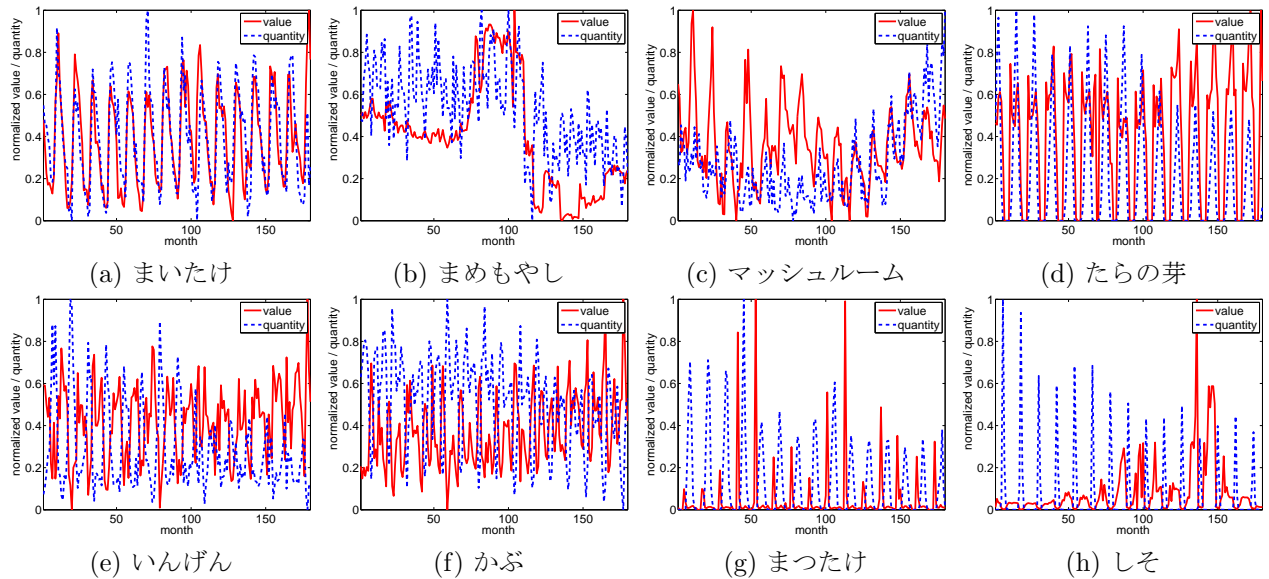


図1: 正規化した平均価格 (value) と数量 (quantity) の時系列の比較

価格と数量で負の相関を示しているのに対し、図1(g)と(h)に示すコサイン類似度  $c$  で最下位群の「まつたけ」と「しそ」では、数量が殆ど0のときに平均価格が跳ね上がるような傾向が見取れる。本実験より、類似度定義としてピアソン相関係数  $p$  とコサイン類似度  $c$  を採用した場合、上位については比較的類似した品目が選ばれる傾向となるものの、下位の品目については、それぞれの類似度定義の特性を反映した結果となることが観測された。

このように品目により類似度が異なる理由については、コサイン類似度とピアソン類似度の上位の「まめもやし」や「まいたけ」等の品目は、15年間を通してみても価格と数量共に調整が容易であることが示唆される。これに対し、コサイン類似度の下位の「しそ」や「まつたけ」は、数量と平均価格の調整が容易ではないことが示唆される。ただし、本考察の妥当性を検証するには、各野菜品目に対して、栽培や価格形成プロセスなどについての実態調査が必要である。

#### 4. おわりに

本稿では、農産物(野菜)の品目別取扱実績の金額、数量、平均価格のそれぞれを異種時系列として扱い分析した。また、時系列間の類似度定義として、ピアソ

ン相関係数とコサイン類似度の2つを用い、使用する類似度定義の違いが分析結果に与える影響についても調査した。15年間の月別市場データを用いた評価実験では、特に、平均価格と数量の異種時系列ペアに焦点を絞った分析を行った。その結果、上位については類似した品目が選ばれる傾向となるものの、下位の品目については、それぞれの類似度定義の特性を反映した結果となることが分かった。今後は、さらに多様な異種時系列データにおいて、提案分析法の有効性などを検証する予定である。

謝辞 本研究は、科学研究費補助金基盤研究(C)(No.15K00429)の助成を受けた。

#### 参考文献

- [1] Mantegna, R.N. (1999), Hierarchical structure in financial markets, Euro. Phys. J. B Vol.11, pp.193-197.
- [2] C. M. Bishop (2010), Pattern Recognition and Machine Learning (Information Science and Statistics), Springer.