

統数研「経済物理学とその周辺」@CGI
9/27-28,2012

**RMT利用による東証
4半期毎の主成分追跡
~業種分類の再考で見えるもの~**

鳥取大学大学院工学研究科
情報エレクトロニクス専攻
田中美栄子, 山本敦史

はじめに

大量データの採取、集積→統計分析



膨大なデータ間の相関関係

ランダム行列理論(RMT)
を用いた主成分分析(PCA)

株式、地震、気象、薬品、...

はじめに：株式投資

RMT-PCAを株式市場に適用

ランダムな変動をする銘柄をノイズとして除去

他銘柄と相関の高い**複数**銘柄が抽出される

業種の固まりによって主要な業種を調べる

銘柄選択の指標になる

参考文献

V. Plerou, P. Gopikrishnan, B. Rosenow,
L.A.N. Amarmal, T. Guhr, H.E. Stanley,
“Random matrix approach to cross correlation
in financial data”,
Physical Review E 65, 066126, 2002.

はじめに: ランダム行列理論

- ランダム行列の固有値分布が $Q(= L/N)$ で決まる

L : データ長, N : データ数

- 数値がランダムに並ぶ行列ならば相関行列の固有値分布が理論式と一致する

固有値分布式

$$P_{RMT}(\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}$$

はじめに: ランダム行列理論

$$\underline{N \rightarrow \infty, L \rightarrow \infty}$$

$$Q = \frac{L}{N} > 1$$

$$\lambda_{\pm} = \left(1 \pm \frac{1}{\sqrt{Q}} \right)^2$$

■ N : データ数

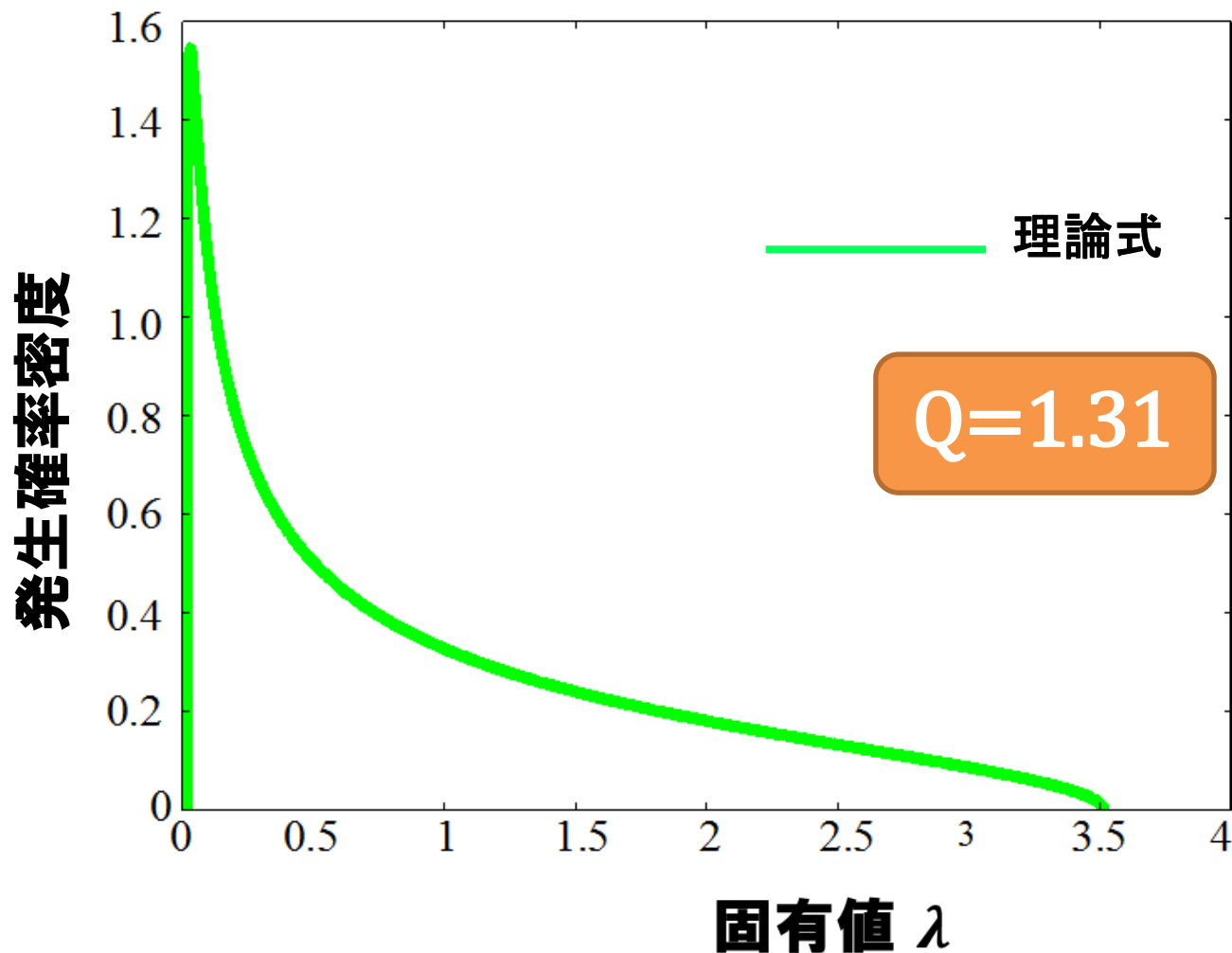
■ L : データ長

■ λ_{\pm} : 最大、最小固有値

固有値分布式

$$P_{RMT}(\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}$$

はじめに：固有値分布



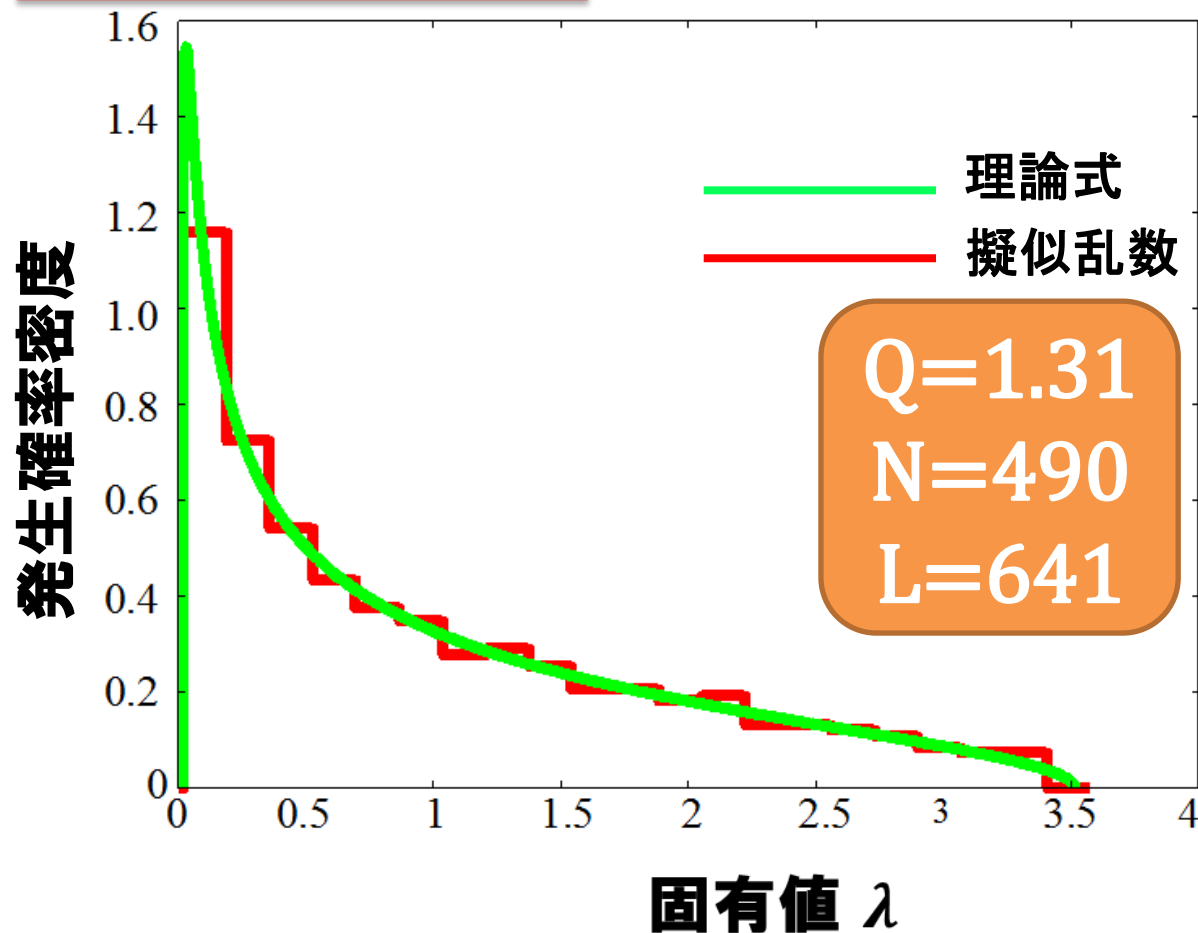
$$Q = \frac{L}{N} > 1$$

N : データ数

L : データ長

はじめに: 固有値分布

$N \rightarrow \infty, L \rightarrow \infty$

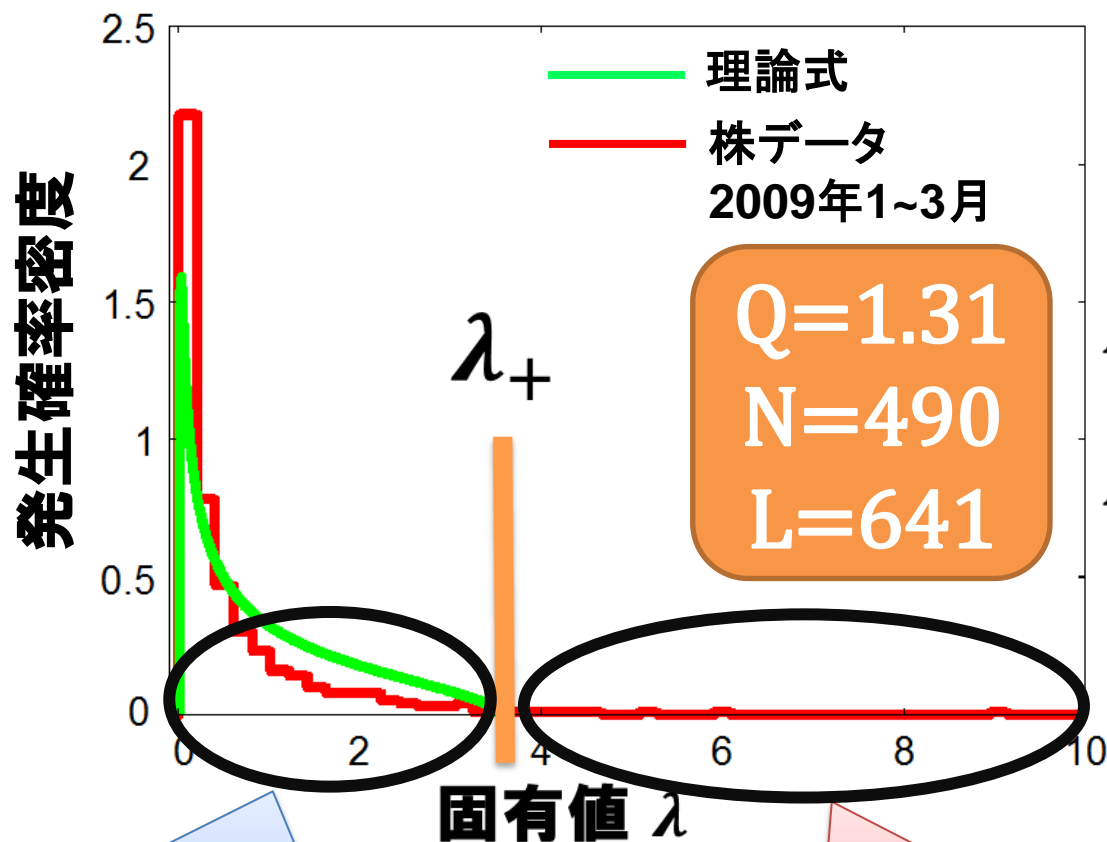


$$Q = \frac{L}{N} > 1$$

N : データ数

L : データ長

はじめに：固有値分布



$$Q = \frac{L}{N} > 1$$

N : データ数

L : データ長

ランダム成分

主成分

目的

■ 先行研究*

➡ 1年間毎の価格変動(1994-2009:16年分)

■ 短期間売買の増加 (TICKデータ)

➡ **短期間**の動向を知りたい

*木戸,楊,田中,高石,情報処理学会論文誌数理モデル化と応用,4巻4号,104-110頁, 2011

目的

■ 3ヶ月毎の主要業種の分析



細かい主要業種の変遷を見ることができる

■ 評価

経済動向と対比

(抽出データの有効性を証明)

手順

手順

1. データの抽出、整形
2. 正規化
3. 固有値の算出
4. 固有値の分離
5. 固有ベクトル成分の分析

使用データ

■ TOPIX500の構成銘柄(2007~2009)

東証市場第一部銘柄の中から、時価総額及び流動性の高い500銘柄で構成される株式群

手順(1): データの抽出、整形

Tick価格を**30分**毎に区切り**最初**の取引価格を抽出

時間	株価		
	銘柄1	銘柄2	銘柄3
09:00	390	200	300
09:30	390	200	290
10:00			
10:30	400	220	290
11:00	400	230	290
12:30	400	230	300

データが少なく揃わない銘柄は除外

一つ前の値を補正

全ての銘柄で取引が無い場合は考慮しない

各期間データ：整形後のL/N

年	月	L(時系列長)	N(銘柄数)	Q(L/N)
2007	1-3	642	486	1.32
	4-6	681	486	1.4
	7-9	681	489	1.39
	10-12	675	492	1.37
2008	1-3	642	488	1.32
	4-6	681	491	1.39
	7-9	692	492	1.41
	10-12	664	487	1.36
2009	1-3	641	490	1.31
	4-6	659	486	1.36
	7-9	681	485	1.4
	10-12	670	483	1.39

手順(2):正規化

■ 30分毎の価格の対数差を算出

■ $X_i(t) = \ln S_i(t+1) - \ln S_i(t)$

時刻 $t(= 1, 2, \dots, L)$ 銘柄 $i(= 1, 2, \dots, N)$

日付	t	銘柄1 $S_1(t)$	銘柄2 $S_2(t)$
4/20 09:30	1	300	30000
4/20 10:00	2	370	30070
4/20 10:30	3	390	30090
4/20 11:00	4	350	30050



t	銘柄1 $x_1(t)$	銘柄2 $x_2(t)$
1	0.2097	0.00233
2	0.0526	0.00065
3	-0.1082	-0.00133

変動量は同じでも
投資額によって違う

手順(2):正規化

$$g_i(t) = \frac{x_i(t) - \langle x_i \rangle}{\sigma}$$

時刻 $t (= 1, 2, \dots, L)$ 銘柄 $i (= 1, 2, \dots, N)$

t	銘柄1 $x_1(t)$	銘柄2 $x_2(t)$
1	0.2097	0.00233
2	0.0526	0.00065
3	-0.1082	-0.00133



t	銘柄1 $g_1(t)$	銘柄2 $g_2(t)$
1	1.219862	1.186411
2	0.009709	0.073369
3	-1.22957	-1.25978

手順(3): 固有値の算出

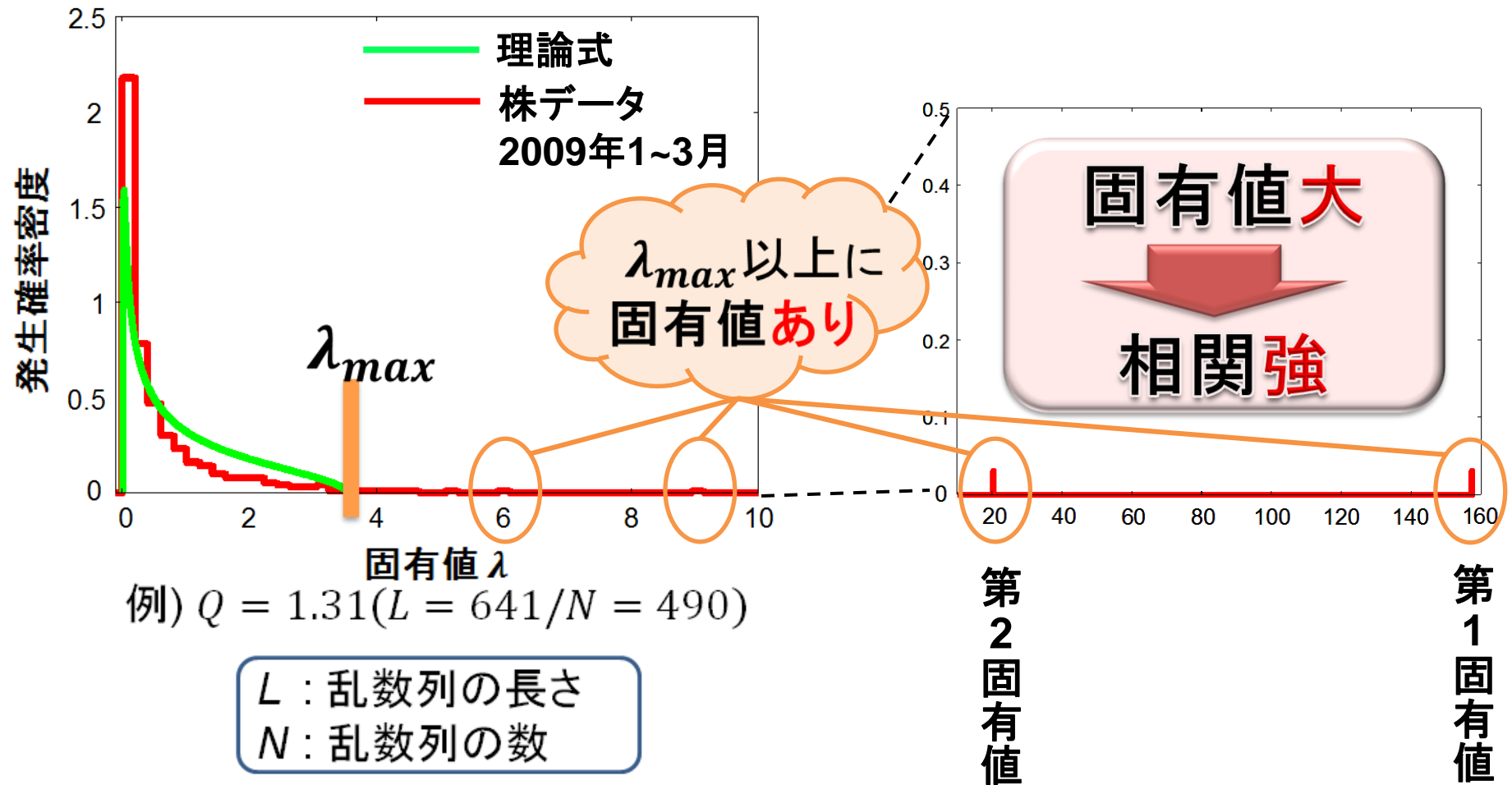
■ 正規化により作成した行列 G の自己相関行列

$$\mathbf{C} = \frac{1}{L} \mathbf{G}^T \mathbf{G}$$

時刻 $t (= 1, 2, \dots, L)$ 銘柄 $i (= 1, 2, \dots, N)$

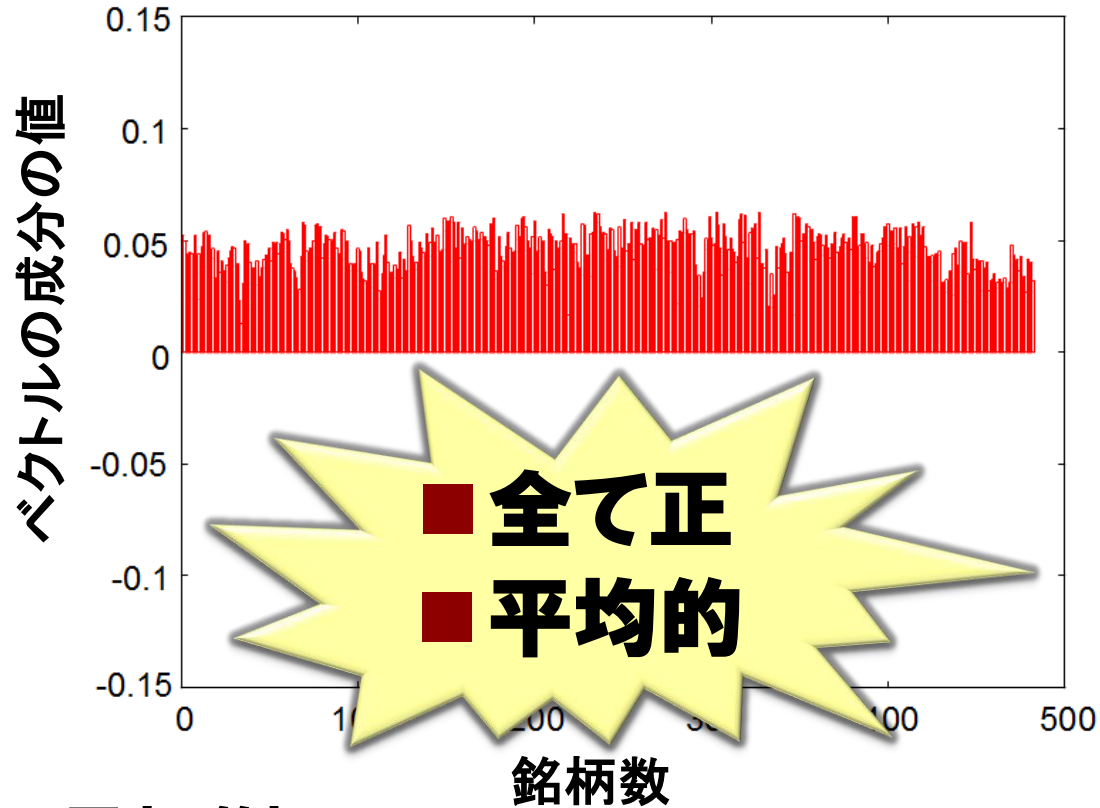
$$= \frac{1}{L} \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,L} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N,1} & g_{N,2} & \cdots & g_{N,L} \end{pmatrix} \begin{pmatrix} g_{1,1} & g_{2,1} & \cdots & g_{N,1} \\ g_{1,2} & g_{2,2} & \cdots & g_{N,2} \\ \vdots & \vdots & \ddots & \vdots \\ g_{1,L} & g_{2,L} & \cdots & g_{N,L} \end{pmatrix}$$
$$= \begin{pmatrix} 1 & g_{1 \cdot} g_{2 \cdot} & \cdots & g_{1 \cdot} g_{N \cdot} \\ g_{2 \cdot} g_{1 \cdot} & 1 & \cdots & g_{2 \cdot} g_{N \cdot} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N \cdot} g_{1 \cdot} & g_{N \cdot} g_{2 \cdot} & \cdots & 1 \end{pmatrix} \rightarrow \text{固有値}$$

手順(4): 固有値の分離



手順(5): 固有ベクトルの分析

■ 第1固有ベクトル成分

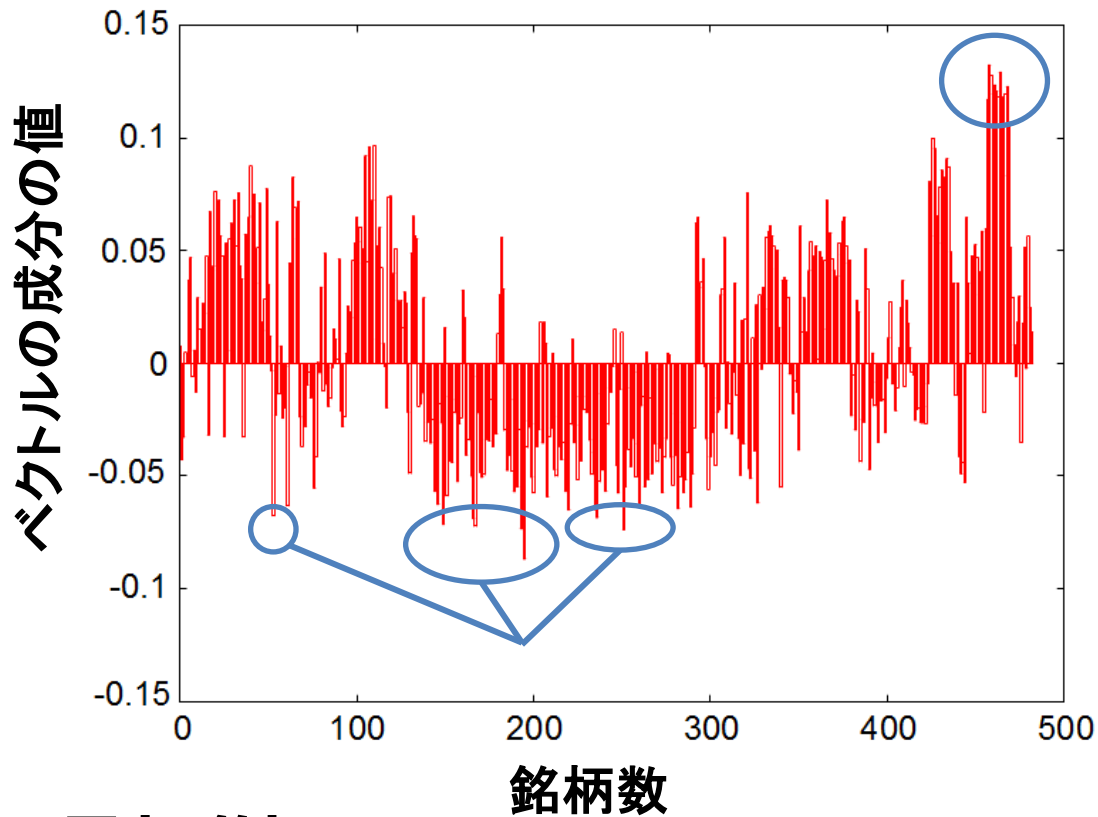


1要素1銘柄

手順(5): 固有ベクトルの分析

10銘柄ずつ抽出
降順と昇順のそれぞれ

■ 第2固有ベクトル成分



1要素1銘柄

10銘柄の内訳(09年1-3月)

金融、電力成分

コード	銘柄名
9502	中部電力
9508	九州電力
9503	関西電力
9505	北陸電力
9532	大阪ガス
9506	東北電力
9531	東京ガス
9504	中国電力
9507	四国電力
9509	北海道電力

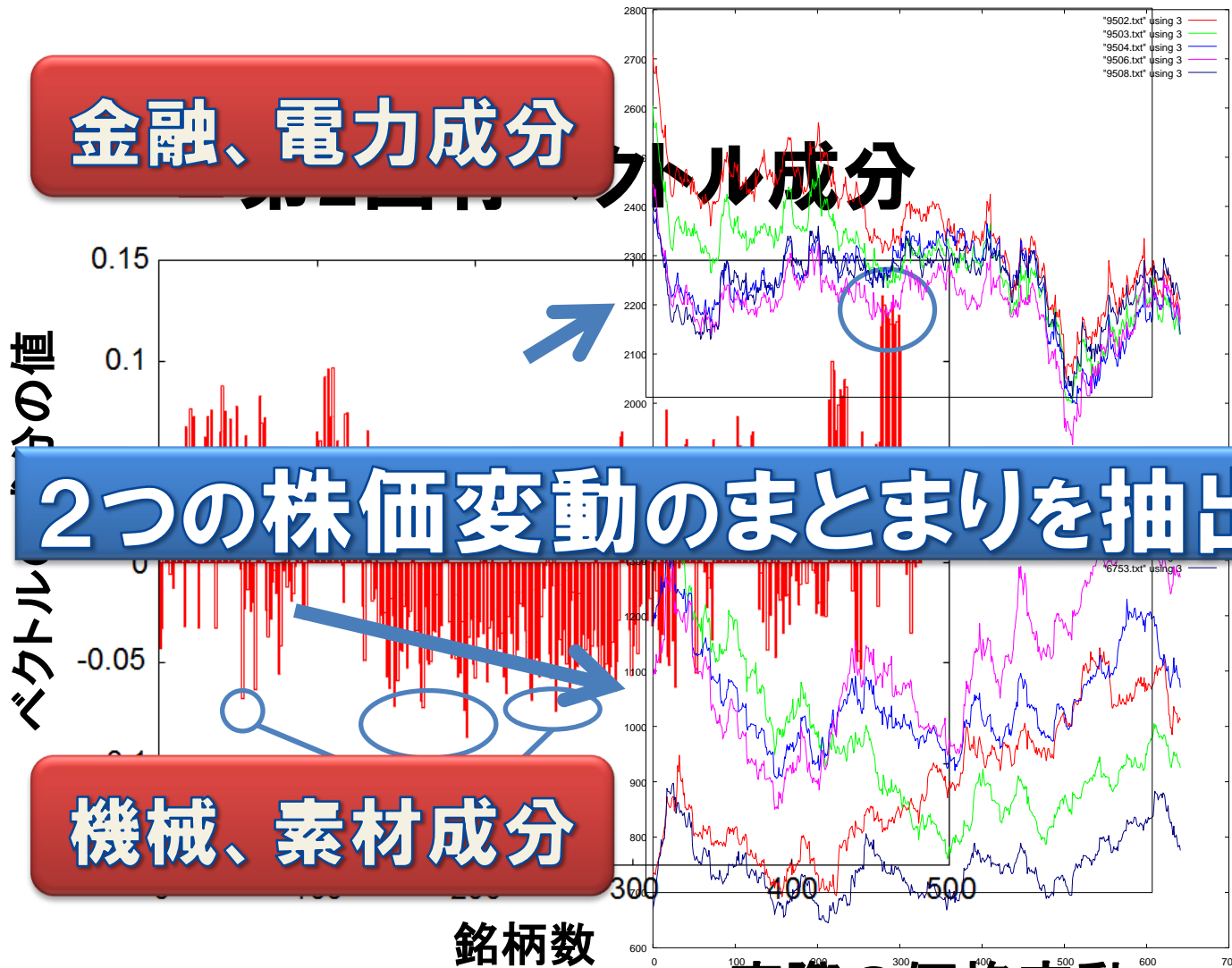
機械、素材成分

コード	銘柄名
6753	シャープ
6586	マキタ
3116	トヨタ紡織
6758	ソニー
5541	大平洋金属
5301	東海カーボン
5631	日本製鋼所
6301	コマツ
6902	デンソー
6305	日立建機

手順(5): 固有ベクトルの分析

金融、電力成分

ベクトル成分



2つの株価変動のまとまりを抽出

機械、素材成分

降順と昇順のそれぞれ
10銘柄ずつ抽出

分類名称	構成銘柄(既存の33業種)	銘柄数
20:食品	「13:水産・農林業」,「20:食料品」	23/500
50:エネルギー資源	「15:鉱業」,「50:石油・石炭製品」	8/500
17:建設・資材	「17建設業」,「52:ガラス・土石製品」,「59:金属製品」	32/500
30:素材・科学	「30:繊維製品」,「37:パルプ・紙」,「40:化学」	52/500
45:医薬品	「45:医薬品」	18/500
70:自動車・輸送機	「51:ゴム製品」,「70:輸送用機器」	29/500
54:鉄鋼・非鉄	「54:鉄鋼」,「57:非鉄金属」	26/500
60:機械	「60:機械」	32/500
65:電機・精密	「65:電気機器」,「77:精密機器」	64/500
94:情報通信・サービスその他	「78:その他製品」,「94:情報・通信業」,「96:サービス業」	48/500
80:商社・卸売	「80:卸売業」	19/500
81:小売	「81:小売業」	27/500
83:銀行	「83:銀行業」	42/500
85:金融(銀行除く)	「85:その他金融業」,「86:証券業」,「87:保険業」	30/500
88:不動産	「88:不動産業」	10/500
90:運輸・物流	「90:陸運業」,「91:海運業」,「92:空運業」,「93:倉庫・運輸関連業」	27/500
95:電力・ガス	「95:電気・ガス業」	13/500

分類法を変えてみる

■ TOPIX17

TOPIXの構成銘柄を、「証券コード協議会」が定める33業種をもとに、33業種を17業種に集約した株価指数。

17	建設・資材	20	食品	30	素材・科学	45	医薬品
50	エネルギー資源	54	鉄鋼・非鉄	60	機械	65	電機・精密
70	自動車・輸送機	80	商社・卸売	81	小売	83	銀行
85	金融	88	不動産	90	運輸・物流		
94	情報通信・サービスその他			95	電力・ガス		

10銘柄の内訳(09年1-3月)

金融、電力成分

銘柄名	業種
中部電力	電力・ガス
九州電力	電力・ガス
関西電力	電力・ガス
北陸電力	電力・ガス
大阪ガス	電力・ガス
東北電力	電力・ガス
東京ガス	電力・ガス
中国電力	電力・ガス
四国電力	電力・ガス
北海道電力	電力・ガス

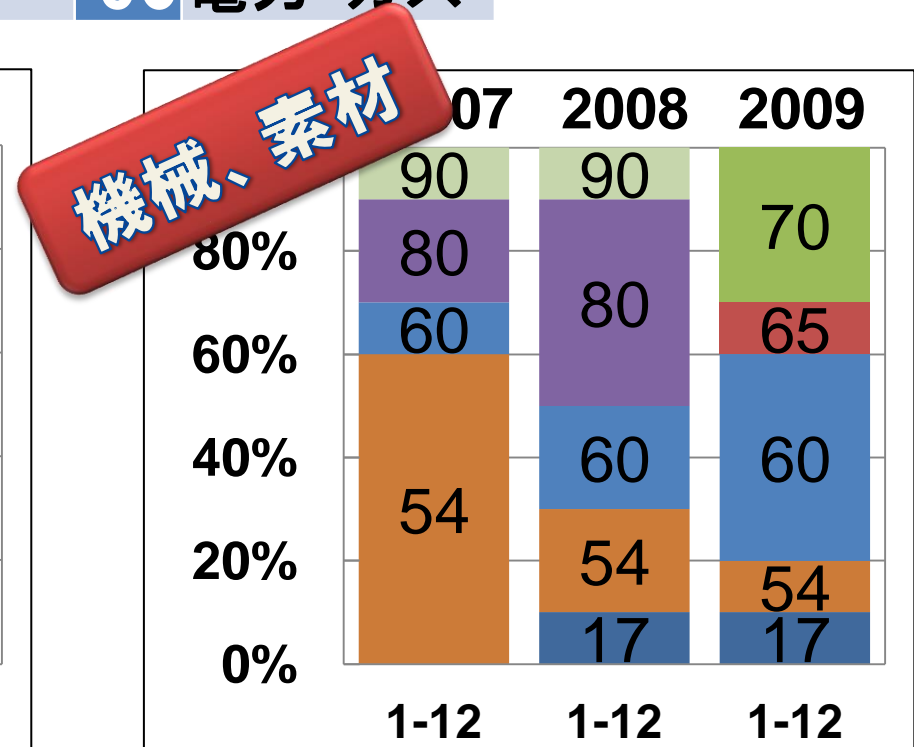
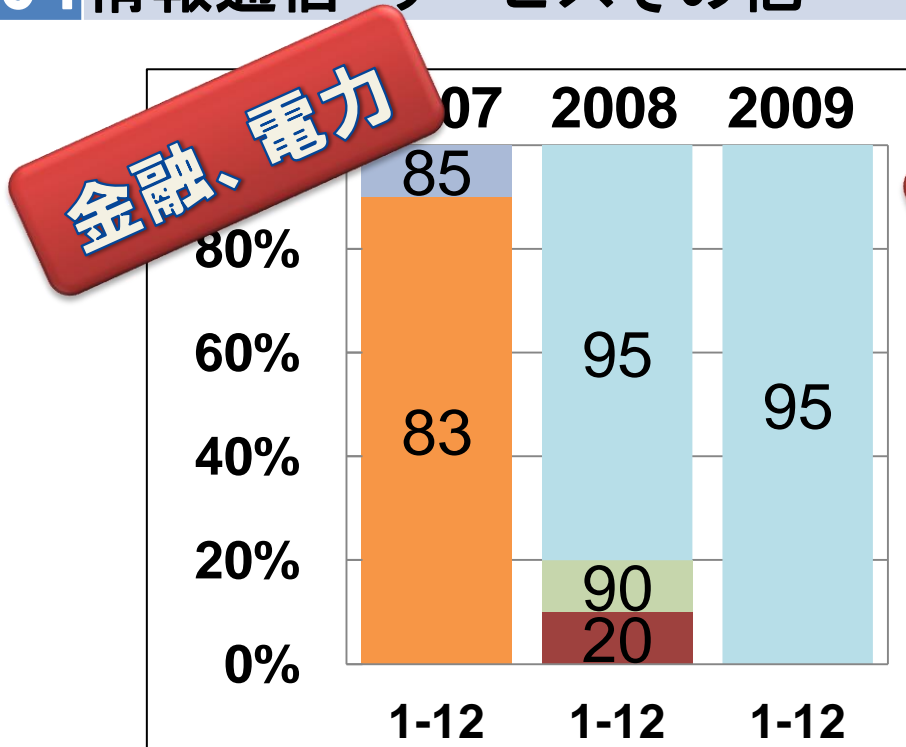
機械、素材成分

銘柄名	業種
マキタ	機械
日本製鋼所	機械
コマツ	機械
日立建機	機械
トヨタ紡織	自動車・輸送機
デンソー	自動車・輸送機
シャープ	電機・精密
ソニー	電機・精密
大太平洋金属	鉄鋼・非鉄
東海カーボン	建設・素材

結果

結果：1年毎のデータ(先行)

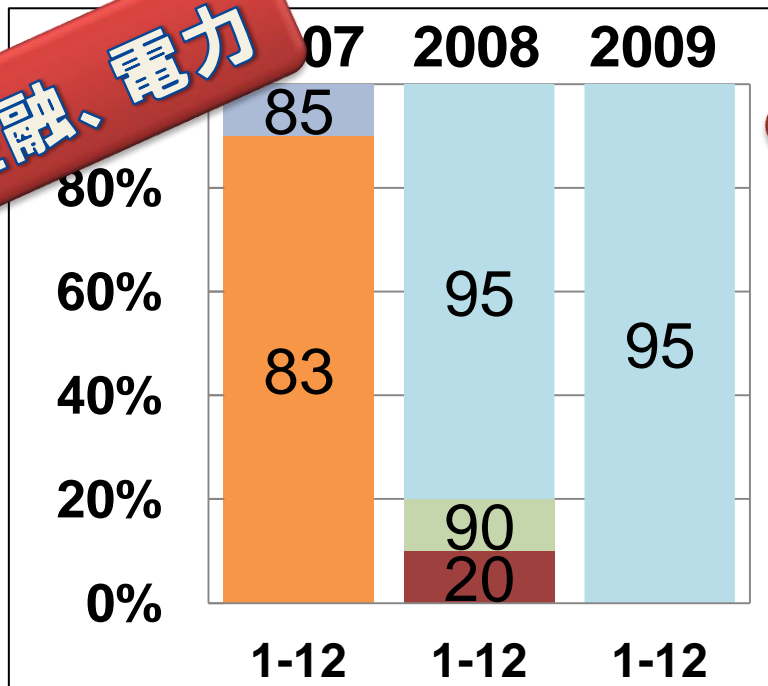
17	建設・資材	20	食品	30	素材・科学	45	医薬品
50	エネルギー資源	54	鉄鋼・非鉄	60	機械	65	電機・精密
70	自動車・輸送機	80	商社・卸売	81	小売	83	銀行
85	金融	88	不動産	90	運輸・物流		
94	情報通信・サービスその他			95	電力・ガス		



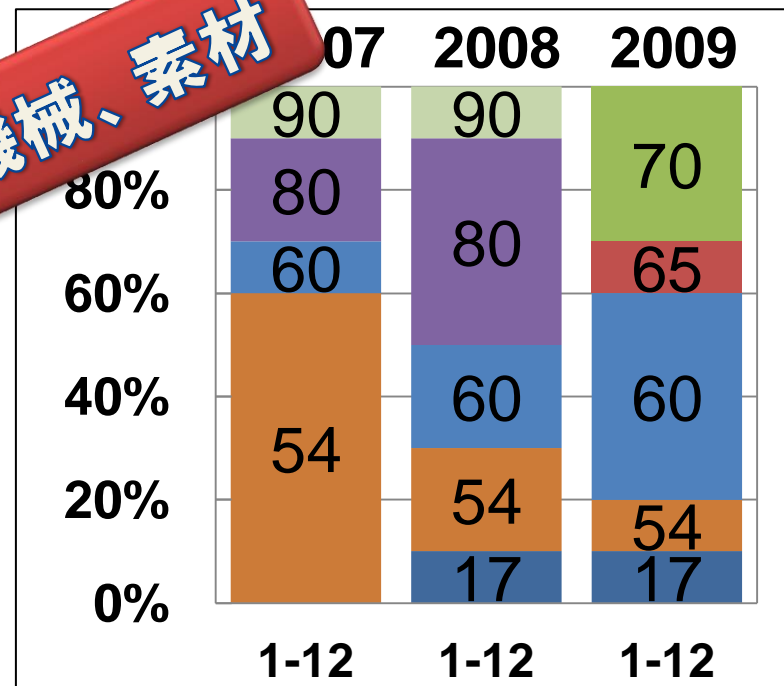
結果：3ヶ月毎

17	建設・資材	20	食品	30	素材・科学	45	医薬品
50	エネルギー資源	54	鉄鋼・非鉄	60	機械	65	電機・精密
70	自動車・輸送機	80	商社・卸売	81	小売	83	銀行
85	金融	88	不動産	90	運輸・物流		
94	情報通信・サービスその他			95	電力・ガス		

金融、電力

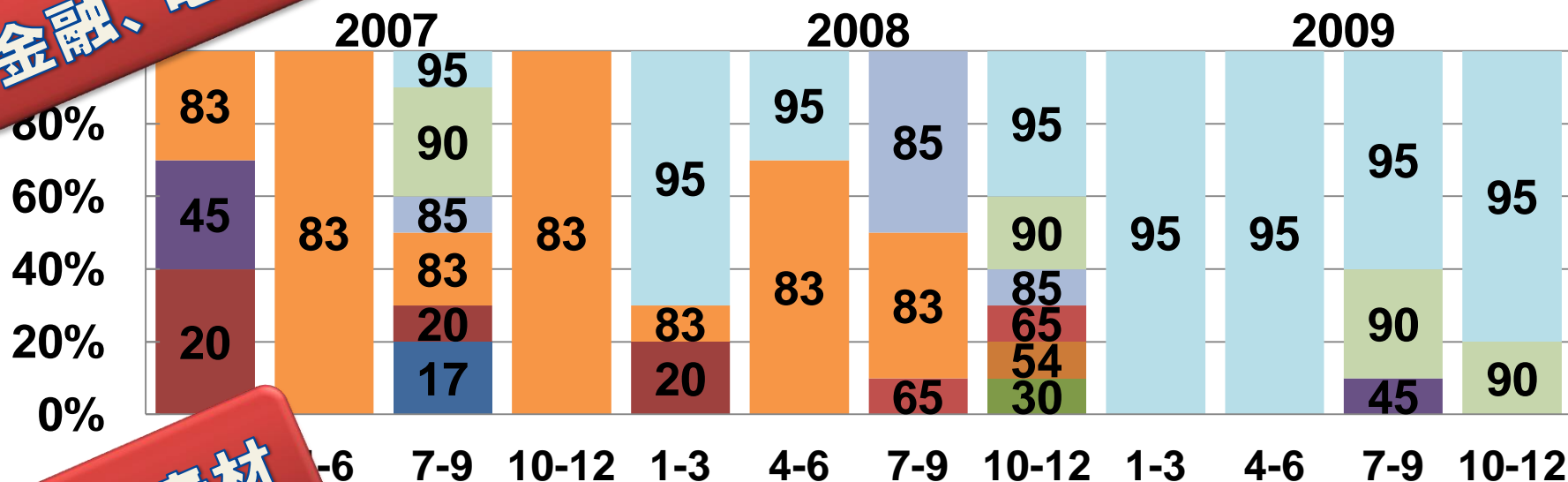


機械、素材

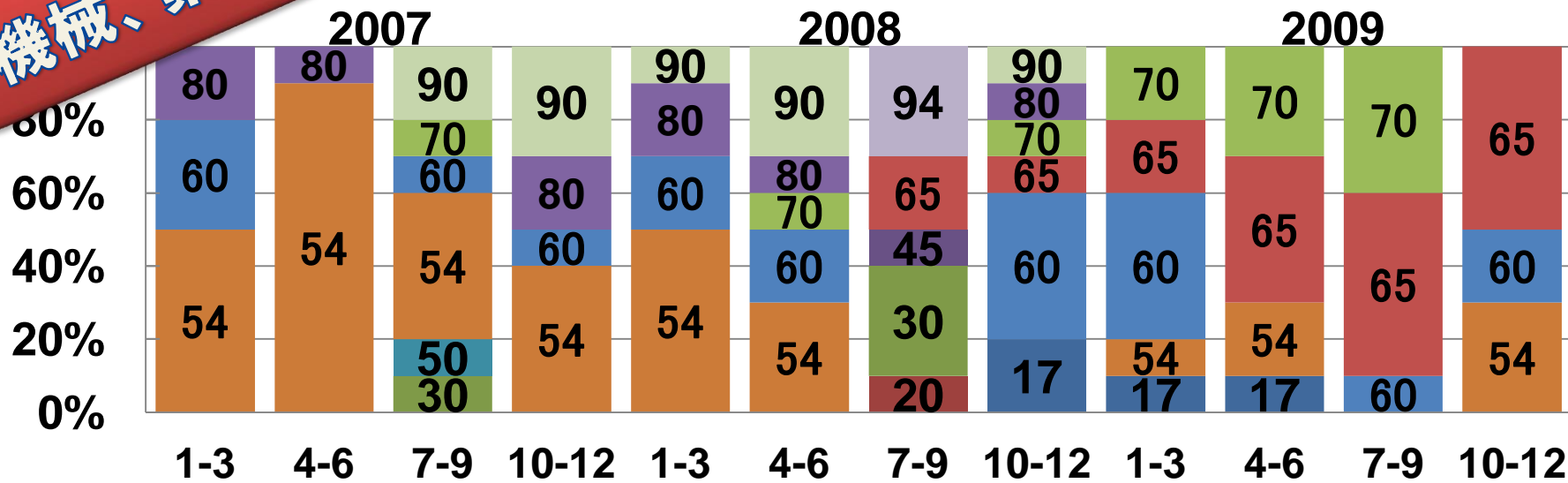


結果・3ヶ月毎の変遷(new)

金融、電力

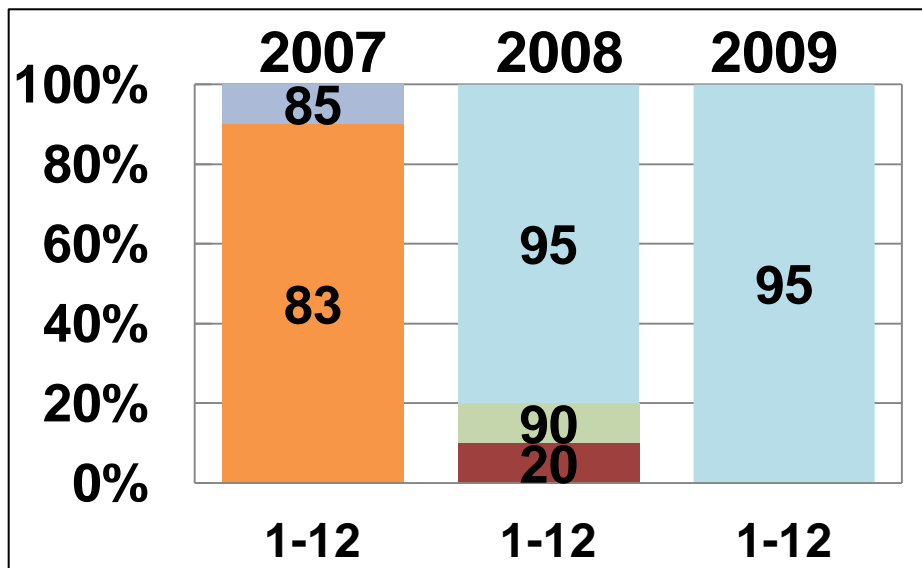
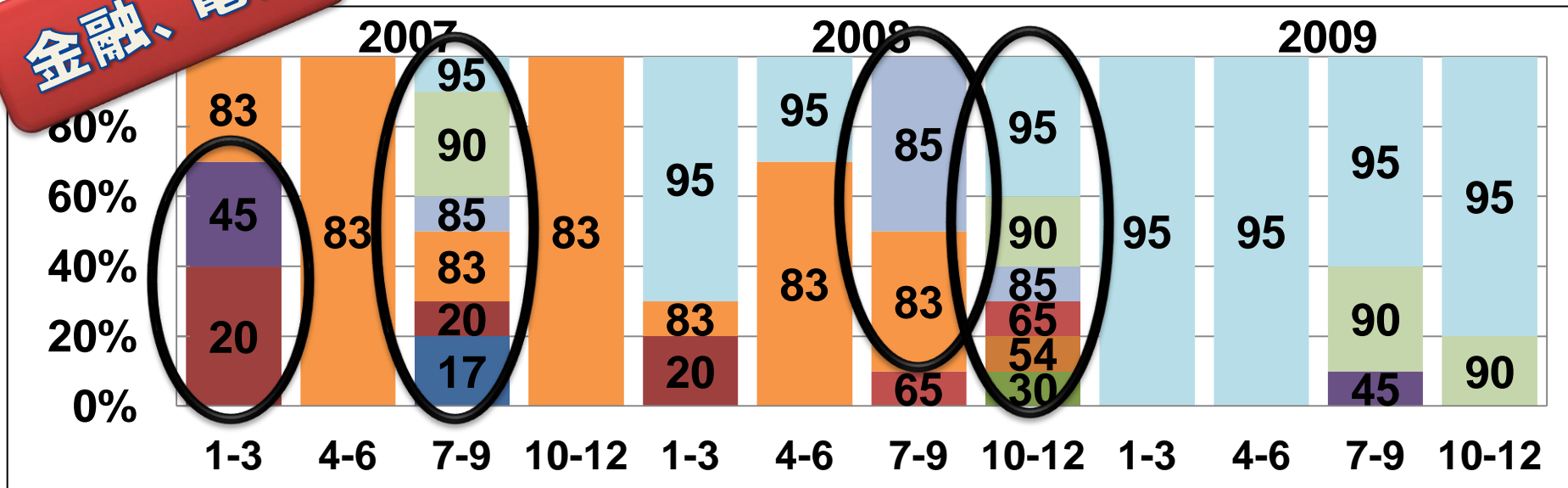


機械、素材



結果・3ヶ月毎(20,45,SP問題, RB倒産)

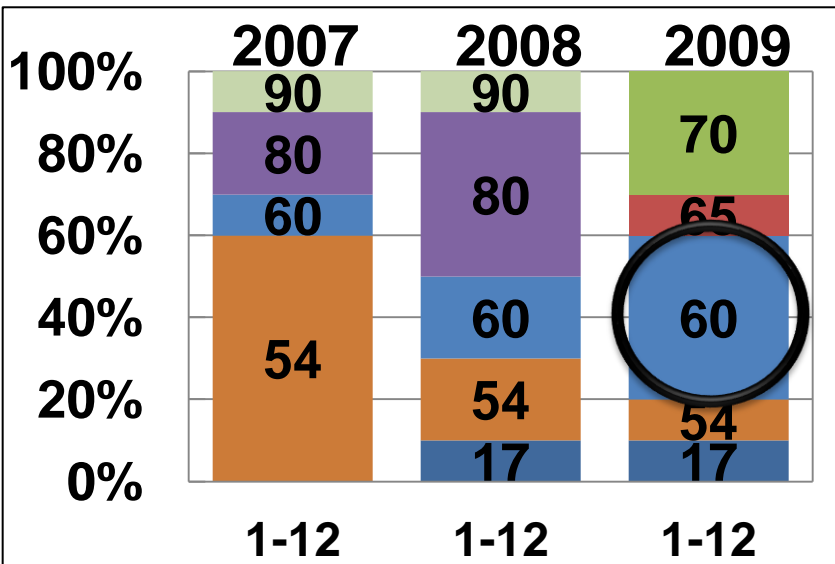
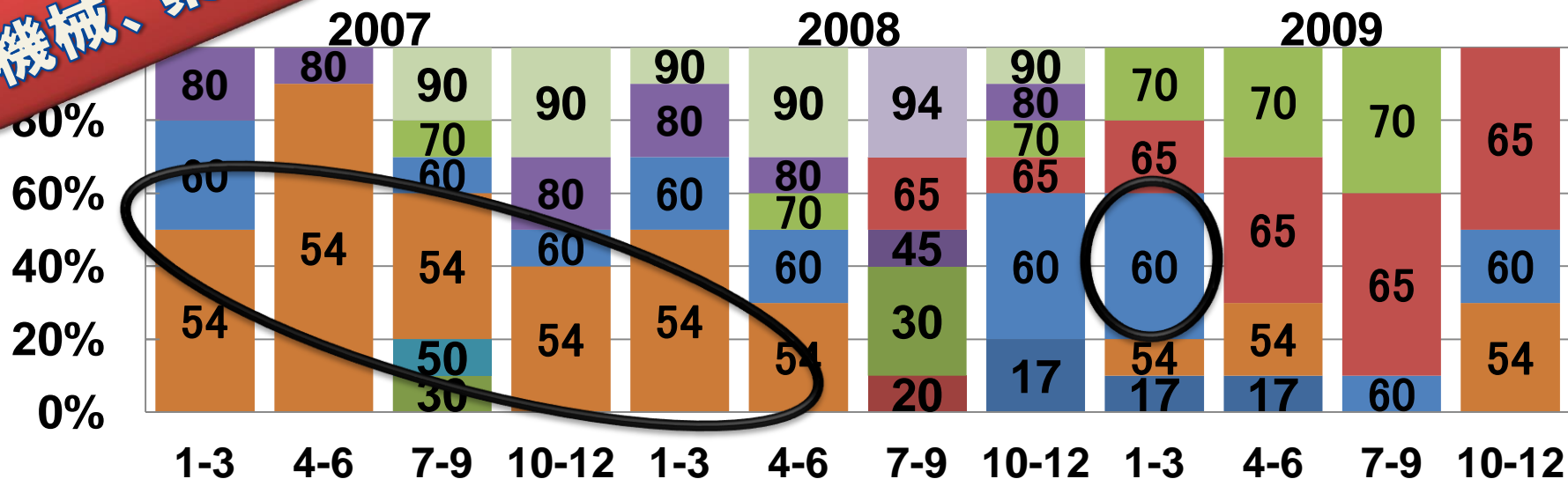
金融、電力



17	建設・資材	20	食品
30	素材・科学	45	医薬品
54	鉄鋼・非鉄	65	電機・精密
80	商社・卸売	83	銀行
85	金融	90	運輸・物流
95	電力・ガス		

結果・3ヶ月毎(2007.7-9, 2008.10-12)

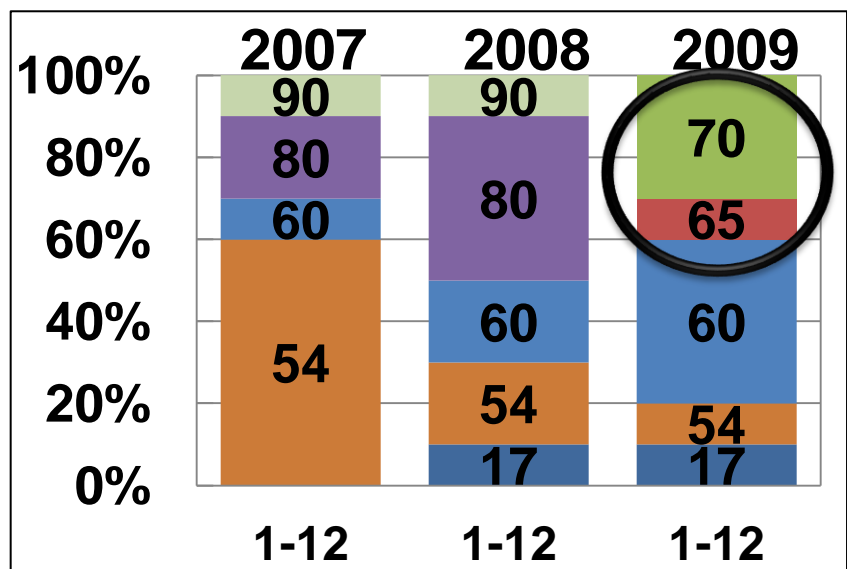
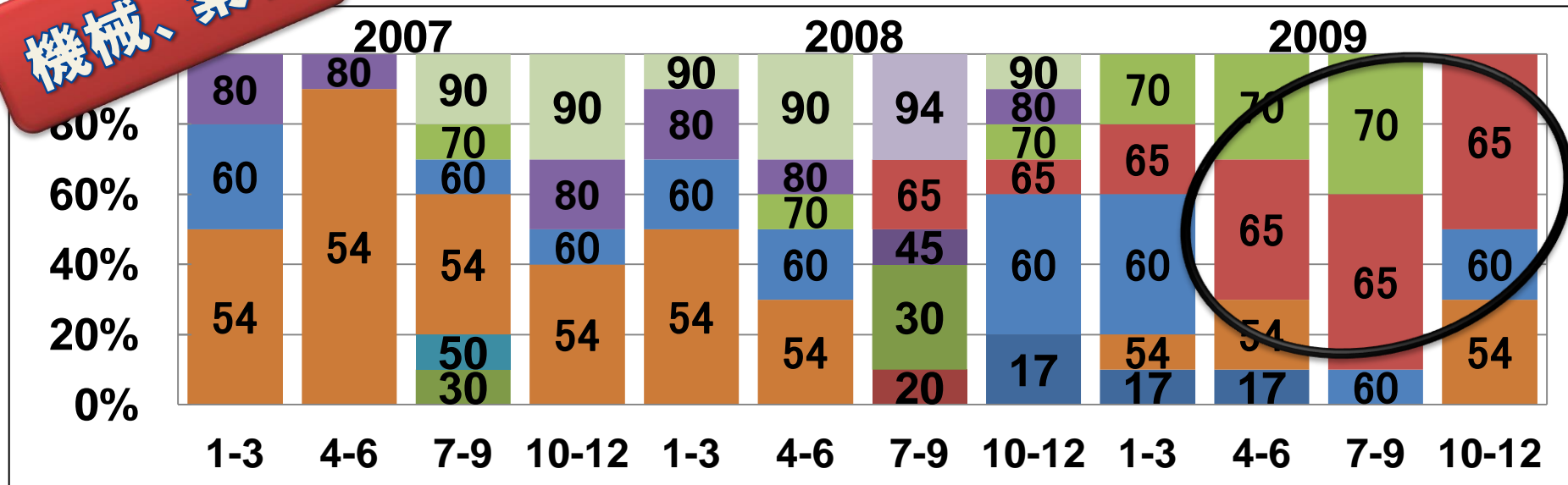
機械・素材



17	建設・資材	20	食品
30	素材・科学	45	医薬品
50	エネルギー資源	54	鉄鋼・非鉄
60	機械	65	電機・精密
70	自動車・輸送機	80	商社・卸売
90	運輸・物流		

結果・3ヶ月毎

機械・素材



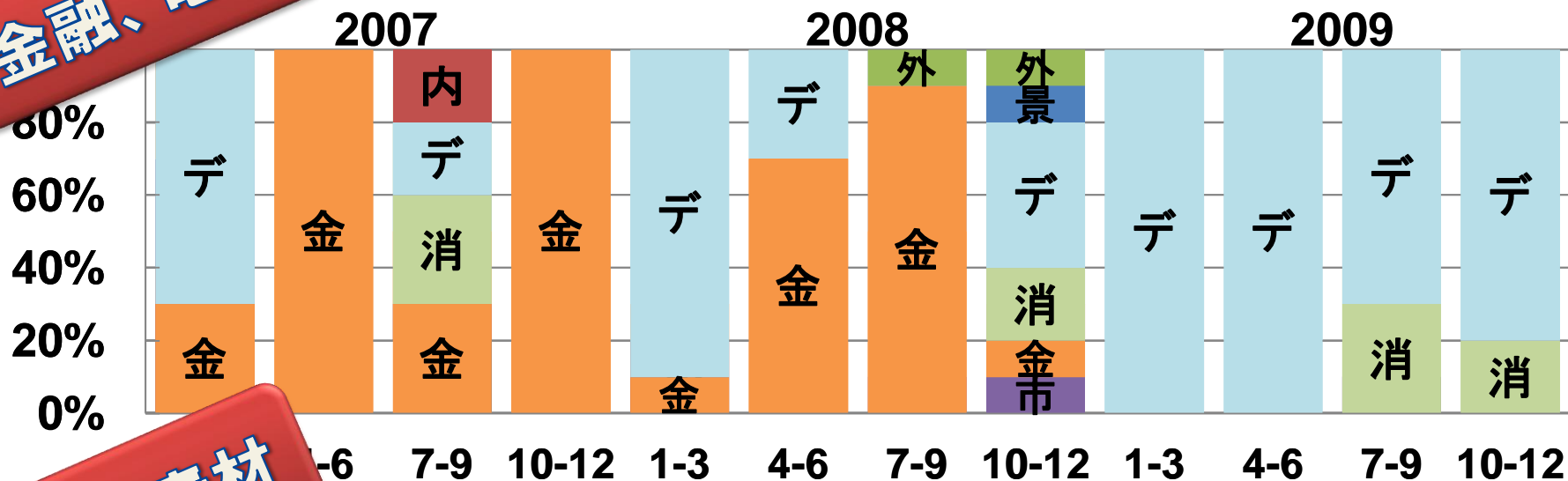
17	建設・資材	20	食品
30	素材・科学	45	医薬品
50	エネルギー資源	54	鉄鋼・非鉄
60	機械	65	電機・精密
70	自動車・輸送機	80	商社・卸売
90	運輸・物流		

関連株分類表で見直すと、

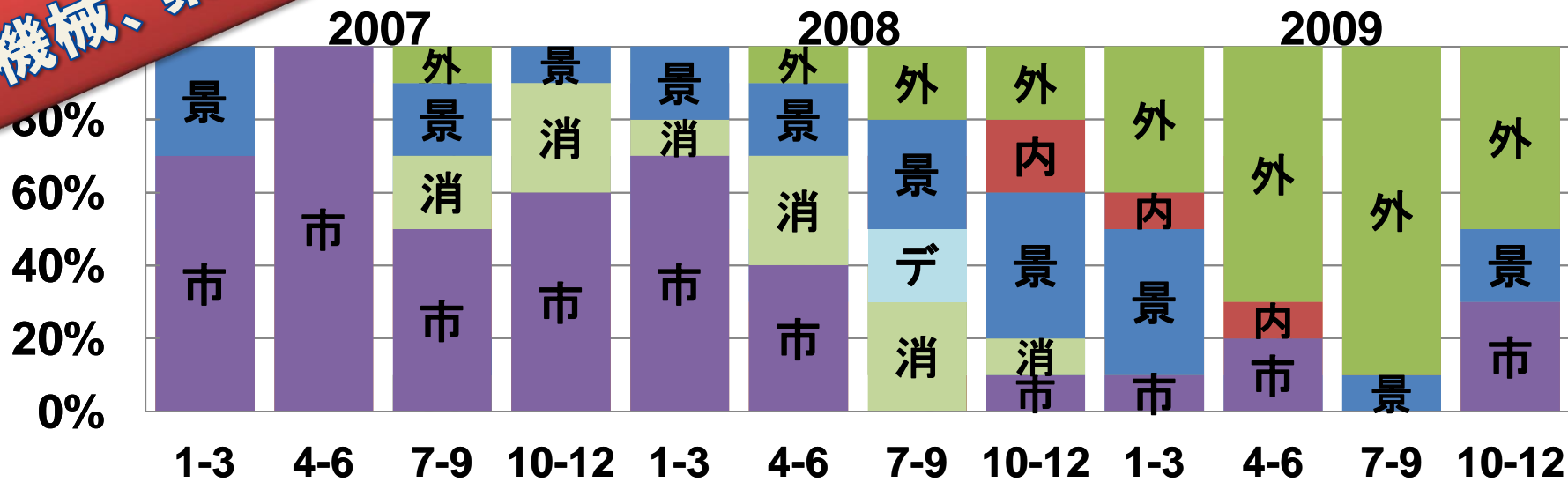
外需関連株	「65:電機・精密」,「70:自動車・輸送機」
内需関連株	「17:建設・資材」,「88:不動産」
景気循環株	「30:素材・化学」,「60:機械」
ディフェンシブ株	「20:食品」,「45:医薬品」,「95:電力・ガス」
消費関連株	「81:小売」,「90:運輸・物流」, 「94:情報通信・サービスその他」
金利敏感株	「83:銀行」,「85:金融」
市況関連株	「50:エネルギー資源」,「54:鉄鋼・非鉄」, 「80:商社・卸売」

結果・3ヶ月毎(関連株分類)

金融、電力

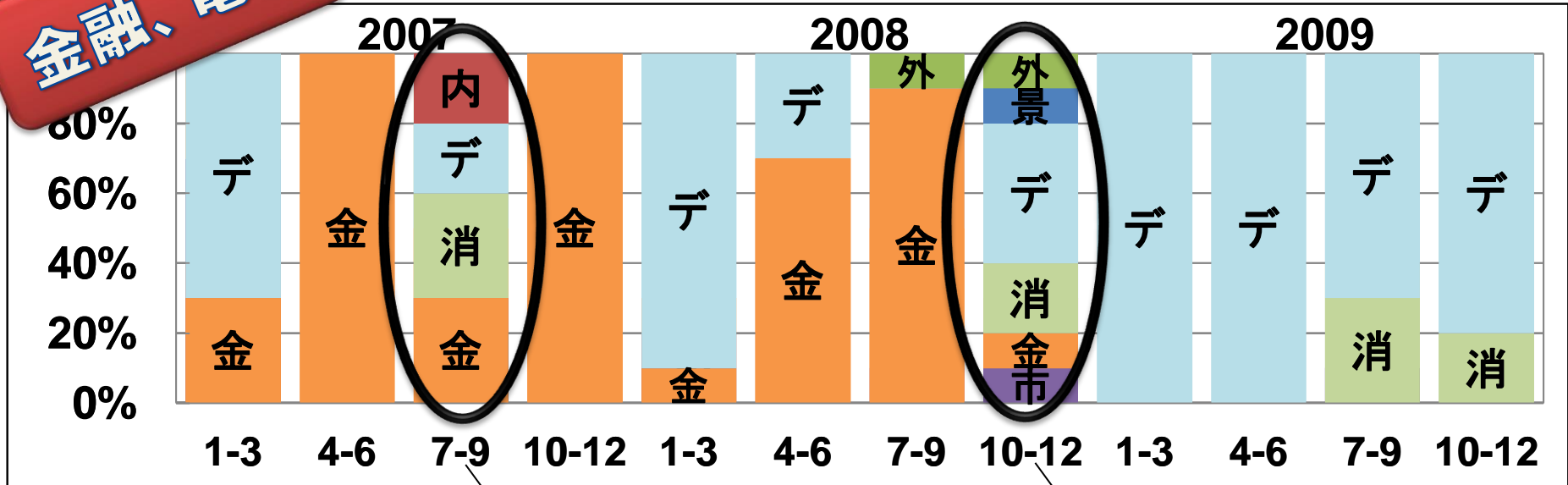


機械、素材



結果・3ヶ月毎(関連株分類)

金融、電力



8月 サブプライムローン

10月 リーマンショック

ディフェンシブ株

業績が景気の**影響を受けにくい**と言われる銘柄

消費関連株

消費者向けの事業を展開する銘柄

金利敏感株

金利の動向によって業績が大きく変わる銘柄

結果：3ヶ月毎(関連株分類)

外需関連株

国外での収益(輸出や現地の売上等)が影響する銘柄

内需関連株

事業基盤が国内にあり, 内需拡大にウエイトをおく銘柄

景気循環株

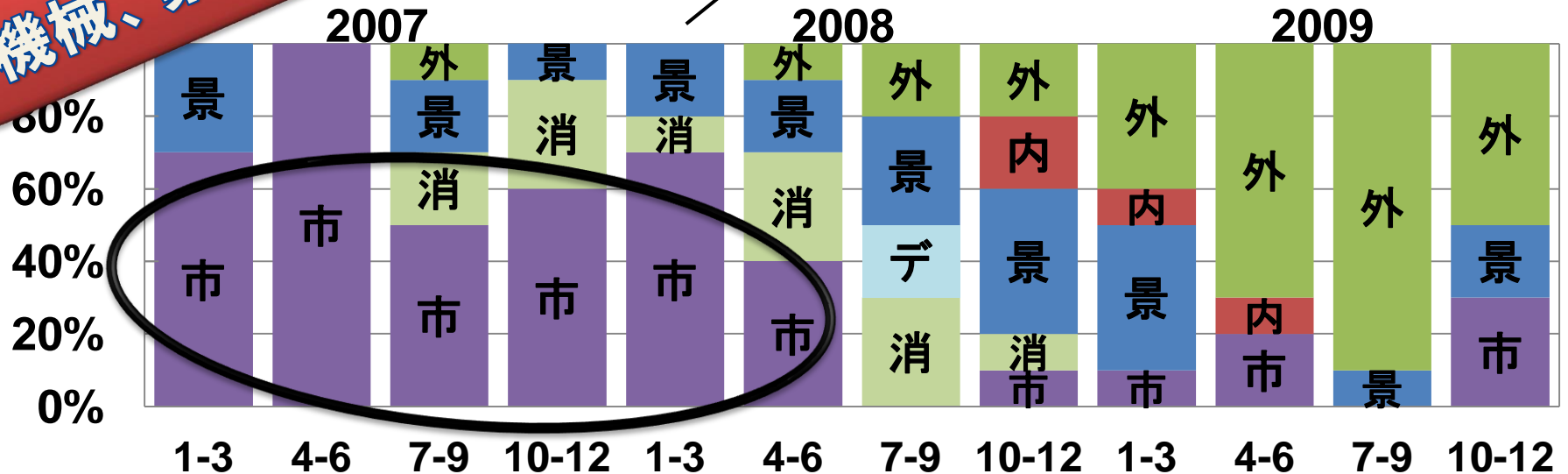
景気の波により受注動向が動き, 業績に影響する銘柄

市況関連株

原油など商品価格に業績が左右される銘柄

機械、素材

1月 原油価格100\$突破



結果：3ヶ月毎(関連株分類)

外需関連株

国外での収益(輸出や現地の売上等)が影響する銘柄

内需関連株

事業基盤が国内にあり, 内需拡大にウエイトをおく銘柄

景気循環株

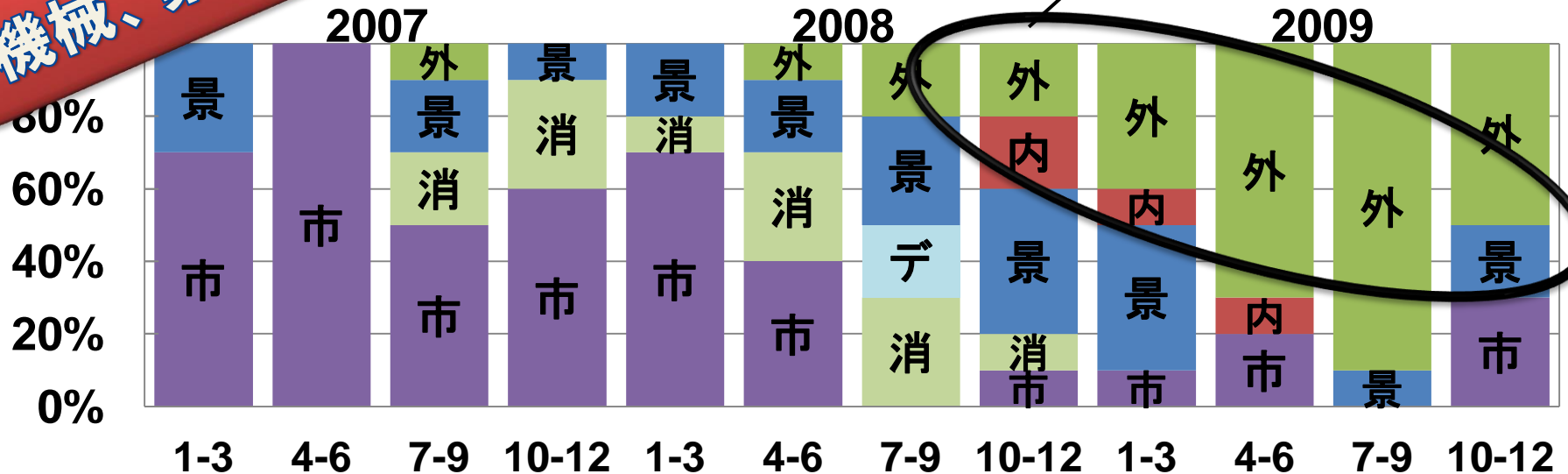
景気の波により受注動向が動き, 業績に影響する銘柄

市況関連株

原油など商品価格に業績が左右される銘柄

機械、素材

11月 トヨタショック



結果：3ヶ月毎(関連株分類)

外需関連株

国外での収益(輸出や現地の売上等)が影響する銘柄

内需関連株

事業基盤が国内にあり, 内需拡大にウエイトをおく銘柄

景気循環株

景気の波により受注動向が動き, 業績に影響する銘柄

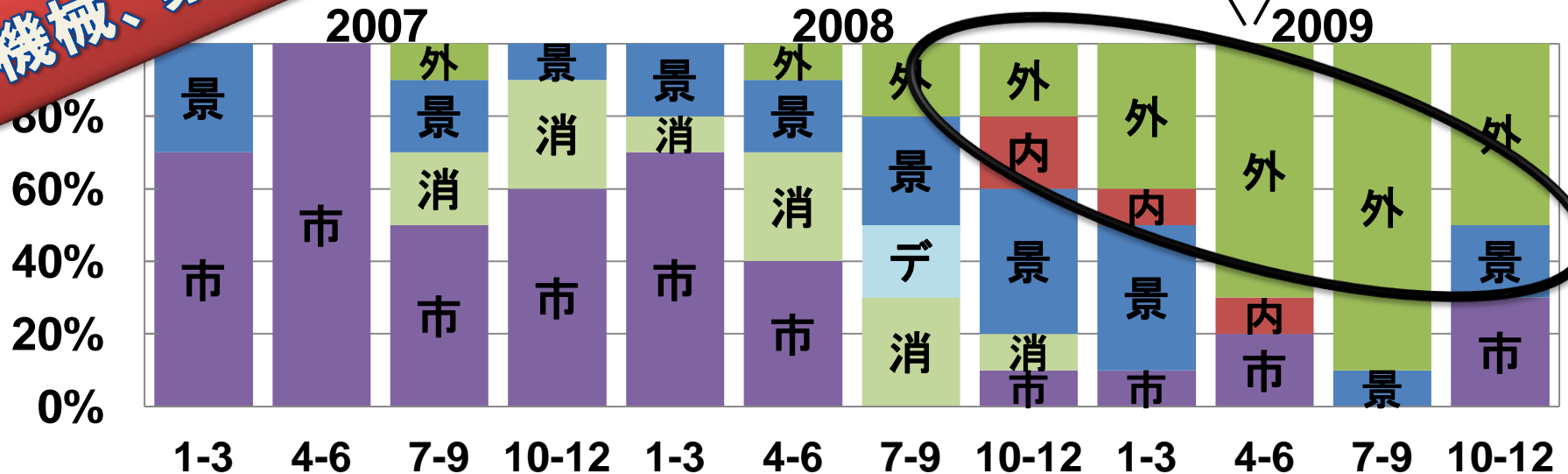
市況関連株

原油など商品価格に業績が左右される銘柄

機械、素材

4月 クライスラー破綻

6月 GM破綻



まとめ

まとめ

- 3ヶ月毎の主要セクタの抽出ができた
- 1年毎の欠点を解消できた
- 第2固有ベクトル成分の変遷を追うことで
経済動向に沿った結果を得られた

今後の課題

- データ数: N を減らし1ヶ月毎に短縮
- 指標を参考にした価格予測の研究
- RMT-PCAの株式市場以外への利用