

株式市場における株価大崩落の兆し検知への挑戦

A challenge to obtain the crystal ball in the stock market - A candle chart based graph density approach -

岡田 克彦 *1 羽室 行信 *2
Katsuhiko OKADA Yukinobu HAMURO

*1*2 関西学院大学 経営戦略研究科

Institute of Business and Accounting, Kwansei Gakuin University

This paper presents the effectiveness of using graph density as a proxy for biased investor sentiment. The peak of the graph density of various movement successfully predicted the short-term stock index performance. Specifically, we used candle chart similarities among stocks and created 50 different graph density based signals. We tested these signals in the Japanese, US and UK stock market for the period between 2000-2016 and found them successful in predicting the market in the short run.

1. はじめに

株式市場には多くの銘柄が上場しているが、その一部の銘柄の単純平均や時価総額加重平均の値が株価指数である。株価指数の上昇や下落の背景には何千(米国では何万)という数の銘柄変動が存在しているが、一般的な投資家が目にする指標は株価指数への寄与度一覧や騰落レシオ程度である。現状では圧倒的多数である隠れた銘柄群の動向について網羅的に知る術はない。しかし、そこには重要な情報が含まれている可能性がある。例えば、ある一時点で引値と安値の差が大きい時、一定価格以下では買いの需要が強いことを意味する。ただ、株価指数でこのような価格パターンが出たとしても、主要銘柄や大型株の銘柄動向を示しているだけかもしれない。同時点において、圧倒的多数を占める非主要銘柄群でも同様の傾向を示しているのであれば、投資家全体の感情を表していると考えられ、株式市場は真に底堅いと判断できよう。本研究では、株式市場を網羅的且つ多面的に捉えるために、10個の変数と5個の局面ベクトルから構成される50の類似度グラフを、日・米・英の上場銘柄から生成した。そして、類似度グラフの枝密度の時系列変化から株価指数先物の短期の予測可能性を検証した。

2. 手法

日 t における銘柄集合 N_t について、類似した銘柄間に枝を張ったグラフ $G_t = (N_t, E_t)$ を考える。枝集合 E_t は式 (1) で示されるように、任意の二つの銘柄 $a, b \in N_t$ について、日 t から過去 d 日間の株価推移の類似度が、与えられた閾値 ρ 以上であるような銘柄間に枝を張ることで構成する。

$$E_t = \{(a, b) | \text{sim}(\mathbf{x}_{t,d}^a, \mathbf{x}_{t,d}^b) \geq \rho; a, b \in N_t\} \quad (1)$$

ここで $\mathbf{x}_{t,d}^a$ は、銘柄 a の日 $t-d+1$ から日 t までの数値ベクトルで、表 1 に示された変数 v のいずれか一つに対応する。類似度関数 $\text{sim}(\cdot, \cdot)$ としては、ピアソンの積率相関係数を用い、変数に応じて $\rho = 0.8$ もしくは $\rho = 0.9$ を設定した(表 1)。また、 $d = 10$ に固定して実験をおこなった。ただし、定義上一日前の情報を参照している変数 ($kerb$, $barb$, $aReturn$) については、 $d = 9$ としている。相関係数の計算を効率的に計算するにあたって、田部井らにより開発された SketchSort 法 [Tabei10] を用いた。

連絡先: 岡田克彦, 関西学院大学, 西宮市上ヶ原一番町 1-155, TEL:0798-54-6572, hamuro@kwansei.ac.jp

表 1: シグナル変数一覧

変数名	定義	ρ	d
$close_t$	c_t	0.8	10
$kerb_t$	o_t/c_{t-1}	0.9	9
$barb_t$	c_t/o_t	0.9	9
$range_t$	$h_t - l_t$	0.8	10
$volume_t$	v_t	0.9	10
kMb_t	$kerb_t - barb_t$	0.9	10
$shadowU_t$	$(h_t - \max(o_t, c_t))/range_t$	0.8	10
$shadowL_t$	$(\min(o_t, c_t) - l_t)/range_t$	0.8	10
$shadow_t$	$shadowU_t \cdot shadowD_t$	0.8	10
$aReturn_t$	$c_t/c_{t-1} - p_t/p_{t-1}$	0.8	9

o_t :始値, h_t :高値, l_t :安値, c_t :終値, v_t :出来高, p_t :先物終値

以上の方法を全データ期間に適用することで、市場内で生じている全銘柄の関係性の変化をグラフの構造変化として表現することが可能となり、TOPIX のような株式指数では捉えることのできない市場の動きを検知できることが期待できる。グラフの構造変化を捉える方法としては、日 t に生成されるグラフ G_t から特徴量 f_t を抽出し、それを時系列データ (f_1, f_2, \dots) として構成し、そこからシグナル(後述)を定義することで実現する。グラフ特徴量としては、近年様々な特徴量が提案されているが、本研究では枝密度(edge density)を用いる。枝密度が高くなるということは、多くの銘柄が、与えられた変数において同じような変動を示していることを意味する。

ただし、類似度として相関係数を用いているので、株価の上昇局面で類似しているのか、もしくは乱高下の局面で類似しているのかは区別できない。そこで、式 (1) に示された類似度の条件に加え、式 (2) に示される条件を加えることで、局面を考慮した関係性グラフを構成する。

$$\text{sim}(\mathbf{x}_{t,d}^a, \mathbf{s}_{t,d}) \geq \sigma \text{ and } \text{sim}(\mathbf{x}_{t,d}^b, \mathbf{s}_{t,d}) \geq \sigma \quad (2)$$

ここで $\mathbf{s}_{t,d}$ はユーザにより指定されるサイズ d の定数ベクトルであり、本研究では表 2 に示される 4 つの局面を定義し、それに局面の制約なし (all) を加えた 5 つの局面を考えた。 σ は全ての局面において 0.7 を用いた。

以上、10 の変数と 5 つの局面の計 50 の組み合わせそれぞれについて枝密度を計算し、そこからシグナルを定義する。図 1 は、 $close$ (終値) 変数の sd (突然の下降局面) に関する 1987 年から 2015 年までの枝密度の推移、および TOPIX の推移を示したものである。枝密度はスパイク状に推移しており、またそ

表 2: 局面ベクトル一覧

局面名 $s_{t,d}$	定義	意味
ld	$(d, d-1, \dots, 1)$	単調な下降局面
lu	$(1, 2, \dots, d)$	単調な上昇局面
sd	$(1, 1, \dots, 1, 0)$	突然の下降局面
su	$(0, 0, \dots, 0, 1)$	突然の上昇局面
all	-	局面の制約なし

の多くのピークは TOPIX の底に対応しているように見える。これが本研究のきっかけであり、個別株の 4 本値データから導出される銘柄間の関係性指標によって TOPIX の底、もしくは天井を検知できるのではないかと仮説を得た。以前の研究では、日本市場で $close$ 変数のみに焦点を当て、その予測可能性を示した [Hamuro2016]。本研究は、それを 3 カ国 (日本、米国、英国)、10 変数に拡張し、さらなる予測精度の向上と普遍性を求める研究として位置づけることができる。

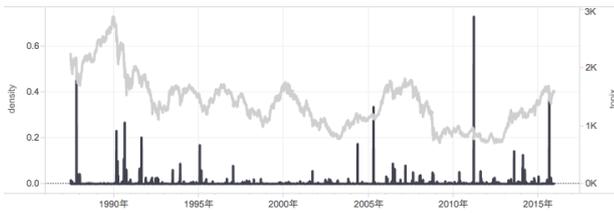


図 1: 1987~2014 年の $close-sd$ (終値-突然の下降局面) の枝密度 (黒線) と TOPIX(灰線) のトレンド

次に、シグナルの定義を示す。ある変数-局面 v における日 t にシグナル $Q_t(v)$ を、式 (3) で示される通り、グラフ特徴量としての枝密度 f_t が日 t においてピークを示しているかどうかの論理関数として定義する。なお、シグナルは、ピークを示す日 t の翌日 $t+1$ に初めて評価できることに注意する。またピークを示した時の枝密度のことをシグナル強度と呼称する。

$$Q_t(v) = (f_{t-1}^v < f_t^v) \wedge (f_t^v > f_{t+1}^v) \quad (3)$$

次に、シングルシグナルとダブルシグナルについて定義する。変数 v のシグナルが日 t から過去 m 日間に少なくとも一回発生することを、

$$Q_t^m(v) = \bigvee_{i=t-m+1}^t Q_i(v) \quad (4)$$

で表すと、シングルシグナルとダブルシグナルは、それぞれ式 (5)、(6) の通り定義される。すなわち、シングルシグナルとは、ある日 t に閾値 α を超える枝密度のピークが観測されることであり、ダブルシグナルとは、日 t にシングルシグナルが観測され、かつ日 t から過去 m 日間にシングルシグナルとは異なる変数のシグナルが観測されることである。本研究では $m=3$ に固定している。なお、ダブルシグナルにおける v_1 のシグナルをメインシグナル、 v_2 のシグナルをサブシグナルと呼称する。またサブシグナルのシグナル強度については特に基準を設けておらず、強度 (枝密度) が非常に小さな値であってもサブシグナルとして扱っている。本研究で扱う変数 v の数は 50 種であるので、シングルシグナルで 50 種、ダブルシグナルで 2,450 種 ($= 50 \times 50 - 50$) のシグナルを扱うことになる。

$$SS_t(v) = Q_t^1(v) \wedge (f_t^v \geq \alpha_t^{s,k}) \quad (5)$$

$$DS_t^m(v_1, v_2) = SS_t(v_1) \wedge Q_t^m(v_2) \wedge (v_1 \neq v_2) \quad (6)$$

$\alpha_t^{s,k}$ は、日 t においてシグナルとして識別するかどうかの基準となる閾値で、その設定方法としては、枝密度の絶対値を与えたり、過去の平均値や中央値で設定するなど様々な設定方

表 3: 分析対象とした銘柄数の推移。毎日に全上場銘柄の時価総額上位 20% を対象銘柄として選択し、年初の銘柄数を 4 年毎に国別に示している。

年	日本 (jp)	米国 (us)	英国 (gb)
2000	314	2,192	673
2004	339	3,606	632
2008	425	3,538	711
2012	460	3,668	667
2016	527	4,227	716

法が考えられるが、ここでは、日 t より前に観測された s 個のピークの第 k 位の枝密度として与える。 s を大きな値に設定すると、より過去のシグナル値を考慮することになり、また小さな値に設定すると、遠い過去のことは忘れ、現在に近い期間におけるシグナル値を基準としてシグナルの発生を定義することになる。すなわち、 s はある種の忘却パラメータとして機能することとなる。一方で k については、その値を小さくすることで、過去の際立ったシグナルを基準とすることとなりシグナル強度を直接定義するパラメータとして機能する。例えば、 $s=100, k=1$ の場合、過去 100 のシグナルのうち最もその強度 (枝密度) が高かったシグナルを超えなければシグナルとして識別しないことになる。本研究では $s=50, 100, 200, k=1, 5, 10, 15, 20$ で実験を行っている。

3. 利用データ

トムソン・ロイターの Datastream より取得した株価データを用いる。本研究で対象とした国は、日本 (jp)、米国 (us)、英国 (gb) の 3 カ国で、2000 年~2016 年 11 月の 17 年間弱の期間のデータである。分析の対象とした銘柄は、全上場銘柄の時価総額上位 20% の銘柄である。時価総額は国別日別に算出し、それぞれで上位銘柄を選択している。選択された銘柄数の年別推移を表 3 に示す。また、一年間の営業日数はおおよそ 250 日であり、17 年間で 4,131 日 (jp)、4,234 日 (us)、4,253 日 (gb) となっている。各国の株価指標については、日本は TOPIX、米国は S&P500、英国は FTSE100 を用いた。それぞれの値動きを図 2 に示す。



図 2: TOPIX(jp), S&P500(us), FTSE100(gb) の値動き。

4. 実験結果

4.1 シグナルの評価方法

シグナルの評価としては、シグナルの平均収益率、p 値、取引シミュレーションにおける Sharpe 比と max drawdown を用いる。平均収益率とは、シグナルが観測された日 ($t+1$) に株価指標を購入し、 h 日後に売却したときの収益率の平均である。この値がプラス (マイナス) であれば、市場の上昇 (下降) 傾向を検知するシグナルと考える。またその精度は、帰無仮説「収益率=0」、対立仮説「収益率 $\neq 0$ 」の検定における p 値を用いる (t 検定)。p 値がより低いシグナルは精度の高いシグナルと評価できる。

表 4: 国別の上位モデル (シングルシグナル, p 値 <0.1%)。h はシグナル観測後の先物保有日数、s はシグナルの基準値を決める過去の参照シグナル数、k はその中での基準値となるシグナルの順位。件数は、期間中のシグナルの観測回数で、return が平均収益率を示している。

国	シグナル	h	s	k	件数	return(%)
jp	barb-all	2	50	5	124	0.88
	barb-sd	3	200	10	55	1.68
us	shadowH-lu	3	100	20	216	0.52
	shadowL-ld	5	50	20	452	0.38
	kMb-sd	5	50	1	53	1.55
gb	barb-su	1	200	15	60	0.61
	kMb-sd	2	200	20	90	0.75
	kerb-ld	4	50	15	185	0.55

さらに、任意のシグナルを選択してトレーディングシミュレーションをした時の累積収益率を求め、その Sharpe 比と max drawdown を求める。Sharpe 比とは、日々の収益率 (取引のない日の収益率は 0 とする) の標準偏差に対する比によって定義され (一般的には、日ではなく年率に換算して計算する)、リスクで標準化した平均的な業績を評価できる。また max drawdown とは累積収益率の最も大きい落ち幅のことで、この値が小さいということは、大規模な市場の崩落/高騰による損失が小さいということで、崩落/高騰の予測性を評価できる。

4.2 国別の上位モデル

最初の実験としてシングルシグナル 50 種、ダブルシグナル 2450 種について、 $s = 50, 100, 200$ 、 $k = 1, 5, 10, 20$ 、 $h = 1, 2, 3, 4, 5$ をパラメータとして収益率を計算し、国別に精度の高いシグナルを求めた。すなわち、 $(50 + 2450) \times 3 \times 4 \times 5 = 150,000$ 通りのシグナルから、先物の予測可能性を示すシグナルを選択した。なお、 s, k, h をパラメータとして定義される 15 万通りのシグナルを、以下では「モデル」呼ぶことにする。表 4 と表 5 に、シングルシグナルおよびダブルシグナルを用いた国別の上位モデルの結果を示す。

まず表 4 はシングルシグナルの結果であり、p 値 $\leq 0.1\%$ の精度を持つモデル一覧である。例えば、日本においては、barb-all すなわち、多くの銘柄の取引のされ方が日中に似てくればそのタイミングで先物を購入することによって平均 0.88% のリターンが得られることを示している。このモデルは過去 50 回のシグナルピーク ($s = 50$) の内、上位 5 位 ($k = 5$) の枝密度以上であるものをシグナルとして識別しており、その基準を満たすものがシミュレーション期間中に 124 回発生したことが示されている。ただし、保有期間は 2 日であるので ($h = 2$)、シグナルの発生日 (枝密度のピーク日 t の翌日) に TOPIX 先物を購入し、2 日間保有した後に売却するという比較的短期の傾向を示唆するといえよう。barb-sd は日中の取引において急激に下落する状況を表現している。日中に多くの銘柄の株価が急落しているため、それらの銘柄に関連する悪いファンダメンタルズ情報が入った可能性がある。しかし、概ねその時点で悪材料が織り込まれ、株式市場の短期の底打ちを示していると考えられる。日本市場のこうしたリバーサル傾向の存在は [Jagadeesh11] と整合的な結果と言えよう。

表 5 は、ダブルシグナルの結果であり、p 値 $\leq 0.05\%$ の条件を満たすモデルを記載している。列挙されているモデルを全て直感的に解釈することは困難であるが、米国においては kerb-all がメインシグナルとなっているものが多い点に着眼したい。kerb 系のシグナルは overnight の動きを示す指標である。多くの銘柄が overnight で動きを示し、且つサブシグナルの kerb-su に見られる様に、近傍に overnight で上昇する銘柄

表 5: 国別の上位モデル (ダブルシグナル, p 値 <0.05%)。項目の意味については表 4 に同じ。

国	mainSignal	subSignal	hp	s	k	件数	return(%)
jp	kerb-all	range-all	3	50	15	50	1.24
	shadowH-su	kMb-all	1	50	15	91	0.52
	close-su	shadowH-lu	2	50	15	73	0.50
	shadowH-su	shadow-su	1	50	15	113	0.48
	shadowH-su	kMb-sd	1	50	15	102	0.48
	shadowH-su	aReturn-lu	1	50	15	101	0.42
	kerb-su	range-ld	2	50	20	62	-0.71
	kerb-sd	barb-sd	2	50	15	61	-0.86
us	shadowH-ld	range-lu	5	50	20	65	-0.97
	kerb-all	kerb-su	5	50	15	71	0.98
	kerb-all	kMb-lu	4	50	15	59	0.95
	kerb-all	range-lu	4	50	15	53	0.90
	kerb-all	volume-su	4	50	15	62	0.87
gb	kerb-all	kerb-ld	4	50	15	56	0.80
	shadow-su	range-su	5	100	10	51	1.12
	shadowH-lu	aReturn-sd	5	100	20	76	0.97
	shadowH-lu	shadow-ld	4	50	20	71	0.84
	kerb-sd	shadow-su	3	50	20	76	0.60

表 6: long/short 戦略別シミュレーションの評価。ret は年率換算の収益率で SD はその標準偏差を示す。max dd は、最大下落率を示す。

有意水準 (%)	モデル数		延べ取引日数	ret (%)	SD (%)	Sharpe 比 ret/SD	max dd (%)	max dd の期間		
	sgl (%)	dbl (%)						開始	終了	
0.1	0.05	8	15	3195	15.5	13.6	1.14	-27.0	20080615	20081010
0.1	0.1	8	32	3452	15.6	13.9	1.13	-27.3	20080615	20081010
0.5	0.5	20	169	4121	11.4	13.9	0.81	-40.8	20080519	20081010
1	1	30	306	4193	11.7	13.7	0.86	-39.1	20080519	20081010
5	5	59	1138	4212	10.6	13.3	0.80	-46.8	20080519	20081010
0.1	0.05	0	3	548	6.1	5.7	1.07	-7.7	20040225	20040601
0.1	0.1	0	4	616	6.5	6.0	1.08	-8.8	20050627	20050926
0.5	0.5	3	42	2955	19.3	13.1	1.48	-34.2	20041025	20060505
1	1	5	93	3685	14.6	14.3	1.02	-35.4	20120725	20130522
5	5	14	629	4192	8.5	13.7	0.62	-50.9	20030312	20060405
0.1	0.05	8	18	3335	21.6	14.4	1.50	-18.2	20080615	20081010
0.1	0.1	8	36	3560	22.1	14.7	1.51	-19.1	20090108	20090309
0.5	0.5	23	211	4185	30.7	13.2	2.32	-17.8	20080916	20081010
1	1	35	399	4205	26.3	11.6	2.27	-7.9	20130404	20130415
5	5	73	1767	4216	19.1	7.7	2.47	-8.9	20090107	20090305

が増えていると、全体の地合いが強いと判断して良いのではないが、[Jagadeesh11] は日本人のリバーサル傾向に対して米国人はモーメンタム傾向を示すと報告しているが、有効なシグナルが国ごとに異なるのは、各国の投資家特性の差異に起因するのかもしれない。

4.3 上位モデルによるシミュレーション

先に述べた 150,000 のモデルの内、p 値上位のモデルを用いた取引シミュレーションを実施した。p 値の上限値を変化させた結果を表 6 に示す。この表には 3 通りの実験結果が示されている。上段は、プラスの平均収益率を達成するモデルを選択し long 戦略を実施した結果。中段はマイナスの平均収益率を達成するモデルを選択し short 戦略を実施、下段は両戦略を同時に実施した結果である。

表 6 の左列から採用する戦略の基準 (有意性)、取引日数、年率リターン、年率リスク、シャープ比、最大下落率 (max dd)、最大下落率の期間を記している。最大下落率は投資家がファンド等を選択する際の重要な基準であり、当初からファンドを購入した顧客が、自らの資産額の最高値からどの程度最大損失を被ったかを示す。例えば、資産が 100 から 120 になって 20% のリターンを得たとしても、A の戦略では単調増加し、B の戦略では 100 が 150 になって 120 になったとする。A の戦略は max dd が 0% であるが、B では 20% となり、同じリターンでも A の戦略の方が高い評価を得る。表 6 の一番上の long 戦略

では、平均年率リターンが 15.5%と高く、シャープ比も 1.14 と良い数字であるが、max dd が-27%でありそれが発生した時期はリーマンショックを含む 2008 年 6 月 15 日から 10 月 10 日だったことを示している。

表 6 から明らかな様に、最もシャープ比の高い long 戦略は、 p 値 $< 0.1\%$ でシングル戦略の中から選択し、 p 値 $< 0.05\%$ でダブル戦略から選択した方法である。最もシャープ比の高い short 戦略は、 p 値 $< 0.5\%$ でシングル戦略の中から選択し、 p 値 $< 0.5\%$ でダブル戦略から選択した方法である。取引方法としては、各シグナルが発信されたタイミングで、各国の株価指数先物 (TOPIX 先物、FTSE100 先物、S&P500 先物) を取引する。定義上シングルシグナルはダブルシグナルも包含するが、ダブルシグナルが出た場合は、その分先物を複数枚買うことで重みをつけることにする。

代表的なシミュレーション結果を図 3, 4, 5 に描画する。モデルとしてはシングルシグナル、ダブルシグナルとも p 値 $< 1\%$ 基準で 150,000 モデルの中から long 戦略用のモデルとして 336 個採用し、short 戦略用のモデルとして 98 個採用した。図 3 は可視化のためにランダムに 20 個を選択して描画している。左が long 戦略、右が short 戦略である。barb-ld と close-sd の様にそれぞれのシグナル発生パターンにはある程度の相関があるかもしれないが、多くの戦略を組み合わせることで、パフォーマンスのスムージングが可能となる。図 4 で確認できるのは、シングルシグナル、ダブルシグナルとも p 値 $< 1\%$ 基準で選択された long モデルと short モデルを全て合算した結果である。組み合わせにより分散効果が発揮され、パフォーマンスが安定することがわかる。さらにロングとショートを組み合わせた、開始資産を 100 として描画した結果が図 5 である。ここに採用した long モデルからのシグナルは 1 日平均 25 回発生し、short モデルからのシグナルは 1 日平均 7 回発生している。全期間を通じてマーケットを long している期間が支配的であるが、金融危機時には short モデルからのシグナルの方が多期間が散見される。また 2008 年 9 月から 10 月末までの金融危機の影響が最も顕著だった期間だけにフォーカスしても、勝率が 6 割に達しており、相場の大崩落をうまく捉えていた可能性が高い。その結果、表 6 に示す様に max dd の 7.9%が発生したのは 2013 年 4 月であり、むしろ金融危機時には順調に資産を増加させていたことがわかる。

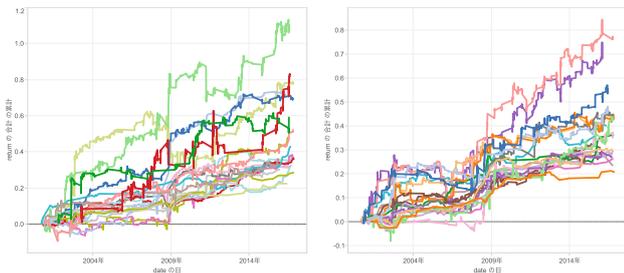


図 3: p 値 $\leq 1\%$ の個々のシグナルで株式指標を取引した場合の累積収益率の推移

5. おわりに

日・米・英の主要株式市場について、個別銘柄の変動パターンに含まれている情報を網羅的に調査し、銘柄間の関係性の変化を類似度グラフの枝密度変化としてとらえ、シグナル化した。計 50 個のグラフ枝密度をベースに各種パラメータの組み合わせも含めた合計 150,000 個のモデルを 2000 年から 2016

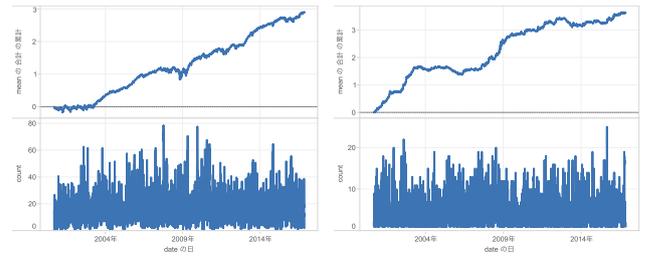


図 4: 有意なシグナルを合成した結果。図 3 に例示された個々のシグナルの収益率を合計/累積したもの。左図が long 戦略で右図が short 戦略の結果。図 3 に見られる個々のシグナルの変動がお互いに相殺され安定した伸びを示していることが分かる。

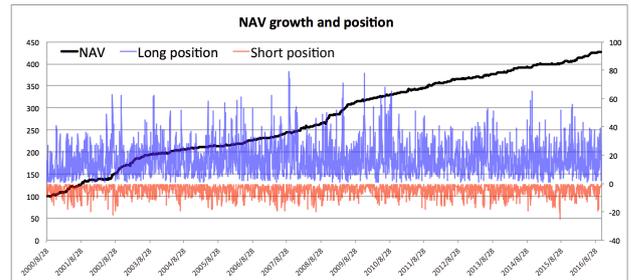


図 5: long/short 戦略を合成した結果。さらに、図 4 の long/short 戦略を合成した結果。それぞれの戦略に見られる変動が更に相殺され、より安定した伸びを示していることが分かる。

年の期間で調査した。次にそれらのモデルのもたらす平均リターンがゼロであるという帰無仮説をどの程度強く棄却するかという基準 (p 値) でモデル選択し、long モデルで 336 個、ショートモデルで 98 個抽出した。それらを組み合わせで運用シミュレーションを行った結果、金融危機時にもダイナミックに対応できるモデルが生成可能なことが明らかになった。この結果は、価格形成に投資家の期待の偏りが反映されており、その偏りが近未来の株価水準の方向性を示すという仮説と整合的である。

謝辞

本研究は、JST CREST、科研費基盤研究 (B) 25285127、及び、統計数理研究所平成 27 年度公募型共同研究「重点研究 28-共研-4101」の研究助成を受けている。

参考文献

- [Hamuro2016] 羽室行信, 岡田克彦, Cheung Stephane, 銘柄類似度グラフの時系列構造変化に基づく株価予測, 2016 年度人工知能学会 (第 30 回), 北九州国際会議場, 2016/6.
- [Jagadeesh11] Narasimhan Jegadeesh, Sheridan Titman, Momentum. Annual Review of Financial Economics, 3(1):493-509. 2011.
- [Tabei10] Yasuo Tabei, Takeaki Uno, Masashi Sugiyama, Koji Tsuda: Single Versus Multiple Sorting in All Pairs Similarity Search, The 2nd Asian Conference on Machine Learning (ACML2010), Tokyo, Japan, 2010.