収益基準に基づく株価不正操作発見システム

LF-011

Discovery System for Fraudulent Stock-Price Manipulation Based on Profits Criterion

下平 剛志 Takeshi Shimodaira　鈴木 英之進 Einoshin Suzuki

横浜国立大学大学院工学府物理情報工学専攻電気電子ネットワークコース

Department of Electrical and Computer Engineering, Yokohama National University

1. はじめに

近年株式市場において，人目につかないように不正が行われていることはほぼ確実であり，不正取引 [1] の発見が重要な問題となっている．不正取引は公正な価格形成を阻害し，投資者に不測の損害を与えることになるため証券取引法で禁止されている．ただし明確な基準が無いこともあり，巧妙に隠され発見が困難である．さらに多種の銘柄 [2,3] から構成される株価データは膨大であり，全てを人間が検査するのは難しい．そこで本研究は，データマイニング [4] による，株価データからの不正発見を目的とする．

不正検出システム [5,6] を構築するにあたり，銘柄毎に取引回数，価格，株数等のデータが異なる点に注意しなければならない．よって幅広い銘柄に対し実行可能にするためには，汎用性が高い検出基準を設定すべきである．さらに不正者がシステムに発見されないよう，巧妙な手口で不正を行うであろうという点を考慮する必要がある．以上の問題に対処するために，平均株価に基づき“不正者の利益”を算出し，不正検出を行う手法を提案する．

2. 対象問題

以下において，取引成立のことを約定と呼ぶ．株価データは，日付を s，時間を t，株価を m，株数を a とした N 回の約定から構成されるとする．i 番目の約定は $d(i) = (s(i), t(i), m(i), a(i))$ となる．図 1 に株価データの例を示す．

d(1)=	1月 6日	9:00	220円	100株
	s(1)	*t*(1)	*m*(1)	*a*(1)
⋮	⋮	⋮	⋮	⋮
d(*N*)=	3月18日	15:00	215円	200株
	s(*N*)	*t*(*N*)	*m*(*N*)	*a*(*N*)

図 1: 株価データの例

ここで，データ集合を $D = \{d(1), d(2), ..., d(N)\}$，約定 $d(i)$ が不正取引であるかを判定する手続きを L，L により検出された n 番目の不正約定を $v_{n(D,L)}$ とする．本研究が対象とする問題は，D を入力とし，L を用いて不正約定集合 $V = \{v_1, v_2, ..., v_{n(D,L)}\}$ を求めることとして表せる．

横浜国立大学大学院工学府物理情報工学専攻電気電子ネットワークコース 鈴木研究室，〒 240-8501 横浜市保土ヶ谷区常盤台 79-5，Tel: 045-339-4135，E-mail: simo@slab.dnj.ynu.ac.jp

3. 提案手法

本手法では不正者の利益を算出することにより不正を判定する．これは，不正者にとって重要なのは自己の利益であり，巧妙な手口で不正を行っても利益は隠せないと考えたからである．よって幅広い銘柄に対し検出が可能で，汎用性が高い基準だと考える．

われわれは移動平均 [7] による平均株価 [8] を用い，それに基づき利益を算出する．次に閾値 θ を用いて利益を，不正利益と一般取引によるものに分け，1 日の総不正利益を不正者の収益 Q として求める．利益の高い約定が集中している箇所ほど不正の可能性が考く，Q が高くなると考えられる．よって収益 Q はユーザが結果を評価する際に有用な情報となる．

3.1 平均株価

一般の移動平均株価は，1 日の最後の約定価格である終値から算出される．ただし終値を不正に操作された場合を考えると，終値だけから平均株価を求めることは危険である．本研究では株価不正操作の発見を目的としているため，終値以外の約定価格も使用し平均株価を算出する．

i 番目の約定の平均株価を，例数 n を用い，$i-n$ 回から $i-1$ 回までの約定価格をもとに算出することにする．ここで例数 n の最大値 n_{max} を定義するにあたり，銘柄により 1 日の約定回数に差がある点を考慮しなければならない．例えば 1 日の約定回数が約 1 回という人気が低い銘柄から，約 1000 回ほどの人気が高い銘柄まで存在し，その差は大きい．株価変動に対し，n_{max} が小さければ平均株価は不安定になり，逆に大きければ安定しすぎて柔軟な対応が困難になると考えた．以上をふまえた上で，検査期間を f 日間とし，検査銘柄の 1 日の平均約定回数 $g_{ave} = \frac{N}{f}$ を得る．そして式 (1) を用い n_{max} を定義する．

$$\begin{aligned} \text{if}(25\sqrt{g_{ave}} > 2g_{ave}) \quad & n_{max} = 25\sqrt{g_{ave}} \\ \text{else} \quad & n_{max} = 2g_{ave} \end{aligned} \tag{1}$$

株式の売買において，銘柄毎に最小の約定株数である売買単位株が定められており，それを A とする．株数が多い約定の方が平均株価を求める際に信頼性が高いと考え，$\sqrt{\frac{a(i)}{A}}$ 倍の重みを乗ずる．移動平均を用いるため例数が n_{max} に達した場合は，逐次古いデータを削除し，新しいデータで平均株価を計算する．以上より求める，

$$\text{if}(i < n_{max}) \quad n = i, \quad \text{else} \quad n = n_{max}$$

$$m_{ave}(i) = \frac{\sum_{k=i-n}^{i-1} \left(m(k) \sqrt{\frac{a(k)}{A}} \right)}{\sum_{k=i-n}^{i-1} \sqrt{\frac{a(k)}{A}}} \tag{2}$$

3.2 収益の導出

i 番目の約定の利益 $P(i)$ を式 (3) より得る．

$$P(i) = \mid m_{ave}(i) - m(i) \mid a(i) \tag{3}$$

$P(i)$ は株価が平均株価より高ければ売り，安ければ買うと仮定した理想の利益を表す．

次に $P(i)$ に対し閾値 θ を用い判定を行う．$P(i) \geqq \theta$ ならば不正取引と考え収益に組み入れ，$P(i) < \theta$ ならば一般の取引と考える．これは各約定に対し，利益が大きければ不自然な取引とみなすことを意味する．ただし不正側は取引を分割して行うことも考えられるため，同時刻の約定が複数回存在する場合はそれを考慮し，収益を求めるようにする．まず，約定単独で閾値を越える利益を単一不正利益 R_1 として式 (4) で定義する．

$$R_1 = \sum_{k \text{ s.t. } P(k) \geqq \theta} P(k) \tag{4}$$

次に $P(k) < \theta$ となる $P(k)$ について，約定時間を $t(k)$，同時刻の約定の回数を j とし，集合不正利益 R_2 を式 (5) で定義する．ただし，$R_2 < 2\theta$ ならば収益とはみなさない．

$$R_2 = \sum_{l=k}^{k+j} P(l) \tag{5}$$

ただし，$t(k) = t(k+1) = \cdots = t(k+j)$
かつ，$P(l) < \theta$，$R_2 \geqq 2\theta$ である．

R_2 は単独では閾値を越えない約定を，同日かつ同時刻の条件で加算し，総合した値が閾値の 2 倍を越えた場合に発生する利益を意味する．2 倍という値は，最低でも 2 つ以上の $P(k)$ を加算することを考慮し設定した．これにより，不正側が取引を分割して行う場合に対応できると考えられる．式 (4)，(5) を用いた収益導出の概念図を図 2 に示す．なお，業績修正等の大きなニュースが発生した際，株価が変動する場合が多い．よってこのようなニュースが発生した日は収益を削除し，誤検出を防ぐ．

3.3 閾値 θ の設定

以上の処理により，適切な θ を設定することで 1 日の不正者の収益が推定できることが分かる．ただし各銘柄について，不正を区別する適切な θ を設定することは困難である．

そこで本手法では，最大閾値 θ_{max} と，最小閾値 θ_{min} を定義し，θ_{max} から θ_{min} まで等間隔で θ を変化させ，前述した収益計算を複数回実行する処理を提案する．この処理の長所は，結果を平均的に評価することにより，適切な θ を設定しなければならない問題を解消した点である．θ_{max} と θ_{min} については，幅広い銘柄に通用するよう次節で定義している．

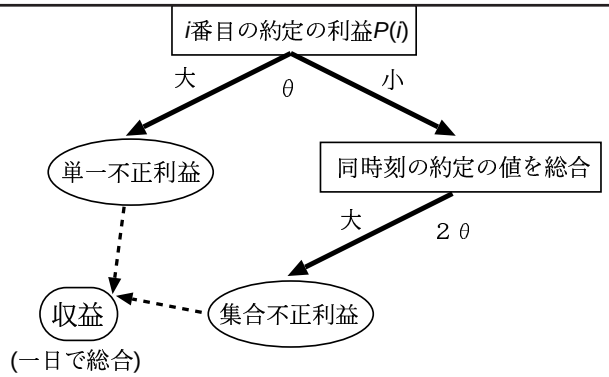

図 2: 閾値 θ を用いた利益分割

3.4 最大閾値と最小閾値

前処理によって，期間内で高い順に例数 n_{max} 個分の高利益 R_h と，高投資額 I_h を得る．投資額 I は，株価 m と株数 a の積で $I = ma$ とする．式 (6) に示すように各値を平均し，平均高利益 R_{hav} と，平均高投資額 I_{hav} を得る．

$$R_{hav} = \frac{\sum_{k=1}^{n_{max}} R_h(k)}{n_{max}}, \quad I_{hav} = \frac{\sum_{k=1}^{n_{max}} I_h(k)}{n_{max}} \tag{6}$$

ここで不正の存在する可能性を表す評価値として G を式 (7) で定義する．

$$G = \frac{R_{hav}}{I_{hav}} \tag{7}$$

例えば R_{hav} が高く I_{hav} が低い銘柄は G が高くなり，不正の存在する可能性が高いと考える．逆に I_{hav} が高いにも関わらず，R_{hav} が低い場合は不正の可能性が低い．

期間内の最大投資額を I_{max} とし，GI_{max} を最大利益の理論値として用いることにする．次に最大利益を p_{max}[1] とし，最大閾値 θ_{max} を式 (8) で定義する．

$$\theta_{max} = \text{MIN}\{GI_{max}, \quad p_{max}\} \tag{8}$$

式 (8) において MIN を用いる理由について述べる．$GI_{max} < p_{max}$ の場合に GI_{max} を用いるのは，p_{max} では約定 1 つしか参照しないため，もし極端に大きい p_{max} が存在した場合，他の箇所の検出が正確に行われないからである．逆に $GI_{max} > p_{max}$ の場合，不正取引が検出されないため p_{max} を用いる．なお，GI_{max} を求める際に使用する I_{max} も単独の約定から求められた値であり，もし I_{max} が極端に大きければ同様の問題を生ずる可能性がある．これは，株数が大きい取引は目立ち，不正側は極力行わないと考えられるため，I_{max} に制限を付与することで対処できる．本手法では株数が売買単位株の 100 倍以下 ($m < 100A$) という制限を設けた．

特定した約定数が全体の $\frac{1}{3}$ を越えた場合の閾値を θ_e とし，$\theta \leqq \theta_e$ となった際に計算を終了することにする．こ

[1] p_{max} は期間内で θ が 0 の時の最大利益に相当する．

れは，閾値を小さくしていくと最終的にはほとんどの約定を特定することになり，無意味だと考えたからである．最小閾値 θ_{min} を式 (9) で定義する．

$$\theta_{min} = \mathrm{MAX}\{0.01\theta_{max},\quad \theta_e\} \tag{9}$$

最後に，求めた θ_{max} を θ_{min} まで 1 %ずつ減少させ，最大 100 回計算を行い約定単位の重要度を得る．計算を行った回数を α，α 回中収益として組み入れた回数を β とする．α に対する β の割合が高い約定ほど重要度が高いことが分かる．

3.5 出力方法

計算回数 α に対し，利益として組み入れた回数 β について次のように表示する．β が多い約定ほど，☆の数が多く重要であることを示す．

- $\beta \geqq 0.8\alpha$ ならば, ☆☆☆
- $0.8\alpha > \beta \geqq 0.5\alpha$ ならば，☆☆
- $0.5\alpha > \beta \geqq 0.2\alpha$ ならば，☆

本手法では，☆の数が 1 つ以上の約定を不正約定とみなし，不正約定集合 V を得る．さらに収益について，計算回数 α で平均し，期間内で 1〜10 位まで順位付けして表示する．この順位付けにより，ユーザは収益の高い日，すなわち本手法が不正の可能性が高いと判定した日を，容易に探し出すことができる．

4. 実験評価

4.1 実験結果

本手法を実際の 5 銘柄に対し実行し，図 3，4 に検出結果の一部を示す．検査期間はある年の 1 月 6 日から 3 月 18 日の 50 日間とした．

☆印の後の () 内は，不正側が行ったと考えられる取引内容を示す．(買) は買った，(売) は売ったを示す．「前日までの平均株価」は前日の終値の時点で算出した平均株価を表し，対象とする日の株価変動を評価する情報となる．

4.2 考察

図 3 の銘柄 A は 3 月×日の収益が期間内で 1 位であることを示している．3 月 5 日に 700 円以下であった株価が 3 月×日に 750 円まで高騰し，3 月 11 日には再び 700 円以下に急落している．さらにこの日は特に大きなニュースもなかったにもかかわらず，最大 55 円も値上がりした．午前中に少ない株数で株価を上げ，情報を得て買いにきた一般投資家に，午後高値で大量に売った等，不正の疑いが考えられる．

図 4 の B は，g_{ave} が約 1 回という取引回数の少ない銘柄である．第一に取引回数が少ないのに，約定が無い日が存在せずに，50 日間中 47 日 (94 %) に必ず 1 回ずつ約定が存在している点が不自然であると考えた．次に，残りの 3 日は図中の収益順位 1 位から 3 位の日である．1 月 22 日から 2 月 10 日，12 日と少しずつ株価を上げ，2 月×日は 50 円も高騰している．約定時間も 9 時 1 分に 4 回と，13 時 5 分に 3 回というように集中し，銘柄 A と同様にその後株価は急落している．よって不正の疑い有りと考えた．

5. おわりに

本論文では，平均株価を用いて算出した収益に基づき，株価不正操作を発見する手法を提案した．提案手法により，株価データから不正約定集合 V と，収益 Q という不正判定に有用と思われる情報を得ることができた．

V と Q をもとに，前後の株価の動き，ニュースの有無等から最終的に不正と判断することはユーザに任されている．これは問題の性質上当然であると考える．パラメータについては，現在統計的アプローチを用いて減らすことに取り組んでいる[2]．

参考文献

[1] 不公正取引について，http://www.iwaisec.co.jp/top/hukouseitorihiki.html

[2] 東京証券取引所，http://www.tse.or.jp

[3] JASDAQ，http://www.jasdaq.co.jp

[4] P．Adriaans and D．Zantinge 著，山本英子，梅村恭司: データ マイニング，共立出版, 2001

[5] 櫻井泰子，鈴木輝，塚本真誠，斉藤純平，金山達来，蛭間久季，渡部章：不正プログラムおよび不正ドキュメントの調査と検索ソフトウェアの開発, Check Point Experience Seminar, FIT2002, Security Solution Expo 2002, 2002

[6] 松尾真一朗，尾形わかは：データ交換可能な多対多マッチングプロトコル, The 2002 Symposium on Cryptography and Information Security, 2002

[7] K．Brannas and S．Quoreshi: Integer-Valued Moving Average Modelling of the Number of Transactions in Stocks，Department of Economics & USBA，Umea University, 2004

[8] 株価指標を使ったチャートの活用法，http://www.miller.co.jp/member/chart/chart_guide.html

[2] 取引株数だけで不正を判定することは可能であるかもしれないが，われわれは検討した結果困難であると判断した．

```
   時間     株価    株数
========= 3月5日 =========
                ⋮
 (12:57)   695,   300
 (14:24)   695,   200
 (14:39)   695,   200
                ⋮
========= 3月×日 =========
 平均収益          315,547円( 1位)
 前日までの平均株価              671円
 ( 9: 5)   695,   100
 ( 9: 8)   700,   100
 ( 9: 8)   710,   300
 ( 9:10)   700,   200
 ( 9:12)   710,   100
 ( 9:12)   720,   200
 ( 9:13)   725,   400 ☆(売)
 ( 9:14)   725,   200
 ( 9:17)   730,   500 ☆(売)
 ( 9:28)   730,   200
 ( 9:28)   730,   200
 ( 9:29)   730,   100
 ( 9:30)   725,   100
 (10: 3)   725,   100
 (10: 9)   725,   100
 (10:10)   725,   100
 (10:26)   725,   700 ☆(売)
 (13:12)   735,  1000 ☆☆(売)
 (13:19)   735,  2400 ☆☆☆(売)
 (13:34)   715,   100
 (13:45)   740,  1100 ☆☆(売)
 (13:50)   740,   500 ☆(売)
 (14: 2)   740,   200
 (14:24)   750,  1800 ☆☆☆(売)
 (14:34)   740,   200
 (14:36)   750,   700 ☆(売)
 (14:46)   750,  1000 ☆☆(売)
 (14:47)   750,   100
 (14:50)   750,   100
 (14:52)   750,   100
 (14:52)   740,   100
 (14:56)   740,   100
                ⋮
========= 3月11日 =========
 平均収益          6,816円(low)
 前日までの平均株価              724円
 ( 9: 0)   710,   200
 ( 9: 7)   710,   200
 ( 9: 8)   710,   600
 (10:46)   690,   200
 (10:51)   690,   500
 (13: 7)   690,   100
```

図 3: 検出結果 (銘柄 A)

```
   時間     株価    株数
========= 1月22日 =========
 (14:45)   270,   1000
                ⋮
========= 2月10日=========
 (14:29)   301,   1000
========= 2月12日 =========
 (14:46)   330,   1000
                ⋮
========= 2月×日 =========
 収益          600,450円( 1位)
 前日までの平均株価              290円
 ( 9: 1)   370,   1000 ☆☆☆(売)
 ( 9: 1)   380,   1000 ☆☆☆(売)
 ( 9: 1)   380,   1000 ☆☆☆(売)
 ( 9: 1)   380,   1000 ☆☆☆(売)
 (13: 5)   410,   1000 ☆☆☆(売)
 (13: 5)   410,   1000 ☆☆☆(売)
 (13: 5)   420,   1000 ☆☆☆(売)
                ⋮
========= 2月○日 =========
 収益          133,802円( 3位)
 前日までの平均株価              320円
 ( 9:25)   410,   1000 ☆(売)
 (12:50)   430,   1000 ☆☆(売)
 (12:50)   430,   1000 ☆(売)
                ⋮
========= 2月△日 =========
 収益          181,253円( 2位)
 前日までの平均株価              343円
 ( 9: 0)   405,   1000
 (12:49)   425,   2000 ☆☆☆(売)
 (12:49)   425,   1000 ☆(売)
 (12:49)   425,   1000
                ⋮
========= 3月13日 =========
  (14:27)   330,   1000
```

図 4: 検出結果 (銘柄 B)