

# 金融市場での 人工知能技術の応用と 人工市場研究の紹介

和泉 潔

[izumi@sys.t.u-tokyo.ac.jp](mailto:izumi@sys.t.u-tokyo.ac.jp)

東京大学大学院 工学系研究科

# 人工知能学会 金融情報学研究会(SIG-FIN)

- 研究発表会 年2回(10月と3月)
- セミナー 年1回 (9月)
- JSAI全国大会 企画セッション(5月)

2016年9月8日 金融情報学セミナー

(1) 18:00-19:00

文書の分散表現と深層学習を用いた  
日銀政策変更の予想

塩野 剛志 (クレディ・スイス証券)

(2) 19:00-20:00

価格時系列を用いた株式市場の価格  
変動予想

野村 至紀 (シンプレクス・アセッ  
ト・マネジメント)

2017年3月10日(金)10:00-17:00(予定)

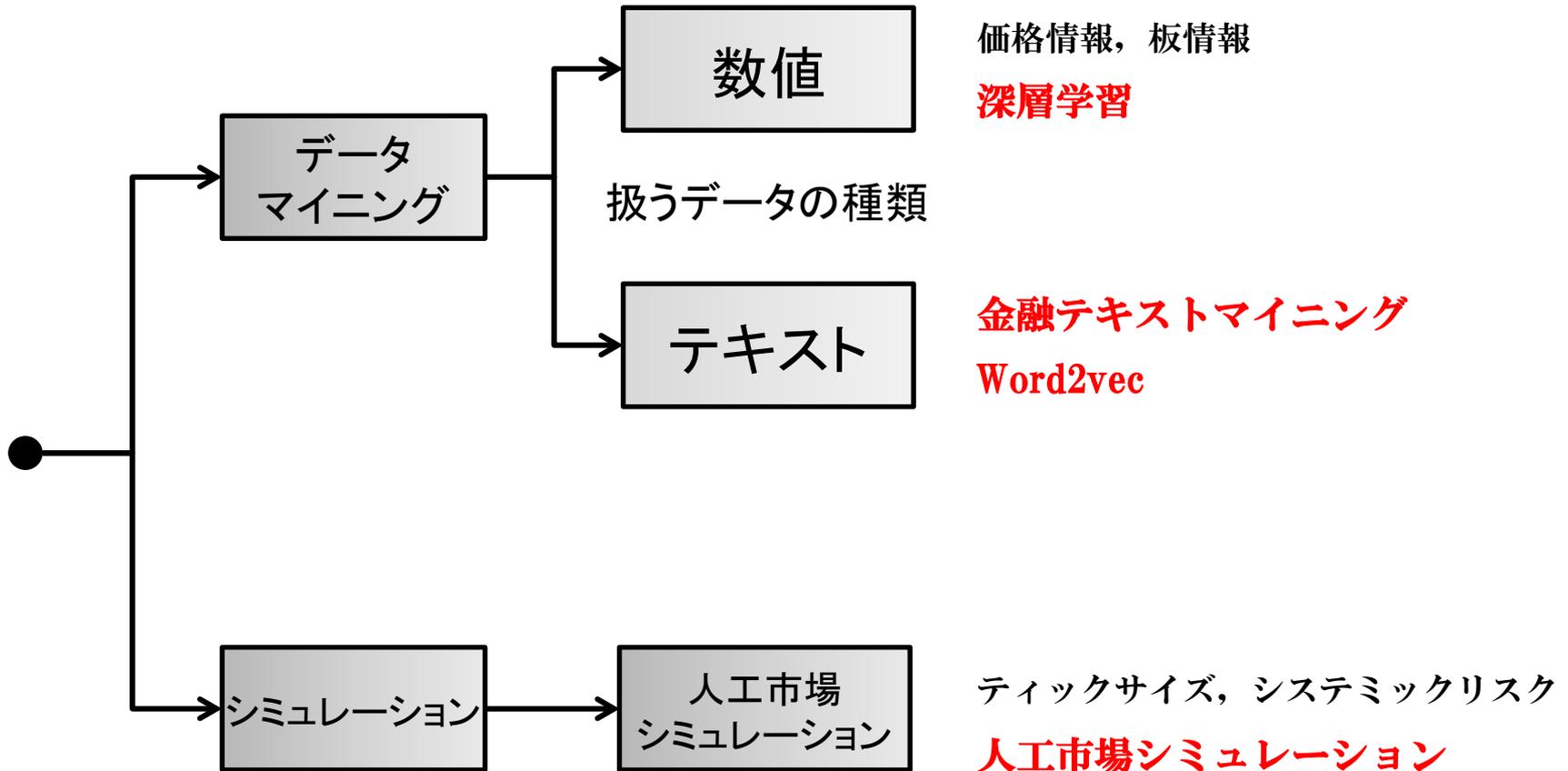
第18回研究会@FinGate (茅場町)

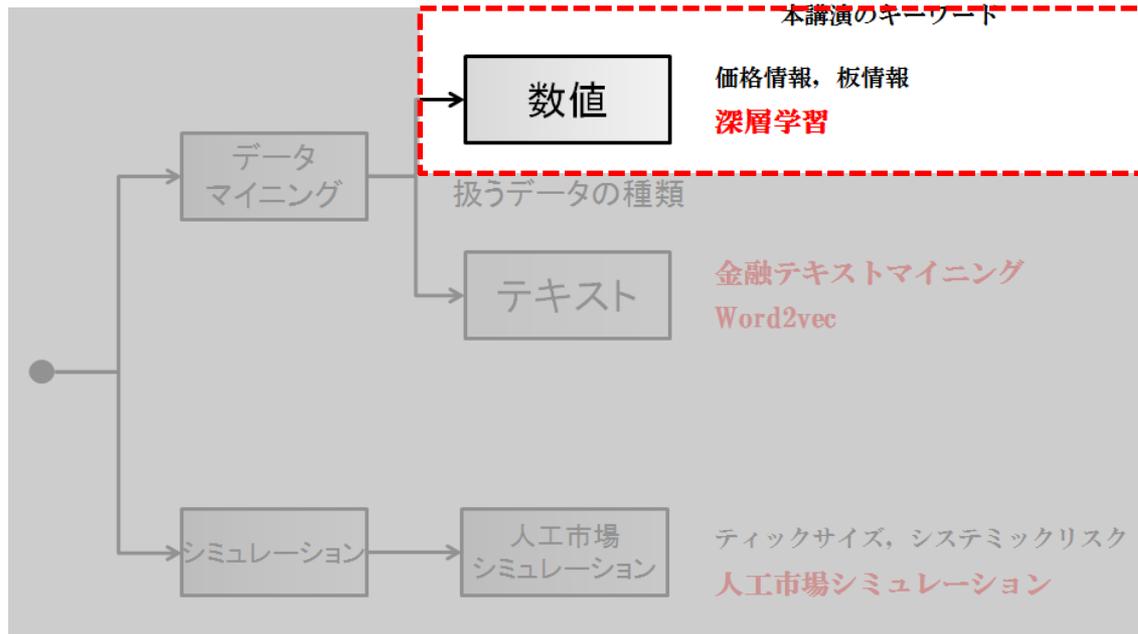
発表申込締切：2017年 2月6日 (月)

原稿提出締切：2017年 2月24日 (金)

# 金融市場でのAI技術の応用事例

本講演のキーワード



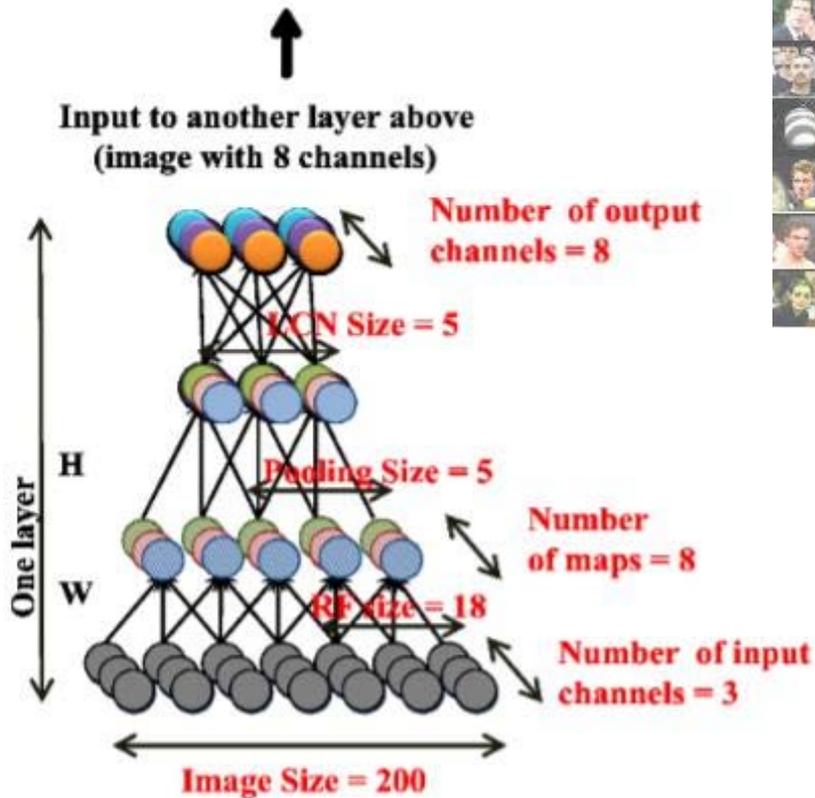


# 板情報の画像化 と 深層学習

中山・山田・和泉：高頻度板情報の時空間パターン分析による株価動向推定,人工知能学会全国大会(2015)

<https://kaigi.org/jsai/webprogram/2015/paper-328.html>

# 深層学習(deep learning)



人の顔や猫に対応するユニットが得られる

似ているものの認識や抽出が可能

入力:YouTubeの  
画像1000万枚



[http://static.googleusercontent.com/media/research.google.com/ja/archive/unsupervised\\_icml2012.pdf](http://static.googleusercontent.com/media/research.google.com/ja/archive/unsupervised_icml2012.pdf)

# 深層学習(deep learning)

Deep Learningによる為替市場の時系列データ分析事例(Alpaca社)

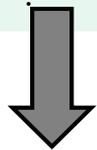
# 深層学習(deep learning)

国内外の金融機関で市場予測への利用例が出ている

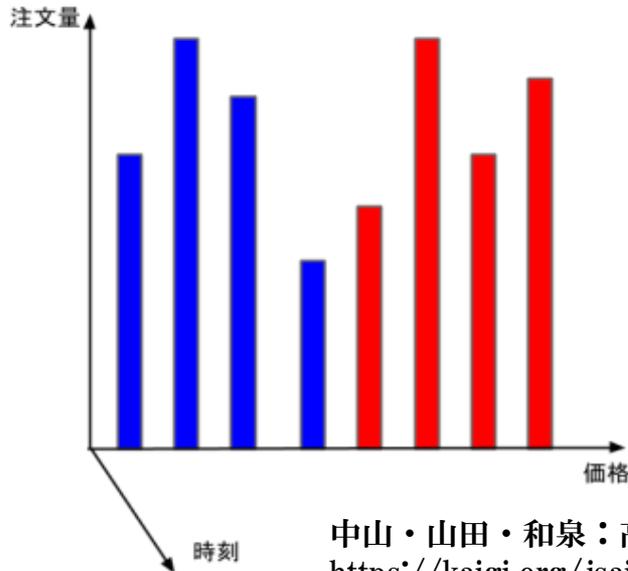
- ディープラーニングを用いたアルゴ取引
  - 7800次元の入力データから1時間後の株価の騰落を予想
  - 予測精度を平均+2.48%の改善

# 板情報の画像化

売り注文 株数(株)	気配値(円)	買い注文 株数(株)
	⋮	
150,000	130	
50,000	129	
	128	70,000
	127	200,000
	⋮	



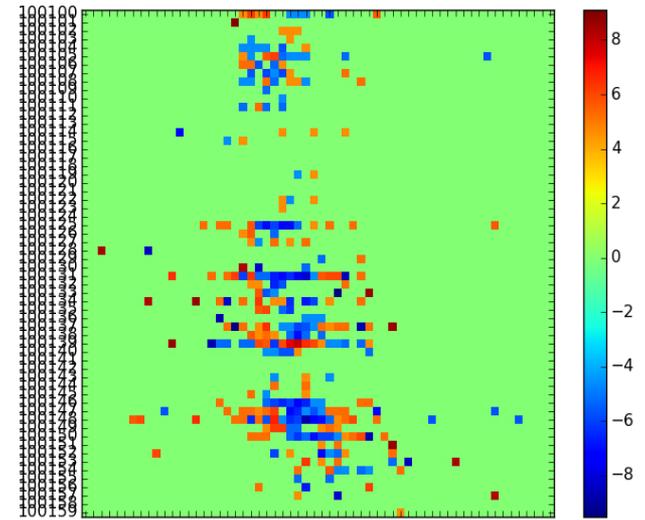
板を1秒毎の時間方向  
に並べていく



時間方向  
20/40秒  
or  
20/40 tick時間

1秒毎の各価格での板の変化  
を一つの点で表す

赤: Ask増/Bid減 青: Ask減/Bid増



価格方向  
20/40 tick



予測

10秒後の株価が上昇する/しない

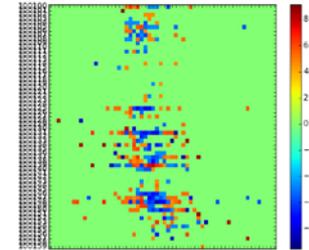
中山・山田・和泉：高頻度板情報の時空間パターン分析による株価動向推定,人工知能学会全国大会(2015)  
<https://kaigi.org/jsai/webprogram/2015/paper-328.html>

# 実験1

- 比較的流動性の高い5銘柄
  - 2013年07月01日から2014年06月30日までの1年間
  - 約75,000個ずつのデータを訓練とテストに分割
  - テストデータは、10秒後に上昇と下降の例が半々となるようにサンプリング

- 板情報を画像化したデータを入力
  - 20秒×20tick
  - 40秒×40tick

時間方向  
20/40秒



価格方向  
20/40 tick

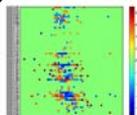
↓ 予測

10秒後の株価が上昇する/しない

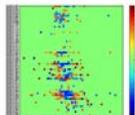
- それぞれを入力として、畳込みニューラルネットワークで10秒後のリターンの学習を行った

# 畳み込みニューラルネットワーク

注文  
キャンセル  
約定



10秒後の  
株価が  
上昇する/  
しない

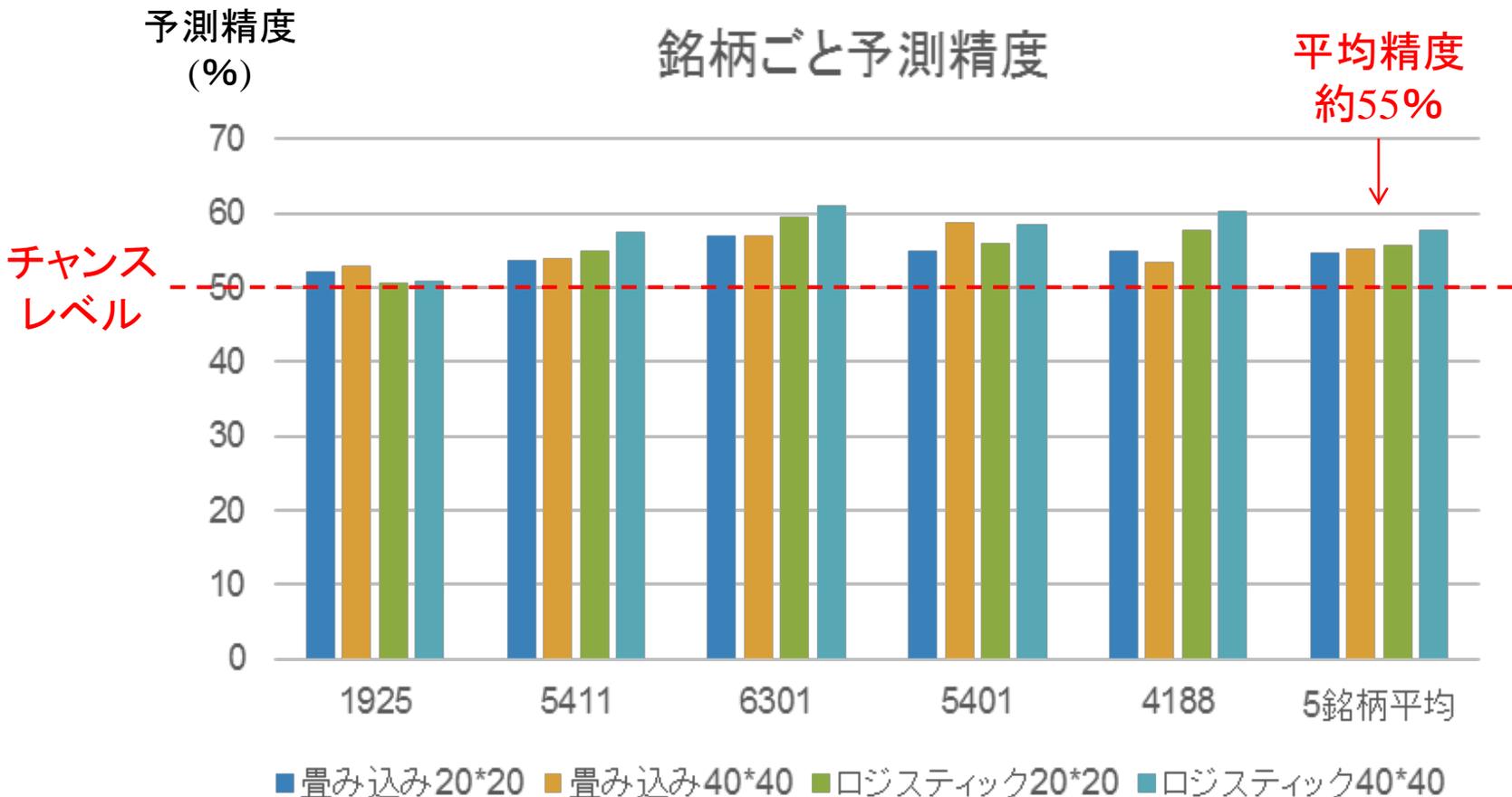


人間の視覚野を模倣した手法

前半の畳み込み層の各フィルタの係数とバイアス項、およびその後のニューラルネットワーク層の係数とバイアス項を、確率的勾配降下法で学習

# 実験1 20/40秒間の板画像

- 20×20画像よりは40×40画像の方が精度が高い傾向が見られた
- 銘柄ごとのばらつきの方が大きい



# 実験2 板の変化、ティック時間、 大きな注文時に着目

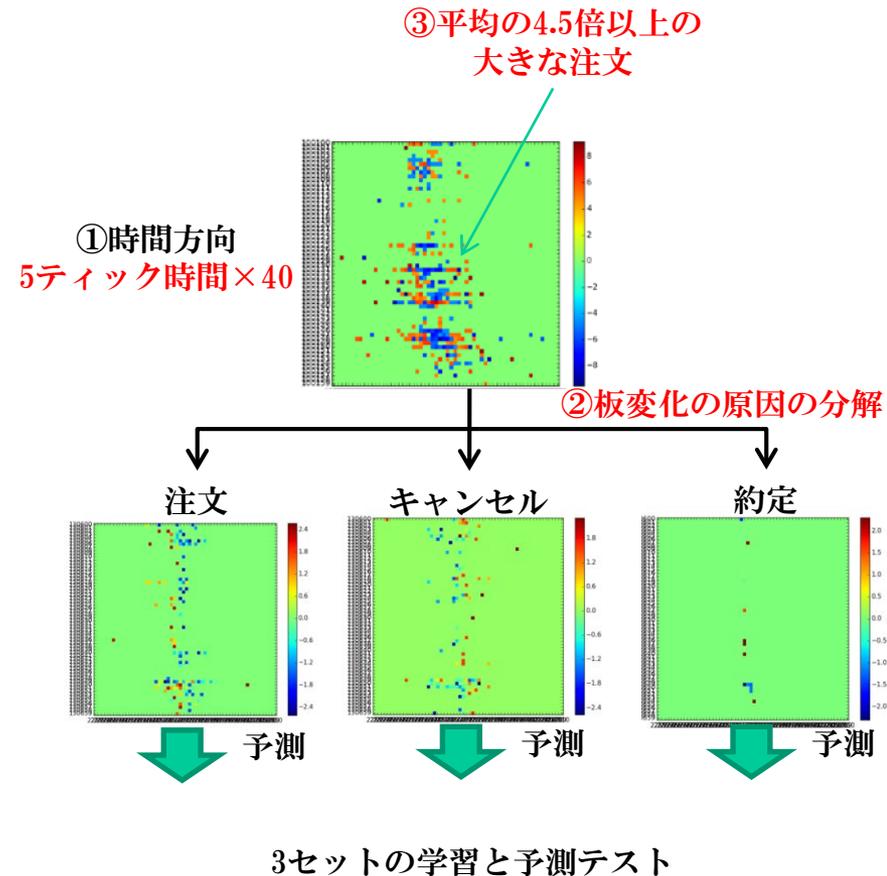
## ① 板の変化の原因を区別する

- 3種類の情報に分解
  - ・ 注文を表す画像
  - ・ キャンセルを表す画像
  - ・ 約定を表す画像

## ② 注文の多い時間帯と少ない時間帯での違いへの対応

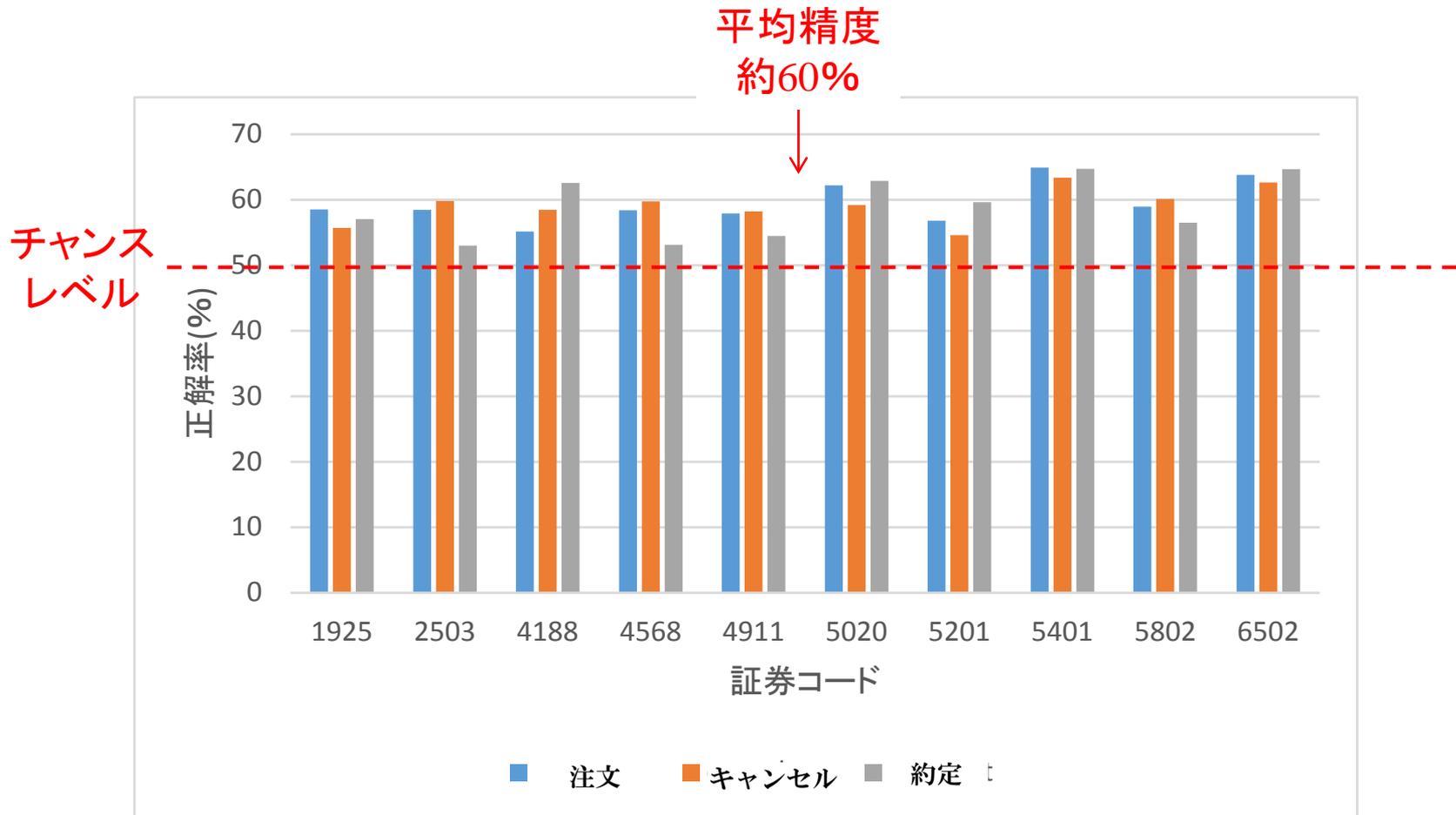
- 画像の縦を40秒から5ティック時間×40へ変更

## ③ 直近の5ティック時間×40で、平均注文量の約4.5倍の大きな注文が入った時のみの画像データで、学習・テストを行った



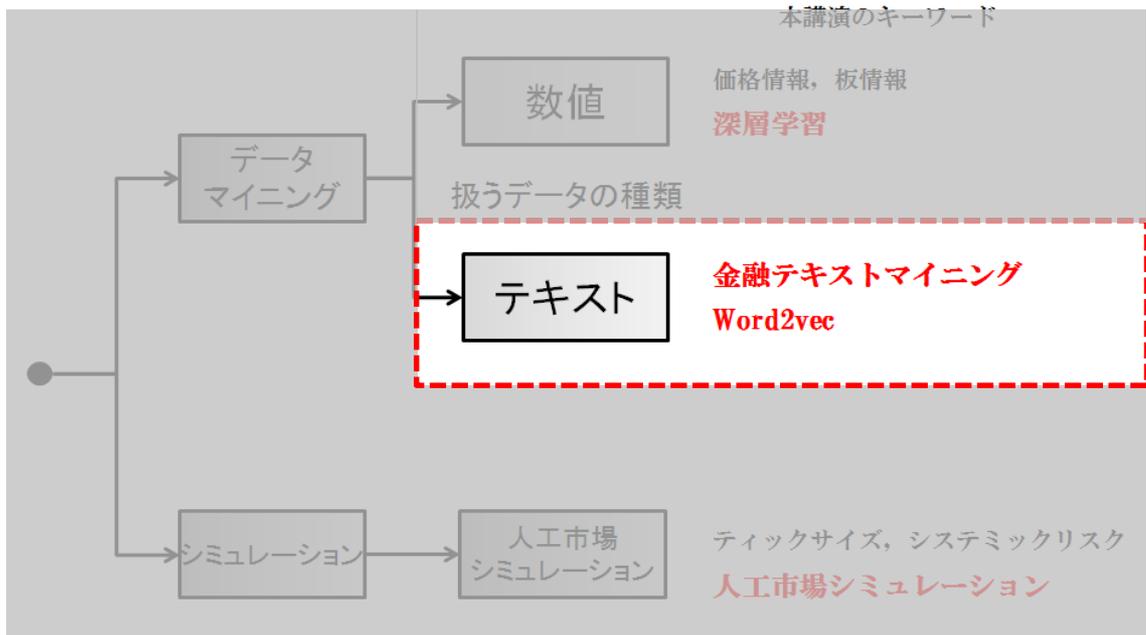
# 実験2 板の変化、ティック時間、 大きな注文時に着目

- 直近の5ティック時間×40で、平均注文量の約4.5倍の大きな注文が入った時のみで、学習・テストを行った



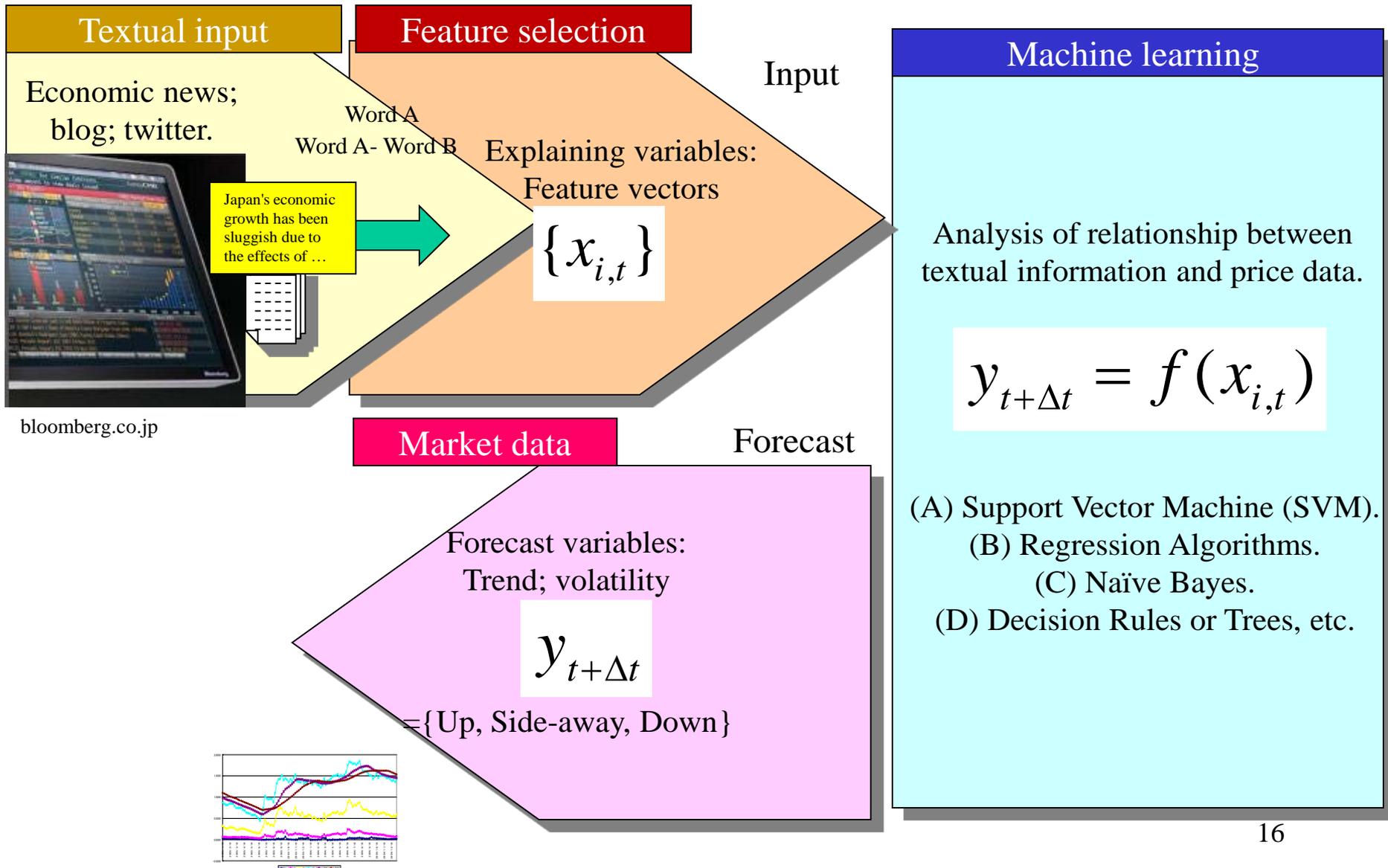
# まとめ 1

- 数値データの解析については、人工知能が得意
  - スピード+膨大なデータの取り扱い
  - ルーチンな予想は人間では勝てない
- しかし、市場の潮目の移り変わりは不得意
  - ニュースなどによって市況が急激に変わると大きく負ける
  - 長期的な見通しは苦手
- ニュースが与える市場への影響分析
  - 金融テキストマイニング



# 金融テキストマイニングの 最新研究事例

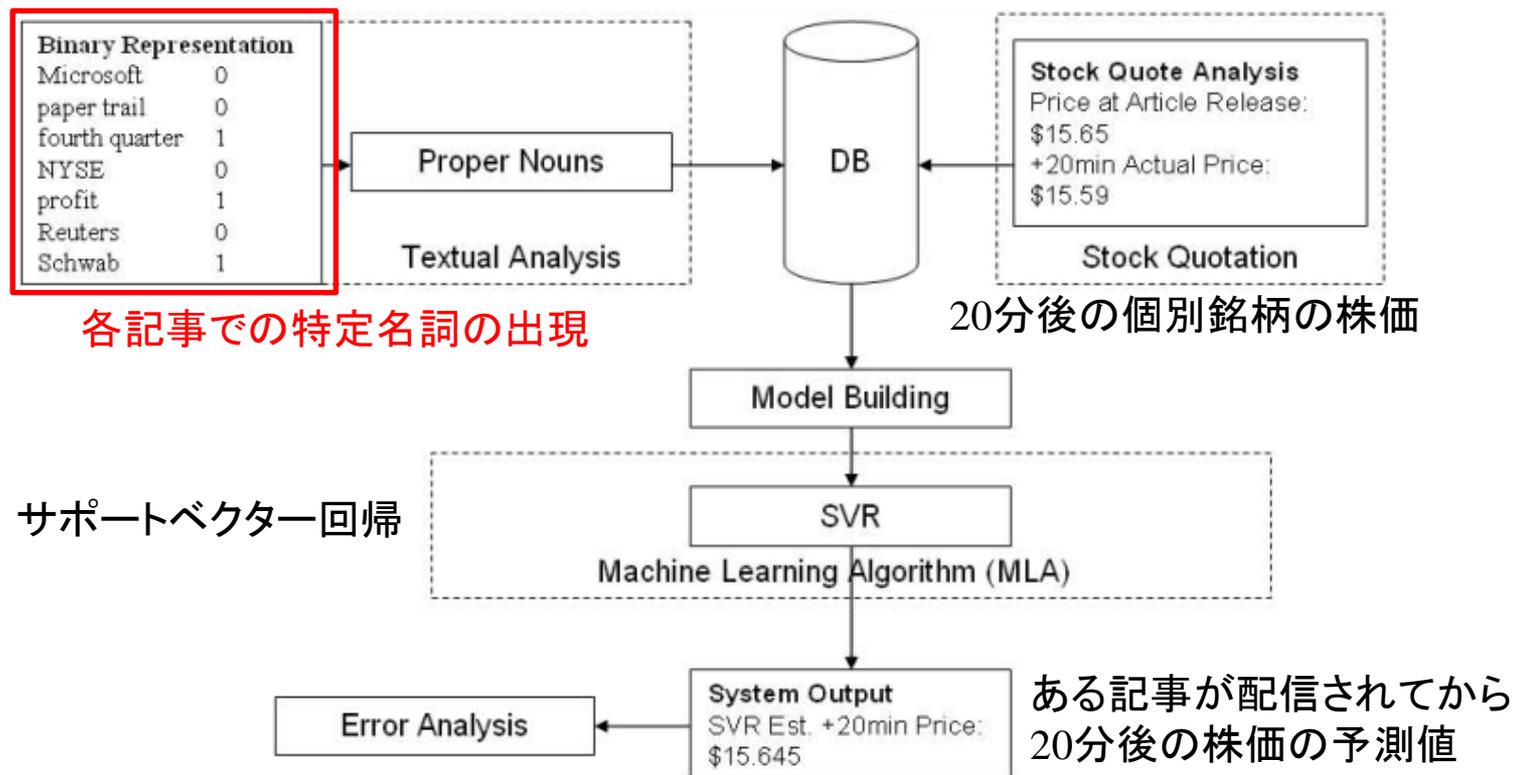
# Framework of text mining in financial markets



# ニュース記事からの特徴量抽出

R. Schumaker, H. Chen, Computer, vol. 43, no. 1, pp. 51-56, 2010

- 2005年10月26日～11月28日(5週間)
  - S&P500の構成銘柄
  - 取引時間(10:30am-3:40pm)のニュースに限定 2809記事



# 運用テスト

R. Schumaker, H. Chen, Computer, vol. 43, no. 1, pp. 51-56, 2010

- 2005年10月26日～ 11月28日(5週間)
  - S&P500の構成銘柄
  - 20分間に1%以上の変動が予想された銘柄を売買
  - ファンドとの取引結果(年率リターン)の比較

上位10社のファンドのうち6社よりも好成績

	Return
ProFunds Ultra Japan Inv (UJPIX)	24.73%
ProFunds Ultra Japan Svc (UJPSX)	24.59%
American Century Global Gold Adv (ACGGX)	12.96%
American Century Global Gold Inv (BGEIX)	12.93%
<b>AZFinText</b>	<b>8.50%</b>
Quantitative Advisors Emerging Markets Instl (QEMAX)	8.16%
Quantitative Advisors Emerging Markets Shs (QFFOX)	8.15%
<b>S&amp;P 500 Index</b>	<b>5.62%</b>
Lord Abbett Small-Cap Value Y (LRSYX)	5.22%
Lord Abbett Small-Cap Value A (LRSCX)	5.19%
Quantitative Advisors Foreign Value Instl (QFVIX)	4.99%
Quantitative Advisors Foreign Value Shs (QFVOX)	4.95%

Table 1. Simulated Trading results of the Top 10 Quants

S&P500銘柄の  
ファンドでは1位

	Return
<b>AZFinText</b>	<b>8.50%</b>
Vanguard Growth & Income (VQNPX)	6.44%
BlackRock Investment Trust Portfolio Inv A (CEIAX)	5.48%
RiverSource Disciplined Equity Fund (ALEIX)	4.69%

Table 2. Simulated Trading results of S&P 500 quants

# Tweetsからの特徴量抽出

J. Bollen;H. Mao;X.-J. Zeng. Journal of Computational Science 2 (1), 1-8, 2011

- 2008年2月28日～11月28日の 9,853,498 tweets (約1千万)
  - ユーザ数 2.7百万
  - 3.2万tweets/1日
- 心的状態を明言しているtweetsだけを分析
  - “i feel”, ”i am feeling”, “i’m feeling”, “i dont feel”, “I’m”, “I am”, “makes me” を含む
- Google-Profile of Mood States (GPOMS)指数(日次)を抽出
  - 気分プロフィール検査(POMS)@心理学の6尺度がベース
    - 「友好的な」「不機嫌な」「活発な」「限界ギリギリの」「パニック状態の」等の72表現の質問紙調査
    - 平穏、警戒、確信、活気、善意、幸福
  - POMSの72表現に関連する964語の辞書とのマッチ
    - Google(2006)の4,5-gram共起語(25億語)を使用

# ダウ平均株価(DJIA)との関係性

J. Bollen;H. Mao;X.-J. Zeng. Journal of Computational Science 2 (1), 1-8, 2011

## • GPOMS指数とDJIAとのGranger 因果性検定

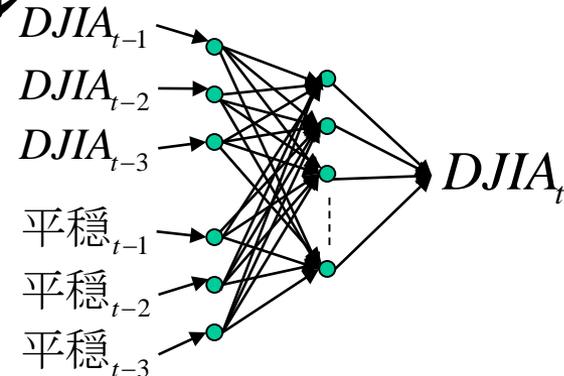
- 2008年2月28日～11月3日
- 「平穩」が2-5日後のDJIAとの因果性があった

Lag	OF	平穩	警戒	確信	活気	善意	幸福
		Calm	Alert	Sure	Vital	Kind	Happy
1 day	0.085*	0.272	0.952	0.648	0.120	0.848	0.388
2 days	0.268	0.013**	0.973	0.811	0.369	0.991	0.7061
3 days	0.436	0.022**	0.981	0.349	0.418	0.991	0.723
4 days	0.218	0.030**	0.998	0.415	0.475	0.989	0.750
5 days	0.300	0.036**	0.989	0.544	0.553	0.996	0.173
6 days	0.446	0.065*	0.996	0.691	0.682	0.994	0.081*
7 days	0.620	0.157	0.999	0.381	0.713	0.999	0.150

(p-value < 0.05: \*\*, p-value < 0.1: \*)

## • Self-organizing Fuzzy Neural Network(SOFNN)による予測モデル

- 訓練: 2008年2月28日～11月28日
- テスト: 2008年12月1日～19日
- 方向性の予測: 86.7%
- DJIAだけでも73.3%



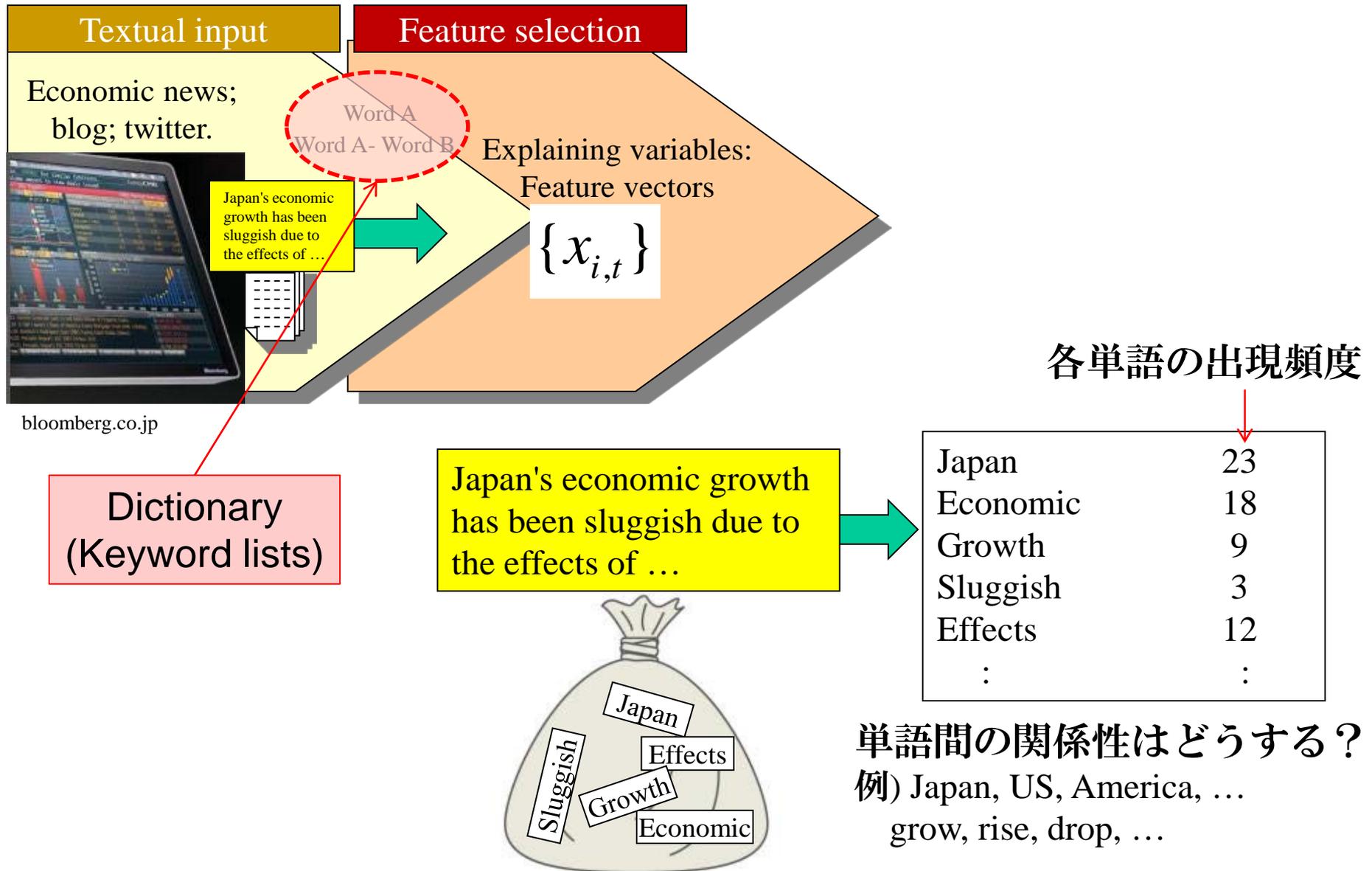
# 金融テキストマイニングの方向性

- 直接、価格を予測することは難しい

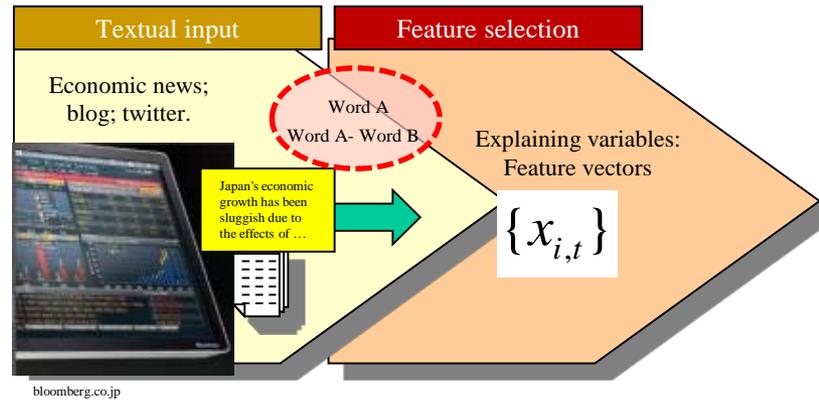


- まず市場に関連する経済環境・状況を判別する方向に
  - 日銀テキスト・インデックス(クレディ・スイス証券)
  - 野村 A I 景況感指数(野村証券)
  - 日銀金融経済月報を用いた金利動向推定  
和泉・余野 他, 英文経済レポートのテキストマイニングと長期市場分析, ジャフイージャーナル 実証ファイナンスとクオンツ運用, pp.12-31, 朝倉書店, 2013.
- 金融の文脈に応じた単語辞書を大規模なテキストデータから自動作成
  - 最近の技術: word2vecによる単語の分散表現

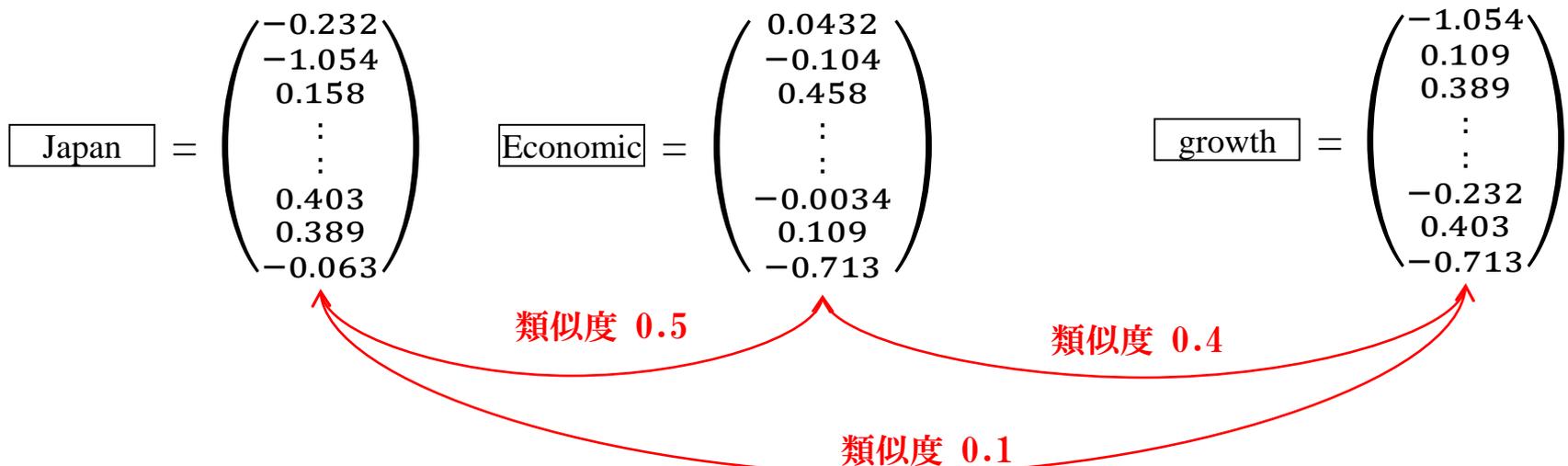
# 既存の手法: Bag-of-words



# Word2vecによる単語の分散表現

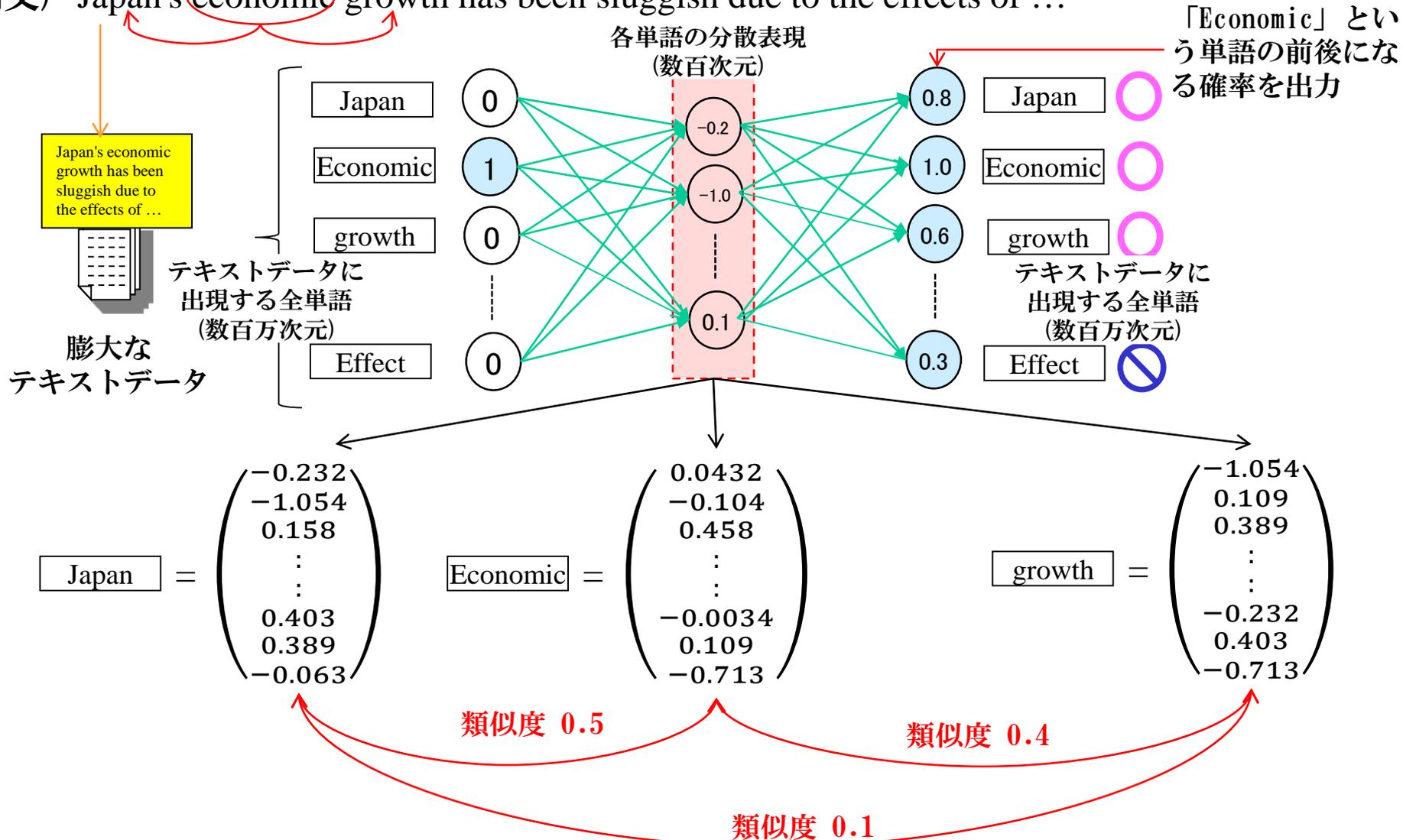


## 単語の分散表現



# Word2vecによる単語の分散表現

例文) Japan's economic growth has been sluggish due to the effects of ...



テキストデータの中で、前後に来る単語が似ている単語同士は類似度が高い

# Word2vecによる単語の分散表現の例

- Textual data (10年間分の経済ニュース記事)
  - financial news from 2003 to 2013
  - 24.5GB
- Vector representation: 200 dim.
- レクサス - トヨタ + 日産 = ?  
「レクサス」から「トヨタ」という要素を抜いて  
「日産」という要素を足した単語を一言で表すと

順位	単語	類似度
1.	インフィニティ	0.572
2.	RAV	0.514
3.	カローラ	0.511
4.	キャデラック	0.500
5.	プリウス	0.497

どれが商品名、どれが会社名  
と教えてなくても、  
コンピュータが自動的に  
単語間の関係性を学習する

## Word2vecによる単語の分散表現の例

### • 金融危機+開始

1	着手	0.440
2	バブル崩壊	0.436
3	再開	0.428
4	コモディティロング	0.411
5	完了	0.407

### • 金融危機 + 規制

1	監督	0.476
2	規制強化	0.469
3	規則	0.440
4	バーゼル	0.432
5	システミックリスク	0.432

### • ホンダ + 欧州

1	キャノン	0.593
2	ソニー	0.550
3	トヨタ	0.549
4	自動車	0.483
5	日産	0.459

### • インフレ + 財政赤字

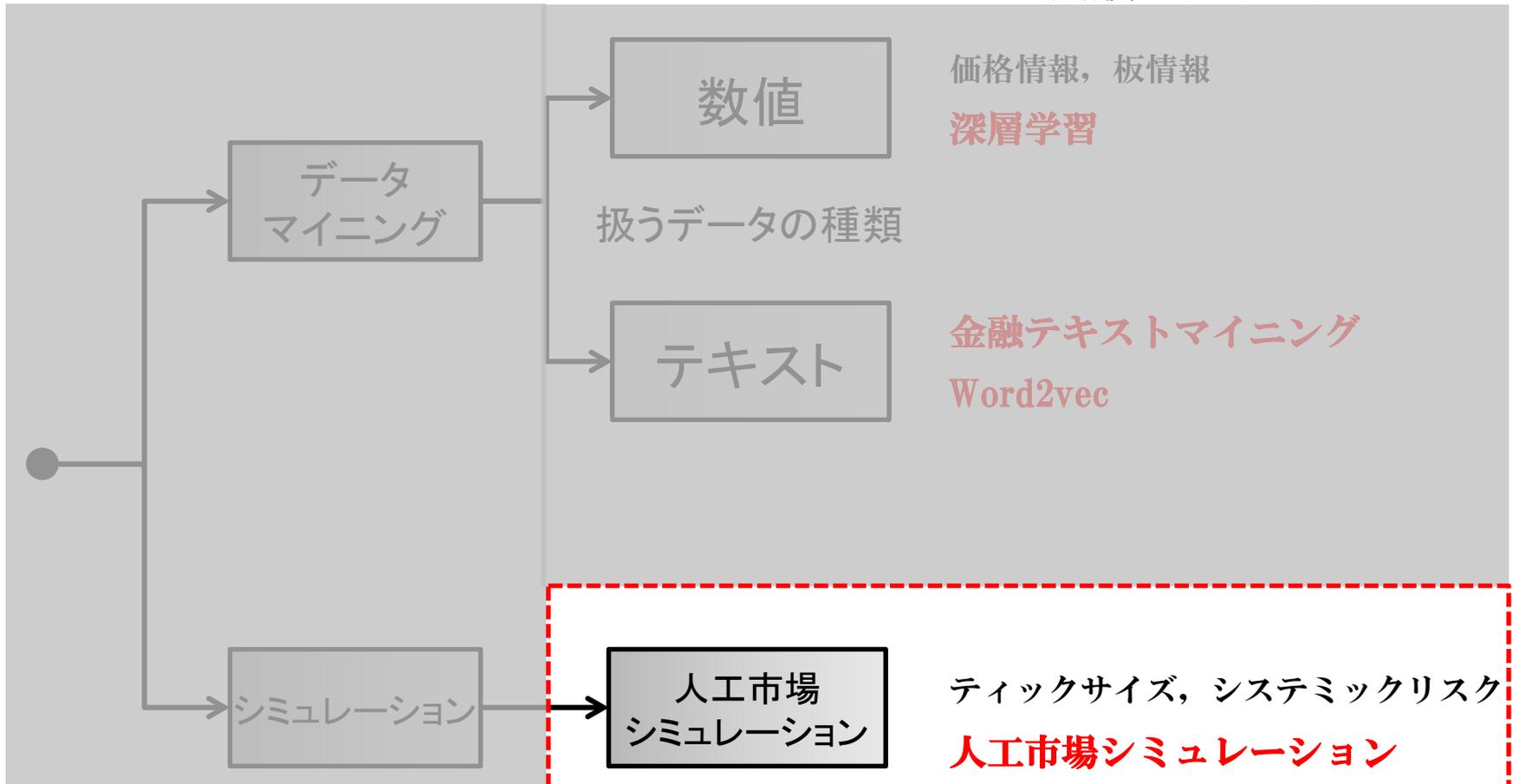
1	経済成長	0.535
2	経済成長率	0.481
3	財政支出	0.475
4	物価上昇率	0.469
5	田植え	0.458

# まとめ 2

- 決まり切ったキーワードのマッチングは、人工知能が得意
  - スピード+膨大なデータの取り扱い
  - ルーチンな予想は人間では勝てない
- しかし、常識の獲得・文脈理解 がまだ発展途上
  - 金融単語間の関係性
- 新技術(word2vecによる単語の分散表現)で、金融文脈での単語の関係を自動的に学習できるかもしれない

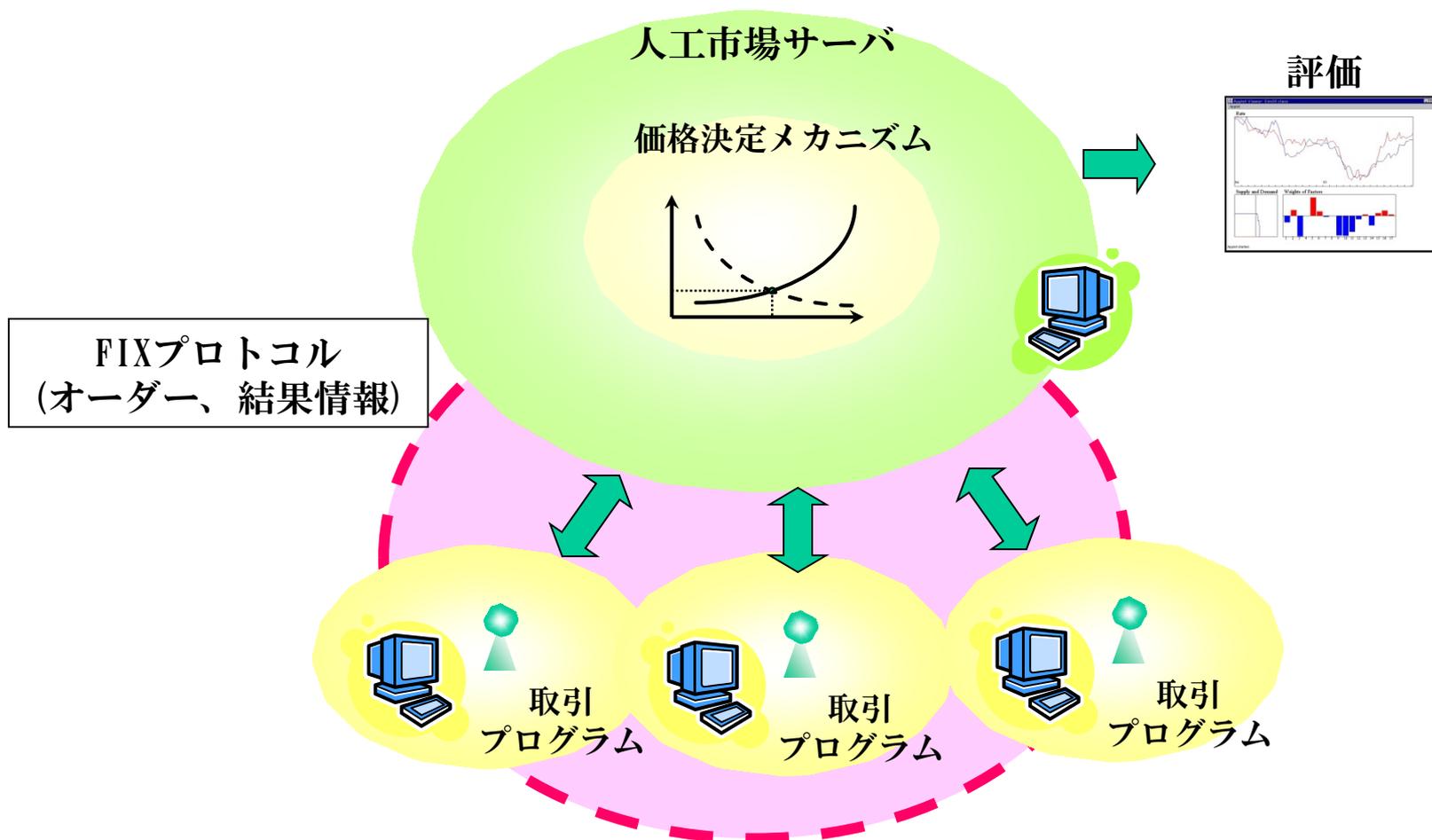
# 金融市場でのAI技術の応用事例

本講演のキーワード



# 人工市場による取引アルゴリズムの評価

- 人工市場サーバに自動売買プログラムを参加させる
  - 多様な市場環境でのテストを行える。



# 自動取引プログラムの評価

- 8種類のアプローチ × 3種類のパラメータ
- 1. ゴールデンクロス: 長短トレンドの交差
- 2. HLBand: 過去の最高値(最安値)の突破
- 3. MACD (Moving Average Convergence / Divergence): 長短  
トレンドの差とその移動平均の比較
- 4. Envelope: (移動平均 ± P%)との比較
- 5. Psychological line: 値上日数と値下日数の比率
- 6. RSI (Relative Strength Index): 値上幅と値下幅の比率
- 7. Bayesian: 直近の値動きパターンとの相関
- 8. テクニカル分析組合せ: RSIと移動平均の混合

# 2005年と2006年の 過去データを用いたバックテスト

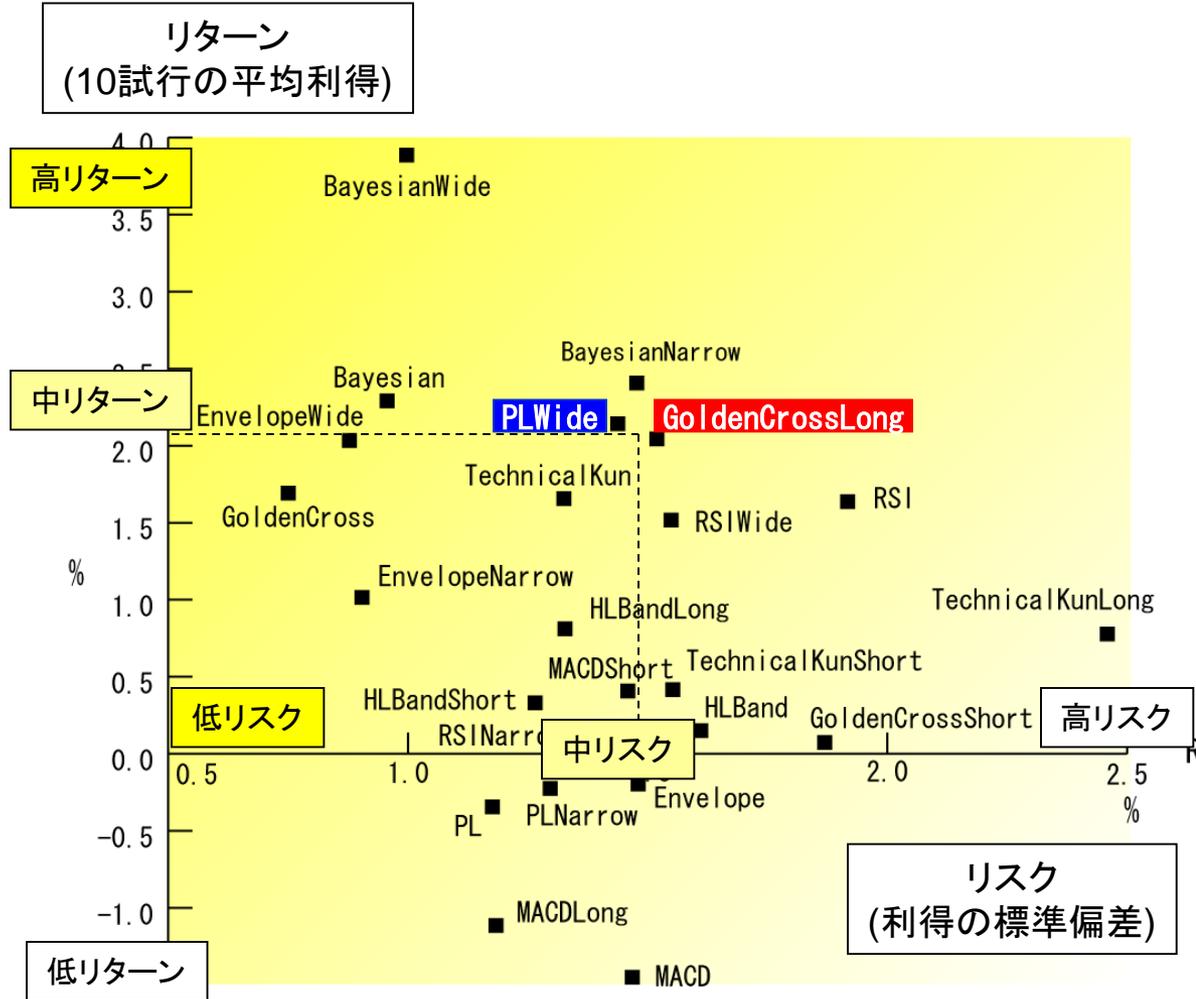
2005年

1	<b>GoldenCrossLong</b>	<b>12.92</b>
2	BaysianWide	6.35
3	HLBandShort	3.78
	:	
19	<b>PsychologicalLineWide</b>	<b>-0.89</b>
20	PsychologicalLine	-1.15
21	RSINarrow	-1.26
22	RSIWide	-1.64
23	RSI	-4.40
24	Envelope	-4.48

2006年

1	<b>PsychologicalLineWide</b>	<b>15.49</b>
2	TechniKunLong	14.32
3	Envelope	12.61
	:	
19	HLBandLong	-0.07
20	PsychologicalLineNarrow	-0.79
21	<b>GoldenCrossLong</b>	<b>-1.16</b>
22	MACDLong	-1.23
23	HLBandShort	-2.02
24	GoldenCrossShort	-5.64

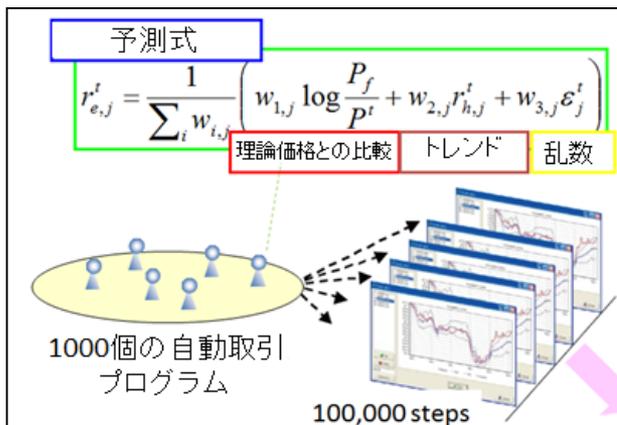
# 人工市場による評価 (10試行)



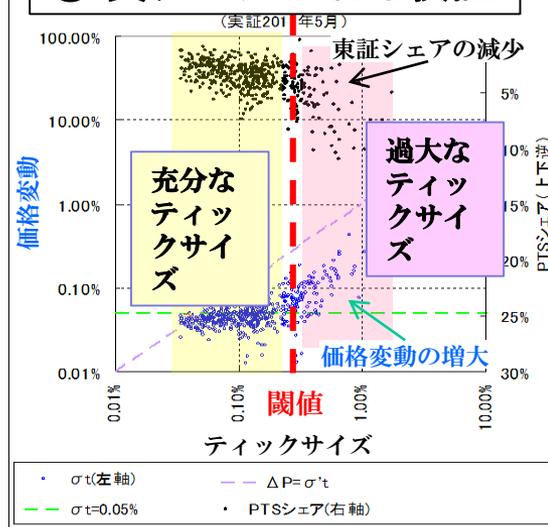
# 人工市場シミュレーションを用いた 現実の金融市場の制度検証

東京証券取引所との共同研究で、シミュレーションと実データ分析により、ティックサイズ(注文価格の最小単位)の閾値を新たに発見した\*1

## ① シミュレーション



## ② 実データによる検証

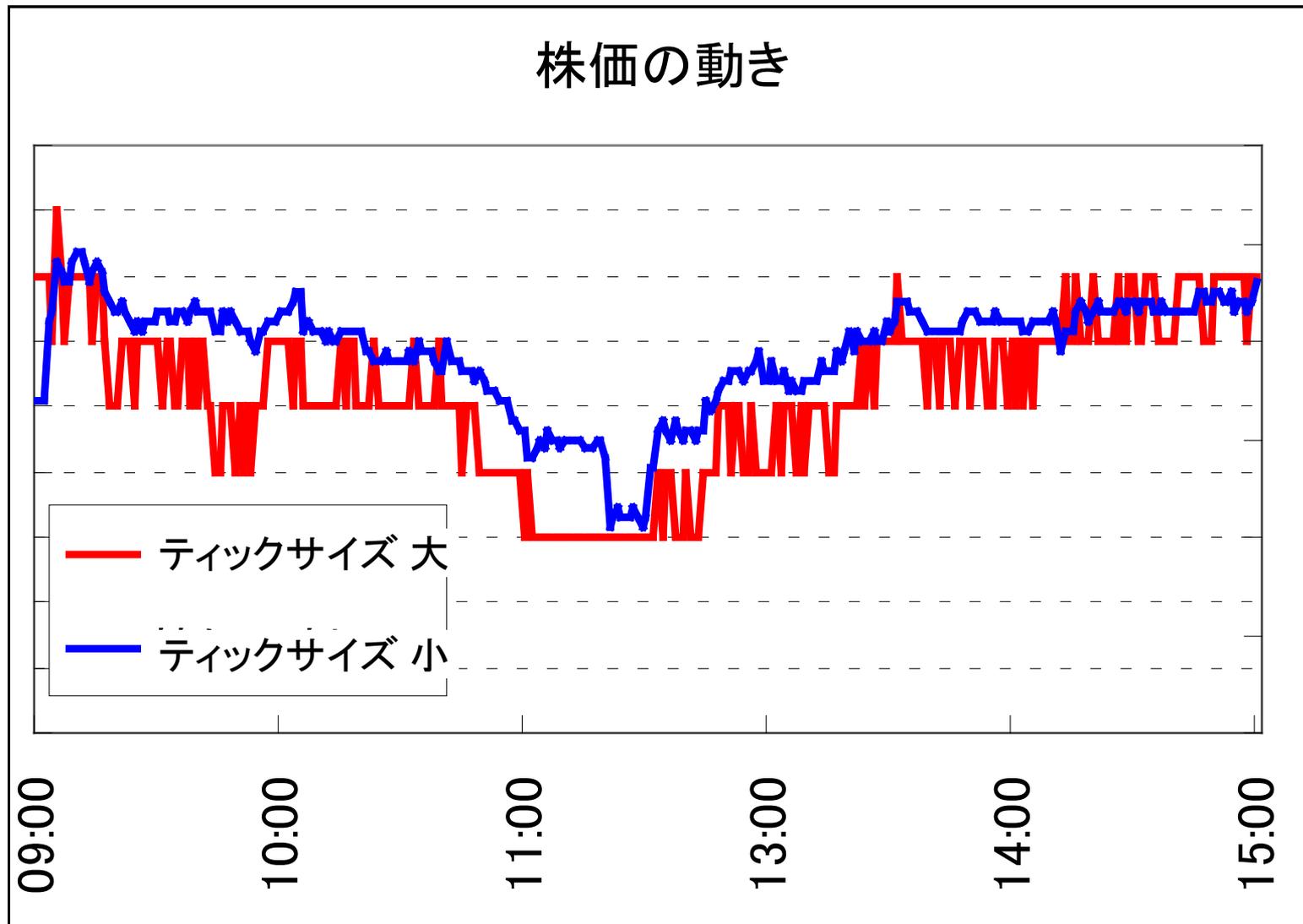


## ③ 実社会の制度設計

2014年の東証ティックサイズ変更において  
銘柄の絞り込みの方針に  
反映された

\*1 Mizuta, Izumi, et.al., The 8th International Workshop on Agent-based Approach in Economic and Social Complex Systems , 2013

# 同じような事業を行っている2つの企業の株価の動き



東証には株価の刻み幅(ティックサイズ)が大きすぎる会社がある  
↑ 投資家が困る場合も ⇒ 他の取引所で取引されてしまう?!

# 米国におけるニューヨーク証券取引所のシェア



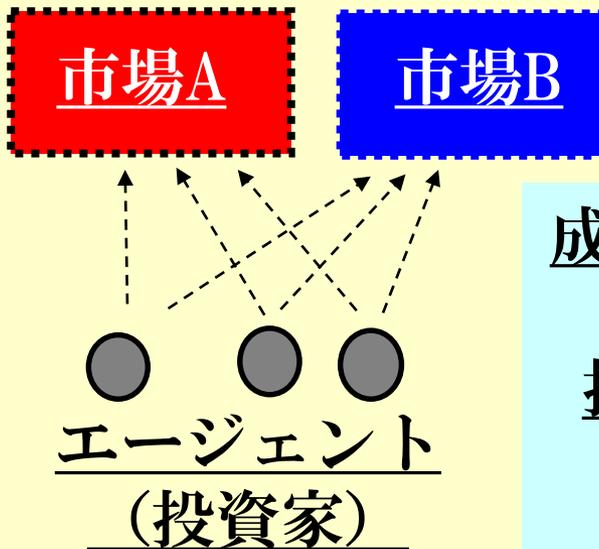
清水葉子, 金融庁金融研究センター ディスカッションペーパー 2013年5月

<http://www.fsa.go.jp/frtc/seika/discussion/2013/01.pdf>

2年程度で支配的地位から陥落

# 1銘柄2市場の人工市場モデル

ティックサイズのみ異なる市場A、Bで  
どのように出来高シェアが移り変わるかを分析

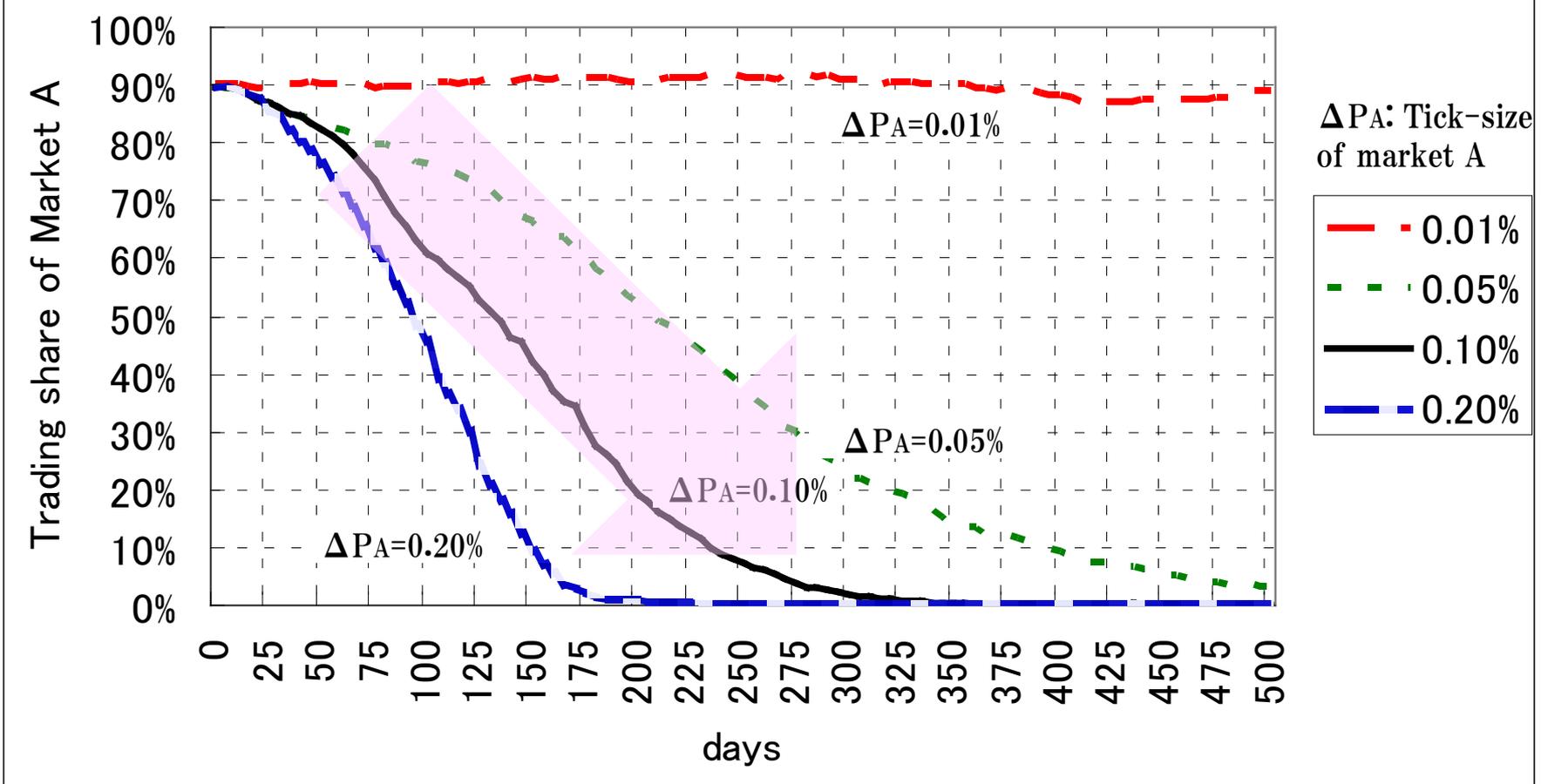


成行注文：有利な価格で売買  
できる市場を選択  
指値注文：各市場の過去の  
総出来高に比例して配分

コンピュータ内で  
シミュレーション

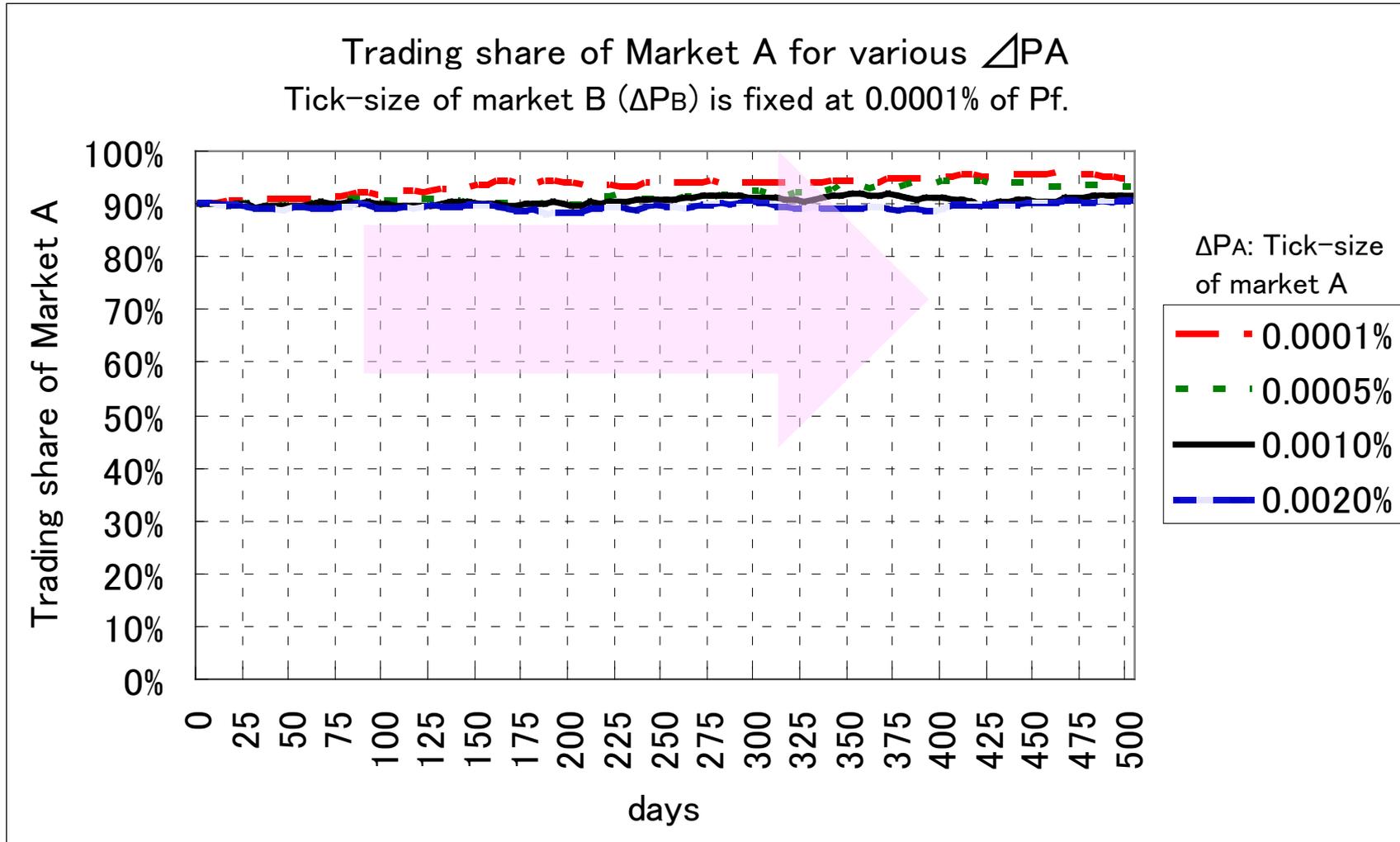
# Tick Size of Market B $\Delta P_B=0.01\%$ , Tick Size is not small

Trading share of Market A for various  $\Delta P_A$   
Tick-size of market B ( $\Delta P_B$ ) is fixed at 0.01% of Pf.



Tick Size of Market A,  $\Delta P_A$  is larger,  
Market A is taken trading volume share faster

# $\Delta PB = 0.0001\%$ , Tick Size is enough small

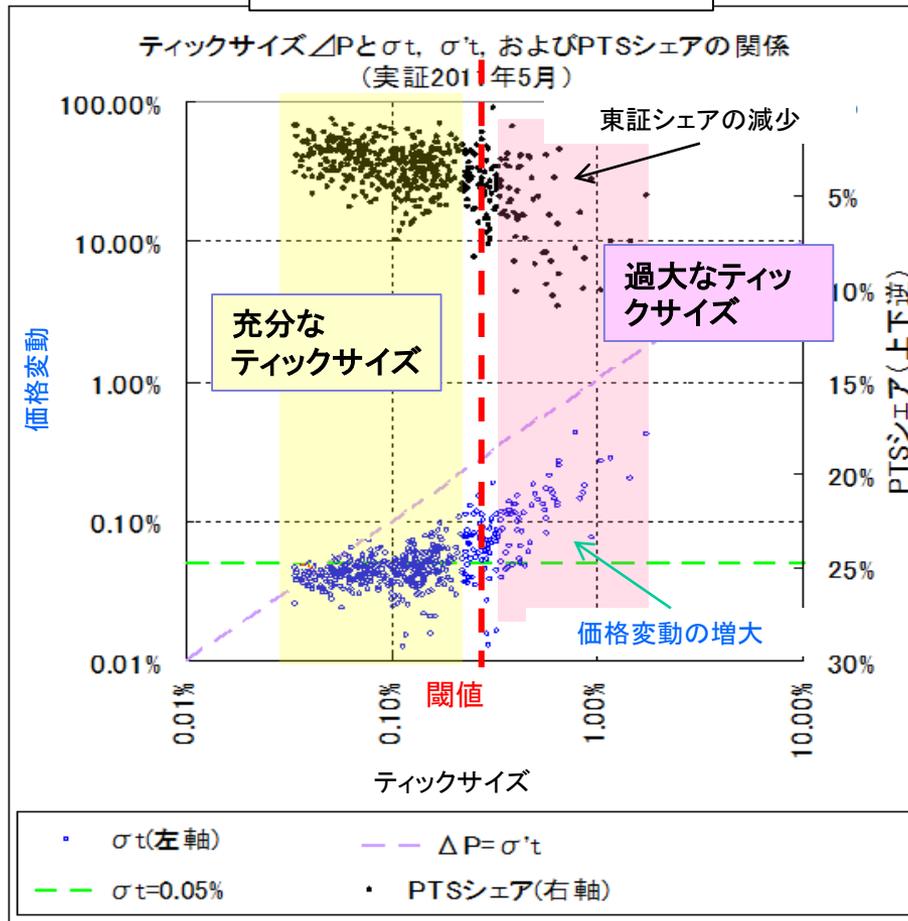


Market B can hardly take the share in spite that  $\Delta PA$  is very larger than  $\Delta PB$

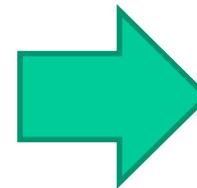
# 人工市場シミュレーションを用いた 現実の金融市場の制度検証

東京証券取引所との共同研究で、シミュレーションと実データ分析により、ティックサイズ(注文価格の最小単位)の閾値を新たに発見した\*1

## 実データによる検証



## 実社会の制度設計



2014年の東証ティックサイズ変更において  
銘柄の絞り込みの方針に  
反映された

# 瞬間大暴落の再現(100銘柄・5万エージェントの市場シミュレーション)

- Dow Jones plunged about 1,000 points (about 9%) only to recover those losses within minutes.
- CFTC-SEC report says
- ...HFTs and other traders drove the price of the **E-Mini S&P 500** down approximately 3%...
- ...During this same time **cross-market arbitrageurs** who did buy the E-Mini S&P 500 simultaneously sold equivalent amounts in the equities markets, driving the price of SPY also down approximately 3%

\*U.S. Commodity Futures Trading Commission and U.S. Securities & Exchange Commission

Picture source: [http://money.cnn.com/2010/10/01/markets/SEC\\_CFTC\\_flash\\_crash/index.htm](http://money.cnn.com/2010/10/01/markets/SEC_CFTC_flash_crash/index.htm)



Shock transfer

<http://www.abeldanger.net/2010/10/flash-crash-arbitrage-trades-contract.html>

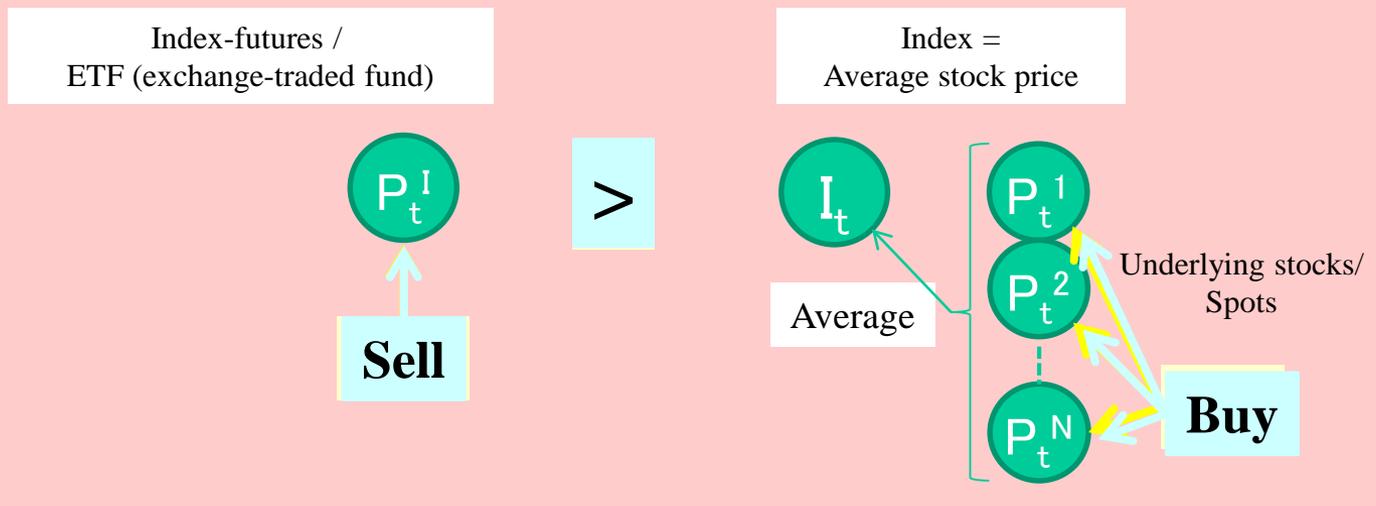
# 高速裁定取引が市場を不安定化させるのか？

## ETF Arbitrage May Be Driving Market Volatility

By Caitlyn Grudzinski | 08/02/12 - 10:47 AM EDT

NEW YORK (TheStreet) -- Since their creation in 1993, ETFs have opened up a new channel of arbitrage for market participants and

Arbitrage is a style of trading, which can profit from a market imbalance, a price difference among similar assets.



simply be due to macro events that affect the whole market.

外的ショック: リコール、工場火災、誤発注

# 人工市場シミュレーション

Stock 3..100

$$P_t^{ETF*} = \frac{P_t^{1*} + P_t^{2*}}{2}$$

**Stock 1** (20% ↓)

Fundamental price:  $P_t^{1*}$   
Market price:  $P_t^1$

Sell	Price	Buy
84	101	
	100	
	99	
	98	124
	97	77

**Stock 2**

Fundamental price:  $P_t^{2*}$   
Market price:  $P_t^2$

Sell	Price	Buy
84	311	
103	310	
	309	156
	308	124
	307	77

**ETF** (10% ↓)

Fundamental price:  $P_t^{ETF*}$   
Market price:  $P_t^{ETF}$

Sell	Price	Buy
32	205	
56	204	
	203	86
	202	74
	201	44

ショック伝播?



100 銘柄シミュレーション  
個別銘柄トレーダ (50,000 体)  
裁定取引トレーダ (100 体)

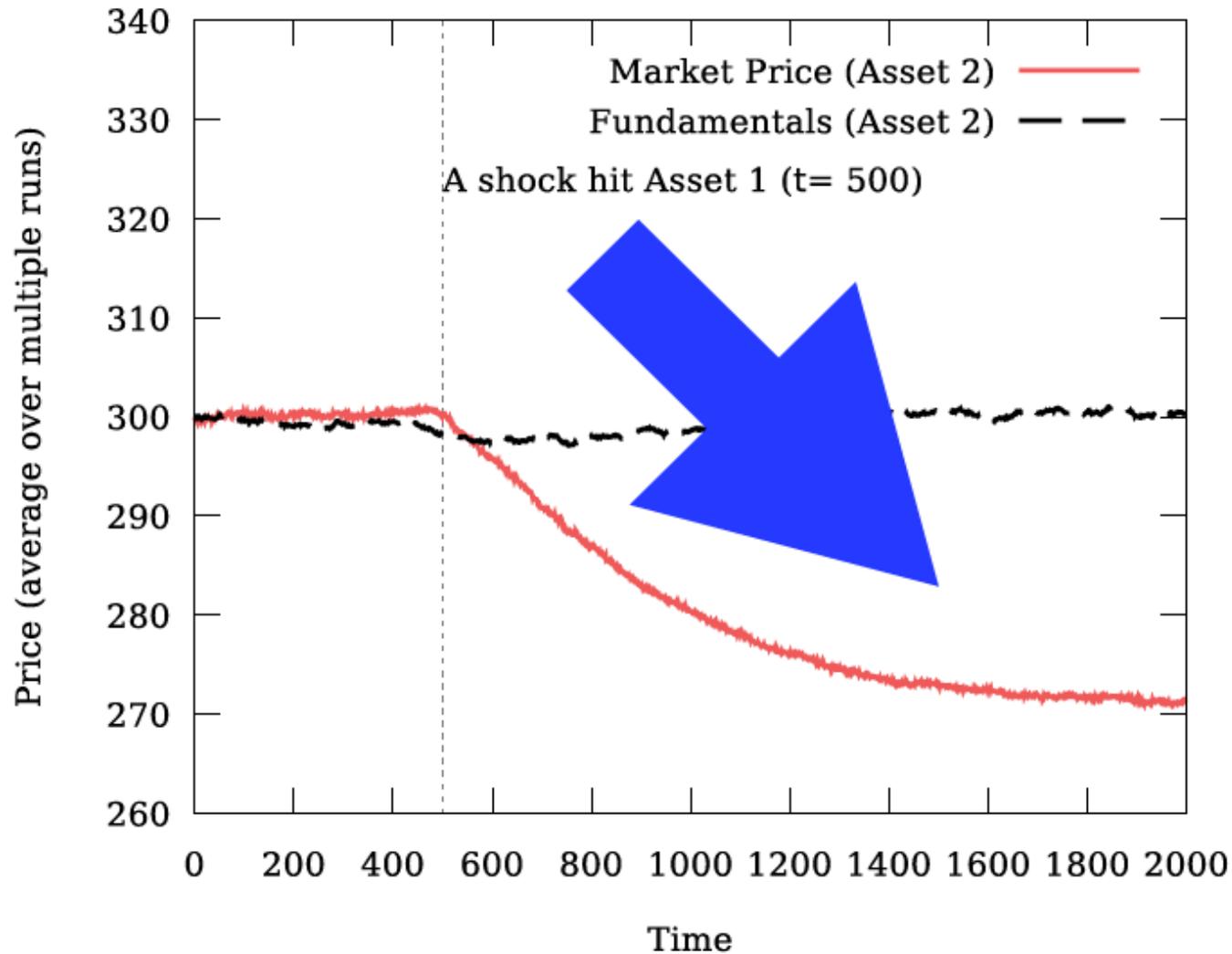
裁定取引の対象銘柄: 10 銘柄

ショック周辺 20 分間のシミュレーション  
あるショックが他の銘柄へ伝搬するか?

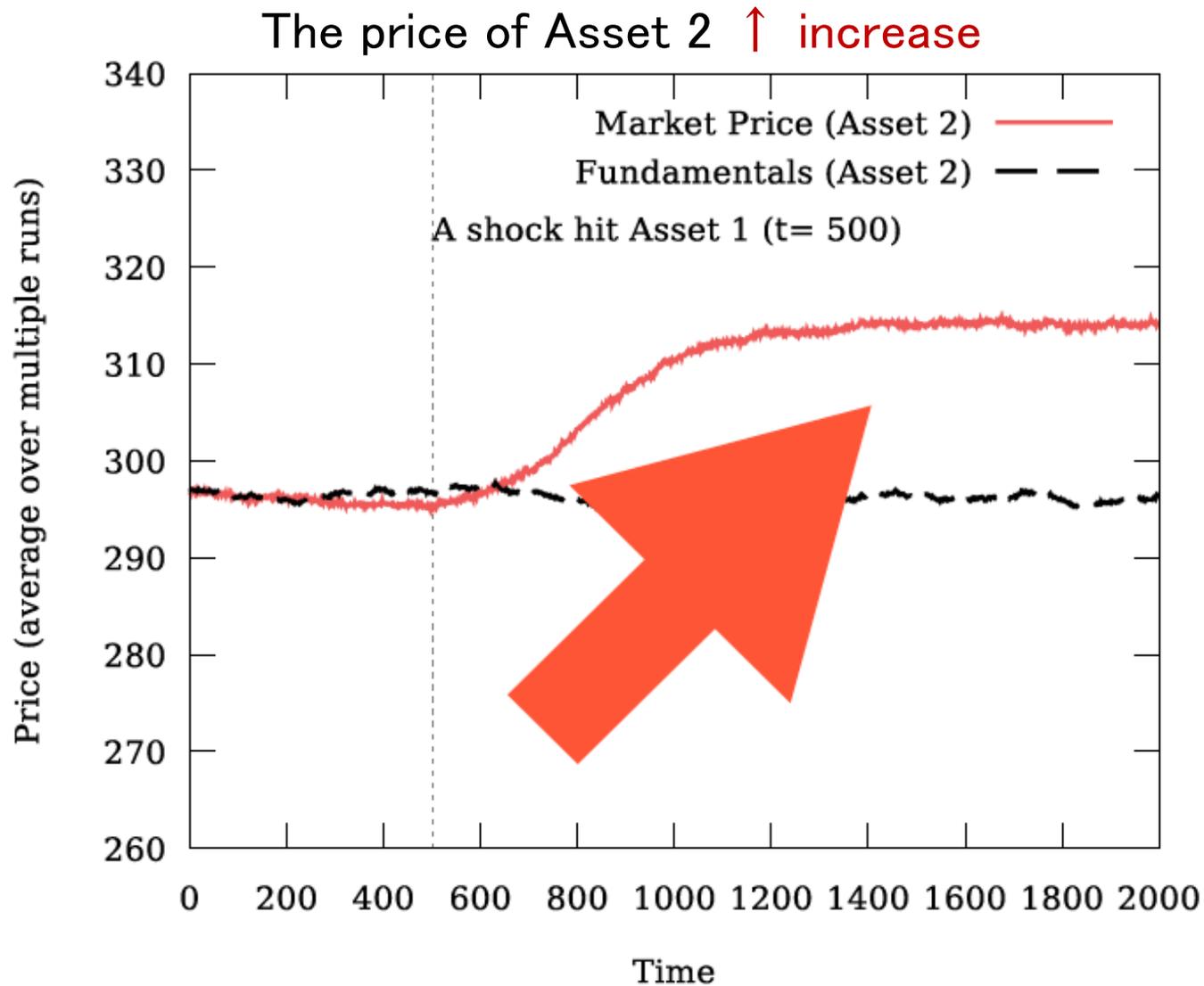
裁定取引(超高速取引)

# Arbitrage promotes spreading of shocks

The price of Asset 2 ↓ decrease



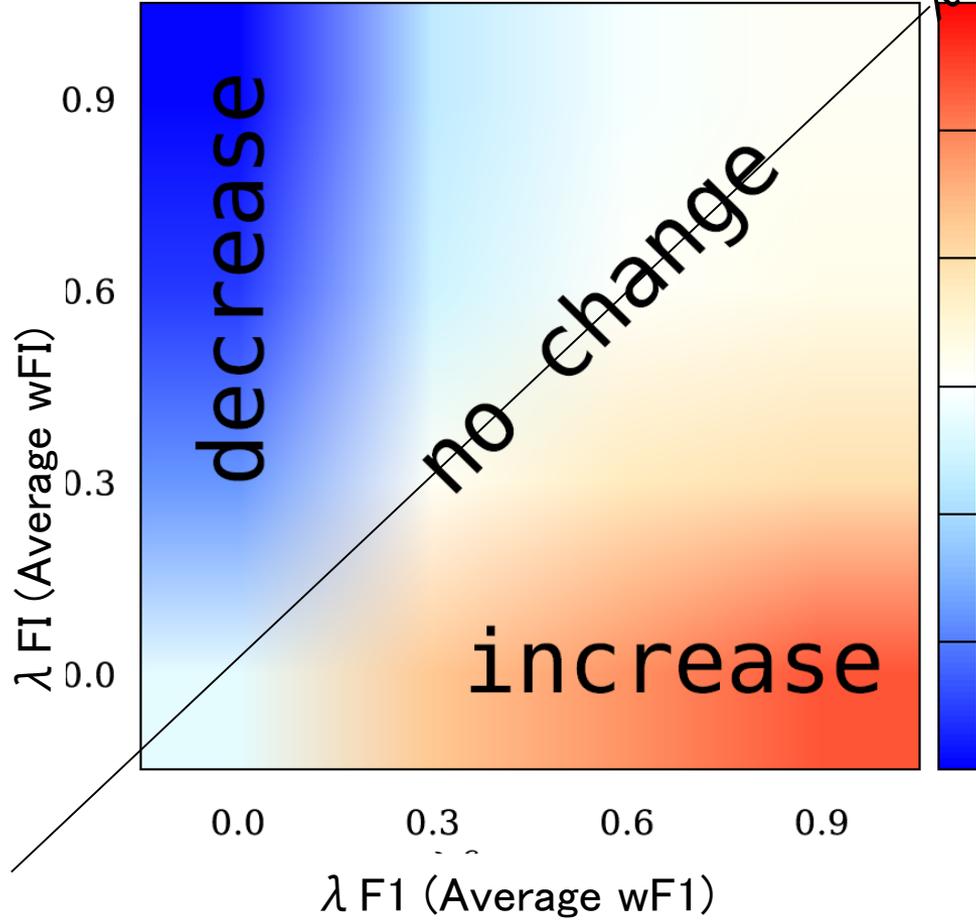
# Arbitrage promotes spreading of shocks



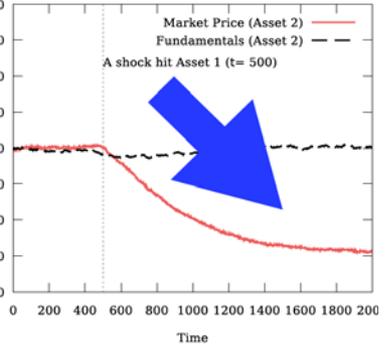
# Effects from trading styles of other assets

The price of **Asset 2**

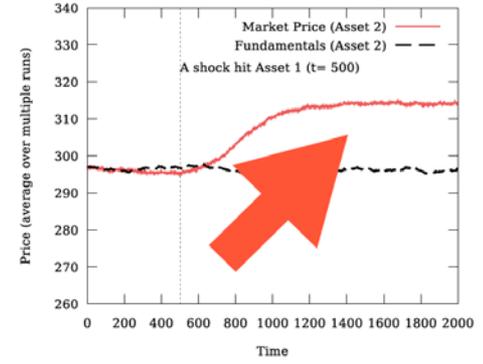
$$\lambda_{F1} = \lambda_{FI}$$



if  $\lambda_{F1} < \lambda_{FI}$   
 → decrease



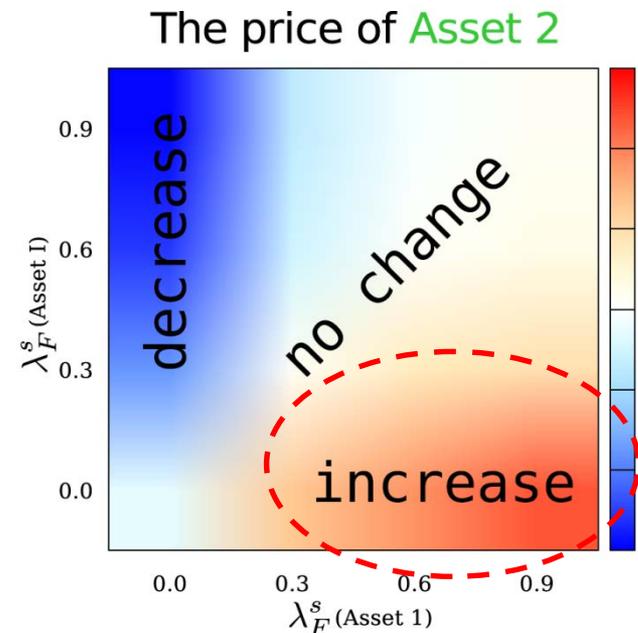
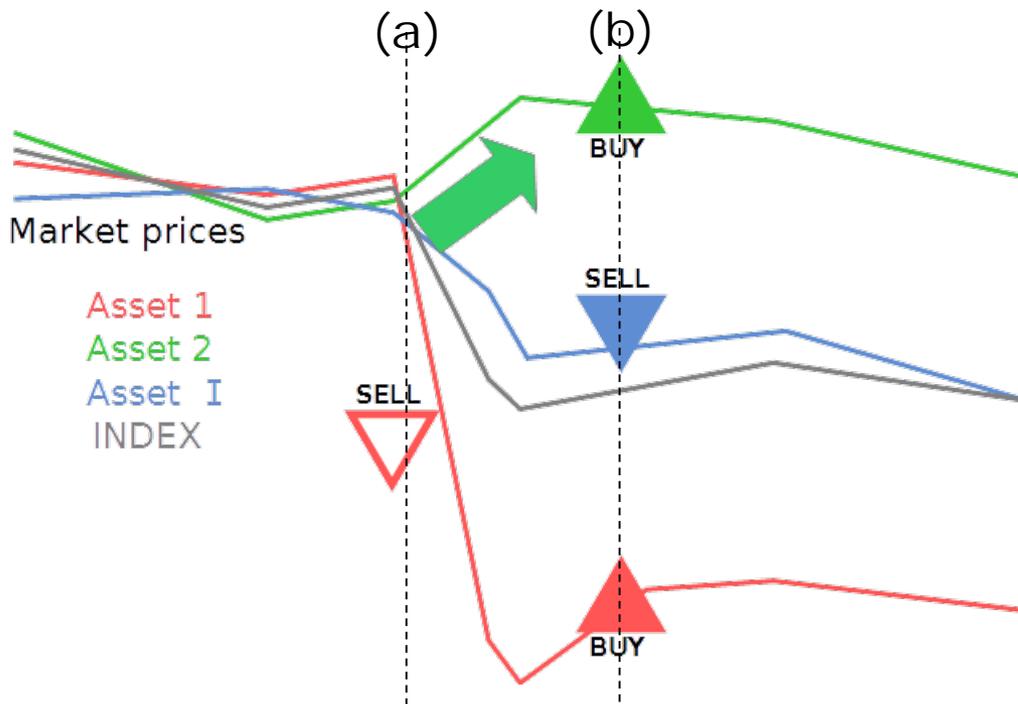
if  $\lambda_{F1} > \lambda_{FI}$   
 → increase



# Mechanism of shock transfer

## Spreading of shocks by arbitrage

- The stronger fundamentalism at Asset 1,  $\lambda_{F1} > \lambda_{F2}$
- (a) When the fundamental price of Asset 1 & I dropped,
  - Local traders at Asset 1 place *sell* orders to follow it
- (b) Asset 1 becomes *comparatively cheaper*,  $p_1 > \text{INDEX}$ ,
  - Arbitrageurs *sell* Asset I and *buy* Asset 1 & 2



# VaR Shock of JGB market in 2003

## Crash of 10-year JGB price (VaR Shock)

- Many banks started VaR based risk management in 2003
- Banks w/VaR based risk management might sell risky assets when the market prices' fluctuations are high.
- They simultaneously started selling Japanese Government Bonds (JGBs).



# 人工市場シミュレーション

## Stock 1

Fundamental price:  $P_t^{1*}$   
Market price:  $P_t^1$

Sell	Price	Buy
84	101	
	100	
	99	
	98	124
	97	77

## Stock 2

Fundamental price:  $P_t^{2*}$   
Market price:  $P_t^2$

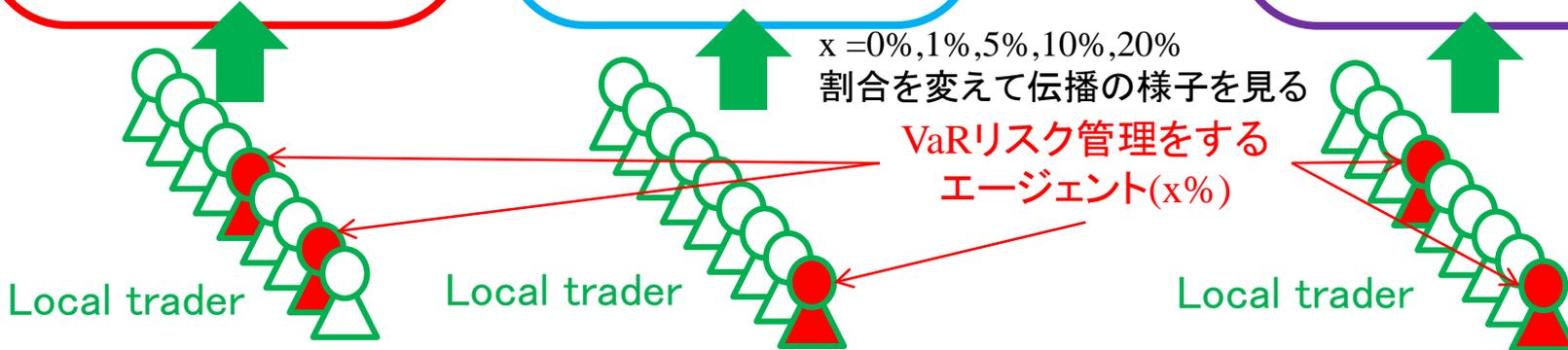
Sell	Price	Buy
84	311	
103	310	
	309	156
	308	124
	307	77

-----

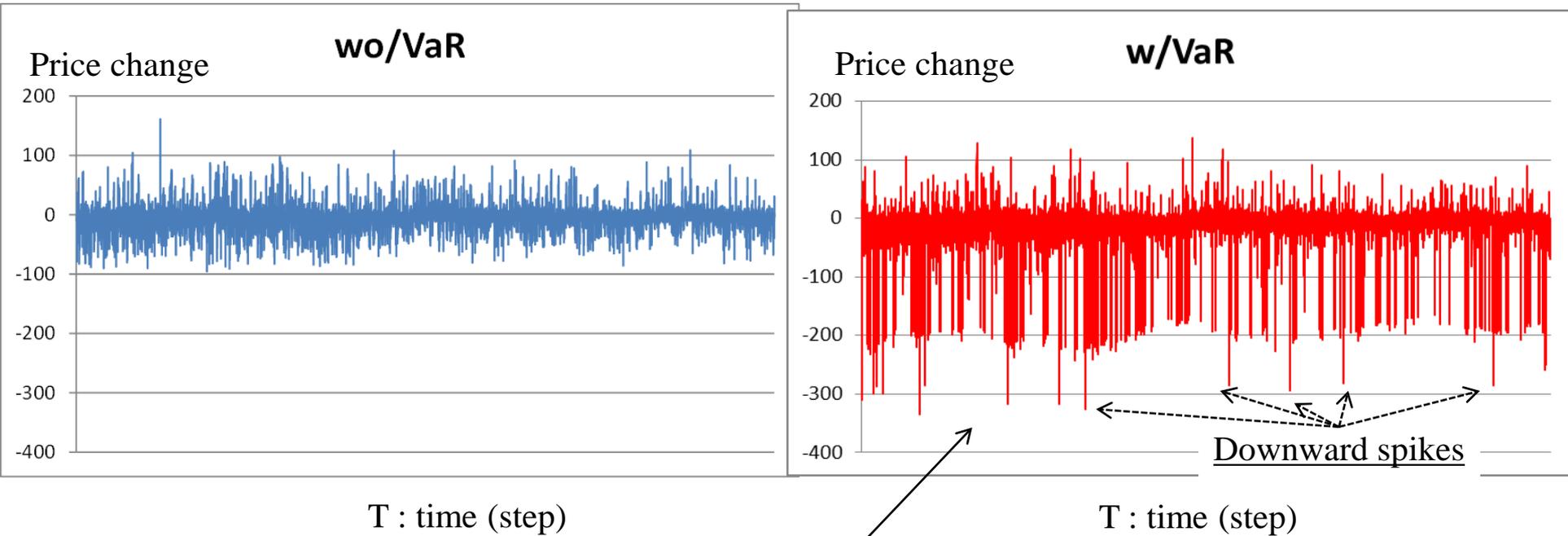
## Stock G

Fundamental price:  $P_t^{G*}$   
Market price:  $P_t^G$

Sell	Price	Buy
32	205	
56	204	
	203	86
	202	74
	201	44



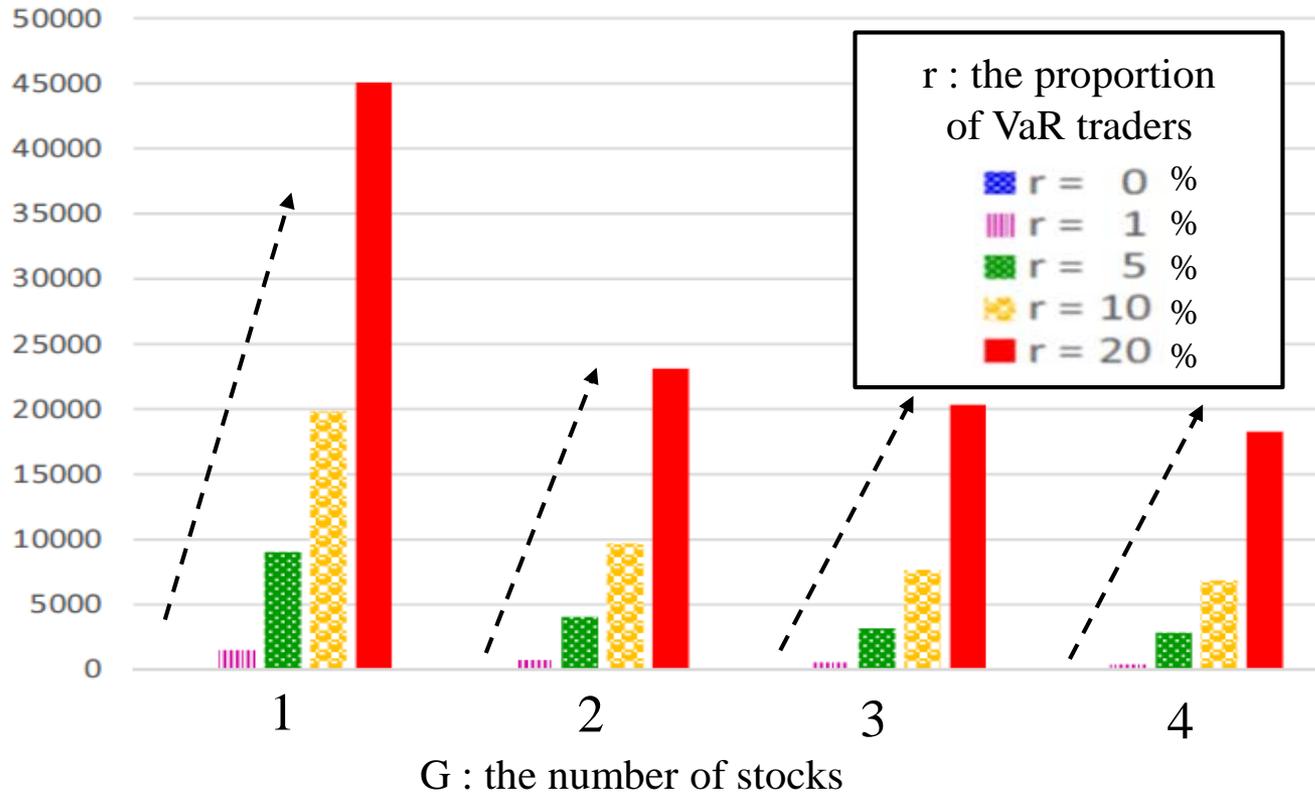
# Experiment 0 : The effect of VaR-based Risk Management in a single-asset market



Single-asset market is destabilized by the introduction of VaR-based Risk management.

# Experiment 1 : The effect of the proportion of VaR traders on a G-asset market (G=1,2, 3, 4)

Total number of spikes among 20 simulation runs



The multiple-asset market is destabilized along with the increase of the proportion of VaR traders.