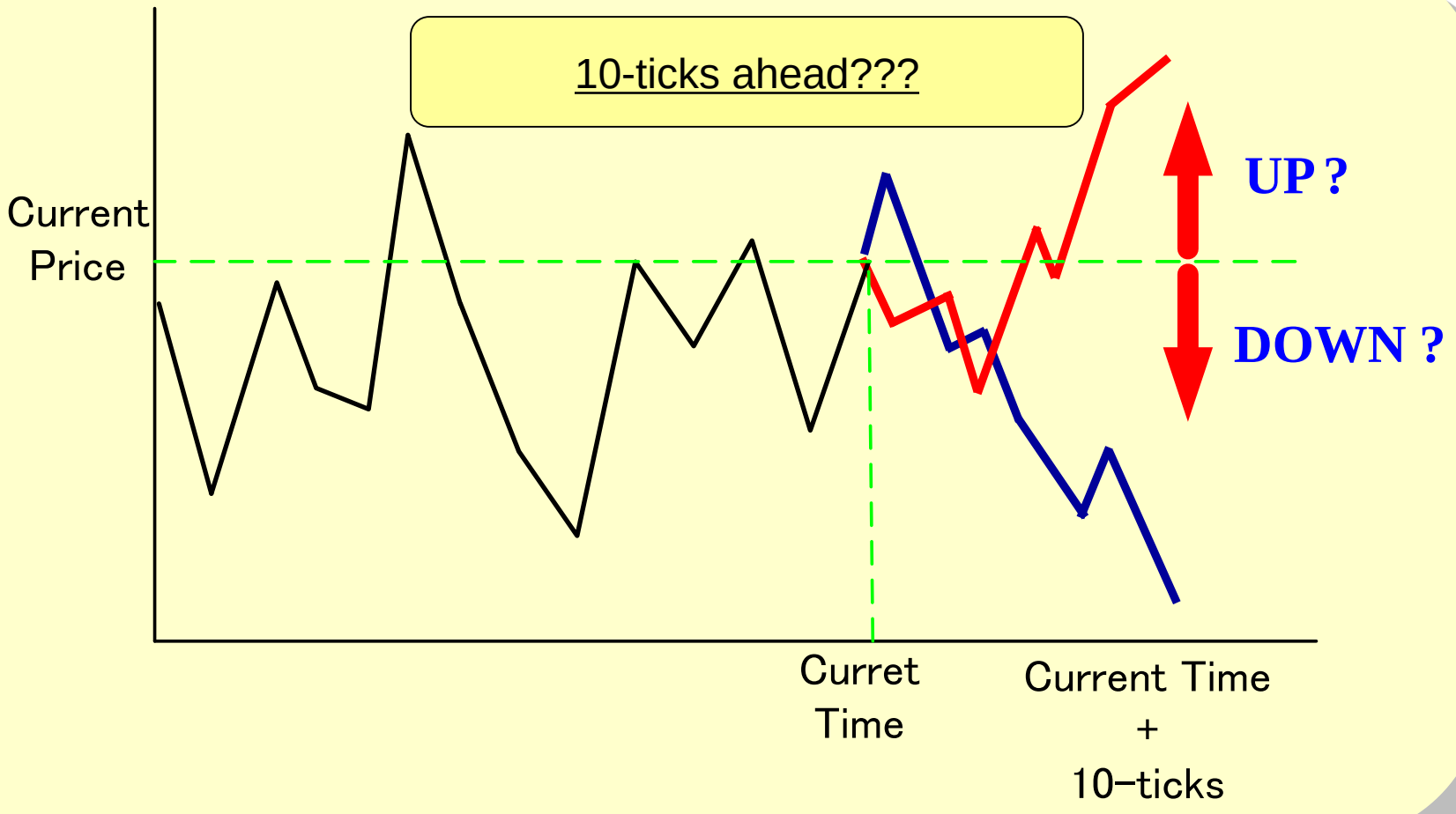
Econophysics Colloquium, Erice, Oct. 26-30, 2009

Application of Evolutional Computation, and Other Techniques to Understand the Intra-day Financial Prices

Mieko TANAKA-YAMAWAKI
and her students in Tottori University
mieko@ike.tottori-u.ac.jp

Tick Price Changes may be Predictable?



Objectives

- Forecasting Tick-wise Price
- Evolutional Strategy
- Improving the Forecast by Recognizing Two Stochastic Phases
- Labeling Different Phases by Means of Technical Indicators
- Constructing a Prediction Generator

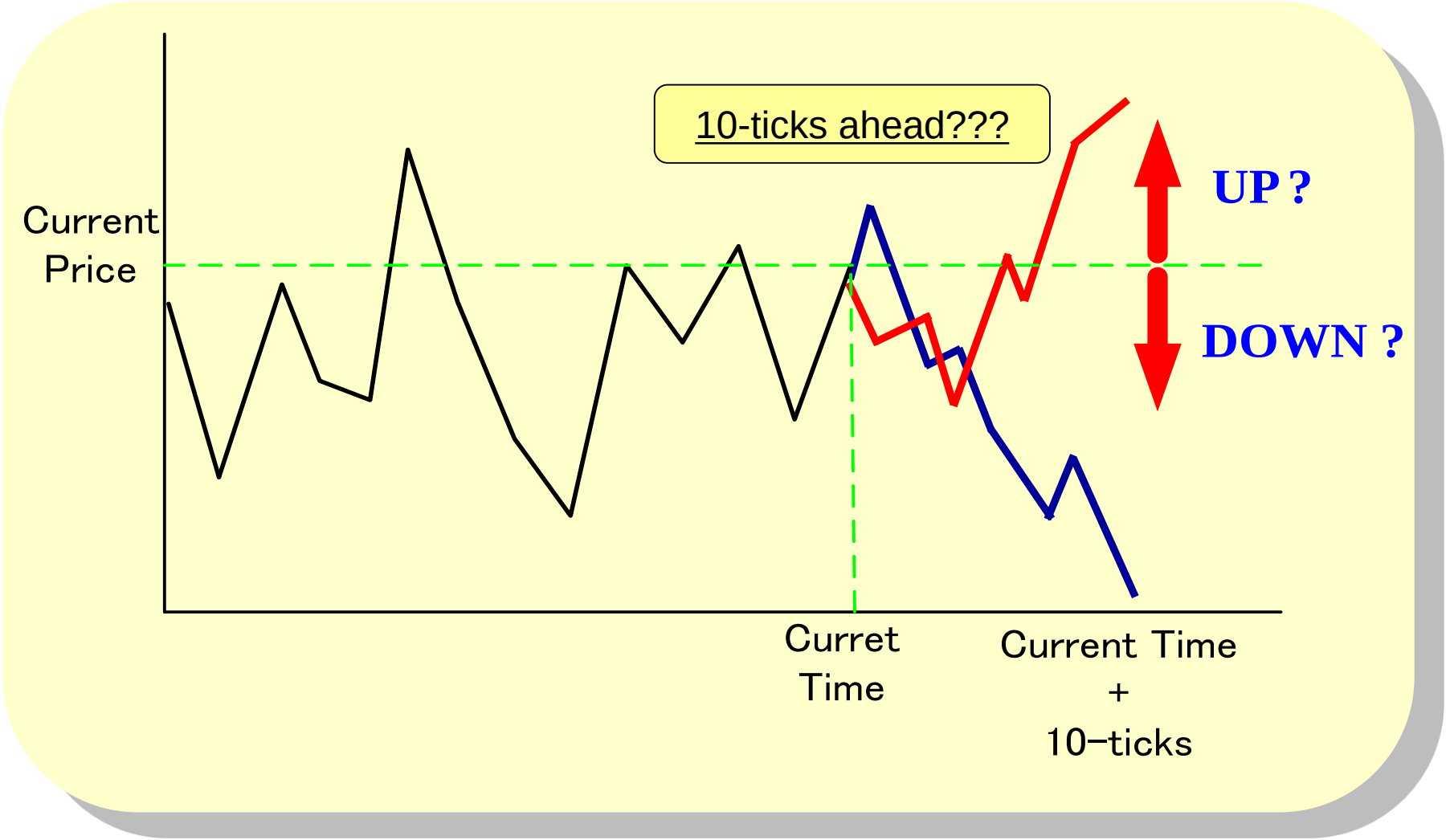
Outline

- the existence of two different stochastic phases in the tick-wise price fluctuations
- improve our old method of developing the evolutional strategy to predict the direction of the tick-wise price movements
- obtain a stable predictive power even in the region where the old method had a difficulty
- Labeling the states by combination of indicators

1. Motivation

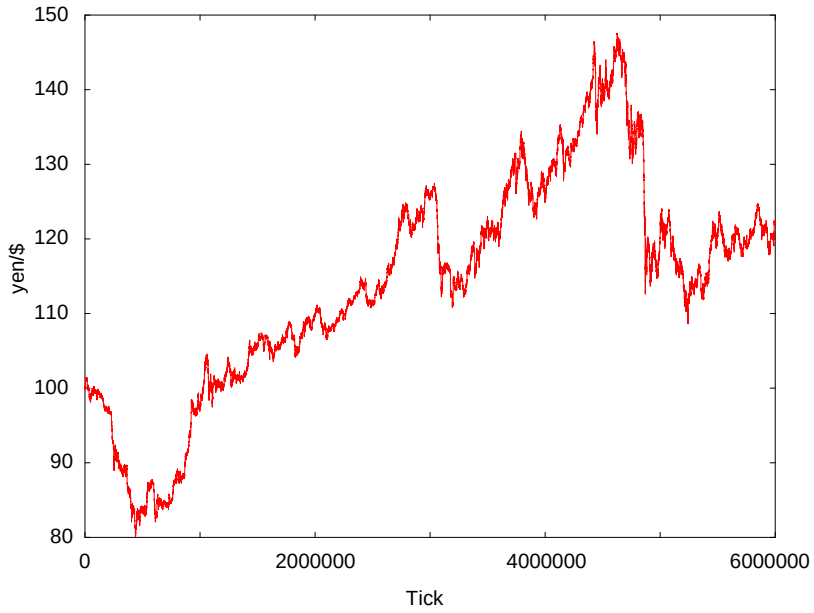
- **Can we predict the price change?**
- **Prices change randomly**
- **Rapid changes in a short time (ticks) are not the pure random walk**
- **Predict direction 2 minutes (～ hour) later**

Is Price Change Predictable?



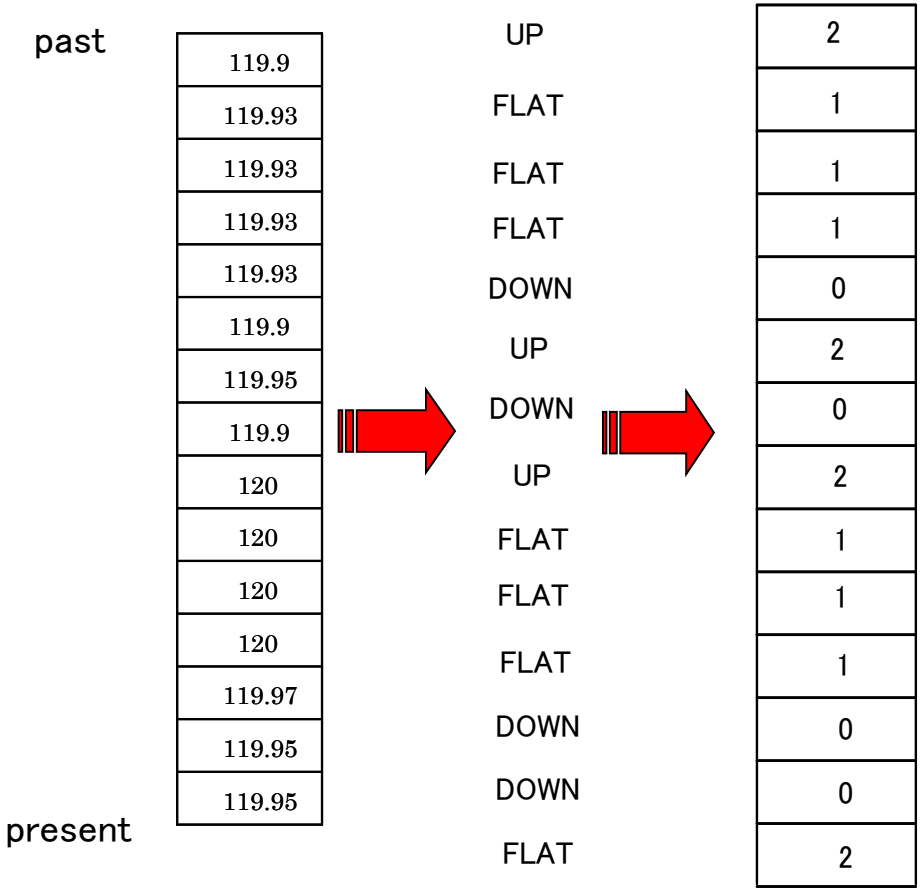
Tick-wise price fluctuation

- $/¥ tick data (1995～2001) 6-million ticks



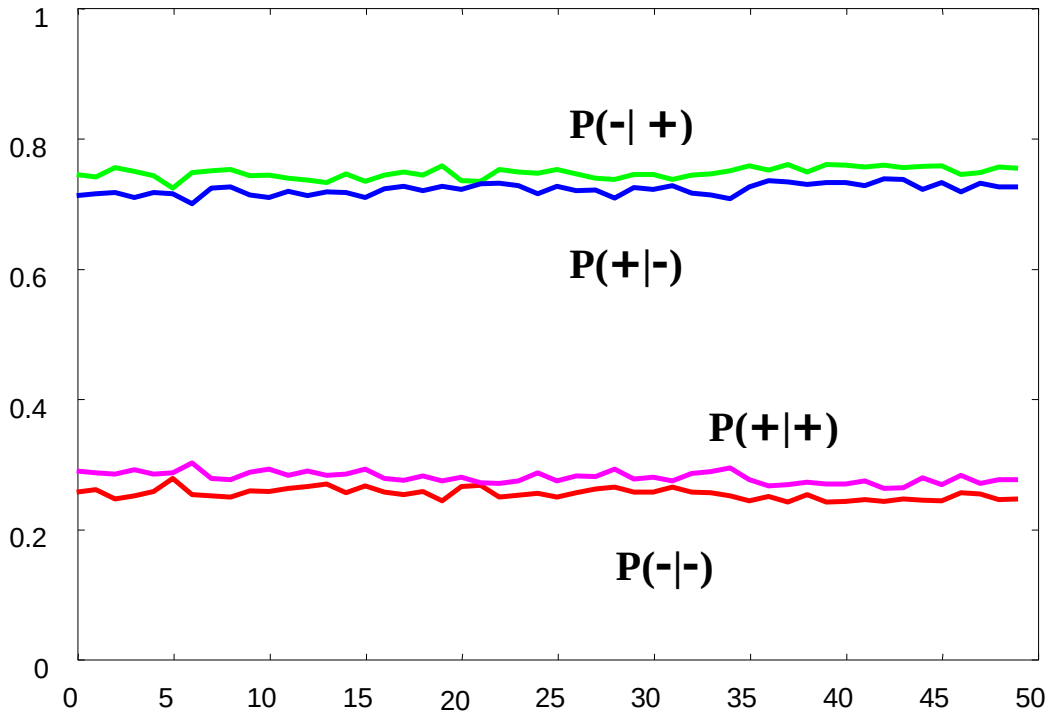
Date	Time	bid	ask
1992/10/1	0:00:32	119.9	120
1992/10/1	0:00:42	119.93	119.98
1992/10/1	0:00:48	119.93	120
1992/10/1	0:00:54	119.93	120
1992/10/1	0:01:06	119.93	120.03
1992/10/1	0:01:12	119.9	120
1992/10/1	0:01:30	119.95	120.05
1992/10/1	0:01:42	119.9	119.95
1992/10/1	0:02:06	120	120.05
1992/10/1	0:02:12	120	120.1
1992/10/1	0:02:18	120	120.1
1992/10/1	0:02:24	120	120.05
1992/10/1	0:02:30	119.97	120.02
1992/10/1	0:02:42	119.95	120.05
1992/10/1	0:03:00	119.95	120

Up, Flat, Down : Coding



Down : 0
Flat : 1
Up : 2

Long-term stability in conditional probabilities in tick-wise price fluctuation (averaged over 2000-ticks)



Repulsive motion at 1-tick ahead

P(−|+)=0.75

$$\overline{P} = \begin{pmatrix} 0.25 & \underline{0.73} \\ \underline{0.75} & 0.27 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.008 & 0.008 \\ 0.008 & 0.008 \end{pmatrix}$$

Time series of conditional probabilities
95.01/02～95.09/05

Learning the prediction strategy

$P(\uparrow|\downarrow)=P(\downarrow|\uparrow)=0.7$

$P(\uparrow|\uparrow)=P(\downarrow|\downarrow)=0.3$

- Forecast of 1 tick ahead based on the motion of immediate past

$P(\downarrow|\uparrow\uparrow)=?$

$P(\downarrow|\uparrow\uparrow\uparrow)=?$

- Forecast of 1 tick ahead based on the motions of several ticks : GA-based automated learning

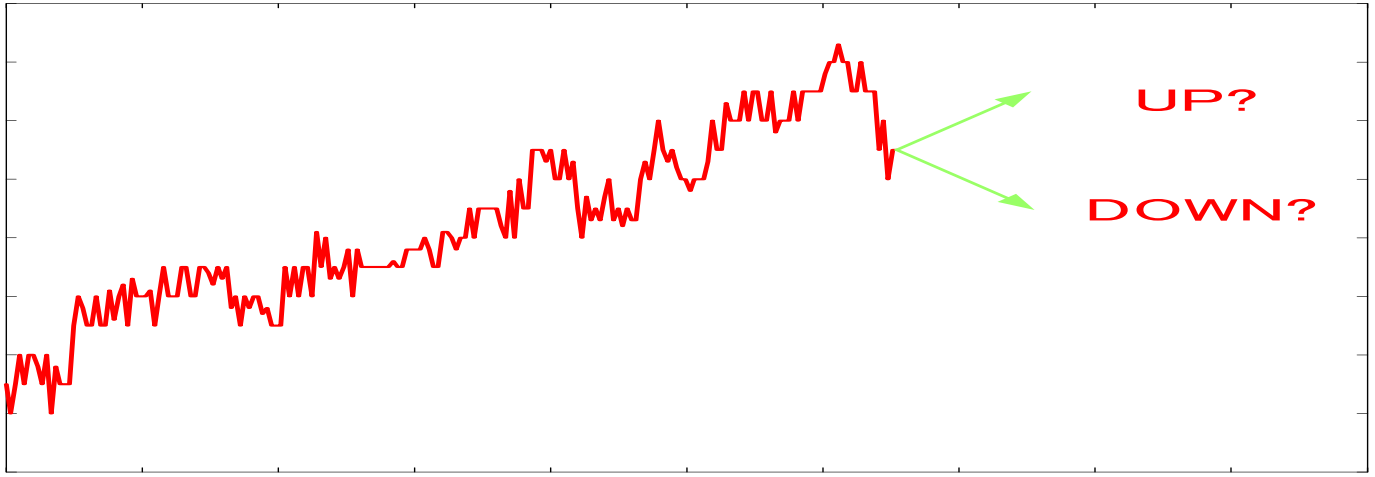
longer condition

bring such knowledge in to a practical use : forecast

- by adaptively learning the pattern : evolutional computation
- a strategy is regarded as a chromosome
- that evolves to adapt to the environment offered by the real-world tick-wise price data

Tick price has patterns→prediction

- Extracting patterns → forecast
- many patterns observed : memory length?
- T-ticks head ?



One of {down, flat, up}, very simple question !

2. Use Evolutional Algorithm

- Learn the strategy from data online
- Need sufficient time (2500 ticks in our case) to fix the strategy
- Change strategy depending on the situation
- Quick labeling of the situation is called for
- Use currently use a combination of technical indicators

Basic Method : Prediction Genome (P)

Design Gene for all the possible history

Ex). N = 3,H=012 case

$P = 012122\cdots002$

$3^3 = 27$

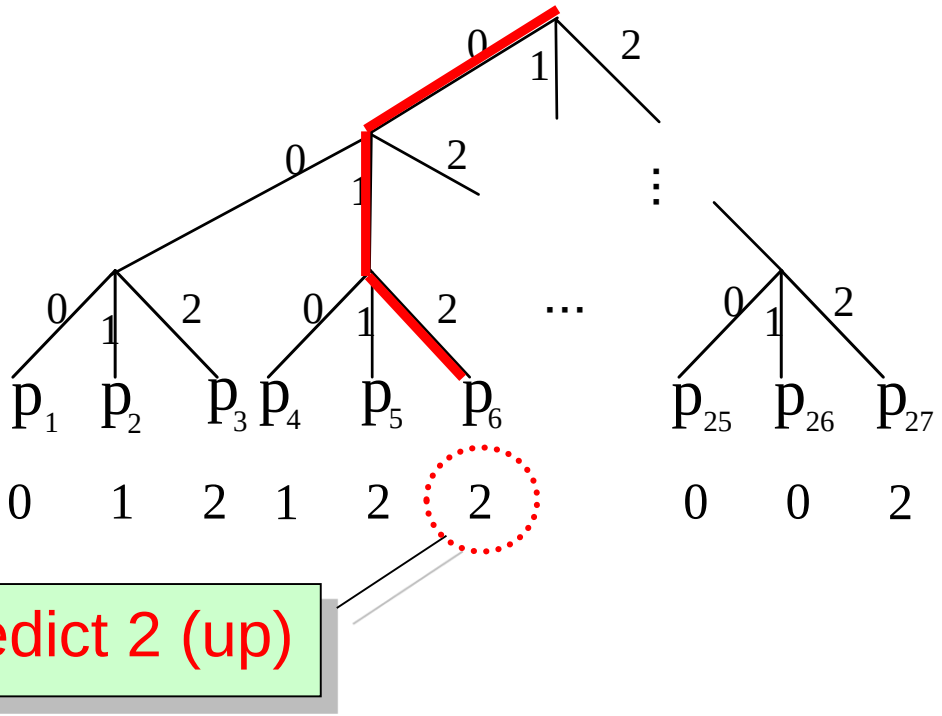
Gene = Strategy

$P = p_1 p_2 \cdots p_{3^N}$

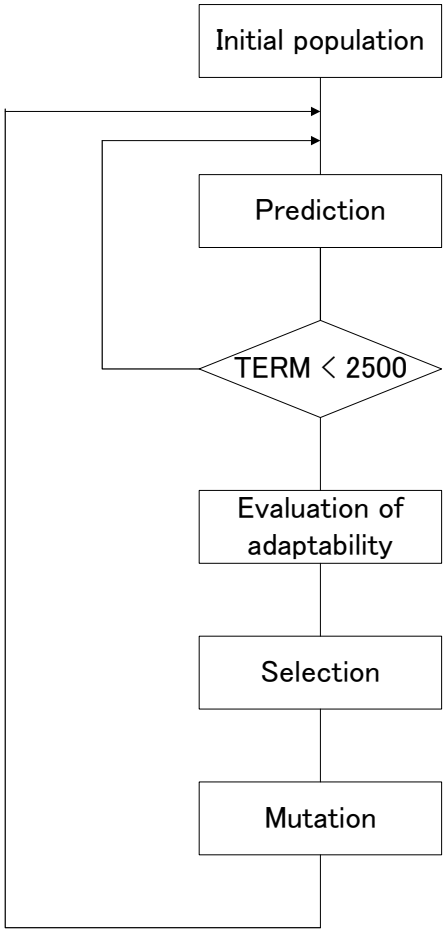
$p_j \in \{0,1,2\}$

Genome length $= 3^N$

Genome types $= 3^{3^N}$

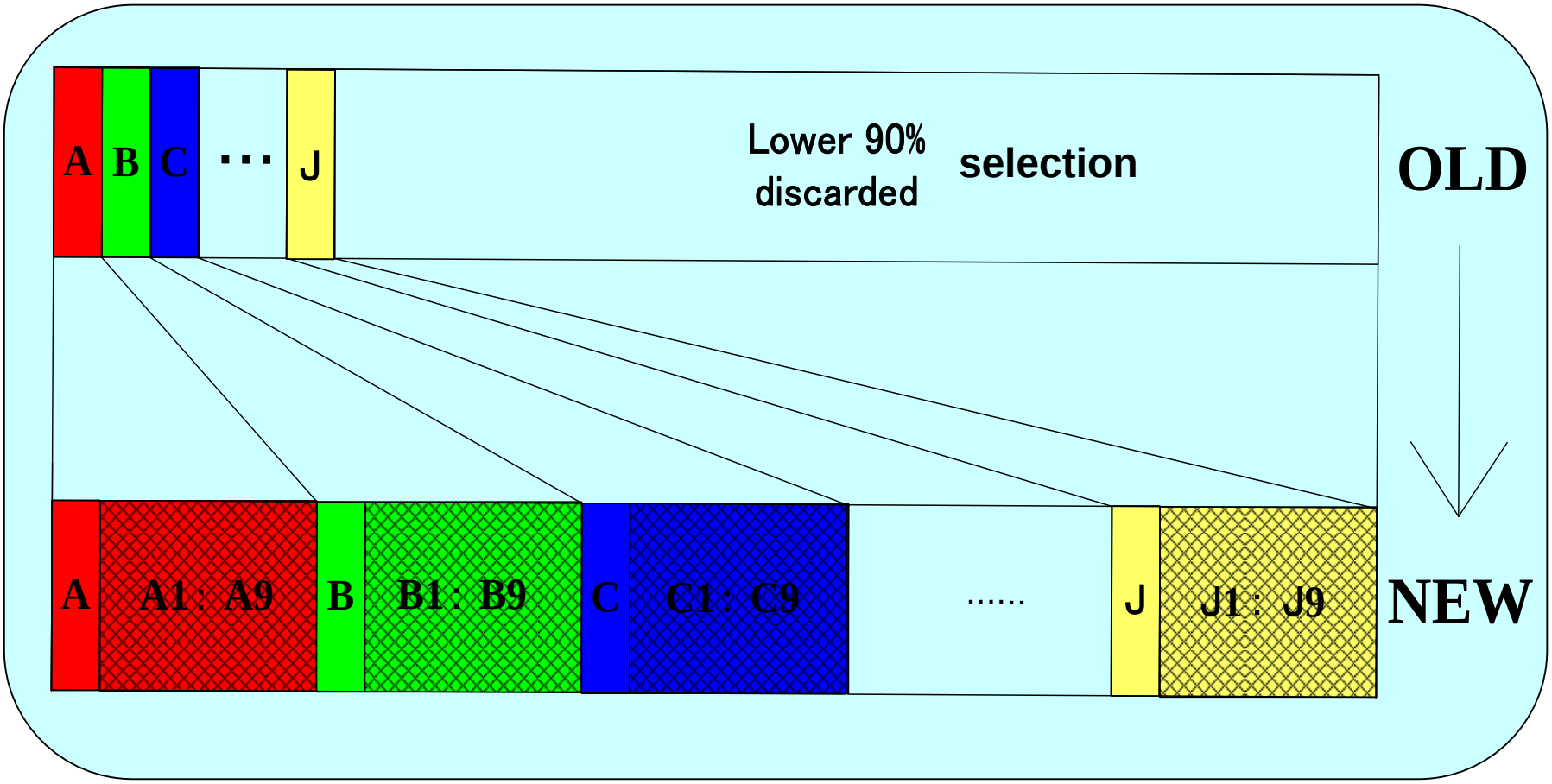


Evolutional Algorithm



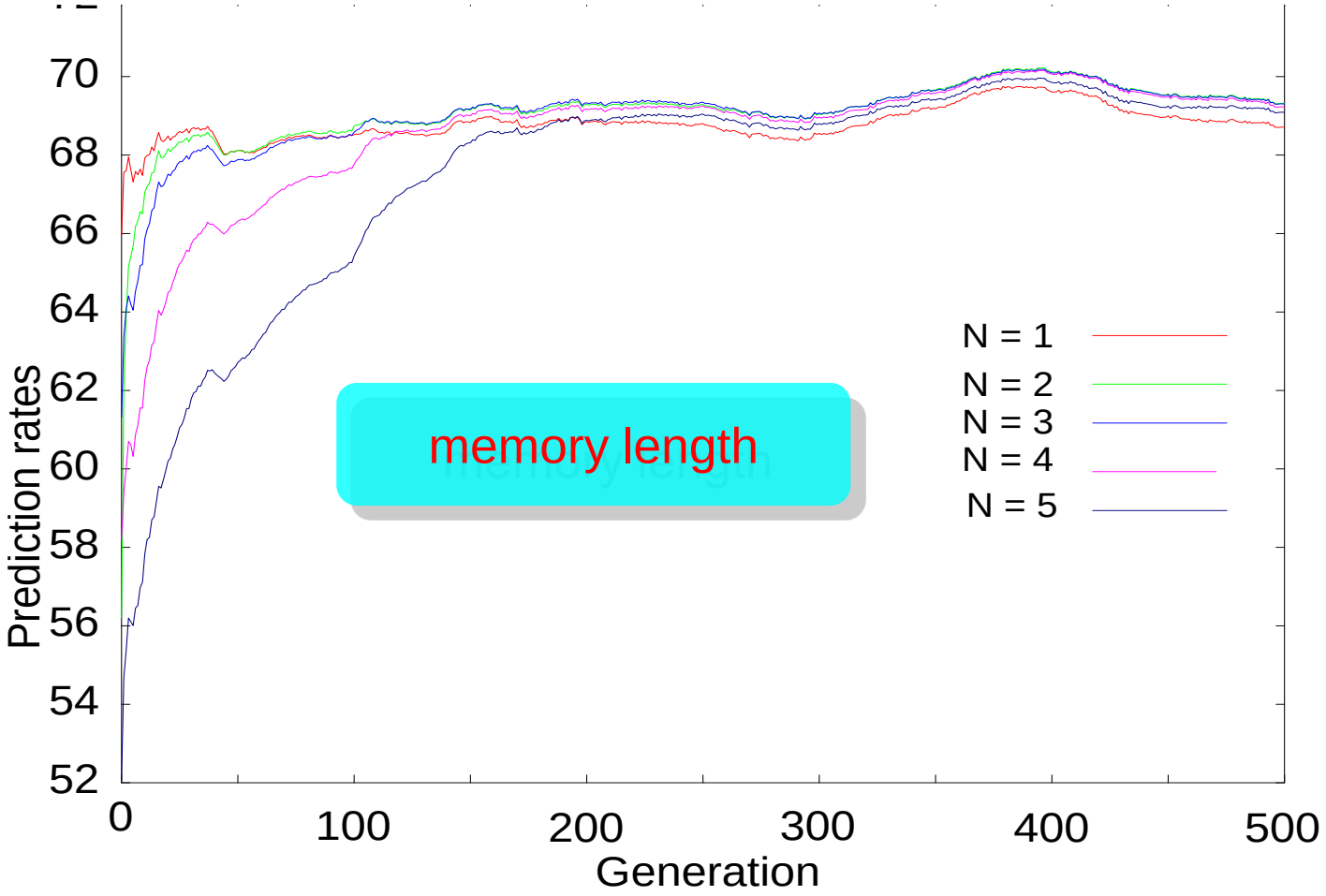
- Prepare 100 genome: initial values are random
- Count the rate of correctly predicted events out of 2500
- Evaluation function is the correctly hitting rate of predictions
- Select the best 10% genome : keep them
- Mutation: Generate mutants by changing one site randomly from each genome to fill the rest 90%

Generation change



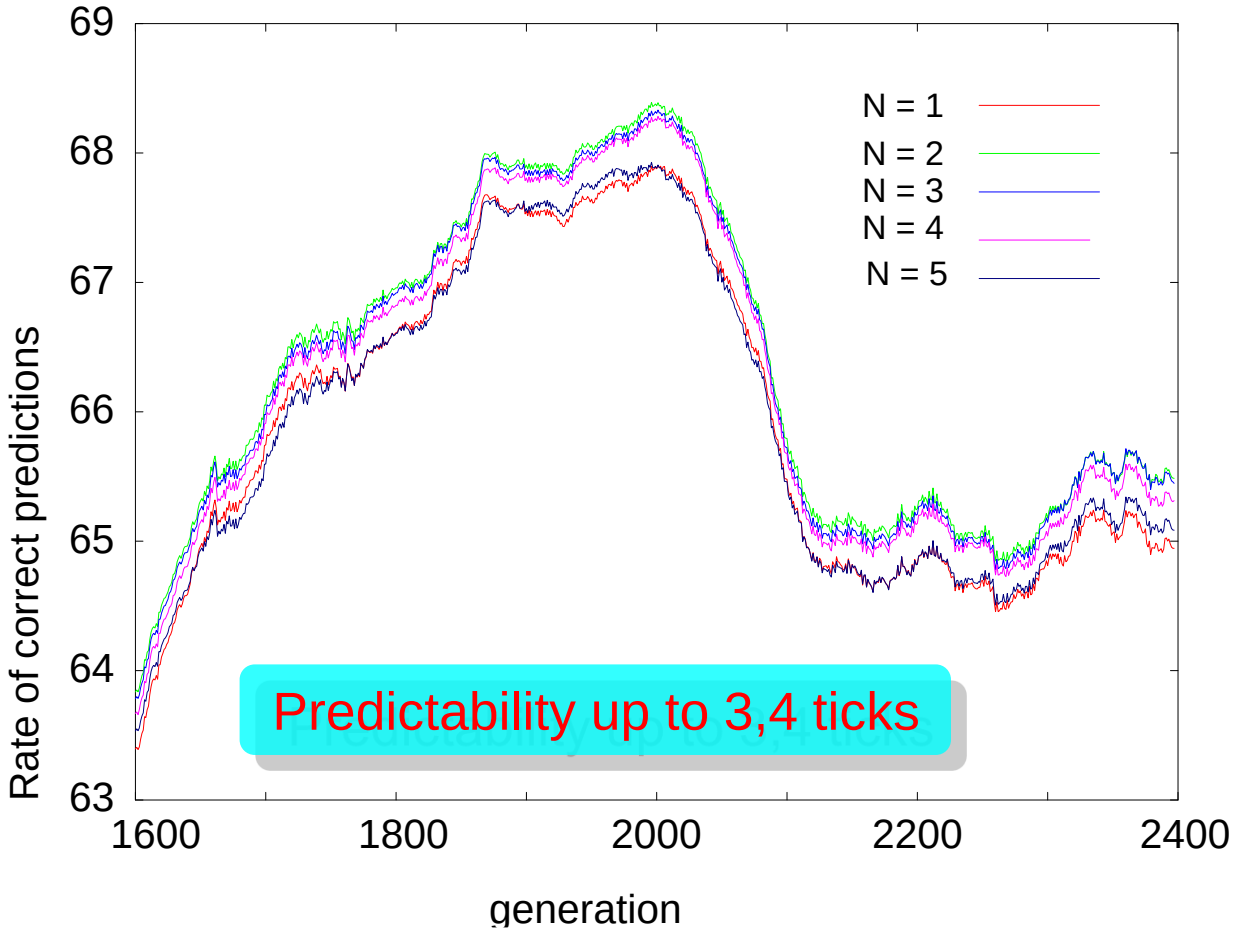
Result: Predictability based on GA leaning

(The first 500 generations are shown)



Predictability after sufficient learning period (1600-2400)

From the top to bottom N=2,3,4,5,1

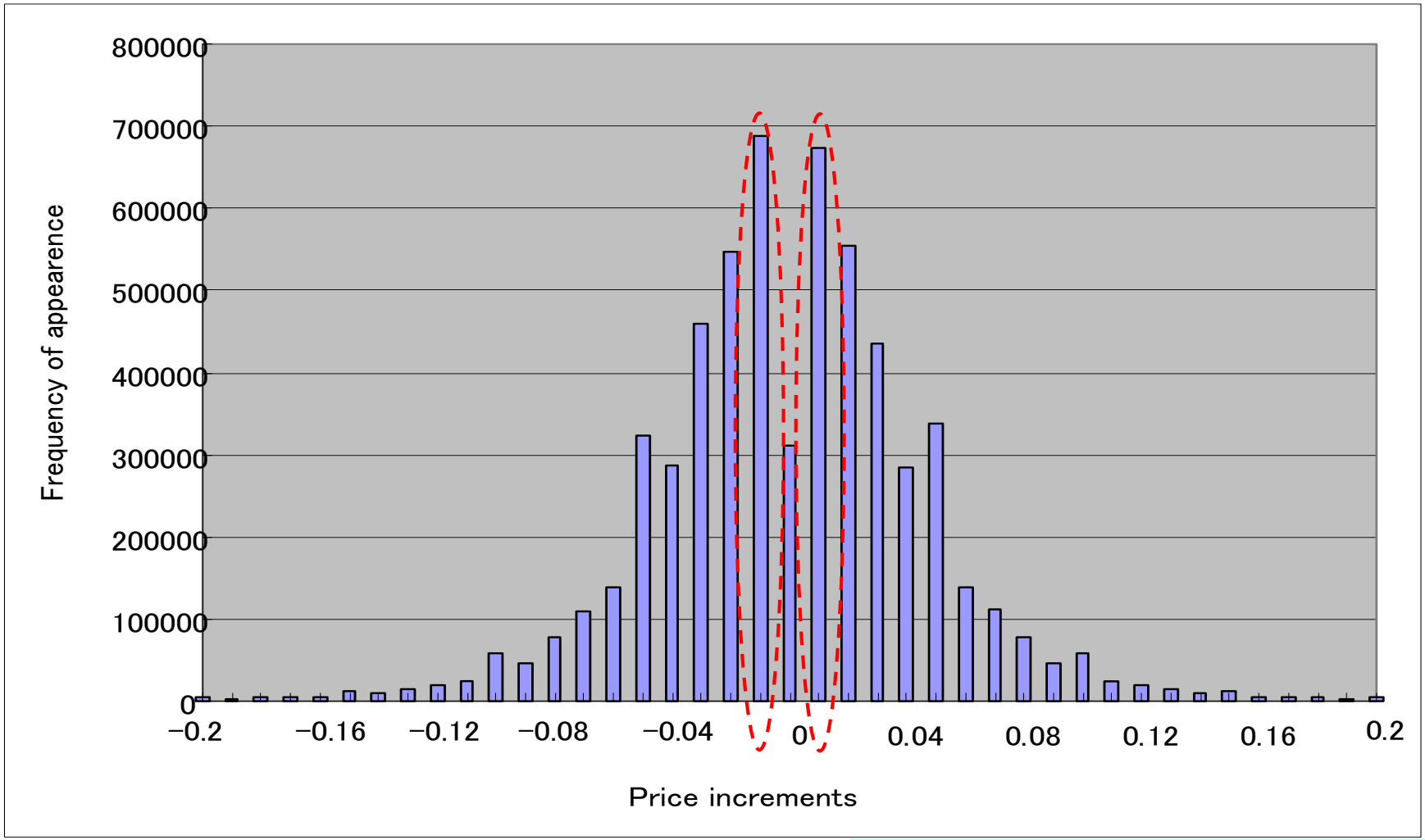


3. Improvement : size effect

- Evolutional algorithm works, but there is another phase seen in the tick-wise price motions
- We go back to the conditional probability analysis and obtain a new knowledge
- Conditional probability P(x|m) plays an important roll. (m = the minimum of Δprice)
- Crossing of two time series

  P(UP|m) and P(DOWN|m)

Histogram of price increments

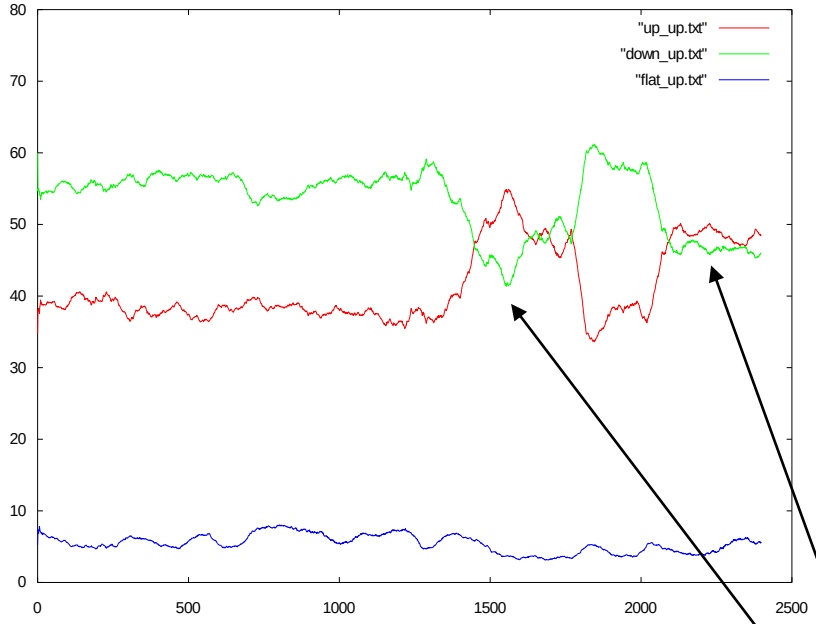


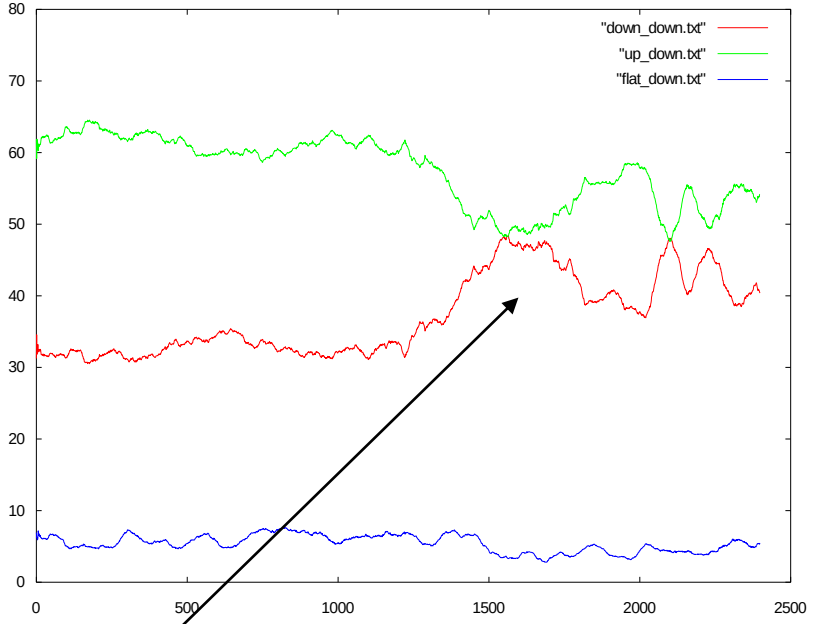
Conditional probability after ±0.01

±0.01 occurs most frequently

After the move of ±m

One point represents condtional probability over 2500ticks



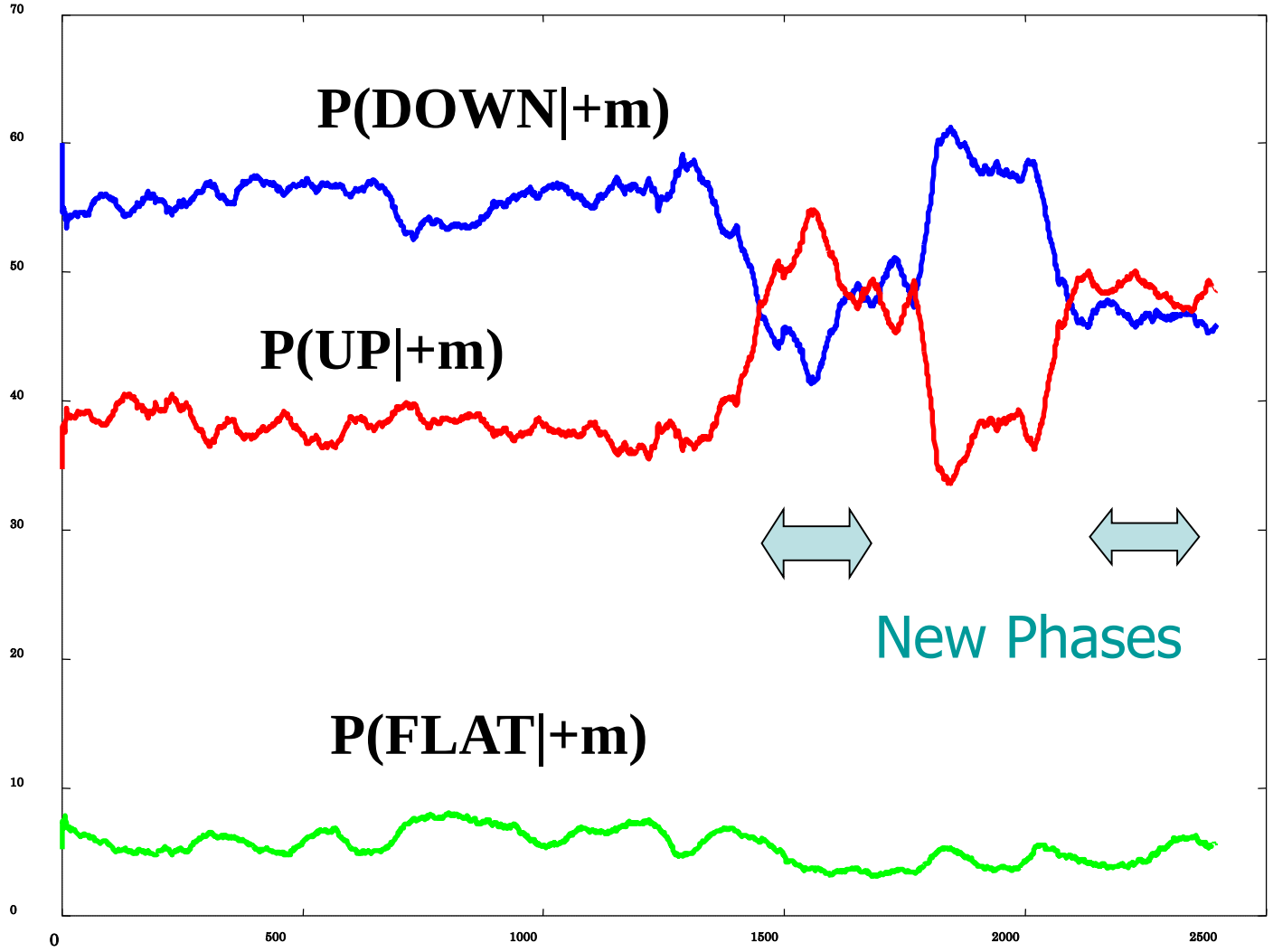


P(x|+m)

P(x|-m)

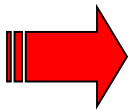
Crossings of two lines

Conditional Probabilities After a Rise by the Minimum Increment, m=0.01.



More states

down: 0
flat : 1
up : 2



Down more than 0.01 : 0
Down by 0.01 : 1
flat : 2
Up by 0.01 : 3
up more than 0.01 : 4

$H = h_1 h_2 h_3 \cdots h_N$

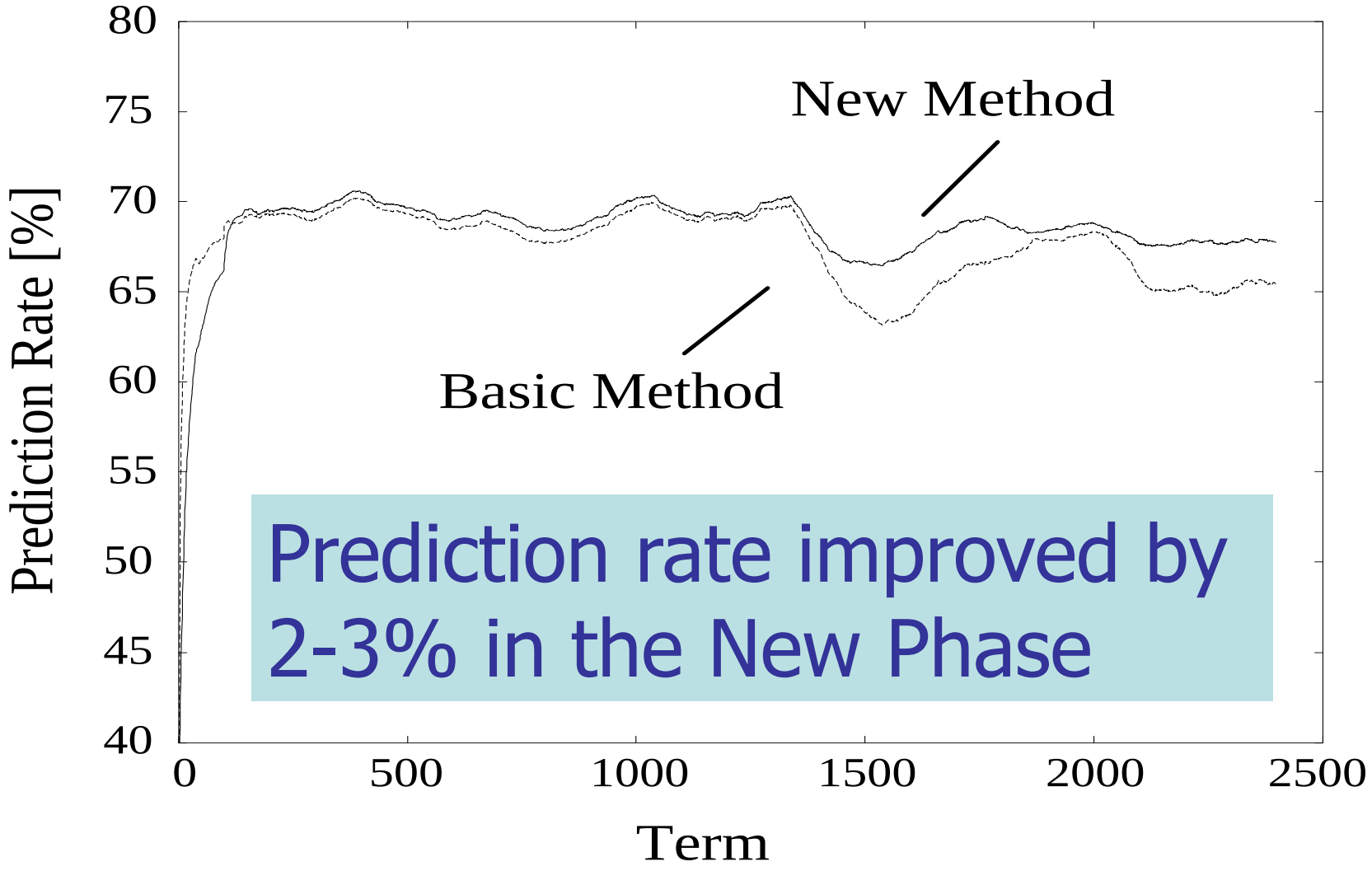
$h_i \in \{0,1,2,3,4\}$

History types $= 5^N$

New Method with Size Effect and Discovery of Two Phases

- Not only the direction of move but also the size of price moves must be taken into account in this new phase.

Comparison：Basic Method (dashed) vs. New Method (solid) for N=3.



4. Labeling the different states

- Labeling by technical indicators
- Combination of technical indicators
- Optimal combination by evolutional algorithm
- Successful in forecasting stock prices and FX prices

Technical Indicators

- **Trend Indicators**
  - **MA (Moving Average)**
  - **MAD (Deviation MA)**
  - **EMA (Exponential MA)**
  - **MO (Momentum)**
  - **MACD**

    (Moving Average Convergence and Divergence)

price direction

- **Oscillator Indicators**
  - **RSI (Relative Strength Index)**
  - **RCI (Rank Correlation Index)**
  - **PL (Psychological Line)**
  - **Etc…..**

overbought and oversold signals

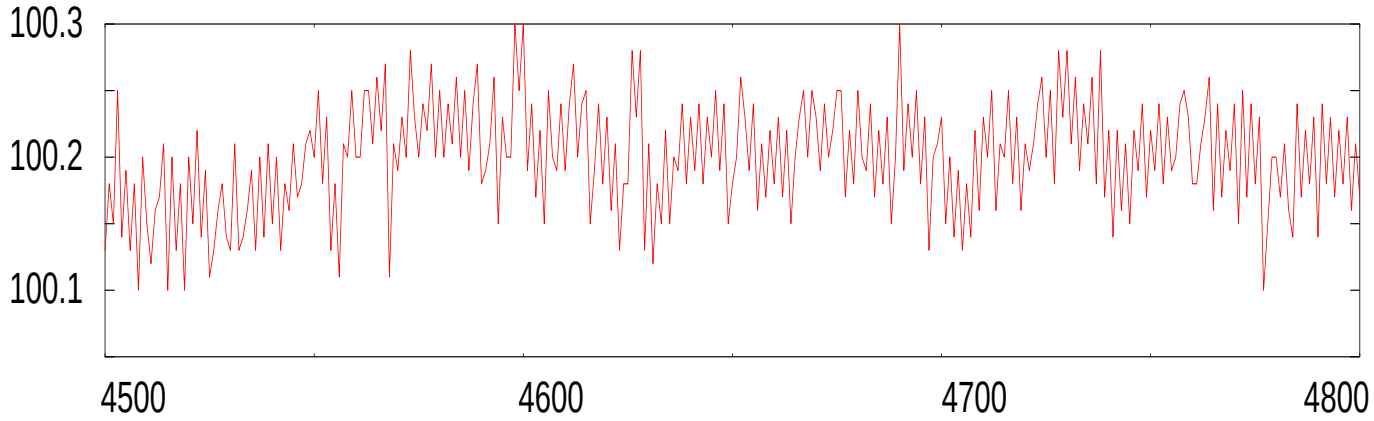
Data: Tick-wise Stock Price

NYSE, 8 symbols, 1993/1/1~1993/12/31

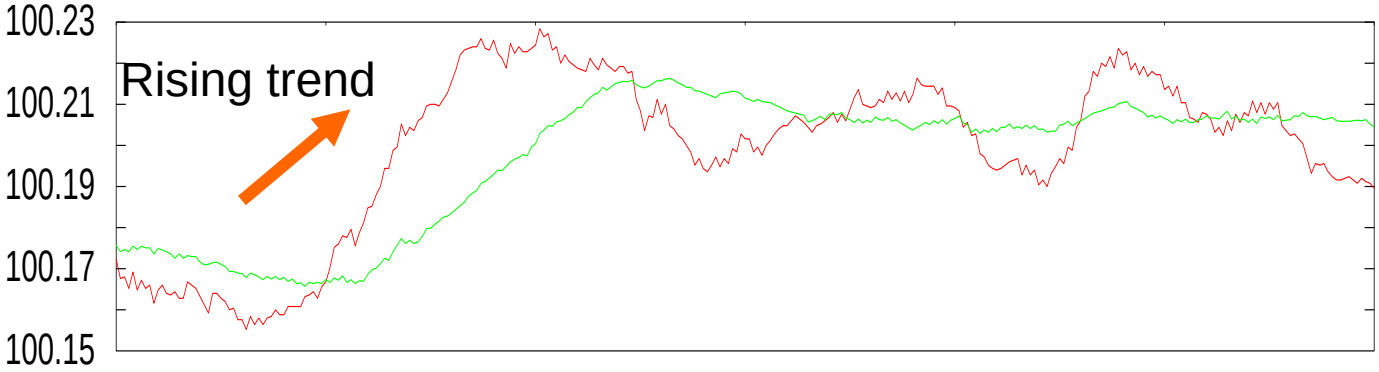
Stock symbol	Business type	Data size (ticks)	Tick interval (s/ticks)	10-ticks (minute)
BBY	retail	54821	109	18
SMRT	retail	12525	473	78
APC	oil	23685	253	42
BP	oil	73562	83	14
CA	computer	65051	92	15
IBM	computer	455233	14	2
F	automobile	194561	32	5
GM	automobile	277241	23	4

Example of trend indicator : MA

Tick price

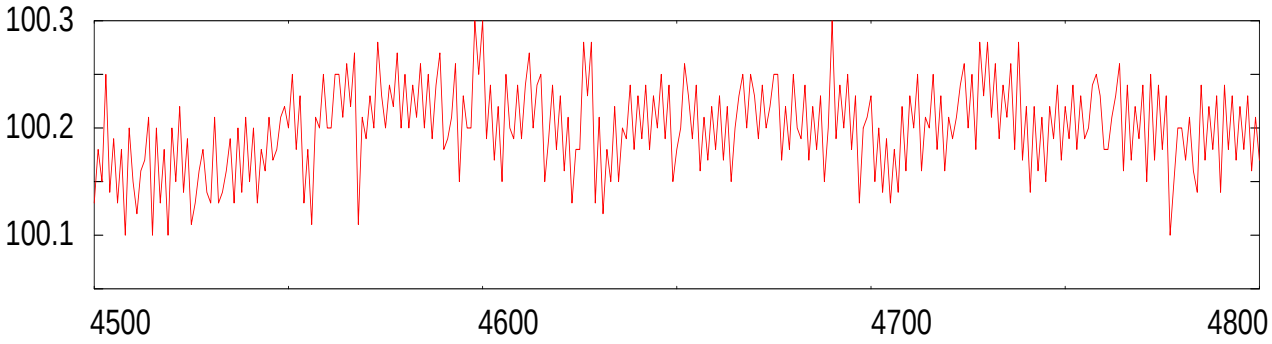


MA
(Trend Indicators)

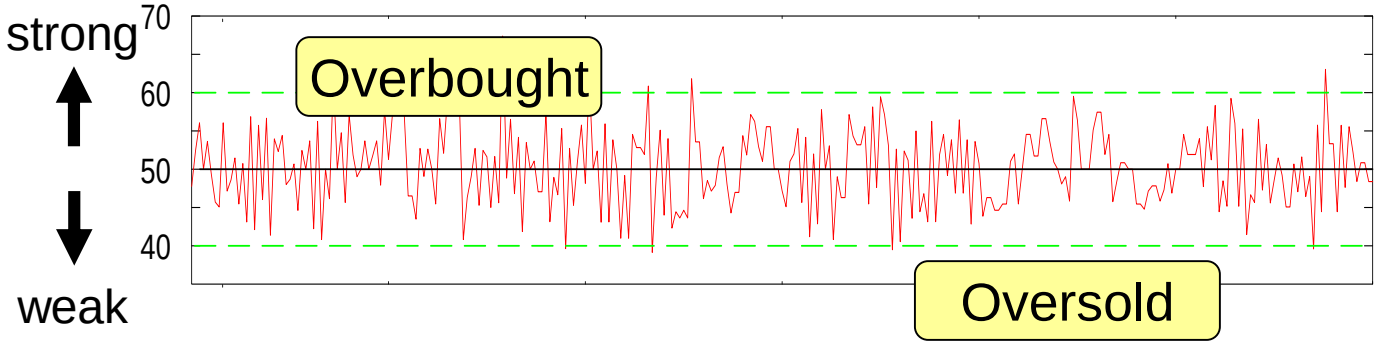


Example of oscillator indicator : RSI

Tick price

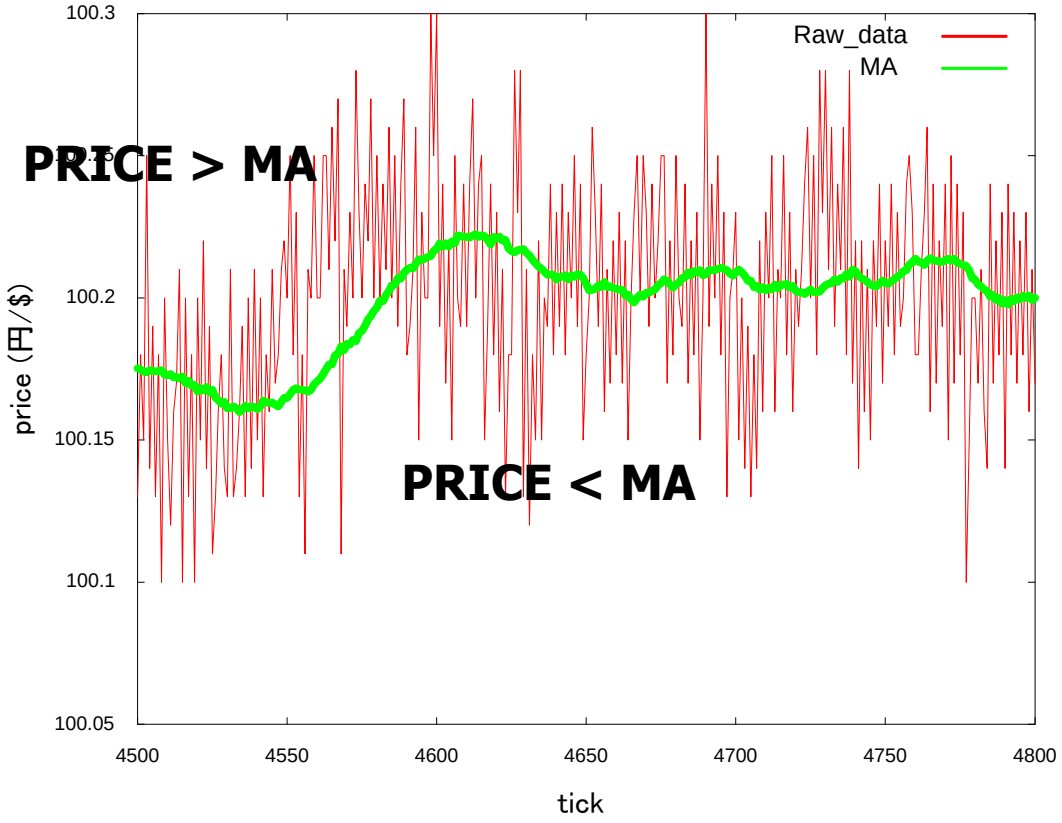


RSI (Oscillator Indicators)



Quantization of indicators : MA

- **MA takes 2 states:**



① PRICE > MA

or

② PRICE < MA

Strategy based on MA only

4 possible strategies

If (① PRICE > MA) UP && if (② PRICE < MA) UP

If (① PRICE > MA) UP && if (② PRICE < MA) DOWN

If (① PRICE > MA) DOWN && if (② PRICE < MA) UP

If (① PRICE > MA) DOWN && if (② PRICE < MA) DOWN

Tick data

Job flow

- Step1

  Evaluate Indicators

- Step 2

  Evaluate and select combination of indicators

- Step 3

  Generate strategies

Learning data

- Step 4

  Prediction

Testdata

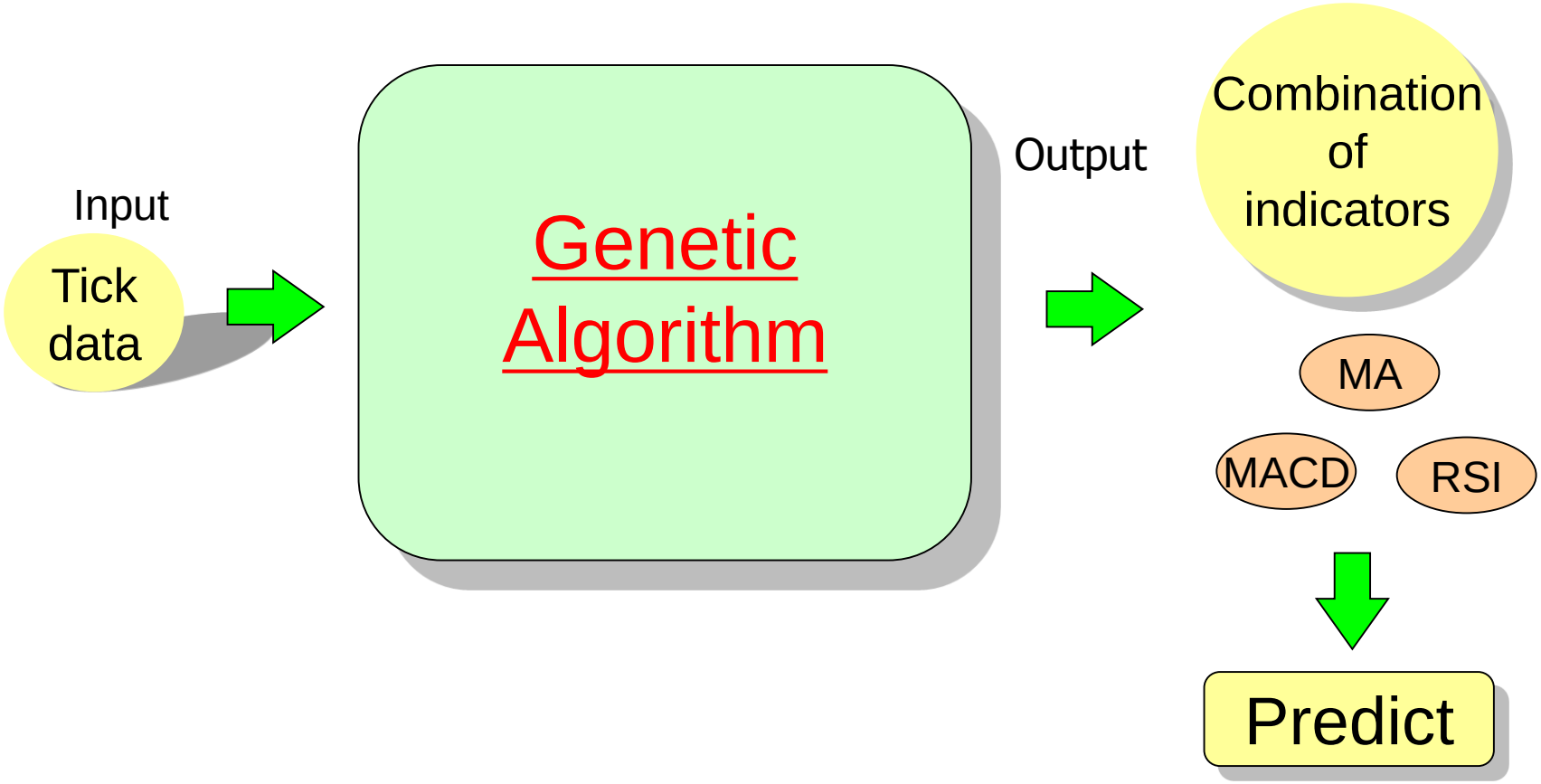
Three assumptions

1. **Different indicator for different stock**

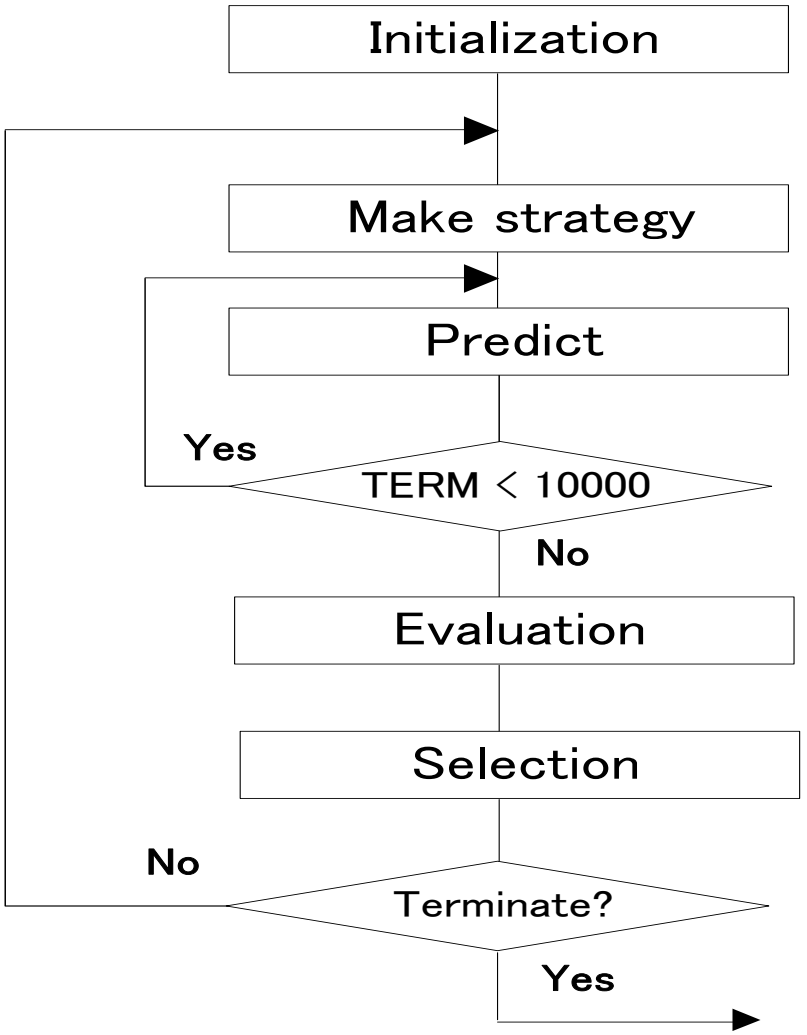
2. **Combination of indicators improves the forecast**

3. **Some indicators are more effective on tick-data**

Selection of indicators by GA



Flow of selecting indicators by GA



Example 1) | MO1 | MA2 | RSI | EMA |

Example 2) | MA1 | MO2 | PL | MACD | RCI | RSI | EMA |

Learning data

Test data

Output

Combination of indicators

Genetic operations & parameters

Genetic operations

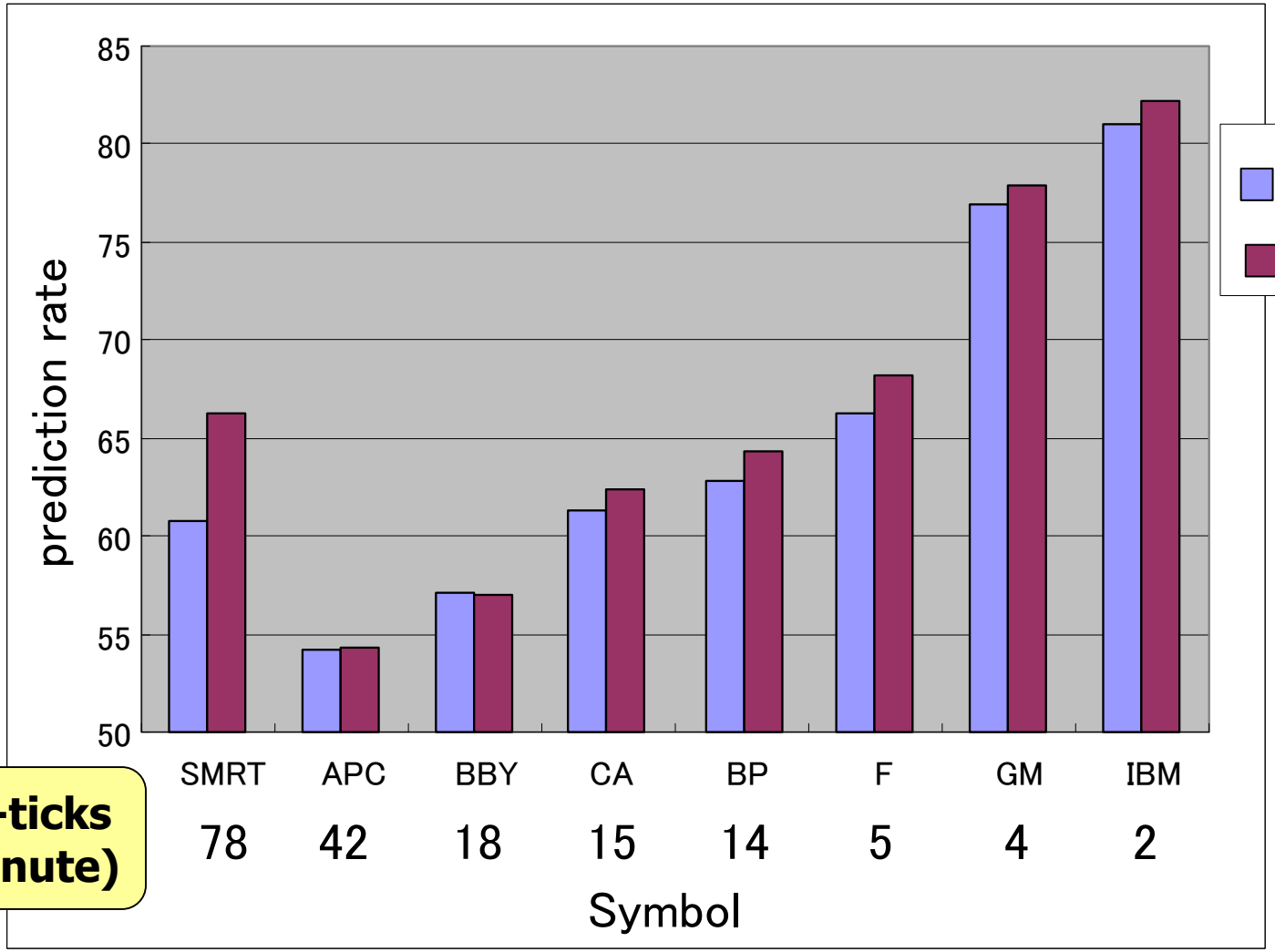
- Selection ： Elite selection(10%) +
  Roulette-wheel selection (90%)
- Crossover ： Two-point crossover
- Parameters
  - Population：100
  - Generation：500
  - Crossover rate：90%
  - Mutation rate：1%

Combination of indicators selected by GA

	APC	BBY	BP	CA
1st	(MO1,MA1,MACD)	(MO2,RSI)	(MO2,MA2)	(MO1,MA2)
2nd	(MO1,MA1,MACD,RCI)	(MO1,RSI)	(MO2,EMA)	(MO1,MA2,MAD)
3rd	(MO1,MA1,RSI)	(RSI,RCI)	(MO2,MA1)	(MO1,MA1)

	F	GM	IBM	SMRT
1st	(MO2,MA2,EMA)	(MO2,MAD,EMA)	(MO2,MAD,EMA)	(MO2,MACD)
2nd	(MO2,MA2)	(MO2,MAD,PHL)	(MO2,MAD)	(MA2,MACD)
3rd	(MO2,MA1,EMA)	(MO2,MAD)	(MO2,EMA)	(MA2,EMA)

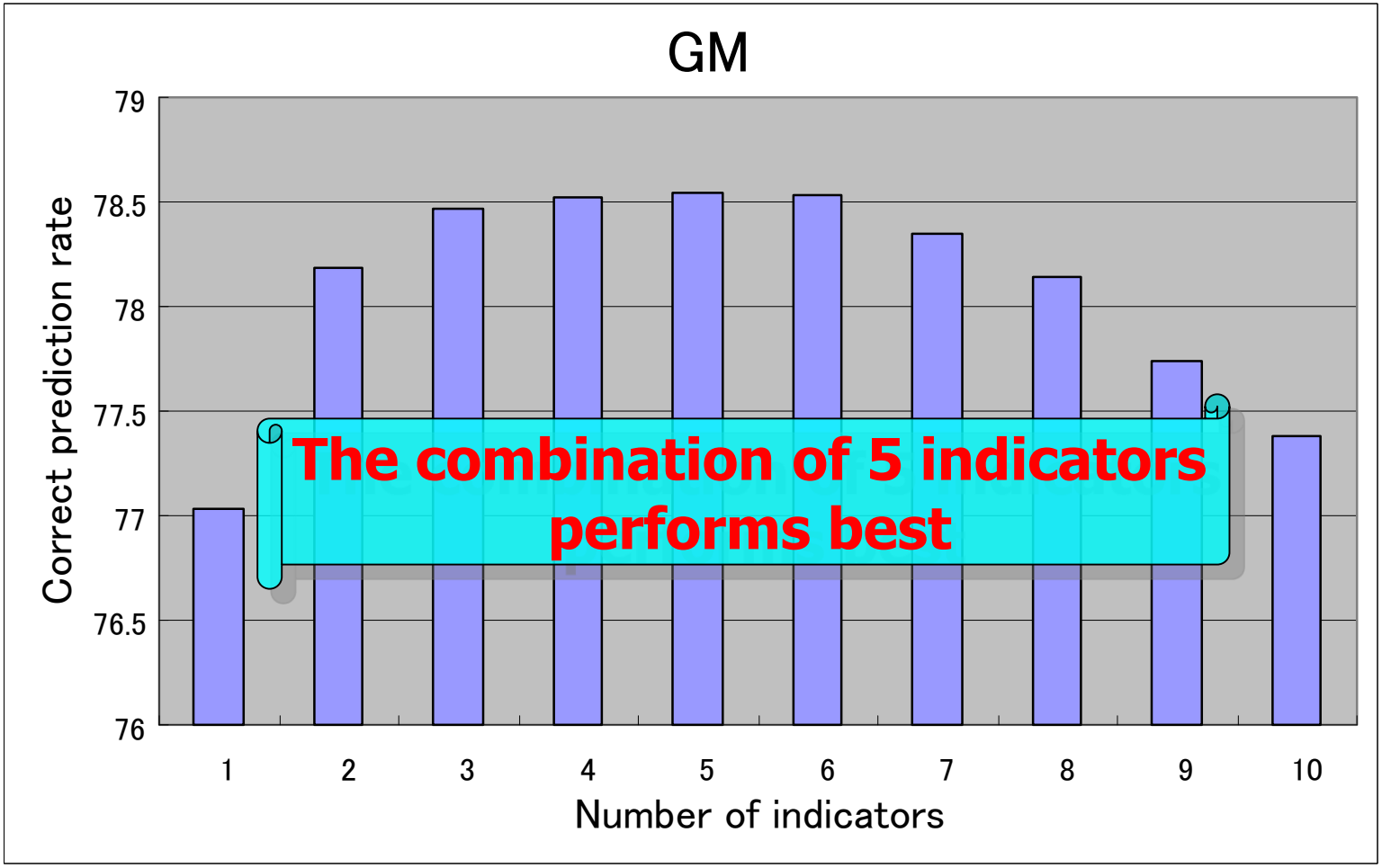
GA vs. Non-GA



10-ticks (minute)

How many indicators should be combined ?

Prediction rates vs. number of indicators



The prediction rate improves when two or more indicators are used

Evaluation

◎：very effective
○：effective
△：effective only to a part of symbol
×：Non-effective

	Indicator	Single	Multiple
Trend type	MA1,MA2	◎	◎
	MAD	◎	◎
	EMA	◎	◎
	MACD	×	○
	MO1,MO2	○	◎
Oscillator type	RSI	△	○
	RCI	△	○
	PHL	○	○

Trend type × Oscillator type is effective

Technical indicators seem to work

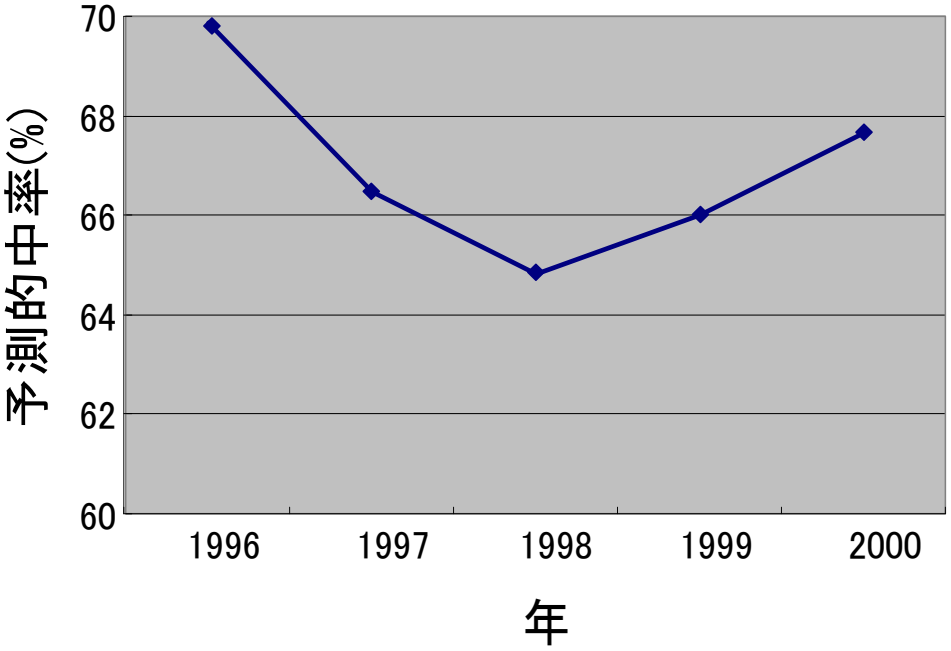
- **Predict price level 10-ticks ahead**
- **Combination of several technical indicators by GA**

  - **Correct prediction in 80% for busy stocks (IBM, GM)**
  - **65% for average of eight stocks**
  - **Best result given by combination of 3-5 indicators**
  - **Poor performance on slow stocks can be explained by the large tick intervals (approx.1-hour)**
  - **Our method should perform well for other popular stocks or foreign exchange rates**

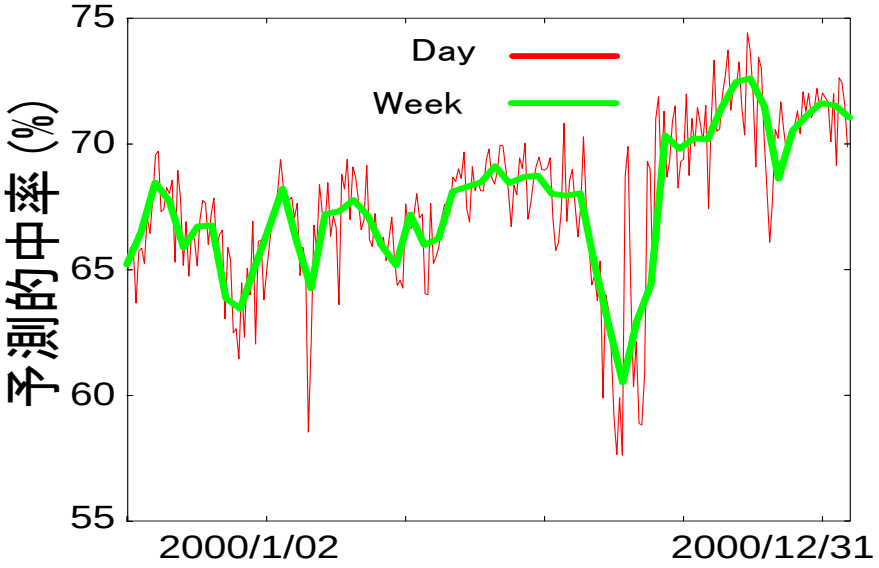
FX data USD/JPY 1996-2000

Result: Forecast of 10 ticks ahead

- Hitting rate for FX (%)



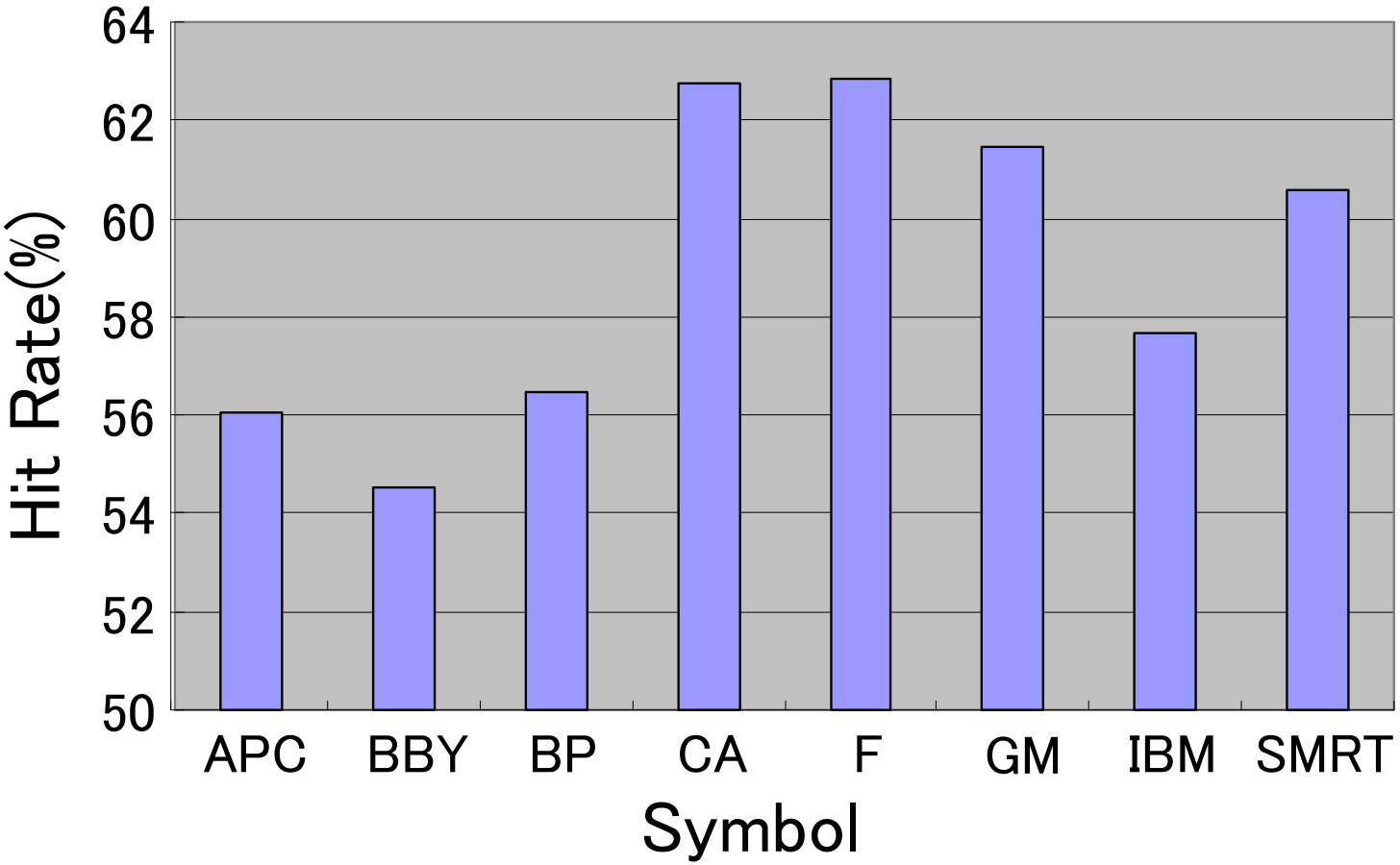
1996-2000



Monthly change (2000)

➡ 70%predictability

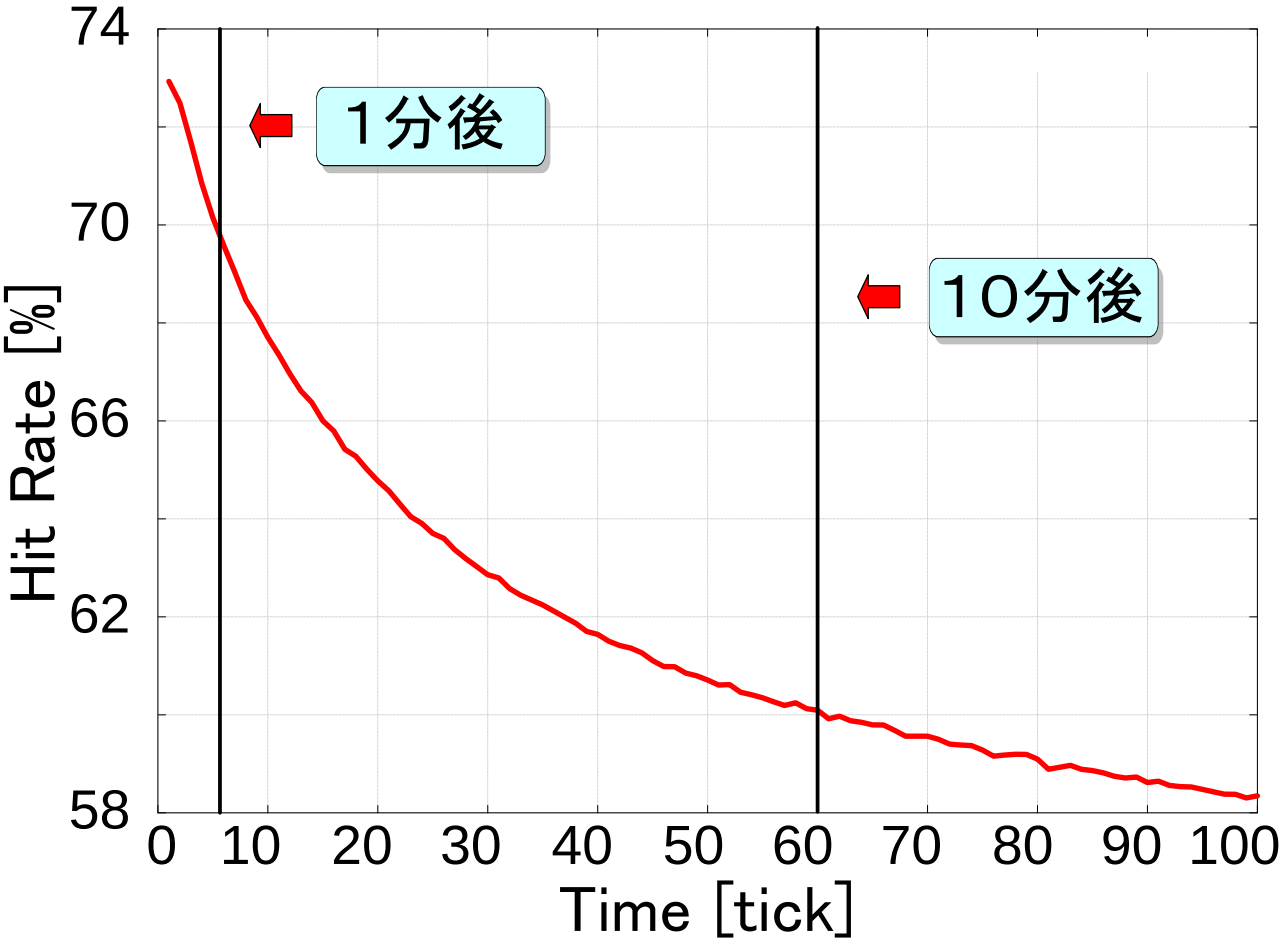
10-ticks先の予測結果(株価)



➡ 60%前後で予測可能

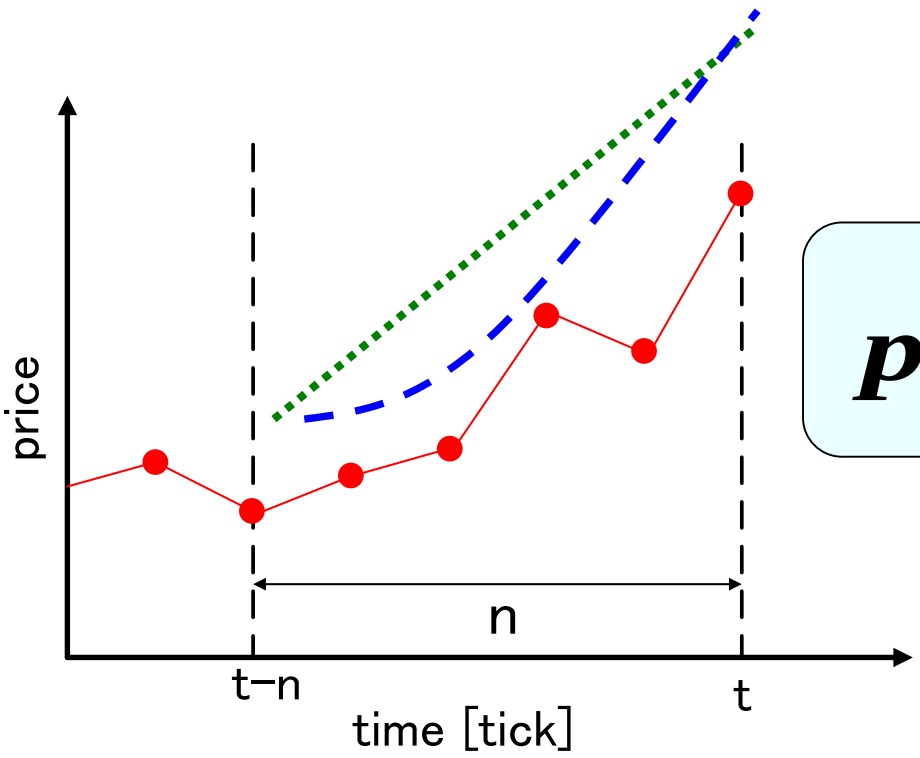
Hit Rate vs. Prediction Term

$/¥ rate(2000)



- Looking for new indicators?

Evaluate velocity and acceleration



$$p(t) = A + Bt$$

Initial Price

Velocity

$$p(t) = \alpha + \beta t + \frac{1}{2}\gamma t^2$$

Initial Price

acceleration

Initial Velocity

2 Dimensionless parameters

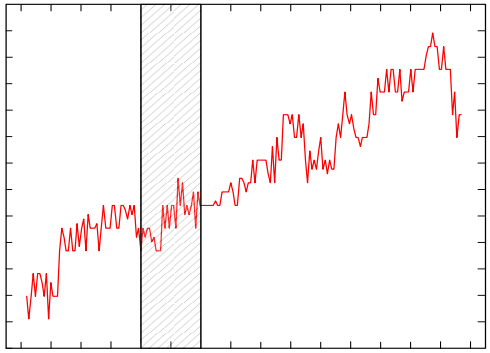
$$F = \frac{\beta^2}{\alpha\gamma}$$

$$T = \frac{nB}{A}$$

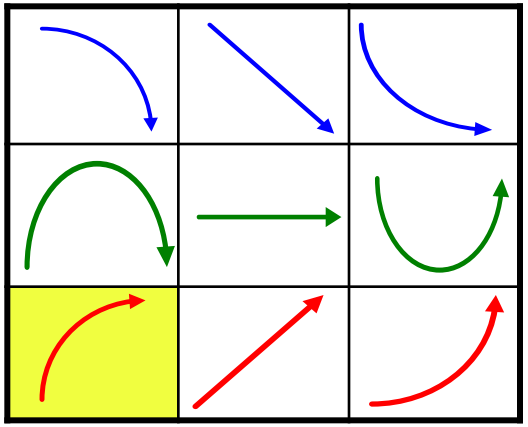
Velocity-Acceleration Patterns

Pattern Classification in terms of v,a, F,T

Dynamical Parameters



Find strong correlation to price !



Short-term Prediction

① Improve Prediction Rate by v,a,F,T
② Search for effective patterns

Job flow

Step 1

Set up conditions

Pre-Process

Step 2

Pattern Classification by dynamical parameters

Step 3

Construct Prediction Strategies

Step 4

Prediction

Predictions

Patterns classification

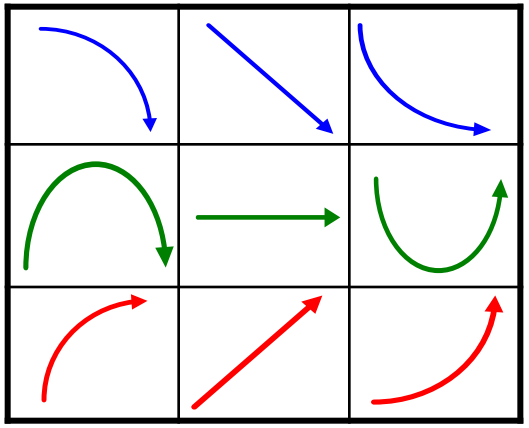
Evaluate

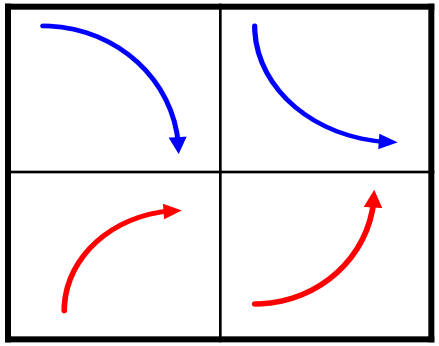
B, β, γ

F, T

Set up

- data
- threshold
- # of patterns





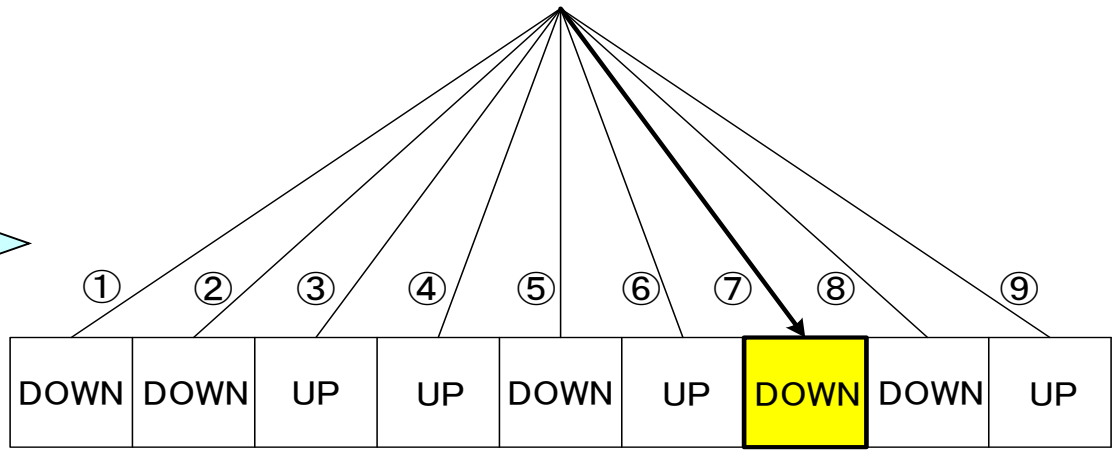
Find the best patterns

Generating Predictions

	$B < B_{down}$	$B_{down} < B < B_{up}$	$B_{up} < B$
$\gamma < \gamma_{down}$	①	②	③
$\gamma_{down} < \gamma < \gamma_{up}$	④	⑤	⑥
$\gamma_{up} < \gamma$	⑦	⑧	⑨

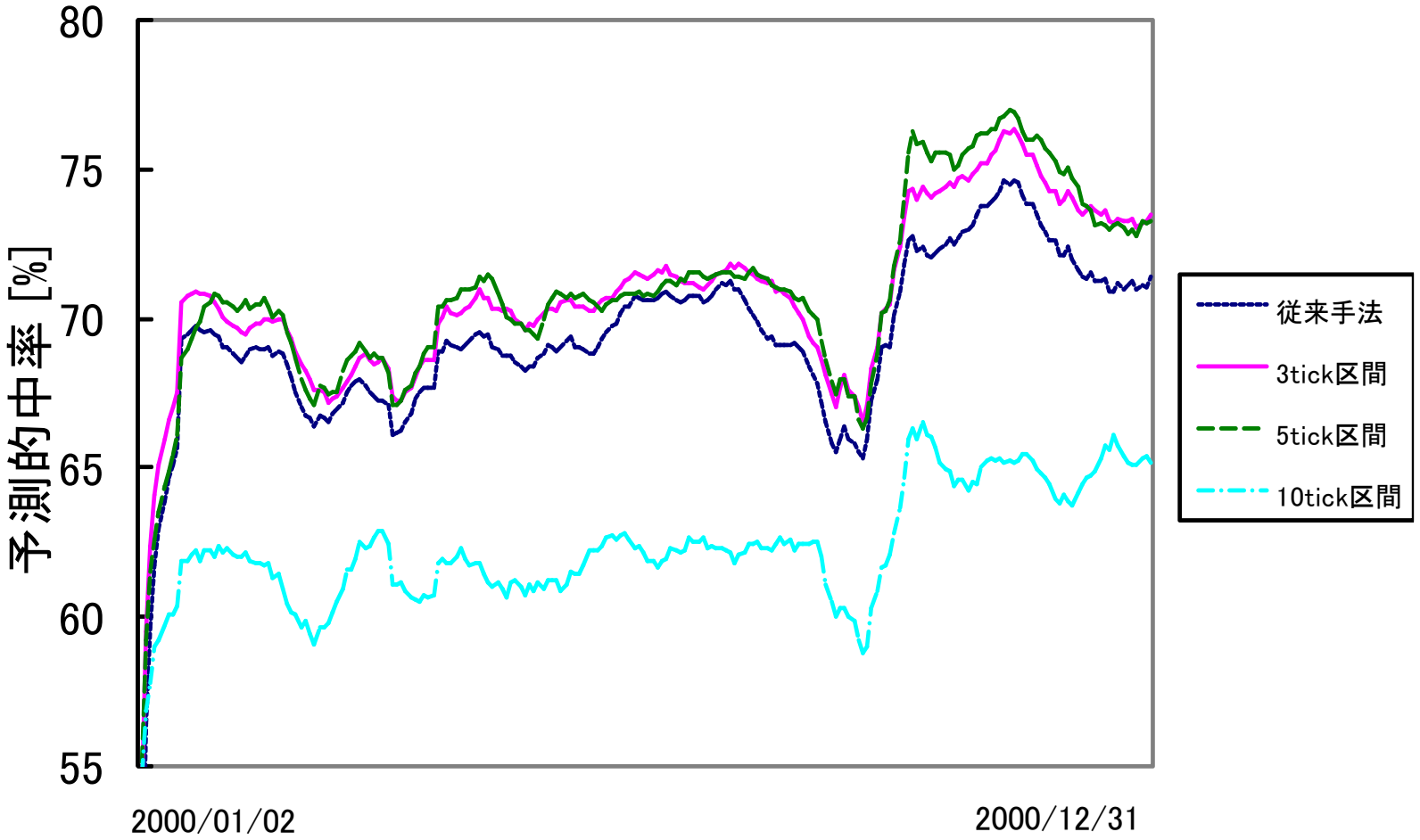
Observed patterns

Prediction

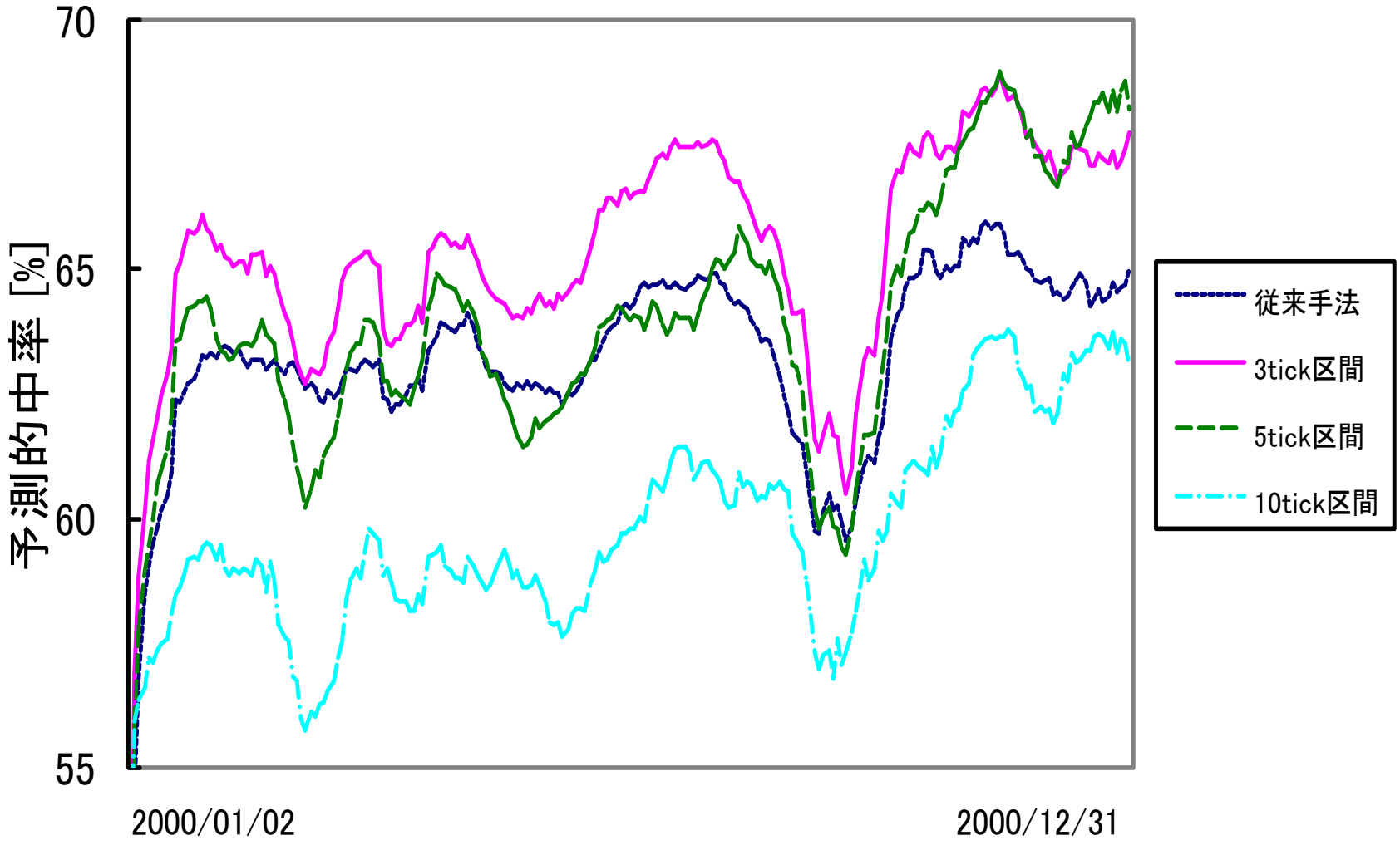


patterns = p ⇒ # strategies = 2^p

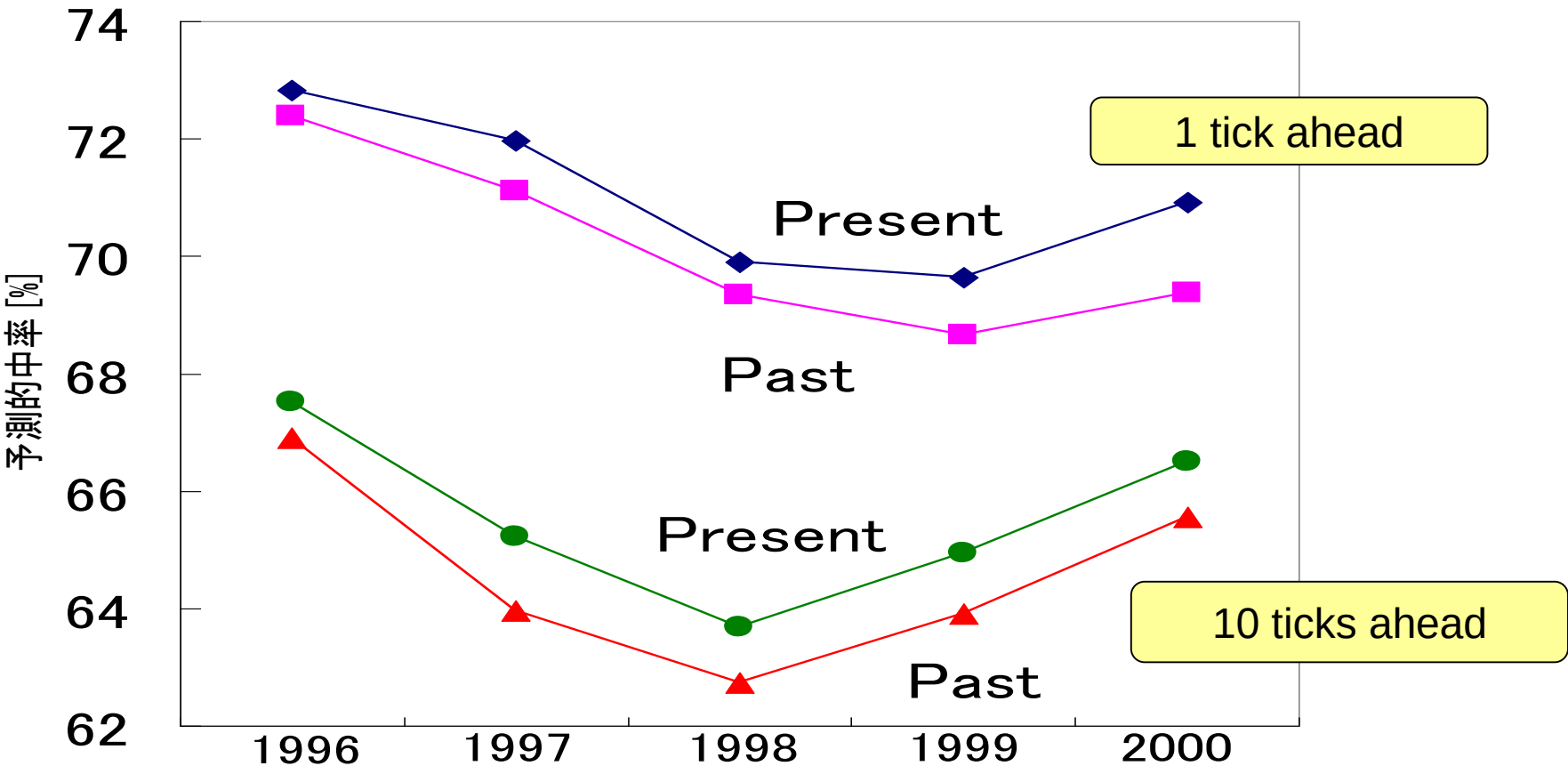
Hitting rate at 1 tick ahead (USD/JPY 2000)



Hitting rate at 10 ticks ahead (USD/JPY 2000)

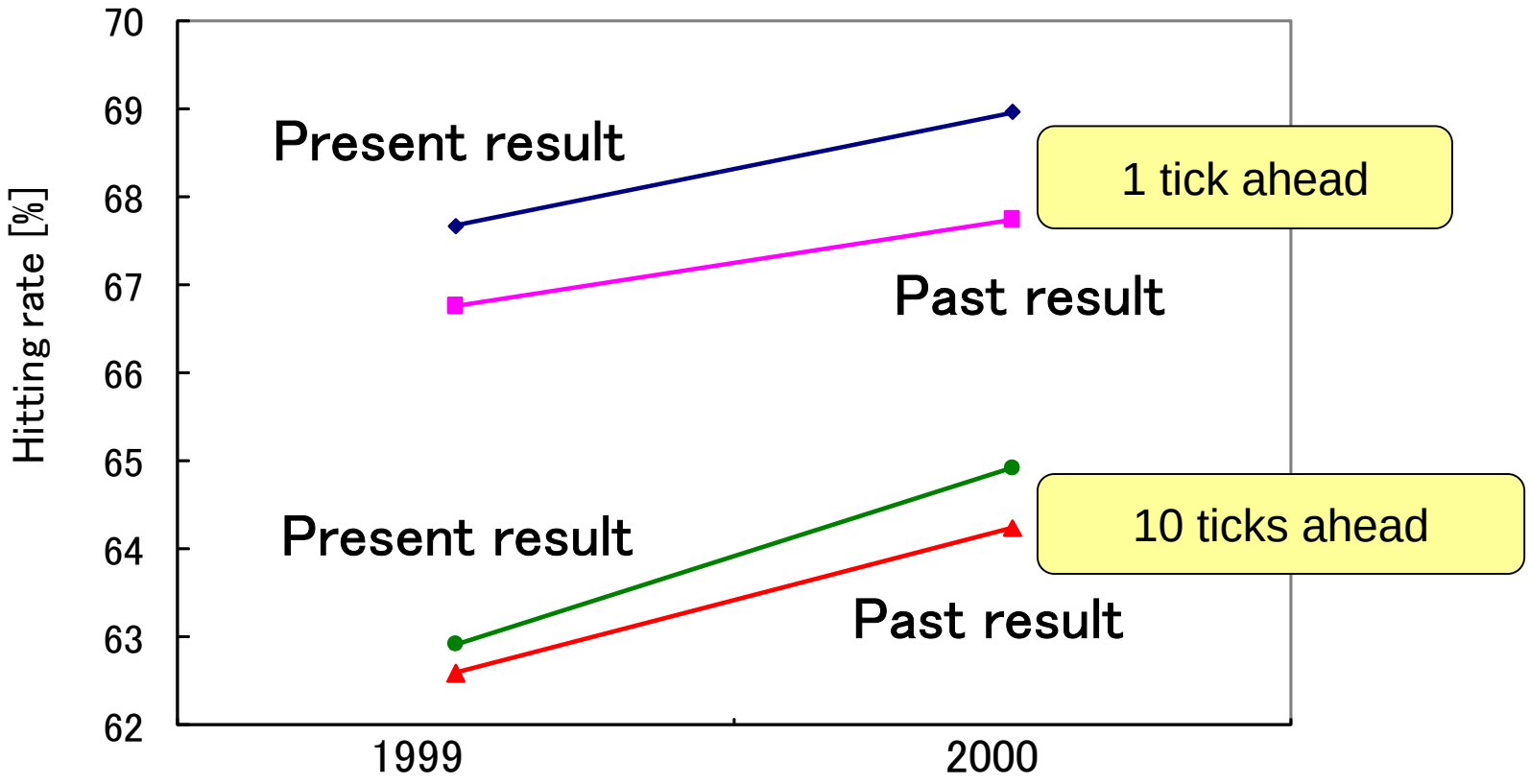


Improved Prediction Rates



Hitting rates improved for all years (’96-’00)

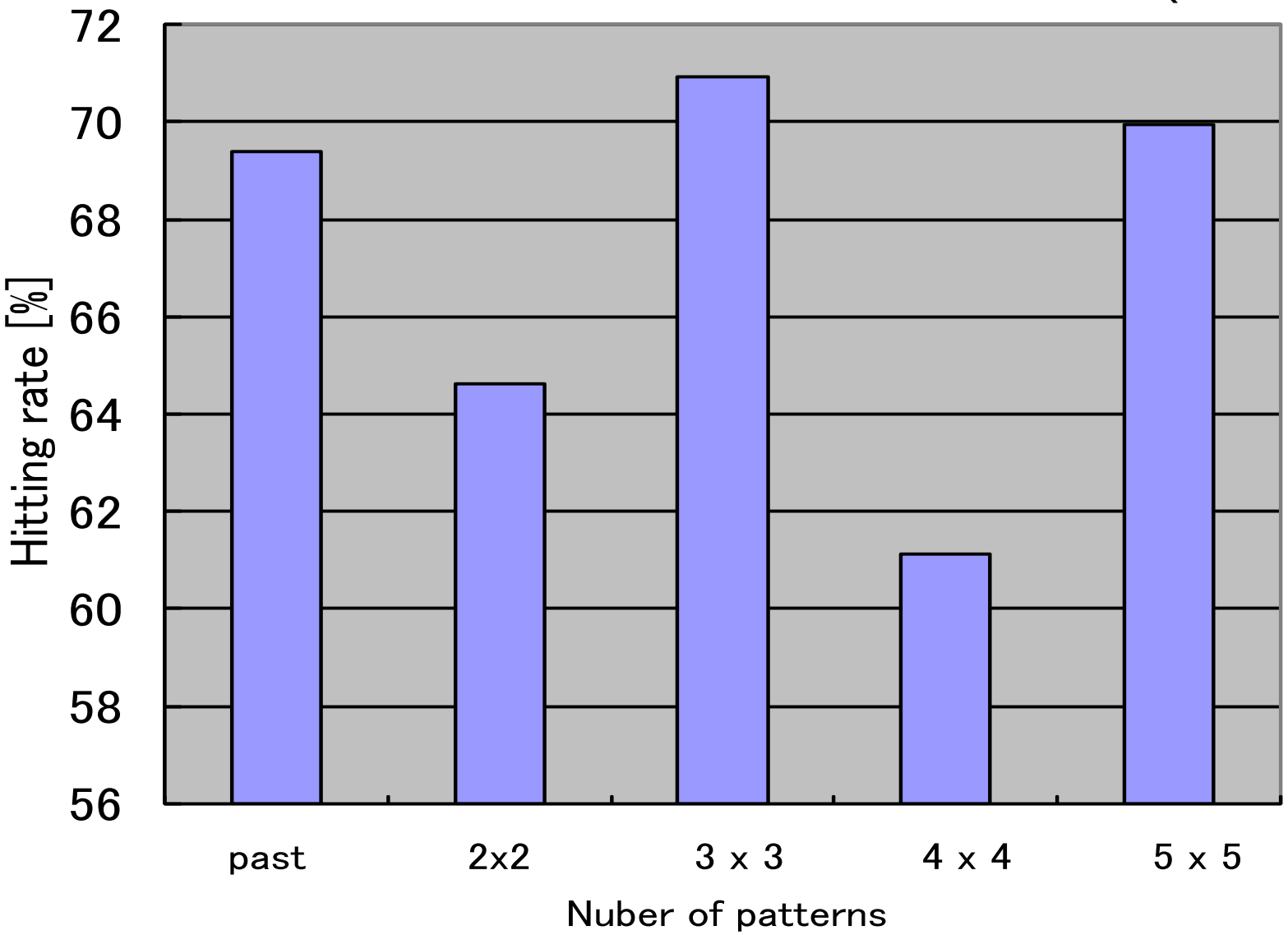
Improved hitting rate USD/JPY



Current result vs. old result in 1999-2000

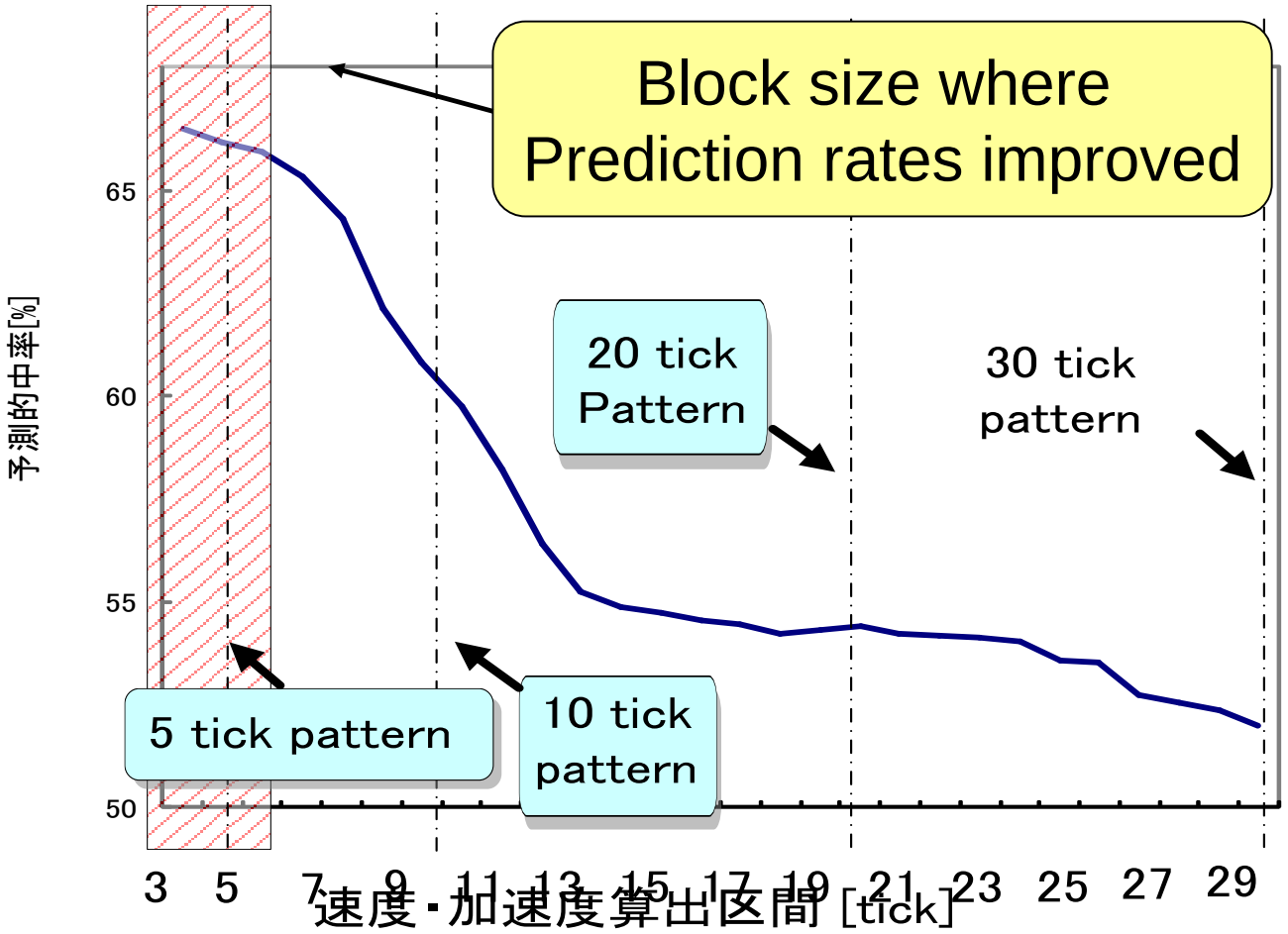
Hitting rate vs. number of patterns

USD/JPY (2000)



Prediction Rate vs. Pattern Length

- USD/JPY(2000) / Prediction at 10 ticks



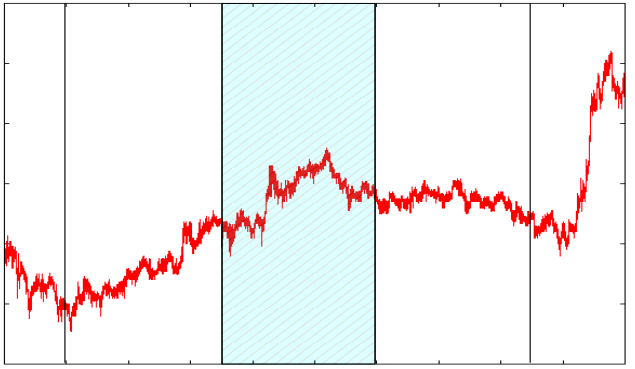
SOM classification

of weekly pieces

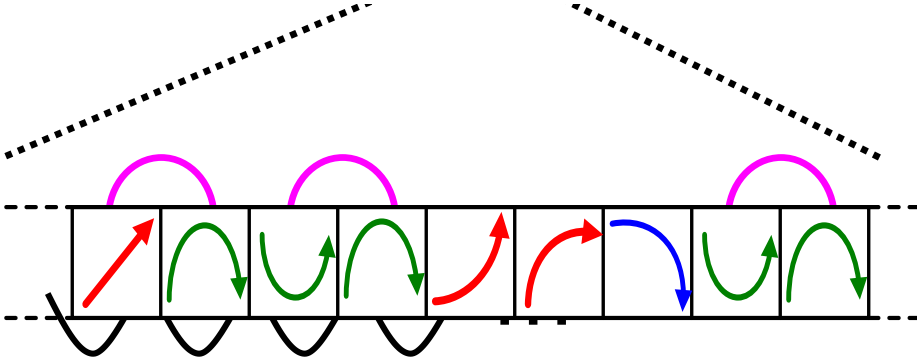
in 2000

Classifying time series pieces by SOM

① Cut tick data into pieces



② Pattern classification

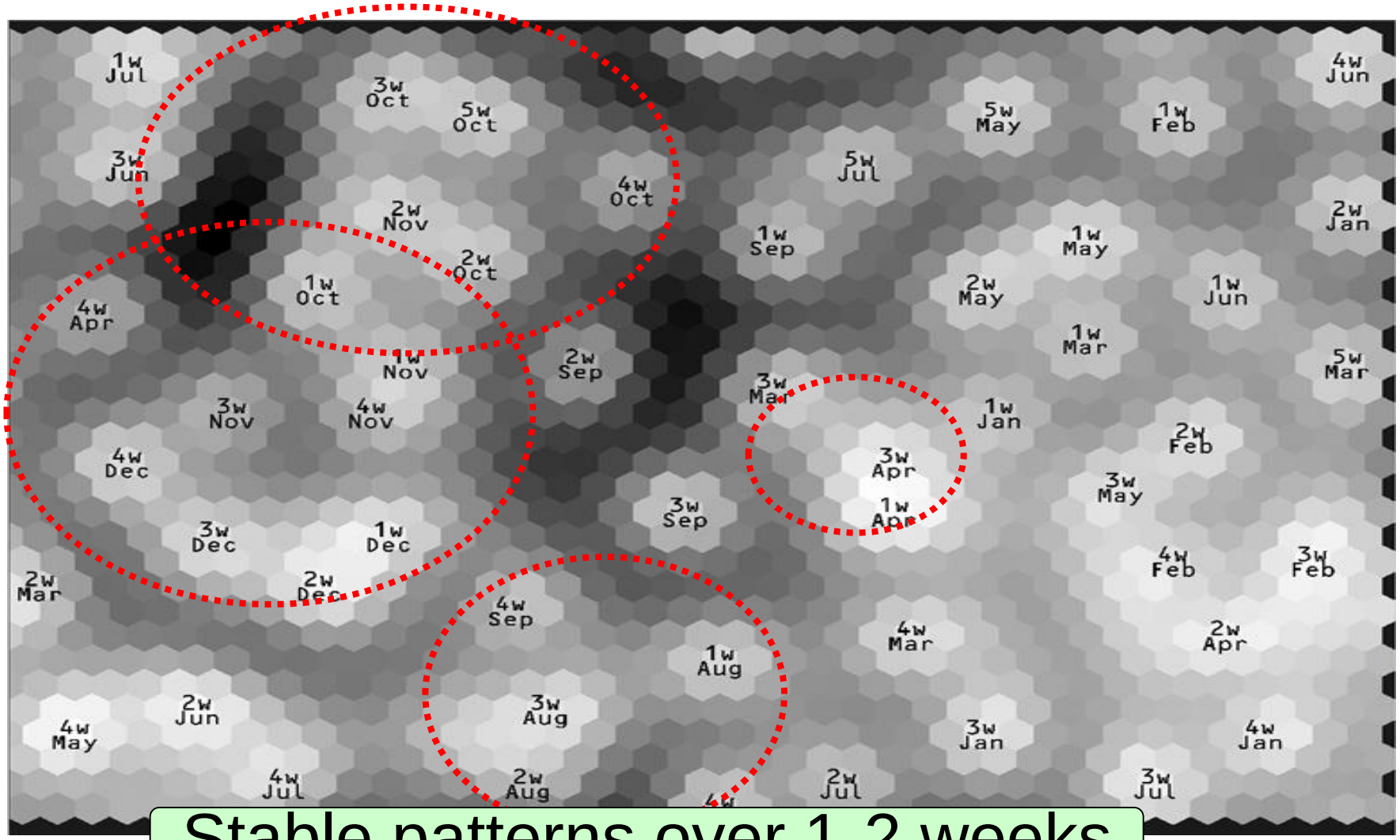


③ SOM classification
of time series pieces

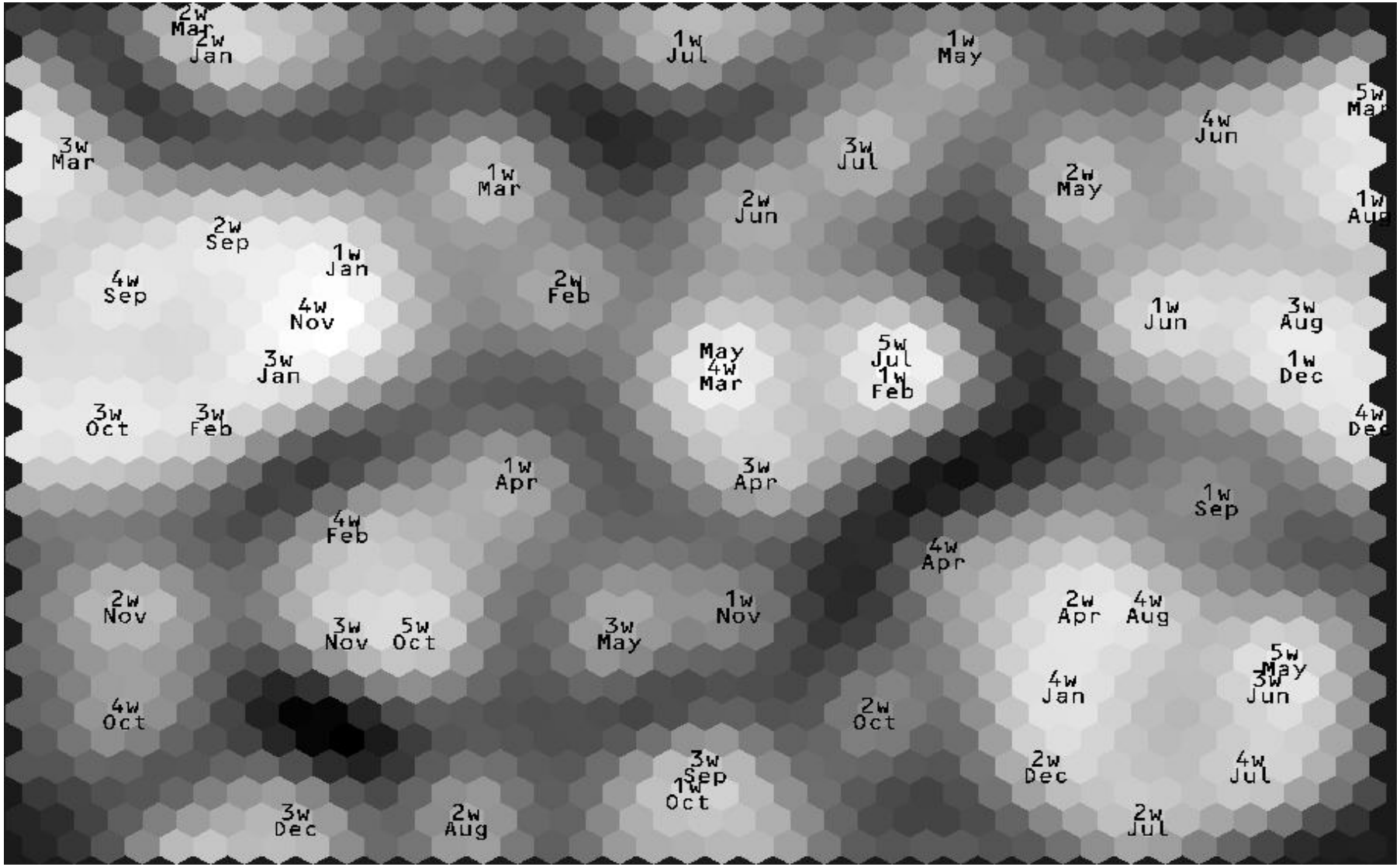
Prob. occurrence of 9 patterns
9 × 9 = 81 transition prob.

Result of classification

- SOM classification of weekly pieces in 2000

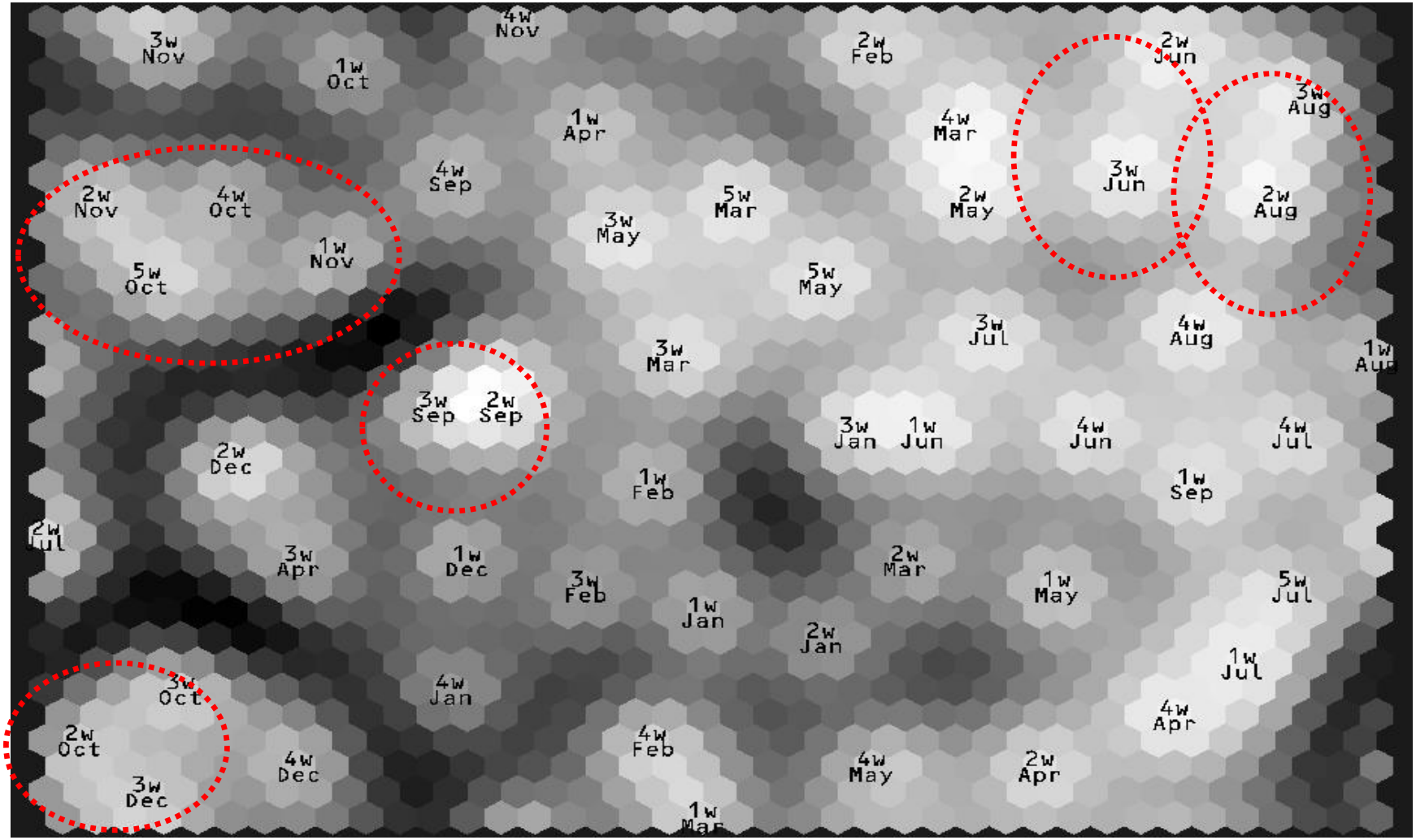


Stable patterns over 1-2 weeks


2w Mar
2w Jan
1w Jul
1w May
5w Mar
4w Jun
3w Mar
3w Jul
1w Mar
2w May
2w Jun
1w Aug
2w Sep
1w Jan
4w Sep
2w Feb
4w Nov
1w Jun
3w Aug
3w Jan
May 4w Mar
5w Jul
1w Feb
1w Dec
3w Oct
3w Feb
4w Dec
1w Apr
3w Apr
1w Sep
4w Feb
4w Apr
2w Nov
1w Nov
2w Apr
4w Aug
3w Nov
5w Oct
3w May
5w May
3w Jun
4w Jan
4w Oct
2w Oct
3w Sep
1w Oct
2w Dec
4w Jul
3w Dec
2w Aug
2w Jul

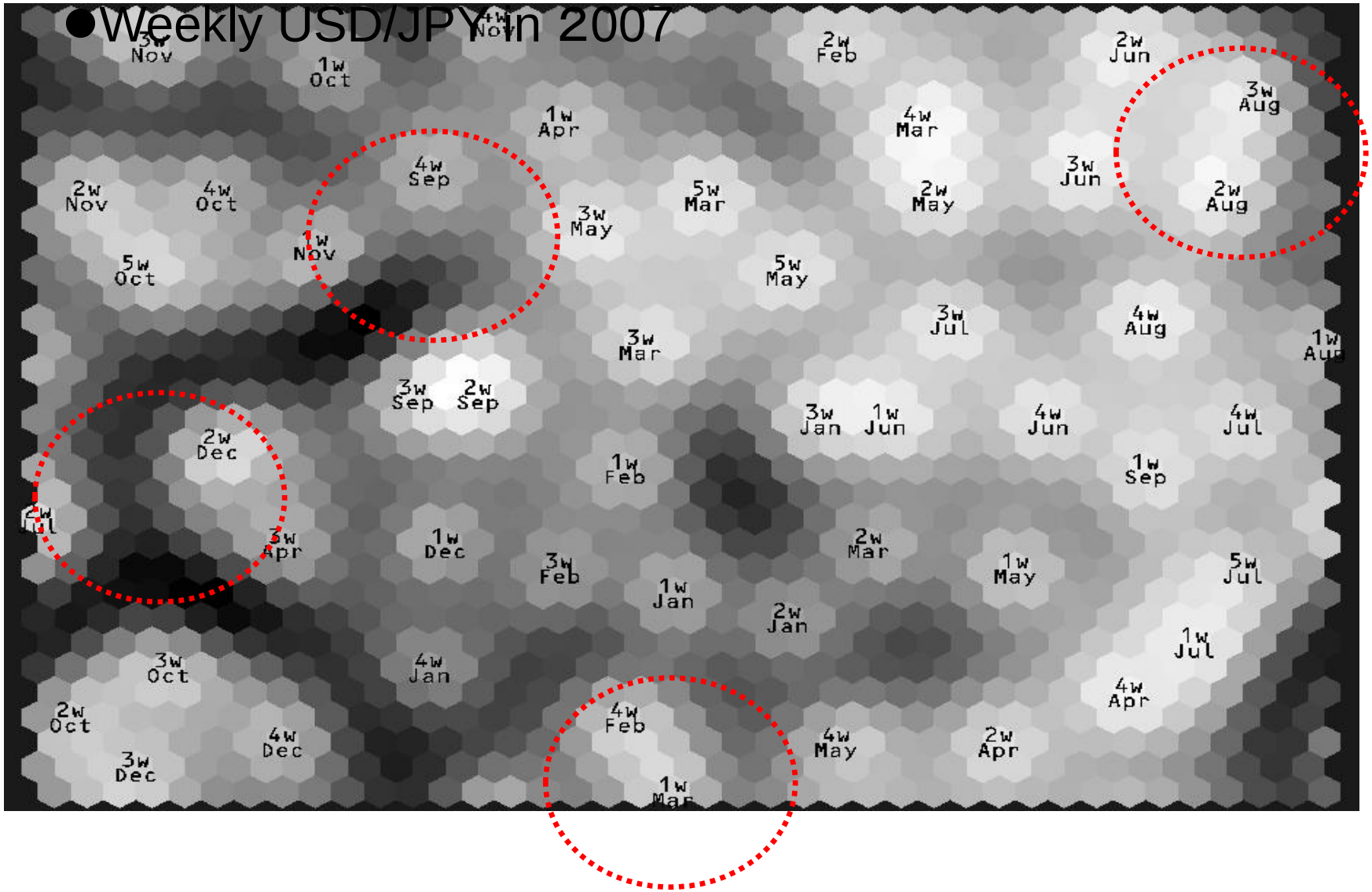
Result : SOM Classification

- Weekly USD/JPY in 1999

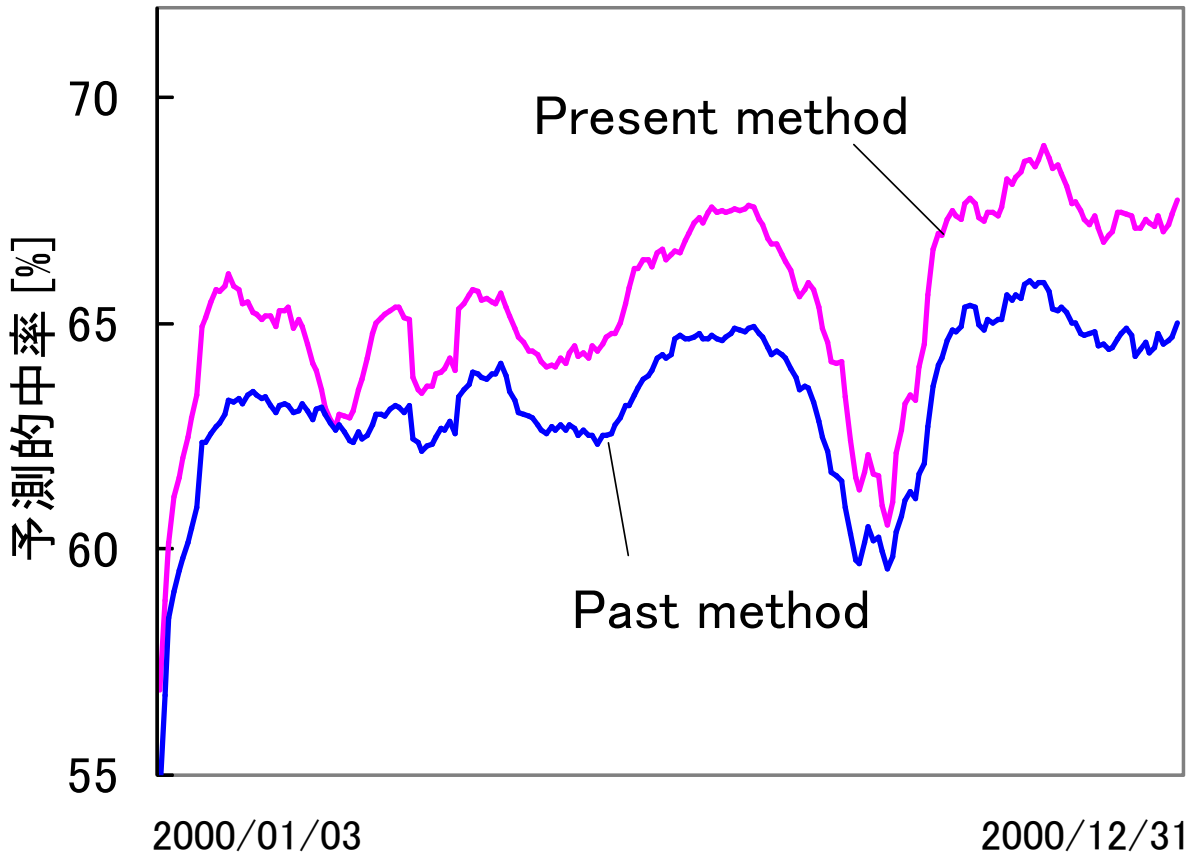


Result : SOM classification

- Weekly USD/JPY in 2007



Hitting rate in a year of 2000
(Predicting 10tick ahead data of USD/JPY)



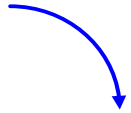
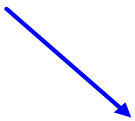
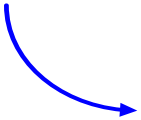
Hitting rate in a year (2000)

Probability distribution of patterns

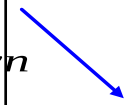
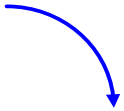
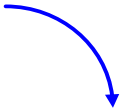
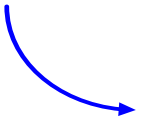
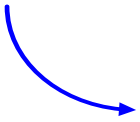
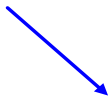
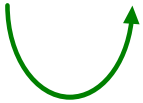
Patterns to obtain best result

Parameters	F, γ, T
Length of p	3～6 ticks
Number of p	9

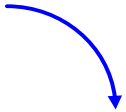
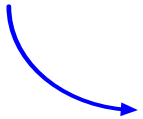
Occurrence probability of patterns for USD/JPY 2000

									
Prob.	10.60	1.52	10.00	24.69	4.38	26.82	10.52	1.46	10.04

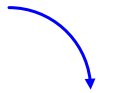
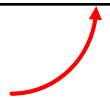
Patterns (1)

	$F < F_{down}$		$F_{down} < F < 0$		$0 < F < F_{up}$		$F_{up} < F$	
	$\beta < 0$	$0 < \beta$	$\beta < 0$	$0 < \beta$	$\beta < 0$	$0 < \beta$	$\beta < 0$	$0 < \beta$
$T < T_{down}$		×						×
$T_{down} < T < T_{up}$								
$T_{up} < T$	×						×	

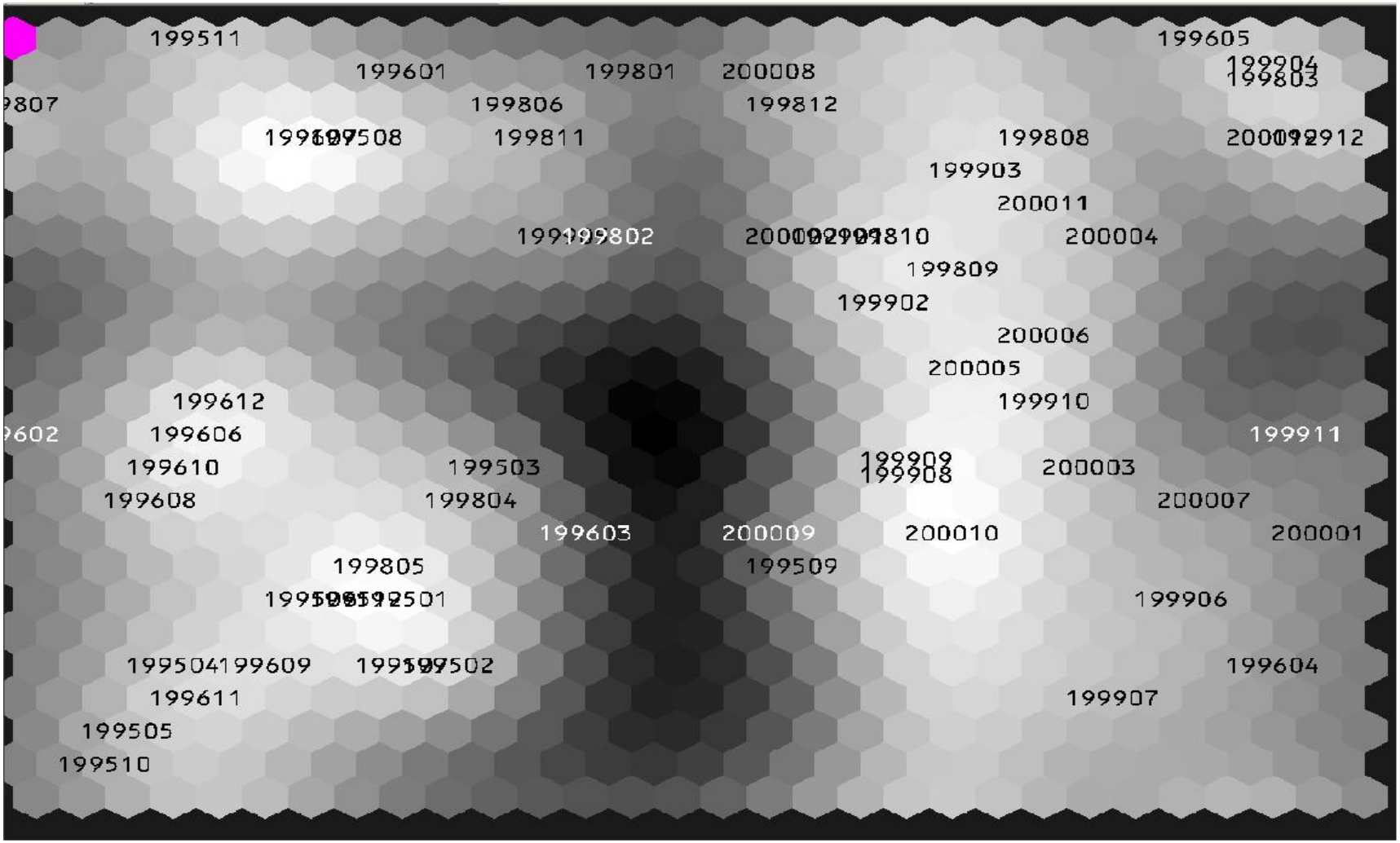
patterns(2)

	$No\ F$ $(\gamma = 0)$	$F = 0(\beta = 0)$		
		$\gamma < \gamma_{down}$	$\gamma_{down} < \gamma < \gamma_{up}$	$\gamma_{up} < \gamma$
$T < T_{down}$				
$T_{down} < T < T_{up}$				
$T_{up} < T$				

Conditional probabilities P(x|y) USD/JPY 2000

									
	12.0	12.04	11.6	10.44	9.08	9.96	10.6	10.7	10.6
	1.91	2.48	1.73	1.52	1.34	1.33	1.38	1.7	1.45
	11.71	12.34	11.56	10.07	8.58	8.93	9.81	8.57	9.67
	22.1	22.52	23.44	24.4	26.42	26.34	24.47	24.75	24.81
	3.67	3.85	3.7	4.73	6.89	4.69	3.85	3.57	3.74
	27.13	26.49	26.03	28.71	28.53	26.65	25.38	24.54	24.32
	9.64	9.21	10.48	9.36	9.2	10.87	11.81	12.32	12.54
	1.33	1.3	1.32	1.3	1.36	1.48	1.68	2.59	1.75
	10.51	9.79	10.15	9.46	8.6	9.75	11.01	11.26	11.11

Monthly USD/JPY(1995-2000)



Summary

- **Automatic forecast of several ticks ahead**
- **Using the patterns observed in tick-wise price fluctuation**
- **Evolutional approach successful in terms of direction**
- **Noticing the new phase helps improving the forecast.**

Cross Correlation Spectra compared with Random Matrix Theory

Applied on tick data

NYSE-TAQ 1994 vs. 2002

Use of Random Matrix Theory

How can we separate significant information from the flood of randomness ?

A good recipe is given by

V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N.Amarmal, T. Guhr, H.E. Stanley,

"Random matrix approach to cross correlation in financial data",

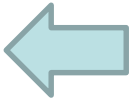
Physical Review E 65, 066126, 2002.

Recipe 1 Correlation Matrix

Cross correlation between Stock- i and Stock- j

$$C_{i,j} = \frac{1}{L} \sum_{k=1}^{L} x_{i,k} x_{j,k}$$

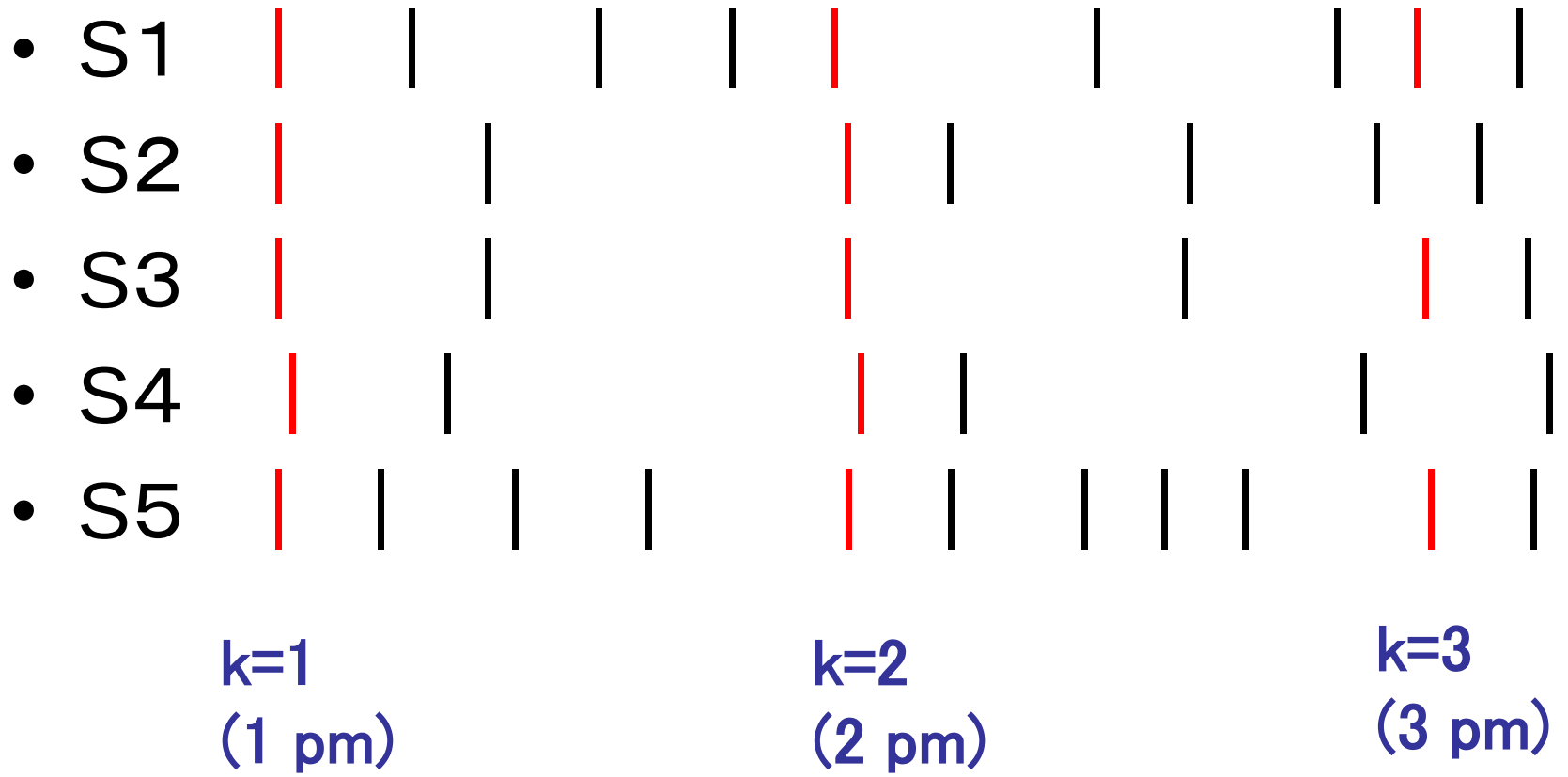
$$x_{i,k} = \begin{vmatrix} x_{1,1} & x_{1,2} & \cdot & x_{1,L} \\ x_{2,1} & x_{2,2} & \cdot & x_{2,L} \\ \cdot & \cdot & \cdot & \cdot \\ x_{N,1} & x_{N,2} & \cdot & x_{N,L} \end{vmatrix}$$



Price time series of Stock-1

Data at every tick (k=1,..,L) for all N is needed!

Trouble of tick-wise stock prices

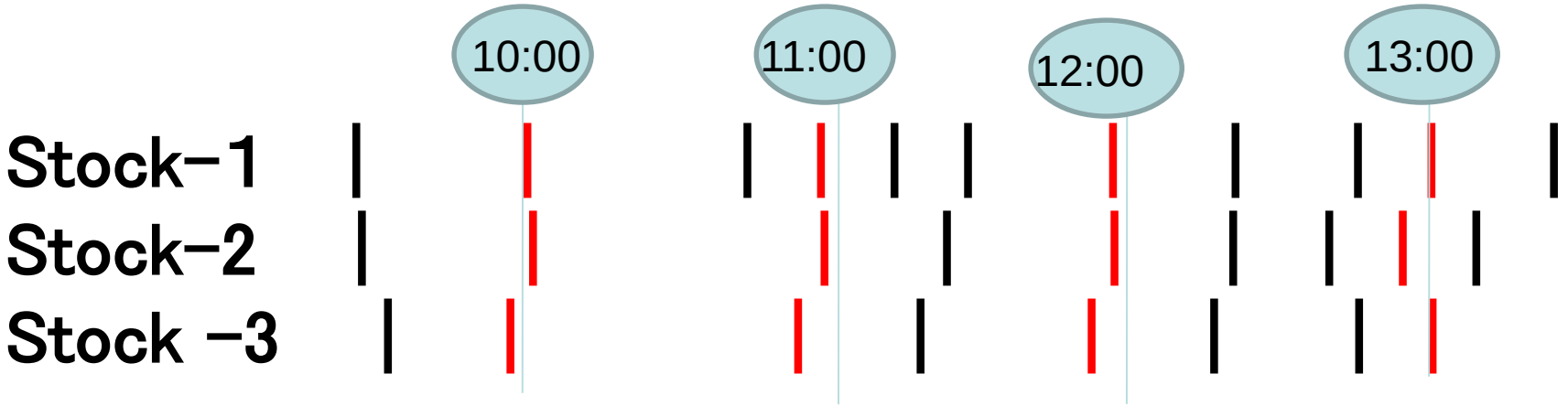


1) Time intervals between ticks are not regular

2) Not all symbols have values at every k

N ↑ means L ↓

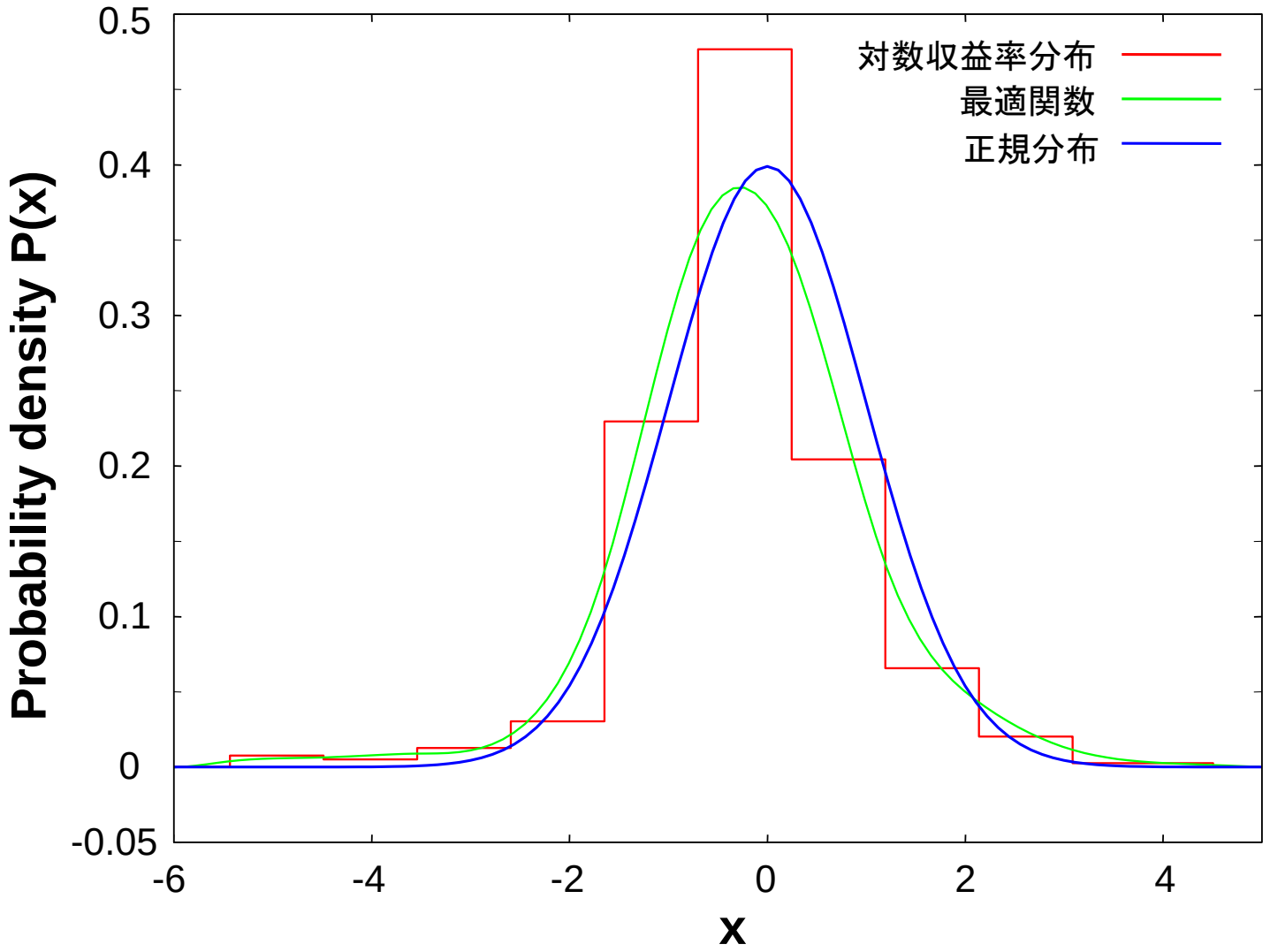
- L=1,512 for every 1 hour
  （10 am-3 pm = 6 pts × 252days = 1,512)
- 1994 NYSE-TAQ : N=419
- 2002 NYSE-TAQ : N=569



Use the nearest prices to every hour

Normalized distribution of x compared to N(0,1)

$$X_{i,k} = \ln(S_{i,k+1} / S_{i,k}) \approx \Delta S_{i,k} / S_{i.k}$$



7 step recipe

1) Get price time series $S_{i,k}$ (i=1,.., L) for k=1,.., N stocks

2) Compute return time series

$$X_{i,k} = \ln(S_{i,k+1} / S_{i,k}) \approx \Delta S_{i,k} / S_{i.k}$$

3) Compute $x_{i,k}$ by normalizing $X_{i,k}$ (mean=0, variance =1)

4) Compute cross correlation matrix $C_{i,j} = \frac{1}{L} \sum_{k=1}^{L} x_{i,k} x_{j,k}$

5) Solve eigenvalue problem of

$$CV = \lambda V$$

6) Compare the eigenvalues with corresponding RMT

7) Discrepancy is the SIGNIFICANT INFORMATION

RMT tells us..

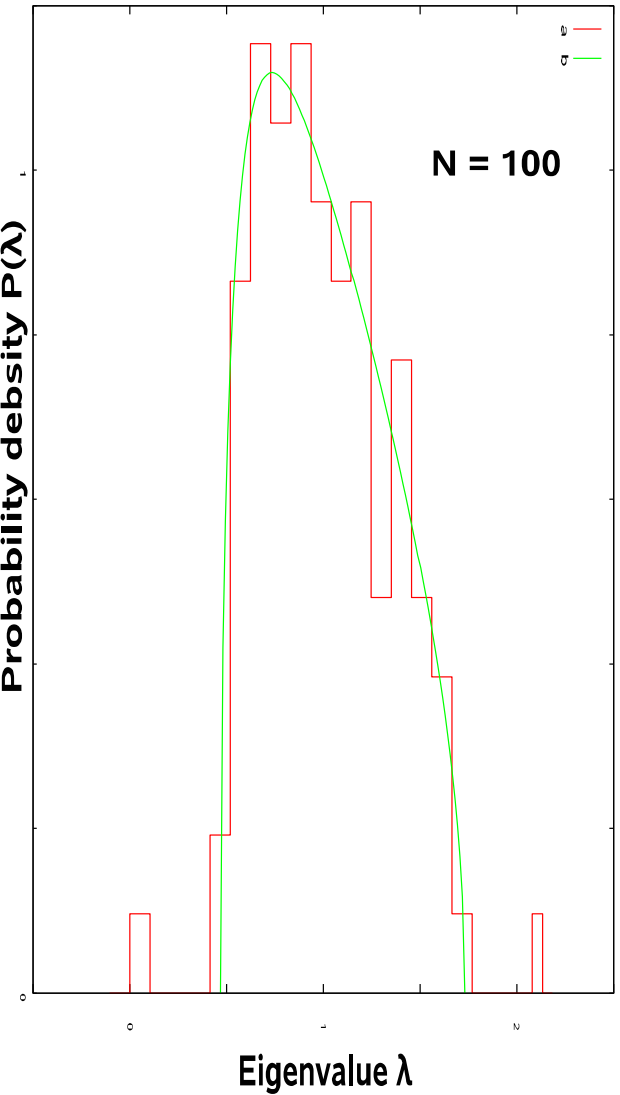
$$N \to \infty,\ L \to \infty,\ Q = L/N = \text{const.}$$

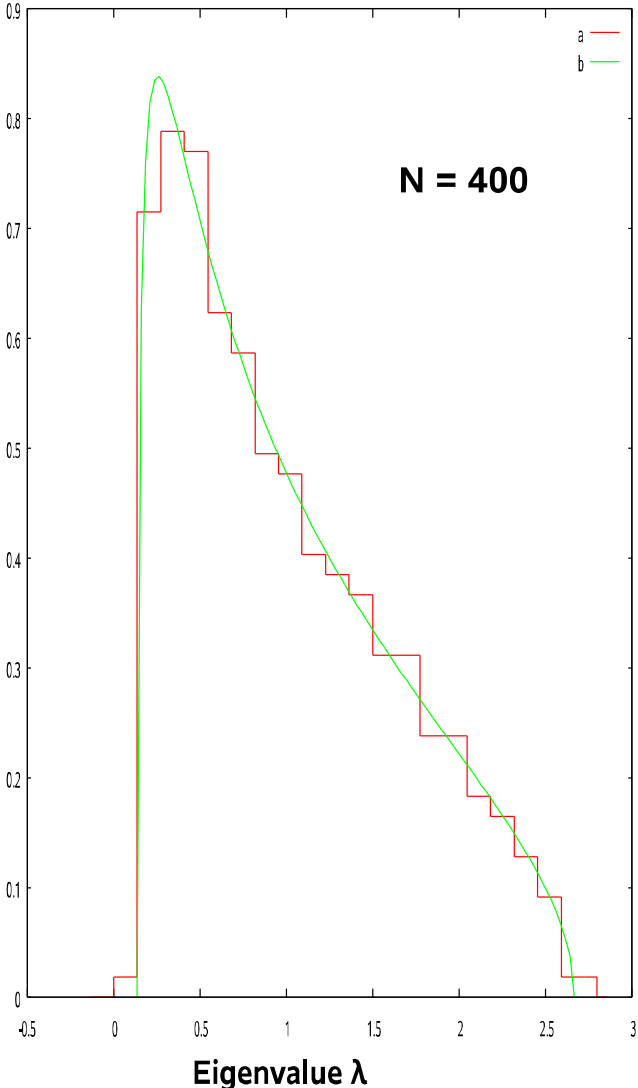
Eigenvalue (λ) distribution is given by

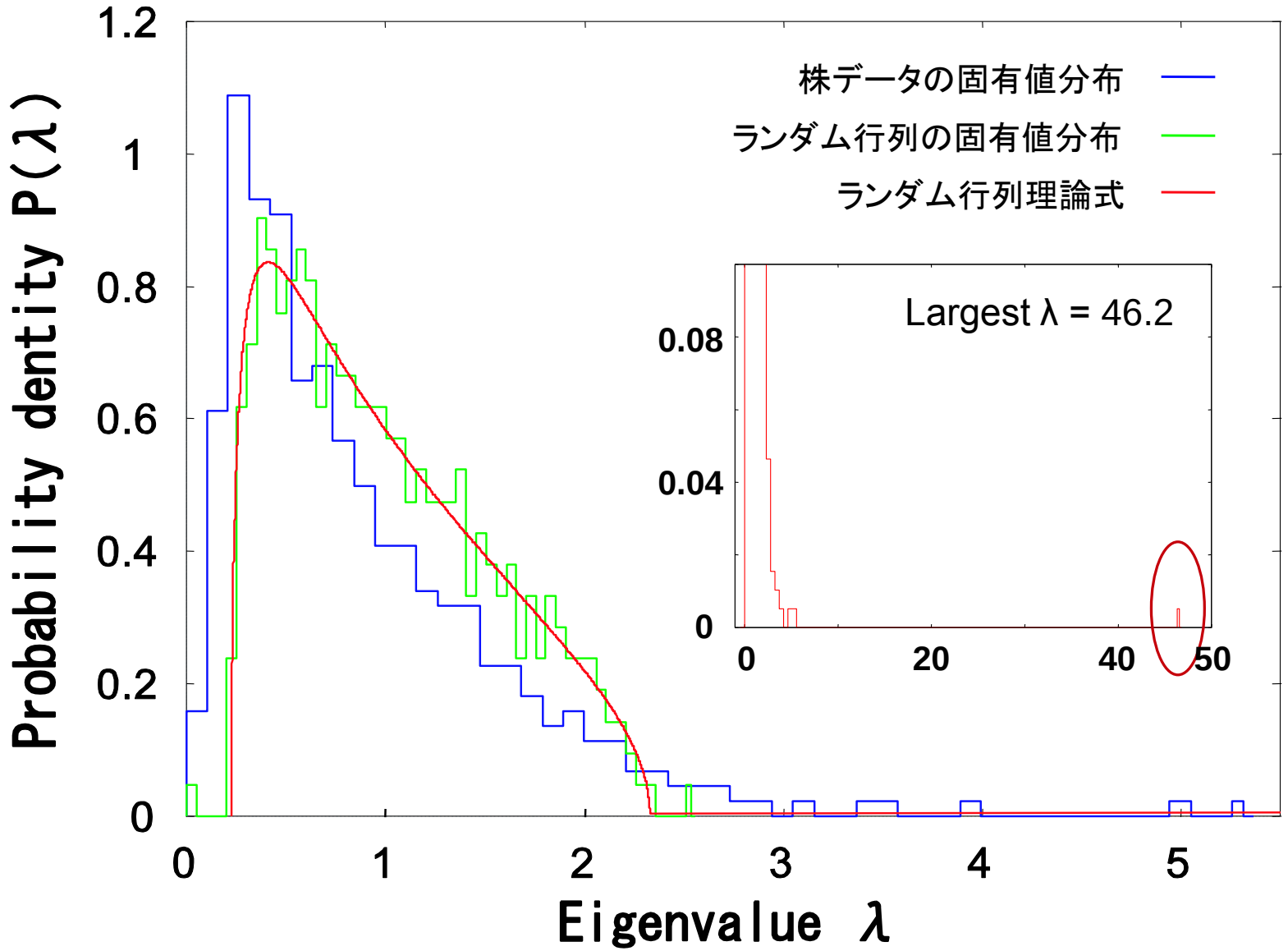
$$P_{RMT}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}, \quad (\lambda_- < \lambda < \lambda_+)$$

Check
Comparison to RMT for various N







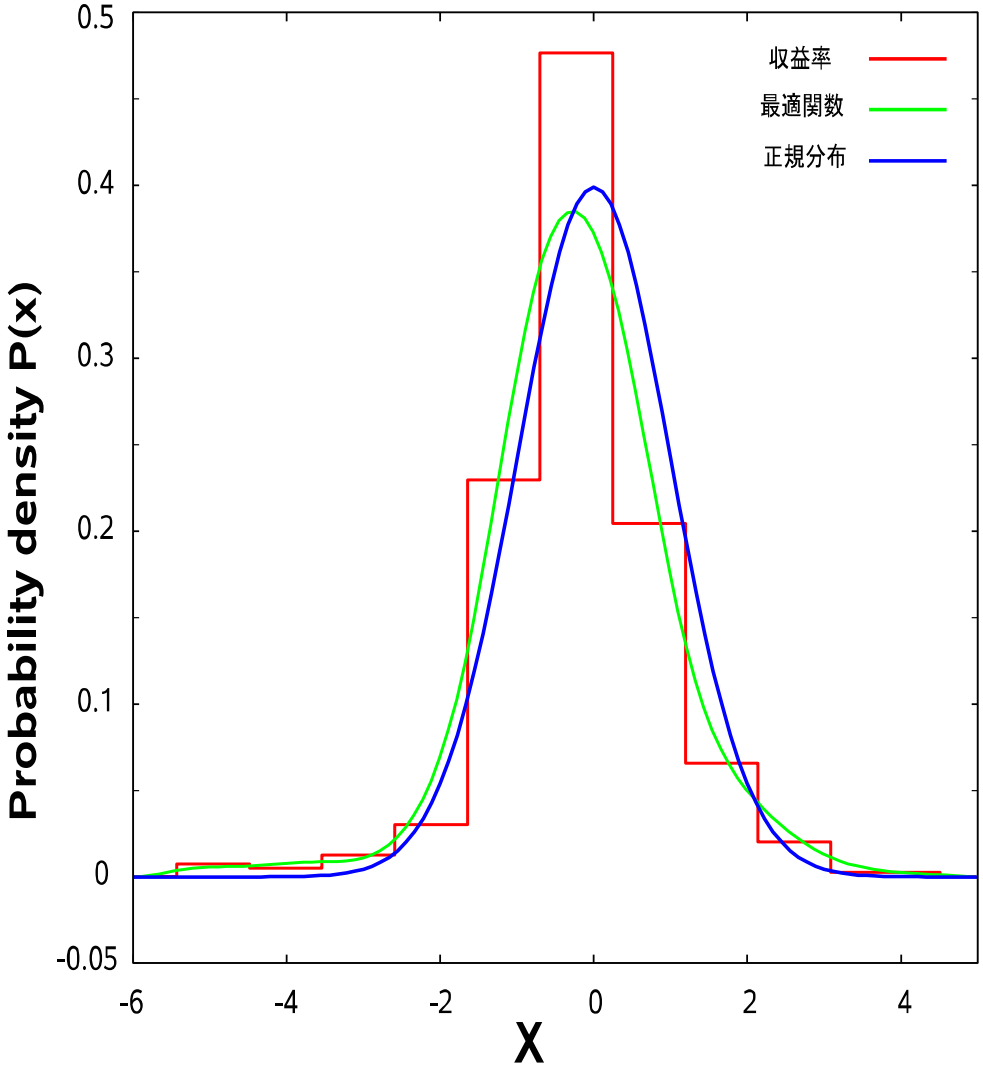
λ_3 SIGNAL $\lambda_4 = 3.9, \lambda_5 = 3.5, \lambda_6 = 3.4, \lambda_7 = 3.1$

NOISE: λ_8

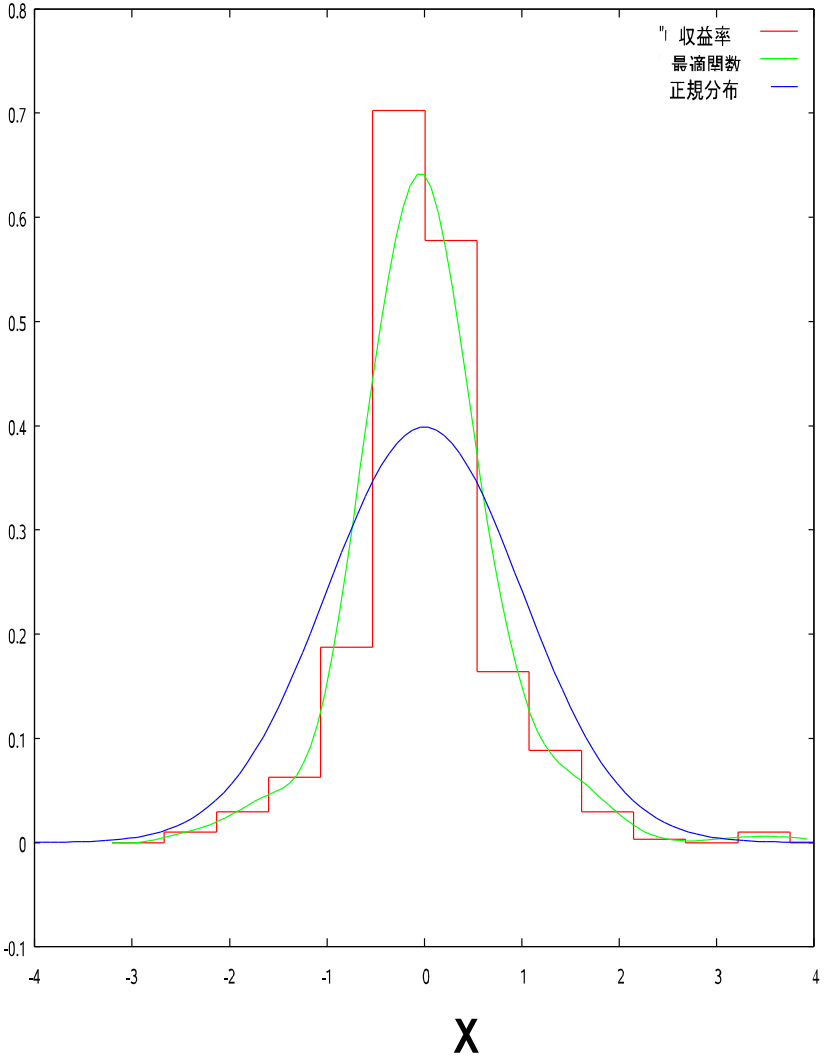
Eigenvector Configulations(94)

EIGENVALUE	EIGENVECTOR		
$\lambda_1 = 46.2$	BIG COMPANY	BIG COMPANY	BIG COMPANY
$\lambda_2 = 5.25$	GOLD MINING	---	---
$\lambda_3 = 5.04$	SEMICONDUCTOR	SEMICONDUCTOR	SEMICONDUCTOR
$\lambda_4 = 3.90$	SEMICONDUCTOR	SEMICONDUCTOR	SEMICONDUCTOR
$\lambda_5 = 3.51$	OIL	OIL	OIL
$\lambda_6 = 3.41$	---	---	---
$\lambda_7 = 3.11$	PAPER	PAPER	---
λ_8 IN NOISE	---	---	---
λ_9 IN NOISE	FINANCE	FINANCE	---
λ_{10} IN NOISE	AUTOMOBILE	AUTOMOBILE	COMMUNICATION

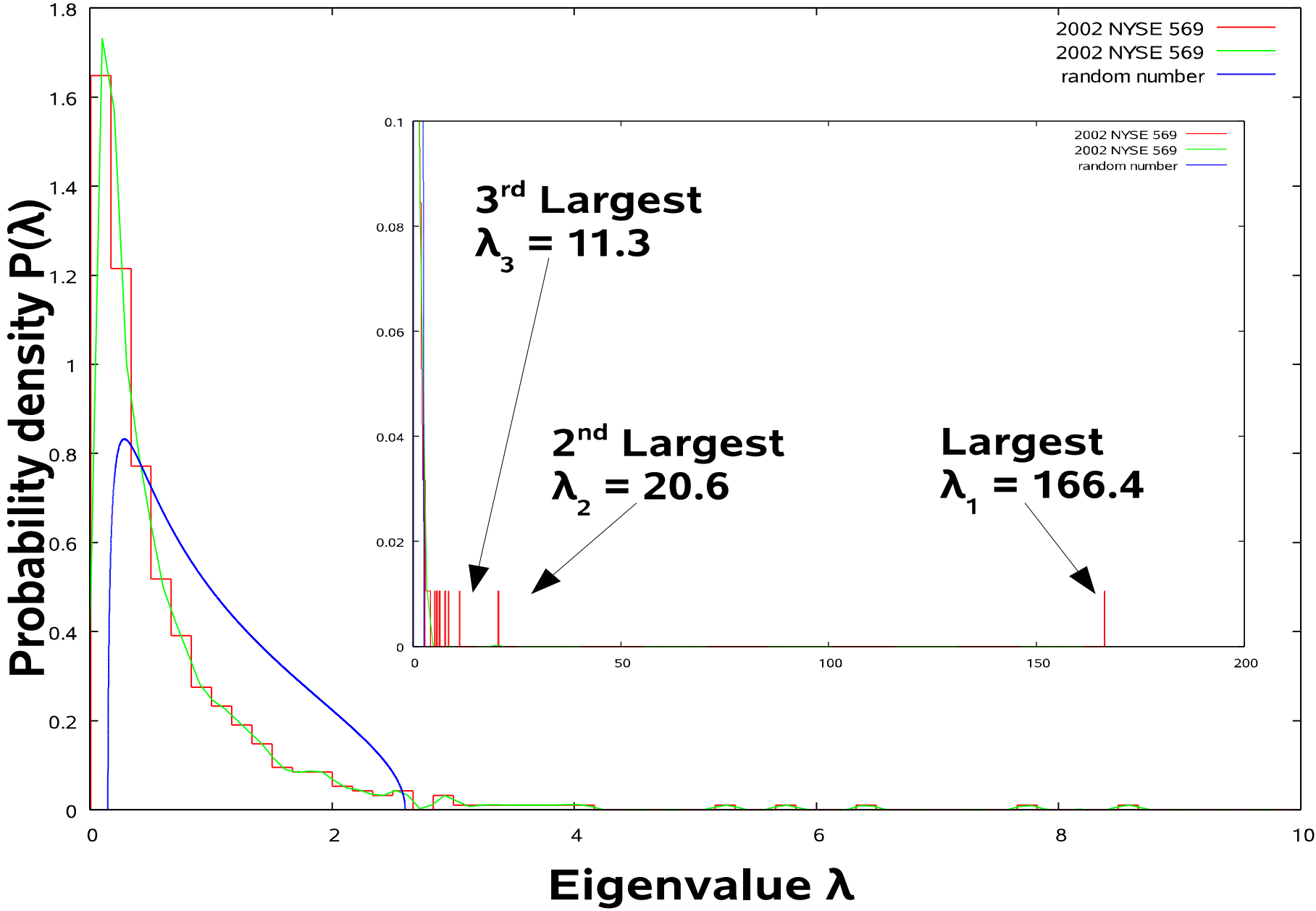
NYSE-TAQ of 2002 N(0,1)



1994年



2002年


2002 NYSE 569
2002 NYSE 569
random number
Probability density P(λ)
Eigenvalue λ
3rd Largest
λ3 = 11.3
2nd Largest
λ2 = 20.6
Largest
λ1 = 166.4

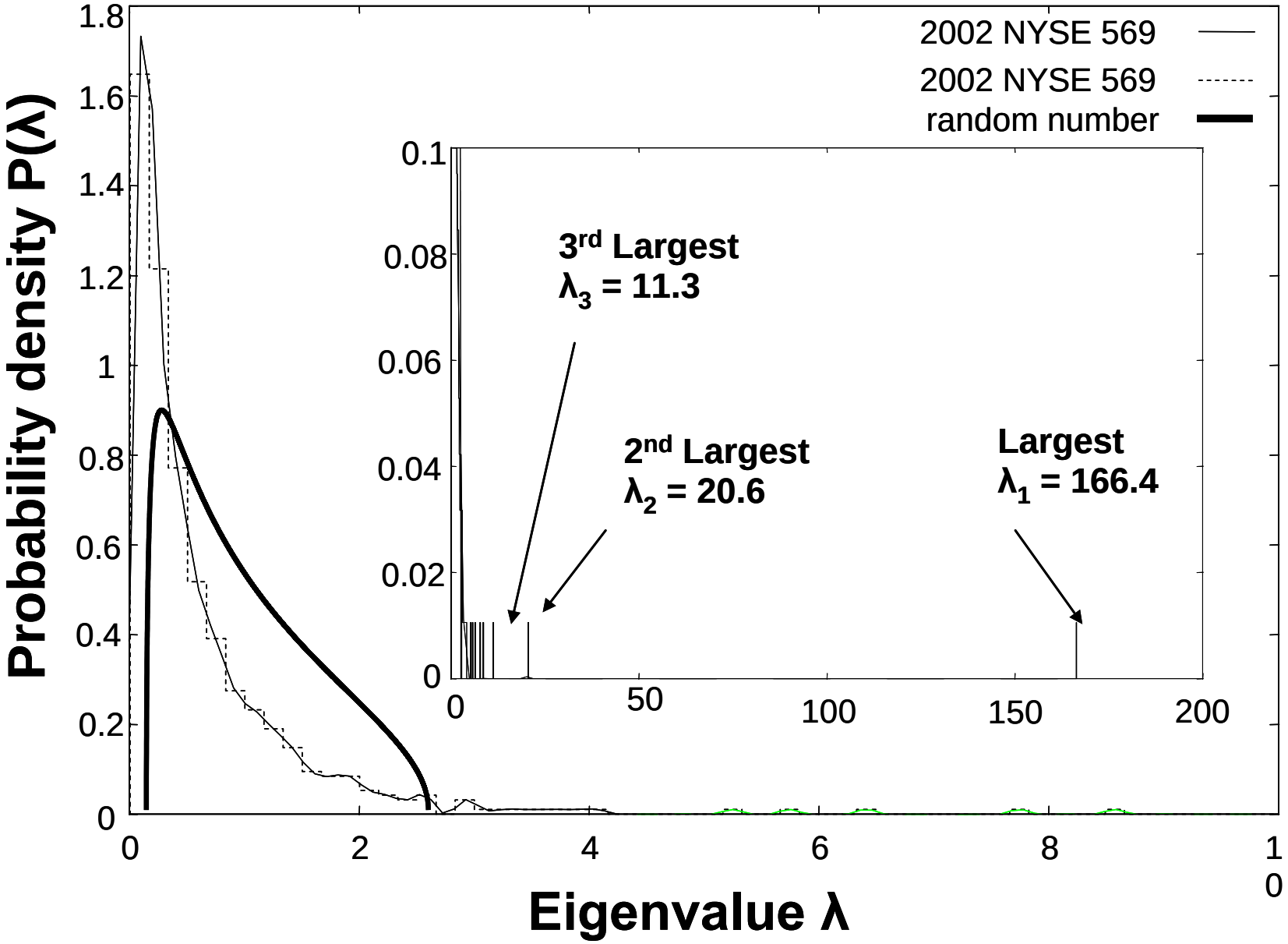

2002 NYSE 569
2002 NYSE 569
random number
Probability density P(λ)
Eigenvalue λ
3rd Largest
λ3 = 11.3
2nd Largest
λ2 = 20.6
Largest
λ1 = 166.4
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
4
6
8
10
0.02
0.04
0.06
0.08
0.1
50
100
150
200

Table2. 1994年固有ベクトル上位10成分の内訳

固有ベクトル	傾向
u1	業種を選ばず大企業数十社が並ぶ(各企業の比重は0.08程度)
u2	1位のBarrick Goldに集中(0.3), 2位以下は特徴なし
u3	10社中7社が半導体に関する企業
u4	10社中5社が半導体に関係する企業
u5	10社中9社が石油に関係する企業
u6	業種に一貫性がない
u7	紙に関する企業が2社だが, 他の業種には一貫性がない
u8	業種に一貫性がない
u9	10社中6社が金融に関する企業、2社は通信に関する企業
u10	車に関する企業が2社, 通信関係が3社

Eigenvector Configulations(02)

EIGENVALUE	EIGENVECTOR		
$\lambda_1 = 166.4$	BANK	BANK	BANK
$\lambda_2 = 20.6$	FOOD	FOOD	FOOD
$\lambda_3 = 11.3$	ENERGY	ENERGY	ENERGY
$\lambda_4 = 8.6$	FOOD	FOOD	FOOD
$\lambda_5 = 7.7$	ENERGY	ENERGY	ENERGY
$\lambda_6 = 6.5$	ELECTRIC	ELECTRIC	ELECTRIC
$\lambda_7 = 5.8$	FOOD	FOOD	FOOD
$\lambda_8 = 5.3$	RETAIL	RETAIL	RETAIL
$\lambda_9 = 4.1$	METAL	METAL	METAL
$\lambda_{10} = 4.0$	COMMUNICATION	COMMUNICATION	COMMUNICATION

Table3. 2002年固有ベクトル上位10成分の内訳

固有ベクトル	傾向
u1	10社中6社が金融関係、2社が銀行
u2	10社中6社が食品関係
u3	10社すべてが電気・エネルギー関係
u4	10社中5社が食品関係、3社が電気・エネルギー関係
u5	10社すべてが電気・エネルギー関係
u6	10社中3社が電気関係
u7	10社中3社が食品関係
u8	10社中8社が小売業関係
u9	10社中5社が鉱業関係、3社が通信関係
u10	10社中8社が通信関係

一時間足と日次変動の違い

- 定性的には似たような振る舞いをする
- 一時間足のほうがランダム性が強い
- 2002年には収益時系列

$$X_{i,k} = \ln(S_{i,k+1} / S_{i,k}) \approx \Delta S_{i,k} / S_{i.k}$$

を正規化した変数が正規乱数から大きく逸脱

（新しい変数が必要？）

今後の課題

参考文献

[1]M.L.Mehta, “RandomNMatrices”、Academic Press 3rd edition, 2004.

[2] 今野浩,“理財工学I ”, 日科技連出版社, 1995.

[3] V. Plerou, et.al., “Random matrix approach to cross correlation in financial data”, Physical Review E 65, 066126, 2002.

[4] Bouchaud，他 ”金融リスクの理論”

[5] 青山秀明，他 “経済物理学”,共立出版,2008

Eigenvector Configulations(94)

EIGENVALUE	EIGENVECTOR		
$\lambda_1 = 46.2$	BIG COMPANY	BIG COMPANY	BIG COMPANY
$\lambda_2 = 5.25$	GOLD MINING	---	---
$\lambda_3 = 5.04$	SEMICONDUCTOR	SEMICONDUCTOR	SEMICONDUCTOR
$\lambda_4 = 3.90$	SEMICONDUCTOR	SEMICONDUCTOR	SEMICONDUCTOR
$\lambda_5 = 3.51$	OIL	OIL	OIL
$\lambda_6 = 3.41$	---	---	---
$\lambda_7 = 3.11$	PAPER	PAPER	---
λ_8 IN NOISE	---	---	---
λ_9 IN NOISE	FINANCE	FINANCE	---
λ_{10} IN NOISE	AUTOMOBILE	AUTOMOBILE	COMMUNICATION