



関西学院大学  
KWANSEI GAKUIN UNIVERSITY

# Detecting stock market seasonality A period mining approach

Stephane Cheung Yukinobu Hamuro Katsuhiko Okada



関西学院大学  
KWANSEI GAKUIN UNIVERSITY

# Detecting stock market seasonality

## A period mining approach

Stephane Cheung\* / Yukinobu Hamuro\* / Katsuhiko Okada \* \*\*

\*Kwansei Gakuin University Business School

\*\*Magne-Max Capital Management, Japan FSA registered Investment Advisor





# Market seasonality

## Fact finding

- **Agrawal and Tandon (1994) JIMF**  
18 countries Seasonality, daily, weekend effect, last trading day of the month, large pre- and inter holiday return. Jan, return.
- **Bouman and Jacobsen (2002) AER**  
36 global markets out of 37 examined have “sell in May effect”, or “Halloween effect”, including Japan
- **Kamstra, Kramer and Levi (2003) AER**  
Markets in northern hemisphere demonstrate “sell in May” but not in southern hemisphere. SAD effect
- **Sakakibara, Yamasaki and Okada (2013) IRF**  
Japan is unique as Jun. is a good month while most other financial markets demonstrate lower June return -> “Dekansho-bushi” effect



## Market Seasonality

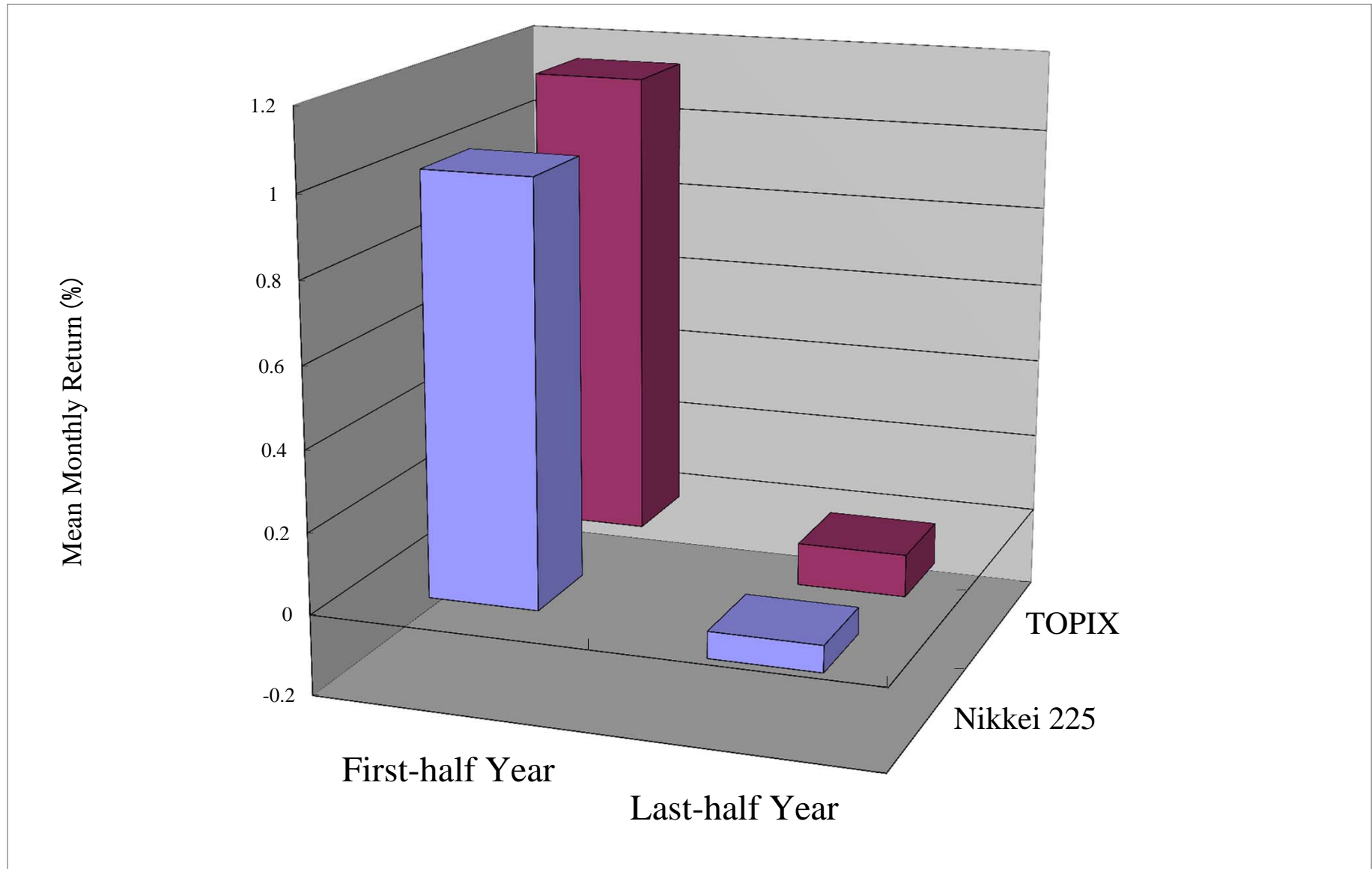
- Mark Twain famously observed that October is the most dangerous month to invest in stock market. ... *The tragedy of pudd'nhead Wilson. 1894.*
- The Stock Exchange world is in a sort of twilight state at the moment. The potential buyers seem to have “sold in May and gone away” ..... *Financial Times May,30 1964*
- ‘Sell in May and Go away’, famous Wall Street adage is once again in focus. Would the market behave as the old saying goes? Many investors in the market have anxiety in a corner of their minds.....*Nihon Keizai Shimbun, April 30, 2013*



	<b>316 years Jacobsen Global Fin. Data index</b>	<b>82 years Dow Jones index</b>	<b>50 years Dow Jones index</b>
	<b>1693-2009</b>	<b>1929-2011</b>	<b>1961-2011</b>
<b>Jan</b>	<b>0.69</b>	<b>1.0</b>	<b>1.2</b>
<b>Feb</b>	<b>0.09</b>	<b>0.0</b>	<b>0.0</b>
<b>Mar</b>	<b>-0.03</b>	<b>0.4</b>	<b>1.1</b>
<b>April</b>	<b>0.49</b>	<b>1.4</b>	<b>2.0</b>
<b>May</b>	<b>0.02</b>	<b>-0.2</b>	<b>-0.1</b>
<b>June</b>	<b>-0.12</b>	<b>0.5</b>	<b>-0.6</b>
<b>July</b>	<b>-0.31</b>	<b>1.5</b>	<b>0.9</b>
<b>Aug</b>	<b>0.44</b>	<b>0.8</b>	<b>0.2</b>
<b>Sep</b>	<b>-0.49</b>	<b>-1.3</b>	<b>-0.8</b>
<b>Oct</b>	<b>-0.5</b>	<b>0.0</b>	<b>0.5</b>
<b>Nov</b>	<b>0.35</b>	<b>0.8</b>	<b>1.2</b>
<b>Dec</b>	<b>0.81</b>	<b>1.5</b>	<b>1.5</b>
<b>Nov-April</b>	<b>2.42</b>	<b>5.2</b>	<b>7.2</b>
<b>May-Oct</b>	<b>-0.96</b>	<b>1.28</b>	<b>0.09</b>
<b>Diff</b>	<b>3.38</b>	<b>3.92</b>	<b>7.11</b>



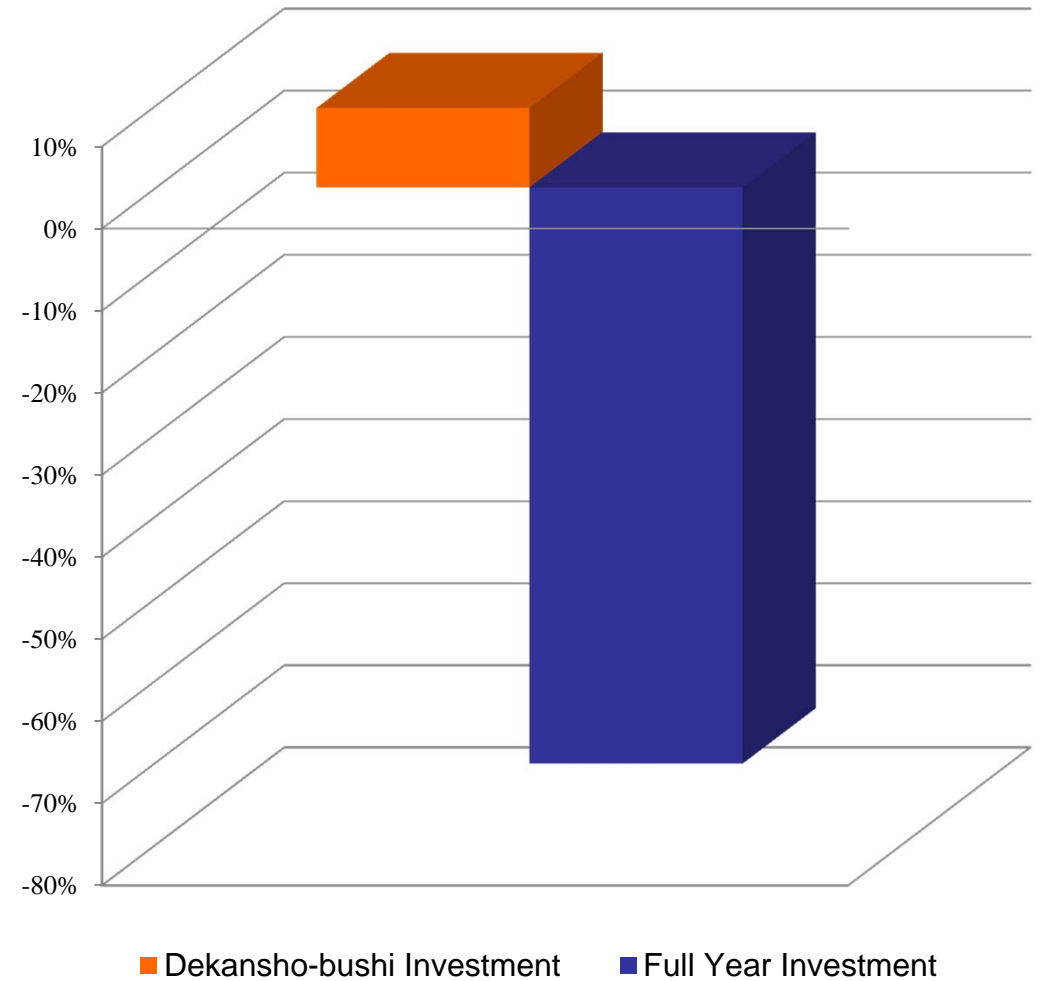
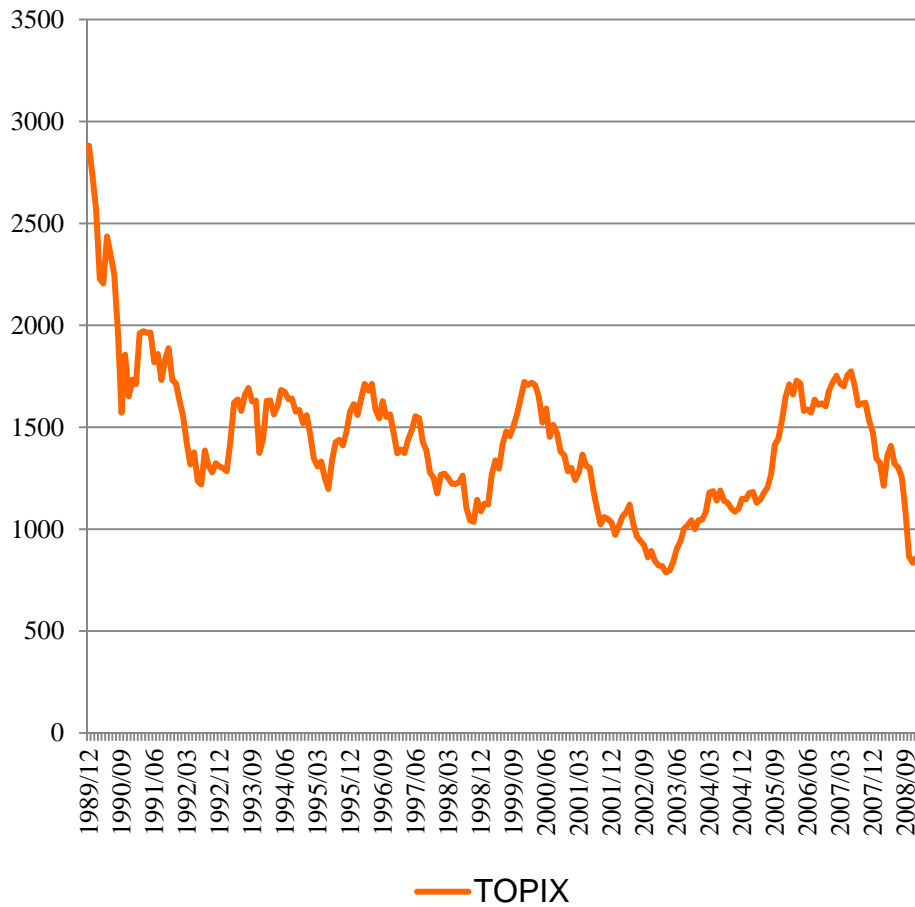
# Mean monthly return of Nikkei 225 and TOPIX Jan 1971-Dec 2014





If fund managers had followed the “words of wisdom”...

Japan's Lost 2 decades





## Research question and expectations

- ❑ Seasonality found in an index level implies that seasonality does exist in individual stock level, or industry level.
- ❑ The best season for holding stock  $i$  may be different from stock  $j$ .
- ❑ Finding seasonality in individual stock level enable us to create a portfolio durable for trading throughout the year.





## Data centric approach

- Traditional financial economist approach
  - a. Researchers come up with some insights
  - b. Create a model (Hypothesis building)
  - c. Collect data
  - d. Conduct an empirical test to prove or disprove the model (accept or reject the hypothesis)
  - e. Conclusion and conjecture
- Data centric approach
  - a. There is no model. We don't even have the hypothesis
  - b. Large scale data
  - c. Methodologies to detect correlation, potential predictability, to handle sparse data structure.
  - d. Pattern implies hypothesis



## The Period Mining Model

- Look for stocks that has high propensity to perform well at a given date. (Period Mining)
- Use previous **-4 to -1** years for training data
- Rolling window up to present
- Mining universe is TOPIX 500. Minimum market value of its composite is **130bil yen (\$1.2bil)**



# Methodology (Image)

Model Building Period



2001-2004

Portfolio formation based on the model created in 2001-4



2005

Model Building Period

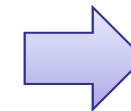


2002-2005

Rolling window up to present



2006

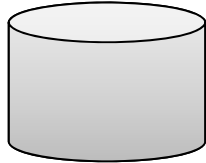


Portfolio simulation  
up to 2015.



# Period Mining: 4 steps in model building

Historical data of Stock Price  
for a stock (eg. Toyota)



- ✓ term: 125days starting from Jan 1<sup>st</sup> or July 1<sup>st</sup>
- ✓ 28 terms from 2001 to 2014

## 1. Enumeration Phase

Enumerating statistics (eg. abnormal return) of all periods defined by combination of starting date and holding period for each term.

(We call this period as "item")

- ✓ starting date: 1,2,...,125
- ✓ holding period: 1,2,...,125
- ✓  $125 \times 125 = 15,625$  items per stock per term

## 2. Aggregation Phase

Aggregating the statistics on each item for last  $n$  years ( $n = 4$  in this exp.)

## 3. Filtering Phase

Filtering items by aggregated statistics which match the given conditions (eg. lower bound of duration days).

## 4. Organizing Phase

Select items so that there is no overlap and maximize the total abnormal return.

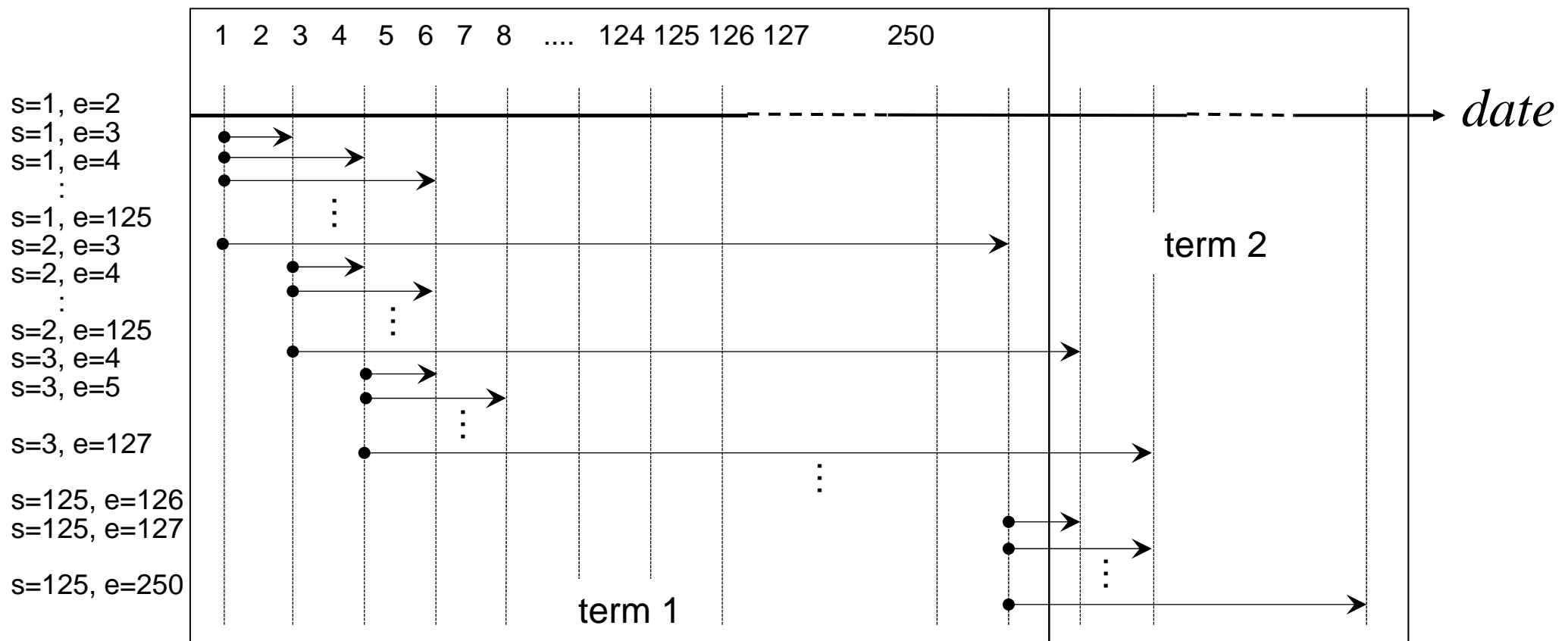
Selected Items (Periods)  
to hold



# Enumeration Phase

Enumerate stock price statistics on period  $p_{s,e}$

- ✓ starting date: 1,2,...,125
- ✓ holding period: 1,2,...,125
- ✓  $125 \times 125 \times 500$  equities  $\hat{=}$  **7.8M items per term**

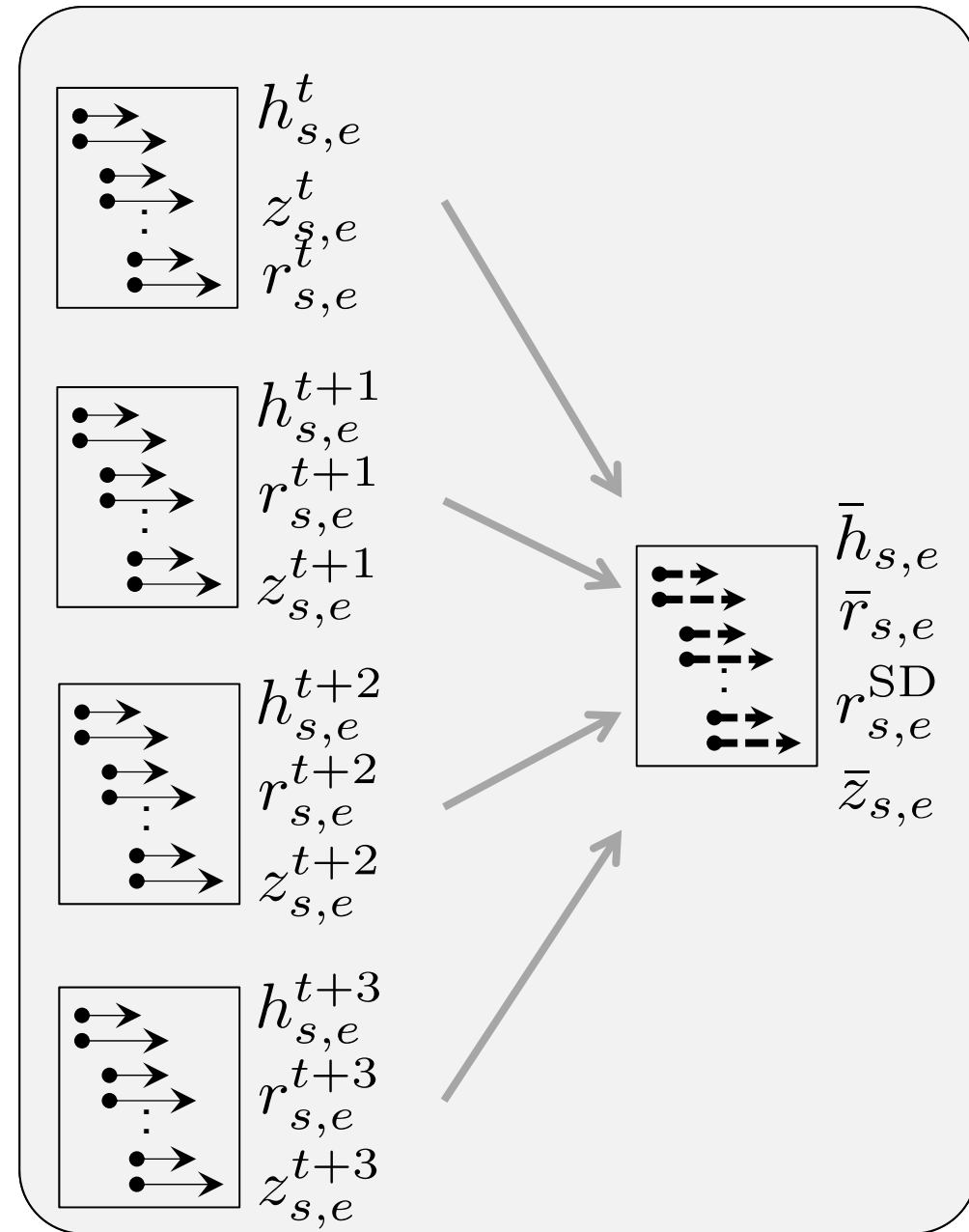




# Aggregation Phase

Aggregate all the items based on the following four criteria

- a) Holding period
- b) Average of abnormal return
- c) SD of abnormal return
- d) Average of zigzag rate





# Filtering Phase

Select periods matching the given 6 conditions.

- ✓ The optimal parameters will be estimated using machine learning technique (Bayesian Global Optimization).

declare

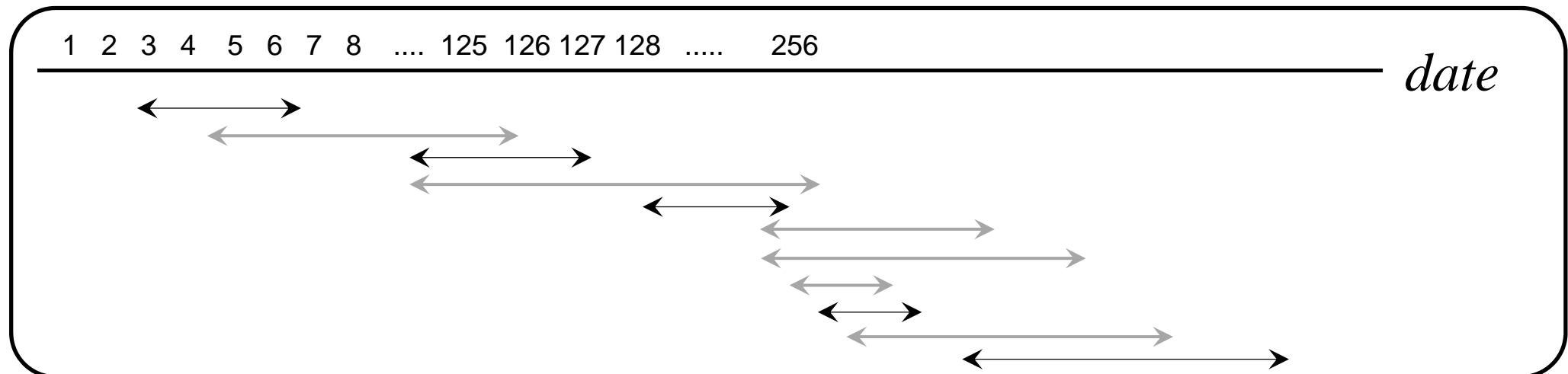
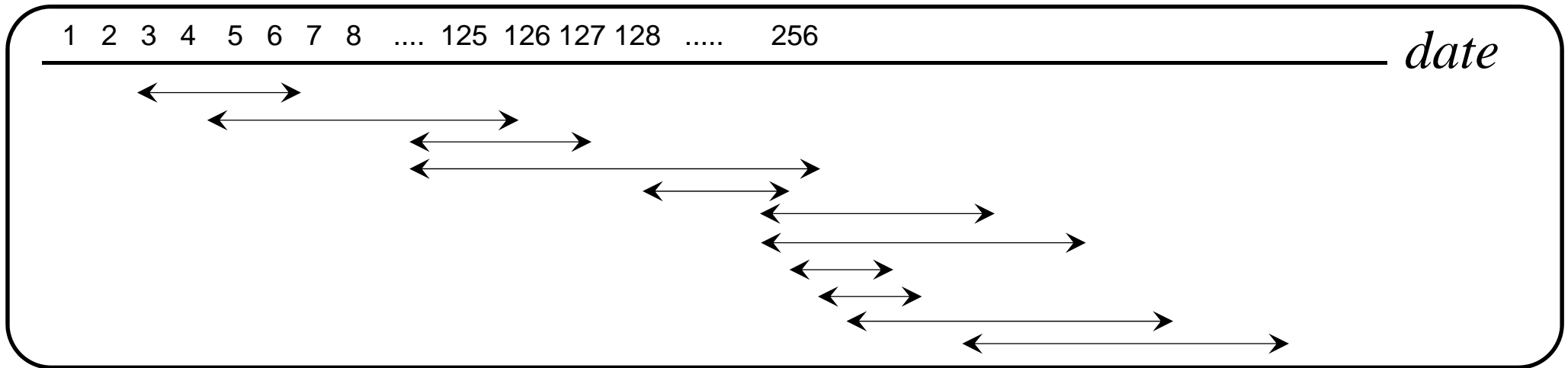
notation		min	max	Rationale
$lh$	lower bound for holding period	0	20	Lower bound: Too short holding period is not favorable due to transactions cost. Upper bound :Index level seasonality is 6 month
$uh$	upper bound for holding period	10	125	
$lr$	lower bound for average of abnormal return	0.01	0.1	Abnormal return is higher the better. but extremely high abnormal return may be due to other reasons than seasonality.
$ur$	upper bound for average of abnormal return	0.11	2.1	
$us$	upper bound for SD of abnormal return	0.01	1.0	Prefer items that generate stable abnormal return in the four year training period
$lz$	lower bound for zigzag rate	0	1	It is preferable if average trend of stock price movement is monotonously increase in the holding period.



# Organizing Phase

Objective : maximize the sum of all abnormal return

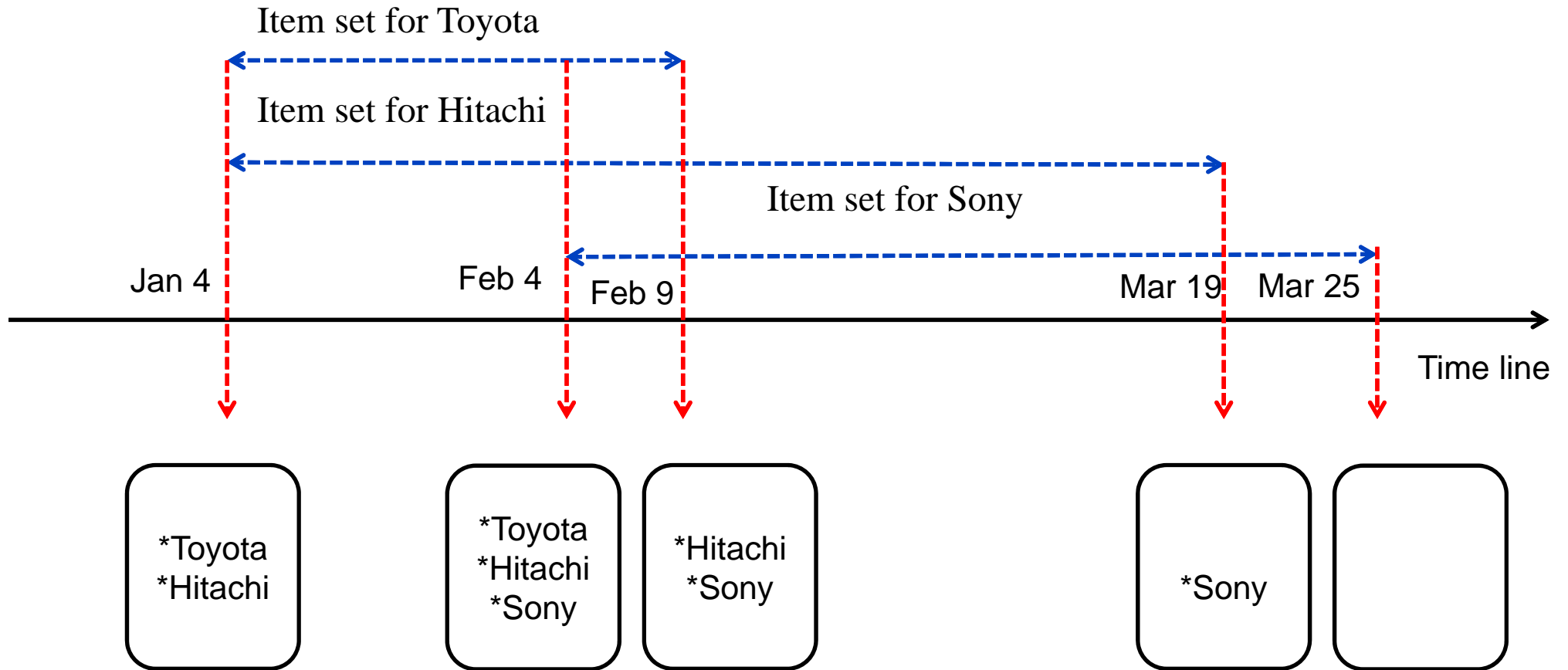
Subject to : selected periods are not overlapped each other







# Creating calendar time portfolio





# Results



## Performance summary

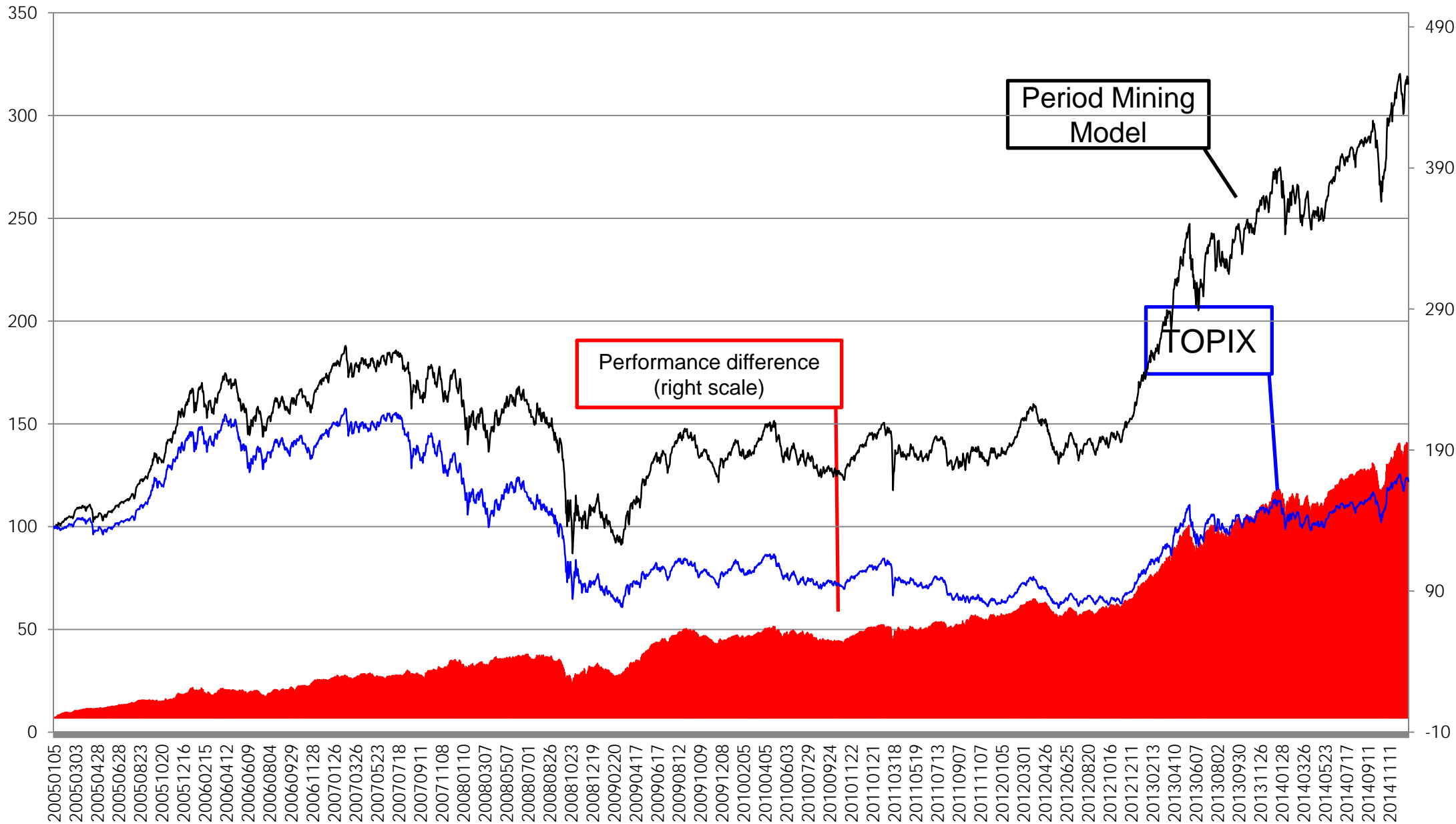
2005年1月—2014年12月

	Period Mining Model EW	Period Mining Model VW	TOPIX	TOPIX Mid 400
<b>Annualized daily return</b>	<b>14.32%</b>	<b>9.25%</b>	<b>4.67%</b>	<b>5.65%</b>
<b>Annualized daily risk</b>	<b>22.80%</b>	<b>23.47%</b>	<b>22.95%</b>	<b>21.99%</b>
<b>Maximum daily gain</b>	<b>12.60%</b>	<b>12.39%</b>	<b>13.73%</b>	<b>12.87%</b>
<b>Maximum daily drawdown</b>	<b>-10.51%</b>	<b>-9.20%</b>	<b>-9.52%</b>	<b>-10.62%</b>
<b>Sharpe ratio</b>	<b>0.628</b>	<b>0.394</b>	<b>0.204</b>	<b>0.257</b>



# Calendar time portfolio performance since inception.

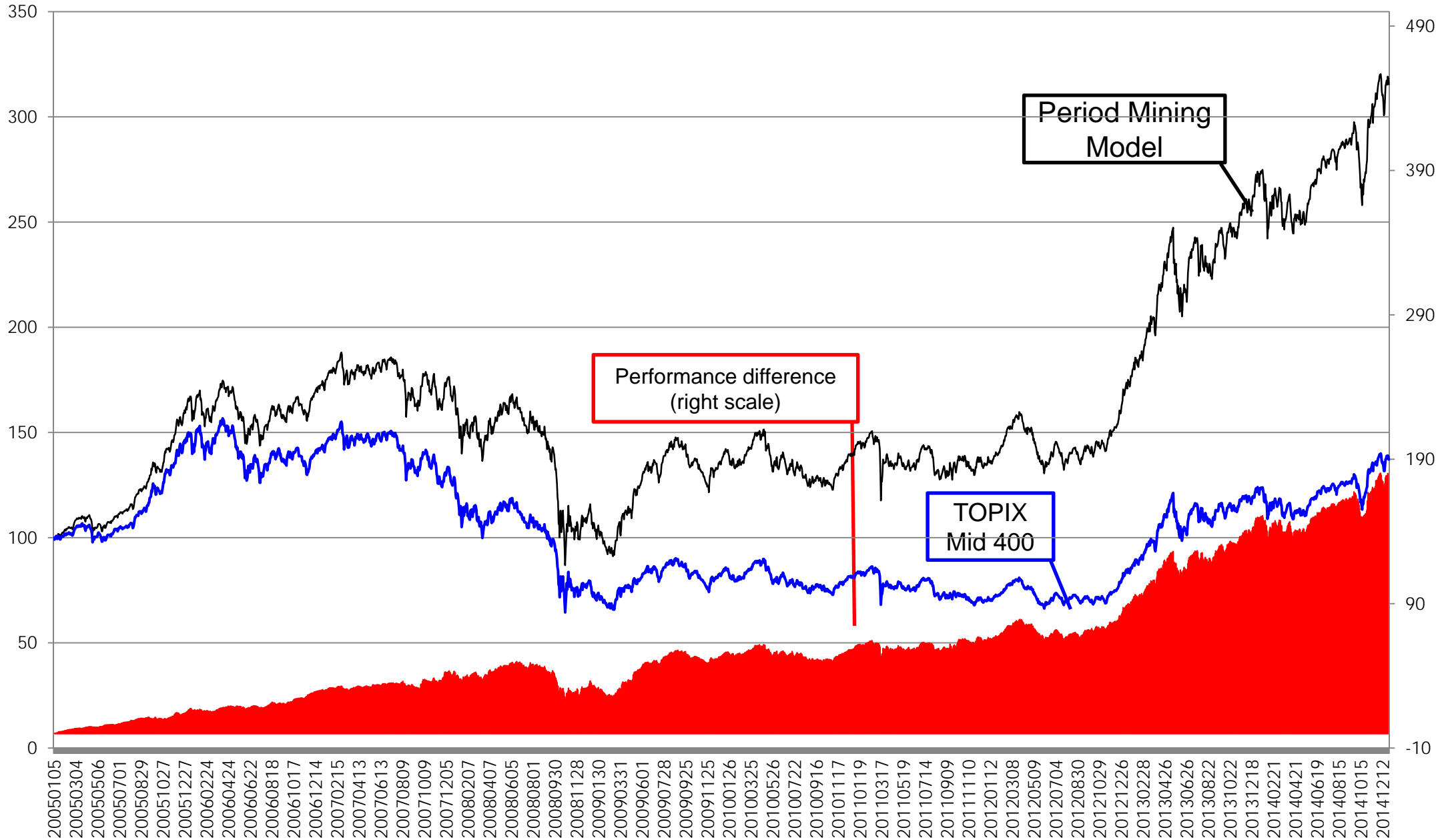
Benchmark index: **TOPIX** , Initial NAV:100, 2005-2014





# Calendar time portfolio performance since inception.

Benchmark index: **TOPIX Mid 400**, Initial NAV:100, 2005-2014





# Controlling for “size” and “book-to-market ratio”

Three factor model

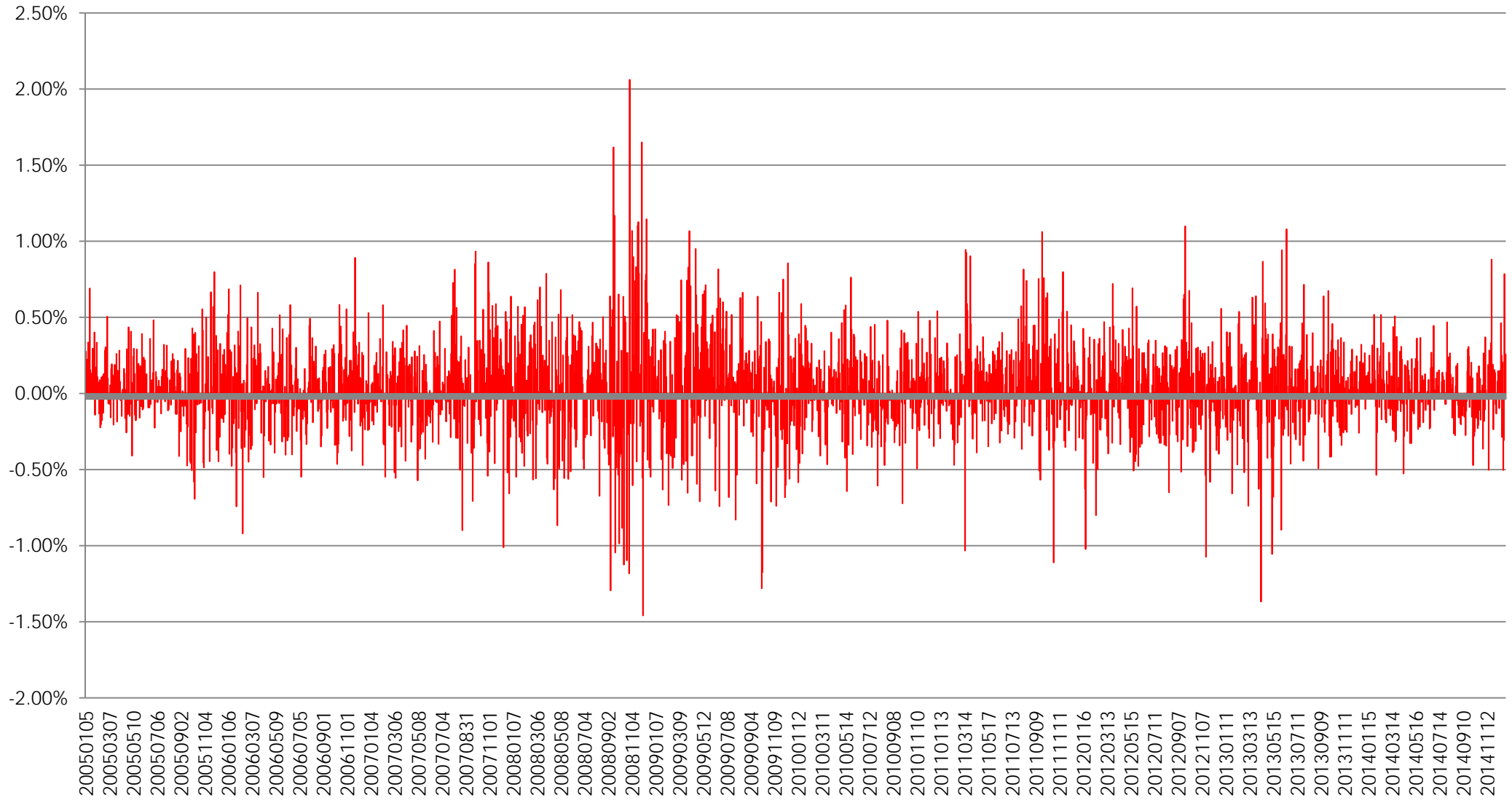
$$R_{p,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_iSMB_t + h_iHML + \varepsilon_{i,t}$$

	Coefficient	Standard error	t-value	p-value
Intercept	0.00026	5.675E-05	4.63	0.00000
Rm-Rf	0.01061	4.674E-05	226.92	0.00000
SMB	0.00131	1.009E-04	12.998	0.00000
HML	0.00169	1.492E-04	7.834	0.00000

**Annualized alpha 6.50%**

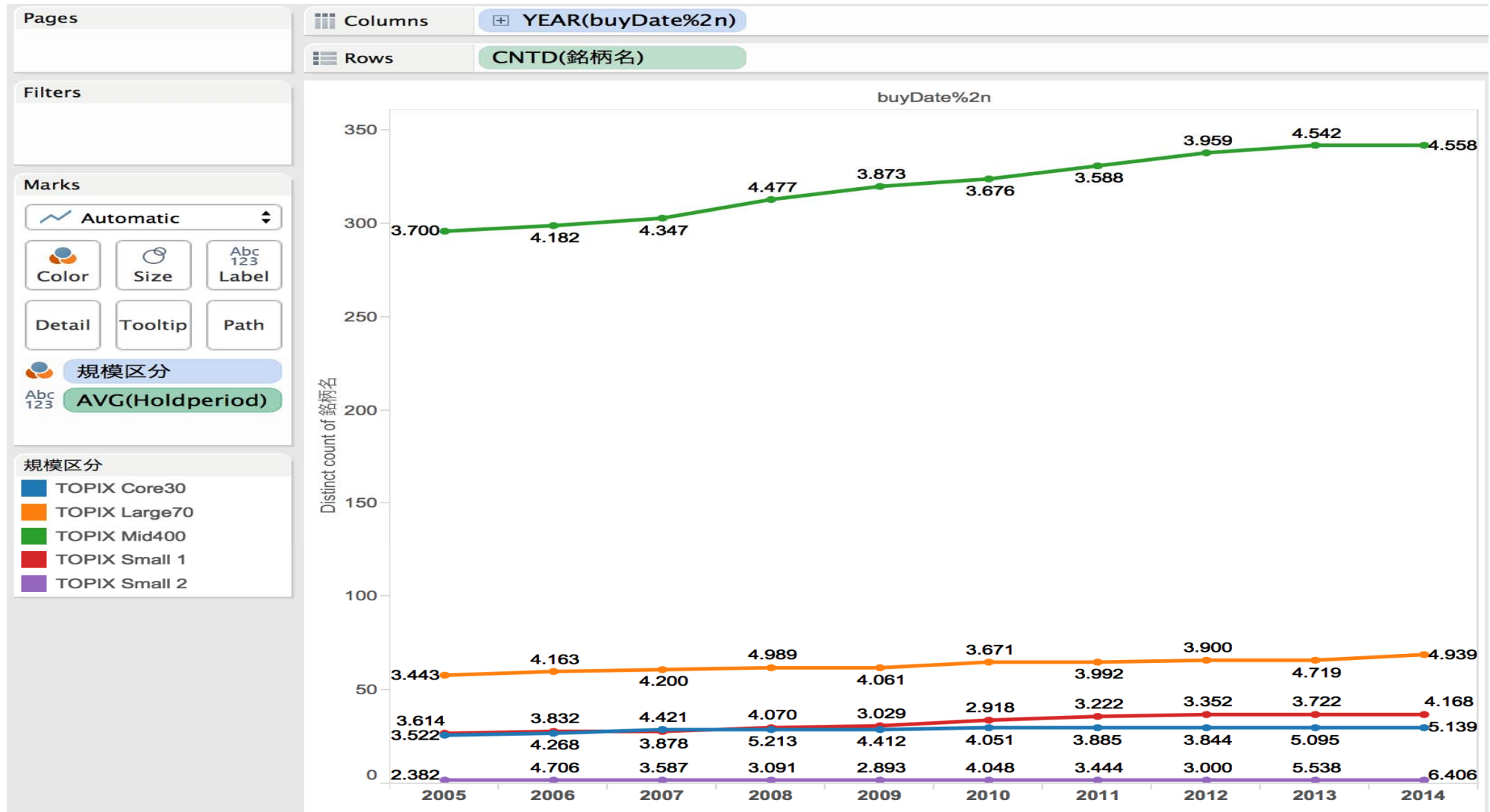


# The daily performance against the benchmark





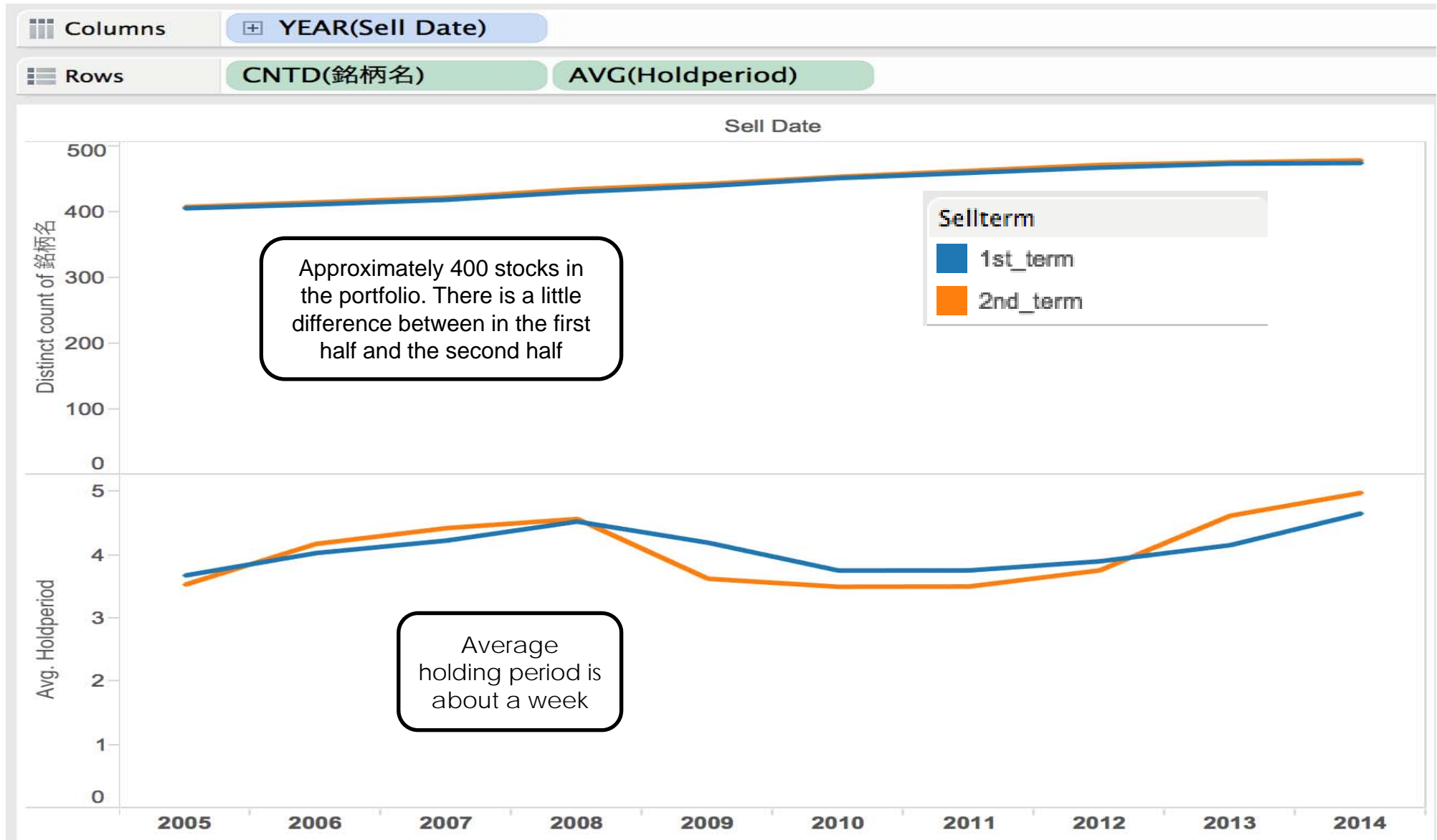
# Period Mining Portfolio, Composite of stocks and holding period





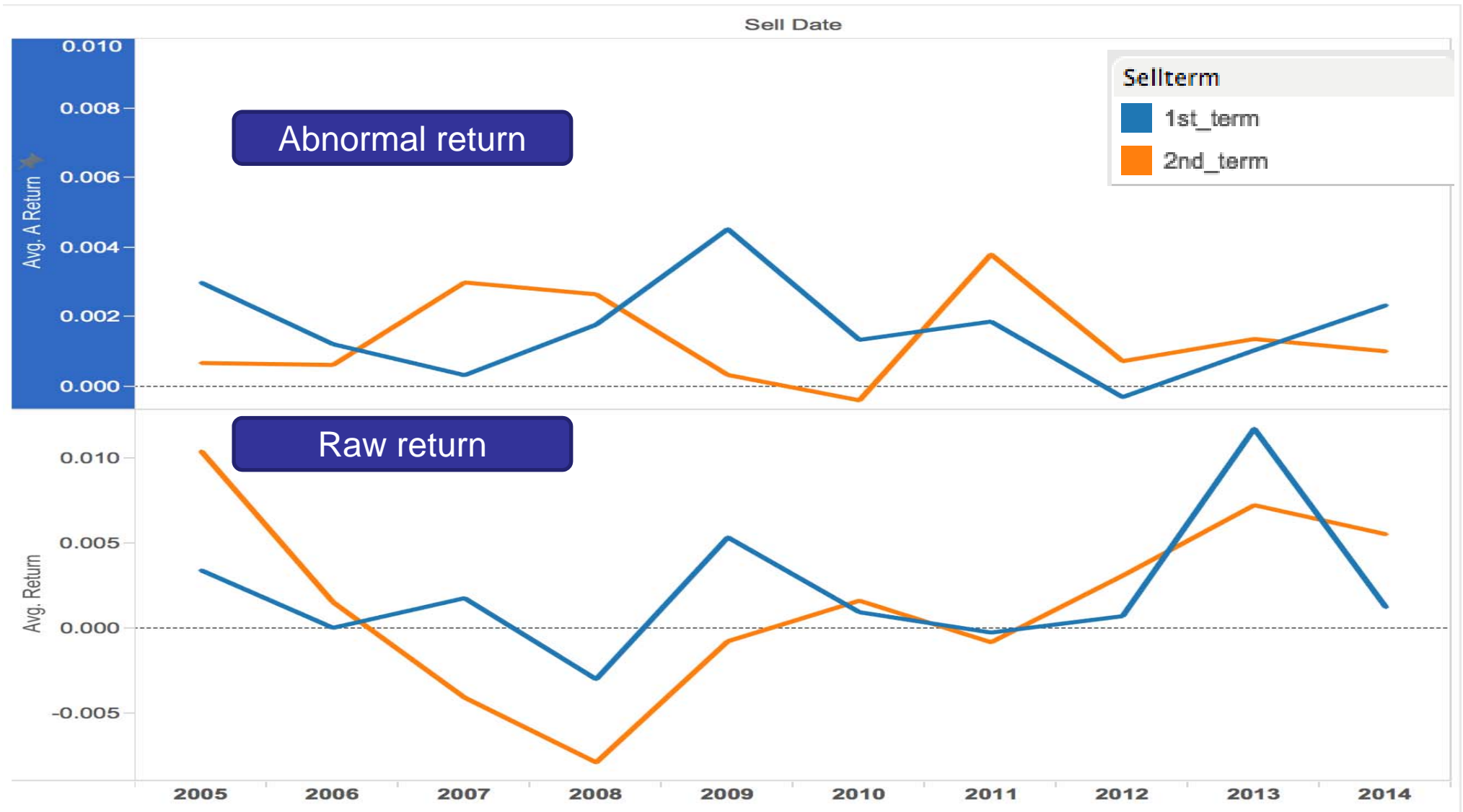


# Number of shares in the portfolio and average holding period





# Returns are not so concentrated in the first half.





# Are we holding more stocks in the earlier month? Maybe...

Columns: MONTH(buyDate..)

Rows: DAY(buyDate%0)

buyDate%0

Day of buyD..	January	February	March	April	May	June	July	August	September	October	November	December
4	206.3	39.1	40.3	43.1		47.0	12.3	40.9	38.1	39.0	39.0	36.0
5	95.1	33.8	45.0	52.4		42.3	100.6	49.4	44.1	36.1	42.3	49.6
6	46.0	27.9	43.0	42.0	49.0	39.9	109.0	52.3	55.0	39.0	53.1	37.3
7	40.0	28.5	60.4	41.6	36.4	42.9	86.3	47.0	54.6	50.0	53.1	46.9
8	43.6	33.4	50.3	43.1	53.7	33.4	28.7	49.0	60.3	34.6	34.4	50.6
9	40.8	41.3	65.9	40.1	45.4	42.3	37.6	39.7	39.0	41.5	57.6	33.0
10	39.0	48.9	52.4	33.6	44.3	42.0	36.0	42.1	33.3	35.2	48.7	34.6
11	43.7		56.3	35.6	36.9	39.7	40.0	52.6	40.1	32.3	36.1	44.7
12	53.5	41.8	46.3	33.9	29.7	38.7	36.1	35.3	48.0	81.0	37.7	39.8
13	34.2	42.3	39.7	38.6	44.0	35.8	44.1	43.6	36.6	48.0	40.1	33.3
14	50.4	43.3	29.8	44.1	44.3	37.7	36.9	30.0	50.4	84.3	42.9	48.1
15	35.3	32.4	50.7	41.4	36.7	37.4	41.7	37.1	47.0	59.1	30.0	35.0
16	36.6	40.9	39.0	40.1	42.6	42.3	37.8	37.4	26.2	28.6	60.0	28.9
17	45.4	27.7	46.9	31.9	47.6	30.6	30.7	37.1	34.4	44.9	57.1	33.9
18	48.0	28.9	49.9	37.9	43.3	45.7	39.3	34.7	41.0	45.6	40.6	32.7
19	35.3	34.7	43.6	28.6	37.6	43.6	31.2	47.9	44.5	42.7	41.9	28.3
20	30.6	47.0	39.3	31.3	43.0	43.3	39.2	33.0	60.5	44.7	36.3	35.0
21	42.0	48.5	33.3	42.7	40.0	36.7	33.6	31.3	38.2	38.7	38.0	32.0
22	42.2	45.8	35.7	40.7	34.4	40.9	41.6	37.3	32.2	41.0	49.0	35.3
23	34.4	46.1	33.4	45.3	35.1	43.6	38.1	34.9		27.7		
24	43.5	61.3	34.3	36.1	35.9	40.6	39.6	40.4	38.2	33.4	55.2	24.2
25	42.8	46.3	50.0	42.3	45.3	43.7	37.8	36.4	34.3	56.7	37.1	35.0
26	49.0	41.8	34.0	35.7	36.9	32.0	42.1	37.1	61.9	47.3	36.1	33.3
27	46.9	34.6	47.7	34.0	33.4	31.1	40.4	35.7	43.7	52.0	34.0	59.0
28	49.7	36.4	34.4	42.9	40.1	34.4	39.9	42.4	41.7	62.1	42.1	73.6
29	48.3	45.5	37.0		24.7	32.6	39.9	41.5	50.9	45.7	47.1	78.5
30	47.4		37.9	42.0	35.6	25.4	49.4	31.7	29.3	49.6	33.7	95.2
31	37.4		43.0		49.1		29.3	35.4		46.0		



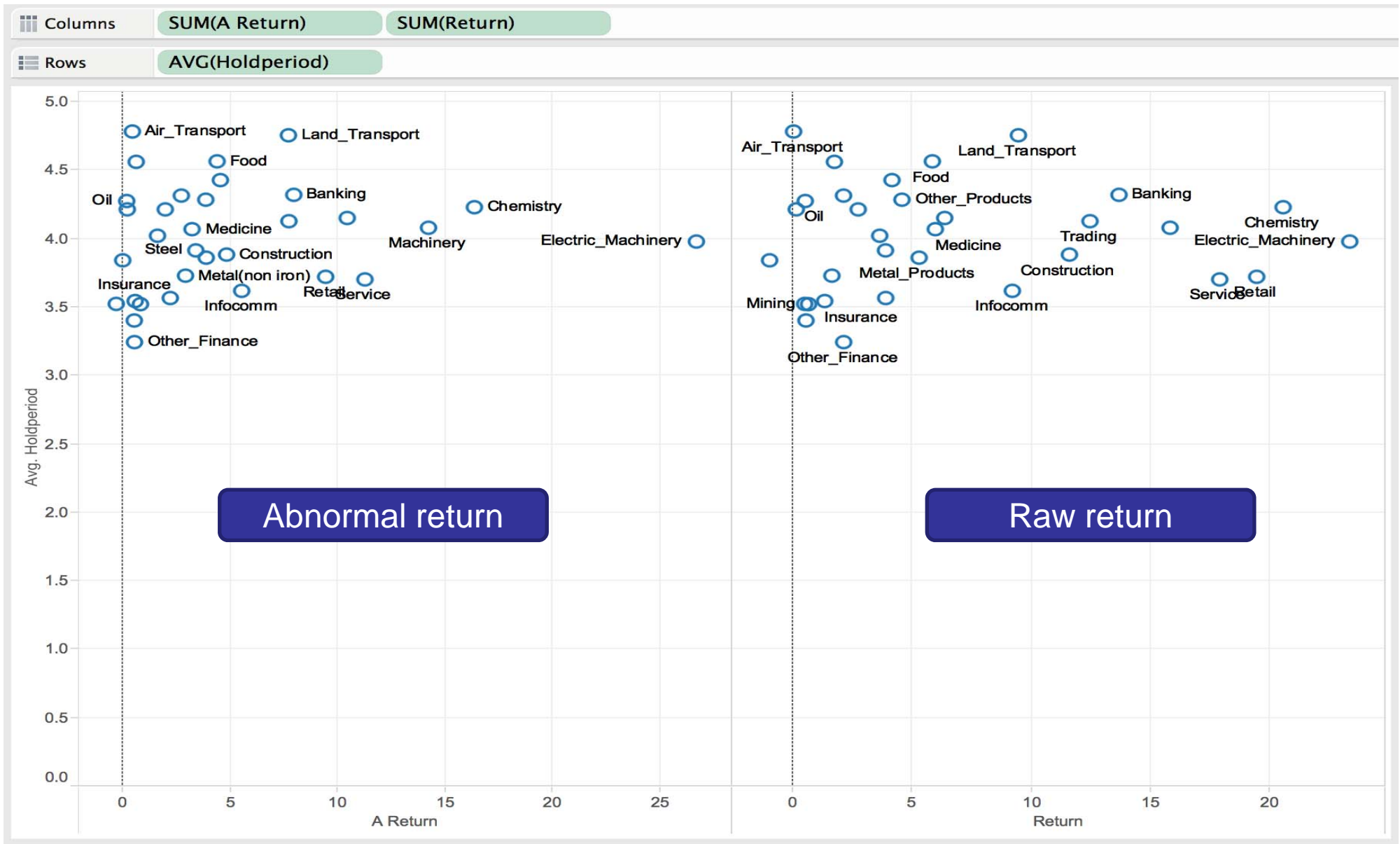
# Where does the profit (abnormal return) come from?

Columns Sellterm 規模区分										
Rows 33Industries										
33Industries	1st_term					2nd_term				
	TOPIX C..	TOPIX L..	TOPIX M..	TOPIX S..	TOPIX S..	TOPIX C..	TOPIX L..	TOPIX M..	TOPIX S..	TOPIX S..
Air_Transport		0.36					0.02			
Banking	1.12	-0.53	2.15			-0.88	0.28	5.75		
Chemistry	0.76	1.26	3.00	2.40		0.42	3.15	4.59	0.73	
Construction		0.66	2.34	-0.44			1.50	0.30	0.40	
Electric_Machinery	0.10	2.55	13.55	1.09		0.65	1.01	7.74	-0.03	
Fiber_Products		0.01	0.10				0.14	0.30		
Food	0.82	-0.73	0.96			0.15	-0.11	3.23		
GlassStone_Products			1.39					0.16		
Infocomm	-0.57		0.24	0.82		1.80		1.39	1.78	
Insurance	0.77	-0.21	0.12			0.05	-0.08	-0.15		
Land_Transport	-0.12	0.52	2.96	0.57		0.23	0.65	2.69	0.16	
Machinery	0.21	0.21	8.75	0.83		0.43	1.61	1.88	0.24	
Medicine	-0.10	-0.26	0.27	-0.48		0.72	0.48	2.36	0.18	
Metal_Products			2.72					1.09		
Metal(non iron)		0.81	0.63	0.24			0.09	0.82	0.27	
Mining		0.21	0.76				0.18	-0.39		
Oil			0.01				-0.13	0.23		
Other_Finance		0.28	1.80	0.44	1.11		-0.41	-2.66	0.25	-0.32
Other_Products		0.51	1.67				0.68	0.93		
Power_Gas		-0.48	-0.26				1.77	3.45		
Precision_Machinery		0.18	2.58	0.97			0.60	-1.49	-0.19	
Pulp_Paper			-0.45					0.92		
Real_Estate	0.79	0.36	1.02	-0.38		0.51	0.26	-0.13	-0.29	
Retail	0.07	-0.44	2.06	0.63	0.71	0.01	1.03	3.57	1.33	0.41
Rubber_Products		0.01	0.37				0.44	1.08		
Sea_Transport			0.64					-0.71		
Securities_Commodities	0.36	-0.11	-1.57			0.05	0.28	1.13		
Service		0.13	3.45	3.05			0.29	4.52	-0.23	
Steel	-0.25	0.41	0.78			0.20	0.02	2.16		
Trading	1.06	0.65	2.97	0.08		0.22	0.70	1.43	0.56	
Transport_Machinery	0.73	2.18	4.72	1.23		0.30	-0.20	1.56	-0.12	
Warehouse_Logistics			-0.51	-0.32				0.25	0.21	





# Longer the holding period, better the return? Not really.





## How long will the strategy work?

- ❑ Pairs trading (Relative value arbitrage strategy) demonstrate “profitability” beyond various risk measures, short-sale constraints and transaction costs. *Gatev et.al. RFS, 2004*
- ❑ However, the profitability has been waned as more and more hedge funds employ similar “pairs trading” strategies.
- ❑ Seasonality trading is an unexplored approach; chances of enjoying hefty profit could potentially be high.



## Conclusion

- Seasonal anomaly in the stock market has been documented in the literature for quite a while, and yet, the anomaly hasn't been arbitrated away by professionals even today.
- The difficulty lies in the potential arbitrage profit is only guaranteed in the years, not in the months.
- We endeavored to look for patterns of seasonal investor behavior in large 500 firms listed in TSE 1<sup>st</sup>.
- To detect the pattern, we used period mining technique and other related techniques commonly used in the computational science.
- Portfolio of stocks in their best season of the year outperform the benchmark index by a substantial margin.
- This “seasonal arbitrage” will remain to be profitable as few participants are playing in this market.