

How can you short-sell the 'Tear of Ant?'

Short-Sell Strategy Using NDR(Naver Stock Discussion Room) Data

BFBH GBA3043

Team JTH

Executive Summary

The number of daily NDR post is proxy that shows average investors' overreaction to a news / earnings announcement / etc.

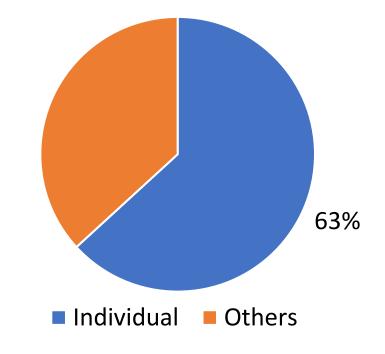
When there is **good news** on a stock and the Stock's **NDR** is **boomed** on the event day (above STDEV3), take short position on it for 20 trading days.

On Average, the CAR will drop about 4% for next 20 days.

Individual Investors in Korea

Individual investors has been major player on Korean Market. But sadly, they keep losing money.

Trading Volume by Investors



10년 주식투자 수익... 외국인 78%, 개인 -74%

이진석 기자 김지섭 기자

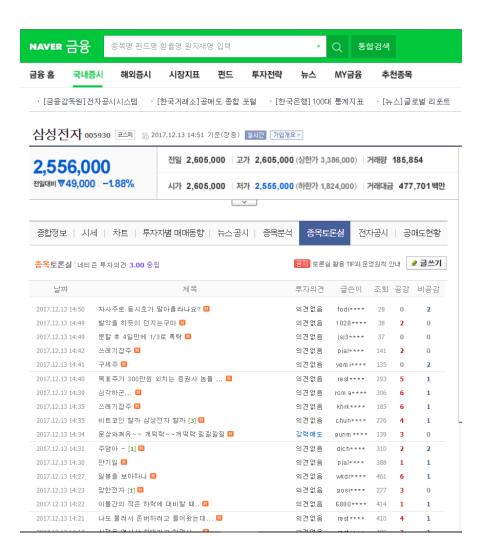
입력: 2017.03.06 19:00



주식 시장에서 개미(개인 투자자)들과 외국인, 기관 투자자들은 치열한 수익률 전쟁을 벌인다. 개미들의 성적표는 어떨까? 개인 투자자들이 선호하는 30개 종목을 골라서 지난 10년간 지속적으로 투자한 경우를 가정한 시뮬레이션(가상 실험)을 해보니 수익률이 -74%로 나타났다. 증시는 '개미지옥'이라는 말이 나올 만하다.

반면 같은 기간 같은 방식으로 계산한 외국인 투자자들의 수익률은 78%였다. 연기금·펀드 등 기관 투자가들이 주로 거래한 30개 종목의 수익률은 9%로 집계됐다.

What is NDR?



Naver Stock Discussion Room

This is subsection of 'Naver Finance', the online community where investors share their opinion in constructive way.

However, the reality is...

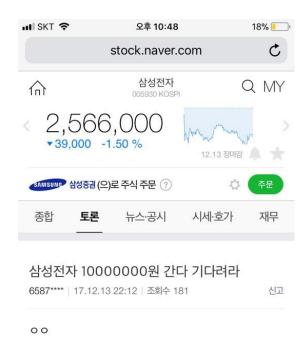
Most of NDR posts are from individual investors,

and the contents are groundless information and optimism.

(even information phishing is going on there.)

You can easily access it through finance.naver.com

Here are some examples.



SE will hit W10million Just hold it.



I told you guys that this will hit W200,000? Now Target Price is W300,000. Let's go for it guys.



Shut up and Hold. If u don't hold this, you are a dumbass.

I told you... it's full of groundless opinion and optimism.

And when there is a special event to a listed company, the stock's NDR go crazy.

Ex) Jayjun Cosmetic – Maskpack export company (90% sales from china)



13th November, Jayjun Maskpack hit jackpot on Chinese black Friday.

고품질 마스크팩, 사드 해빙타고 중국 소비자를 사로잡았다!



Posted on 11/12/2017 by 헤럴드경제 in 뉴스, 라이프, 메인 헤드라인, 뷰티&패션, 사회 with 0 Comments

-제이준코스메틱, 中 광군제 하루 만에 매출 156억 기록

[헤럴드경제=이진용 기자]국내 마스크팩 전문 브랜드 제이준코스메틱(대표 이진형, 판나)은 중국판 블랙 프라이데이로 불리는 광군제에서 최대 매출 실적을 달성했다.

제이준코스메틱은 광군제 기간 동안 티몰(Tmall), 타오바오, VIP 등 중국 전체 온라인 플랫폼에서 인텐시브 샤이닝 마스크, 블랙 물광마스크, 안티 더스트 화이트닝 마스크 등의 대표 제품을 판매하며, 약 156억원의 매출을 기록했다고 전했다.

작년 매출 기준, 월 평균 매출(약 150억원)액에 해당하는 거래가 단 하루 만에 온라인에서 일어난 것이다.

What happened to Jayjun's NDR that day?

There were more than 200 posts. (Average daily posts: 30)

<mark>종목</mark> 토론실 네티즌	<mark>종목</mark> 토론실 네티즌 투자의견 3.00 중립 공제 토론실 활용 TIP과 운영원칙 안내 				<mark>종목</mark> 토론실 네티	즌 투자의견 3.00 중립	공지 토론실	활용 TIP과 운	영원칙 인	tut d	₹ 글쓰기		
날짜	제목	투자의견	글쓴이	조회	공감	비공감	날짜	제목	투자의견	글쓴이	조회	공감	비공감
2017.11.13 12:46	8000원몸 산다	의견없음	znfk****	352	2	0	2017.10.05 09:50	북한 미사일 움직임 확인함. [1]	의견없음	ohso***	2046	4	10
2017.11.13 12:46	좋구나 좋아 ~~~~~~	의견없음	thdc****	275	1	0	2017.10.03 18:39	대박이지요 [1]	의견없음	joll****	3042	3	10
2017.11.13 12:45	분봉이 죽어가네	의견없음	vnfj* * * *	335	1	0	2017.10.03 00:18	고점 8천원에 물려서 ㅜㅜㅜ [2]	의견없음	siho****	2519	4	17
2017.11.13 12:43	단기목표가가어찌되심들? [3]	의견없음	kkmm****	590	2	1	2017.10.02 23:33	어쩐지 느낌이 그렇터라	의견없음	seiz****	1988	3	0
2017.11.13 12:48	┗ 단기목표가가어찌되심들?	의견없음	narn****	202	0	0	2017.10.02 17:34	거래소패쇄시켜라	의견없음	jung****	1443	1	5
2017.11.13 12:43	9천원가겠다	의견없음	dohy* * * *	425	4	0	2017.10.02 15:35	장이 안열리니까	의견없음	ckrg****	1424	9	2
2017.11.13 12:42	문상왔어요	의견없음	pjh7****	336	2	1	2017.10.02 09:34	공매도	의견없음	mesa****	1419	6	1
2017.11.13 12:42	이제 시세분출시작입니다	의견없음	tiaj* * * *	273	1	0	2017.10.01 19:14	상한가 10방 [2]	의견없음	dmfv* * * *	2492	3	4
2017.11.13 12:42	훨훨 나라라~~	의견없음	byc9****	194	1	0	2017.10.01 10:03	낼 주식 개장 안하죠? [1]	의견없음	Zons****	1397	0	5
2017.11.13 12:42	무섭게 양매수네	의견없음	SSPY****	242	1	0	2017.09.30 15:16	이주식 먼가여 [2]	의견없음	kore* * * *	2644	1	18
2017.11.13 12:42	멸치 개쓰 꺼져 라 제발	의견없음	cmj1****	180	1	0	2017.09.30 13:49	소액주주들을 살려주세요. [3]	의견없음	jung****	1830	4	3
2017.11.13 12:41	멸치빠지고 나이스 ㅋㅋ	의견없음	kdhk****	248	2	2	2017.09.29 19:45	ヲキ~~OトOト~~~~ [4]	의견없음	ffsk* * * *	2573	1	4
2017.11.13 12:39	오케이 멸치다빠졌고	의견없음	sung****	293	4	0	2017.09.29 18:48	다들 즐거운 추석 보내세요	의견없음	wdog***	825	2	0
2017.11.13 12:39	아~십알 멸치새끼	의견없음	yes5****	278	2	1	2017.09.29 17:37	시간외랑종가랑왜케오르죠? [5]	의견없음	yoyo****	2211	0	6
2017.11.13 12:39	멸치	의견없음	ktwa* * * *	211	1	1	2017.09.29 15:58	꽉 잡아라	의견없음	SUCC****	1569	4	3
2017.11.13 12:39	오늘 개미 개털림 [1]	의견없음	ldk5***	377	1	0	2017.09.29 15:37	오늘의 매동 [7]	의견없음	m ylo* * * *	2147	28	0
2017.11.13 12:36	후회된다 ㅅㅂ [1]	의견없음	znfk****	438	1	0	2017.09.29 14:53	ㅇㅏ 흔들기에 당할 내가 아니다 [2]	의견없음	shdj****	984	1	1
2017.11.13 12:35	엄청 팔아대네ㅠ	의견없음	∨nfj* * * *	420	1	1	2017.09.29 14:48	또 당하네!	의견없음	cass****	840	1	1
2017.11.13 12:35	에프앤니퍼블릭으로 와라 모회사다 [1]	의견없음	oseo****	369	1	1	2017.09.29 14:45	거봐외국인들 단타라니깐ㅋㅋ [3]	의견없음	rlag****	835	0	1
2017.11.13 12:34	너무 급히 올라가서 현기증 난다	의견없음	jbj2****	287	3	1	2017.09.29 13:47	내 맘이 급하다 ~ 빨리 쫌 칠천가자![2]	의견없음	obg5****	653	0	1

Our Assumption - Overreaction

If the number of NDR post is more than STDEV3, It is signal that market is overreacting to the news.

So on good news, CAR will skyrocket and slowly be normalized.

So on bad news, CAR will plummet and slowly be normalized.

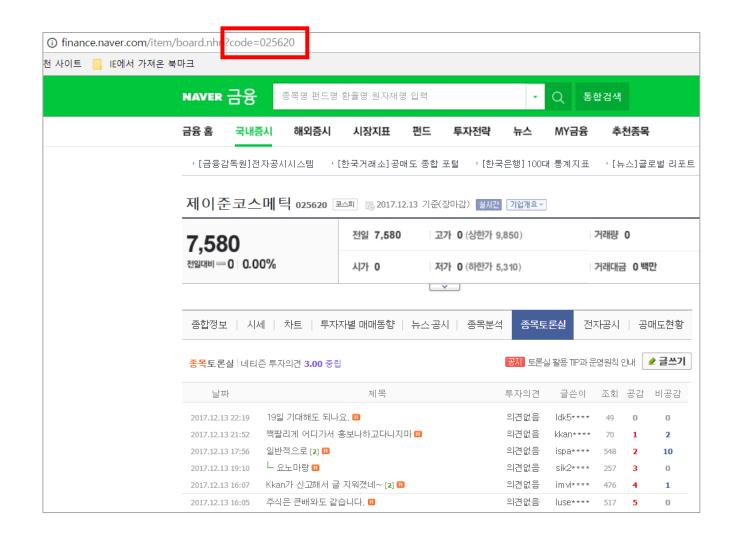
We can find out arbitrage chance in this phenomenon.

1. Ticker Symbols(Kospi)



	Α
	, , , , , , , , , , , , , , , , , , ,
1	271980
2	035720
3	271560
4	268280
5	267290
6	251270
7	079440
8	267270
9	267250
10	267260
11	264900
12	145720
13	111110
14	143210
15	241560
16	207940
17	234080
18	241590

2. NDR posts(Date)



2. NDR posts(Date)

```
def Crawling(code):
   url = "http://finance.naver.com/item/board.nhn?code={0}&page=1".format(code)
   HEADERS={'user-agent':('Mozilla/5.0 (Windows NT 6.1)' 'AppleWebKit/537.36 (KHTML, like
   r = requests.get(url, headers=HEADERS)
   soup = bs(r.text, 'html.parser').select("tbody")[2]
       end_page=int(re.search('page=(.*?)"',str(soup.select('.pgRR'))).group(1))
   except:
       end_page=1
   resultList=[]
   for i in range(1,end page+1):
       url = "http://finance.naver.com/item/board.nhn?code={0}&page={1}".format(code, i)
       r = requests.get(url)
       soup = bs(r.text,'html.parser')
       search_table = soup.findAll("span", class_="gray03")
       list1=[date.text for date in search_table]
       list2=[s for s in list1 if len(s) > 10]
       list3=[I[:10].replace('.','-') for I in list2]
       for u in list3:
           resultList.append(u)
       counts=dict(Counter(resultList))
       data=pd.Series(counts,name=code)
       data2=data.to_frame()
       print(i)
   return data2
```

```
kospi=pd.read_excel('kospiCode.xlsx',header=None, dtype=object)
kospiCode=kospi[0].tolist()

# 크롤링 함수 실행코드 (한 번에 돌리면 오류 발생 확률이 매우 높으
appended_data=[]
for i in kospiCode:
    data=Crawling(i)
    appended_data.append(data)
    print("종목코드 "+i+" 리퀘스트완료...")
appended_data=pd.concat(appended_data, axis=1)
print("표 통합완료")

appended_data = appended_data.fillna(0)
appended_data.to_excel('0전환.xlsx')

df = pd.read_excel('0전환.xlsx')
```

2. NDR posts(Date)

	051900	051910	047040	047050	042660	042670	051630	049770	034020
2016-10-30	0	0	0	0	0	0	0	0	0
2016-10-31	25	14	4	2	20	3	1	1	6
2016-11-01	11	1 5	9	12	26	18	0	1	7
2016-11-02	9	12	4	10	21	26	0	13	3
2016-11-03	8	10	1	6	9	28	0	0	9
2016-11-04	10	9	7	6	8	29	0	0	5
2016-11-05	0	2	0	7	6	2	0	0	1
2016-11-06	2	0	1	1	9	6	1	0	0
2016-11-07	9	1 5	6	6	5	33	0	0	1
2016-11-08	3	16	6	2	34	39	0	0	5
2016-11-09	10	35	12	12	38	97	0	5	16
2016-11-10	8	44	29	3	51	181	0	1	16
2016-11-11	13	42	154	3	60	135	0	3	13
2016-11-12	0	6	12	2	32	27	0	1	5
2016-11-13	0	4	3	0	21	17	0	0	4
2016-11-14	5	28	73	5	32	640	0	4	18
2016-11-15	11	7	304	3	21	350	0	3	18
2016-11-16	5	13	236	7	23	314	0	2	10
2016-11-17	3	12	98	1	19	390	0	2	9
2016-11-18	15	1 5	133	2	14	536	0	2	19
2016-11-19	0	4	24	1	5	78	0	0	4
2016-11-20	0	2	19	3	8	35	0	0	2
2016-11-21	22	44	25	4	14	117	0	1	9

Optional: Automate Crawling

```
def timer_crawling_first(code, num_page):
    url = "http://finance.naver.com/item/board.nhn?code=
    HEADERS={ 'user-agent': ('Mozilla/5.0 (Windows NT 6.1)
    r = requests.get(url, headers=HEADERS)
    soup = bs(r.text, 'html.parser').select('tbody')[0]
    dates = soup.findAll("span", class_="gray03")
    now = datetime.datetime.now()
    p = re.compile(str(now)[:10].replace('-','.'))
    todaylist = p.findall(str(dates))
     length = len(todaylist)
     return length
                                     오늘의 종목: 표준편차3 이상 게시글 개수 Dinbox x
                                                                                                                                                                                                                      def additional(code, num_page
     if timer crawling first(d
                                           acoustic0419@gmail.com
                                                                                                                                                                                         11:43 PM (4 minutes ago)
         return timer crawling
     else:
                                            ['249420', '079550', '039570', '095570', '213500', '133820', '192530', '027410', '183190', '180640', '068400', '053210', '019440', '138930', '082740', '102460', '114090', '093240', '011070', '092200', '095720', '094800'
          return 20 + additional
                                            044380', '009770', '092230', '091090', '023530', '073240', '034310', '020760', '055550', '047050', '030000', '029460', '024890', '018670', '030720', '030210', '006740', '017810', '007460', '017960', '010140', '010690'
                                            '008770', '012600', '004720', '007630', '007340', '008730', '005500', '005500', '005500', '001750', '005420', '016380', '007810', '006490', '002710', '003060', '003350', '001080', '003570', '001820', '000670', '004000', '001360'
                                            '008260', '002320', '001210', '000320', '003200', '002300', '001530', '000120', '000060']
kospi = pd.read excel('C:\\Us
kospiCode = kospi[0].tolist()
                                              Click here to Reply or Forward
resultdict = {}
for i in kospiCode:
    result = additional(i,1)
    resultdict[i] = result
                                    7 GB (46%) of 15 GB used
                                                                                                                          Terms - Privacy
                                                                                                                                                                                                  Last account activity: 0 minutes ago
df=pd.read excel('C:\\Users\\\
mean_plus_std3 = {}
for i in df.columns:
    mean nlus std3[i] = df[i] mean()+df[i] std()*3
```

Data Transformation

1. Filter Dates (std)

	051900	051910	047040	047050	042660	042670	051630	049770	034020
2016-10-30	0	0	0	0	0	0	0	0	0
2016-10-31	25	14	4	2	20	3	1	1	6
2016-11-01	11	15	9	12	26	18	0	1	7
2016-11-02	9	12	4	10	21	26	0	13	3
2016-11-03	8	10	1	6	9	28	0	0	9
2016-11-04	10	9	7	6	8	29	0	0	5
2016-11-05	0	2	0	7	6	2	0	0	1

	000020	000030	000040	000050	000060	000070	000080	000100	000120
std3	['2016-11-	['2016-11-	['2016-11-	['2016-11-	['2016-10-	['2016-10-	['2016-12-	['2016-11-	['2016-11-
std2	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-10-
std_2	['2017-01-	['2016-11-	['2016-11-	['2016-11-	['2016-11-	['2017-01-	['2016-12-	['2016-12-	['2016-11-
std_3	['2017-01-	['2016-11-	['2017-06-	['2016-11-	['2017-03-	['2017-01-	['2017-02-	['2016-12-	['2016-12-

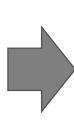
Data Transformation

1. Filter Dates (std3)

	Α	В	С	D	Е	F	G
1		000020	000030	000040	000050	000060	00007
2	std3	['2016-11-	['2016-11-	['2016-11-	['2016-11-	['2016-10-	['2016-1
3	std2	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-10-	['2016-1
4	std_2	['2017-01-	['2016-11-	['2016-11-	['2016-11-	['2016-11-	['2017-0
5	std_3	['2017-01-	17', '2017-()7-14', '201	7-07-20', '	2017-08-14	l', '2017·

```
for i in df.columns:
    dict1[i] = ast.literal_eval(df.loc["std_3",i])
df2= pd.DataFrame.from_dict(dict1, orient="index")
series1 = pd.Series()

for i in range(736):
    list1 = list(dict1.values())[i]
    se1 = pd.Series(list1)
    se1.index = [df.columns[i]]*len(list(dict1.values())[i])
    series1 = series1.append(se1)
```



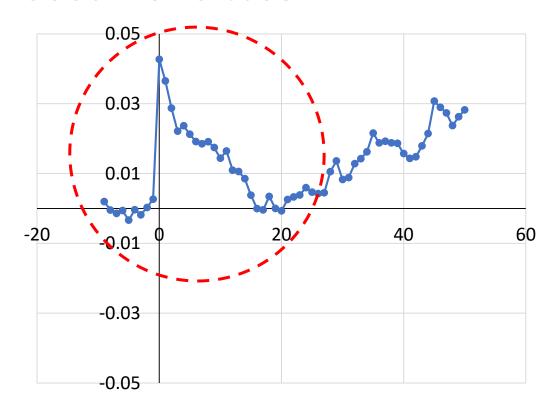
000020	2017-01-17
000020	2017-07-14
000020	2017-07-20
000020	2017-08-14
000020	2017-08-23
000020	2017-09-06
000020	2017-09-25
000020	2017-10-16
000020	2017-10-17
000030	2016-11-11
000030	2016-11-14
000030	2016-12-15
000030	2016-12-28
000030	2017-05-11
000040	2017-06-23
000040	2017-07-06
000040	2017-07-11
000040	2017-07-12
000040	2017-07-13
000040	2017-07-14
000040	2017-10-13
000040	2017-10-16
000050	2016-11-02

Data Analysis

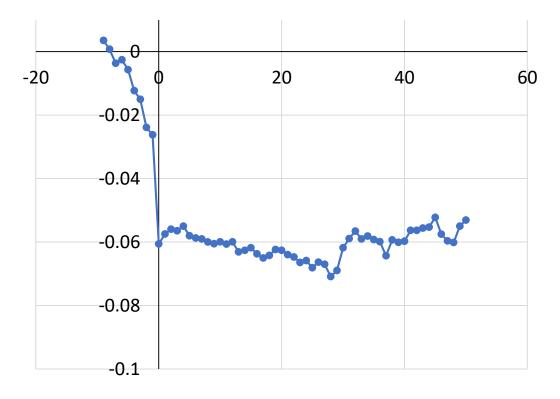
- β(beta) : 1 year of Kospi
- Rf: 5 years from investing.com
- Rm: Daily return of Kospi
- Abnormal return derived using CAPM
- 10 to 50 days from event day(NDR>std_3)
- Eliminate non-trade date
- Result: CAR(Cumulative Abnormal Return)

Result – STD3

Good News Case

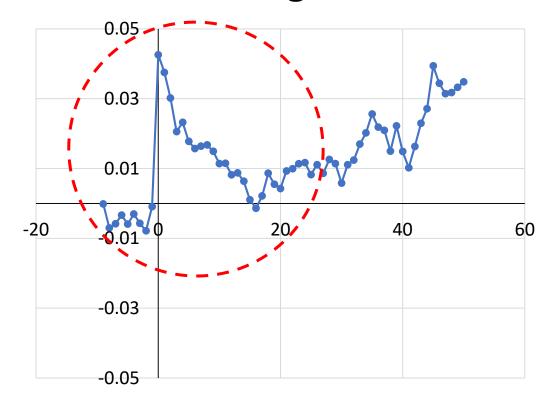


Bad News Case

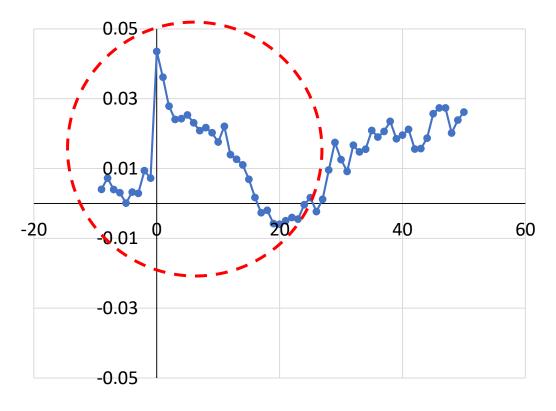


Result – Big & Small Comparison

Good News & Big

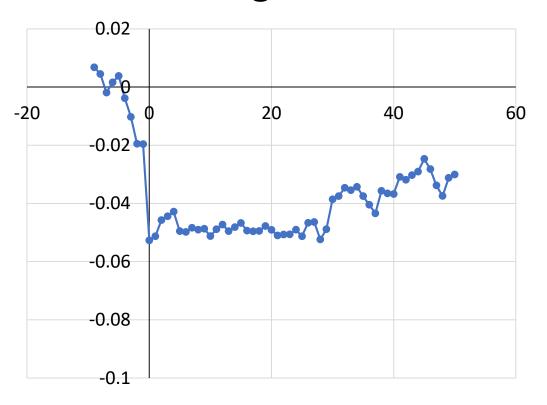


Good News & Small

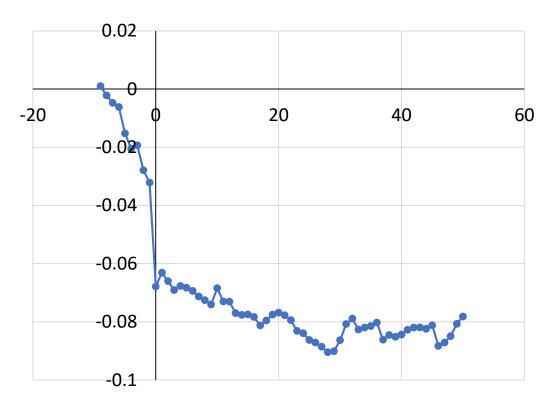


Result – Big & Small Comparison

Bad News & Big

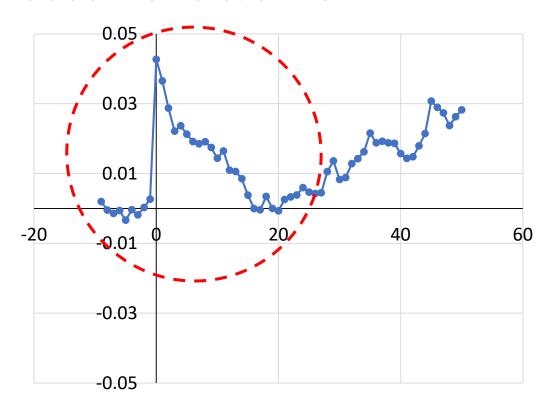


Bad News & Small

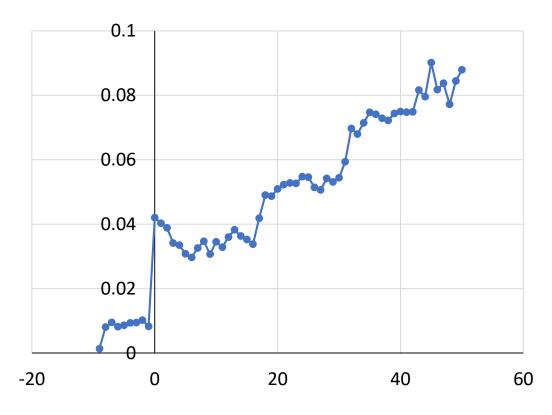


Result – STD3 & STD2 Comparison

Good News & STD3

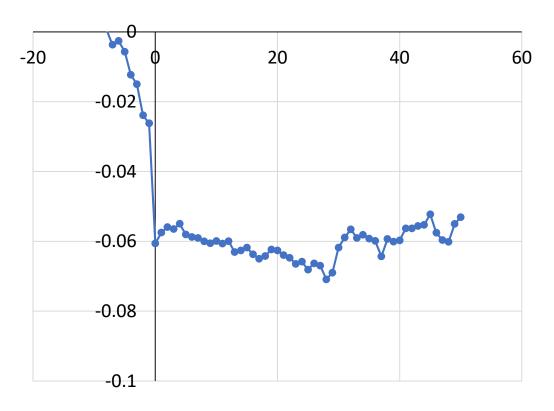


Good News & STD2

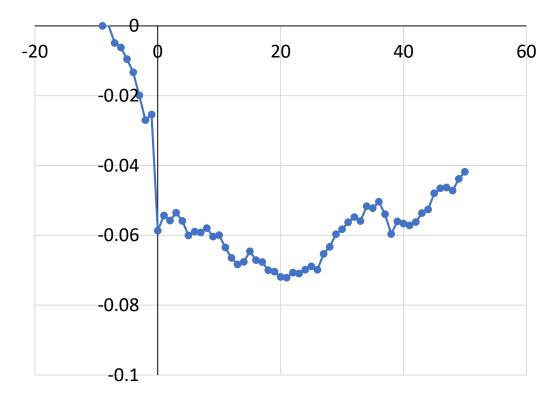


Result – STD3 & STD2 Comparison

Bad News & STD3



Bad News & STD2



Implication & Strategy

Individual Investors have tendency to **over-react** to a good news in short-term, and CAR shows decreasing in range [0,20] (CAR: -4%) In case of bad news, CAR is stable so there is no overreaction observed.

... Your Position should be **Short for 20 trading days**, then, you can get arbitrage chance from market's overreaction.

Risk & Improvement

Need for more database.

There is only 1-year range NDR posts. So, we are accumulating NDR post data from last October. Asking Naver for NDR data can be an option.

Need for Back-testing.

Due to lack of data (only 1 year!), we could not test our strategy with past stock market movement. However, if we back-test it with more wide time horizon, we can get more valid result than this.

Thank you for listening

Now, we'll take your questions.