

# **Forecasting Stock Returns with Search Ranking**

Hyun Ki Kim

# Research Question

## Does search trend precede price change?

순위	종목명	검색비율	현재가	전일비	등락률	거래량	시가	고가	저가	PER	ROE
1	금호타이어	2.43%	7,780	↑ 1,790	+29.88%	45,121,409	7,240	7,780	6,700	-34.12	-3.21
2	NAVER	2.25%	785,000	0	0.00%	68,102	798,000	799,000	780,000	33.48	18.50
3	인스코비	1.91%	8,450	▼ 750	-8.15%	13,795,286	8,910	8,970	8,080	-1,408.33	-1.48
4	셀트리온	1.21%	299,000	▼ 5,000	-1.64%	857,981	299,500	304,500	297,000	93.58	17.84
5	코디엠	1.04%	1,230	▲ 20	+1.65%	97,408,897	1,275	1,370	1,155	-33.24	-9.28
6	텔콘	1.04%	15,600	▲ 150	+0.97%	10,278,395	14,950	16,350	14,850	-65.82	-17.37
7	한농화성	1.04%	6,180	↑ 1,420	+29.83%	7,524,777	4,860	6,180	4,855	11.10	8.30
8	삼성바이오로직스	0.69%	485,500	▼ 500	-0.10%	127,365	484,000	489,500	480,000	-331.17	-2.41
9	제이준코스메틱	0.69%	23,150	▲ 1,600	+7.42%	3,606,017	21,300	23,400	21,100	21.16	25.67
10	삼성전자	0.69%	2,406,000	▼ 21,000	-0.87%	253,716	2,394,000	2,407,000	2,364,000	8.88	21.01
11	SK하이닉스	0.69%	81,800	▲ 500	+0.62%	3,546,192	80,300	81,800	78,700	5.60	36.80
12	아모레퍼시픽	0.52%	345,000	▲ 12,000	+3.60%	367,448	333,000	346,500	332,500	60.43	9.81
13	LG전자	0.52%	110,500	0	0.00%	375,711	109,500	112,000	109,000	11.58	13.69
14	라운시큐어	0.52%	4,015	▲ 110	+2.82%	1,339,573	3,875	4,080	3,830	37.88	23.82
15	카페 24	0.52%	129,100	▲ 2,100	+1.65%	126,323	125,100	130,000	124,700	201.40	N/A
16	파라다이스	0.52%	24,350	▲ 750	+3.18%	4,192,209	23,450	24,900	23,250	-117.07	-1.86
17	KT	0.52%	27,500	▲ 100	+0.36%	525,157	27,250	27,750	27,200	15.06	4.12
18	에이치엘비	0.52%	76,200	▲ 4,500	+6.28%	1,399,700	70,200	76,200	70,100	-166.01	-19.92
19	애경산업	0.52%	41,750	▲ 3,650	+9.58%	2,498,521	37,500	42,850	37,150	41.30	26.08
20	한국전력	0.52%	34,950	▲ 1,700	+5.11%	4,538,306	33,200	35,350	33,000	17.28	1.81

# Introduction

---

## Naver:

1. Largest search engine in Korea
2. Minute-level data

## Search ranking:

1. Heuristic for choosing a stock
2. Real-time web scraping

# Background

---

## Forecasting stock returns:

1. Twitter (Bollen et al., 2011)
2. Google Trend (Preis et al., 2013)
3. Wikipedia (Moat et al., 2013)

## Limitations:

1. Trading cost
2. Backtesting
3. Risk-adjusted return

# Rationale

---

## Why people search:

1. Price fluctuation
2. Limited attention
3. Good or bad news

## Need to measure:

1. Price volatility, newcomer
2. Company size, popularity
3. Price trend

# Rationale

---

## Why people search:

1. Price fluctuation
2. Limited attention
3. Good or bad news

## Need to measure:

1. Price volatility, newcomer
2. Company size, popularity
3. Price trend

# Method and Design

---

## **Data:**

1. High frequency data
2. Training period (4/3 – 4/13)
3. Validation period (4/16 – 4/27)
4. Testing period (4/30 – 5/11)

## **Theory:**

1. Supervised learning
2. Model selection
3. Forwardtesting
4. Risk-adjustment (Fama-French 5-factor model)

# Method and Design

---

## **Data:**

1. High frequency data
2. Training period (4/3 – 4/13)
3. Validation period (4/16 – 4/27)
4. Testing period (4/30 – 5/11)

## **Theory:**

1. Supervised learning
2. Model selection
3. Forwardtesting
4. Risk-adjustment (Fama-French 5-factor model)



# Q&A

---

**Questions?**

# References

---

Bollen, J., Mao, H. & Zeng, X. **Twitter mood predicts the stock market.** *Journal of Computational Science* 2, 1–8 (2011).

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E. & Preis, T. **Quantifying Wikipedia usage patterns before stock market moves.** *Scientific Report* 3, 1801 (2013).

Preis, T., Moat, H. S. & Stanley, H. E. **Quantifying Trading Behavior in Financial Markets Using Google Trends.** *Scientific Report* 3, 1684 (2013).