

# 분야별 관심도에 따른 KOSPI 증가·거래량 변동의 상호 연관성 분석

## I. 서론

주식 구매 행동에 영향을 미치는 요인은 개인마다 매우 다양하다. 최근, 정보통신 기술의 발달로 정보에 대한 접근성이 확대되고 비트코인으로 대표되는 가상화폐, 그리고 부동산 등으로 단기간에 많은 수익을 올리는 이들이 증가함에 따라 주식 투자에 대한 사람들의 관심 역시 증가하게 되었으며, 이는 실제로 많은 주식 구매 행동으로 이어지는 원인이 되었다.

실제 KOSPI 데이터에 따르면, 주식 거래량은 2020년 가량 급등한 후 2021년 2월 19일에 정점을 찍은 것으로 나타났다. 그렇다면 이러한 주식 거래의 변동에 있어 어떠한 분야가 사람들의 구매행동에 영향을 미칠 것일까.

이에 본인은 Granger Causality 개념을 활용해 사람들의 특정 분야에 대한 관심도의 변화가 사람들의 주식 구매 행동에 영향을 미치는지, 혹은 반대로 주식 가격 및 거래량의 변동이 어느 분야의 관심도의 변동에 영향을 미치는지에 대해 알아보고자 본 연구를 진행하게 되었다.

## II. 연구 소개 및 진행 절차

### (1). 데이터 소개

본 연구는 주식 가격 및 거래량에 사람들의 관심도가 영향력을 행사하는지 그 여부를 알아보고자 진행되었다. 이때 주식 가격 및 거래량은 KOSPI 데이터를 통해, 사람들의 관심도는 Google Trends에서 제시하는 주제 별 검색량 데이터를 통해 측정했다.

각 데이터에 대한 설명은 다음과 같다.

#### 1) KOSPI 데이터 (출처: FinanceData)

- KOSPI: 유가증권시장에 상장된 보통주를 기준으로 산출하는 시가총액식 주가지수
  - 기준(100): 1980년 1월 4일 (출처: e-나라지표)
- 기준 날짜: 2004년 1월 ~ 2022년 8월 (단위: 월)
- 변수: 날짜, 증가, 거래량
- KOSPI 데이터는 일별로 제공되지만, 비교 대상인 검색량 데이터가 월별로 제공됨에 따라 계산의 편의를 위해 해당 데이터를 월별 평균값으로 대체해 연구를 진행했다.

#### 2) 주제 별 검색량 데이터 (출처: Google Trends)

- 대상 주제: 건강, 사회, 정치, 금융
- 대상 지역: 대한민국
- 기준 날짜: 2004년 1월 ~ 2022년 8월 (단위: 월)
- 변수: 날짜, 검색량
- 검색량(시간의 흐름에 따른 관심도 변화):  
특정 지역 및 기간을 기준으로 차트에서 가장 높은 지점 대비 검색 관심도.
  - 검색 빈도가 가장 높은 검색어의 경우 100, 검색 빈도가 그 절반 정도인 검색어의 경우 50, 해당 검색어에 대한 데이터가 충분하지 않은 경우 0을 의미. (설명 출처: Google Trends)

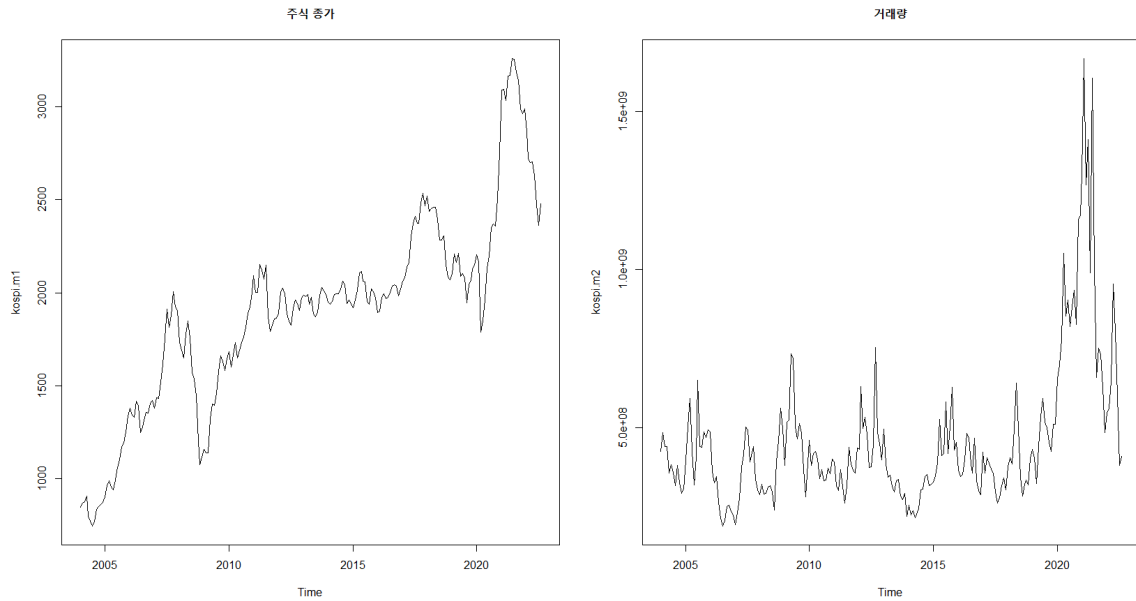
본 연구는 이와 같은 데이터를 바탕으로 진행되었다. 연구의 분석에는 R 프로그램을 사용했으며, 연구에 사용된 가설의 검정 및 모델링 결과와 그 해석 역시 이에 기초하여 진행된 것임을 미리 밝힌다.

### (2). 데이터 설명 및 전처리 과정

#### 1) 시각화 및 ACF, PACF

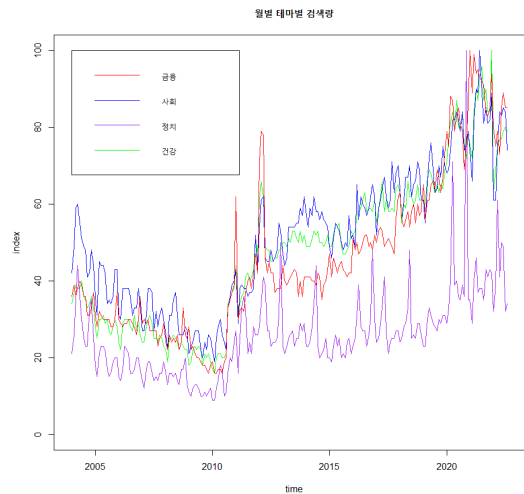
시간의 흐름에 따른 각 변수의 시각화 그래프는 다음과 같다.

## 1. 코스피 데이터



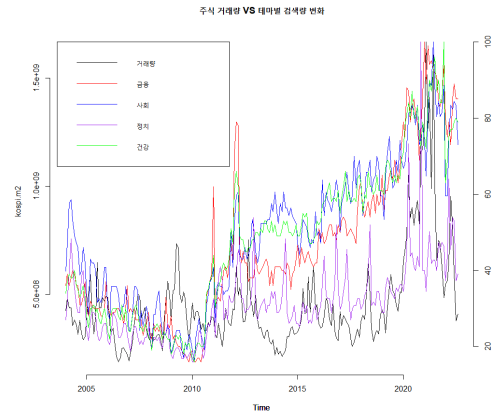
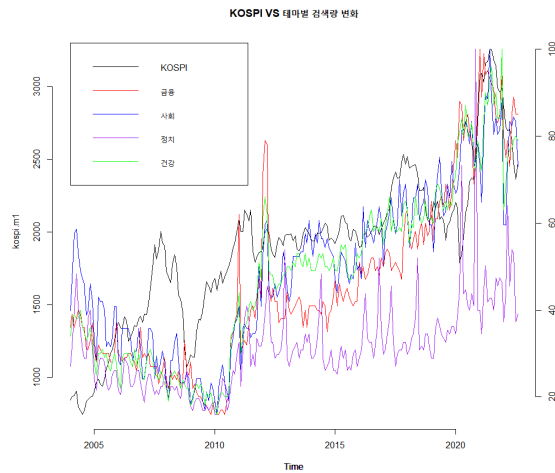
이에 따르면 KOSPI 증가 데이터는 전체적으로 우상향하는 트렌드를 보이고 있으며, 각 지점마다 변동폭이 큰 것을 알 수 있다. 반면 거래량 데이터는 일정한 추세를 보이지 않고 일정 수준을 유지하며 증가 및 감소를 반복하고 있음을 알 수 있다. 다만 거래량 데이터의 경우 2020년 경에 큰 폭으로 증가했다가 크게 감소하는 모습을 보이는데, 이는 코로나19의 발병으로 인해 주가가 크게 변동했던 영향으로 인한 것으로 사료된다.

## 2. 분야별 검색량 데이터



대부분의 주제 별 관심도가 전반적으로 우상향하는 모습을 보이지만, 정치 테마(보라색)의 경우 다른 지수들에 비해 그 값이 작아 사람들의 관심도가 상대적으로 저조함을 파악할 수 있다.

이를 KOSPI 증가 및 주식 거래량 데이터와 종합하여 나타낸 그래프는 다음과 같다.



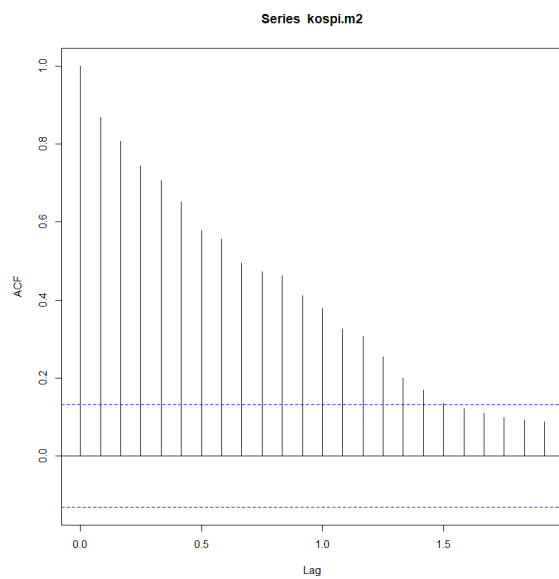
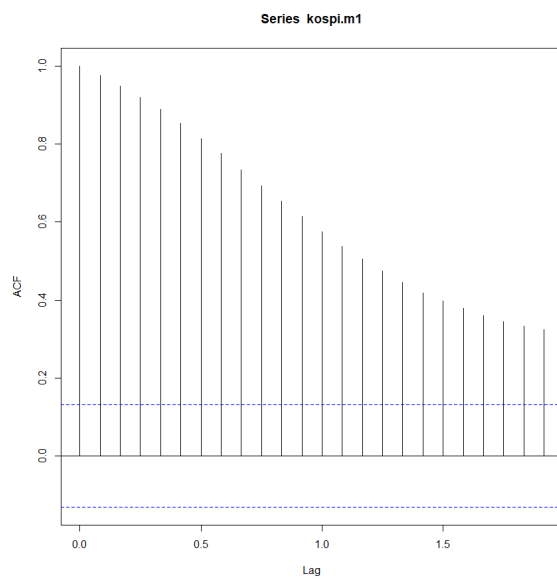
이에 따르면 검색량 데이터는 대략적으로 KOSPI 지수와 유사한 흐름을 보이며, 전반적으로 우상향하는 추세가 드러남을 파악할 수 있다. 이 때, KOSPI 증가(검은색)와 정치 테마에 대한 검색량(보라색)의 차이가 유독 두드러지는 것이 나타난다.

거래량 데이터의 경우는 전반적인 검색량 데이터와 다른 양상을 보였다. 대부분의 검색량 데이터가 우상향한 것에 비해 거래량은 일정 수준을 유지하며 증가 및 감소를 반복했으며, 정치 테마의 검색량(보라색) 데이터가 거래량 데이터의 변동과 가장 유사한 흐름을 보인 것으로 파악되었다.

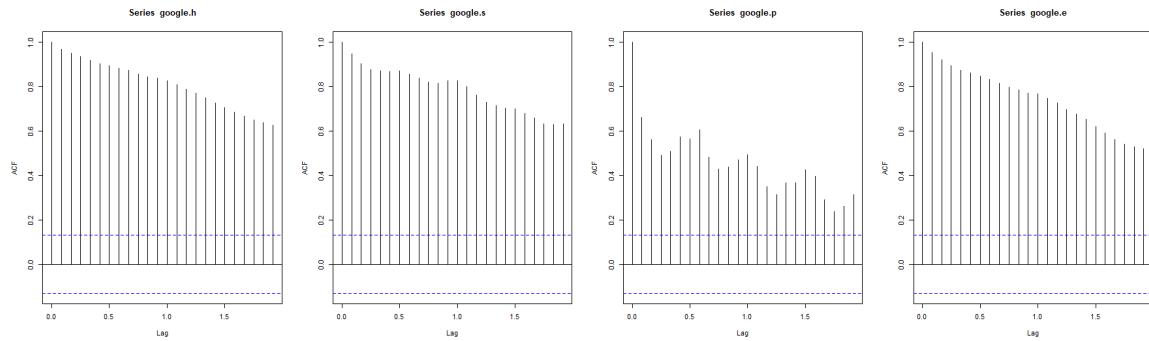
### 3. ACF 및 PACF 그래프

각 변수에 대한 ACF 그래프는 다음과 같다.

- 주식 데이터 (좌측부터 KOSPI 증가, 거래량)

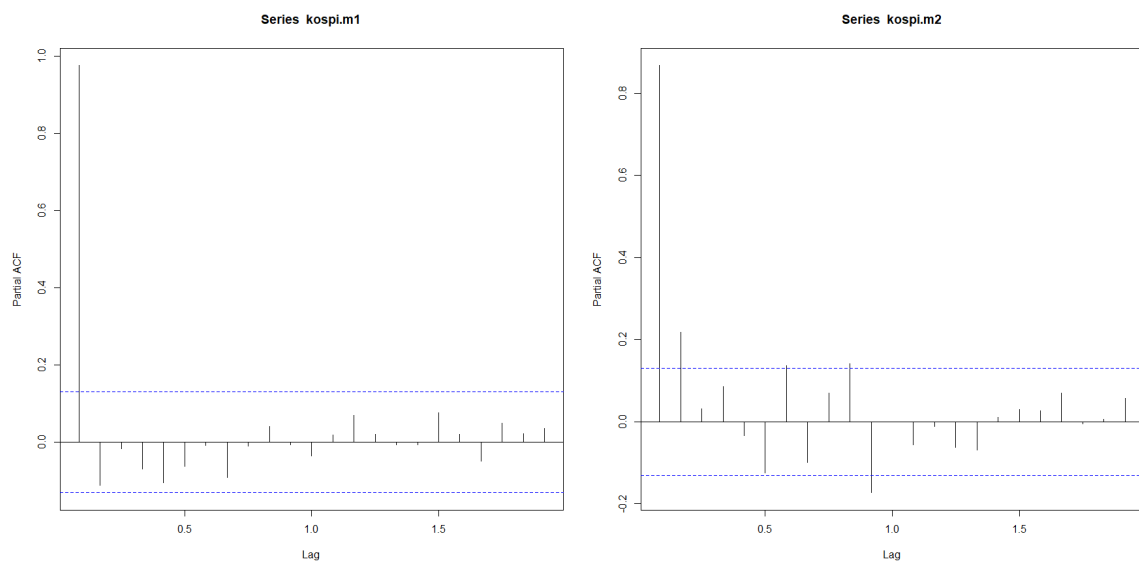


- 검색량 데이터 (좌측부터 건강, 사회, 정치, 금융)

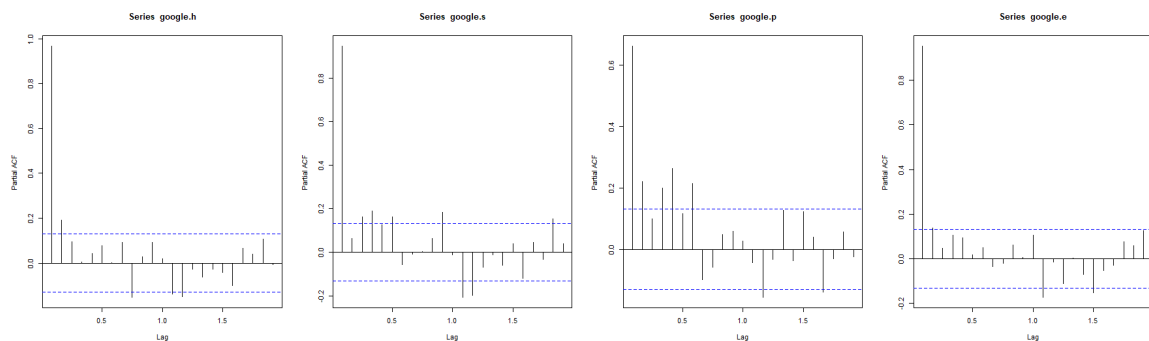


동일 변수에 대한 PACF 그래프는 아래와 같다.

- 주식 데이터 (좌측부터 KOSPI 증가, 거래량)



- 검색량 데이터 (좌측부터 건강, 사회, 정치, 금융)



## 2) 전처리: Stationary 변환

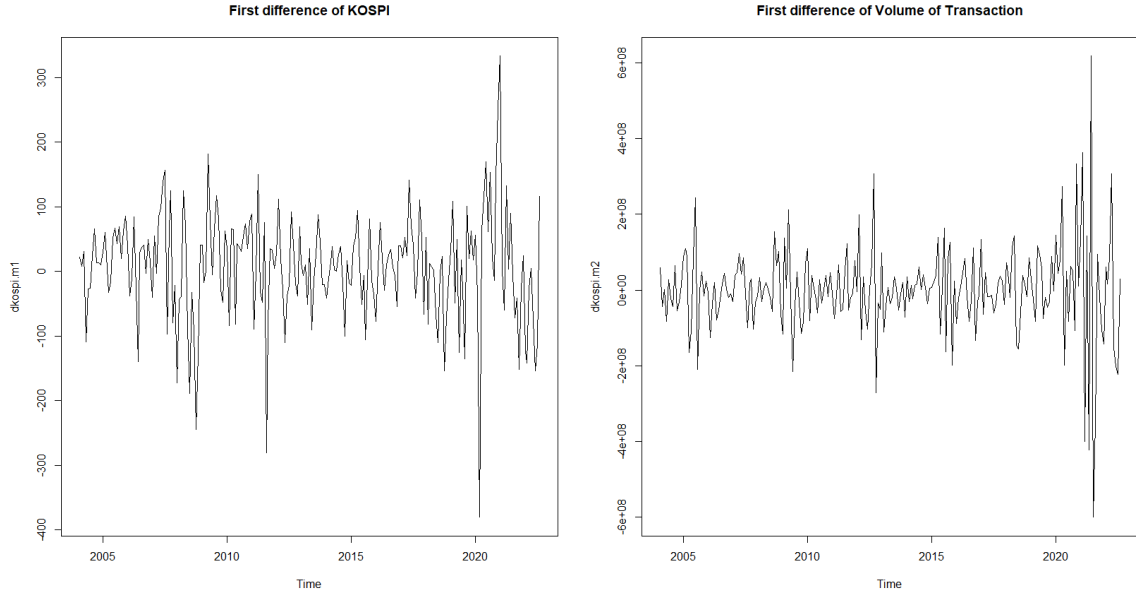
KPSS 검정( $H_0$ : Stationary,  $H_1$ : Non Stationary)을 통해 기존 데이터의 Stationary 여부를 확인한 결과, 모두 유의수준 5% 하에서  $H_0$ 를 기각해 Non Stationary한 데이터임이 드러났다. 이 때, 주식 거래량을 제외한 데이터는 모두 추세가 있는 것으로 판단해 Trend가 있는 것으로 가정하고 KPSS 검정을 진행했다(다만, 추세가 없는 Level 데이터로 가정한 검정에서도 KPSS값과 이에 따른 p-value의 차이만 보였을 뿐,  $H_0$ 를 기각하는 결과는 모두 동일하게 드러났다.).

기존 데이터에 대한 구체적인 KPSS 검정 결과는 다음과 같다.

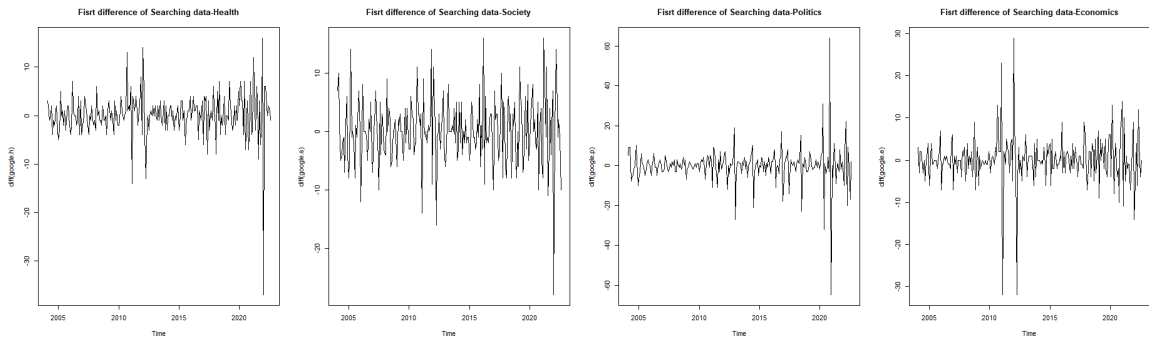
	KOSPI 증가	주식 거래량	건강 검색량	사회 검색량	정치 검색량	금융 검색량
KPSS 값	0.15796	1.3056	0.40425	0.54021	0.33456	0.60586
p-value	0.04004	0.01	0.01	0.01	0.01	0.01
결과	Non Stationary	Non Stationary	Non Stationary	Non Stationary	Non Stationary	Non Stationary

이에 해당 데이터를 Stationary하도록 변환하기 위해 한번 차분하는 과정을 거쳤다. 해당 데이터의 그래프는 다음과 같다.

- 주식 데이터 (1st difference) (좌측부터 KOSPI 증가, 거래량)

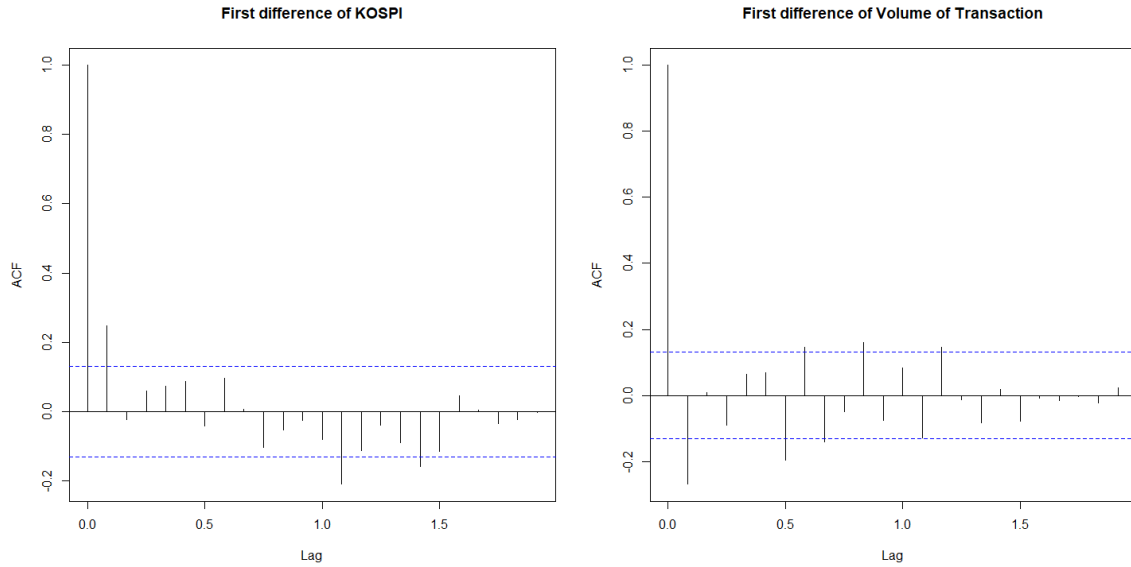


- 검색량 데이터 (1st difference) (좌측부터 건강, 사회, 정치, 금융)

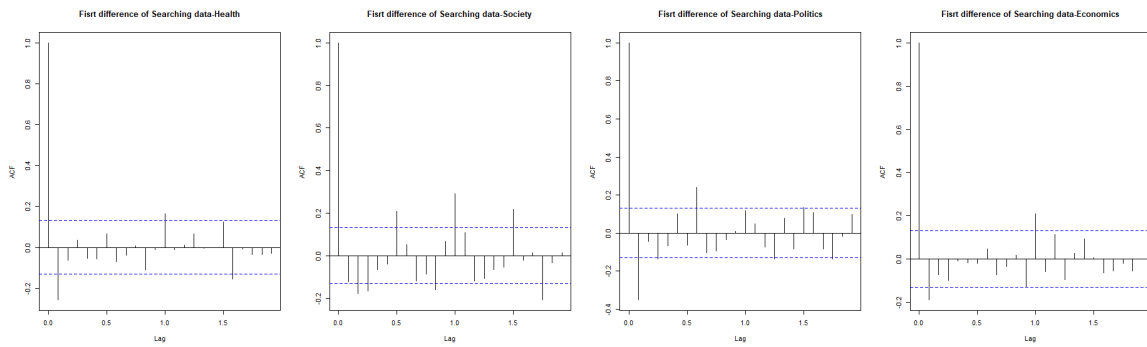


해당 데이터의 ACF는 다음과 같이 나타났다.

- 주식 데이터 (1st difference) (좌측부터 KOSPI 증가, 거래량)

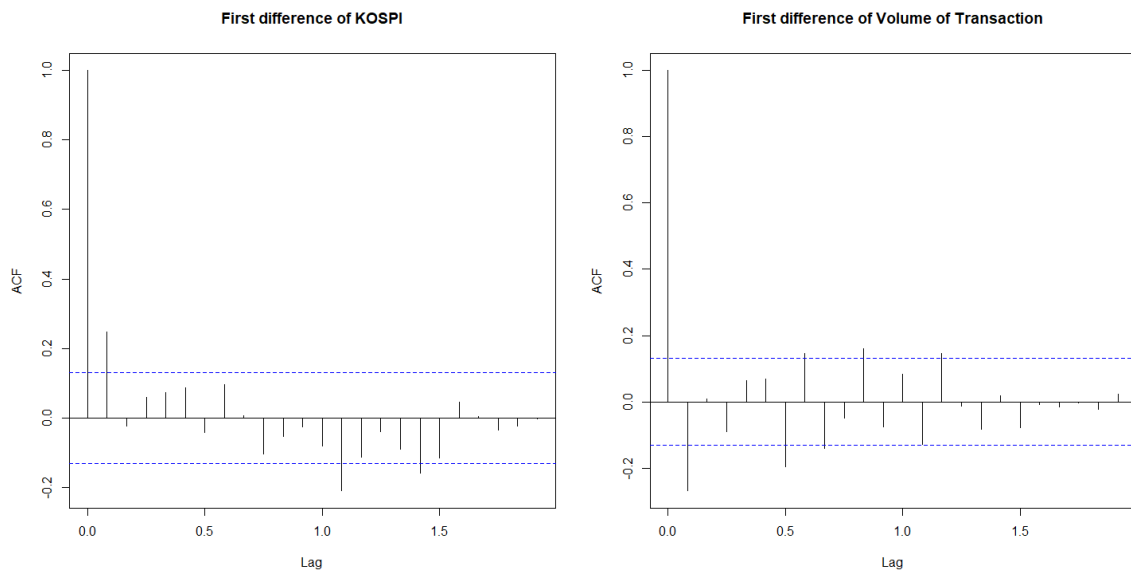


- 검색량 데이터 (1st difference) (좌측부터 건강, 사회, 정치, 금융)

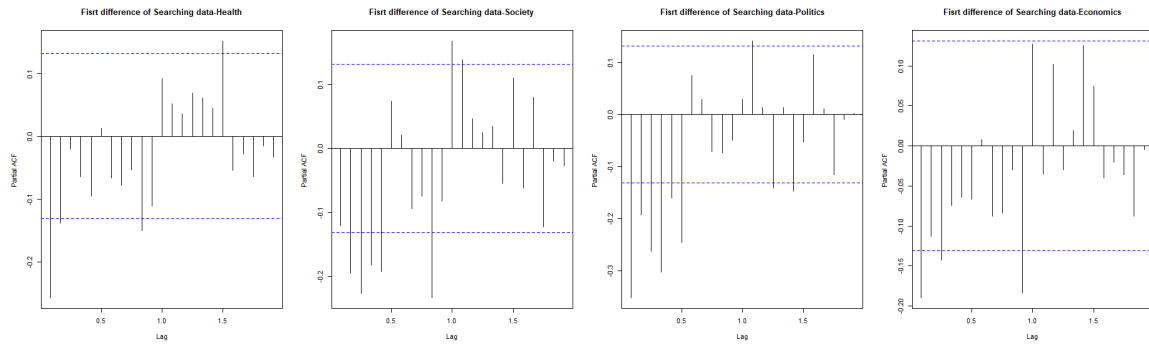


해당 데이터의 PACF는 다음과 같이 나타났다.

- 주식 데이터 (1st difference) (좌측부터 KOSPI 증가, 거래량)



- 검색량 데이터 (1st difference) (좌측부터 건강, 사회, 정치, 금융)



위와 같은 그래프들을 통해, 원본 데이터를 한번 차분한 해당 데이터가 Stationary 조건을 만족한 것으로 확인하고 KPSS 검정을 진행했다. 다만 이번에는 그래프에 따라 모든 데이터를 추세가 없는 Level 데이터로 가정하고 검정을 진행했으며, 모두 유의수준 5% 하에서  $H_0$ 를 기각하지 못해 Stationary함을 확인할 수 있었다.

해당 데이터에 대한 구체적인 KPSS 검정 결과는 아래와 같다.

	KOSPI 증가	주식 거래량	건강 검색량	사회 검색량	정치 검색량	금융 검색량
KPSS 값	0.054072	0.033369	0.074906	0.058547	0.019929	0.09337
p-value	0.1	0.1	0.1	0.1	0.1	0.1
결과	Stationary	Stationary	Stationary	Stationary	Stationary	Stationary

이에 해당 데이터를 사용해 모델링을 진행했다.

### (3). 모델링 및 결과 해석

모델링에는 시간의 흐름에 따른 두 변수의 상호 연관성을 나타내는 벡터자기회귀 모델(Vector Auto Regression, 이하 VAR 모델)을 사용했다. 이 과정에서 Overfitting과 같은 문제를 방지하고 보다 정확한 검정 결과를 얻고자 Stationary하도록 한번 차분한 데이터를 사용해 모델링을 진행했다. 이 때 파라미터 값인 최적의 Lag을 설정하기 위해 R의 var 패키지 내장 함수인 `VARselect` 함수를 사용했으며, Lag 선택 기준은 AIC 값을 사용했다. 다만 일부 Lag이 크게 설정된 모델의 경우, 정보의 효율적인 전달을 위해 마지막 Lag을 제외한 Lag은 결과에서 제외하고 표시했음을 미리 밝힌다.

이에 따른 KOSPI 증가 ( $y_{1,t}$ )와 주식 거래량 데이터 ( $y_{2,t}$ ), 그리고 건강 ( $x_{1,t}$ ), 사회 ( $x_{2,t}$ ), 정치 ( $x_{3,t}$ ), 금융 ( $x_{4,t}$ ) 테마 검색 데이터에 대한 VAR 모델링은 다음과 같이 나타났다.

#### 1) VAR 모델링

1. KOSPI 증가 ( $y_{1,t}$ ) VS 주식 거래량 데이터 ( $y_{2,t}$ ) (Lag: 1)

- 식

$$y_{1,t} = 0.2703y_{1,t-1} - 6.015 \times 10^{-08}y_{2,t-1} + 5.402$$

$$y_{2,t} = 44400y_{1,t-1} - 0.2769y_{2,t-1} - 693500$$

-  $y_{1,t}$ 의  $y_{2,t-1}$ 계수와  $y_{2,t}$ 의  $y_{1,t-1}$ 계수 모두 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

2. KOSPI 증가 ( $y_{1,t}$ ) VS 건강 ( $x_{1,t}$ ) (Lag: 1)

- 식

$$y_{1,t} = 0.24029y_{1,t-1} + 1.41194x_{1,t-1} + 5.32696$$

$$x_{1,t} = -0.0009965y_{1,t-1} - 0.2559054x_{1,t-1} + 0.2490255$$

-  $y_{1,t}$ 의  $x_{1,t-1}$ 계수와  $x_{1,t}$ 의  $y_{1,t-1}$ 계수 모두 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

3. KOSPI 증가 ( $y_{1,t}$ ) VS 사회 ( $x_{2,t}$ ) (Lag: 13)

- 식

$$y_{1,t} = \dots - 0.170813y_{1,t-13} + 0.212741x_{2,t-13}$$

$$x_{2,t} = \dots - 0.008852y_{1,t-13} + 0.2002x_{2,t-13}$$

-  $y_{1,t}$ 의  $x_{2,t-13}$ 계수가 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

#### 4. KOSPI 증가 ( $y_{1,t}$ ) VS 정치 ( $x_{3,t}$ ) (Lag: 9)

- 식

$$y_{1,t} = \dots - 0.053925y_{1,t-9} - 0.192438x_{3,t-9}$$

$$x_{3,t} = \dots - 0.016686y_{1,t-9} - 0.075037x_{3,t-9}$$

-  $y_{1,t}$ 의  $x_{3,t-9}$ 계수와  $x_{3,t}$ 의  $y_{1,t-9}$ 계수 모두 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

#### 5. KOSPI 증가 ( $y_{1,t}$ ) VS 금융 ( $x_{4,t}$ ) (Lag: 2)

- 식

$$y_{1,t} = \dots - 0.08919y_{1,t-2} - 1.96837x_{4,t-2}$$

$$x_{4,t} = \dots - 0.0004453y_{1,t-2} - 0.1194001x_{4,t-2}$$

-  $x_{4,t}$ 의  $y_{1,t-2}$ 계수가 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

#### 6. 주식 거래량 데이터 ( $y_{2,t}$ ) VS 건강 ( $x_{1,t}$ ) (Lag: 10)

- 식

$$y_{2,t} = \dots - 0.2335y_{2,t-10} + 4384000x_{1,t-10}$$

$$x_{1,t} = \dots - 5.794 \times 10^{-9}y_{2,t-10} - 0.1446x_{1,t-10}$$

-  $y_{2,t}$ 의  $x_{1,t-10}$ 계수와  $x_{1,t}$ 의  $y_{2,t-10}$ 계수 모두 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

#### 7. 주식 거래량 데이터 ( $y_{2,t}$ ) VS 사회 ( $x_{2,t}$ ) (Lag: 10)

- 식

$$y_{2,t} = \dots - 0.1802y_{2,t-10} + 3559000x_{2,t-10}$$

$$x_{2,t} = \dots - 1.312 \times 10^{-09}y_{2,t-10} - 0.222x_{2,t-10}$$

-  $x_{2,t}$ 의  $y_{2,t-10}$ 계수가 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

#### 8. 주식 거래량 데이터 ( $y_{2,t}$ ) VS 정치 ( $x_{3,t}$ ) (Lag: 7)

- 식

$$y_{2,t} = \dots - 0.06041y_{2,t-7} + 4415000x_{3,t-7}$$

$$x_{3,t} = \dots - 3.661 \times 10^{-09}y_{2,t-7} - 0.1089x_{3,t-7}$$

-  $x_{3,t}$ 의  $y_{2,t-7}$ 계수가 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

#### 9. 주식 거래량 데이터 ( $y_{2,t}$ ) VS 금융 ( $x_{4,t}$ ) (Lag: 3)

- 식

$$y_{2,t} = \dots - 0.1258y_{2,t-3} + 2874000x_{4,t-3}$$

$$x_{4,t} = \dots - 1.452 \times 10^{-09}y_{2,t-3} - 0.1451x_{4,t-3}$$

-  $x_{4,t}$ 의  $y_{2,t-3}$ 계수가 유의수준 5% 하에서 유효하지 않은 것으로 드러났다.

모형 전체의 유의성을 검정한 결과, 모든 모형이 유의수준 5% 하에서 유의한 것으로 드러났다.

## 2) Granger Causality 확인

이후, 상기한 VAR 모형을 바탕으로 변수 간 Granger Causality를 확인하는 절차를 거쳤다. `grangertest` 함수를 사용해 검정을 진행했으며,  $H_0$ 는 독립변수가 종속변수를 Granger Causal하지 않음을,  $H_1$ 은 독립변수가 종속변수를 Granger Causal함을 의미한다.

Lag는 VAR 모형을 적합할때 사용했던 값을 그대로 사용했으며, 그 결과는 다음과 같이 나타났다.

	결과					
원인	KOSPI 증가	주식 거래량	건강 검색량	사회 검색량	정치 검색량	금융 검색량
KOSPI 증가	-	0.2073	0.0755	1.1577	3.0449*	0.2933
주식 거래량	1.6834	-	4.1769*	1.4651	1.2209	1.9899
건강 검색량	1.3995	1.5283	-	-	-	-
사회 검색량	0.9688	2.0934*	-	-	-	-
정치 검색량	0.9605	7.0054*	-	-	-	-
금융 검색량	3.0471*	4.4309*	-	-	-	-

(\*: 유의수준 5%에서 유의)

즉, 금융에 대한 관심도는 KOSPI 증가에 2기의 시차를 두고, 사회·정치·금융 관심도는 주식 거래량에 각각 10기, 7기, 3기의 시차를 두고 Granger Causality가 나타남을 확인할 수 있다.



또한 반대의 방향으로, KOSPI 증가는 정치에 대한 관심도에, 주식 거래량은 건강에 대한 관심도에 각각 13기, 10기의 시차를 두고 Granger Causality가 나타남이 드러났다.

이를 요약한 결과는 아래와 같이 나타났다.

- 상호 Granger Causality가 나타나지 않은 데이터
  - KOSPI 증가&주식거래량 / KOSPI 증가&건강관심도 / KOSPI 증가&사회관심도
- 일방향의 Granger Causality가 나타난 데이터
  - 정치 관심도  $\leftarrow$  KOSPI 증가 / KOSPI 증가  $\leftarrow$  금융 관심도 / 건강 관심도  $\leftarrow$  주식 거래량  
주식 거래량  $\leftarrow$  사회 관심도 / 주식 거래량  $\leftarrow$  정치 관심도 / 주식 거래량  $\leftarrow$  금융 관심도
- 쌍방향의 Granger Causality가 나타난 데이터
  - 없음

### III. 마무리

본 연구 결과, 일부 기대와는 다른 결과가 나타나는 항목들이 보였다. KOSPI 증가와 주식 거래량은 상호 Granger Causality가 나타나지 않았으며, KOSPI 증가와 거래량의 변동은 금융에 대한 사람들의 관심도의 변동을 Granger Causal하지 못한 점이 그 대표적인 사례이다.

다만, 여건 상 일부 분석이 미진하게 진행된 점이 있었던 점은 보완해야 할 것으로 보인다. 가령, 보다 정확한 분석을 위해서는 본 연구의 주요 연구 대상이었던 KOSPI 지수에 대한 데이터가 비교 대상인 검색량 데이터와 마찬가지로 건강, 사회, 정치, 금융과 같은 테마로 구분해 연구를 진행했다면 보다 정확한 분석 결과를 얻었을 것으로 기대된다. 또한 시간 외 요인의 영향력을 설명하지 못하고, Granger Causality가 있다는 점 이외의 정확한 연관성을 해석해내지 못한 점은 한계점으로 볼 수 있다. 이와 같은 점을 염두에 두고 후속 연구가 진행된다면, 보다 발전되고 심도있는 결과를 얻을 수 있을 것으로 기대된다. 이에 본 연구는 주식 구매 행동 및 주식 가격이 각 분야의 관심도와 어떠한 연관성이 있는지 그 영향력의 방향성을 밝혀냈다는 점에 그 의의가 있었다고 생각한다.