

碩士 卒業報告書

指導教授 李熙祥

구글트렌드의 퀴리볼륨과 주식거래량 및
주가와
의 상관관계 연구
- S&P500 기업을 중심으로 -

成均館大學校 專門大學院

技術經營專門大學院

李 相 一

목차

제1장 서론	1
1.1 연구배경과 목적	1
1.2 연구의 범위와 구성	3
1.2.1 연구 범위	3
1.2.2 연구 구성	3
제 2 장 선행연구	5
2.1 구글트렌드에 의한 사회적, 경제적 활동 예측에 대한 연구	5
2.2 구글트렌드에 의한 금융시장의 변동성 예측에 관한 연구	6
2.2.1 검색 볼륨 데이터와 개별 기업의 주식거래량간의 상관관계 (주간 데이터 사용)	6
2.2.2 검색 볼륨 데이터와 개발 기업의 주식거래량간의 상관관계 (일간 데이터 사용)	7
제 3 장 분석방법	8
3.1 분석종목의 선정	8
3.2 데이터의 수집 기간 및 경로	8
3.3 데이터의 수집 방법	10
3.4 분석 방법	13
제 4 장 분석결과	15
4.1 쿼리볼륨과 주식거래량의 상관관계	15
4.2 쿼리볼륨과 주가와의 상관관계	21
제 5 장 결론	28

5.1 연구결과	28
5.2 추후연구과제.....	29
 참고문헌	 31

표목차

<표 3-1> S&P500 기업 리스트	9
<표 3-2> R 소스 코드 실행 결과 예시: 일일 주식 거래 데이터 수집	11
<표 3-3> 각 종목에 대한 일별 퀴리볼륨과 주식거래량 및 주가	12
<표 4-1> 상위 20개 기업에 대한 퀴리볼륨과 주식거래량의 피어슨 상관계수 ...	16
<표 4-2> 시간지연 상수에 따른 퀴리볼륨과 주식거래량의 피어슨 상관계수	19
<표 4-3> 상위 20개 기업에 대한 퀴리볼륨과 주가의 피어슨 상관계수	22
<표 4-4> 시간지연 상수에 따른 퀴리볼륨과 주가의 피어슨 상관계수	24

그림목차

<그림 3-1> R 소스 코드 예시 : 일일 주식 거래 데이터 수집	10
<그림 3-2> R 소스 코드 예시 : 일일 쿼리볼륨 데이터 수집	12
<그림 3-3> R 소스 코드 예시 : 피어슨 상관관계수 계산	14
<그림 4-1> 쿼리볼륨과 주식거래량의 기업별 피어슨 상관관계수 분포	15
<그림 4-2> 쿼리볼륨과 주식거래량의 시계열 그래프 및 상관관계수 그래프(NVIDIA)...	17
<그림 4-3> 쿼리볼륨과 주식거래량의 시계열 그래프 및 상관관계수 그래프(NETFLIX)	17
<그림 4-4> 상위 12 개 기업에 대한 쿼리 볼륨과 주식거래량의 시계열 그래프...	18
<그림 4-5> 업종에 따른 상관관계수 비교 (쿼리볼륨 vs. 주식거래량)	21
<그림 4-6> (+)상관관계를 가지는 상위 6 개 기업에 대한 쿼리볼륨과 주가와 의 시계열 그래프	23
<그림 4-7> (-)상관관계를 가지는 상위 6 개 기업에 대한 쿼리볼륨과 주가와 의 시계열 그래프	23
<그림 4-8> 업종에 따른 상관관계수 비교 (쿼리볼륨 vs.주가)	27

논문요약

구글트렌드의 쿼리볼륨과 주식거래량 및 주가의 상관관계 연구

본 연구는 미국 S&P500 에 소속되어 있는 505 개의 기업을 대상으로 피어슨 상관관계 분석(Pearson correlation analysis)을 이용하여 개별 기업에 대한 구글트렌드의 쿼리볼륨(Query volume)과 해당 기업의 주식 거래량 및 주가의 상관관계를 연구하였다. 최근 1 년간의 일별 데이터를 구글트렌드와 Yahoo Finance 에서 R 프로그램을 사용하여 수집하고 분석하였다. 연구결과, 쿼리볼륨과 주식거래량의 경우 분석기업 중 62%가 상관관계가 있는 것으로 나타났으며 상위 20 개 기업의 상관계수는 0.695~0.873 로 매우 높은 상관관계가 확인되었고 쿼리볼륨이 주식거래량을 1~4 일까지 예측하는 것으로 나타났다. 쿼리볼륨과 주가의 경우 분석기업 중 69%가 상관관계가 있는 것으로 나타났으며 상위 20 개 기업의 상관계수는 0.721~0.911 로 매우 높은 상관관계가 확인되었고 쿼리볼륨과 주가가 30 일의 지연시간에도 높은 상관관계를 보여주었다. 본 연구는 웹검색 트래픽을 나타내는 수단 중 하나인 구글트렌드의 쿼리볼륨의 변동성을 통해 주식거래량과 주가의 변동성을 예측할 수 있음을 확인하였으며 이를 통하여 금융기관이나 투자자 등이 위기 상황을 사전에 인지하고 적절한 대응을 할 수 있도록 도움을 줄 수 있다.

주제어 : 구글트렌드, 상관관계분석, 쿼리볼륨, 주식거래량, 주가

제 1 장 서 론

1.1 연구배경과 목적

구글트렌드는 인터넷 사용자가 여러 검색어의 상대 검색량을 분석할 수 있게 해주는 구글(Google Inc.)에서 무료로 제공하는 온라인 서비스이다. 구글트렌드의 결과는 절대 검색 빈도가 아닌 상대 검색 빈도 또는 검색량 지수(SVI) 형식으로 표시된다[1]. 구글트렌드에 대한 이해는 광고주, 마케터, 경제학자, 학자 및 자신의 세계에 대해 더 많이 알고 싶거나 현재의 최초상기도 (top-of-mind)가 무엇인지 알고 싶은 사람들에게 유용할 수 있다[2]. Hyunyoung Choi & Hal Varian(2009)의 연구를 기점으로 구글트렌드의 쿼리 검색이나 소셜미디어의 텍스트 분석으로 여러 사회적, 경제적 활동을 예측하는 연구들이 진행되었다[3-7]. 예를 들면 자동차 판매량이나 신규 주택 판매량 및 전염병의 확산 등을 예측하는데 구글트렌드를 사용하는 연구가 진행되었고[3,4] 미국 대통령 선거에서 대통령 당선을 예측하는데 소셜미디어인 twitter 을 사용하는 연구가 진행되었다[5]. Hyunyoung Choi & Hal Varian(2009)은 구글트렌드의 쿼리볼륨이 미래를 예측하는 수단이 아니라 지연되어 발표되는 사회적, 경제적 지표의 현재를 예측하는 수단임을 강조하였다. 구글트렌드의 경우 여러 사회적, 경제적 지표와 관련된 쿼리볼륨에 대한 데이터를 일별이나 주별로 바로 제공하므로 사회적 활동이나 경제적 활동의 현재 수준을 시간 지연 없이 추정할 수 있다고 가정하고 검증할 수 있다는 것이다[3].

구글트렌드의 퀴리 데이터로 경제적 활동을 예측하는 연구는 여러 산업 분야 중 금융 시장에서의 움직임을 예측하는 연구로 이어졌다[6-7]. Tobias Preis & Daniel Reith(2010)는 구글트렌드의 퀴리 볼륨과 금융시장의 변동성이 주간 단위로 관련성이 있는지를 조사하여 “기업명”에 대한 주간 퀴리볼륨과 해당 기업의 주간 주식거래량의 상관관계를 확인하였다. 그들의 연구결과 퀴리볼륨과 주식거래량의 주간 단위의 상관관계는 매우 낮았으며 상관관계를 엄밀하게 확인하는데 있어 주간 단위의 데이터는 불충분하다고 언급되었다[6]. Ilaria Bordino & Stefano Battiston(2012)는 퀴리볼륨과 주식거래량의 상관관계를 더 정밀하게 확인하기 위하여 일별로 데이터를 분석하였고 기존 연구에서 퀴리로 사용한 “기업명” 대신에 기업의 “티커(ticker)”를 퀴리로 사용하여 매우 높은 상관관계를 확인하였다[7]. 본 연구는 Bordino, et al.(2012)의 연구 보다 대상 기업을 확대하여 퀴리볼륨과 주식거래량의 일별 상관관계를 확인하고 퀴리볼륨과 주가의 일별 상관관계를 추가적으로 확인하였다. Bordino, et al(2012)의 경우 주로 기술기업으로 구성된 NASDAQ-100에 속한 100개의 기업을 대상으로 분석하였으나 본 연구는 다양한 업종으로 구성된 S&P500에 속한 505개 기업으로 분석 대상을 확대하여 최근 1년간의 데이터를 분석하였다. 본 연구를 통하여 Bordino, et al(2012)가 검증한 일별 퀴리볼륨과 주식거래량의 상관관계가 대상 기업의 차이에 따라 변화가 없는지를 밝히고 일별 퀴리볼륨과 주가와 상관관계를 추가적으로 확인하는 것을 본 연구의 목적으로 하였다.

1.2 연구의 범위와 구성

1.2.1 연구 범위

웹 검색 트래픽을 이용한 사회적, 경제적 활동 예측에 대한 연구 중 금융 변동성 예측에 대한 연구를 진행하였으며 웹 검색 트래픽은 구글트렌드의 쿼리볼륨을 사용하였다.

2016 년도 7 월에서 2017 년도 6 월까지 최근 1 년간의 구글트렌드의 쿼리볼륨과 주식거래량 및 주가와의 상관관계를 분석하였다. 구글트렌드의 일별 쿼리 볼륨은 구글트렌드(<https://trends.google.com>)에서 수집하였고 일일 재무 데이터는 Yahoo Finance (<http://finance.yahoo.com>)에서 오픈 소스인 R 프로그램으로 수집하였다.

1.2.2 연구 구성

본 연구는 총 5 장으로 구성된다. 2 장에서는 구글트렌드에 의한 사회적, 경제적 활동에 대한 예측과 금융 시장의 변동성에 대한 예측에 대한 선행 연구를 살펴본다. 3 장에서는 본 연구에서 사용되는 분석방법에 대해 알아본다. R 프로그램을 사용하여 Yahoo Finance 와 구글트렌드에서 데이터를 수집하고 수집한 데이터를 피어슨 상관관계분석을 통하여 분석하는 방법을 살펴본다. 4 장에서는 본 연구의 분석 결과들을 보여 준다. 쿼리볼륨과 주식거래량의 상관관계와 쿼리볼륨과 주가와의

상관 관계를 나누어 살펴본다. 5 장은 본 연구의 결과들을 정리하고 추후 연구 과제에 대해 언급한다.

제 2 장 선행연구

2.1 구글트렌드에 의한 사회적, 경제적 활동 예측에 대한 연구

구글트렌드에 의한 사회적 활동 연구 중 성공적인 연구 결과 중 하나는 Ginzberg J & Mohebi M(2009)에 의해 수행된 연구이다. 그들의 연구에 따르면 미국에 살고 있는 사람들 사이에서 유행성 독감 바이러스가 전염되는 것과 관련하여 인플루엔자 및 그 치료와 관련된 키워드에 대한 검색 활동으로 의료기관이 수집한 공식 데이터로 측정한 실제 확산을 예측할 수 있음이 나타났다[4].

Hyunyoung Choi & Hal Varian(2009)은 시계열의 자기회귀 모델에 구글트렌드 변수를 추가한 연구를 진행하였다. 이 연구의 경우 자기회귀 모델보다 상당한 개선을 나타냈는데 제조사별 자동차 판매량의 경우 18%가 개선되었고, 신규 주택 판매의 경우 12% 개선되는 결과를 보였다[3].

2.2. 구글트렌드에 의한 금융시장의 변동성 예측에 관한 연구

2.2.1 검색 볼륨 데이터와 개별 기업의 주식 거래량간의 상관 관계 연구(주간 데이터 사용)

Tobias Preis & Daniel Reith(2010)은 구글 검색 볼륨 데이터와 금융시장의 변동성이 주간 단위로 관련성이 있는지를 조사하였다. 그들은 S&P 500 종목에 대한 ‘기업명’에 대한 주간 검색 볼륨과 S&P 500 기업 주식의 주간 거래량의 상관 관계를 확인하였다. 개별 기업(예 : APPLE)의 주별 가격 변동과 검색 볼륨 변동 및 주별 거래량 변동과 검색 볼륨 변동 사이의 상호 상관 관계를 확인했을 때 가격 변동과 검색 볼륨 변동은 의미 있는 상관 관계가 없었고 거래량 변동과 검색 볼륨 변동 사이에는 시간 지연이 없는 $\Delta t=0$ 에서 중요한 상관관계를 나타내었다. S&P 전체 기업에 대하여 상기와 동일한 상관 관계 확인 시 거래량 변동과 검색 변동 사이에 상관 관계가 확인되었지만 개별 기업보다는 약한 상관 관계를 보였다. 좀 더 엄밀하게 상관 관계를 확인하기 위해서 주별 데이터보다 더 짧은 기간의 데이터 확인이 필요하다고 제안하고 있으며 개별 기업의 상관관계 또한 수치적으로 다소 낮게 나타났다[6].

2.2.2 검색 볼륨 데이터와 개별 기업의 주식 거래량간의 상관 관계 연구(일간 데이터 사용)

Ilaria Bordino & Stefano Battiston(2012)는 검색 볼륨과 주식 거래량 변동성의 상관 관계를 더 정밀하게 확인하기 위하여 일일 빈도로 데이터를 분석하였다. 또한 기존 연구에서 쿼리로 사용한 “기업명” 대신에 각 기업의 “티커 (ticker)” 를 쿼리로 사용하였다. 티커 (ticker) 의 경우 주식에 대한 정보를 취득할 목적으로만 사용하기 때문에 검색 볼륨 변동성과 주식 거래량 변동성 사이의 상관 관계가 더 높아질 것으로 예상할 수 있다. 그들은 NASDAQ-100 에 소속된 87 개 기업에 대해 2010 년 중반에서 2011 년 중반까지 1 년간의 데이터를 일별로 검색 볼륨과 주식 거래량 간의 상호 상관 관계를 분석하였다. 조사한 87 개 기업 중 일부에 대한 상호 상관 관계를 확인하였고, 시간 지연에 따른 상관 관계의 영향성은 최대 1~3 일까지 나타났다. 시간 지연이 없는 $\Delta t=0$ 에서의 상관계수가 큰 순서로 10 개의 기업을 나열할 경우 상관계수는 0.65~0.83 으로 매우 높은 상관 관계를 확인할 수 있으며 시간 지연에 따른 상관계수 또한 1~3 일의 지연에서 0.11~0.68 로 꽤 높은 상관 관계를 확인할 수 있었다. [7]

제 3 장 분석방법

3.1. 분석 종목의 선정

본 연구에서는 S&P 500 지수에 포함된 기업을 분석 종목으로 선정하였다. S&P 500 지수는 국제 신용평가기관인 미국의 Standard and Poors(S&P)가 작성한 주가 지수로 뉴욕증권 거래소에 상장된 기업 중 우량기업주 500 개를 중심으로 산정하는 지수이다. S&P 500 지수는 다우존스 지수와 나스닥 지수와 더불어 뉴욕증시의 3 대 주가지수에 속한다. NADAQ-100 에 소속된 87 개 기업에 대하여 실시한 선행 연구가 시간의 경과에 따라서 유효한지 확인할 뿐 아니라 데이터 범위를 확장하여 기술 기업뿐 아니라 다양한 분야의 기업을 대상으로 할 경우에도 유사한 결과를 도출할 수 있는지를 확인하는데 의의가 있다.

3.2. 데이터의 수집 기간 및 경로

2016 년 7 월 1 일에서 2017 년 6 월 30 일까지 1 년의 기간 동안 일별로 데이터를 분석할 데이터를 수집하였다. 모든 주식에 대한 일일 재무 데이터는 Yahoo Finance (<http://finance.yahoo.com>)에서 수집하였고 일일 거래량과 주가(종가)를 중심으로 수집하였다.

구글 퀴리볼륨 데이터는 구글트렌드(<https://trends.google.com>)에서 수집할 수 있으며 국가와 지역 단위로 데이터를 얻을 수 있다. 본 연구는 S&P 500에 속하는 미국 기업을 대상으로 했기 때문에 구글트렌드에서 수집한 퀴리볼륨 데이터는 국가를 ‘미국’으로 한정하여 수집되었다. 하기 <표 3-1>은 S&P500에 해당하는 505개 기업을 티커의 알파벳 순서로 분류하여 나타낸 것이다. 나머지 기업들은 생략하고 505개 기업 중 앞의 10개 기업과 마지막 10개 기업만을 예시로 나타내었다.

<표 3-1> S&P500 기업 리스트

Num	Ticker	Name	Sector
1	A	Agilent Technologies Inc	Health Care
2	AAL	American Airlines Group	Industrials
3	AAP	Advance Auto Parts	Consumer Discretionary
4	AAPL	Apple Inc.	Information Technology
5	ABBV	AbbVie	Health Care
6	ABC	AmerisourceBergen Corp	Health Care
7	ABT	Abbott Laboratories	Health Care
8	ACN	Accenture plc	Information Technology
9	ADBE	Adobe Systems Inc	Information Technology
10	ADI	Analog Devices, Inc.	Information Technology
~			
495	XL	XL Capital	Financials
496	XLNX	Xilinx Inc	Information Technology
497	XOM	Exxon Mobil Corp.	Energy
498	XRAY	Dentsply Sirona	Health Care
499	XRX	Xerox Corp.	Information Technology
500	XYL	Xylem Inc.	Industrials
501	YHOO	Yahoo Inc.	Information Technology
502	YUM	Yum! Brands Inc	Consumer Discretionary
503	ZBH	Zimmer Biomet Holdings	Health Care
504	ZION	Zions Bancorp	Financials
505	ZTS	Zoetis	Health Care

3.3. 데이터의 수집 방법

모든 데이터의 수집 및 처리는 R program 과 R 의 오픈 소스 통계 패키지를 사용하였으며 다량의 데이터의 저장과 처리를 위해 MSSQL 을 함께 사용하였다. 미국 주식 종목 및 기업의 분류 데이터는 NASDAQ 사이트에서 다운로드 가능하다. (<http://www.nasdaq.com/screening/company-list.aspx>).

수집 종목에 대한 주식의 일일 거래와 관련된 1년간의 주식 데이터의 경우 Yahoo Financial 에서 R 로 수집하여 MSSQL 에 Table 형태로 저장하였고 유사한 방법으로 각 기업에 대한 1년간의 구글 퀴리 볼륨에 대한 일일 데이터를 구글트렌드에서 R 을 이용하여 수집하였다. Yahoo Financial 에서 각 종목의 일일 거래 데이터를 수집하여 데이터베이스에 저장하는 R 의 소스코드 중 일부를 <그림 3-1> 에 예시하였다.

<그림 3-1> R 소스 코드 예시 : 일일 주식 거래 데이터 수집

```
insertIntoPriceDataFromYahoo<-function(ticker)
{
  con =odbcDriverConnect('driver={SQLServer};server=localhost;database=master;trusted_connection=true')
  sql = paste(sep="","delete from RawData.Dbo.PriceDataFromYahoo where ticker =",ticker,"")

  dfYahoo = getSymbols(ticker, auto.assign = FALSE, src='yahoo')
  sqlQuery(con, sql)
  date_list = rownames(data.frame(dfYahoo))
  for(date_ in date_list)
  {
    df = as.vector(dfYahoo[date_==date_list,])
    sql = paste(sep="","Insert into RawData.Dbo.PriceDataFromYahoo Values('", date_,"'",
      , ticker,"'", df[1], ",", df[2], ",", df[3], ",", df[4], ",", df[5], ",", df[6], ",")"
    )
    sqlQuery(con, sql)
  }
  close(con)
}
```


<그림 3-1>의 소스 코드를 살펴보면 Yahoo Financial 에서 각 종목의 주식거래량과 주가를 얻기 위하여 getSymbols() 함수를 사용하였다. getSymbols() 함수를 사용하면 손쉽게 각 종목의 시가, 고가, 저가, 종가, 거래량, 수정주가를 얻을 수 있다. 예시한 코드를 실행시키면 하기 <표 3-2>와 같은 결과물을 얻을 수 있다. <표 3-2>는 실행한 결과의 일부분만을 예시로 나타낸 것이다.

<표 3-2> R 소스 코드 실행 결과 예시: 일일 주식 거래 데이터 수집

Date	Ticker	Open	High	Low	Close	Volume	AdjPrice
2016-07-01	AAPL	93.696	94.657	93.539	94.0882	26026500	95.89
2016-07-02	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
2016-07-03	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
2016-07-04	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
2016-07-05	AAPL	93.598	93.607	92.685	93.20511	27705200	94.99
2016-07-06	AAPL	92.822	93.863	92.597	93.73497	30949100	95.53
2016-07-07	AAPL	93.902	94.687	93.823	94.13727	25139600	95.94
2016-07-08	AAPL	94.677	95.069	94.245	94.86337	28912100	96.68

구글트렌드에서 각 기업의 티커(ticker)를 입력하고 쿼리볼륨 데이터를 수집하여 SQL 테이블에 저장하는 R 의 소스코드의 예시를 <그림 3-2>에 나타내었다. 구글트렌드에서 쿼리볼륨을 일별로 수집하기 위하여 gtrends()함수를 사용하였다. gtrends() 함수는 gtrends(“검색어”, geo=“국가”, “시작날짜 종료날짜”) 형태로 사용할 수 있다. <그림 3-3> 예시 코드에서는 검색어를 각 종목의 티커(ticker)로 설정하고 지역을 미국으로 하고 수집 일자 2016 년 7 월 1 일부터 1 년 간 데이터를 수집하기 위해 변수로 설정하였다. 쿼리볼륨을 일별로 수집할 수 있는 최대 기간은 6 개월이므로 6 개월 이상의 데이터를 수집하기 위한 기능을 for() 함수로 구현하였다.

<그림 3-2> R 소스 코드 예시 : 일일 쿼리볼륨 데이터 수집

```
for(kin 1:duration_times){
  google_query <- gtrends(ticker, geo ="US", period[k])
  google_trend <- google_query$interest_over_time
  google_trend_data <- data.frame(google_trend)
  google_trend_hits_sum <- rbind(google_trend_hits_sum, google_trend_data)
}
```

위에서 예시한 것과 같은 방법으로 R 을 사용하여 각 기업의 일일 거래 데이터와 구글트렌드에서 쿼리볼륨 데이터를 수집하여 결합하고 티커(ticker)와 날짜순으로 분류하여 <표 3-3> 과 같은 형태의 최종 데이터를 얻었다. <표 3-3> 은 S&P500 에 속한 505 개 기업의 일일 거래 데이터와 일간 쿼리볼륨 데이터를 1 년의 기간으로 수집한 데이터의 일부를 나타낸 것이다. QueryVolume, Volume, AdjPrice 열은 각각 쿼리볼륨, 주식거래량, 수정주가를 나타낸다.

<표 3-3> 각 종목에 대한 일별 쿼리볼륨과 주식거래량 및 주가

Num	Date	Query Volume	Ticker	Open	High	Low	Close	Volume	AdjPrice
1091	2017-06-26	38	AAP	120.06	124.62	120.06	121.790001	1544400	121.790001
1092	2017-06-27	40	AAP	121.39	123.3	121.29	121.959999	1652600	121.959999
1093	2017-06-28	44	AAP	119.75	120.6	116.63	116.709999	2106100	116.709999
1094	2017-06-29	42	AAP	116.25	118.25	115.62	116.050003	2184100	116.050003
1095	2017-06-30	31	AAP	116.45	117.82	116.03	116.589996	981200	116.589996
1096	2016-07-01	25	AAPL	93.696	94.657	93.539	94.088203	26026500	95.889999
1097	2016-07-02	6	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
1098	2016-07-03	4	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
1099	2016-07-04	6	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
1100	2016-07-05	26	AAPL	93.598	93.607	92.685	93.205109	27705200	94.989998
1101	2016-07-06	25	AAPL	92.822	93.863	92.597	93.73497	30949100	95.529999
1102	2016-07-07	22	AAPL	93.902	94.687	93.823	94.137268	25139600	95.940002
1103	2016-07-08	27	AAPL	94.677	95.069	94.245	94.863365	28912100	96.68
1104	2016-07-09	6	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
1105	2016-07-10	3	AAPL	NULL	NULL	NULL	NULL	NULL	NULL
1106	2016-07-11	28	AAPL	94.932	95.815	94.912	95.157722	23794900	96.980003

3.4. 분석방법

수집한 기업의 주식 거래 데이터인 일별 거래량 및 주가와 구글트렌드에서 수집한 쿼리볼륨 데이터와의 상관관계를 피어슨 상관관계수 (Pearson Correlation Coefficient) $r(\delta)$ 를 통하여 확인하였다. 피어슨 상관관계수 r 은 -1에서 1의 범위를 가지며 절대값이 높을수록 높은 상관 관계를 가지고 (+)일 경우 양의 상관관계를, (-)일 경우 음의 상관관계를 가진다. <수식 3-1>에서 \bar{Q} , \bar{T} 는 일별 쿼리볼륨과 일별 거래데이터(거래량 및 주가)의 평균을 나타내고 Q_t , T_t 는 시점 t 에서의 일별 쿼리볼륨과 일별 거래 데이터(거래량 및 주가)를 나타낸다. δ 는 시간 지연에 대한 상수로 거래량과 쿼리볼륨의 경우 최대 5일의 지연을 고려하였고 주가와 쿼리 볼륨의 경우 최대 30일의 지연을 고려하였다.

$$r(\delta) = \frac{\sum_{t=1}^n (Q_t - \bar{Q})(T_{t+\delta} - \bar{T})}{\sqrt{\sum_{t=1}^n (Q_t - \bar{Q})^2} \sqrt{\sum_{t=1}^n (T_{t+\delta} - \bar{T})^2}} \quad \text{<수식 3-1> 피어슨 상관관계수}$$

<표 3-3>의 형태로 데이터베이스에 저장되어 있는 각 종목에 대한 일별 쿼리볼륨과 주식거래량 및 수정주가를 R로 불러와서 각 기업의 일별 거래량 및 주가와 일별 쿼리볼륨 데이터와의 피어슨 상관관계수를 R에서 제공하는 통계 패키지를 사용하여 구현하였다. 주식 거래 데이터의 경우 주말과 휴일에는 거래 데이터가 없으므로 거래 데이터가 없는 날짜의 데이터는 제외시켰고 분석 기간

동안 결측값이 있는 기업의 경우에도 분석 대상에서 제외하였다. <그림 3-3>은 R의 통계 패키지를 이용하여 피어슨 상관계수를 구현한 소스 코드를 보여준다. ccf() 함수는 R에서 피어슨 상관계수를 계산해주는 함수로 ccf(x, y, lag.max = 시간지연상수) 형태로 사용한다. 여기서 ‘stocktickers’는 분석하는 전체 기업의 수를 나타낸다.

<그림 3-3> R 소스 코드 예시 : 피어슨 상관계수 계산

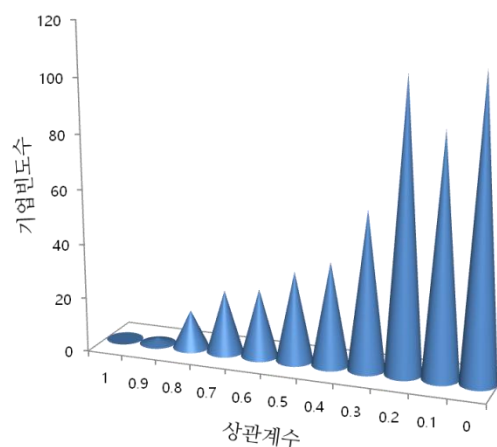
```
for(j in 1:length(stocktickers)){  
  pearson_VoGt[j] <- ccf (VolumeData[,j], GtrendData [,j], lag.max=5)  
  pearson_PrGt [j] <- ccf (PriceData [,j], GtrendData [,j], lag.max=30)  
  pearson_PrVo[j] <- ccf (PriceData [,j], VolumeData [,j], lag.max=30)  
}
```

제 4 장 분석결과

4.1. 퀴리볼륨과 주식 거래량의 상관관계

<그림 4-1>은 피어슨 상관계수 $r(0)$ 값의 절대값 분포를 0.1 단위로 기업의 빈도수로 나타낸 것이다. 상관계수 r 에서 $r < 0.1$ 이면 상관관계가 없다고 보고 $0.1 < r < 0.3$ 이면 낮은 상관관계가, $0.3 < r < 0.5$ 이면 보통의 상관 관계가, $0.5 < r < 0.7$ 이면 높은 상관관계가, $r > 0.7$ 이면 매우 높은 상관관계가 있다고 본다[8]. <그림 4-1>의 퀴리볼륨과 주식거래량의 상관관계 분포를 살펴보면 상관관계 없음이 40%, 낮은 상관관계가 33%, 보통의 상관관계가 14%, 높거나 매우 높은 상관관계가 13%의 분포를 나타낸다. 60% 이상이 상관관계가 있다고 볼 수 있고 27% 이상은 보통의 상관 관계이거나 높은 상관관계가 있다고 볼 수 있다.

<그림 4-1> 퀴리볼륨과 주식거래량의 피어슨 상관계수에 따른 기업 빈도수 분포

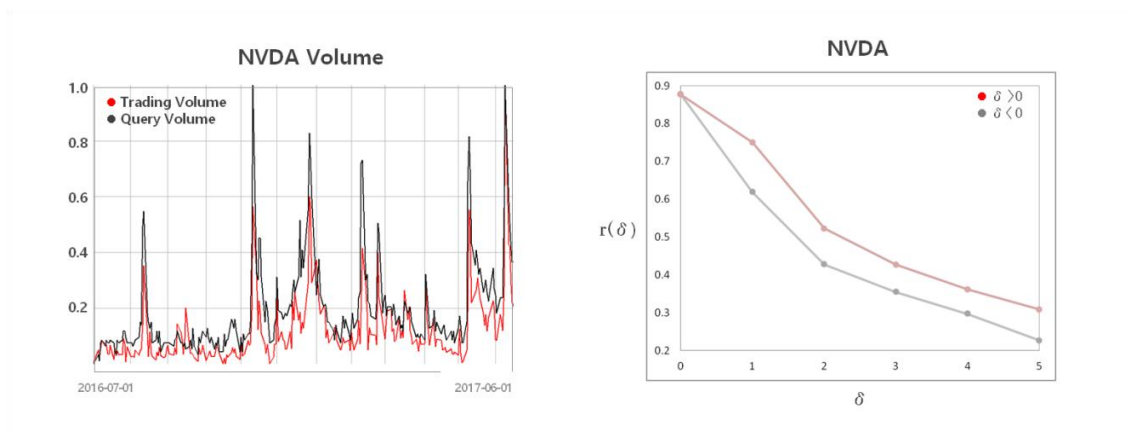


<표 4-1>은 피어슨 상관계수 $r(0)$ 의 값이 높은 20 개의 기업을 나열한 것으로 상관계수가 0.695~0.873 의 값을 나타내며 매우 높은 상관관계가 있음을 확인할 수 있다. <그림 4-2> 및 <그림 4-3>는 20 개의 기업 중 2 개 기업에 대하여 일별 퀴리볼륨과 주식거래량을 시계열 그래프 및 피어슨 상관계수를 시간 지연 상수에 따라 그래프로 나타낸 것이다. 계수 $r(\delta)$ 은 양의 δ 값(적색)이 음의 δ (회색)보다 항상 크며 이것을 통하여 퀴리 볼륨이 주식거래량을 예측하는 경향이 있다고 볼 수 있으며 그러한 기대는 1~4 일까지 나타난다

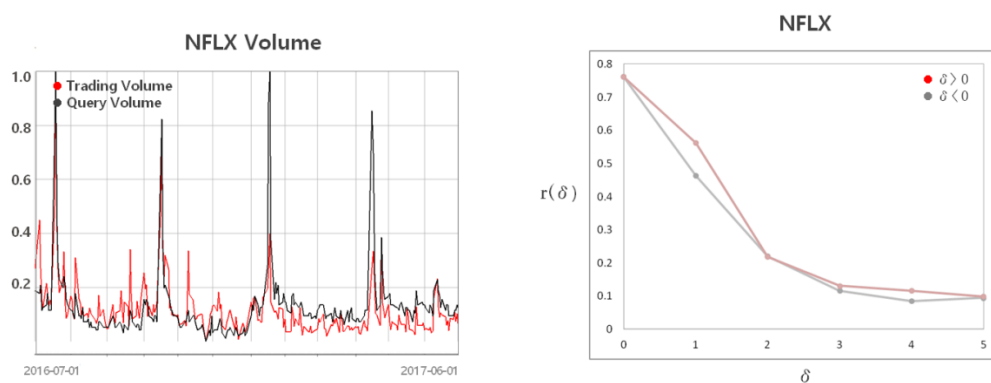
<표 4-1> 상위 20 개 기업에 대한 퀴리볼륨과 주식거래량의 피어슨 상관계수

Num	Company Name	Categoty	Ticker	$r(0)$
1	NVIDIA Corporation	Technology	NVDA	0.873
2	Best Buy Co., Inc.	Consumer Services	BBY	0.808
3	Illumina, Inc.	Capital Goods	ILMN	0.797
4	Kohl's Corporation	Consumer Services	KSS	0.783
5	Amazon.com, Inc.	Consumer Services	AMZN	0.773
6	QUALCOMM Incorporated	Technology	QCOM	0.762
7	Nordstrom, Inc.	Consumer Services	JWN	0.752
8	Netflix, Inc.	Consumer Services	NFLX	0.747
9	Bristol-Myers Squibb Company	Health Care	BMJ	0.746
10	Chipotle Mexican Grill, Inc.	Consumer Services	CMG	0.735
11	First Solar, Inc.	Technology	FSLR	0.735
12	Adobe Systems Incorporated	Technology	ADBE	0.724
13	Wal-Mart Stores, Inc.	Consumer Services	WMT	0.718
14	Red Hat, Inc.	Technology	RHT	0.714
15	Wells Fargo & Company	Finance	WFC	0.707
16	Alexion Pharmaceuticals, Inc.	Health Care	ALXN	0.706
17	Apple Inc.	Technology	AAPL	0.700
18	Starbucks Corporation	Consumer Services	SBUX	0.699
19	Amgen Inc.	Health Care	AMGN	0.695
20	Intel Corporation	Technology	INTC	0.695

<그림 4-2> 쿼리볼륨과 주식거래량의 시계열 그래프 및 상관계수 그래프 (NVIDIA)



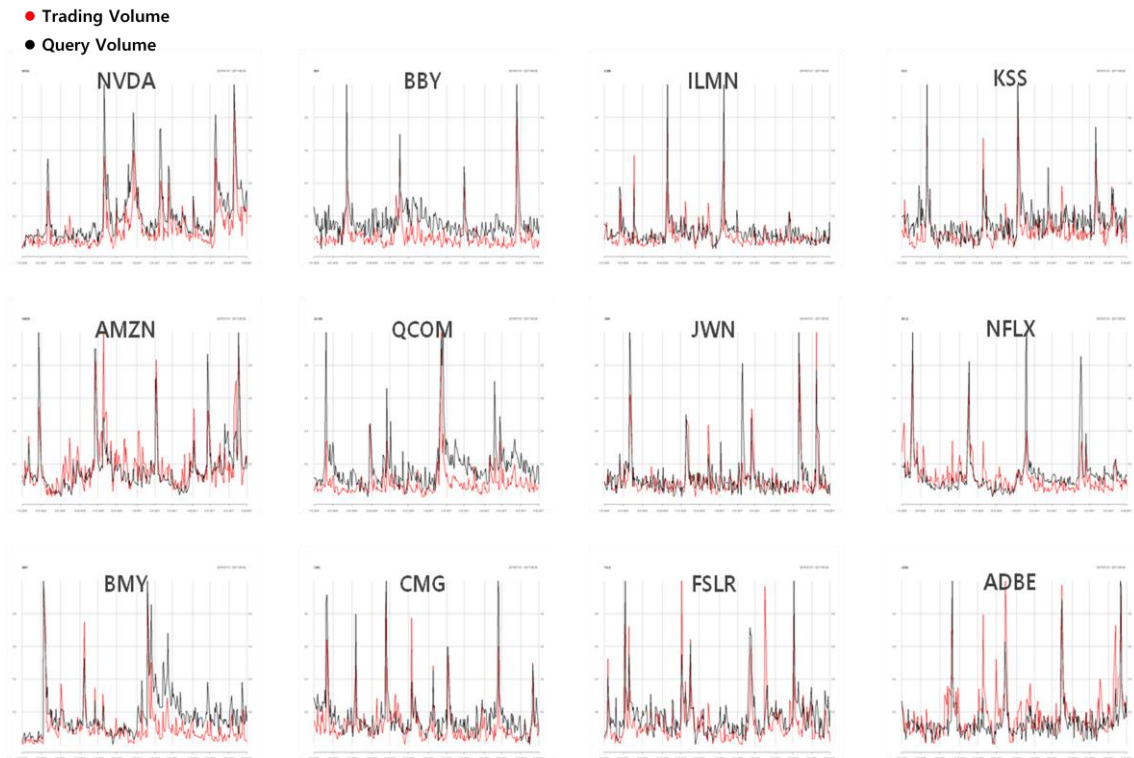
<그림 4-3> 쿼리볼륨과 주식거래량의 시계열 그래프 및 상관계수 그래프(NETFLIX)



<그림 4-4>는 피어슨 상관 관계 계수 $r(0)$ 의 값이 높은 12 개 기업의 쿼리 볼륨과 주식거래량을 시계열 그래프로 나타낸 것이다. 상위 12 개의 기업은 NVIDIA, Best Buy, Illumina, Kohl's corp., Amazon, Qualcomm, Nordstrom, Netflix, Bristol-Myers, Chipotle Mexican Grill, First Solar, Adobe Systems 으로 2 개 기업을 제외하면 모두 consumer service 나 technology 업종에 속하는 기업들이다.

<그림 4-4>의 시계열 그래프를 살펴보면 일별 쿼리볼륨과 주식거래량의 상관관계가 매우 큰 것을 확인할 수 있다.

<그림 4-4> 상위 12 개 기업에 대한 쿼리볼륨과 주식거래량의 시계열 그래프



<표 4-2>는 $r(0)$ 의 값이 높은 20 개 기업의 시간 지연(-5~5 일)에 따른 $r(\delta)$ 를 나타낸 표이다. $r(0)$ 뿐 아니라 1~4 일의 시간 지연이 있는 경우에도 비교적 높은 상관관계가 나타남을 확인할 수 있다.

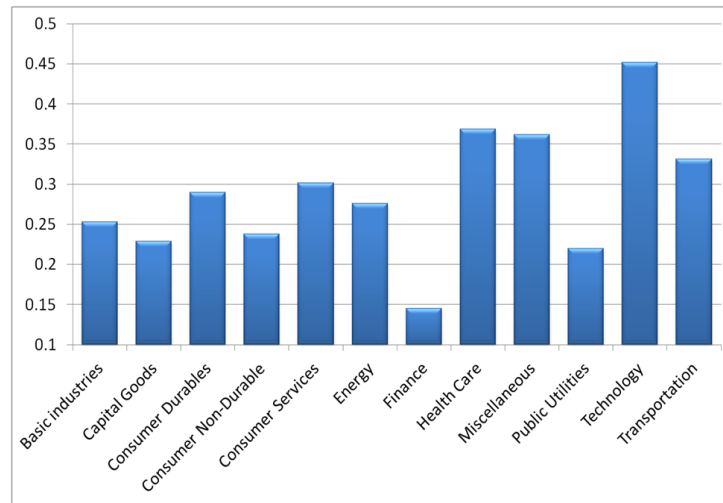
<표 4-2> 시간지연 상수에 따른 쿼리볼륨과 주식거래량의 상관계수

Num	Company Name	Ticker	δ										
			-5	-4	-3	-2	-1	0	1	2	3	4	5
1	NVIDIA Corporation	NVDA	0.216	0.290	0.337	0.423	0.601	0.873	0.736	0.508	0.404	0.345	0.306
2	Best Buy Co., Inc.	BBY	0.040	0.131	0.182	0.271	0.504	0.808	0.395	0.114	0.035	-0.002	-0.061
3	Illumina, Inc.	ILMN	0.046	0.067	0.091	0.148	0.326	0.797	0.421	0.148	0.028	0.025	0.016
4	Kohl's Corporation	KSS	0.026	0.049	0.128	0.260	0.476	0.783	0.383	0.144	0.059	0.004	-0.016
5	Amazon.com, Inc.	AMZN	0.137	0.175	0.164	0.198	0.430	0.773	0.540	0.252	0.189	0.133	0.106
6	QUALCOMM Incorp.	QCOM	0.179	0.335	0.404	0.434	0.551	0.762	0.474	0.219	0.199	0.165	0.081
7	Nordstrom, Inc.	JWN	-0.019	0.019	0.036	0.143	0.361	0.752	0.510	0.132	0.048	-0.010	-0.042
8	Netflix, Inc.	NFLX	0.088	0.084	0.112	0.205	0.455	0.747	0.544	0.207	0.121	0.107	0.104
9	Bristol-Myers Squibb	BMJ	0.118	0.217	0.210	0.245	0.381	0.746	0.326	0.146	0.103	0.138	0.065
10	Chipotle Mexican Grill,	CMG	0.020	0.070	0.021	0.164	0.359	0.735	0.494	0.097	0.056	0.051	0.049
11	First Solar, Inc.	FSLR	0.123	0.182	0.156	0.209	0.414	0.735	0.481	0.144	0.054	0.099	0.037
12	Adobe Systems Incorp.	ADBE	0.126	0.137	0.208	0.291	0.423	0.724	0.585	0.236	0.134	0.094	0.022
13	Wal-Mart Stores, Inc.	WMT	0.077	0.095	0.112	0.135	0.378	0.718	0.241	0.119	0.105	0.058	0.034
14	Red Hat, Inc.	RHT	0.092	0.083	0.086	0.186	0.452	0.714	0.581	0.209	0.153	0.073	0.051
15	Wells Fargo & Company	WFC	0.278	0.299	0.344	0.401	0.505	0.707	0.532	0.344	0.271	0.254	0.193
16	Alexion Pharmaceuticals,	ALXN	0.222	0.268	0.356	0.400	0.587	0.706	0.440	0.339	0.275	0.209	0.110
17	Apple Inc.	AAPL	0.034	0.069	0.068	0.192	0.443	0.700	0.483	0.142	0.131	0.198	0.195
18	Starbucks Corporation	SBUX	0.062	0.150	0.264	0.306	0.403	0.699	0.605	0.283	0.195	0.129	0.069
19	Amgen Inc.	AMGN	0.037	0.036	0.042	0.089	0.227	0.695	0.351	0.175	0.060	0.031	-0.050
20	Intel Corporation	INTC	0.041	0.060	0.133	0.209	0.380	0.695	0.532	0.182	0.135	0.098	0.038

상기의 결과들을 정리하고 통계적으로 확인해 보면, 대상기업 505 개 기업 중 결측값이 있는 9 개의 기업을 제외한 분석 기업 496 개중에 피어슨 상관관계의 p-value 가 0.05 이하인 기업의 수가 307 개로 나타났다. 기업의 61.9%가 95% 신뢰구간에서 상관 관계가 있다고 볼 수 있으며 307 개 기업의 상관계수 $r(0)$ 의 평균과 표준편차가 0.3339 ± 0.1866 로 상관관계가 높은 기업이 다수를 차지하고 상위 20 개 기업의 상관 계수 $r(0)$ 의 평균값이 0.7434 로 매우 높은 상관 관계를 나타낸다.

<그림 4-5>는 기업의 업종에 따라 상관계수를 나타낸 표이다. Technology, Miscellaneous, Health Care 업종 순으로 상관관계가 높게 나타난다. Technology 업종의 경우 NVIDIA, QUALCOMM, Adobe System, Red Hat, Apple, Intel, Cisco 등이 있으며 반도체, 통신, 컴퓨터, 전자기기등의 첨단 기술을 취급하는 기업들이 포함되어 있다. Miscellaneous 업종은 The Priceline Group, Akamai Technologies, PayPal, Western Union, Xerox, Mastercard, Visa 등의 기업들이 있으며 온라인여행, 클라우드 서비스, 전자상거래, 금융/통신, 사무기기등을 취급하는 다양한 기업들이 포함되어 있다. Health Care 업종은 Bristol-Myers, Alexion, Amgen, Biogen, Mallinckrodt, Gilead Sciences 등의 기업들이 있으며 제약, 바이오, 의료기기등을 취급한다. 상관계수가 높은 상위 20 개 기업을 업종별로 확인해 보면 Consumer Service 군이 8 개 기업, Technology 군이 7 개 기업으로 전체의 75%를 차지하며 대부분이 일반 소비자들에게 인지도가 매우 높은 기업들이다. 상기의 내용들을 종합해 보면, 쿼리 볼륨과 주식거래량의 상관관계가 높은 기업들은 일반 소비자들에게 인지도가 높은 첨단 기술 기업이거나 제약, 바이오, 의료기기등 미래 성장성이 높은 기업 및 소비자들이 쉽게 접할 수 있는 금융, 통신, 전자상거래, 유통등의 서비스 기업들임을 알 수 있다. 이러한 기업들의 경우 일반 투자자가 티커(ticker)를 검색한 행위가 주식 거래로 이어지는 경우가 많다고 할 수 있으며 특히 브랜드 인지도가 높은 Consumer Service 나 Technology 기업들의 경우 이러한 경향이 더 높아진다고 볼 수 있다.

<그림 4-5> 업종에 따른 상관관계수 비교 (쿼리볼륨 vs. 주식거래량)



4.2. 쿼리 볼륨과 주가와의 상관관계

<표 4-3>은 쿼리 볼륨과 주가의 피어슨 상관관계수 $r(0)$ 의 값이 높은 상위 20 개의 기업을 정리한 표이다. 쿼리볼륨과 주가의 상관관계는 기업에 따라 (+) 및 (-)의 상관 관계를 나타낸다. 주가의 경우 거래량과 달리 상승 및 하강의 두 가지 방향성을 가지고 움직이므로 이런 양상을 나타낸다.

<표 4-3> 상위 20개 기업에 대한 쿼리볼륨과 주가의 피어슨 상관계수

Num	Company Name	Category	Ticker	r(0)
1	L Brands, Inc.	Consumer Services	LB	-0.911
2	Deere & Company	Capital Goods	DE	-0.847
3	Mastercard Incorporated	Miscellaneous	MA	-0.816
4	XL Group Ltd.	Finance	XL	0.779
5	Michael Kors Holdings Limited	Consumer Non-Durable	KORS	-0.789
6	United Continental Holdings, I	Transportation	UAL	-0.786
7	Norfolk Souther Corporation	Transportation	NSC	-0.803
8	Signet Jewelers Limited	Consumer Services	SIG	0.777
9	CSX Corporation	Transportation	CSX	-0.737
10	Ralph Lauren Corporation	Consumer Non-Durable	RL	-0.721
11	Reynolds American Inc	Consumer Non-Durable	RAI	-0.787
12	Ulta Beauty, Inc.	Consumer Services	ULTA	0.767
13	U.S. Bancorp	Finance	USB	0.762
14	Harris Corporation	Capital Goods	HRS	0.757
15	Dow Chemical Company (The)	Basic Industries	DOW	0.746
16	Rockwell Collins, Inc.	Capital Goods	COL	-0.742
17	Humana Inc.	Health Care	HUM	0.754
18	Zions Bancorporation	Finance	ZION	-0.694
19	Intercontinental Exchange Inc.	Finance	ICE	0.735
20	NRG Energy, Inc.	Public Utilities	NRG	-0.721

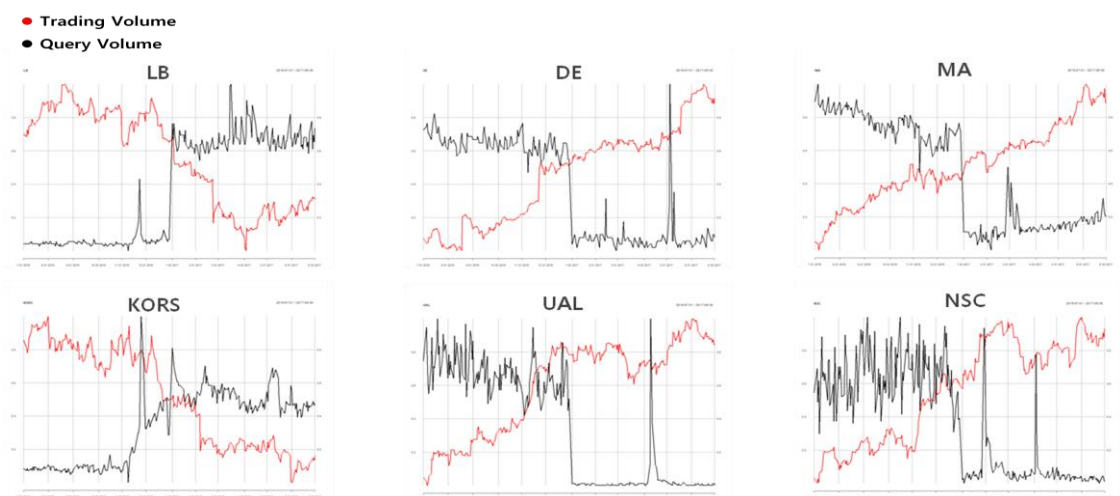
<그림 4-5> 및 <그림 4-6>는 피어슨 상관 관계 계수 $r(0)$ 가 높은 값부터 12 개의 기업 중 (+) 상관관계를 가지는 6 개 기업과 (-) 상관관계를 가지는 6 개 기업의 쿼리 볼륨과 주가를 시계열 그래프로 나타낸 것이다. 상위 12 개의 기업은 L Brands, Inc., Deere & Company, Mastercard, XL Group, Michael Kors Holdings, United Continental, Norfolk Souther, Signet Jewelers, CSX, Ralph Lauren , Reynolds American, Ulta Beauty 로 Consumer service 군이 3 개, Transportation 군이 3 개, Consumer Non-Durable 군이 3 개, 기타 업종이 3 개로 다양한 업종에 골고루 분포되어 있다. <그림 4-5> 및 <그림 4-6>의 시계열

그래프를 살펴보면 일별 쿼리볼륨과 주가의 상관관계가 매우 높은 것을 확인할 수 있다.

<그림 4-6> (+)상관관계를 가지는 상위 6개 기업에 대한 쿼리볼륨과 주가의 시계열 그래프



<그림 4-7> (-)상관관계를 가지는 상위 6개 기업에 대한 쿼리볼륨과 주가의 시계열 그래프



<표 4-4>는 피어슨 상관계수 $r(0)$ 의 값이 높은 상위 20 개 기업의 시간 지연(-30~30 일)에 따른 $r(\delta)$ 를 나타낸 표이다.

<표 4-4> 시간지연 상수에 따른 쿼리볼륨과 주가의 상관계수

Num	Company Name	Ticker	δ												
			-30	-25	-20	-15	-10	-5	0	5	10	15	20	25	30
1	L Brands, Inc.	LB	-0.600	-0.653	-0.709	-0.783	-0.842	-0.887	-0.911	-0.905	-0.881	-0.860	-0.831	-0.797	-0.764
2	Deere & Company	DE	-0.638	-0.685	-0.714	-0.745	-0.785	-0.817	-0.847	-0.814	-0.777	-0.736	-0.681	-0.626	-0.572
3	Mastercard Incorp.	MA	-0.475	-0.524	-0.589	-0.652	-0.705	-0.758	-0.816	-0.779	-0.741	-0.702	-0.672	-0.658	-0.638
4	XL Group Ltd.	XL	0.537	0.584	0.634	0.664	0.717	0.745	0.779	0.732	0.704	0.682	0.684	0.655	0.603
5	Michael Kors Holdings Limited	KORS	-0.440	-0.481	-0.534	-0.602	-0.665	-0.735	-0.789	-0.797	-0.786	-0.808	-0.808	-0.781	-0.750
6	United Continental Holdings	UAL	-0.680	-0.716	-0.753	-0.781	-0.776	-0.782	-0.786	-0.731	-0.676	-0.606	-0.548	-0.497	-0.439
7	Norfolk Souther Corp.	NSC	-0.736	-0.751	-0.785	-0.800	-0.825	-0.813	-0.803	-0.764	-0.719	-0.669	-0.610	-0.538	-0.478
8	Signet Jewelers Limited	SIG	0.353	0.424	0.507	0.591	0.671	0.731	0.777	0.775	0.757	0.738	0.706	0.657	0.615
9	CSX Corporation	CSX	-0.567	-0.595	-0.626	-0.653	-0.695	-0.732	-0.737	-0.737	-0.741	-0.709	-0.668	-0.617	-0.551
10	Ralph Lauren Corp.	RL	-0.429	-0.540	-0.603	-0.622	-0.690	-0.710	-0.721	-0.697	-0.615	-0.556	-0.509	-0.445	-0.420
11	Reynolds American Inc	RAI	-0.584	-0.640	-0.697	-0.741	-0.774	-0.781	-0.787	-0.763	-0.753	-0.736	-0.689	-0.652	-0.621
12	Ulta Beauty, Inc.	ULTA	0.419	0.476	0.527	0.581	0.638	0.702	0.767	0.735	0.723	0.732	0.719	0.705	0.715
13	U.S. Bancorp	USB	0.592	0.656	0.682	0.706	0.743	0.755	0.762	0.705	0.668	0.624	0.574	0.515	0.464
14	Harris Corporation	HRS	0.745	0.758	0.764	0.771	0.773	0.776	0.757	0.683	0.627	0.577	0.529	0.479	0.441
15	Dow Chemical	DOW	0.540	0.596	0.634	0.672	0.689	0.704	0.746	0.711	0.704	0.709	0.703	0.672	0.630
16	Rockwell Collins, Inc.	COL	-0.584	-0.632	-0.663	-0.672	-0.709	-0.711	-0.742	-0.716	-0.703	-0.696	-0.680	-0.681	-0.664
17	Humana Inc.	HUM	0.563	0.595	0.616	0.676	0.721	0.746	0.754	0.716	0.673	0.647	0.634	0.620	0.586
18	Zions Ban corporation	ZION	-0.513	-0.541	-0.573	-0.617	-0.653	-0.660	-0.694	-0.668	-0.650	-0.627	-0.622	-0.594	-0.545
19	Intercontinental Exch.	ICE	0.445	0.490	0.535	0.604	0.643	0.686	0.735	0.686	0.646	0.610	0.572	0.527	0.501
20	NRG Energy, Inc.	NRG	-0.487	-0.554	-0.599	-0.627	-0.642	-0.688	-0.721	-0.735	-0.731	-0.699	-0.660	-0.635	-0.595

대상기업 505 개 기업 중 결측값이 있는 9 개의 기업을 제외한 분석 기업 496 개중에 쿼리볼륨과 주가의 피어슨 상관관계의 p-value 가 0.05 이하인 기업의 수가 341 개로 나타났다. 기업의 68.7%가 95% 신뢰구간에서 상관 관계가 있다고 볼 수 있으며 341 개 기업의 상관계수 $r(0)$ 의 평균과 표준편차가 0.3548 ± 0.1852 로 상관 관계가 높은 기업이 다수를 차지하고 상위 20 개 기업의 상관계수

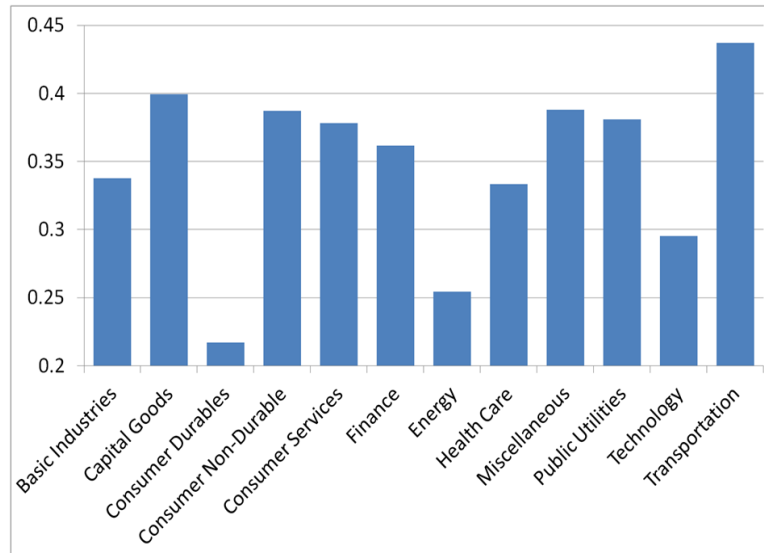
$r(0)$ 는 0.721~0.911 로 매우 높은 상관 관계를 나타낸다. 퀴리 볼륨과 주가의 상관관계가 퀴리볼륨과 주식거래량의 상관관계보다 전체나 상위 20 개 기준으로 했을 때 모두 더 높다. 이것은 이전 문헌에서 확인하지 못했던 결과이다. Preis et al.(2010)은 주간 단위로 퀴리볼륨과 주식거래량의 낮은 상관관계는 확인했지만 퀴리볼륨과 주가의 상관관계는 확인하지 못했고 Bordino et al.(2012)도 일별 퀴리볼륨과 주식거래량의 높은 상관관계는 확인했지만 퀴리볼륨과 주가의 상관관계는 확인하지 못했다. 또한 최대 1~4 일의 지연시간을 가지고 있는 퀴리 볼륨과 주식거래량의 상관관계와는 달리 30 일까지의 시간 지연에도 상관관계가 높게 유지되는 것은 특이한 점이다. 이것은 주식거래량과는 달리 주가의 경우 일별 가격 변동의 제한과 다양한 요인들로 인한 상, 하방 경직성으로 인하여 상대적으로 일별 편차가 크지 않기 때문인 것으로 보인다. 또한 퀴리 볼륨과 주가의 상관관계는 퀴리 볼륨이 증가함에 따라 주가가 상승하는 양의 상관관계뿐 아니라 퀴리 볼륨이 증가함에 따라 주가가 하락하는 음의 상관관계가 유사한 비율로 나타나며 이는 퀴리 볼륨이 증가하는 요인으로 긍정성에 의한 것과 부정성에 의한 것이 함께 나타날 수 있음을 보여준다. 퀴리 볼륨과 주가의 상관계수가 1 개월 이상의 시간 지연에도 높은 값을 유지하는 것을 볼 때 주가의 경우 주식거래량과는 달리 퀴리 볼륨에 따라 일일 단위의 짧은 시간에 따라 단기적으로 변동하지 않고 최소 1 개월 이상의 추세를 가지고 움직인다고 볼 수 있다.

<그림 4-6>은 기업의 업종에 따라 퀴리볼륨과 주가의 상관계수를 나타낸 표이다. Transportation, Capital Goods, Consumer Non-Durable, Consumer 업종 순으로 상관관계가 높게 나타난다. 퀴리볼륨과 주식거래량의 상관계수와는 다르게 업종별 편차가 크지 않은 편이다. 퀴리볼륨과 주식거래량의 상관계수의 경우

0.3 이상인 업종이 4 개이지만 퀴리볼륨과 주가의 상관관계수의 경우 0.3 이상인 업종이 9 개로 대부분의 업종에 골고루 분포되어 있다.

Transportation 업종의 경우 Southwest Airlines, FedEx, Expeditors, Union Pacific, Kansas City 등이 있으며 주로 항공, 철도 서비스를 제공하는 기업들이다. Capital Goods 업종은 Harris, Caterpillar, Lockheed Martin, Mettler-Toledo, Boeing 등의 기업들이 있으며 산업용기계, 건설, 방산, 우주항공 등을 취급하는 다양한 기업들이 포함되어 있다. Consumer Non-Durable 업종은 Leucadia, Altria Group, Mattel, Hershey 등의 기업들이 있으며 농수산물, 농업, 장난감, 특수식품 등을 취급한다. 상기의 내용들을 종합해 보면, 퀴리볼륨과 주가의 상관 관계가 높은 기업들은 퀴리볼륨과 주식거래량의 상관관계가 높은 기업과 비교하여 상대적으로 일반소비자에게 인지도가 높지 않은 기업들이며 다소 경기에 둔감한 업종들이다. 가장 상관관계수가 낮은 Consumer Durables, Energy, Technology 업종들이 경기에 민감한 업종들이며 이를 통하여 퀴리볼륨과 주가와 상관관계가 높은 기업들은 상대적으로 경기에 둔감한 특색을 나타내는 것으로 보여진다. 그러나 퀴리볼륨과 주가의 상관 관계가 높은 기업들의 업종별 특색은 결과론적인 추론으로 검증에 의해 확인된 사항은 아니다. 퀴리볼륨과 주식거래량 및 주가와 상관관계가 업종에 따라 유의차가 있는지는 향후 연구 과제로 좀 더 면밀하게 확인하는 것이 필요하다.

<그림 4-8> 업종에 따른 상관계수 비교 (쿼리볼륨 vs. 주가)



제 5 장 결 론

5.1. 연구결과

본 연구는 피어슨 상관관계 분석을 이용하여 구글트렌드의 퀴리볼륨과 주식거래량 및 주가의 상관관계를 확인하였다. 이를 위해 S&P500 에 속해 있는 505 개 기업의 퀴리볼륨과 주식 데이터를 구글트렌드와 Yahoo Finance 에서 수집하고 분석하였다. S&P500 에 속한 505 개 기업을 대상으로 각 기업의 ‘티커(ticker)’를 사용한 퀴리볼륨과 주식거래량 및 주가에 대하여 2016 년 7 월부터 2017 년 6 월까지 일년간의 일별 데이터를 분석하였다.

다량의 데이터 수집과 분석을 위해서 R 프로그램을 사용하였고 R 의 통계 패키지를 이용하여 피어슨 상관관계 분석을 실시하였다. 분석 결과 퀴리볼륨과 주식거래량의 경우 분석기업 중 62%가 상관관계가 있는 것으로 나타났고 퀴리볼륨과 주가의 경우 분석기업 중 69%가 상관관계가 있는 것으로 밝혀졌다. 상위 20 개 기업의 퀴리볼륨과 주식거래량 및 주가의 상관계수를 확인했을 때 0.70~0.91 의 분포로 매우 높은 상관관계를 확인할 수 있었다. 시간지연상수 δ 를 확인했을 때 퀴리볼륨과 주식거래량의 경우 1~4 일 동안 상관관계를 나타내며 퀴리볼륨과 주가의 경우 30 일까지 높은 상관관계를 보여 주었다. 이것을 통하여 퀴리볼륨이 주식거래량 및 주가를 예측하는 경향이 있다고 볼 수 있으며 주식거래량의 경우 그러한 기대가 1~4 일까지 나타나고 주가의 경우 30 일까지도 나타난다고 볼 수 있다.

본 연구는 다양한 업종을 포함하고 있는 S&P500 기업을 대상으로 쿼리볼륨과 주식거래량 및 주가와의 일별 데이터를 사용하여 높은 상관관계를 확인하였고 특별히 쿼리볼륨과 주가와의 일별 데이터를 사용하여 높은 상관관계를 추가적으로 확인하였다는 데 의의가 있다. 또한 이전 연구에서는 확인하지 못했던 업종에 따른 쿼리볼륨과 주식거래량 및 주가와의 상관관계의 특성을 살펴본 것에 의의가 있다.

이 연구가 시사하는 바는 웹검색 트래픽을 나타내는 수단 중 하나인 구글트렌드의 쿼리볼륨의 변동성을 통해 주식거래량과 주가의 변동성을 예측하는데 실질적인 도움을 받을 수 있다는 것이다. 특별히 개별 기업에 대한 주식 시장의 급격한 변동성 등을 예측할 수 있게 되어 금융기관이나 투자자 등이 위기 상황을 인지하고 적절한 대응을 할 수 있다는 점에서 실질적인 도움을 받을 수 있을 것이다.

5.2. 추후연구과제

본 연구의 한계점과 향후 연구방향을 제시하면 다음과 같다.

첫째, 구글트렌드의 쿼리볼륨과 주식거래량 및 주가와의 상관관계 분석을 통하여 각각의 관계에 높은 상관관계가 있다는 것을 확인했지만 주식거래량 및 주가에 영향을 미치는 인자가 무엇인지는 밝히지 못하였다. 또한 쿼리볼륨과 주식거래량과의 상관관계가 1~4 일까지 나타나는 것과는 다르게 쿼리볼륨과 주가와의 상관관계의 경우 30 일까지 높은 상관관계를 나타내는 이유를 명확하게 설명하지는 못하였다.

둘째, 기업에 따라서 다양한 상관관계를 나타내고 있지만 상관관계가 기업에 따라 다르게 나타나는 원인에 대해서는 확인하지 못하였다. 또한 업종에 따라 상관관계의 차이를 확인하였지만 정성적으로 분석하였을 뿐 통계적인 절차에 의해 정량화하지는 못하였다.

향후, 구글트렌드의 쿼리볼륨이 주식 거래와 주가에 미치는 영향의 요인을 확인하고 기존에 활발하게 연구되고 있는 시계열 예측 모델에 결합한다면 정확성을 높일 수 있는 새로운 모델을 제안할 수 있을 것이다. 또한 구글트렌드의 쿼리볼륨에 더하여 SNS 나 뉴스, 인터넷 게시판 등의 텍스트들을 수집하여 텍스트 마이닝을 수행한 결과를 결합한다면 더 정확하고 정교한 결과를 도출할 수 있을 것으로 생각한다.

참고문헌

- [1] Sam De Smet, Onder Leiding Van, Prof. Dirk Van den Poel, 2016, Predictive power of Google Trends analysis on Euronext Brussels stock performance
- [2] Yossi Matias, Niv Efron, Yair Shimshoni, 2009, On the predictability of Search Trends
- [3] Choi H, Varian H, 2009, Predicting the present with google trends
- [4] Ginzberg J, Mohebi M, Patel R, Brammer L, Smolinski M, et al., 2009, Detecting influenza epi-demics using search engine query data
- [5] Lei Shi, Neeraj Agarwal, Ankur Agrawal, Rahul Garg, 2012, Predicting US Primary Elections with Twitter
- [6] Tobias Preis, Daniel Reith, H.Eugene Stanley, 2010, Complex dynamics of our economic life on different scales: insights from search engine query data
- [7] Ilaria Bordino, Stefano Battiston, Guido Caldarelli, Matthieu Cristelli, Antti Ukkonen, Ingmar Weber, 2012, Web Search Queries Can Predict Stock Market Volumes
- [8] Will G Hopkins, 2002, A New View of Statistics

ABSTRACT

A Study on the Correlations between Query Volume of Google Trend, Stock Trading Volume and Stock Price

Lee, Sang Il

Management of Technology

Sungkyunkwan University

In this study, correlation analysis was used to determine the relationship between query volume, stock trading volume and stock price. Assuming that the investors search the web for information of interested companies and it leads to actual investment, we analyzed the relationship between the query volume, the stock trading volume and stock price for the companies belonging to the S&P500.

For the 505 companies belonging to S&P500, we collected data for the last one year from 'Google Trends' and 'Yahoo Finance', and analyzed them with open source program R. Analyzing the relationship between query volume and stock trading volume, it shows that 62% of analysts have a correlation. Analyzing the relationship between query volume and stock price, it shows that 69% of analysts have a correlation.