



# Stock Price Data Analysis

EunJu Lee, Jisu Kim  
Department of Statistics, Pusan National University

## Introduction

### Objective

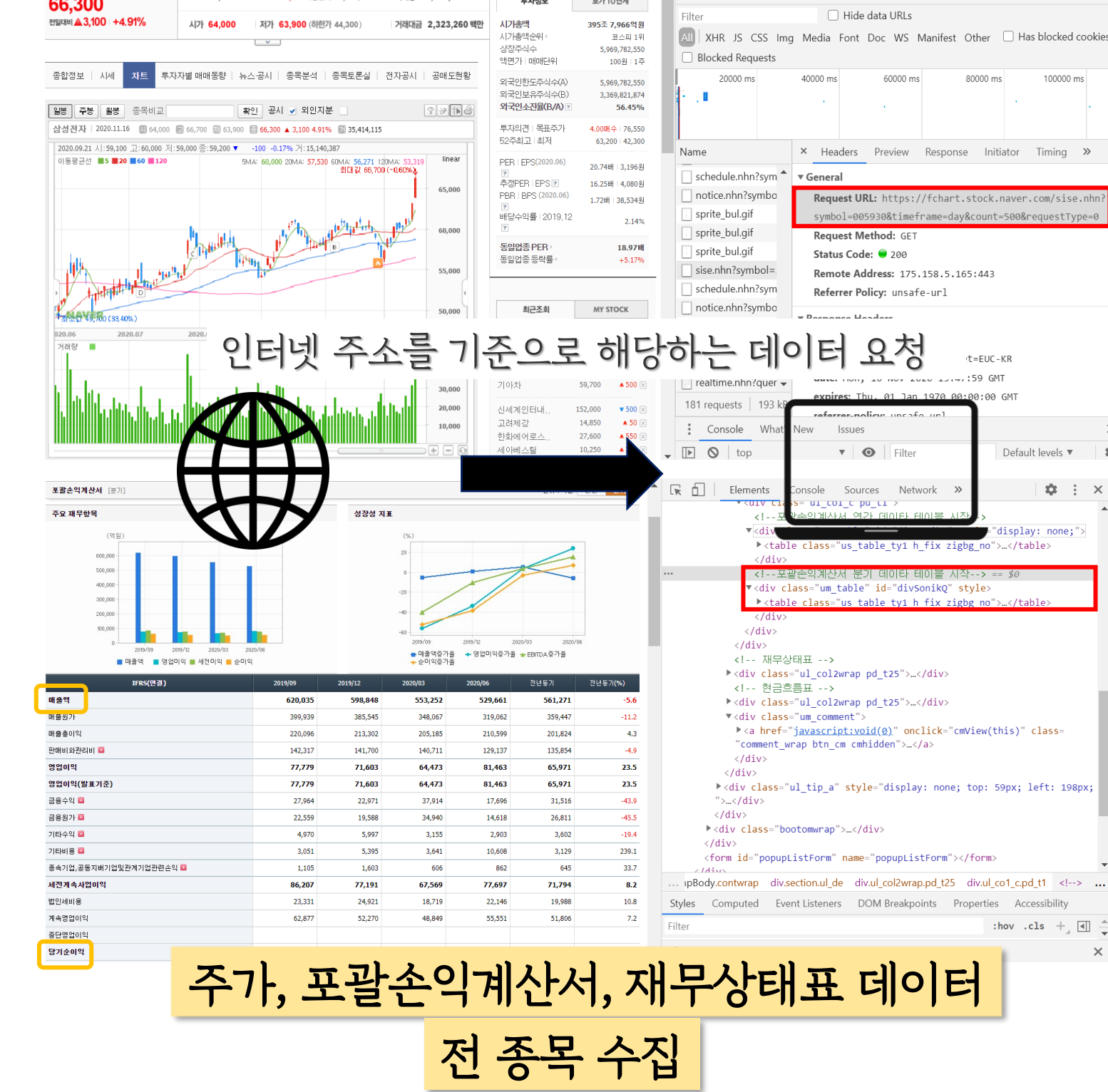
2020년 코로나19로 인해 주식시장은 대부분 폭락을 겪었으며 이로 인해 젊은 연령층의 소액투자자 일명 “개미군단”이 크게 증가하고 있다. 이에 따라 코로나로 인해 변동이 심했을 것이라 추정되는 37지 산업군의 각 대표 기업 5개에 대한 주가와 매출액, 순이익 데이터를 직접 수집하여 분석한다. 먼저 과거의 주가 데이터를 통해 2020년 주가 예측값과 실제값의 차이를 보며 2020년 주가 변동의 특징을 알아본 후, 현재 주식 투자 종목 선정 시 무엇을 고려해야 할 지 기업의 대표 성과 지표인 매출액, 순이익 데이터와 주가의 연관성을 확인함으로써 고민해본다.

### Method

- \* 데이터는 네이버 금융, FNGUIDE에서 수집
- [데이터 수집] - 크롤링(Get 방식)을 이용하여 데이터를 수집한 후 EDA 및 시각화에 대해 모델화 한다.
- [데이터 분석] - ① ARIMA, GARCH 모델을 이용하여 시계열분석을 함으로써 2020년 데이터를 예측해 본다.
- ② 상관계수와 단순선형회귀모델을 이용하여 주가 증감률과 매출액, 순이익 증감률에 인과성 또는 상관성이 존재하는지 확인한다.

## Result : Data Collection & EDA

### GET 방식의 크롤링

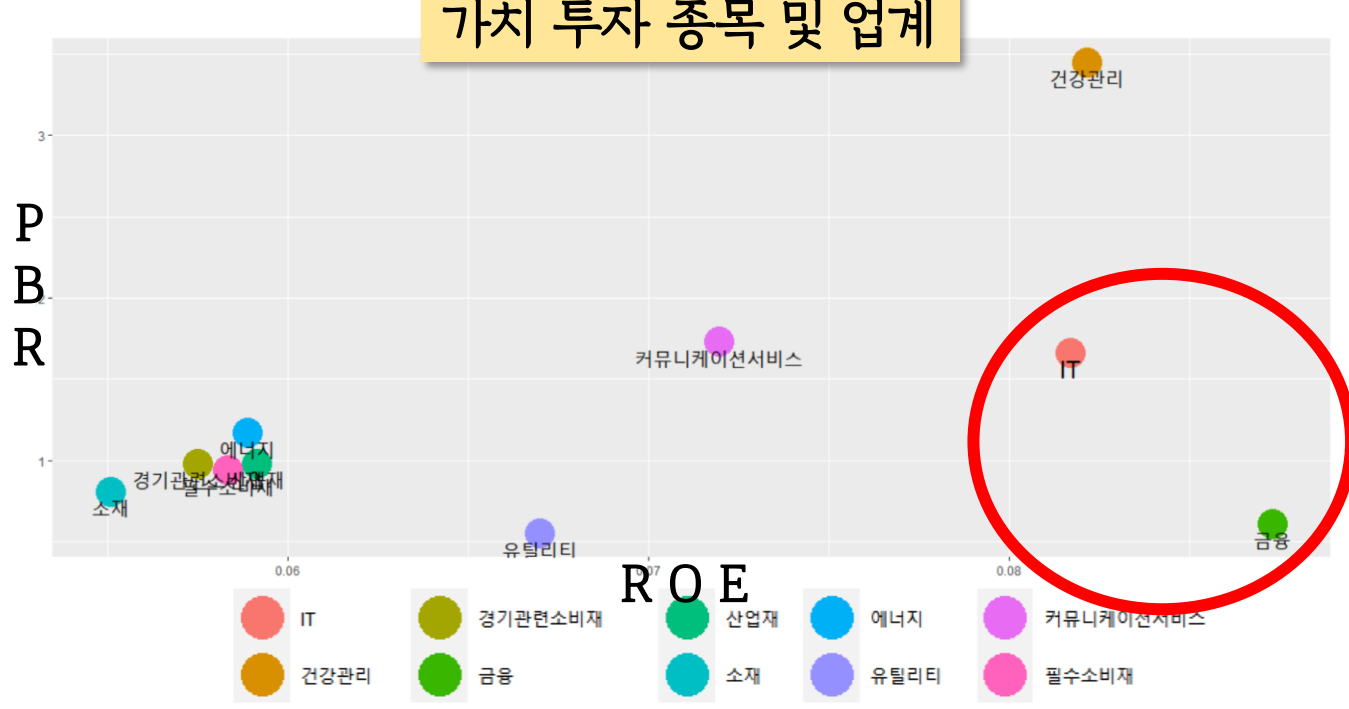


### EDA

#### ① 가치지표에 따른 Best5 종목 및 업계

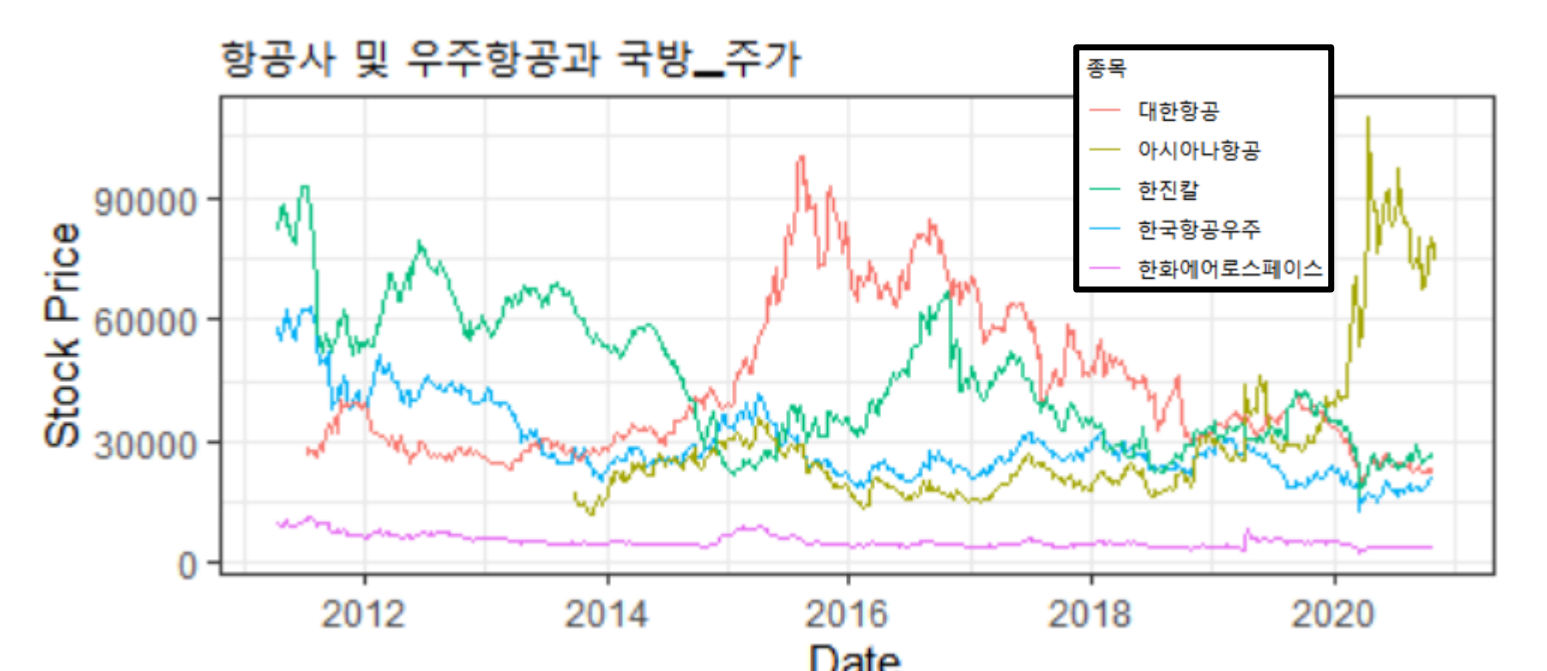
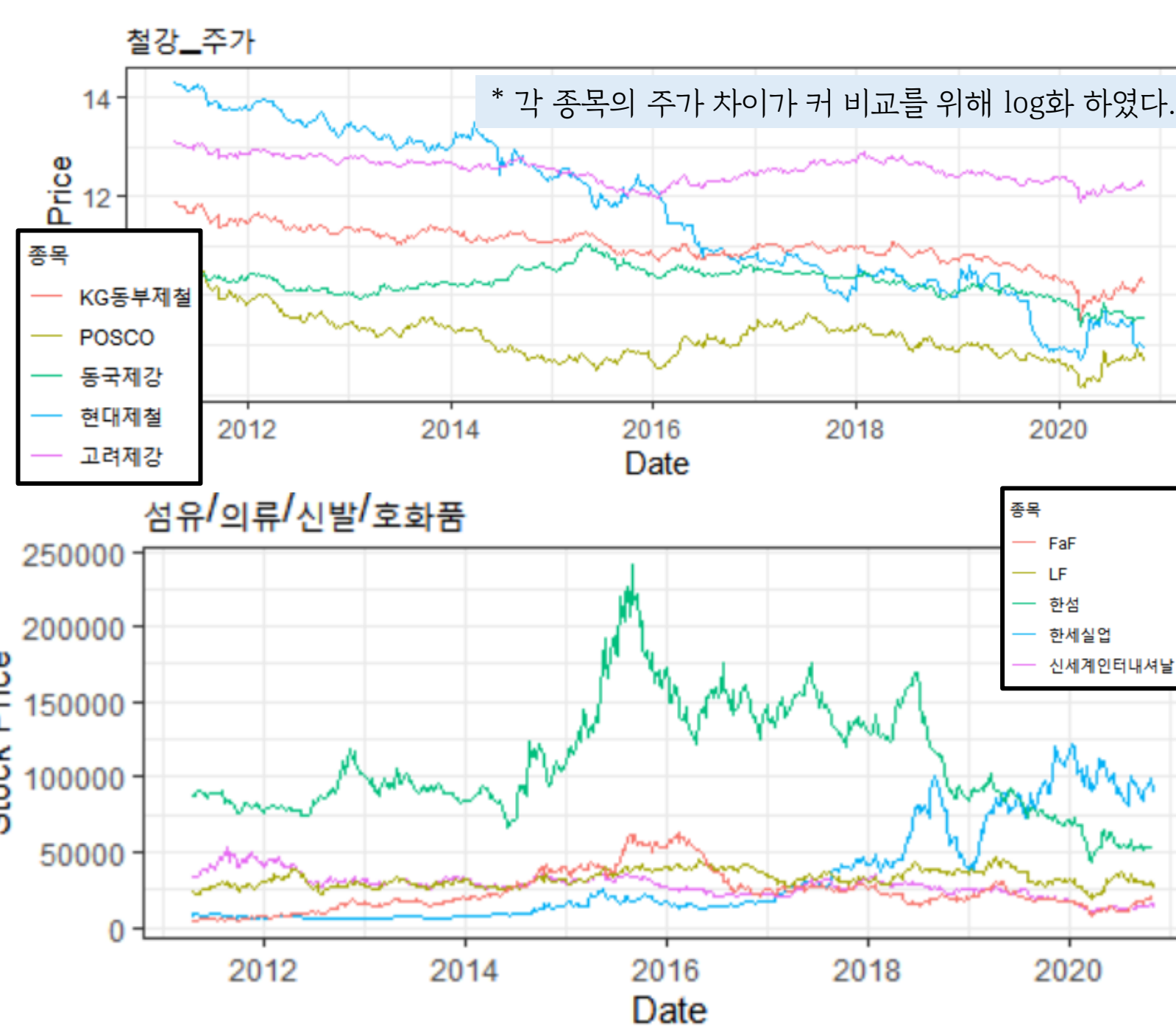
\* PBR(주당순이익) : 주가/순자산 => PBR ↓ : 기업의 가치 > 주가  
PER(주가순자산비율) : 주가/주당순이익 => PER ↓ : 기업의 가치 > 주가  
ROE : PBR/PER => ROE ↑ : 자본으로 많은 이익을 도출

종목명	PBR	PER	ROE
사조대림	0.41	1.17	0.3504
메디알스	0.61	1.27	0.4803
원익	0.58	1.41	0.4113
KG ETS	0.67	1.62	0.4136
푸드웰	0.68	1.67	0.4072



#### ② 주가 데이터 시각적 비교

\* 각 산업군의 주요 지표라고 할 수 있는 가동률 지수에 대해 Boxplot을 이용하여 이상치로 판단된 데이터 중 2020년 데이터에 해당하는 산업군이 코로나로 인해 수요의 변동이 크게 나타났다고 추정한다. 이에 따라 선택된 산업군(철강, 섬유/의류/신발/호화품, 항공사 및 우주항공과 국방)의 대표 기업 5개를 KOSPI 200을 중심으로 선택하여 주가 변동은 어떠한지 비교한다.



철강, 섬유/의류/신발/호화품, 항공사 산업군,  
2020년 가동률 지수가 급격히 하락

- => 주가 또한 변동이 있는가?  
2020년을 기준으로 주가가 하락하나 **전반적으로 다시 회복세**  
=> 더욱 자세히 분석해보자.
1. 2020년 주가의 예측값과의 비교를 통해 2020년 주가의 특징 확인
  2. 기업의 성과 지표인 매출액, 순이익과 주가의 연관성 분석

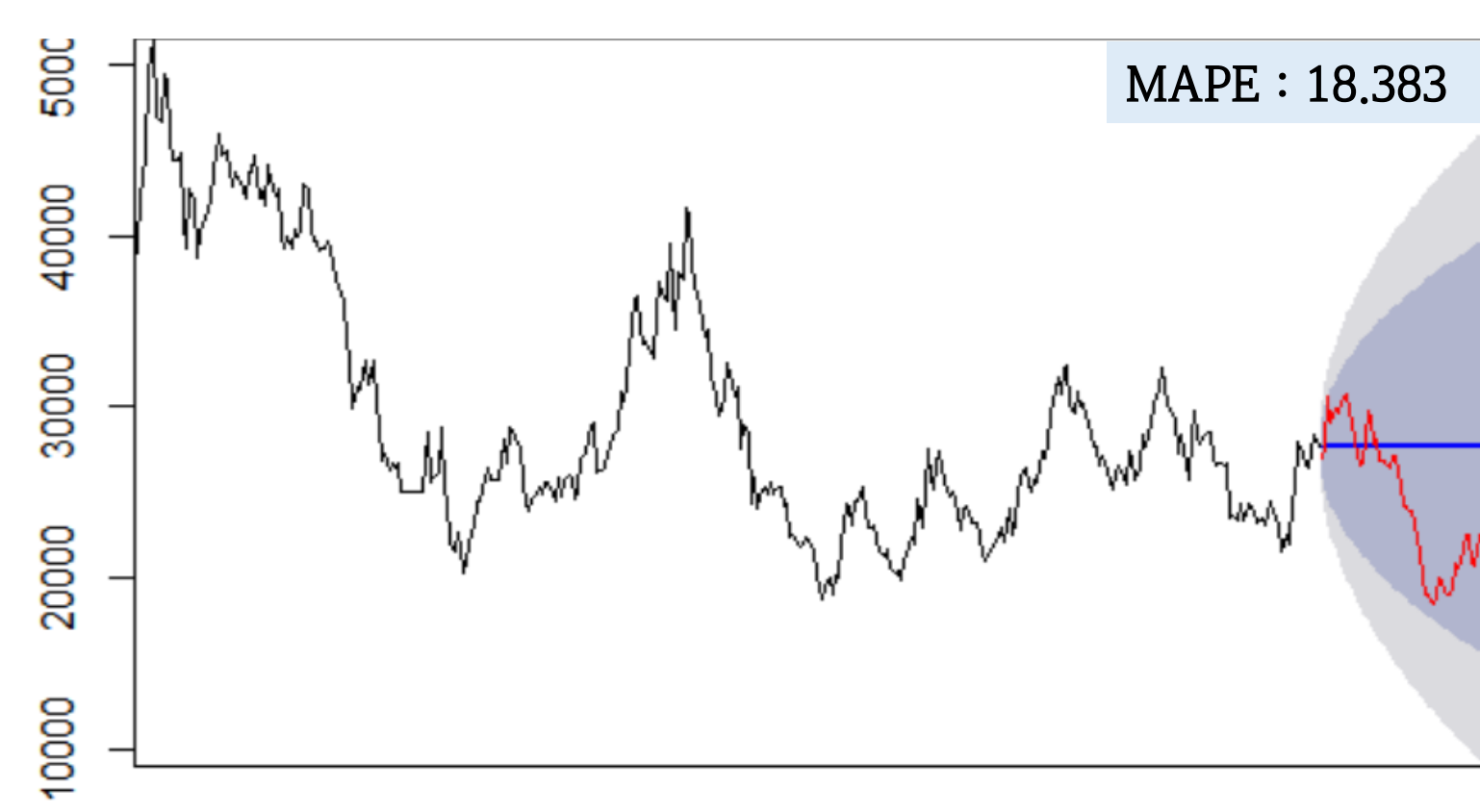
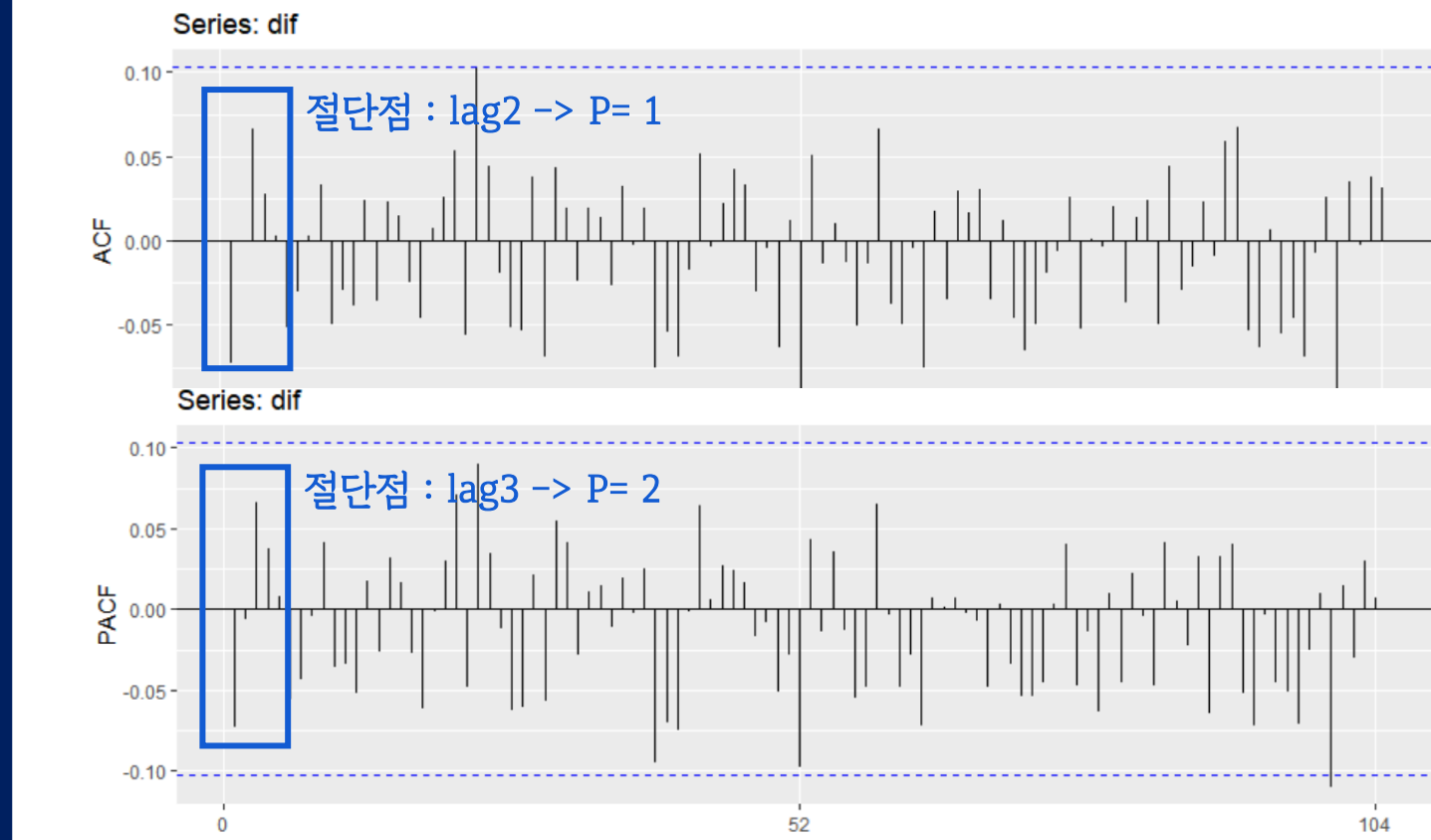
## Result : Data Analysis

### 1. Time Series Analysis

\* 2012년~2018년 주가 데이터 : train data, 2019년 주가 데이터 : validation data, 2020년 주가 데이터 : test data  
\* 모델 성능 판단 척도로는 척도에 의존성이 없는 MAPE를 이용하며, 예측오차가 작을수록 작은 값이다.  
\* 15개의 기업 중 임의로 대한항공에 대해 살펴본 후 전종목의 2020년 주가에 대해 살펴본다.

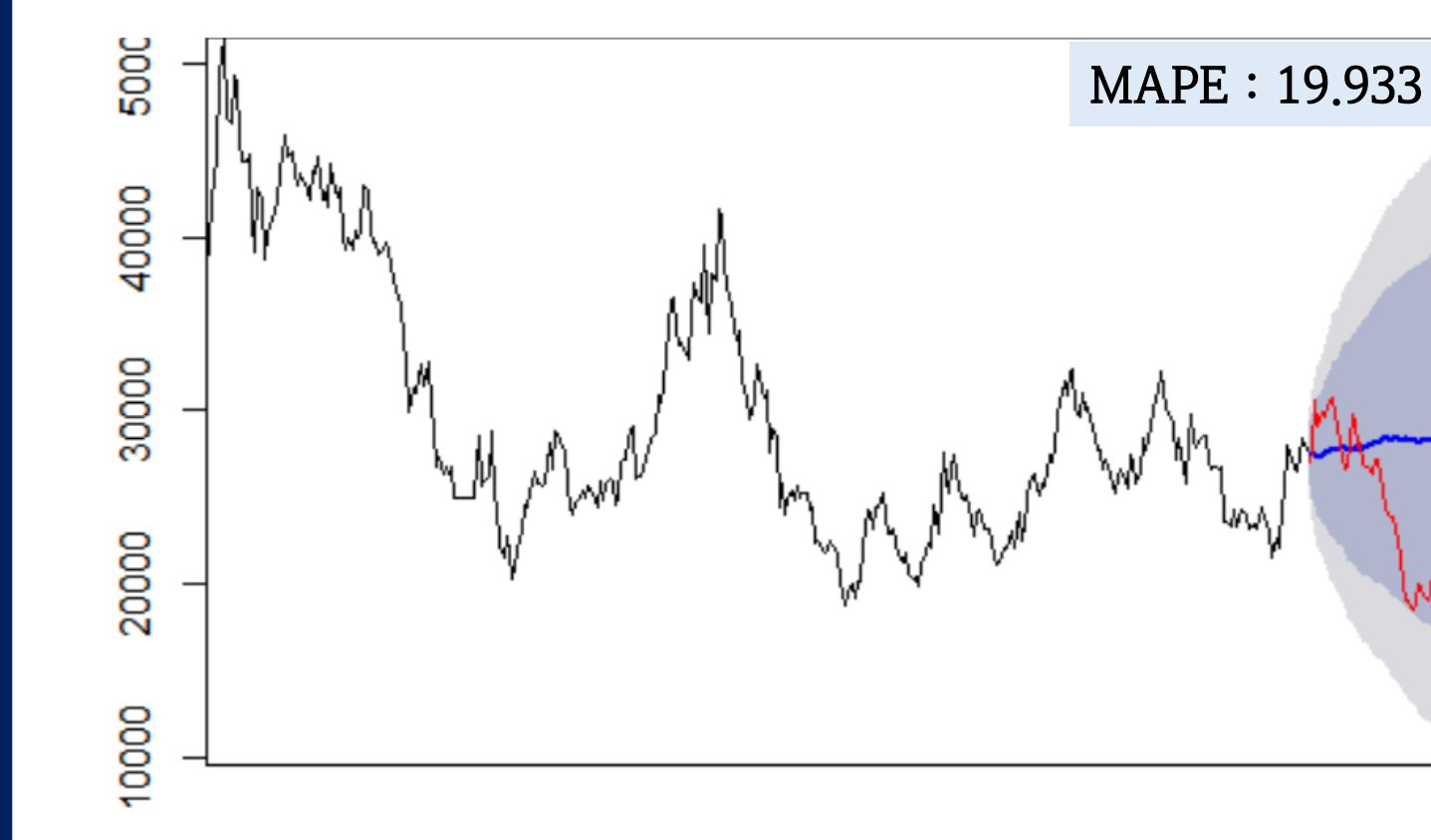
#### ① ARIMA (1,1,2) 모델

정상성 만족을 위해 차분 1번을 진행 후, ACF, PACF를 통해 절단점을 찾아 차수를 정한다.



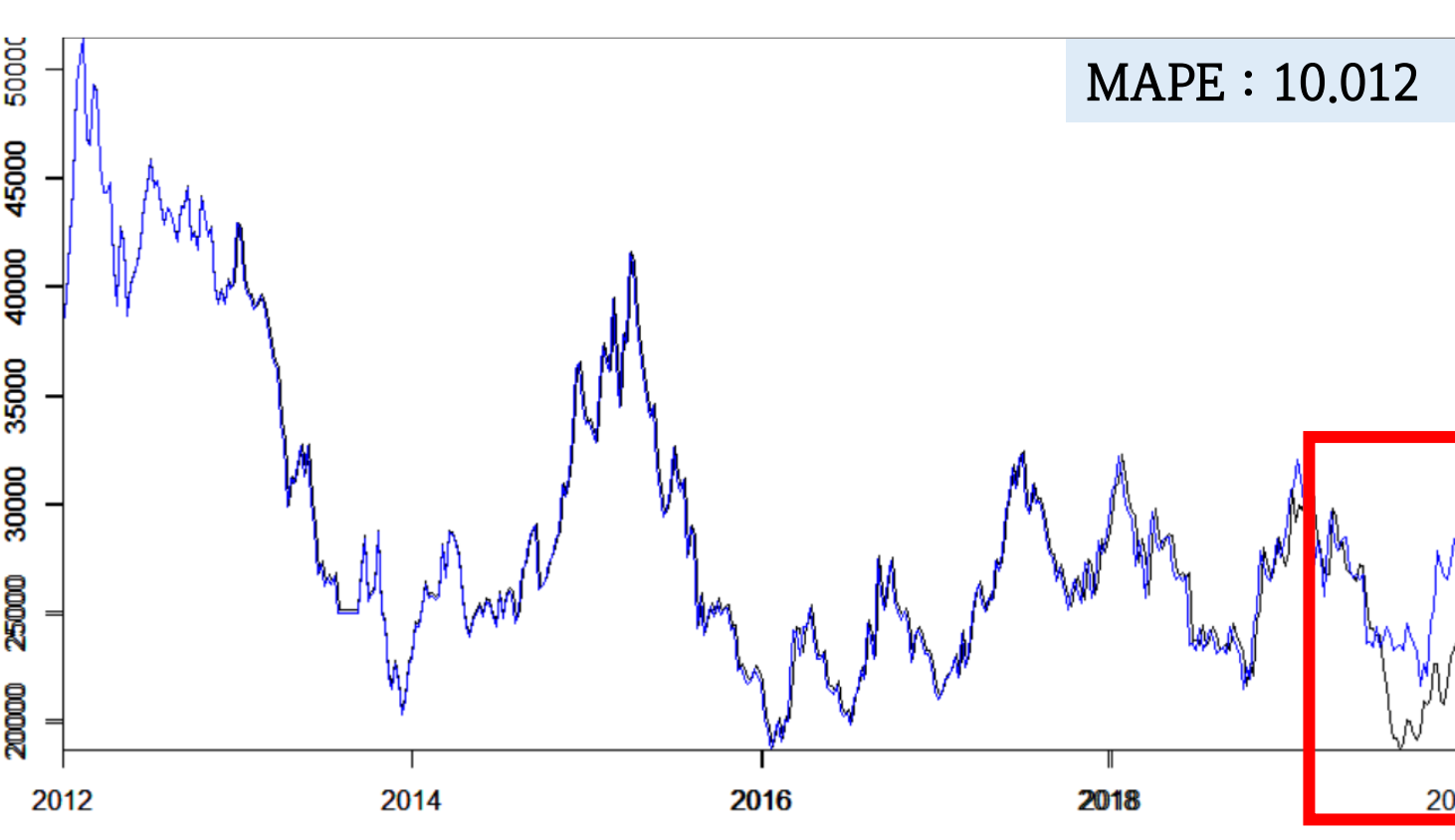
#### ② ARIMA (1,1,0)(2,0,0) 모델

Auto.arima를 이용하여 차수 결정

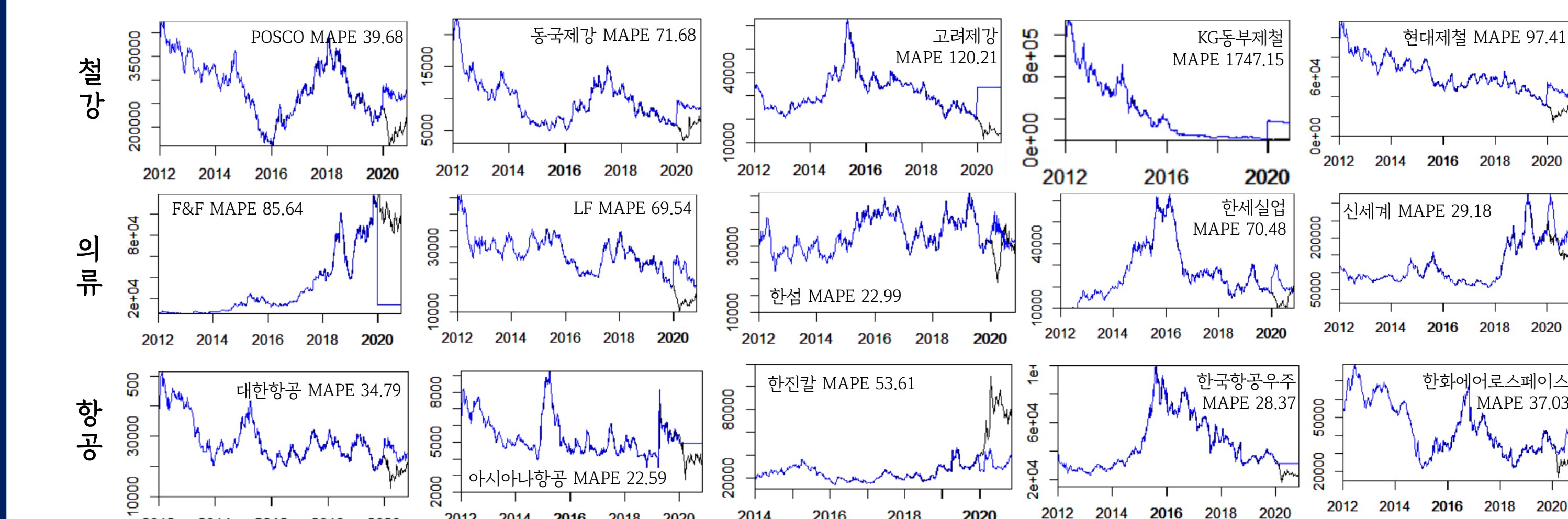


#### ③ ARMA(1,0)-GARCH(1,1)모델

Auto.arima를 이용하여 차수 결정



가장 예측이 잘 된 모델은 MAPE값이 10.012로 가장 낮은 **③ARMA(1,0)-GARCH(1,1) 모델** 이다.  
이에 따라 15개의 기업에 대해 ARMA-GARCH모델로 2020년 예측된 주가와 실제 주가의 차이를 확인해보자.



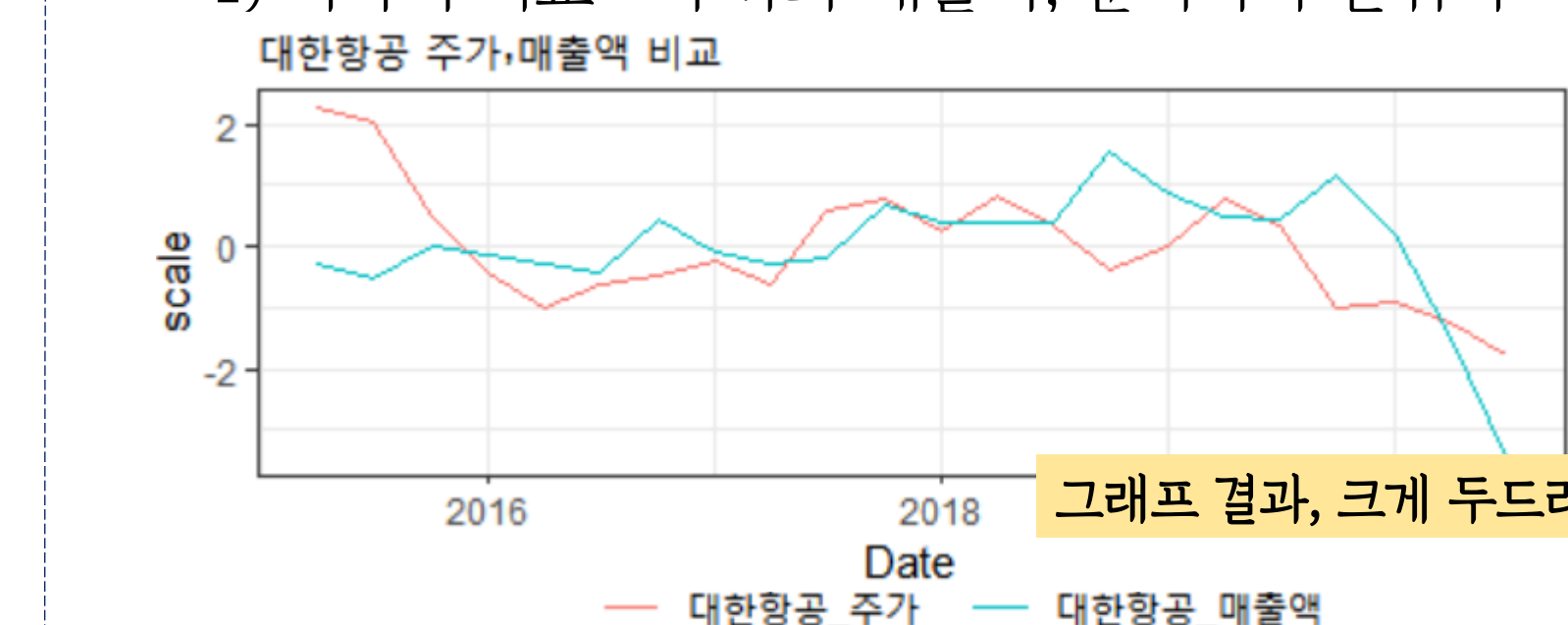
2020년, 대부분 변동을 보이므로 사회적 변화는 주식 시장에 영향을 미칠 것이라고 생각할 수 있다.

### 2. Correlation & Regression Analysis

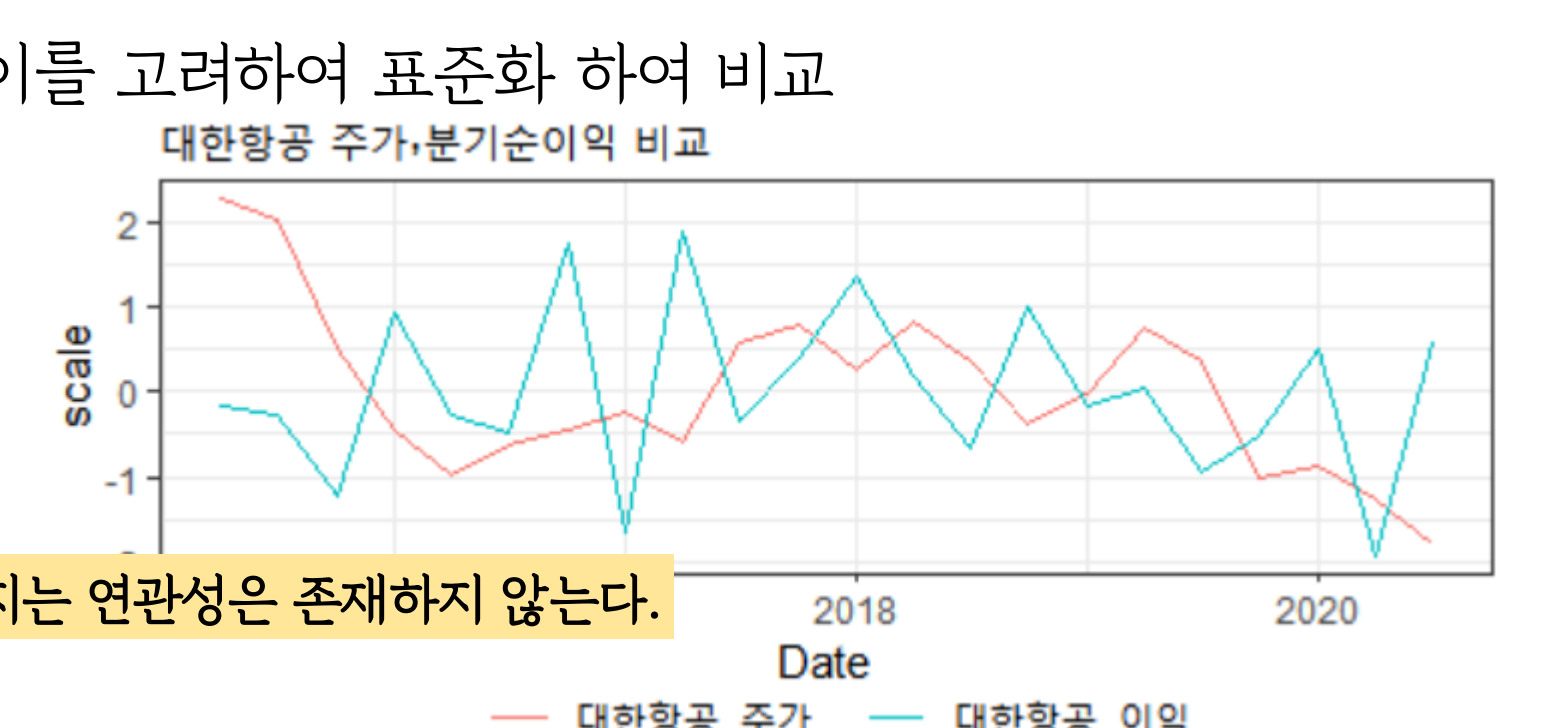
\* 최근 5개년, 분기별 평균 주가와 연결재무제표의 매출액, 분기순이익의 상관성, 인과성을 분석한다.  
\* 15개의 기업 중 임의로 대한항공에 대해 살펴본 후, 전종목에 대해 분석해본다.

#### ① 주가와 매출액의 연관성

1) 시각적 비교 - 주가와 매출액, 순이익의 단위의 차이를 고려하여 표준화 하여 비교



#### ② 주가와 순이익의 연관성

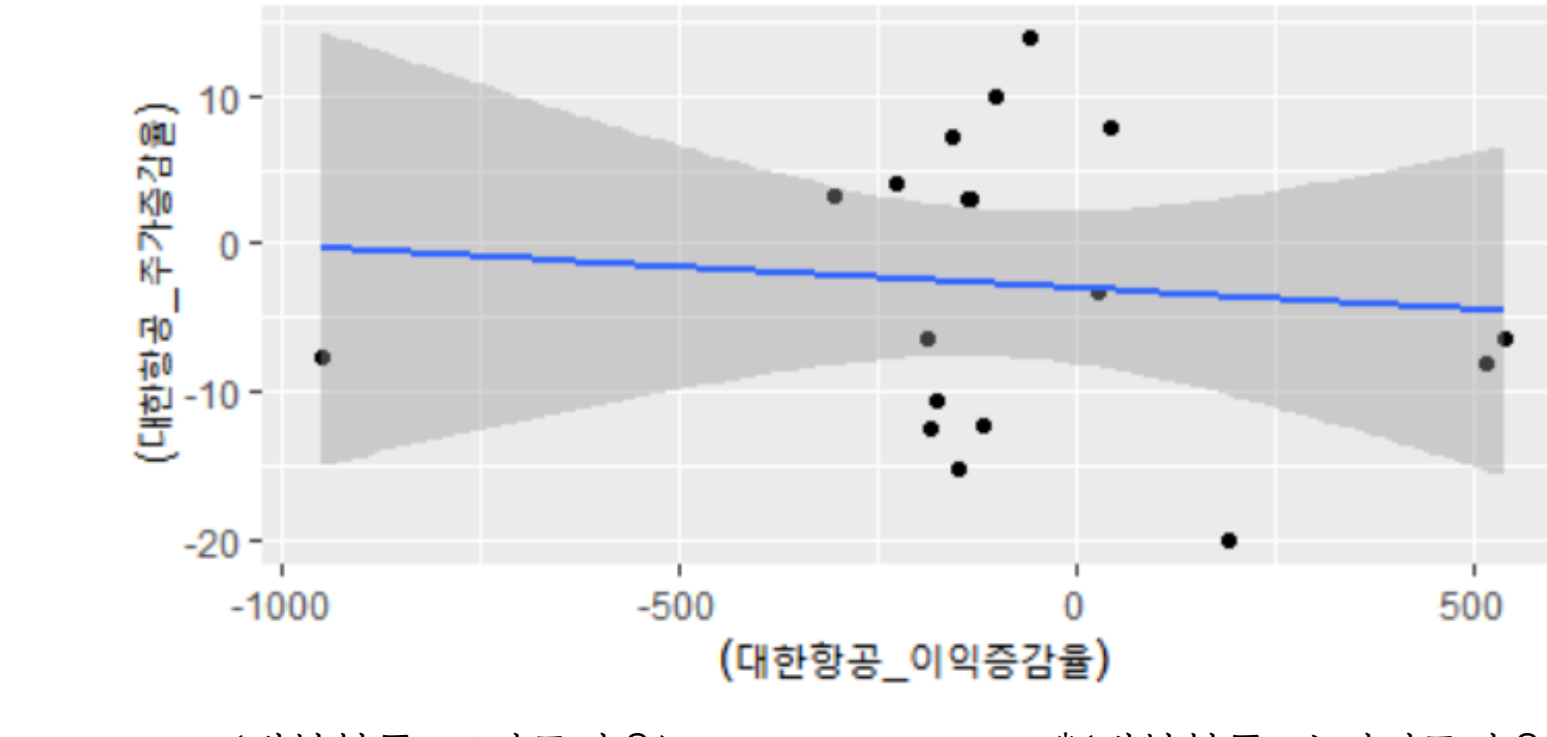
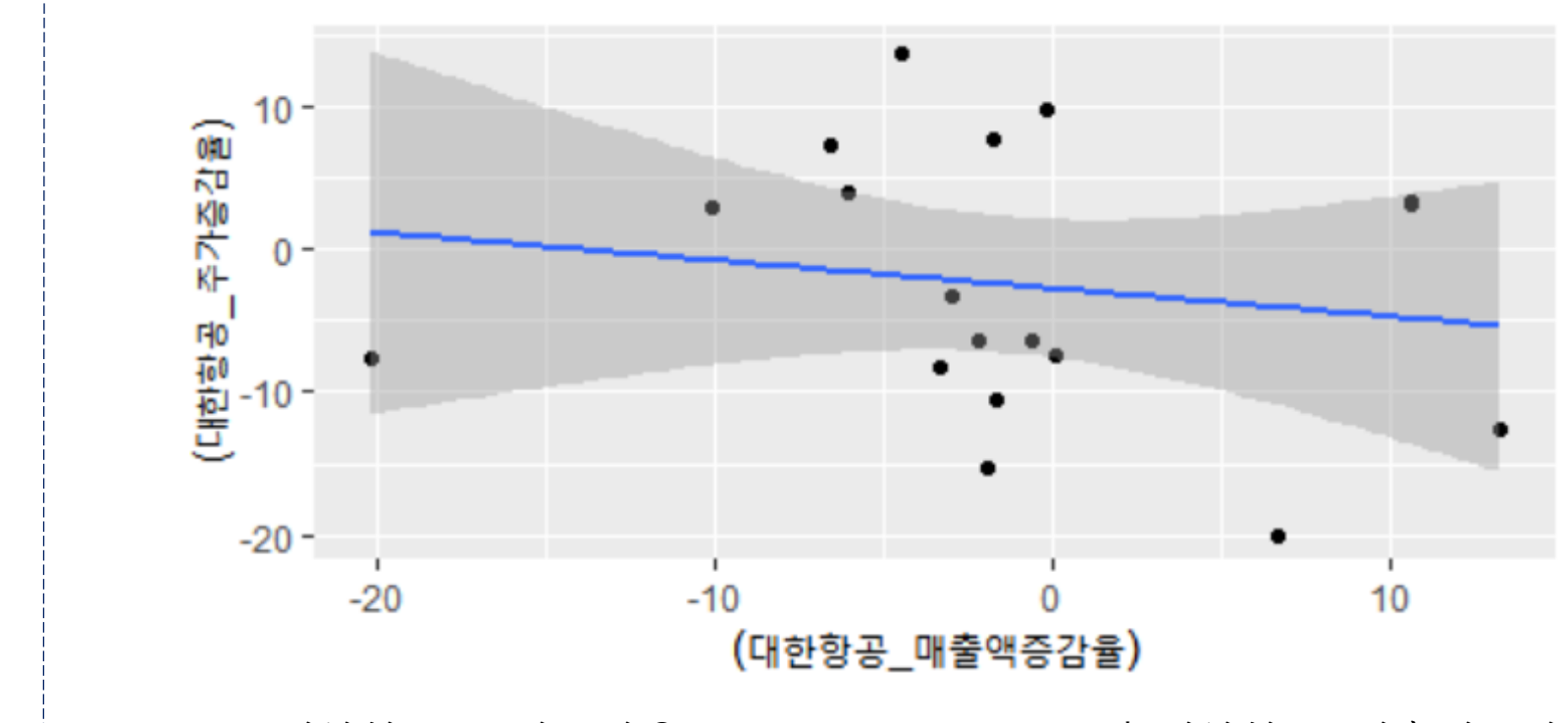


2) 상관성 - 매출액, 순이익의 증감율과 주가의 증감율 간의 선형적 관계가 존재하는가?

Cor(매출액 증감율, 주가 증감율) = -0.0327  
Cor(순이익 증감율, 주가 증감율) = -0.12787  
상관계수가 0에 매우 가까우므로 상관성이 없다고 할 수 있다.

3) 인과성 - 매출액, 순이익의 증감율은 주가의 증감율에 영향을 미치는가?

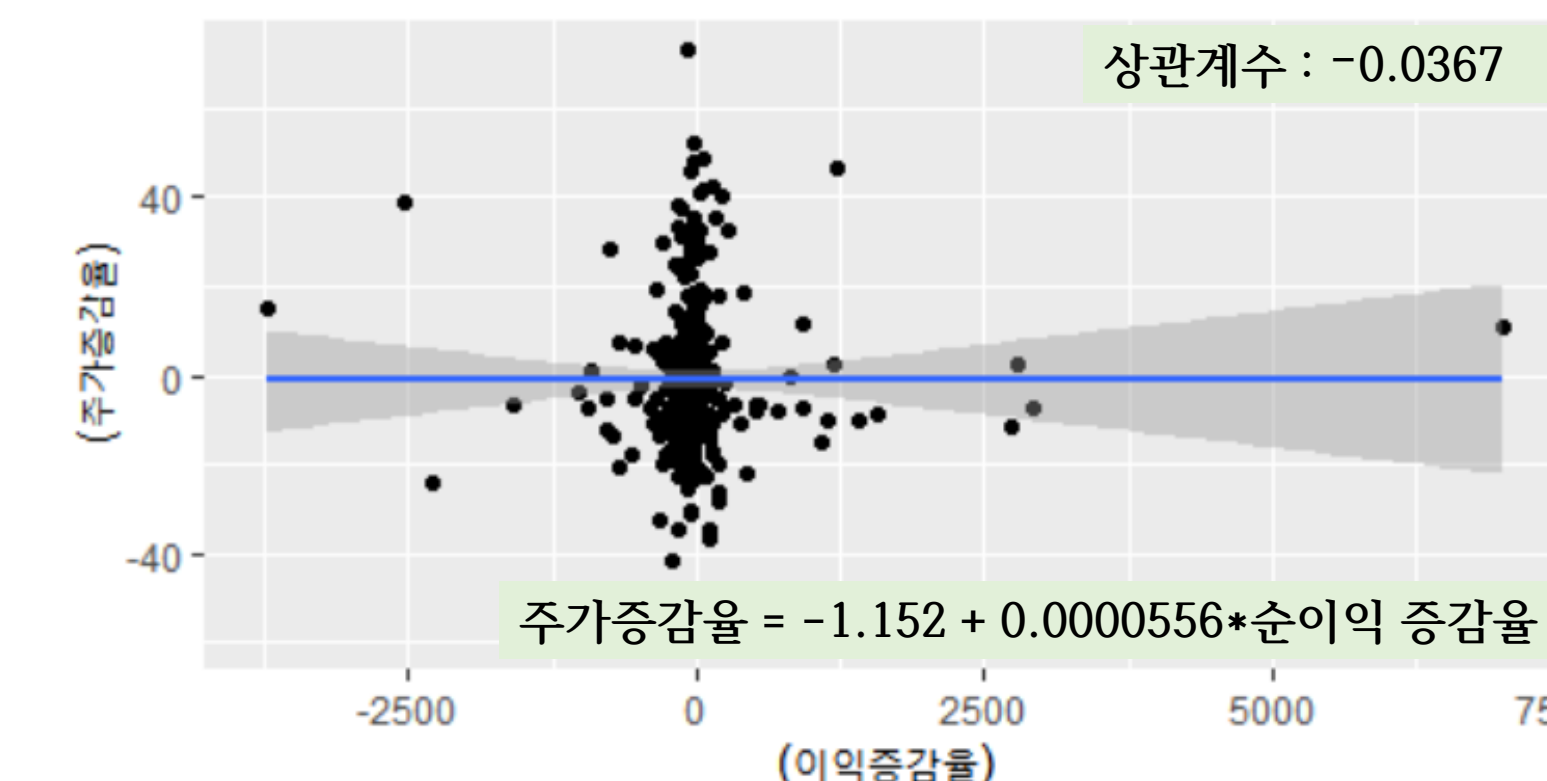
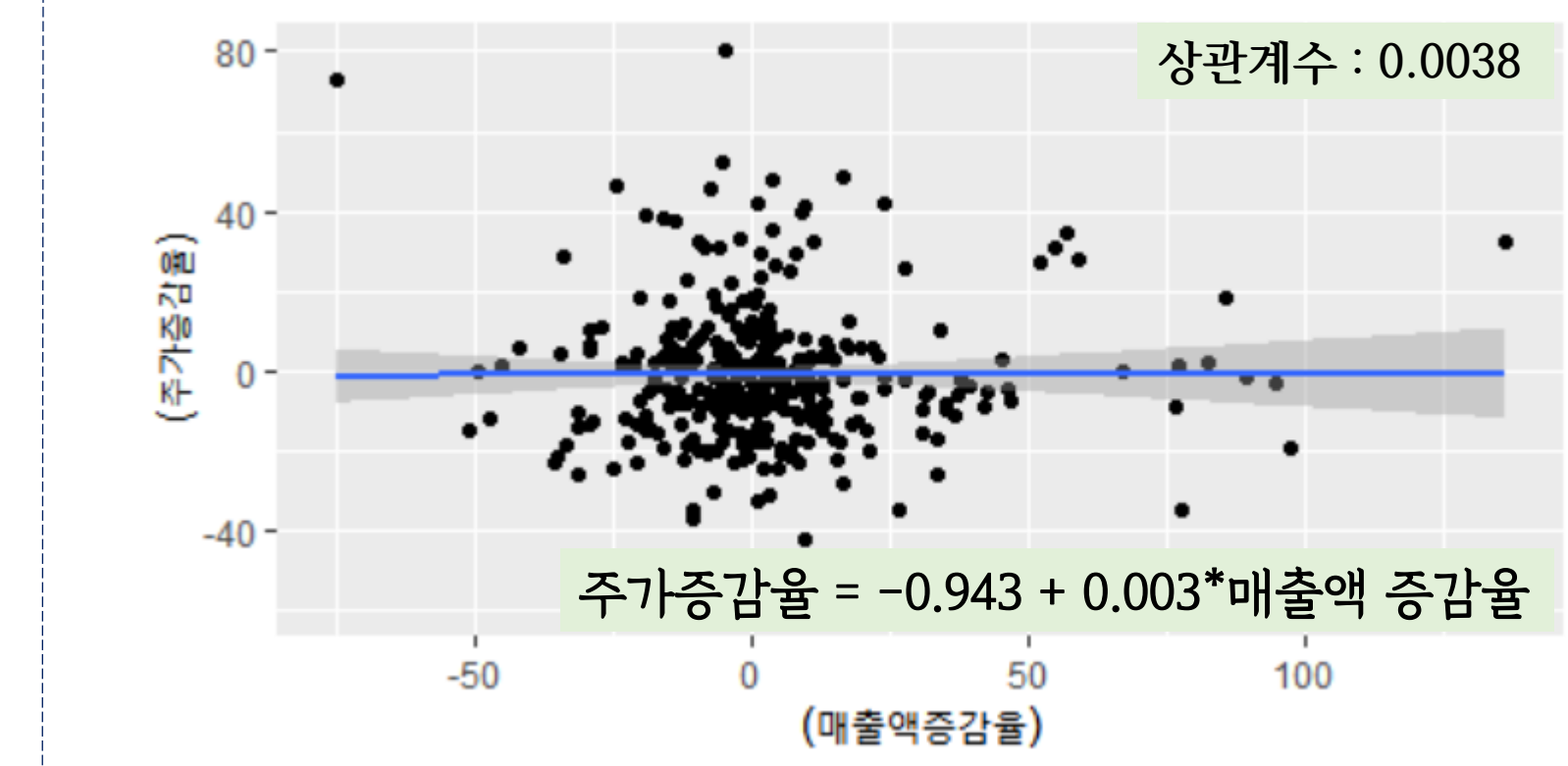
\* 이상치를 제거한 후 인과성에 대해 분석하였다.



$\beta$ 에 대한 유의성 검정 결과 P-value=0.513으로  $H_0: \beta = 0$ 을 기각하지 못한다.  
 $\beta$ 에 대한 유의성 검정 결과 P-value=0.711로  $H_0: \beta = 0$ 을 기각하지 못한다.

매출액, 순이익의 증감율은 주가의 증감율에 영향을 미치지 않는다.

4) 15개의 기업의 인과성 및 상관성



기업의 성과 지표인 **매출액, 당기순이익은 주가에 영향을 미치지 않는다.**  
즉, 투자 종목 선정 시 이에 의존하는 것은 위험하다.

## Conclusion

1. 가치지표에 따르면 IT, 금융 업계 또는 사조대림, 메디알스, 원익 등에 가치 투자할 수 있다. 하지만 이는 보수적인 지표로 의존하기 어렵다.
  2. 과거의 데이터를 통해 주가를 예측하는 것은 매우 어려우므로 현재, 더 나아가 미래의 사회/경제학적 변화에 대해 보아야 한다.
  3. 주식 투자 종목 선정 시 매출액, 순이익 등의 기업의 성과 지표에 의존하는 것은 위험하다.
- => 현재, 젊은 연령층의 소액투자자들이 증가하고 있는 상황에서 이들은 신중한 투자가 필요로 하다.
- 이 때 단순한 지표만으로 투자를 결정하는 것은 매우 제한적이므로 빠르게 변화하는 정보를 파악하는 것이 중요하다.

## Reference

- Wilkinson, Leland. 2012. The Grammar of Graphics. Springer.
- Grolemund, Garrett, and Hadley Wickham. 2018. R for Data Science.
- Hyndman, R.J., Athanasopoulos, G. 2018 Forecasting: principles and practice, 2nd edition, OTexts.