

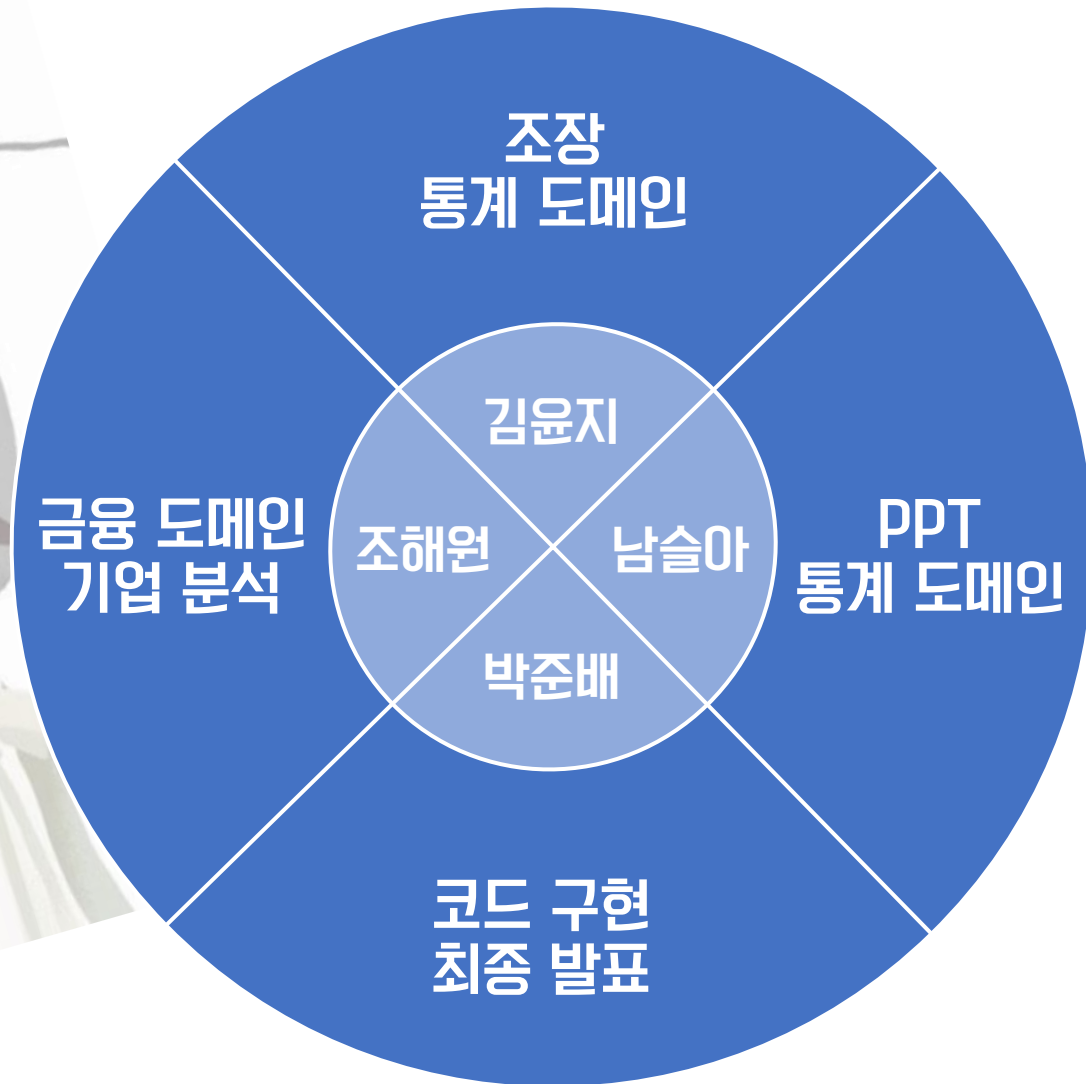
지표를 이용한 삼성전자 주가 예측 및 상관관계분석

주식4조

김윤지, 남슬아, 박준배, 조해원

삼성전자

[팀원 소개]



[목 차]

- 0. 중간 발표 피드백**
- 1. 목표 및 결과**
- 2. 회귀 변수 선정**
- 3. 회귀 분석**
- 4. 시계열 분석**
- 5. 결론 및 한계**
- 6. 의의**



0. 중간 발표 피드백

- 무위험 수익률 변수로 장기 금리 미국채를 선택한 이유는 무엇인가
- 장기 금리만 포함할 경우 PCR의 영향을 모형에 어떻게 고려할 것인가
- 금을 독립변수에 포함한 이유는 무엇인가

0. 중간 발표 피드백

Q. 무위험 수익률 변수로 장기 금리 미국채를 선택한 이유는 무엇인가

무위험[^]수익률 無危險收益率

+ 단어장 저장

1. 경제 투자 위험이 전혀 포함되지 않는, 투자의 순수한 기대 수익률. 화폐의 시간적 가치만을 고려한 것으로 정기 예금, 국채 등의 이자율이 이에 해당한다.

- 무위험 수익률은 ‘무엇’으로 한다는 합의된 단일한 정의가 없음.
- 조사당시 최초로 발견한 미국채10년물(T-Note 중에서 특히 10년물은 은행같은 큰 기관들이 선호)과 국고채3년(한국 벤치마크 무위험 이자율)을 무위험 이자율로 ‘정의’ 내림

0. 중간 발표 피드백

Q. 장기 금리만 포함할 경우 PCR의 영향을 모형에 어떻게 고려할 것인가

- 삼성전자 재무재표의 현금흐름표와 현금수지분석표의 차이: 4조원
- 삼성전자는 제조업으로 분류되는 기업 : 제조업의 특성상 원가회계에 녹아 있는 Dep을 구하기 어려움.
 - 즉, 실질적인 현금흐름의 정확한 파악 자체가 어려움. 가정치로 대략적인 확인만 가능.

0. 중간 발표 피드백

Q. 금을 독립변수에 포함한 이유는 무엇인가

- 최초 분석 계획 당시 삼성을 반도체 기업으로 생각하여 그에 대한 원자재 성격으로 금을 추가함.
- 안전자산인 금과 위험자산인 주가는 음의 상관관계를 보일 것이 통상적으로 예상되지만, 양의 상관관계가 나타남: 해당 시기에 주가가 전반적으로 크게 상승하는 시기였기 때문에 주가 상승과 동반하여 금값도 같이 상승한 것으로 생각됨.

1. 목표 및 결과



[프로젝트 목표]

- 수업에서 학습한 내용을 최대한 활용하여, 향후 프로젝트에 필요한 역량을 강화한다.

[분석 목표]

- 삼성전자 주가와 상관관계가 높은 지표를 파악한다.
- 상관관계가 높은 변수들을 독립변수로 회귀 모델을 적용하여 삼성 전자 주가 예측에 유의한 모형인지 검증한다.
- 시계열 모델을 적용하여 삼성 전자 주가 예측에 유의한 모형인지 검증한다.

1. 목표 및 결과



[분석 결과]

- 다중회귀분석 ◦ 다항회귀분석 결과, 외부 변수를 독립변수로 한 회귀 모형으로는 삼성전자 주가 예측이 어려움.
- 시계열 분석 결과, 시계열 모형은 삼성전가 주가의 단기 예측에는 적합하였으나 장기 예측은 어려움.

2. 회귀 변수 선정



[상관 분석 결과]

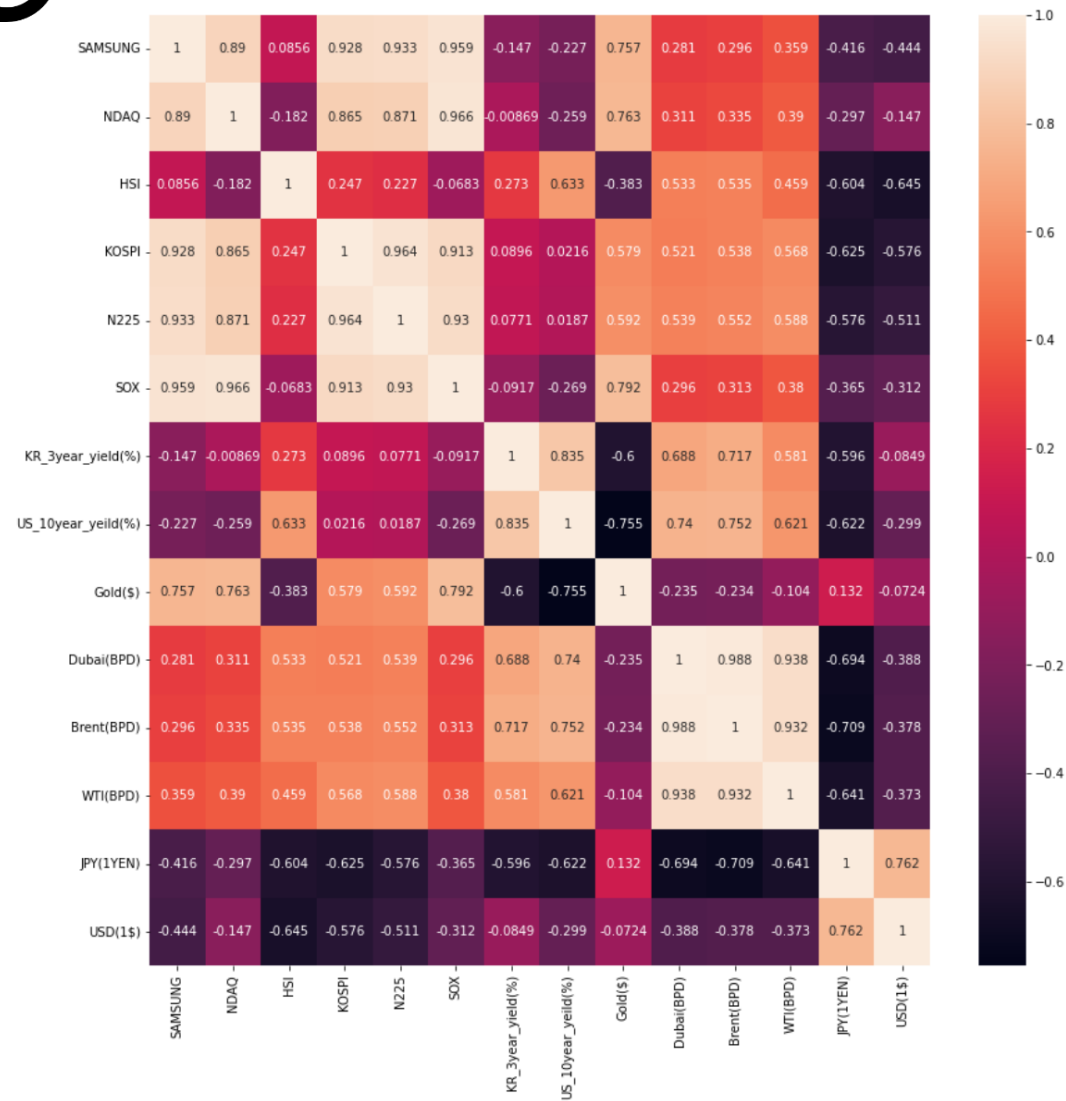
1. '독립변수 - 삼성전자 주가' 상관관계수

- 항생을 제외한 주가 지수: 강한 양의 상관관계(0.8~0.9)
- 채권2종: 음의 상관관계(-0.1~-0.3)
- 금: 삼성 주가와 강한 양의 상관관계(0.7)
- 유가3종 : 양의 상관관계(0.3~0.5)
- 환율2종: 음의 상관관계(-0.4)

2. 독립변수 간 상관관계수

다중공선성

- 항생을 제외한 주가 지수 4종: 강한 양의 상관관계(0.8~0.9)
- 무위험이자율(2종): 강한 양의 상관관계(0.8)
- 유가3종: 강한 양의 상관관계(0.9)가 나타난다.
- 금과 주가지수(항생 제외): 양의 상관관계(0.6~0.8)
- 금과 채권: 음의 상관관계(-0.8~-0.6)
- 외환: 다른 자산과 음의 상관관계



2. 회귀 변수 선정



- Lasso, Ridge, RandomForest : Feature Selection

	니케이	반도체 지수	코스피	항생	엔화	한국 3년채	미국 10년채
Lasso							
Ridge							
RandomForest							

2. 회귀 변수 선정



- PCA: Feature Selection

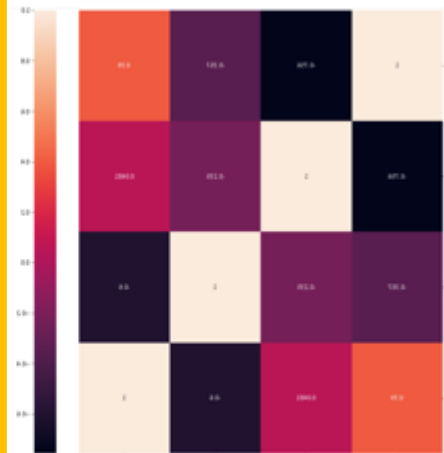
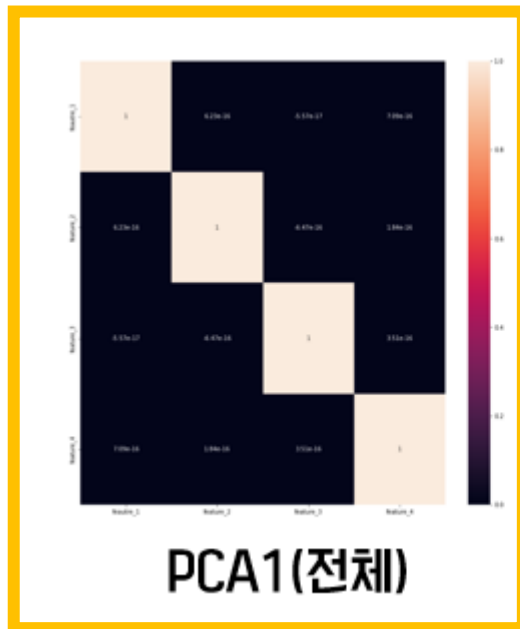
1. 전체지표(주가지수(4종), 환율(2종), 무위험 이자율(2종), 원자재(금, 유가 3종))를 PCA(component=4)로 차원축소
2. 금을 주식 자산에 포함시켜 자산군별로 PCA
3. 금을 채권 자산에 포함시켜 자산군별로 PCA
4. 상관관계가 ± 0.3 이하인 지표를 제외 후 자산군별로 PCA

3. 회귀 분석

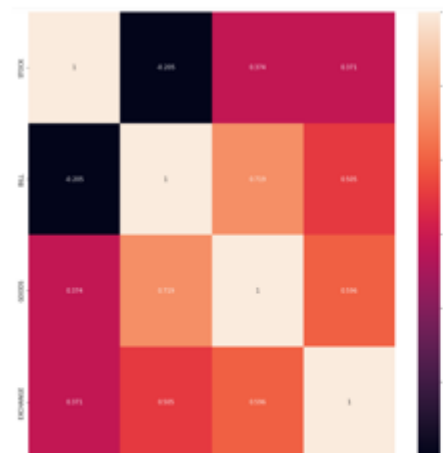


- PCA: Feature Selection

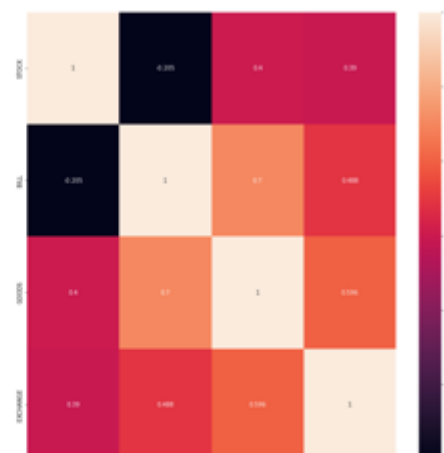
➡ PCA1 유형 외에는 변수 간 높은 상관관계가 해결되지 않음.



PCA2(자산군별)



PCA3(금&주가지표)



PCA3(금&채권)

3. 회귀 분석



1. Lasso, Ridge, RandomForest: 선형 회귀 분석 결과

	MSE	RMSE	R2(결정계수)
Lasso	0.070	0.265	0
Ridge	0.003	0.057	1
RandomForest	0.361	0.601	-4

➡ 모형이 적합하지 않음.

3. 회귀 분석



2. PCA1: 다중 선형 회귀 분석 결과

	MSE	RMSE	R2(결정계수)
Lasso	0.070	0.265	0
Ridge	0.004	0.060	1
RandomForest	0.001	0.034	1
LinearRegression	0.004	0.061	1

➡ 모형이 적합하지 않음.

3. 회귀 분석



2. PCA1: 다항 선형 회귀 분석 결과

	MSE	RMSE	R2(결정계수)
Degree=2	0.00	0.04	1
Degree=3	0.16	0.41	-1

➡ 모형이 적합하지 않음.

4. 시계열 분석



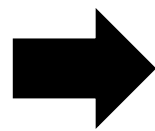
1. 데이터 전처리: 스케일링

[학습]

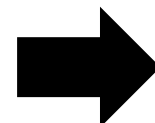
- 삼성전자 주가 25개년 데이터
(1996-01-01 ~ 2021-12-31)
- 삼성전자 주가 최근 3개년 데이터
(2019-01-01 ~ 2021-12-31)

[예측]

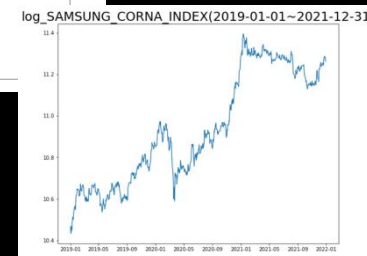
- 삼성전자 주가 최근 3개월 데이터
(2022-01-01 ~ 2022-04-01)



로그
스케일링



추세

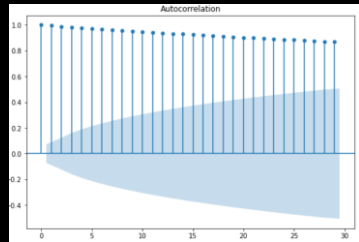


4. 시계열 분석



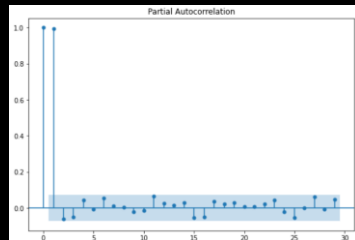
1. 데이터 전처리: 정상성

- 삼성전자 주가 최근 3개년 데이터
(2019-01-01 ~ 2021-12-31)

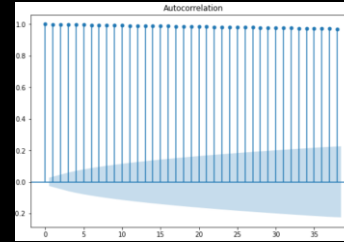


ACF

PACF

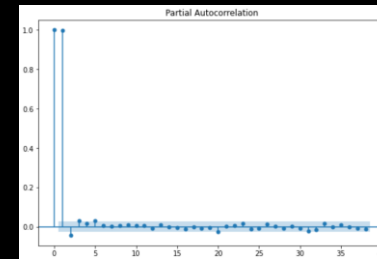


- 삼성전자 주가 25개년 데이터
(1996-01-01 ~ 2021-12-31)



ACF

PACF



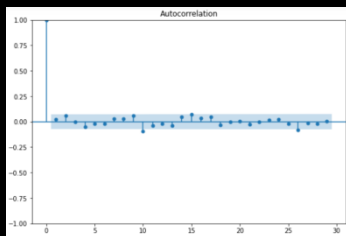
로그 스케일
데이터

4. 시계열 분석



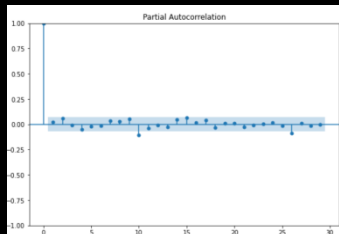
1. 데이터 전처리: 정상성

- 삼성전자 주가 최근 3개년 데이터
(2019-01-01 ~ 2021-12-31)

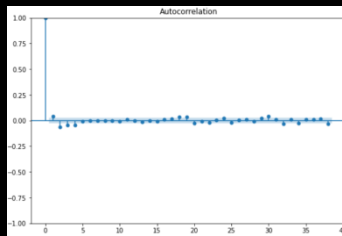


ACF

PACF

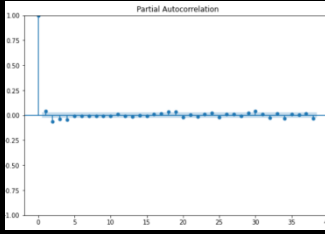


- 삼성전자 주가 25개년 데이터
(1996-01-01 ~ 2021-12-31)

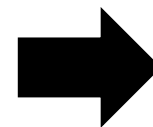


ACF

PACF



1차 차분



[3개년 데이터]

KPSS Statistic: 0.075
p-value: 0.1
num lags: 2
Critical Values:
10% : 0.347
5% : 0.463

[25개년 데이터]

KPSS Statistic: 0.054
p-value: 0.1
num lags: 22
Critical Values:
10% : 0.347
5% : 0.463

정상성

4. 시계열 분석



2. 최적 시계열 모형 도출: 삼성전자 주가 25개년 데이터

`auto_arima(y=log_whole_samsung, d = 1, start_p=0, max_p=3, start_q=0, max_q=3, seasonal=False, stepwise=True, trace=True)`

Best model: ARIMA(1,1,2)(0,0,0) [0] intercept

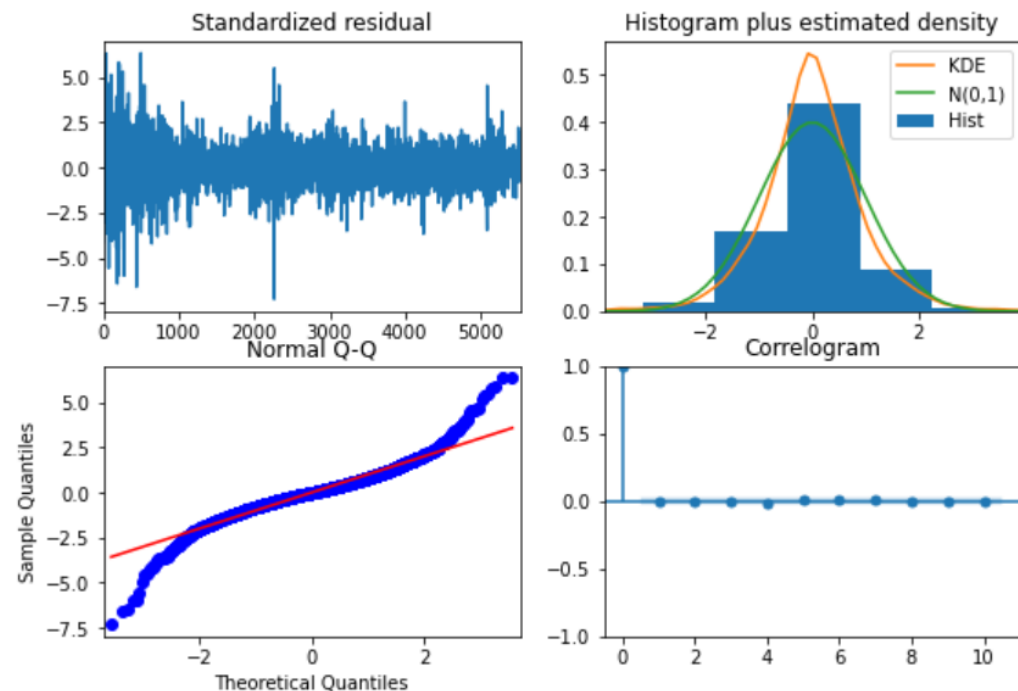
Total fit time: 18.540 seconds

SARIMAX Results

```
=====
Dep. Variable:          y      No. Observations:      5528
Model:                SARIMAX(1, 1, 2)      Log Likelihood:    13226.596
Date:                Tue, 03 May 2022      AIC:              -26443.192
Time:                10:40:02      BIC:              -26410.105
Sample:              0      HQIC:             -26431.654
                  - 5528
Covariance Type:      opg
=====
```

	coef	std err	z	P> z	[0.025	0.975]
intercept	0.0002	0.000	1.817	0.069	-1.43e-05	0.000
ar.L1	0.6403	0.060	10.614	0.000	0.522	0.759
ma.L1	-0.6003	0.060	-10.065	0.000	-0.717	-0.483
ma.L2	-0.0906	0.010	-9.397	0.000	-0.109	-0.072
sigma2	0.0005	5.08e-06	96.109	0.000	0.000	0.000

```
=====
Ljung-Box (L1) (Q):                0.01      Jarque-Bera (JB):          5436.44
Prob(Q):                          0.94      Prob(JB):                   0.00
Heteroskedasticity (H):            0.35      Skew:                       -0.02
Prob(H) (two-sided):              0.00      Kurtosis:                   7.86
=====
```



4. 시계열 분석



2. 최적 시계열 모형 도출: 삼성전자 주가 최근 3개년 데이터

```
auto_arima(y=log_corona_samsung, d = 1, start_p=0, max_p=3, start_q=0, max_q=3, seasonal=False, stepwise=True, trace=True)
```

Best model: ARIMA(0,1,0)(0,0,0)[0] intercept
Total fit time: 1.033 seconds

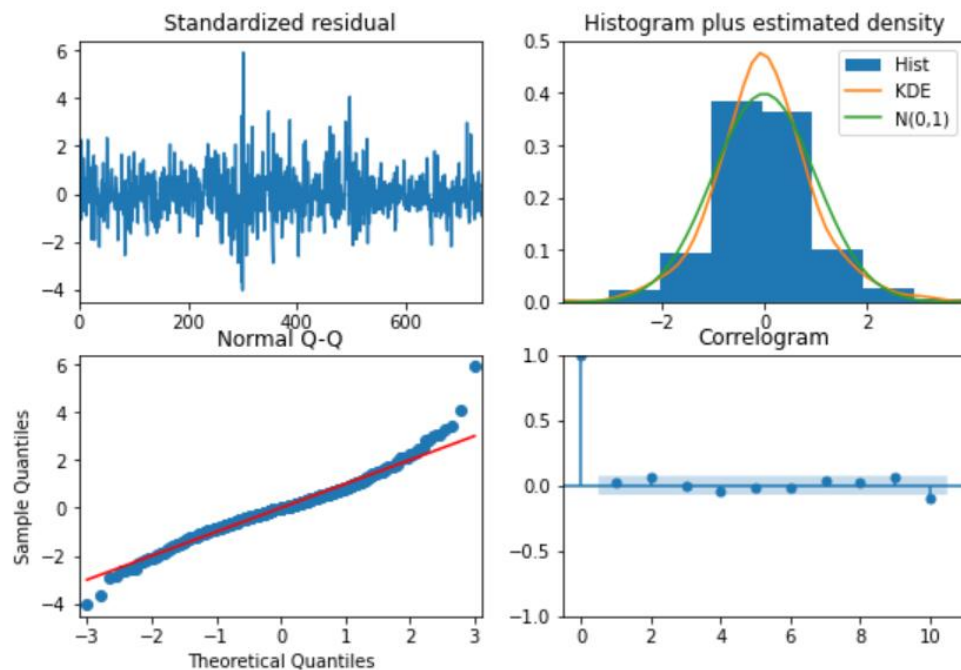
SARIMAX Results

Dep. Variable:	y	No. Observations:	742
Model:	SARIMAX(0, 1, 0)	Log Likelihood	1982.192
Date:	Mon, 02 May 2022	AIC	-3960.384
Time:	19:28:34	BIC	-3951.168
Sample:	0	HQIC	-3956.831
	- 742		

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
intercept	0.0011	0.001	1.728	0.084	-0.000	0.002
sigma2	0.0003	9.54e-06	29.139	0.000	0.000	0.000

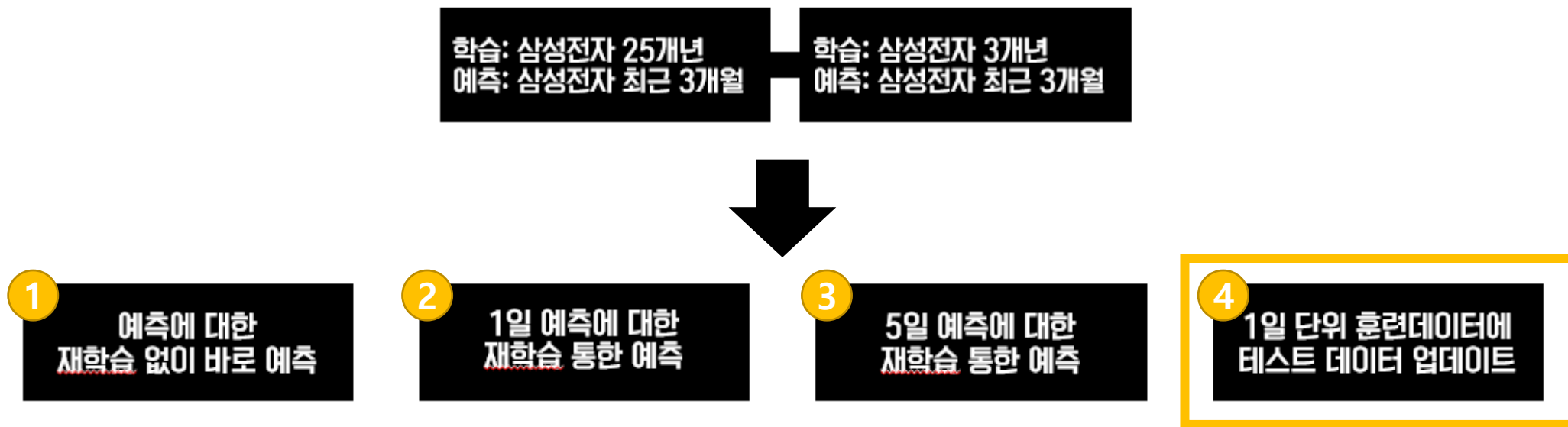
Ljung-Box (L1) (Q):	0.47	Jarque-Bera (JB):	245.39
Prob(Q):	0.49	Prob(JB):	0.00
Heteroskedasticity (H):	0.88	Skew:	0.37
Prob(H) (two-sided):	0.31	Kurtosis:	5.72



4. 시계열 분석



3. 분석



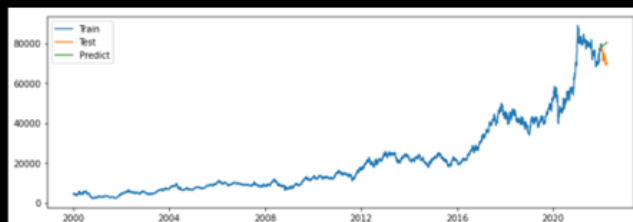
4. 시계열 분석



3. 분석: 예측에 대한 재학습 없이 바로 예측

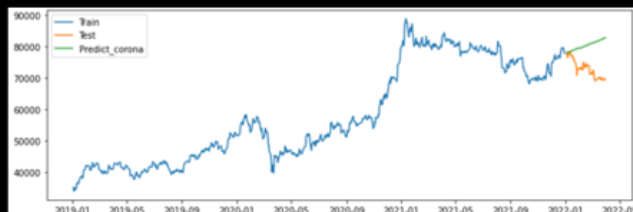
1

- 학습: 삼성전자 주가 25개년
- 예측: 삼성전자 주가 최근 3개월



2

- 학습: 삼성전자 주가 최근 3개년
- 예측: 삼성전자 주가 최근 3개월



1 과 2 결과를 비교



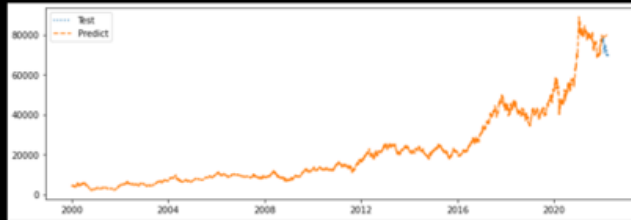
4. 시계열 분석



3. 분석: 1일 예측에 대한 재학습 통한 예측

1

- 학습: 삼성전자 주가 25개년
- 예측: 삼성전자 주가 최근 3개월



2

- 학습: 삼성전자 주가 최근 3개년
- 예측: 삼성전자 주가 최근 3개월



1 과 2 결과를 비교



4. 시계열 분석



3. 분석: 5일 예측에 대한 재학습 통한 예측

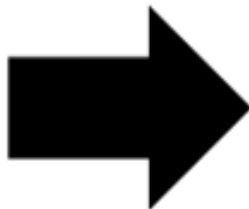
1

- 학습: 삼성전자 주가 25개년
- 예측: 삼성전자 주가 최근 3개월



2

- 학습: 삼성전자 주가 최근 3개년
- 예측: 삼성전자 주가 최근 3개월



1 과 2 결과를 비교



4. 시계열 분석



3. 분석: 1일 단위 훈련 데이터에 테스트 데이터 업데이트

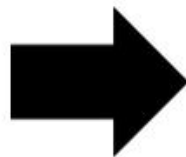
1

- 학습: 삼성전자 주가 25개년
- 예측: 삼성전자 주가 최근 3개월



2

- 학습: 삼성전자 주가 최근 3개년
- 예측: 삼성전자 주가 최근 3개월



1 과 2 결과를 비교



5. 결론 및 한계



[결론]

1. 다중회귀분석 ◦ 다항회귀분석 결과, 외부 변수를 독립변수로 한 회귀 모형으로는 삼성전자 주가 예측이 어려움.
2. 시계열 분석 결과, 시계열 모형은 삼성전가 주가의 단기 예측에는 적합하였으나 장기 예측은 어려움.

6. 의의



- 프로젝트 목표인 ‘학습 내용 적용’은 달성하였음.

- 분석 목표의 경우 효과적 모형은 발견하지 못하였으나,

- > 회귀 분석 모형은 데이터 선정 단계에서 통계적 검증, 사전 조사가 중요함.

- > 외부 변수를 독립변수로 한 주가 예측은 회귀 모형으로 구현이 어려움.

- > 시계열 데이터는 장기 예측은 어려웠으나 단기 예측은 가능하였음.

상기의 결과를 얻었다는 점에서 분석 의의가 있음.



감사합니다 