Introduction to Time Series Analysis

Introduction to Artificial Intelligence - 2023 Summer

Aug 10, 2023
Thu 4 PM

Kwangwoon University MI:RU
Artificial Intelligence Study

Agenda

In this course, you will learn

Part 1 – Introduction to Time Series Analysis

Part 2 – Time Series Components and Patterns

Part 3 – Time Series Modeling: ARIMA

Part 4 – Time Series Modeling: Seasonal ARIMA

Part 5 – Time Series Modeling: Exponential Smoothing

Part 6 – Time Series Modeling: ARIMA vs. Exponential Smoothing

Introduction to Time Series Analysis

Definition and characteristics of time series

time series data?

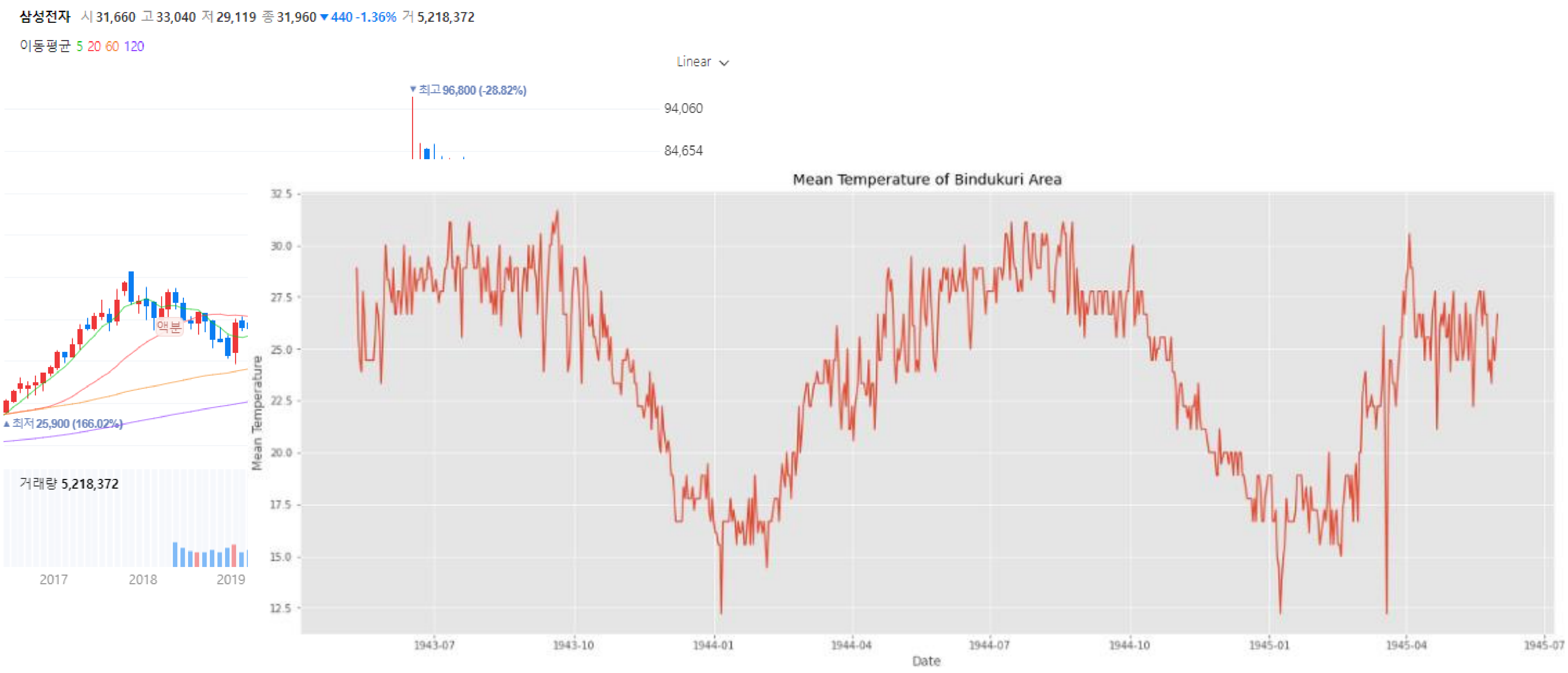
-> 일정한 시간 동안 수집된 일련의 순차적으로 정해진 데이터 셋의 집합

characteristics?

-> 시간에 관해 순서가 매겨져 있고, 연속된 관측치는 서로 상관관계를 가짐

Introduction to Time Series Analysis

Applications of time series analysis in various domains



Introduction to Time Series Analysis

Challenges and considerations in analyzing time series data

Challenges?

-> 시계열 모형은 여러 변수를 고려할 수 없음

-> 오차를 피할 수 없음

considerations?

-> 특성에서 알 수 있었듯, 관측치가 항상 독립적인 것은 아님(상관관계)

Time Series Components and Patterns

Time Series Components and Patterns

Trend, seasonality, and cyclicity in time series data

Trend?

->데이터가 장기적으로 증가하거나 감소할 경우 추세가 존재함.

추세의 방향이 변화하는 모습

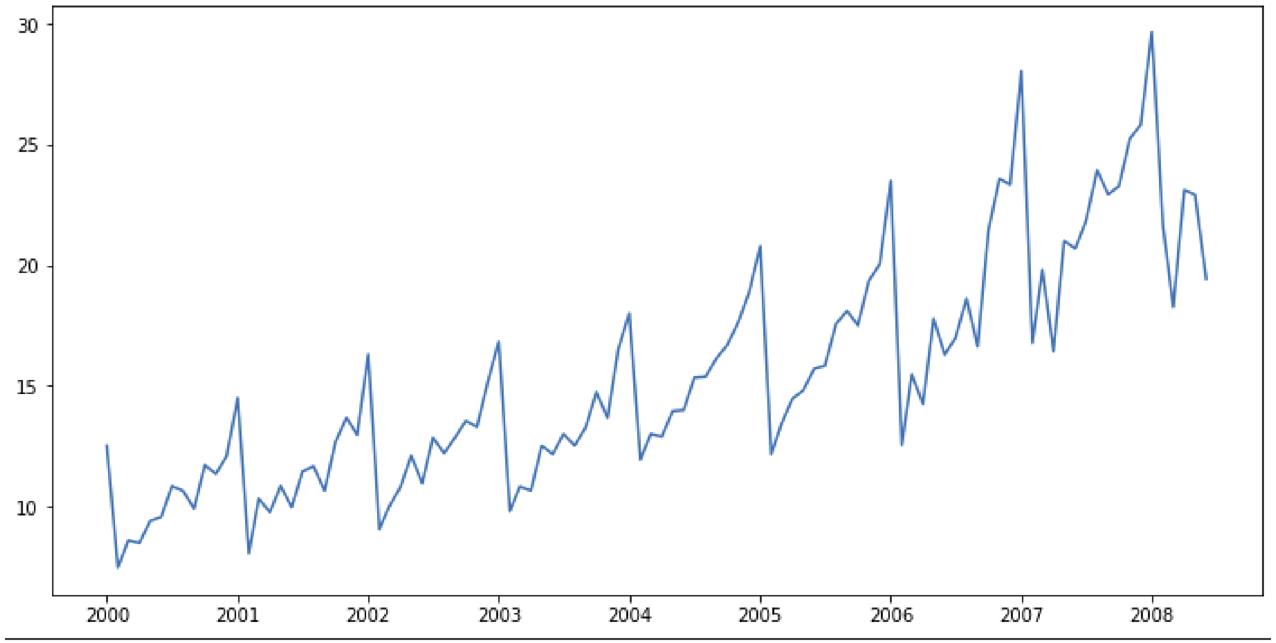
Time Series Components and Patterns

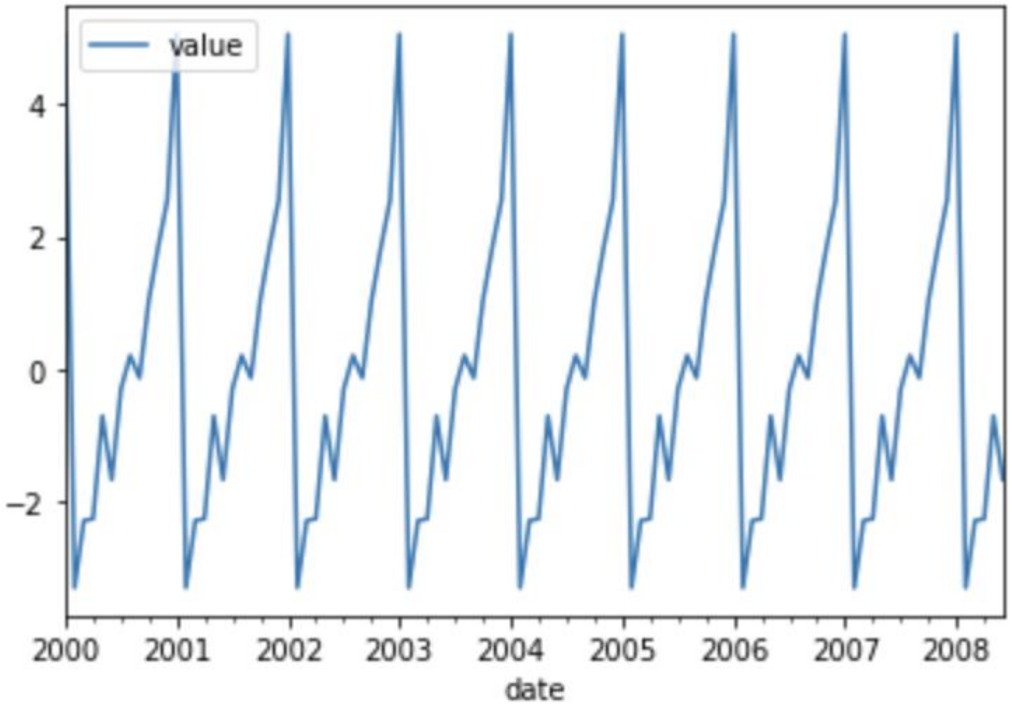
Trend, seasonality, and cyclicity in time series data

seasonality?

-> 계절성 요일이 시계열에 영향을 줄 때, 계절성 패턴이 나타남.

시계열 데이터에서 계절성을 추출한 모습





Time Series Components and Patterns

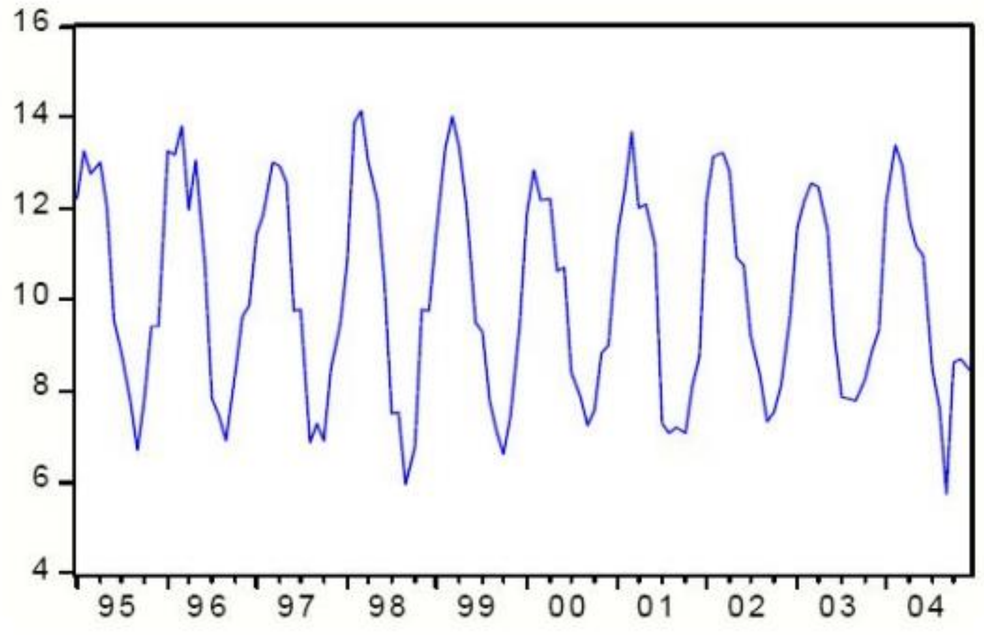
Trend, seasonality, and cyclicity in time series data

cyclicity?

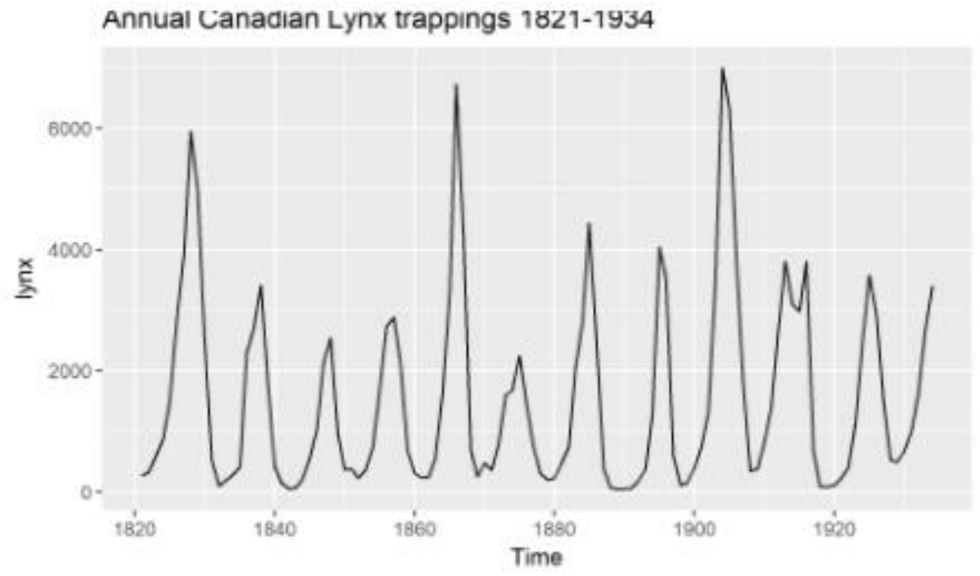
->고정된 빈도가 아닌 형태로 증가하거나 감소하는 모습을 주기라고 함

->일정하지 않은 빈도로 발생하는 계절성

계절성



주기성



Time Series Components and Patterns

Stationarity and its importance in time series analysis

Stationarity? 정상성?

->일정하여 늘 한결같은 성질

->시간에 무관하게 과거, 현재, 미래의 분포가 같다.

->시계열의 평균과 분산이 일정하고, 추세와 계절성이 존재하지 않는 성질

Why is Stationarity important?

->정상성을 가진 데이터는 여러 구간에 걸친 관찰 값 간의 관계가 일정함

->시계열 분석의 모델은 정상화된 시계열을 가정함

Time Series Components and Patterns

Autocorrelation and partial autocorrelation functions

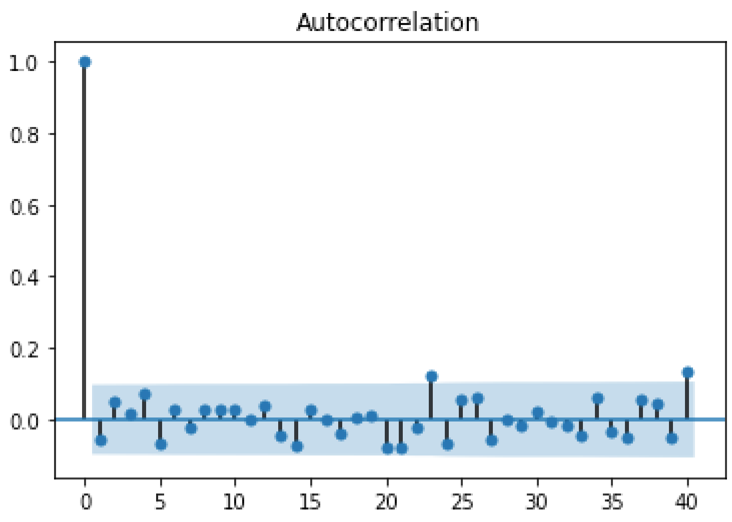
Autocorrelation?

->시계열의 시차 값 사이의 연관 정도

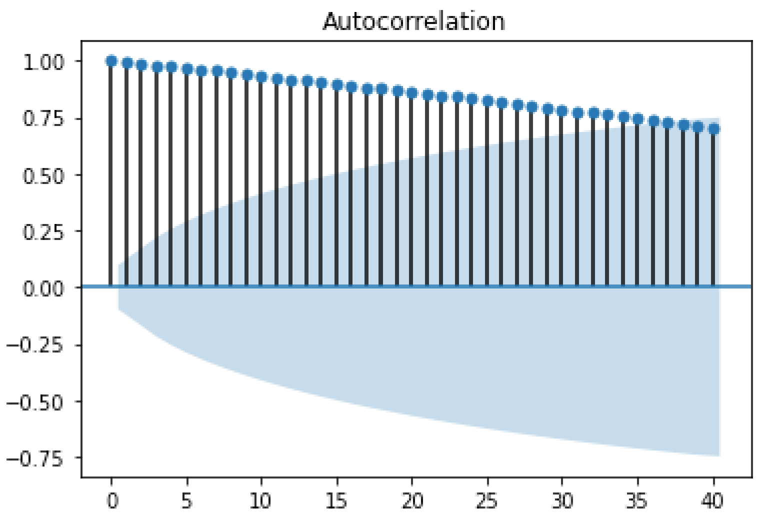
->과거의 관찰 값이 현재 관찰 값에 미치는 영향을 확인할 수 있음

->ACF(AutoCorrelation Function)을 통해 시각화

정상성



비정상성



Time Series Components and Patterns

Autocorrelation and partial autocorrelation functions

Partial autocorrelation?

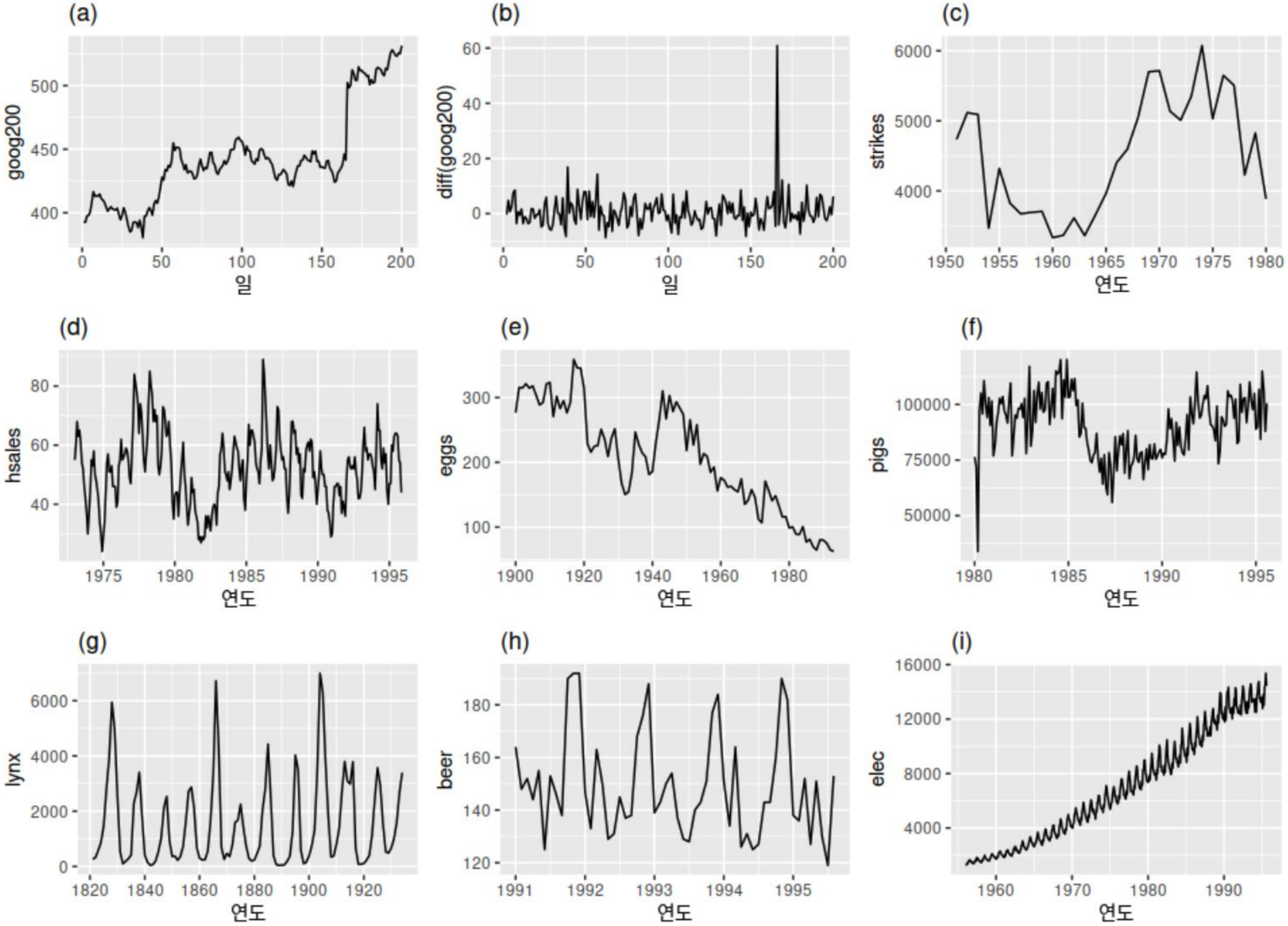
->시계열의 시차 값 사이의 순수한 상호 연관성

->중간 관측 값과의 연관성은 무시하고 순수한 시차 값 사이의 연관성

->PACF(Partial AutoCorrelation Function)을 통해 시각화

Time Series Components and Patterns

Identifying and understanding different patterns in time series data



Time Series Modeling: ARIMA

Time Series Modeling: ARIMA

Introduction to Autoregressive Integrated Moving Average (ARIMA) models

ARIMA model?

->시계열을 예측하는 방법 중 가장 널리 사용하는 방식

->데이터에 나타나는 자기상관을 표현하는데 목적이 있음

->현재 값을 과거 값과 과거 예측 오차를 통해 설명함

Time Series Modeling: ARIMA

Differencing for achieving stationarity

How to make stationarity?

->정상성을 나타내기 위해 연이은 관측 값들의 차이를 계산할 수 있음

->이 방법을 differencing(차분)이라고 함.

differencing?

-> 이어진 데이터들의 차이를 구하는 것

$$y_t^* = y_t - y_{t-1}$$

Time Series Modeling: ARIMA

Identification, estimation, and interpretation of AR, MA, and I components

AR(Auto Regressive)?

-> 과거가 미래를 예측한다는 사실에 기반

-> AR의 차수는 얼마나 먼 과거의 데이터까지 고려

$$X(t) = w*X(t-1) + b + u*e(t)$$

Time Series Modeling: ARIMA

Identification, estimation, and interpretation of AR, MA, and I components

MA(Moving Average)?

->평균을 중심으로 각 시계열 값이 가지는 오차를 반영

->MA(q)의 차수는 얼마나 먼 과거의 오차까지 고려

$$X(t) = w*e(t-1) + b + u*e(t)$$

Time Series Modeling: ARIMA

Identification, estimation, and interpretation of AR, MA, and I components

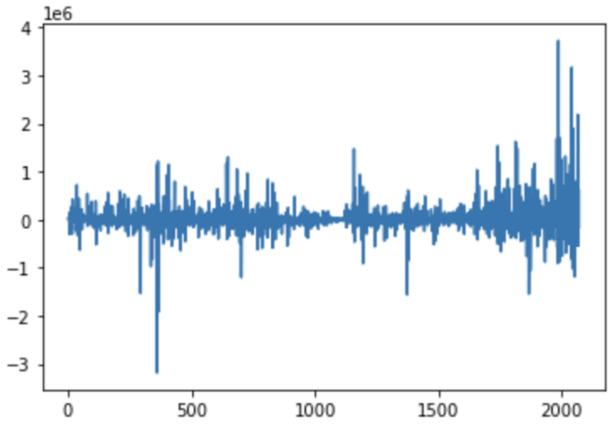
ARIMA(Auto Regressive Integrated Moving Average)?

-> AR(p) + I + MA(q)

-> I는 차분에 대한 정보

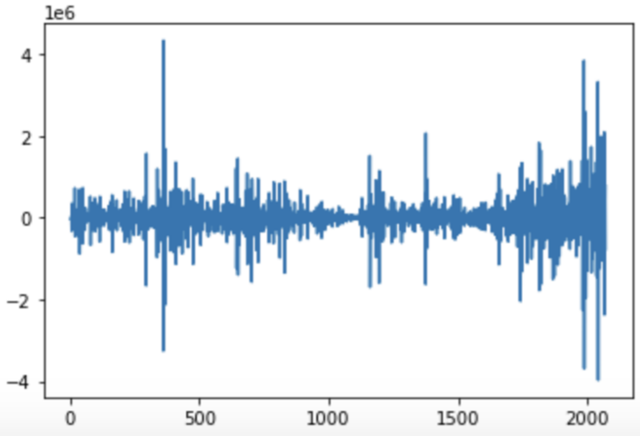
-> ARIMA(p,d,q)

In [4]:
```
diff_1 = train['price'].diff().dropna()
plt.plot(diff_1)
plt.show()
```



1차 차분

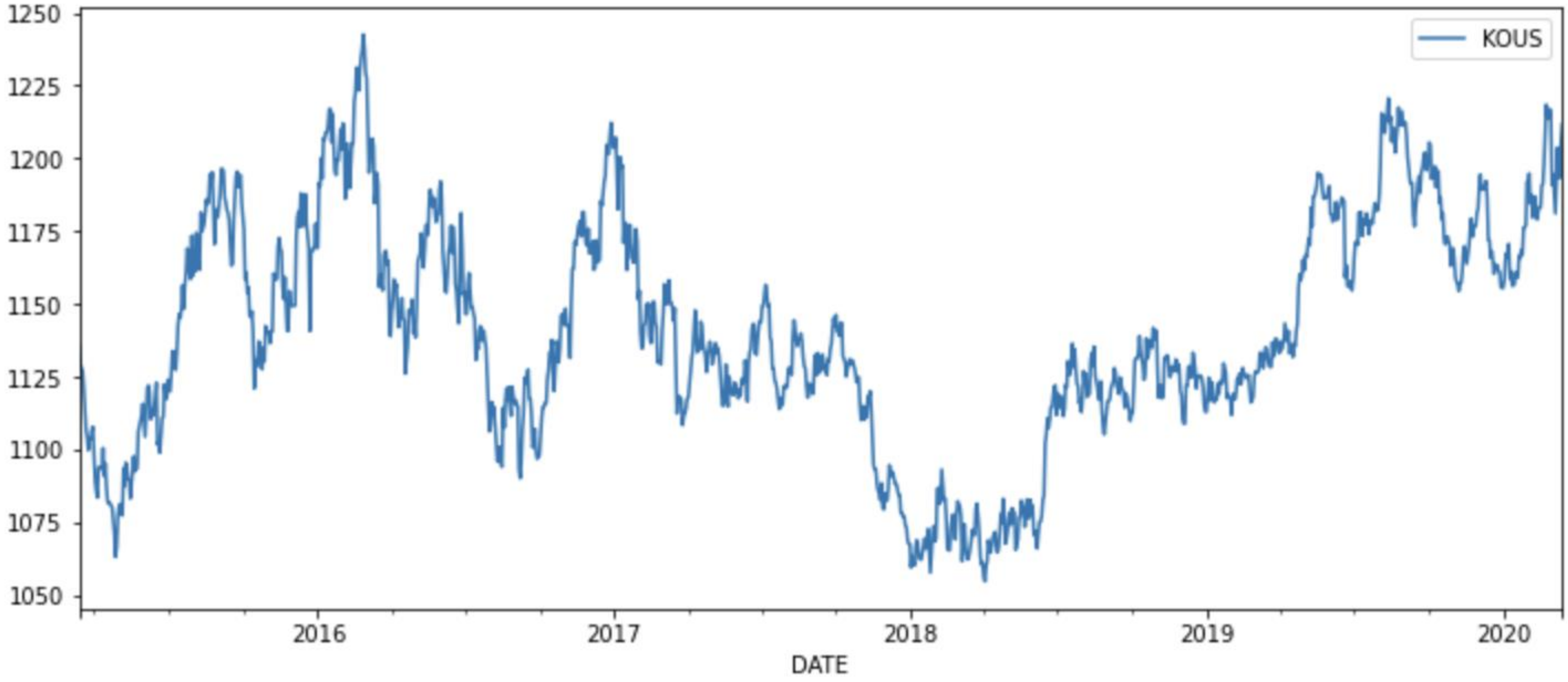
In [19]:
```
diff_2 = diff_1.diff().dropna()
plt.plot(diff_2)
plt.show()
```



2차 차분

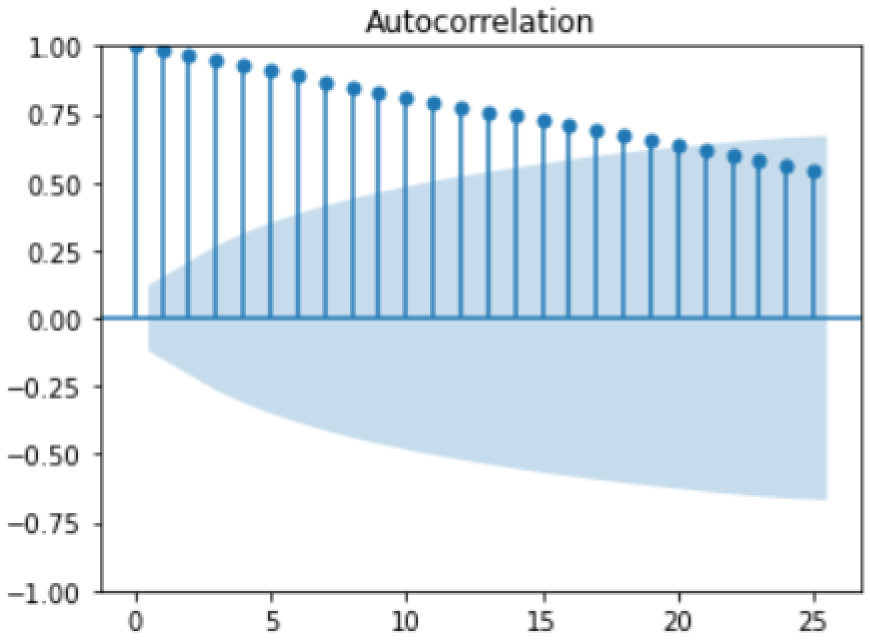
Time Series Modeling: ARIMA

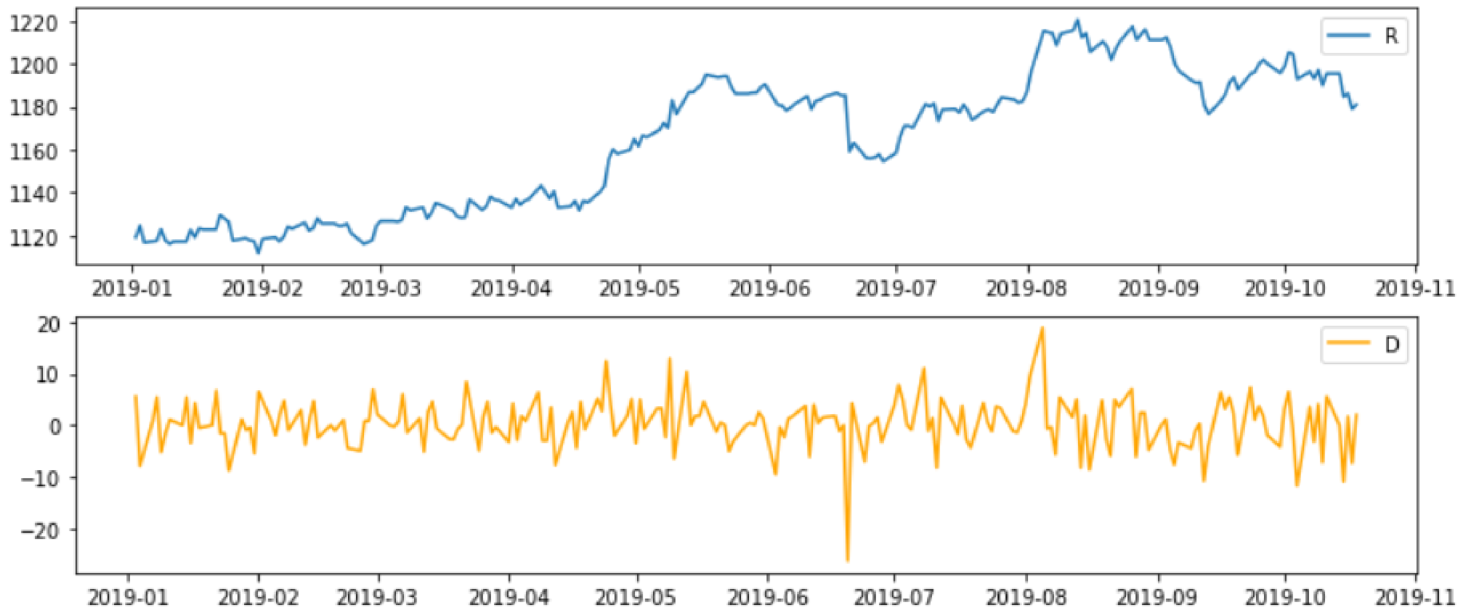
Forecasting with ARIMA models



Time Series Modeling: ARIMA

Forecasting with ARIMA models





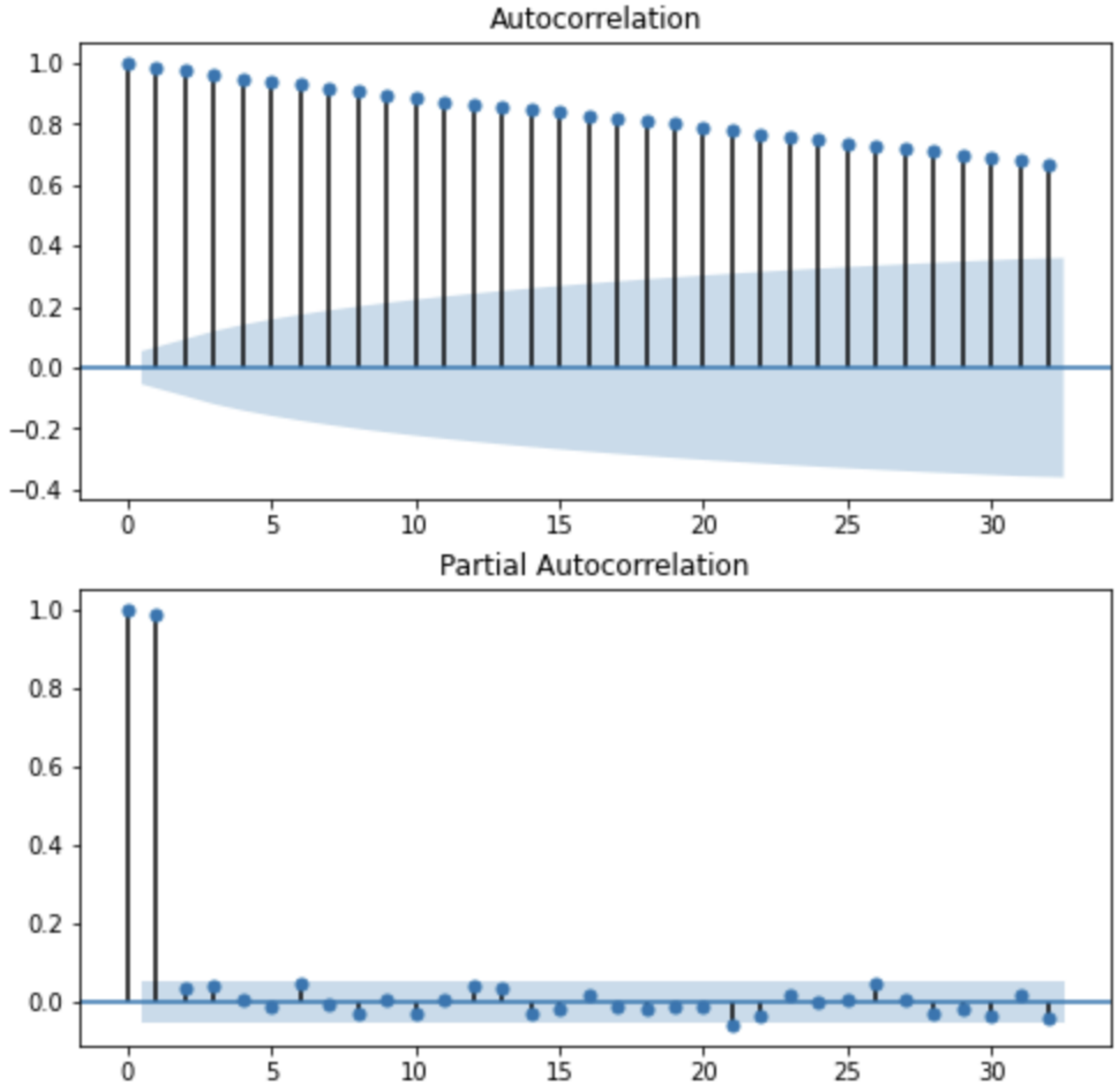
Time Series Modeling: ARIMA

Forecasting with ARIMA models

Model	ACF	Partial ACF
MA(q)	Cut off after lag q (q시차 이후 0으로 절단)	Die out (지수적으로 감소, 소멸하는 sine함수 형태)
AR(p)	Die out (지수적으로 감소, 소멸하는 sine함수 형태)	Cut off after lag p (p시차 이후 0으로 절단)
ARMA(p, q)	Die out (시차 (q-p)이후 부터 소멸)	Die out (시차 (q-p)이후 부터 소멸)

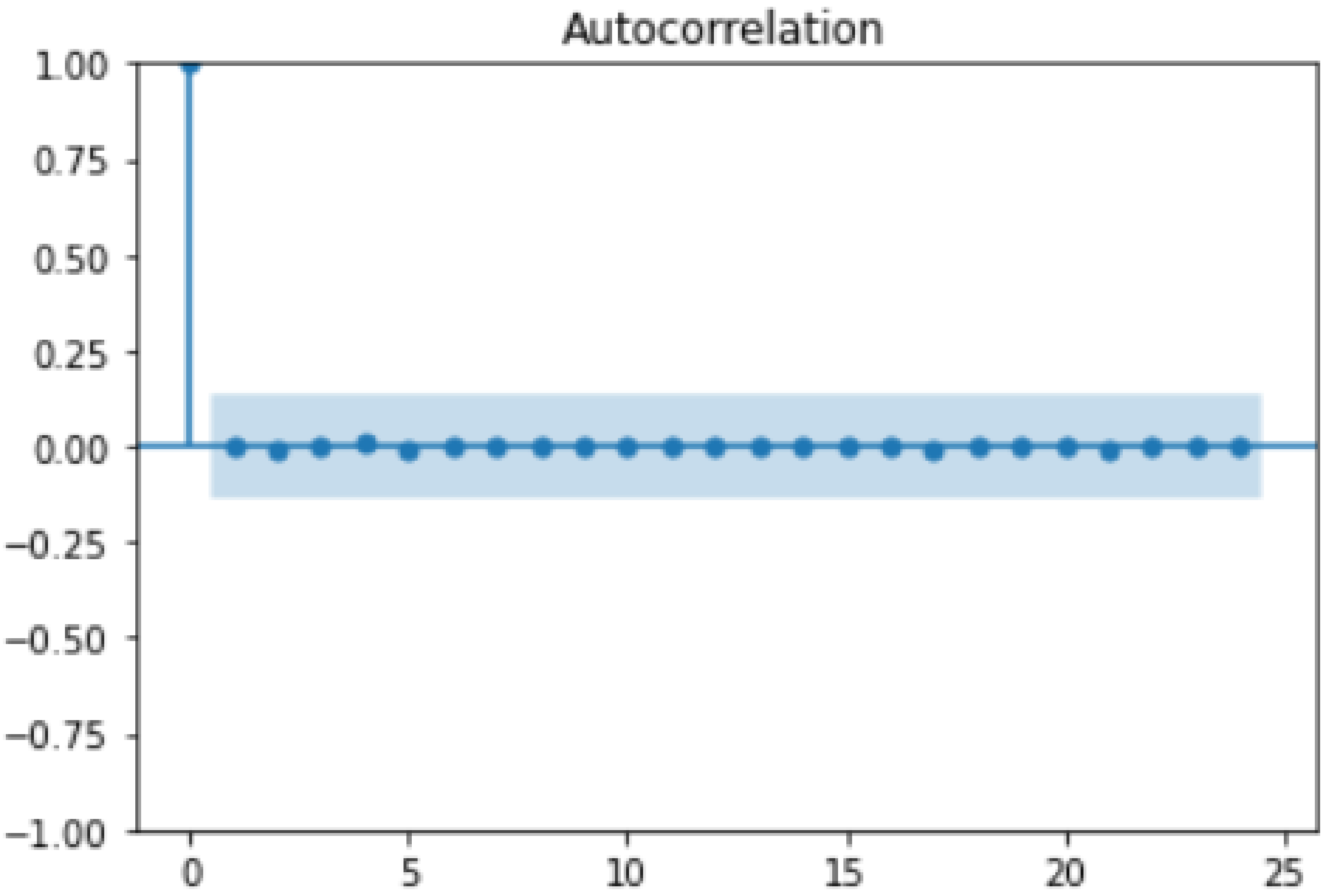
Time Series Modeling: ARIMA

Forecasting with ARIMA models



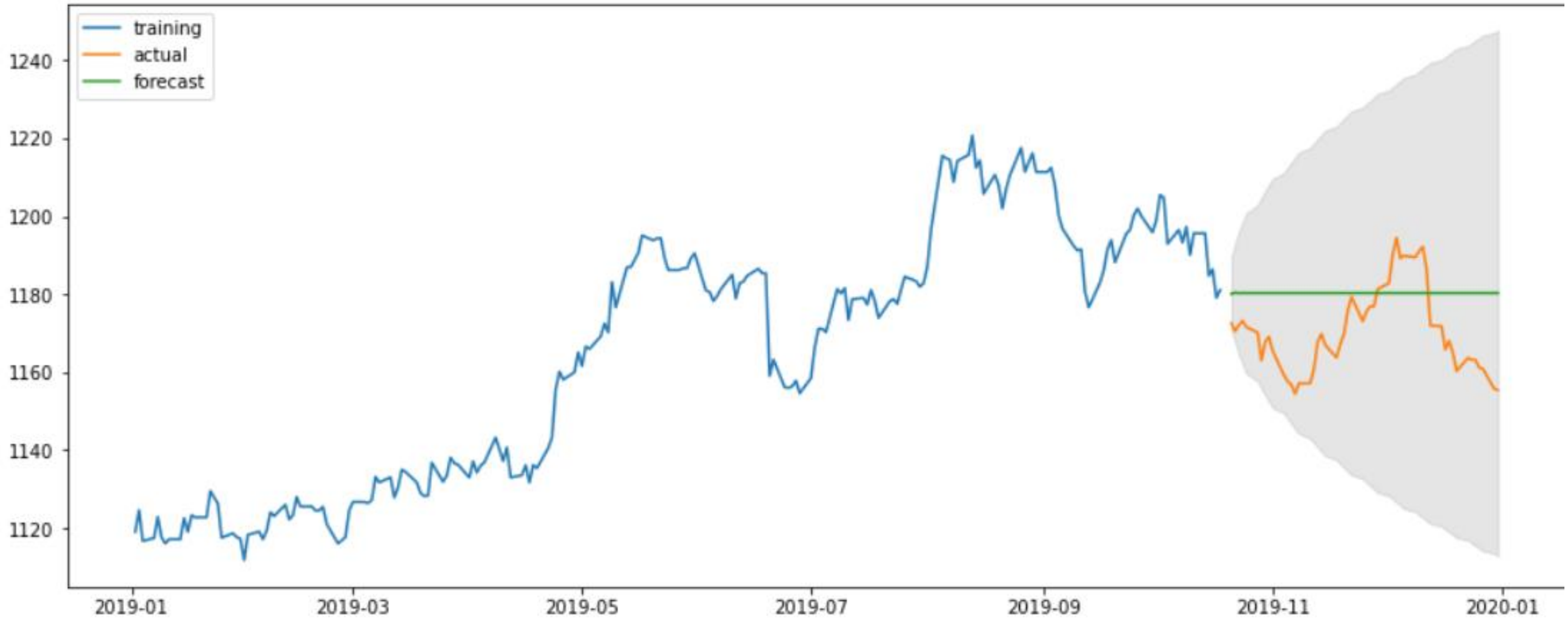
Time Series Modeling: ARIMA

Forecasting with ARIMA models



Time Series Modeling: ARIMA

Forecasting with ARIMA models



Time Series Modeling: Seasonal ARIMA

Introduction to Seasonal ARIMA (SARIMA) models

SARIMA(Seasonal Auto Regressive Integrated Moving Average)?

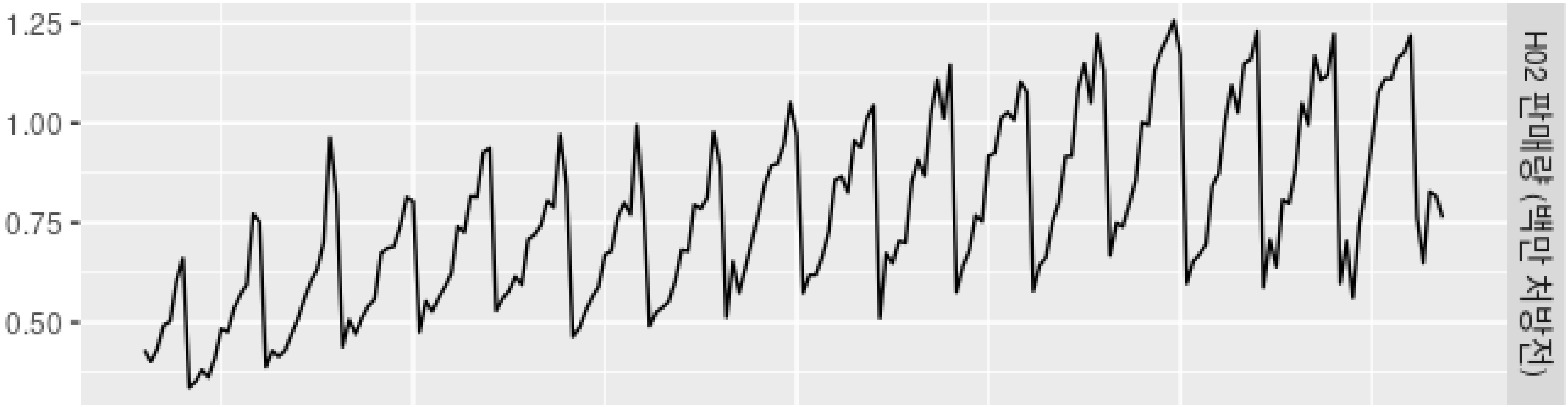
->계절성 데이터를 모델링 하기위한 ARIMA 모델

->m은 매년 관측 값의 개수

$$\text{ARIMA}\quad \underbrace{(p, d, q)}_{\text{모델의 비 계절성 부분}} \quad \underbrace{(P, D, Q)_m}_{\text{모델의 계절성 부분}}$$

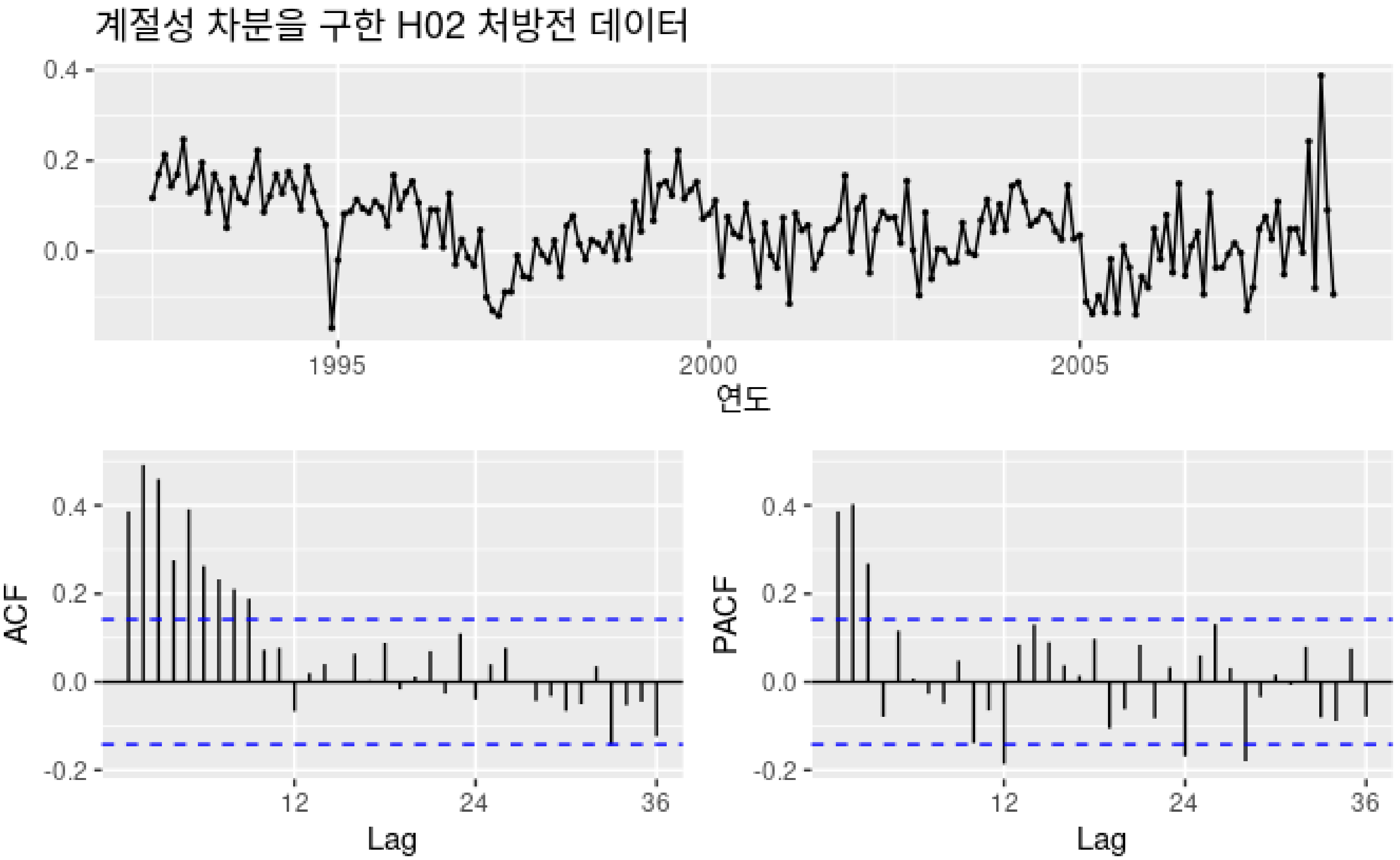
Time Series Modeling: Seasonal ARIMA

Forecasting with SARIMA models



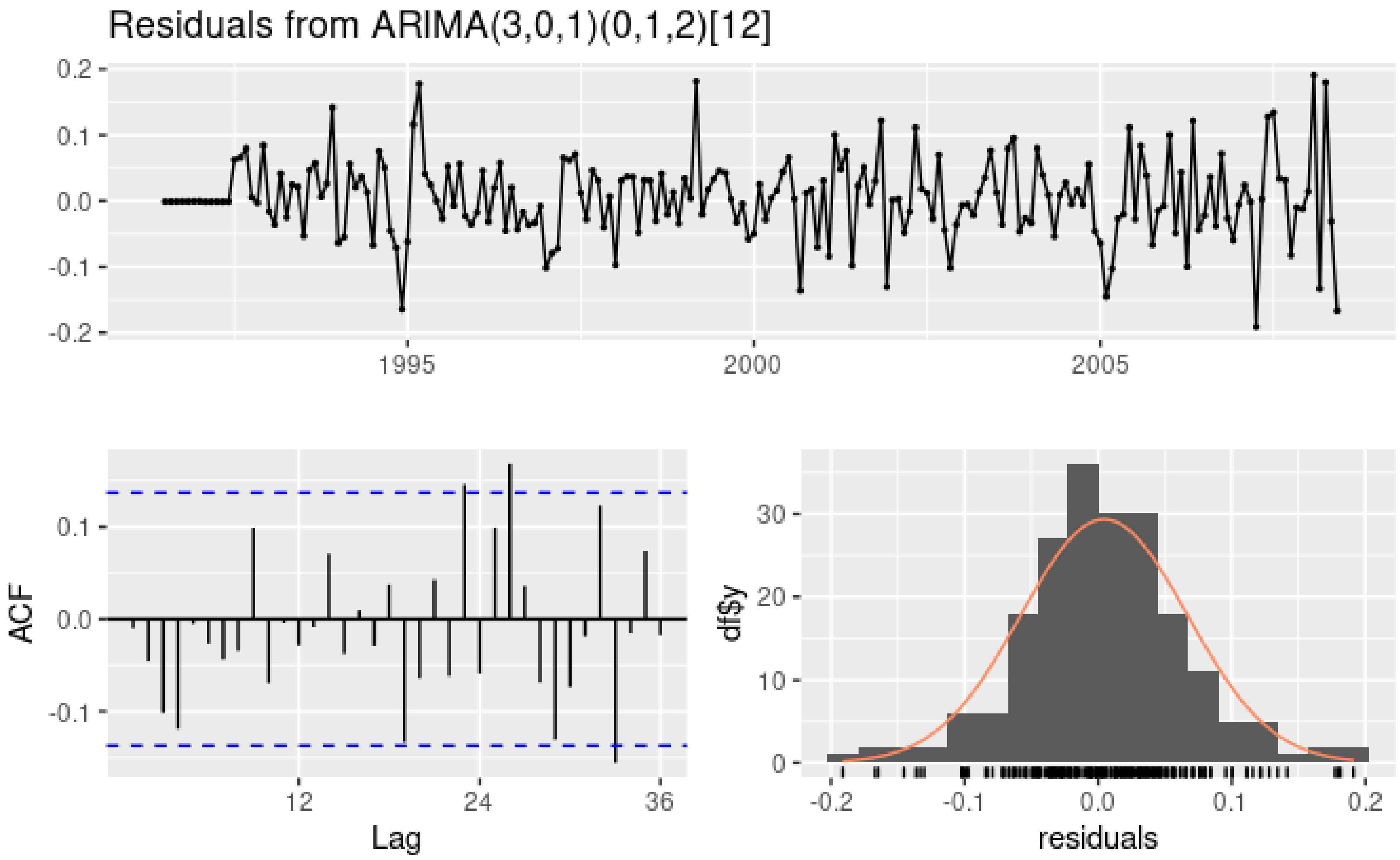
Time Series Modeling: Seasonal ARIMA

Forecasting with SARIMA models



Time Series Modeling: Seasonal ARIMA

Forecasting with SARIMA models



Time Series Modeling: Seasonal ARIMA

Forecasting with SARIMA models