
아파트 매매 실거래 가격지수에 관한 다각적 분석



2019310508 김수빈

2019310157 박지성

INDEX

0. 주제 선정 배경

1. 기간별 분석

2. 단변량 분석

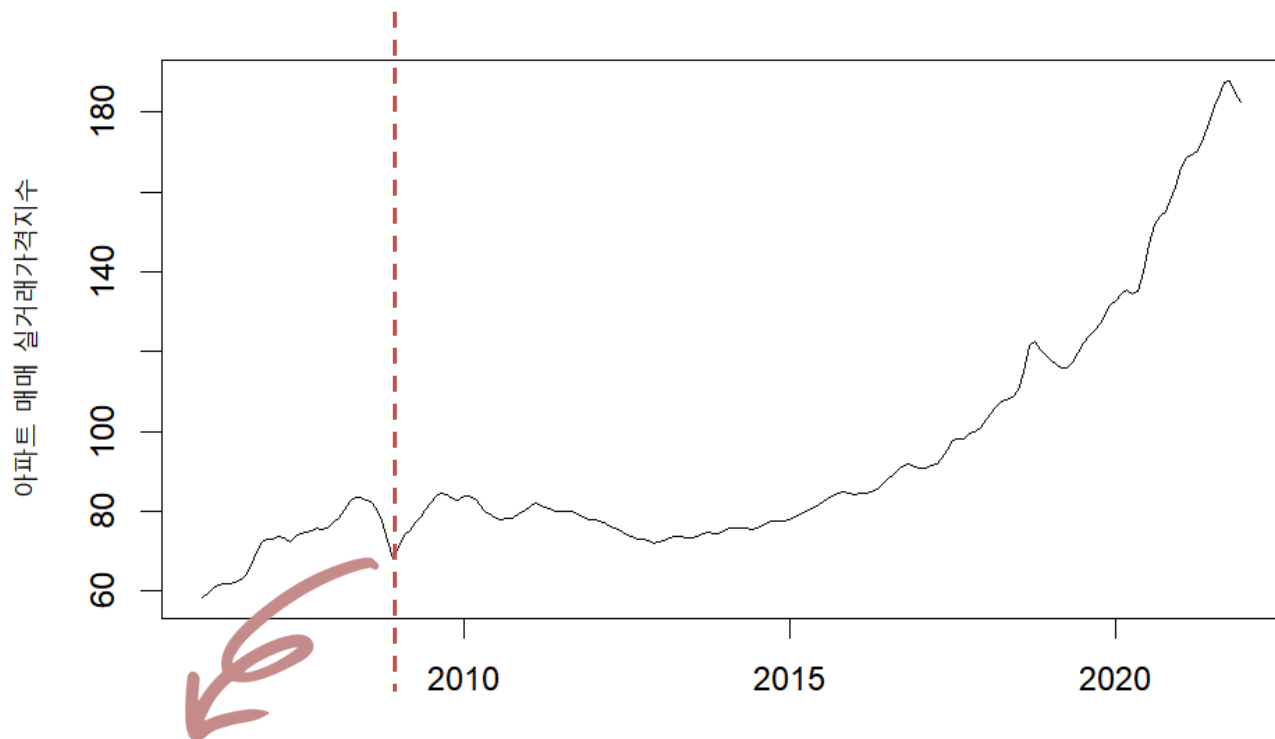
3. 다변량 분석

4. 결론

0

주제 선정 배경

주제 선정



2008년도 미국의 서브프라임모기지 금융위기의 충격으로 발생한 주택시장의 가격하락과 이에 따르는 전세가격의 상승으로 인해서 주택시장에 여러 가지 문제점 이 발생

주제 선정



- 여러가지 부동산제도의 도입과 경제 사건 등으로 부동산 시장에 많은 변화를 가져옴
- 특히, 아파트 거래는 큰 폭의 가격변동으로 인해 주택시장은 더욱더 불규칙일 것으로 예상

분석 목표 및 흐름

분석 목표

변동이 잦은 부동산 시장 파악의 필요성을 느껴 여러 가지
시계열 방법을 이용하여 아파트 가격을 예측하고자 함

분석 흐름

- ① 시계열 특성이 다른 두 기간을 구분하여 기간별 특성 파악
- ② 단변량 분석 및 예측 진행
- ③ 다변량 분석 및 예측 진행
- ④ 모델들 간 비교 분석을 통한 인사이트 도출

데이터 소개

아파트 매매 실거래가격지수란?

전국의 아파트 및 연립. 다세대를 대상으로 하여 실제 거래되어
신고된 아파트 및 연립. 다세대의 거래 가격수준

날짜	가격지수
2006-01-01	58.5
2006-02-01	59.4
2006-03-01	60.9
2006-04-01	61.6
2006-05-01	61.9
2006-06-01	61.8

...

2021-06-01	184.5
2021-07-01	181.3
2021-08-01	184.5
2021-09-01	187.3
2021-10-01	187.8
2021-11-01	184.7
2021-12-01	182.6

출처: 한국부동산원 부동산통계정보시스템(R-ONE)

1

기간에 따른 분석

1

기간에 따른 분석

기간 구분 기준



기간 구분 기준

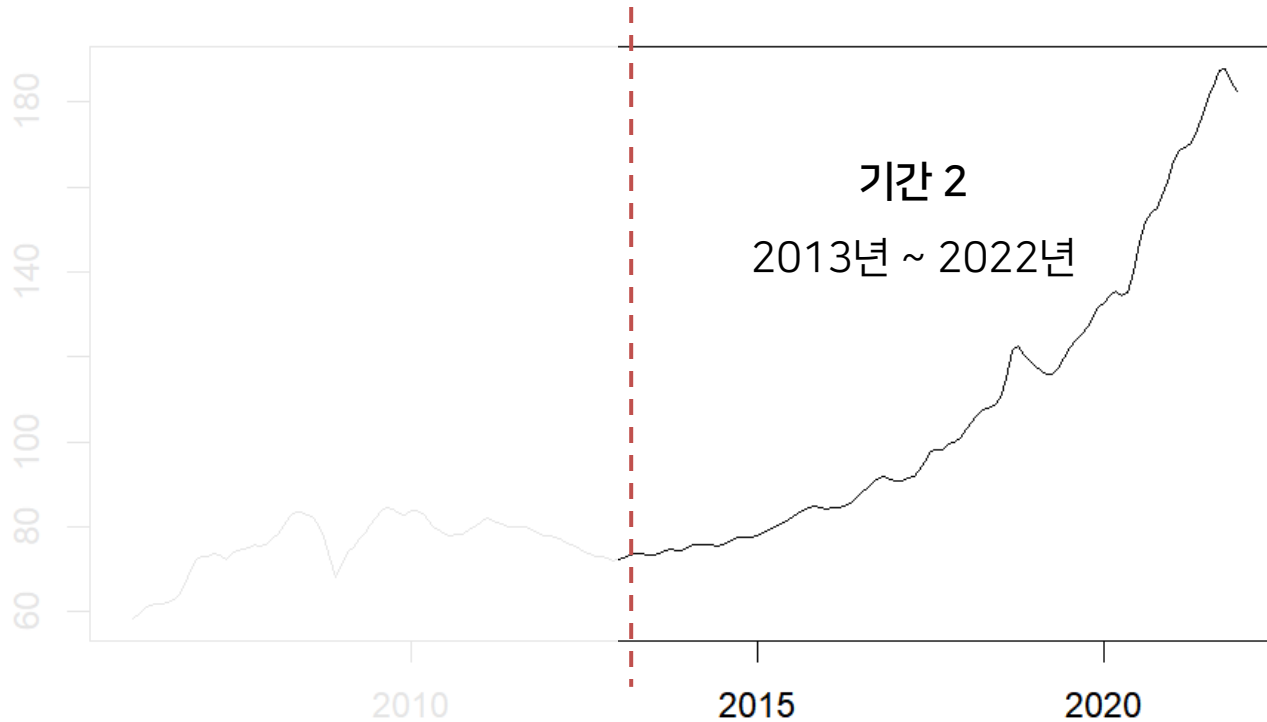


- 2008년에 금융위기로 인하여 아파트 매매 거래 가격 지수가 급하락한 것을 확인
- 그렇지만 금방 경향성을 회복하였다는 점에서 정상성 특징을 지닐 수 있을 것으로 예상

1

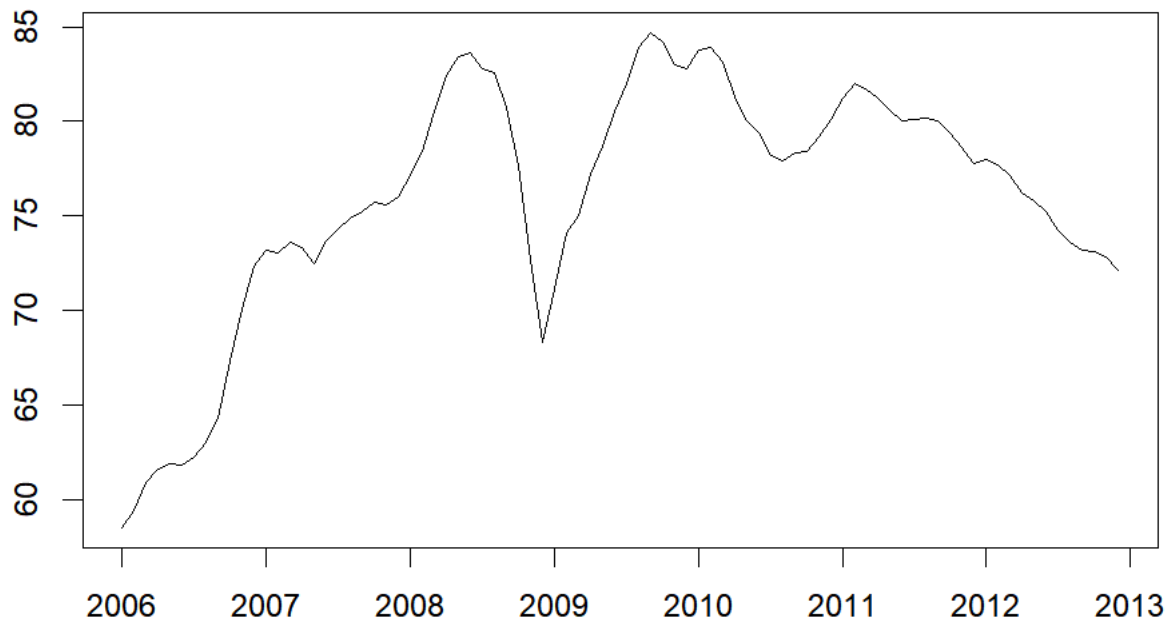
기간에 따른 분석

기간 구분 기준



지속적으로 상승하는 경향성 (linear trend)가 보임

기간 1



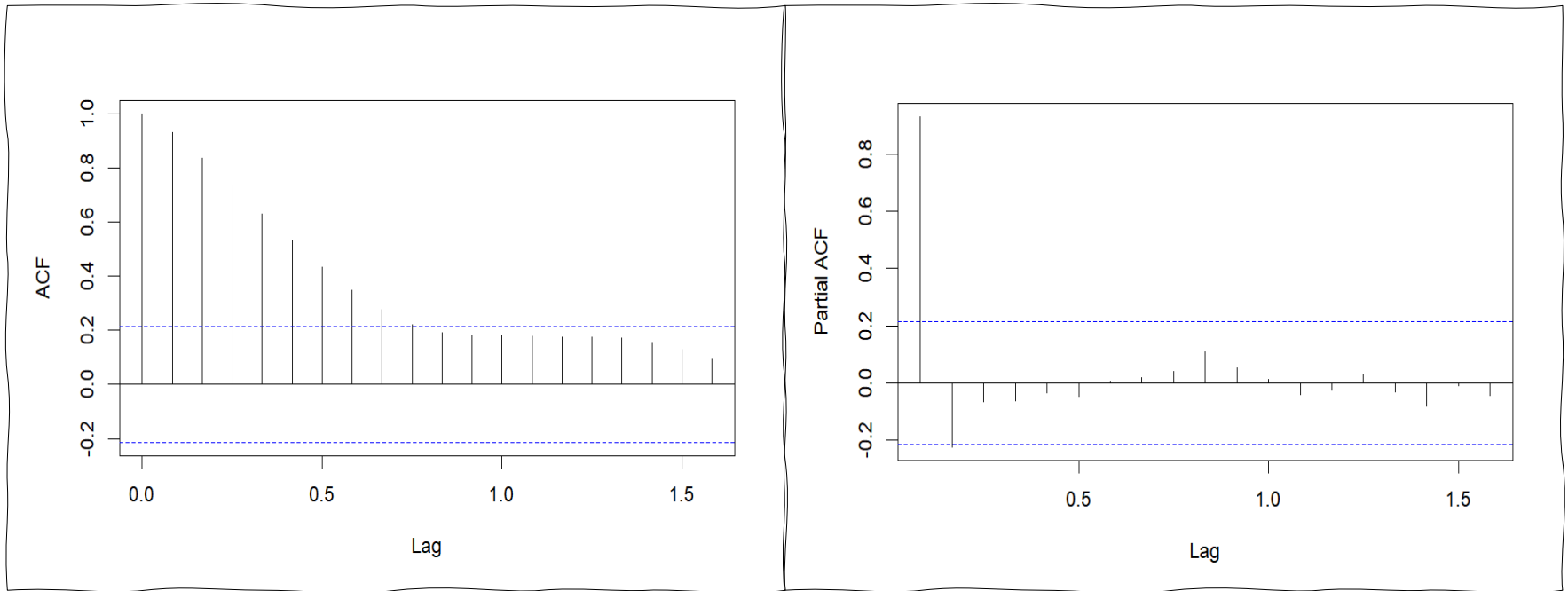
정상성을 띄고 있지 않고, 추세가 존재한다는 것을 알 수 있음

✱ 정상성 검정을 통해 확인이 필요

1

시간에 따른 분석

기간 1 정상성 검정: ACF, PACF 그래프 확인



ACF가 느리게 감소하고, lag 대부분 신뢰구간 밖에 있는 것을 확인

이는 비정상성 특징을 표시

기간 1 정상성 검정: Ljung-Box, ADF, KPSS 검증

Box-Ljung test

```
data: t1  
X-squared = 315.74, df = 20, p-value < 2.2e-16
```

Augmented Dickey-Fuller Test

```
data: t1  
Dickey-Fuller = -2.0533, Lag order = 4, p-value = 0.554  
alternative hypothesis: stationary
```

KPSS Test for Level Stationarity

```
data: t1  
KPSS Level = 0.8976, Truncation lag parameter = 3, p-value = 0.01
```

1

기간에 따른 분석

기간 1 정상성 검정: Ljung-Box, ADF, KPSS 검증

Box-Ljung test

```
data: t1  
X-squared = 315.74, df = 20, p-value < 2.2e-16
```

Augmented Dickey-Fuller Test

```
data: t1  
Dickey-Fuller = -2.0533, Lag order = 4, p-value = 0.554  
alternative hypothesis: stationary
```

KPSS Test for Level Stationarity

```
data: t1  
KPSS Level = 0.8976, Truncation lag parameter = 3, p-value = 0.01
```



차분을 통한 정상성 변환 필요

기간 1 정상성 변환



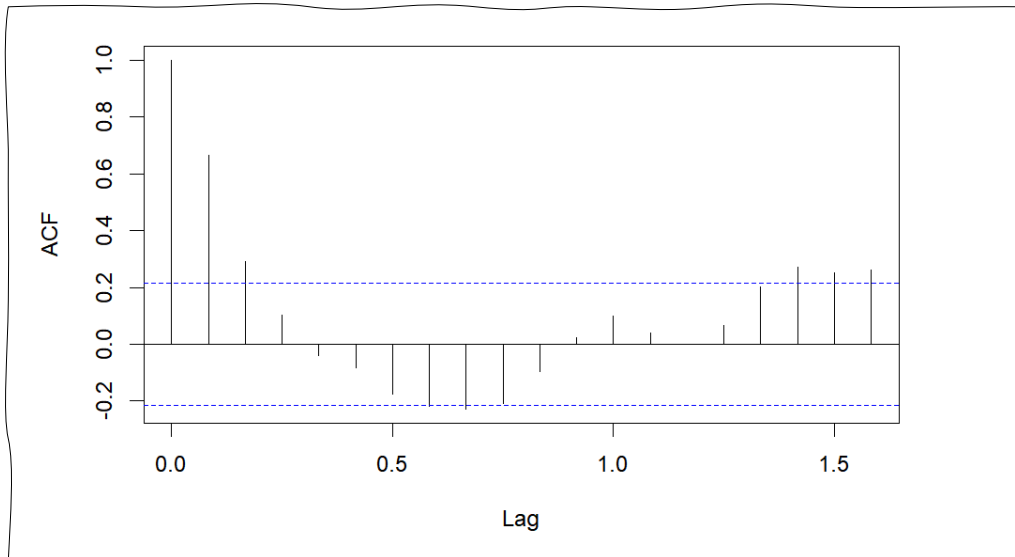
ndiff() 함수: 적당한 차분 횟수를 정하는 방법

확인 결과, 1차 차분하여 진행

1

시간에 따른 분석

기간 1 정상성 재확인



Augmented Dickey-Fuller Test

```
data: t1_fd  
Dickey-Fuller = -3.9592, Lag order = 4, p-value = 0.01557  
alternative hypothesis: stationary
```

KPSS Test for Level Stationarity

```
data: t1_fd  
KPSS Level = 0.40578, Truncation lag parameter = 3, p-value = 0.07466
```

- ACF 빠르게 0으로 감소
- ADF 검정 결과, $p\text{-값} < 0.05$
- KPSS 검정 결과, $p\text{-값} > 0.05$



정상성 확보!

기간 1 모델링

```
Series: t1_1
ARIMA(0,0,1)(0,0,1)[12] with non-zero mean

Coefficients:
          ma1      sma1      mean
          0.8357  0.3458  0.6630
s.e.      0.0783  0.1809  0.3855

sigma^2 = 1.358:  log likelihood = -73.66
AIC=155.32  AICc=156.27  BIC=162.72
```

결과

- 제일 적합한 모델:
ARIMA(0,0,1)(0,0,1)[12]
- AIC값: 155.32

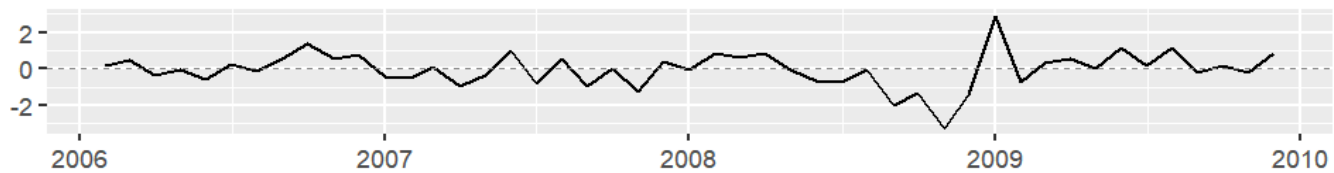
auto.arima() 함수를 이용하여 제일 적합한 모델 선정

1

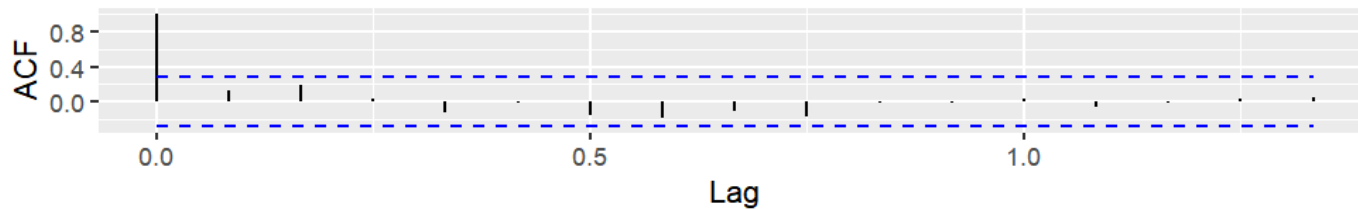
시간에 따른 분석

기간 1 모델링: 잔여 진단

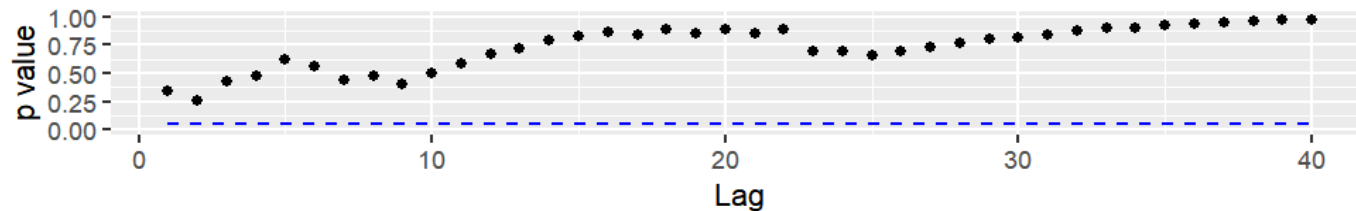
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic

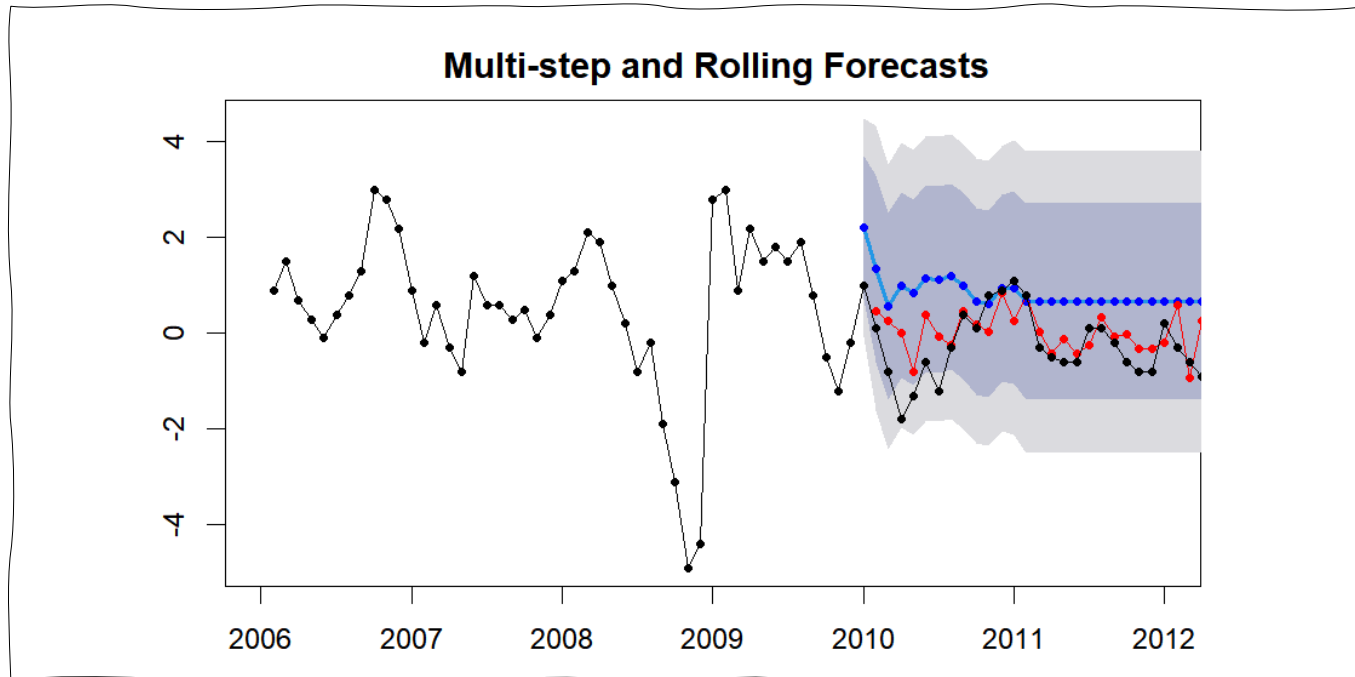


잔여 값들 간 상관관계 가 없음을 확인하여 Forecasting 진행 가능

1

시간에 따른 분석

기간 1 모델링: Forecasting



	Multi_Step	Rolling
RMSE	1.27316	0.6627419

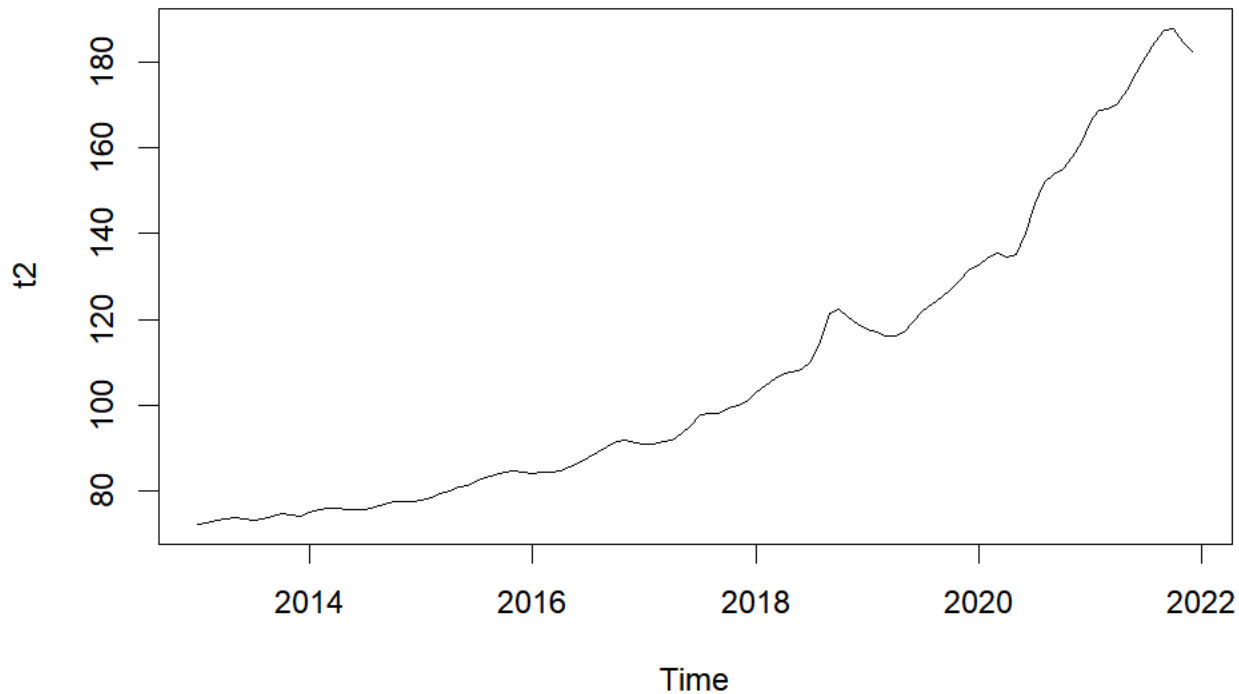
Rolling Forecast 정확도가 더 높음



1

기간에 따른 분석

기간 2 시계열 그래프



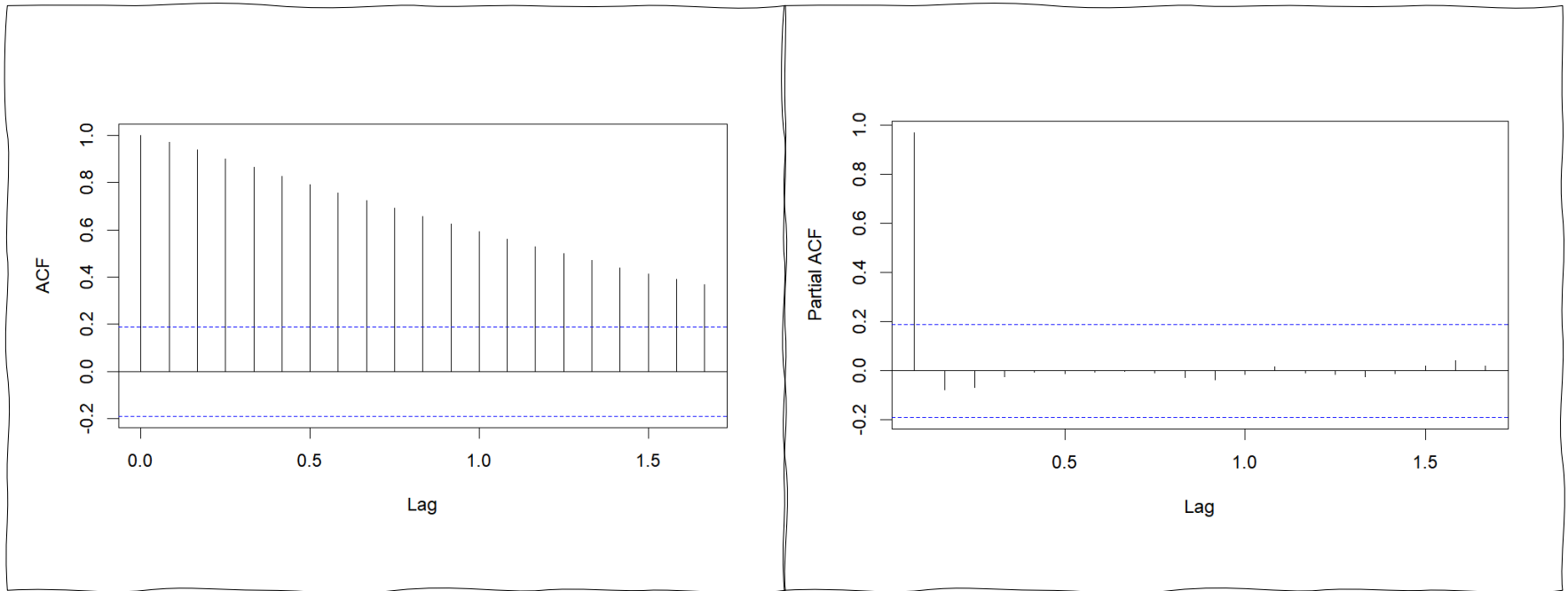
정상성을 띄고 있지 않고, 상승하는 추세는 존재한다는 것을 알 수 있음

이 또한 **정상성 검정**을 통해 확인이 필요

1

기간에 따른 분석

기간 2 정상성 검정: ACF, PACF 그래프 확인



ACF가 매우 느리게 감소하고, 모든 lag가 신뢰구간 밖에 있는 것을 확인
이는 비정상성 특징을 표시

1

기간에 따른 분석

기간 2 정상성 검정: Ljung-Box, ADF, KPSS 검증

Box-Ljung test

data: t2
X-squared = 1083.5, df = 20, p-value < 2.2e-16

Augmented Dickey-Fuller Test

data: t2
Dickey-Fuller = -0.46023, Lag order = 4, p-value = 0.9821
alternative hypothesis: stationary

KPSS Test for Level Stationarity

data: t2
KPSS Level = 2.0362, Truncation lag parameter = 4, p-value = 0.01



비정상성 판단

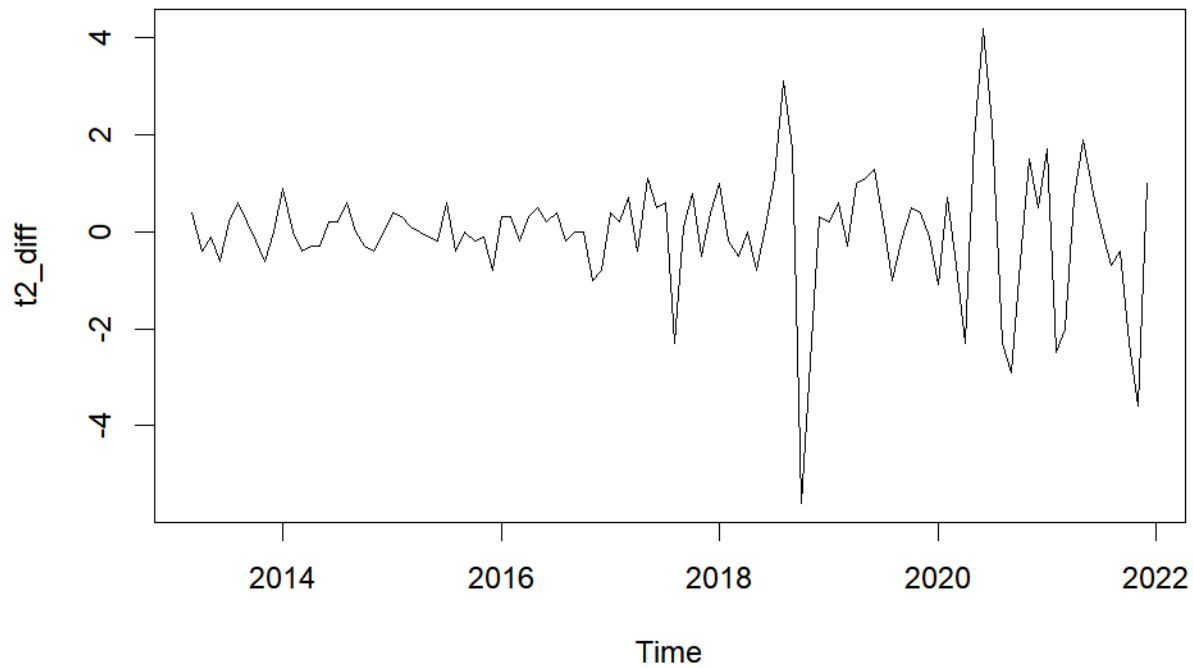


차분을 통한 정상성 변환 필요

1

기간에 따른 분석

기간 2 정상성 변환

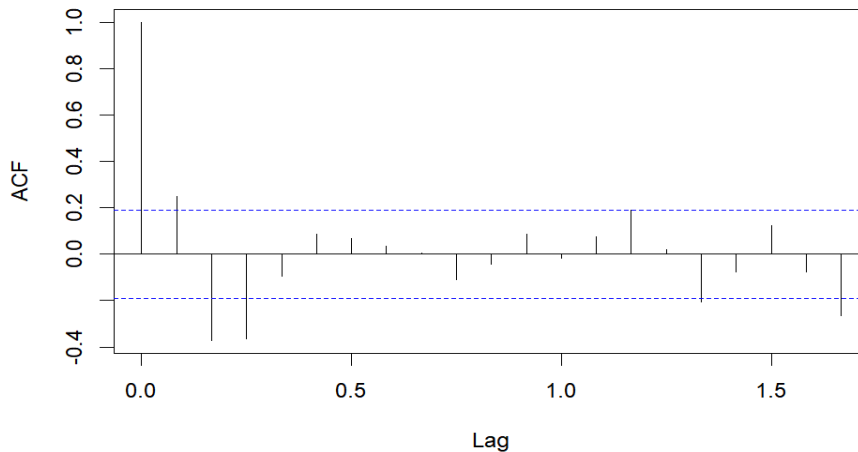


ndiff() 확인 결과, 2차 차분하여 진행

1

시간에 따른 분석

기간 1 정상성 재확인



Augmented Dickey-Fuller Test

```
data: t2_diff  
Dickey-Fuller = -6.7426, Lag order = 4, p-value = 0.01  
alternative hypothesis: stationary
```

KPSS Test for Level Stationarity

```
data: t2_diff  
KPSS Level = 0.075777, Truncation lag parameter = 4, p-value = 0.1
```

- ACF 매우 빠르게 0으로 감소
- ADF 검정 결과, $p\text{-value} < 0.05$
- KPSS 검정 결과, $p\text{-value} > 0.05$



정상성 확보

기간 2 모델링

```
Series: t2_1
ARIMA(2,0,1) with zero mean

Coefficients:
          ar1      ar2      ma1
      0.8087 -0.6432 -0.8800
s.e.  0.0983  0.1077  0.0561

sigma^2 = 0.595:  log likelihood = -80.8
AIC=169.6  AICc=170.21  BIC=178.59
```

결과

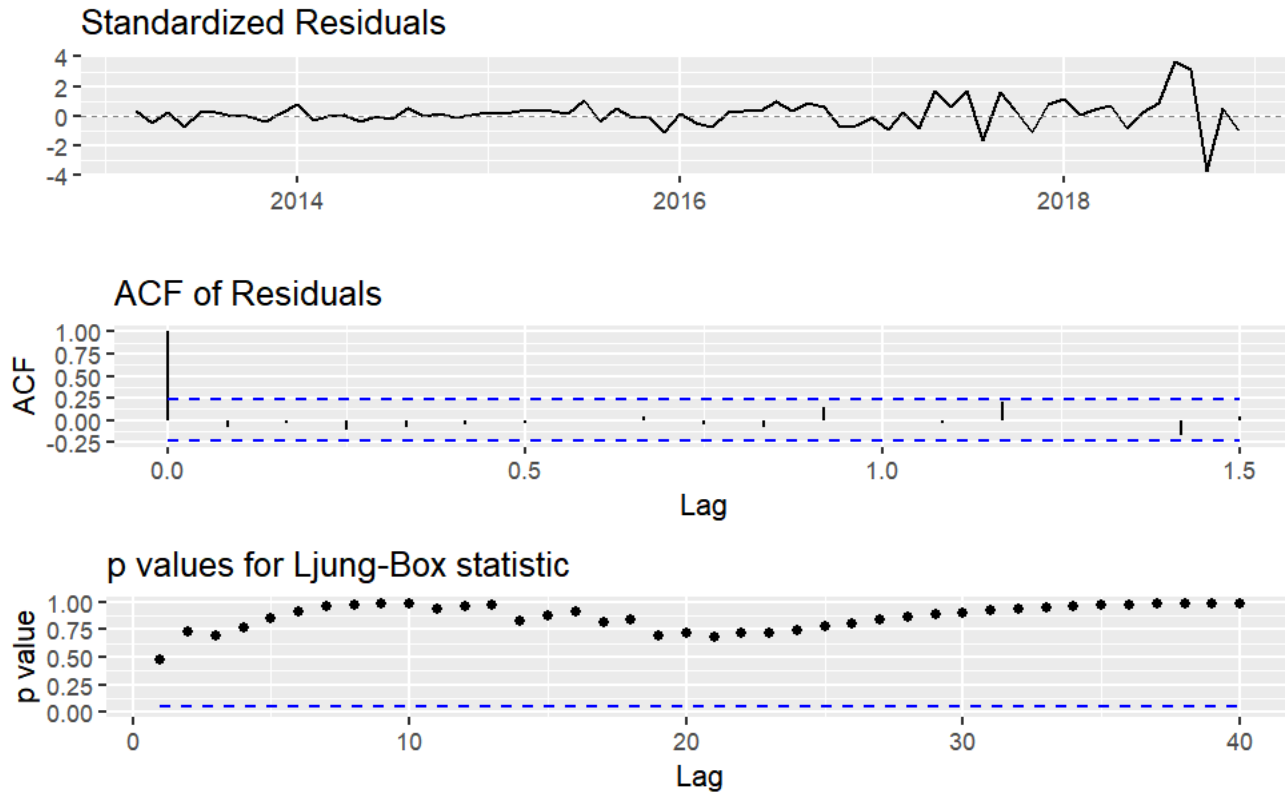
- 제일 적합한 모델:
ARIMA(2,0,1)
- AIC값: 169.6

auto.arima() 함수를 이용하여 제일 적합한 모델 선정

1

시간에 따른 분석

기간 2 모델링: 잔여 진단

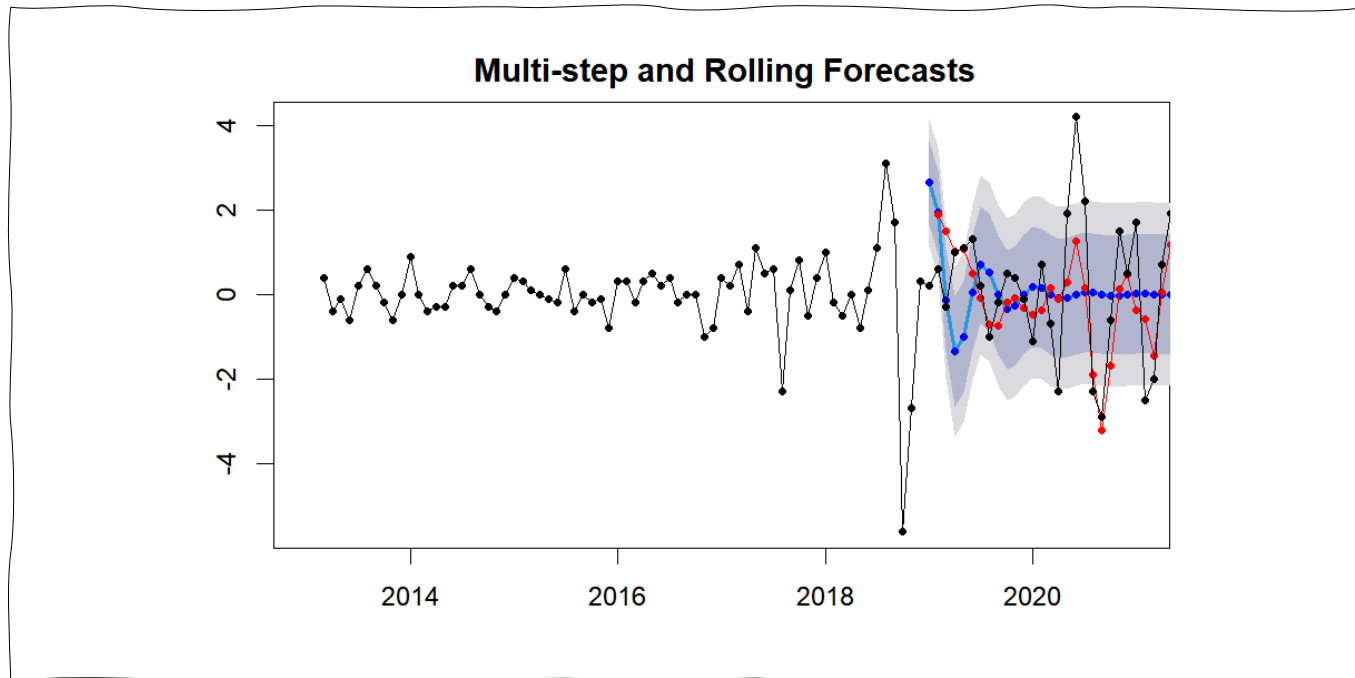


잔여 값들 간 상관관계 가 없음을 확인하여 Forecasting 진행 가능

1

시간에 따른 분석

기간 2 모델링: Forecasting



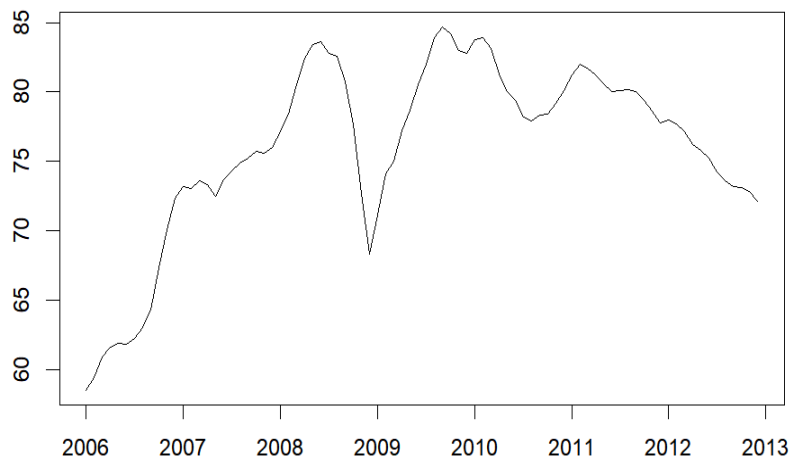
	Multi_Step	Rolling
RMSE	1.27316	0.6178813

Rolling Forecast 정확도가 더 높음



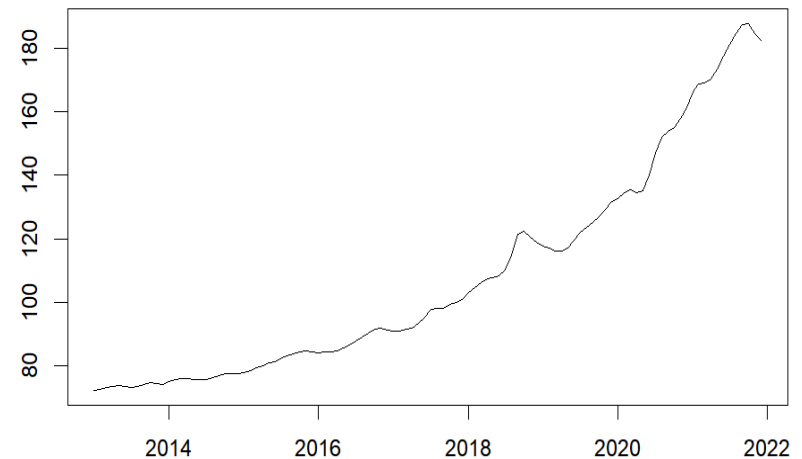
두 모델 간 시계열 특징 비교

기간 1



- Shock이 존재하나 회복하는 양상
 - 계절성 없음

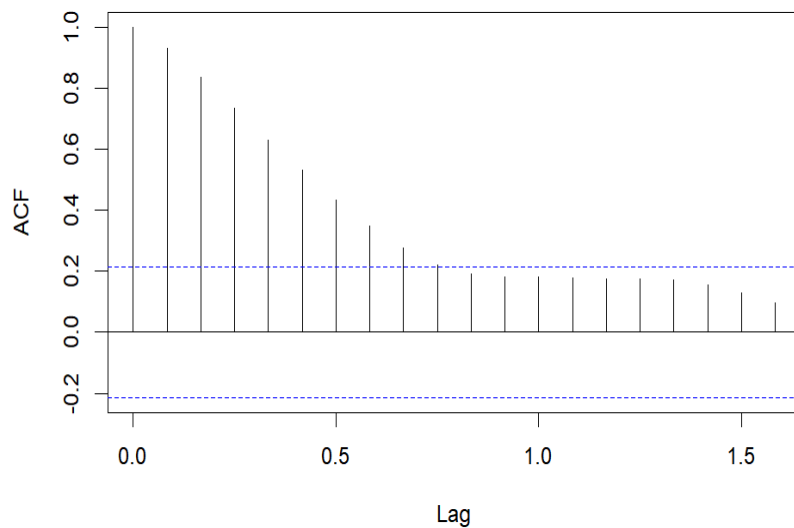
기간 2



- 상승 추세
- 계절성 없음

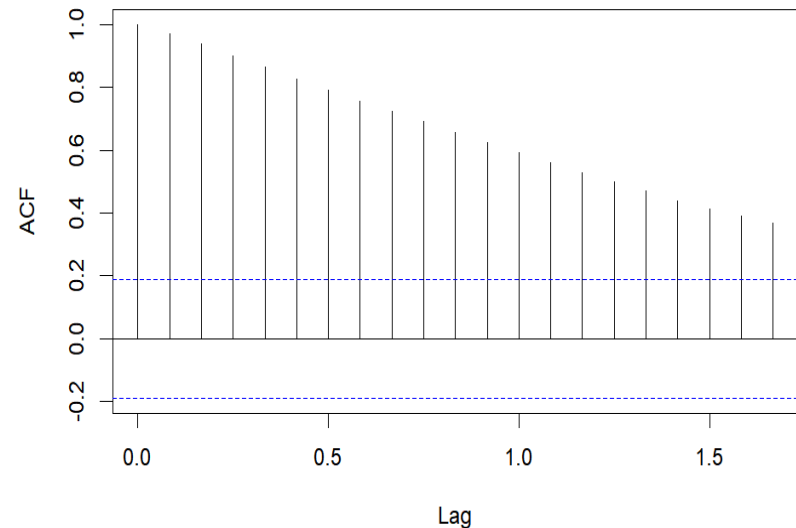
두 모델 간 시계열 특징 비교

기간 1



- ACF가 0으로 느리게 감소
- 1차 차분해서 정상성 확보 후 예측

기간 2

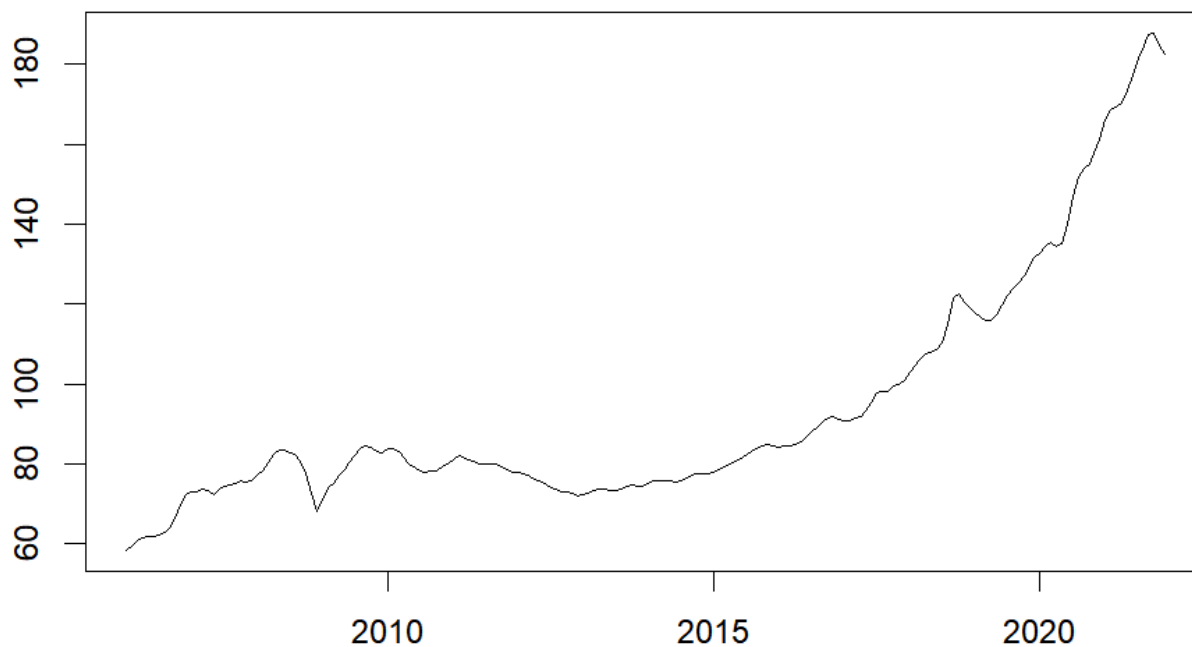


- ACF가 0으로 더 느리게 감소
- 1차 차분으로 정상성의 띄지 않아,
2차 차분하여 추가 진행

2

단변량 분석

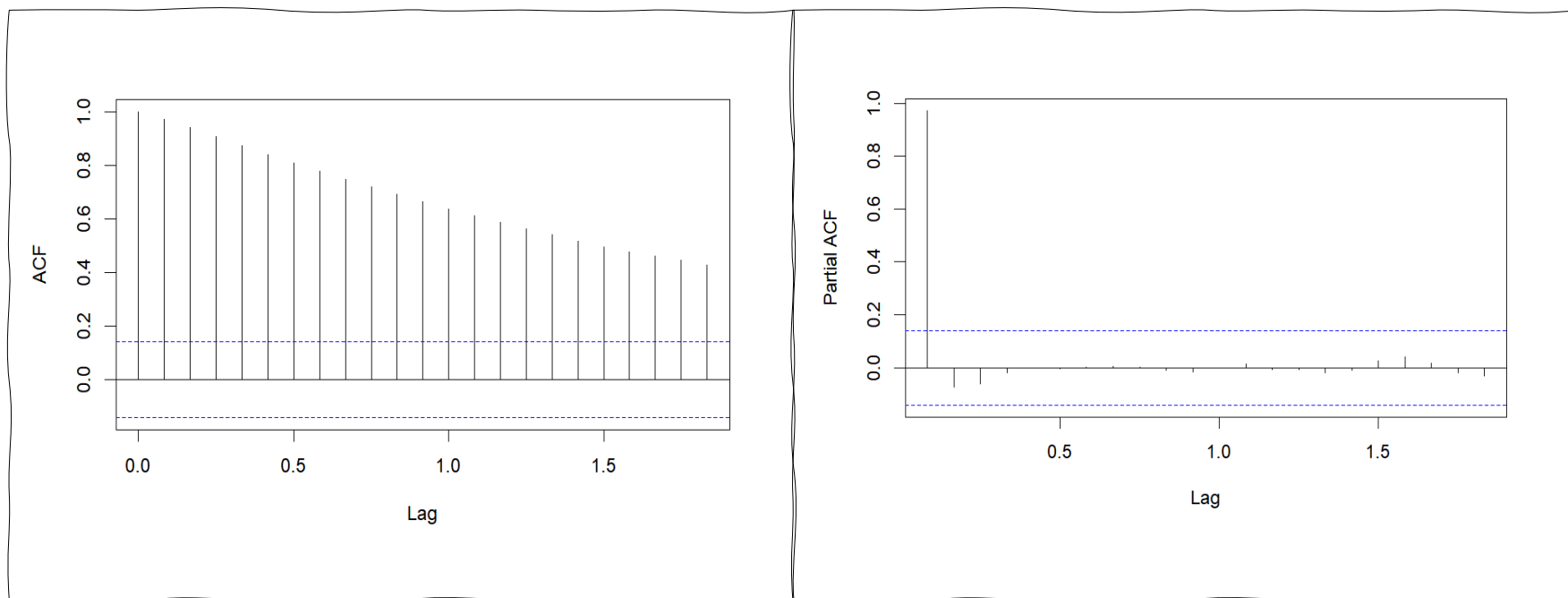
전범위 기간 시계열 그래프



정상성을 띄고 있지 않고, 추세 있다는 것을 확인 가능

✖ 정상성 검정을 통해 확인이 필요

정상성 검정: ACF, PACF 그래프 확인



ACF가 매우 느리게 감소하고, lag 모두 신뢰구간 밖에 있는 것을 확인
이는 비정상성 특징을 표시

정상성 검정: Ljung-Box, ADF, KPSS 검증

Box-Ljung test

data: df_Y
X-squared = 2037.4, df = 20, p-value < 2.2e-16

Augmented Dickey-Fuller Test

data: df_Y
Dickey-Fuller = 0.75281, Lag order = 5, p-value = 0.99
alternative hypothesis: stationary

KPSS Test for Level Stationarity

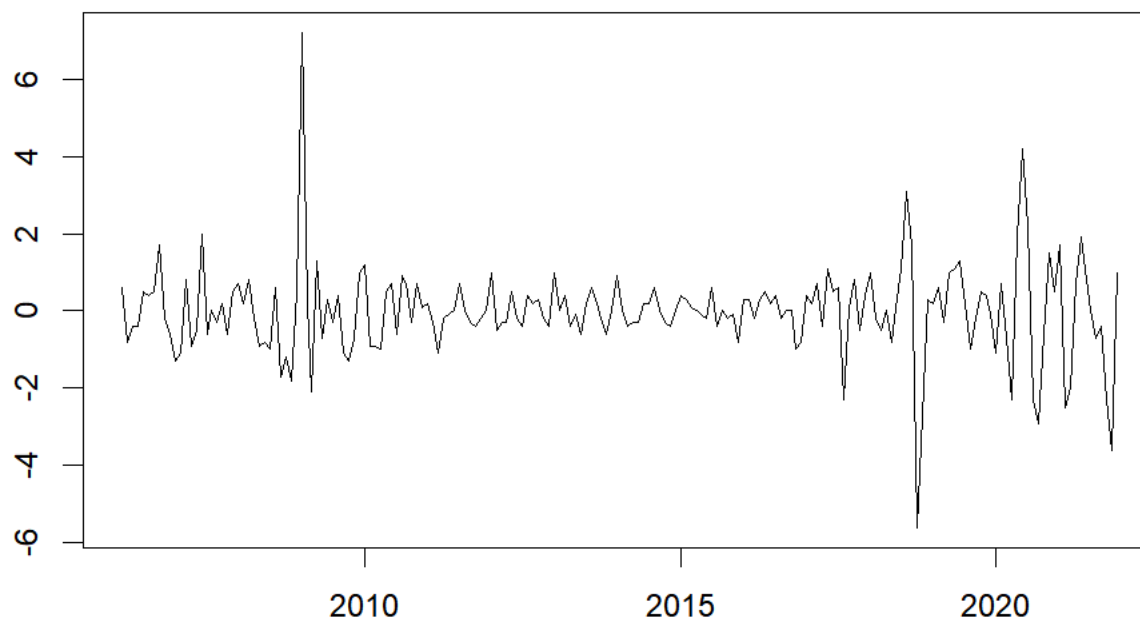
data: df_Y
KPSS Level = 2.6949, Truncation lag parameter = 4, p-value = 0.01



비정상성 판단

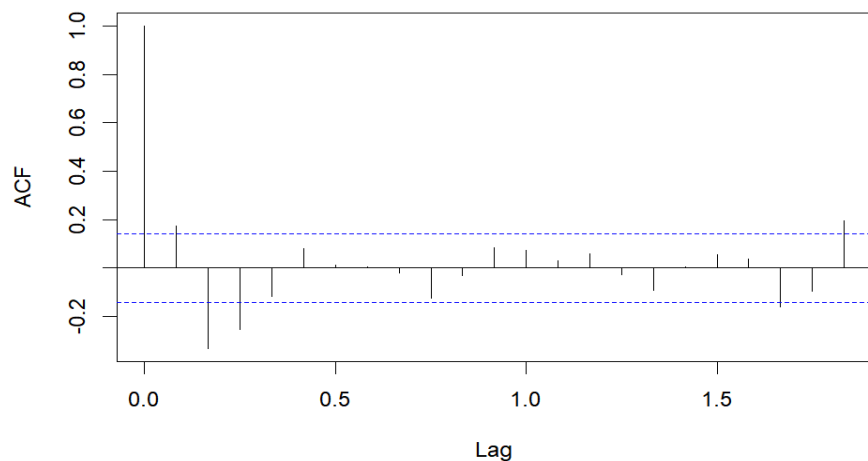


차분을 통한 정상성 변환 필요

정상성 변환 - `ndiff()`

`ndiff()` 확인 결과, 2차 차분하여 진행

정상성 재확인



Augmented Dickey-Fuller Test

```
data: df_diff
Dickey-Fuller = -8.403, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

KPSS Test for Level Stationarity

```
data: df_diff
KPSS Level = 0.033255, Truncation lag parameter = 4, p-value = 0.1
```

- ACF 매우 빠르게 0으로 감소
- ADF 검정 결과, $p\text{-값} < 0.05$
- KPSS 검정 결과, $p\text{-값} > 0.05$



정상성 확보

(1) p와 q값 직접 설정한 ARIMA 모델링

```
Call:
arima(x = df1, order = c(pstar, 0, qstar))

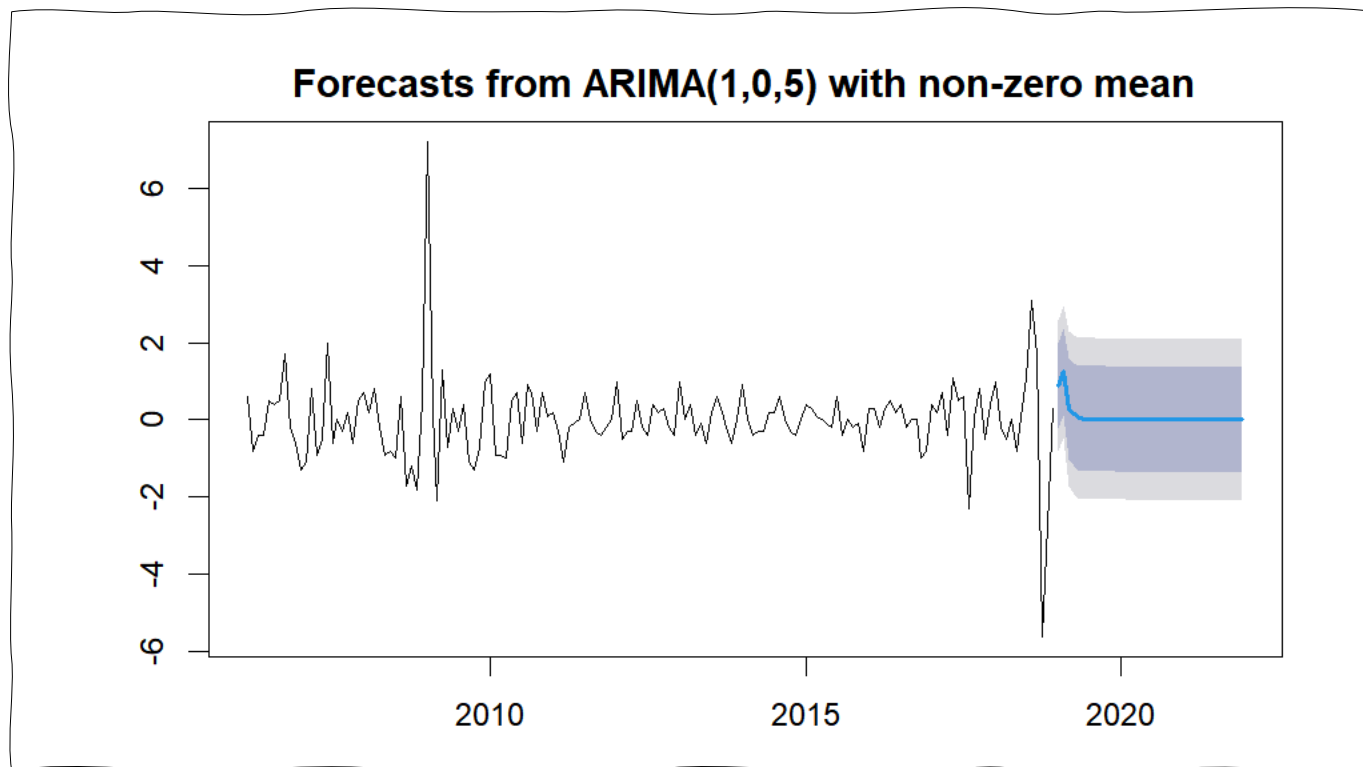
Coefficients:
          ar1          ma1          ma2          ma3          ma4          ma5  intercept
          0.874      -0.9851      -0.5252      0.2645      0.0488      0.2174           0.0010
s.e.      0.053       0.0965       0.1108       0.1321       0.1004       0.0818           0.0108

sigma^2 estimated as 0.7403:  log likelihood = -198.6,  aic = 413.2
```

Pmax와 qmax를 설정하여 제일 작은 AIC값을 가지는 모델을 선정

모델 결과: AIC값 413.2를 가지고 있는 ARIMA(1,5)

(1) p와 q값 직접 설정한 모델링



LjungBox-test를 통해 잔여 값 간 상관관계 없는 것을 확인 후 Forecasting 진행



Accuracy (RMSE): 1.61728

(2) auto.arima를 통한 모델링

```
Series: df1
ARIMA(0,0,4)(0,0,1)[12] with zero mean

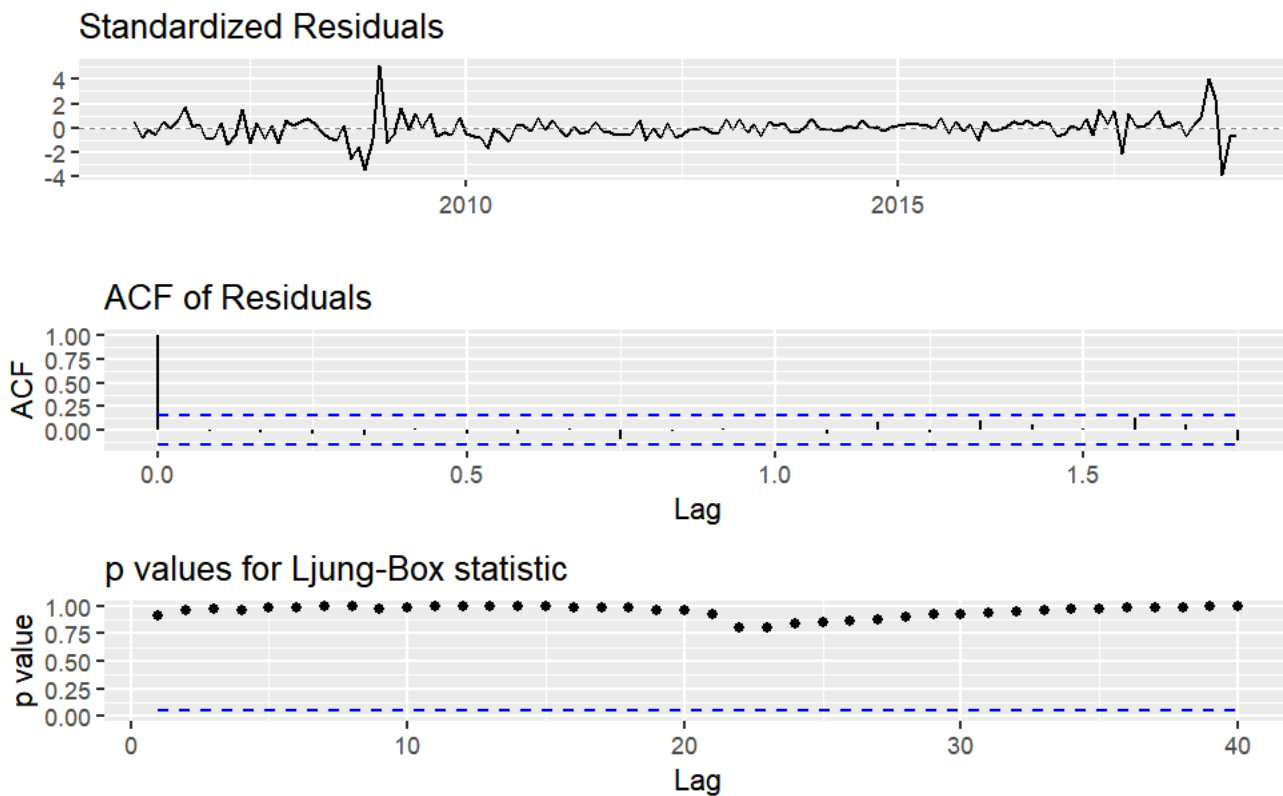
Coefficients:
      ma1      ma2      ma3      ma4      sma1
    -0.0255 -0.5545 -0.2060 -0.1805  0.1649
s.e.   0.0833  0.0857  0.0822  0.0857  0.0945

sigma^2 = 0.8091:  log likelihood = -201.11
AIC=414.21  AICc=414.79  BIC=432.44
```

결과

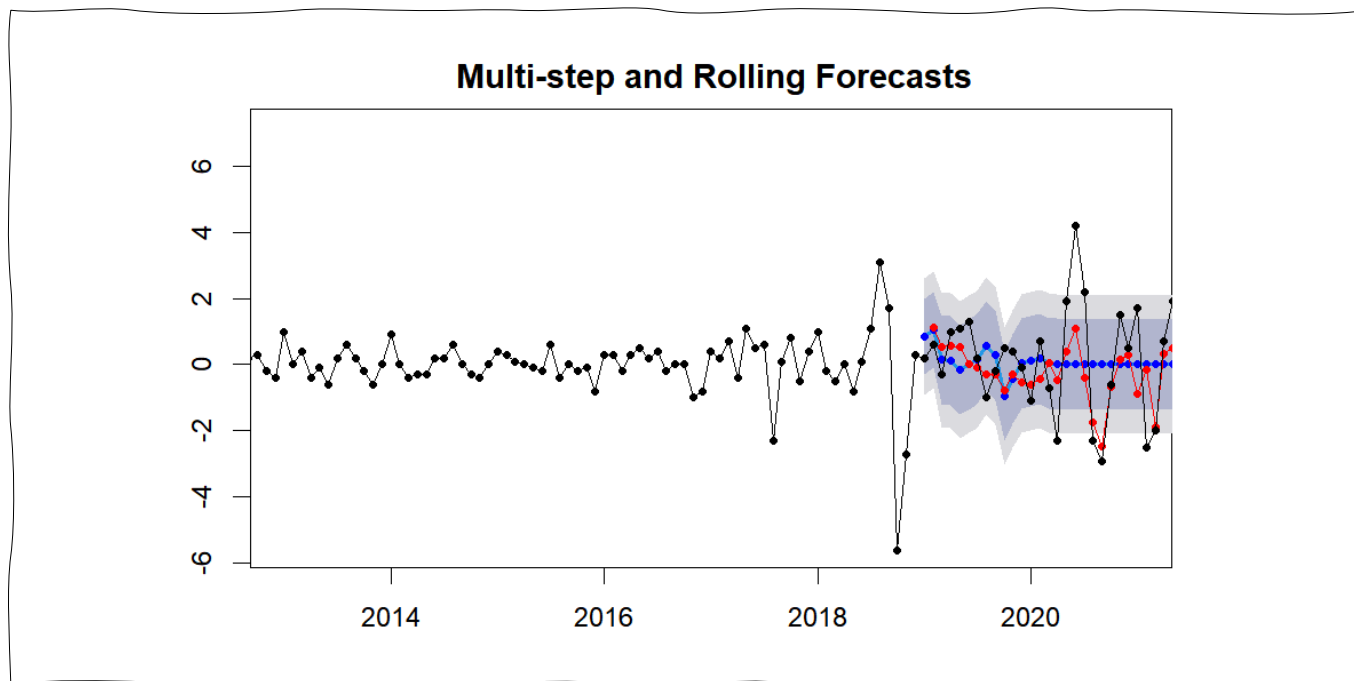
- 제일 적합한 모델:
ARIMA(0,0,4)(0,0,1)[12]
- AIC값: 414.21

(2) auto.arima를 통한 모델링



잔여 값들 간 상관관계 가 없음을 확인하여 Forecasting 진행 가능

모델링: Forecasting



	Multi_Step	Rolling	(p,q) Model
RMSE	1.27316	0.6627419	1.61728

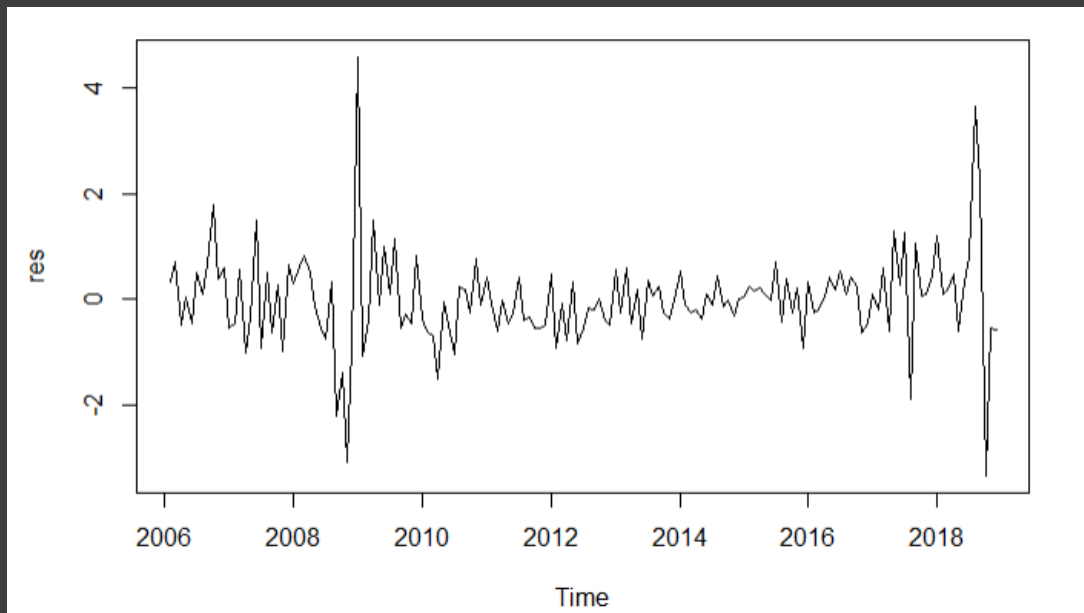


Rolling Forecast > Multi-Step > (p, q) Model 순으로 정확도 높음

2

단변량 분석

이와 더불어 1차 차분한 데이터를 `auto.arima`를 통해 적합 후
 모델링: Forecasting
 변동성 모형을 통해 분산간 이분산성을 모델링하는 방법도 진행



Multi_Step

Rolling

(p,q) Model

1차 차분 후 `arima(0,0,3)(1,0,0)[12]`에 적합한 후 Residual plot



Rolling Forecast을 통해 이분산성이 존재함을 확인 정확도 높음

변동성 모형 (ARCH, GARCH)



ugarchspec() 활용

Variance model : garchOrder = c(**p** , **q**)

Mean model : armaOrder = c(0,3)

fixed.pars : arfima = 1

*Infocriteria()*를 통해 aic value 도출

p \ q	1	2	3
0	5.377	5.3901	5.4029
1	4.054	4.066	4.079
2	4.066	4.079	4.092
3	4.044	4.057	4.07



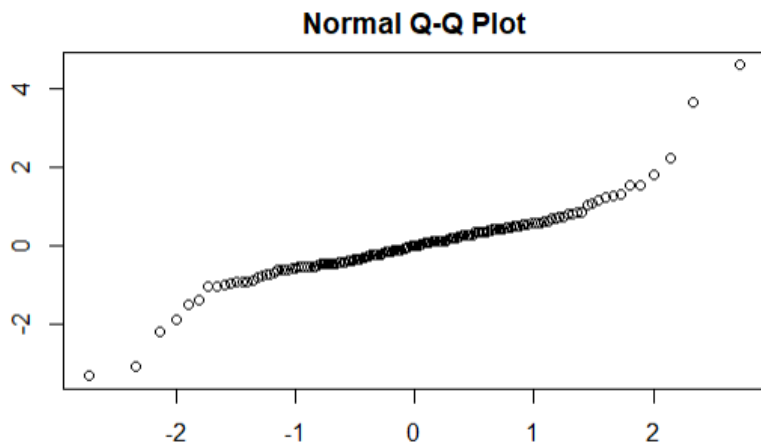
AIC가 가장 작은

garch(1,1)을

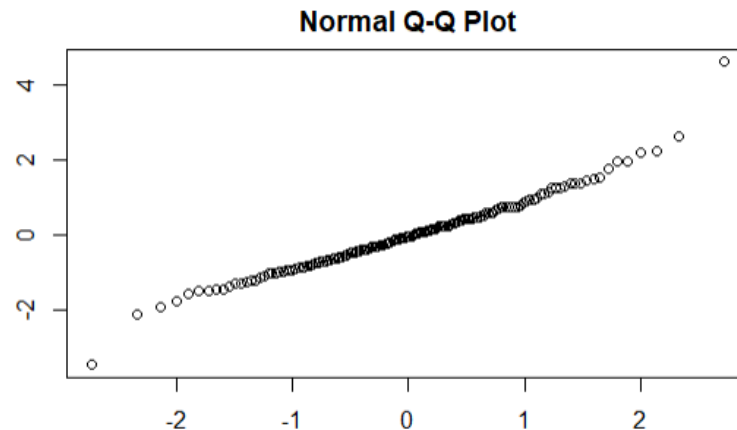
residual에 적합

Fit the residual to GARCH(1,1)

✖ ARIMA 모형 적합 후
Residual



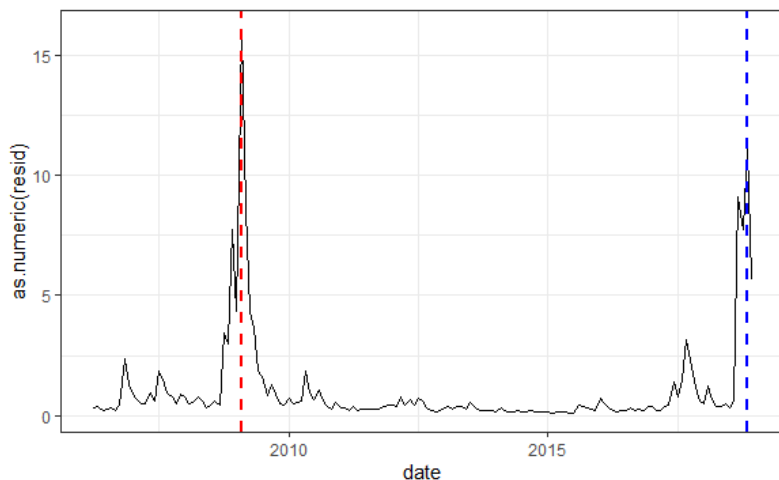
✖ Residual을 GARCH(1,1)에
적합 시킨 후 잔차



여전히 이상치가 존재하지만 변동성 모형에 적합하기 이전보다
비교적 정규분포를 따름을 확인할 수 있음

Fit the residual to GARCH(1,1)

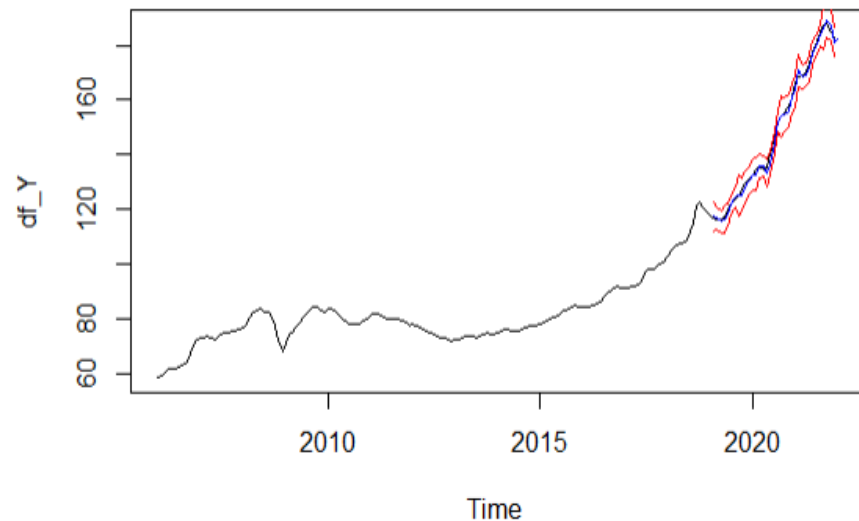
Conditional Variance



Red dotted line : 2009-02
2008년 금융위기로 인한 영향

Blue dotted line : 2018-11
9.13 부동산 종합 정책으로 인한 영향

Forecast with 95% C.I



Fitted 된 garch모델을
활용하여 이후 36개월에 대한
잔차의 variance 추정

3

다변량 분석

변수 선정



여러 논문을 참고하여 아파트 매매 실거래가에 영향을 줄 수 있는 변수 선정

- 기계학습 알고리즘을 활용한 지역 별 아파트 실거래가격지수 예측모델 비교 : LIME 해석력 (조보근, 2020)
- 서울아파트시장과 거시경제변수 요인들 간 동학적 상관관계 분석 (김경민, 2018)

소비자
물가지수

통화량
(M2)

주가지수
(KOSPI)

지가지수
(서울시)

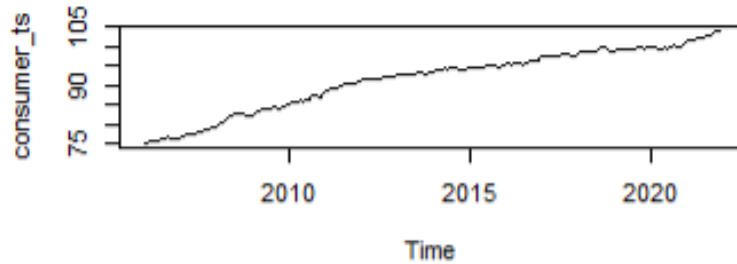
가계대출
금리

3

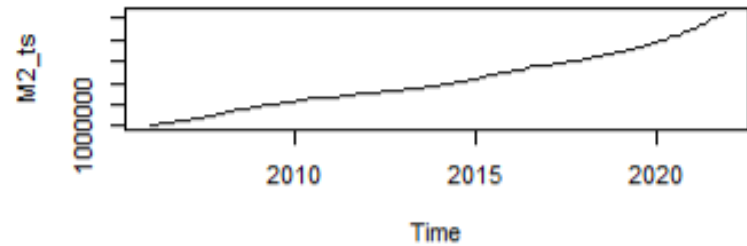
다변량 분석

변수 시각화

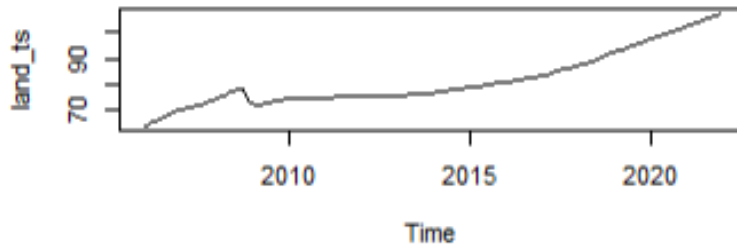
소비자 물가지수



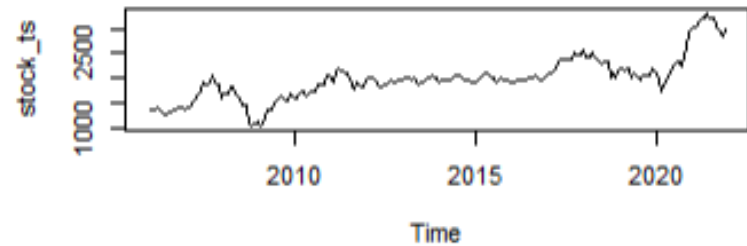
통화량



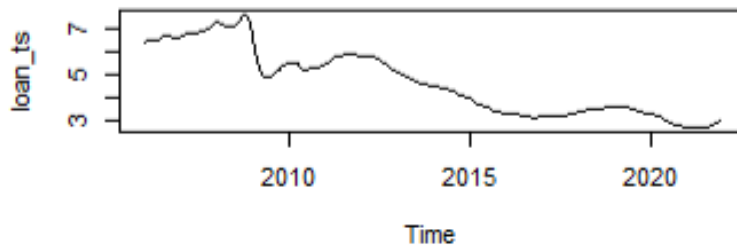
지가지수



주가지수



가계대출 금리

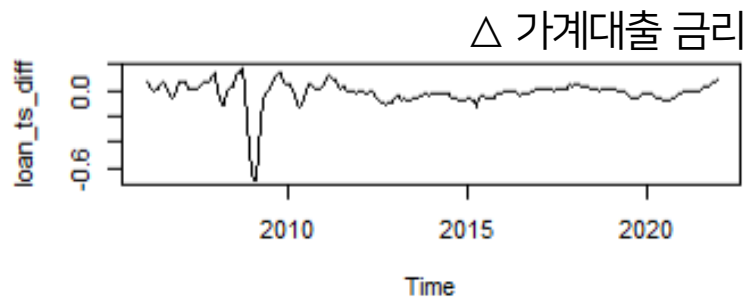
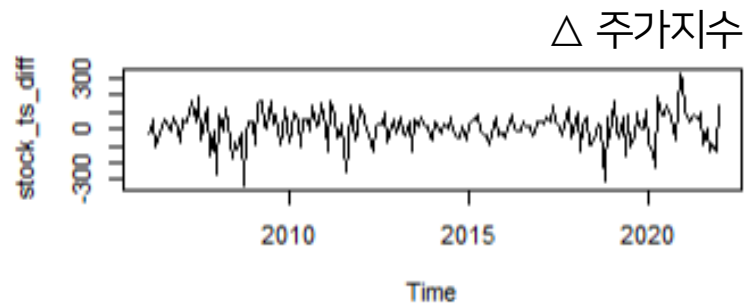
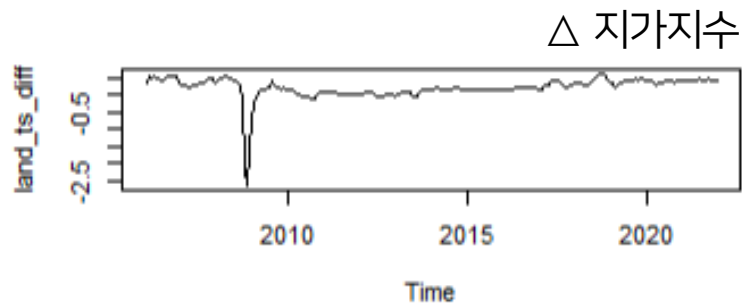
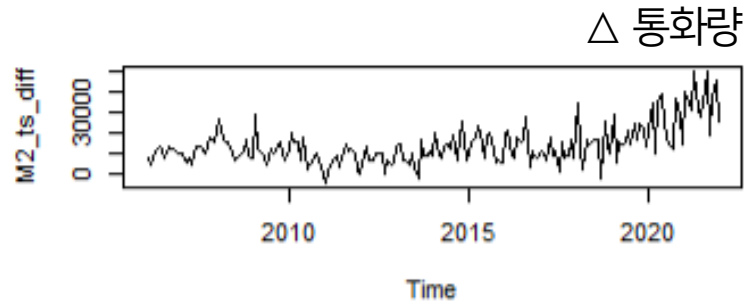
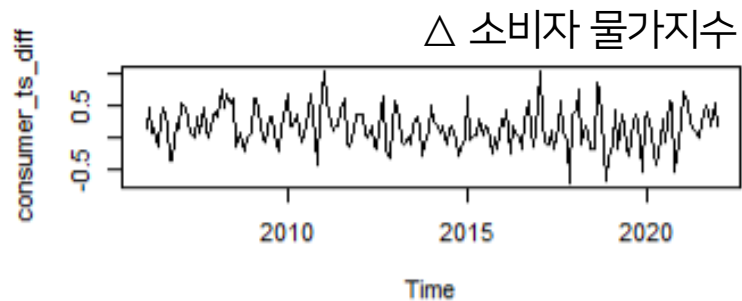


모든 변수에 추세가 존재



1차 차분 진행

변수 시각화 - 1차 차분



ADF test를 통해 정상성 확인
(type = "none")

그레인저 인과검정



VARSelect를 통해 Y(아파트 실거래 지수)와 변수들 간 order 값 결정



Granger 인과검정을 통해 아파트 실거래 지수 예측에 도움이 되는 변수들 파악



Result

	소비자물가지수	통화량	지가지수	KOSPI	가계대출금리
AIC(n)	4	3	5	2	5
Granger	0.1266	0.06	0.0003	0.9331	9.04e-05

최적의 모델 탐색

General Additive Model

각 변수의 비선형함수를 허용하여 표준선형모델을 확장한 일반적 모델

Model 1 : GAM(아파트 매매 실거래 가격 지수 ~ 지가지수 + 가계대출 금리) (full model)

Model 2 : GAM(아파트 매매 실거래 가격 지수 ~지가 지수) (reduced model)

Model 3 : GAM(아파트 매매 실거래 가격 지수 ~가계대출 금리) (reduced model)

Wald 검정

H_0 : 두 모델 간 최대 가능도 추정량의 차이가 같음

Model 1 vs Model 2

P-value : $8.424e-09$



Model 1 selected

Model 1 vs Model 3

P-value : $< 2.2e-16$



Model 1 selected

최적의 모델 탐색

General Additive Model

각 변수의 비선형함수를 허용하여 표준선형모델을 확장한 일반적 모델

Model 1 : GAM(아파트 매매 실거래 가격 지수 ~ 지가지수 + 가계대출 금리) (full model)

Model 2 : GAM(아파트 매매 실거래 가격 지수 ~지가 지수) (reduced model)

Model 3 : GAM(아파트 매매 실거래 가격 지수 ~가계대출 금리) (reduced model)

Wald 검정

H0 : 두 모델 간 최대 가능도 추정량의 차이가 같음

Model 1 vs Model 2

P-value : 8.424e-09



Model 1 selected

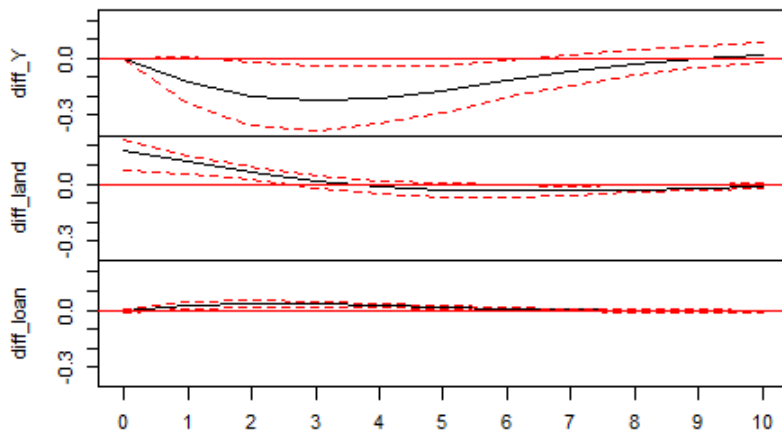
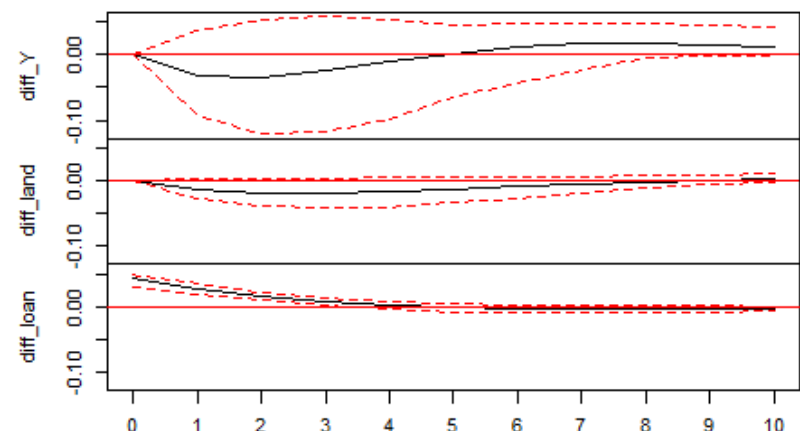
Model 1 vs Model 3

P-value : < 2.2e-16



Model 1 selected

Impuse Response plot

Response from Δ 지가지수Response from Δ 가계대출금리

지가지수와 가계대출금리의 shock은 아파트 매매 실거래가 지수에
상당하고 지속적인 영향을 줄 수 있음

4

결론

정리

단변량

2차 차분 + ARIMA
ARIMA(0,0,4)(0,0,1)[12]



RMSE

0.6627

다변량

지가지수 + 가계대출금리
VAR



2.485

의의

- 다양한 시계열적 모델 기법을 활용하여 아파트 매매 가격거래 지수가 지닌 특성을 파악
- 변동성 모형과 VAR을 통해 가격거래 지수의 큰 영향을 준 일련의 사건과 높은 설명력을 지닌 변수를 파악해 볼 수 있었음



THANK YOU

