

# 아파트 매매지수 회귀모형

# 시작

## ✓ 소개

- 2학년 편입 후 3년간 학습
- 통계나 IT와 전혀 관련없는 직업
- 자료분석 등에 매우 관심이 많음

≡ 한국경제

부동산

### [12·16대책] 또나온 부동산대책...규제 효과없자 초고강도 처방

입력 2019.12.16 13:23 | 수정 2019.12.16 13:30

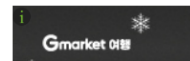
"집값 예년보다 안정" 표정관리하던 정부, 서울 집값 상승 랠리에 기습 발표  
대통령 "집값 자신있다" 발언 이후 극비 진행...세제·대출·청약·공급 등 총망라

정부가 지난해 9·13대책 이후 1년3개월 만이자 민간택지 분양가 상한제 대상지  
선정 이후 한달 만에 또다시 초고강도 부동산 대책을 내놔다.

이번 정부 들어 2017년 6·19대책, 8·2부동산 대책, 지난해 9·13대책에 이어 정부 합동  
종합대책 형태로 발표된 네번째 대책이면서 지난해 12월 주거복지로드맵이나  
지난달 발표된 분양가 상한제 대상지역 지정 등 개별 또는 후속 조치까지 합쳐  
18번째 대책이다.

실시간 2

- 1 [집코노미TV] 돈 안들0
  - 2 [12·16대책 한 달] 수도
  - 3 [박동휘의 베트남은 지
  - 4 현대차 북미본부장 "미
  - 5 '올림픽 본선 자동진출
- 광고 난 주식투자로 매일  
광고 '로또1등' 당첨확률·



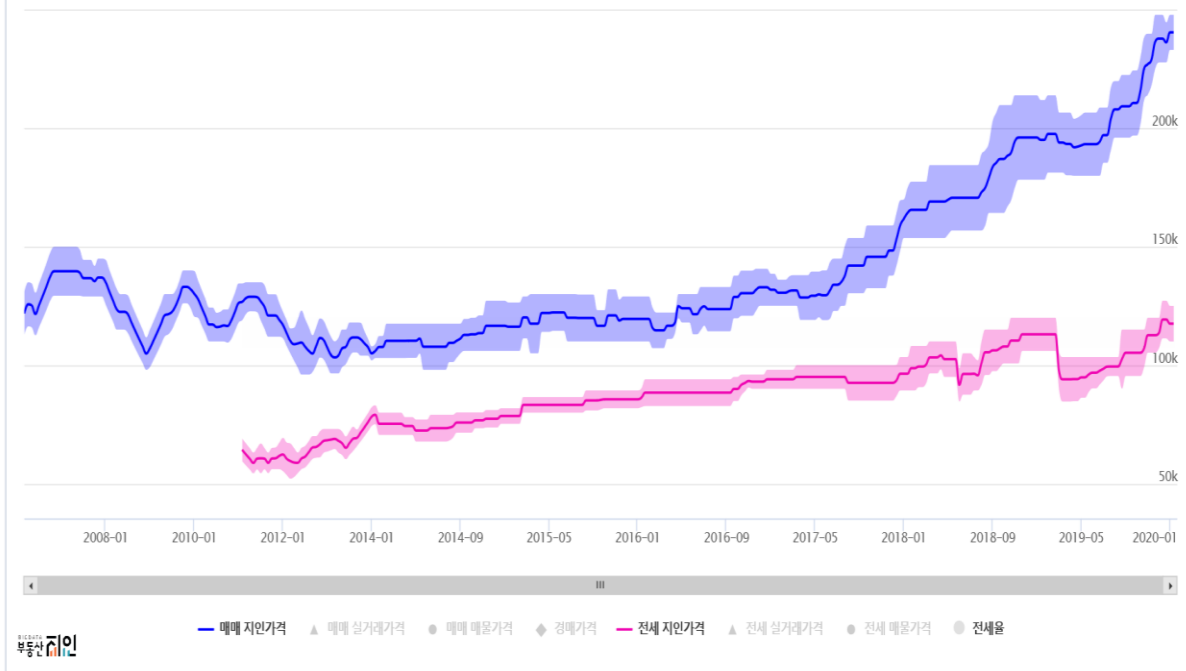
## ✓ 주제선정사유

- 부동산에 대한 아주 약간의 **Insight** (주식과 달리 부동산은 모든 국민이 시장 참여자)
- 집은 나와 우리의 인생에 평생 큰 영향을 미치고 있음 (기숙사 → 하숙 → 월세 → 전세 → 결혼 → 회사 → 자녀 → 출가 → 역모기지)
- 최근 2년간 아파트가격의 급격한 상승과 그로 인한 이번 정부의 부동산대책 18번째 발표 ('12.16 부동산대책)

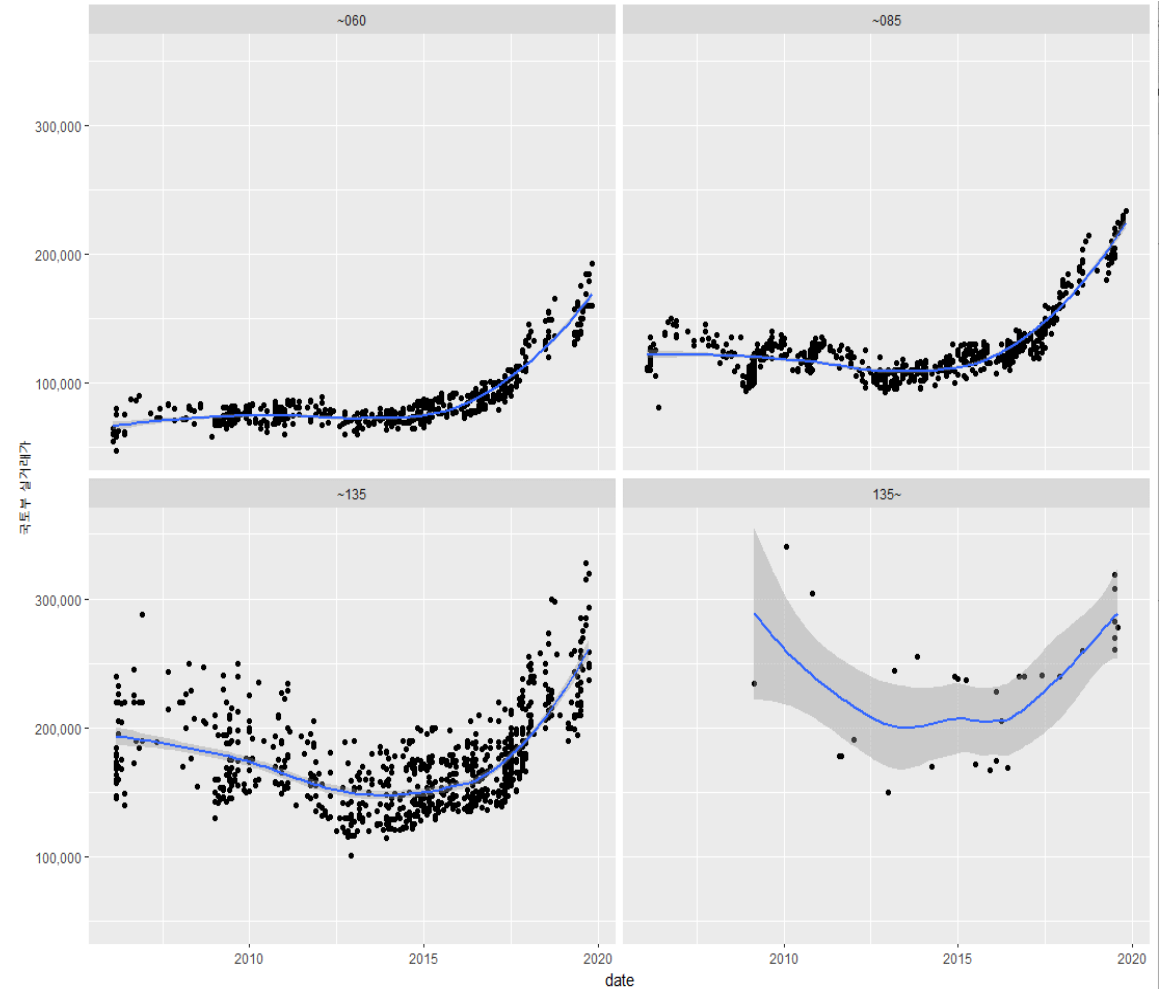
## ✓ 저의 통계분석 목표 : 걱정만 회귀모형 구축이 1차 목표이고 예측은 2차 목표임

# 시작

- ✓ 주택 가격의 특성 : 모든 시장참여자가 여러 영향요인(설명변수)의 영향을 받아서 합리적인 의사결정에 의해 즉시 가격이 변동하는 상품이 아님 → 주택 가격은 상승, 또는 하락의 에너지(돈, 유동성 등)가 쌓이다가 폭발하기도 하고 오랜 침체기를 보내다가 갑자기 상승하기도 함, 특히 단기간에는 정책의 영향을 많이 받음

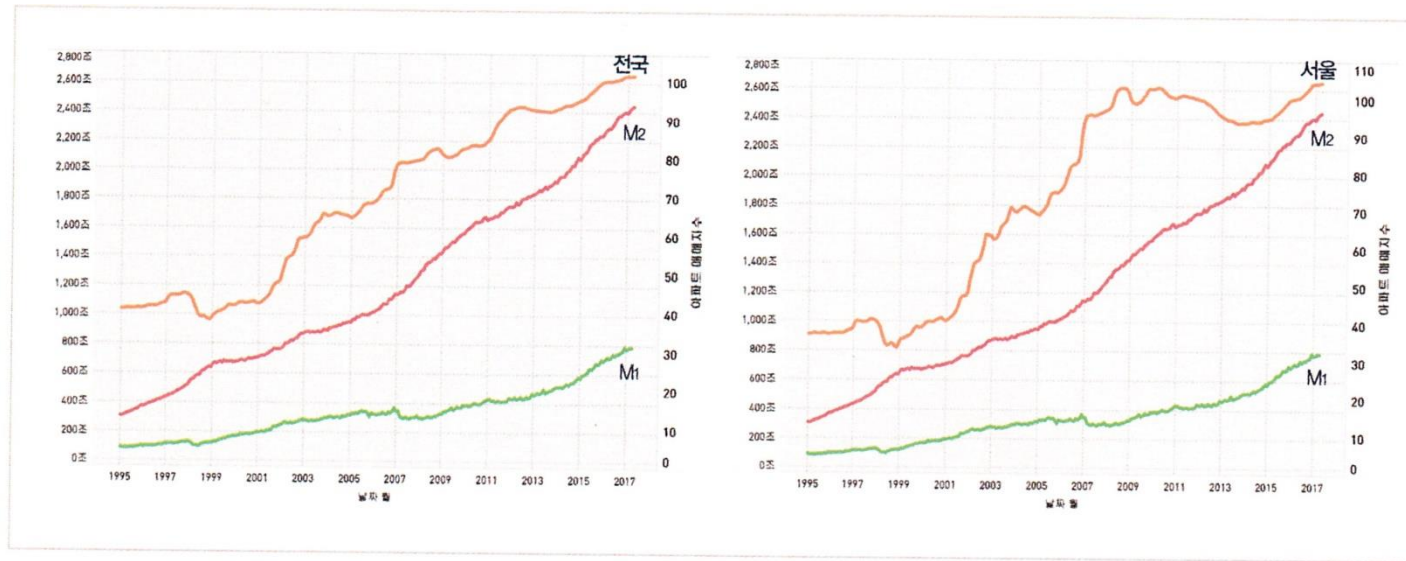


- ✓ 샘플 아파트의 평형대별 실거래가
  - 자료 : 국토부 실거래가 사이트
  - 기간 : '05.01 ~ '19.11



# 기존 부동산 전문가들의 분석방법

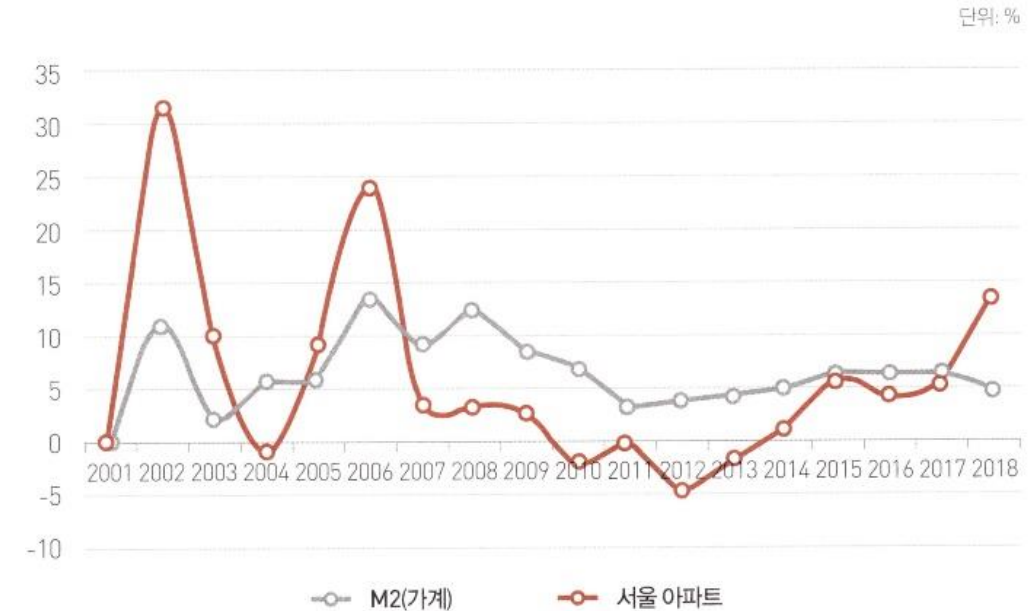
- ✓ 통화량(M2) : 시중에 돈이 많이 풀리면 아파트의 명목가격은 상승한다
  - 통화량과 아파트매매지수 비교
  - 통화량 증감율과 아파트 시세 증감율 비교



1995~2017년 전국 및 서울 아파트 매매지수

< 빅데이터 부동산투자, 김기원 >

가계부문 통화량과 서울 아파트 시세 증감률



< 서울아파트 상승의 끝은 어디인가, 강승우 >

# 기존 부동산 전문가들의 분석방법

- ✓ 글로벌 주택가격 : 글로벌 동조화 현상으로 우리나라의 주택가격은 글로벌 주택가격을 따라간다
  - 글로벌 주택가격과 수도권 주택가격의 비교
  - 미국 주택가격 변동률과 강남구 주택가격 변동률

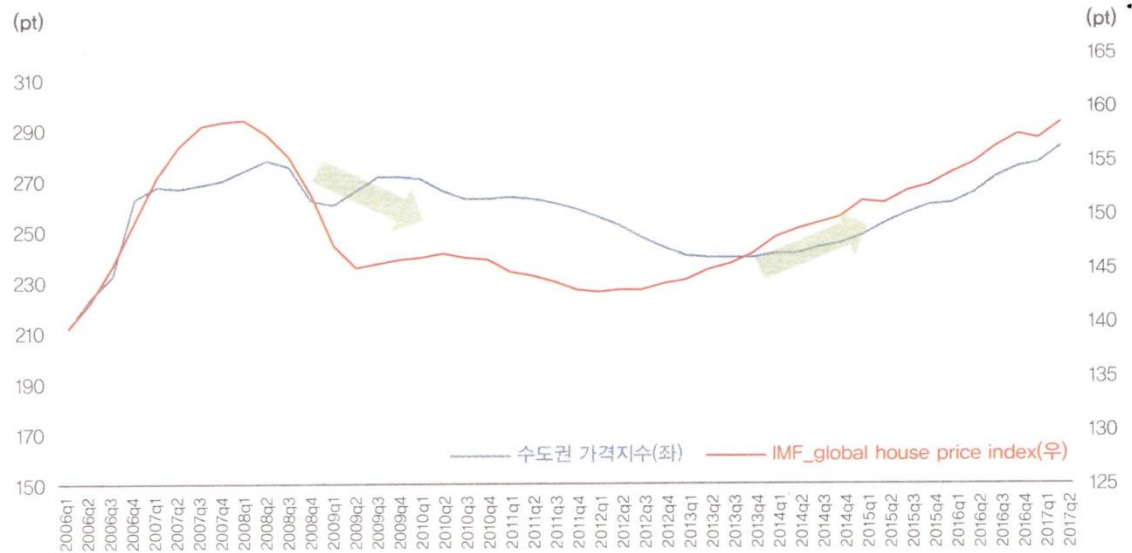


그림 2-1 글로벌 주택 가격과 수도권 주택 가격 추이  
(2006년 1분기~2017년 2분기/부동산114 REPS, IMF)

< 대한민국 부동산의 미래, 조영광 >

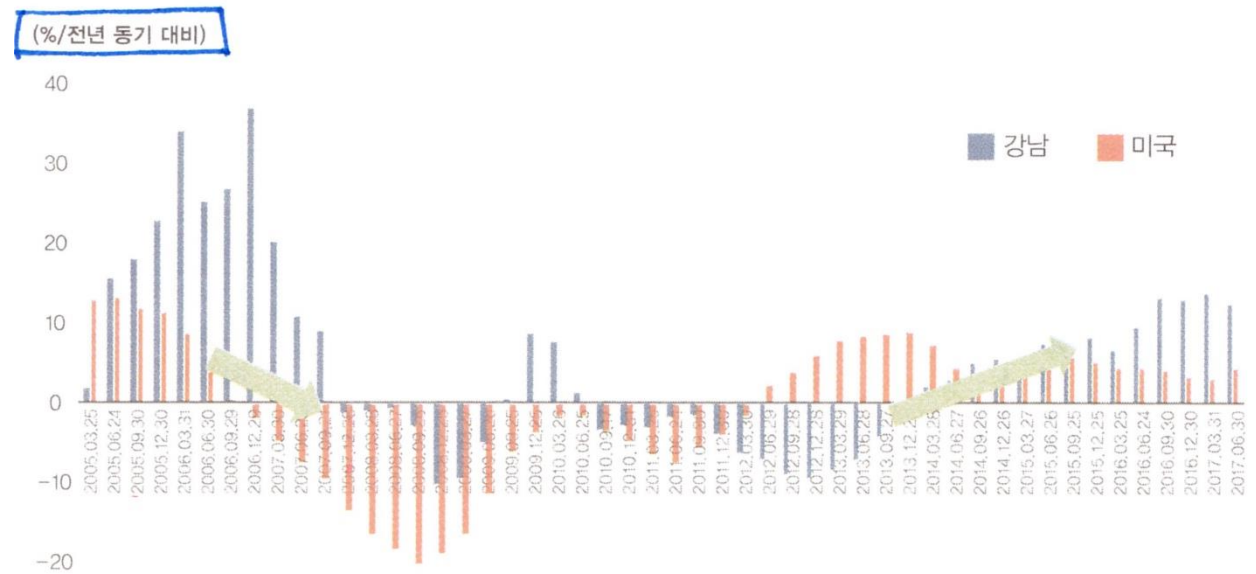
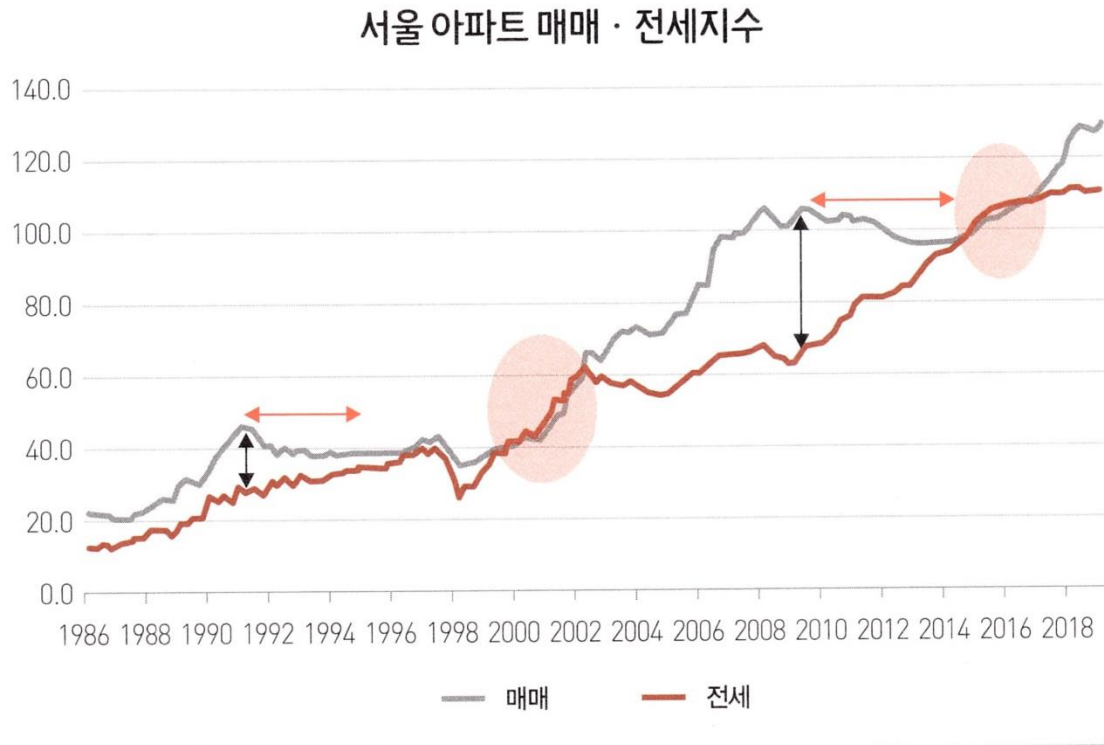


그림 14 강남구와 미국의 주택 가격 변동률 추이  
(2005년 1분기~2017년 2분기/부동산114 REPS, FRED)

< 대한민국 부동산의 미래, 조영광 >

# 기존 부동산 전문가들의 분석방법

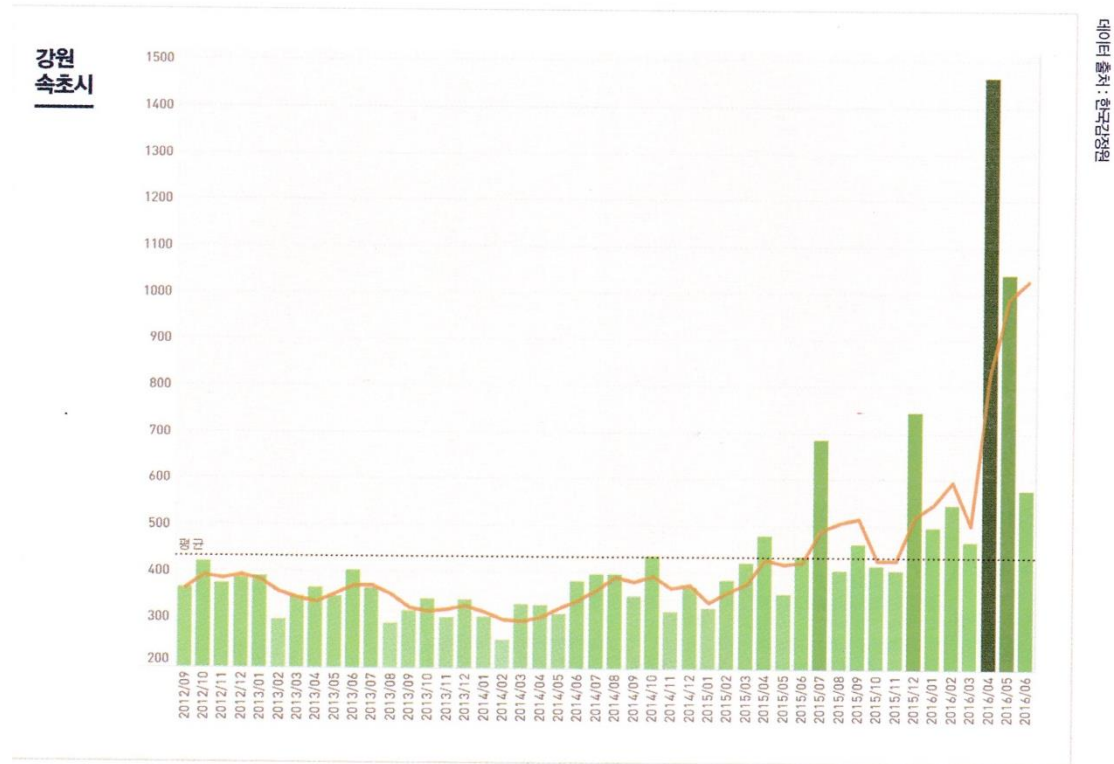
- ✓ 매매 및 전세지표의 관찰 : 매매지수와 전세지수의 차이가 커지면 하락의 전조  
매매거래량이 폭발하면 상승의 전조
- 매매지수와 전세지수의 이격률
- 매매거래량의 급격한 변동



※ 2015년 2분기 매매가·전세가=100

자료: KB부동산

< 서울아파트 상승의 끝은 어디인가, 강승우 >

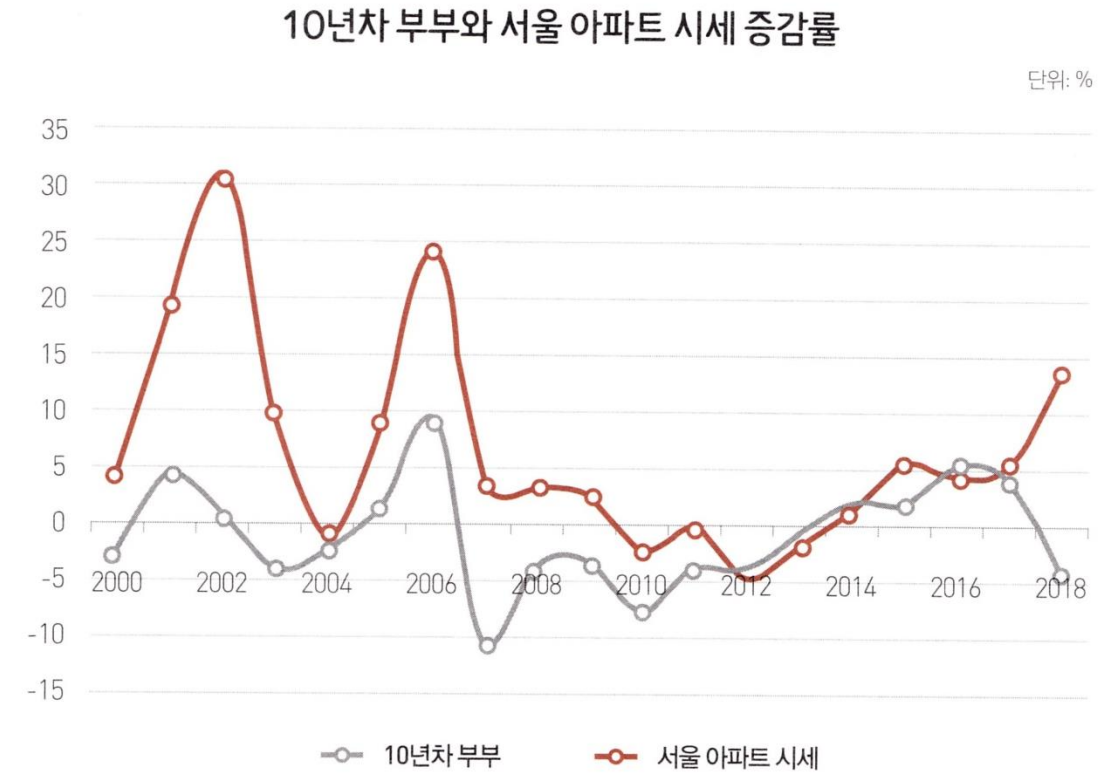


빅데이터 시각화 1-1 | 2012~2016년 속초시 거래량

< 빅데이터 부동산투자, 김기원 >

# 기존 부동산 전문가들의 분석방법

- ✓ 10년차 부부 증감률과 서울아파트 시세 증감률은 비슷한 추세를 따른다
  - 인구론적인 관점
  - 결혼 10년차가 되면 자산도 어느정도 모이고 자녀가 초등학교 입학연령이 되어 아파트 구매 혹은 상급지로 갈아타려는 욕망이 최대



자료: 통계청, KB부동산

< 서울아파트 상승의 끝은 어디인가, 강승우 >



## ✓ 시작

- 4학년 1학기(지난학기) 예측방법론 학습 시 매우 어려움을 느낌, but 모든 경제 데이터는 시계열임을 깨닫고 학습함

## ✓ 준비과정

- 기말고사 후 3주간 약 30개의 시계열데이터(통계청, 국토부, KB, 한국감정원, 한국은행, 부동산정보 사이트)를 수집·데이터전처리·기본그래프 검토 등을 하며,
- 동시에 각종 부동산서적 학습, 예측방법론 교재 재학습, 외부 시계열서적, 인터넷자료 등을 학습하면서 본인의 지식의 한계를 깨달음

## ✓ 시행착오

- 초기에 시계열분석에 대한 이해 부족으로 모든 자료를 계절조정을 한 다음 전월 대비 증감률(안정시계열)로 바꾸어서 분석을 하면 된다고 생각함
- 교재 7장 회귀모형을 이용한 예측에 나온 대로 각종 재료를 분석도구(lm 함수)에 집어넣기만 하면 분석결과가 나오는 것으로 착각함 (회귀모형은 마술상자)

## ✓ 최종 분석방법

- 자료review(그래프, ccf) → 안정시계열로(acf, pacf, 단위근검정) → 회귀모형 → alt1. 자기시차변수 추가/ alt2. 코크란 오컷 방법 → VAR모형 → 공적분검정



# 분석을 위한 준비과정

## ✓ 자료의 수집

- 비슷한 자료 중 어떤 자료를 쓸 것인가?

서울시 자료 / 월별 데이터 / 10년 이상

축적된 시계열 자료가 우선 대상

## ✓ 데이터 정리

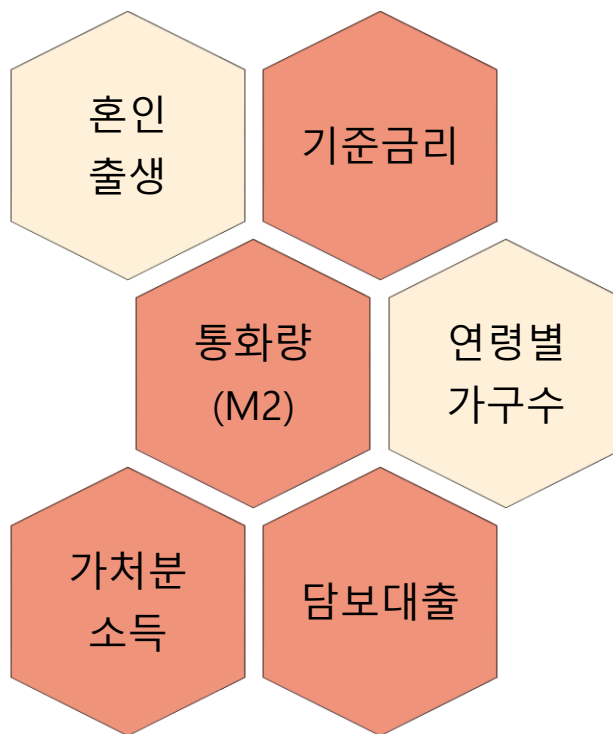
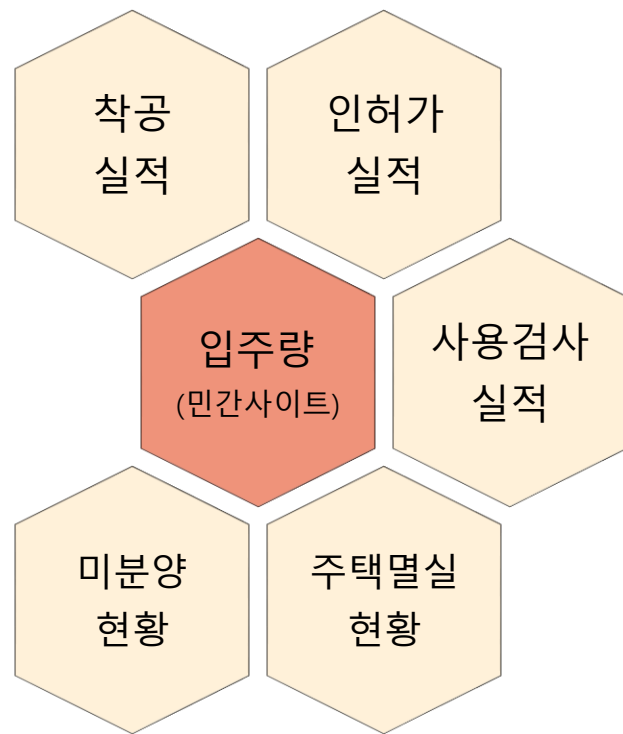
- 이름의 결정, 변수명의 결정 등

(데이터이름)		(변수이름)	
volume	물량	date	날짜
cost	금액	area	면적
index	지수	city,gu,dong	시군구
housing	주택	brand	아파트명
apt	아파트	floor	층
housingsize	주택규모	quantity	물량
housingtype	주택유형	price	단가
permit	인허가	trade	매매
commence	착공	give	증여
completion	준공	ownership	분양권
M2	통화량		
household	가구		
age	연령		
income	소득		
baserate	기준금리		

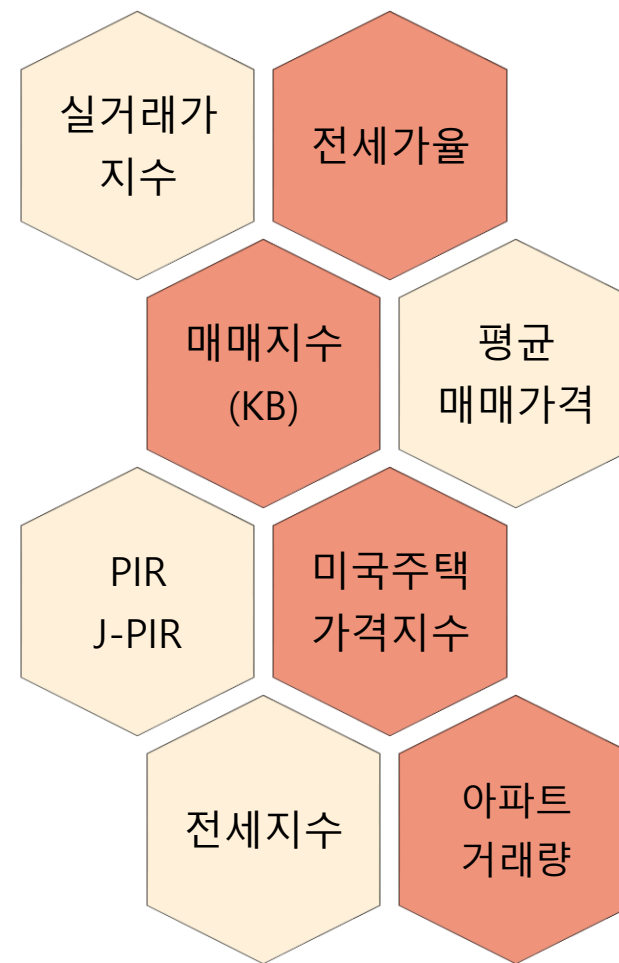
## 자료의 수집 및 데이터정리

구분	통계명	상세도	날짜	기관
공급	주택인허가실적	시 단위	2007/01~	통계청
공급	주택착공실적	시 단위	2011/01~	통계청
공급	주택사용검사실적	시 단위	2010/07~	통계청
공급	주택멸실현황	시	2010~2017	국토교통부
공급	아파트 입주량	시군구동	~'22	부동산지인
현황지표	매매지수(APT)	시군구	1986/01~	KB
현황지표	전세지수(APT)	시군구	1986/01~	KB
현황지표	매매전세비(전세가율)	시군구	1998/12~	KB
현황지표	실거래가 공개시스템	시군구동	2006/01~	국토교통부
현황지표	평균매매가격(아파트)	시군구	2012/01~	한국감정원
현황지표	실거래가격지수(아파트)_지역	서울6대권역	2006/01~	한국감정원
현황지표	실거래가격지수(아파트)_규모	전국,서울	2006/01~	한국감정원
현황지표	실거래가격지수(아파트)_시군구별	시군구	2006/1Q~, 분기	한국감정원
수요	통화량(M2)_명목_평잔	시 단위	1986/01~	통계청
수요	가계부문통화량(M2)_명목_평잔	전국	2001/12~	한국은행
수요	PIR(매매)	전국	2008/12~	KB
수요	J-PIR(전세)	전국	2008/12~	KB
수요	연령별 가구수	시 단위	00,'05,'10,'15~'18,'19~'45	통계청
수요	주택구입부담지수	시 단위	2004/1Q~, 분기	한국주택금융공사
수요	연령별 가처분소득	전국	1990/1Q~, 분기	통계청
수요	한국은행 기준금리	전국	1999.05~	한국은행
수요	주택담보대출	시 단위	2007/12~	한국은행
수요	인구동향(출생,혼인)	시 단위	1981/01~	통계청
수요	혼인건수	시 단위	1997/01~2018/12	통계청
수요	이혼건수	시 단위	1997/01~2018/12	통계청
수요	아파트거래현황	시군구	2006/01~	한국감정원
수요	거래원인별 아파트거래	시군구	2013/01~	한국감정원
수요	매입자연령대별 아파트매매거래	시군구	2019/01~	한국감정원
수요	미분양주택현황보고	시군구	2000/12~	국토교통부
수요	연령별매매가격지수_아파트	서울6대권역	2012/01~	한국감정원
수요	가계대출	전국	2007/4Q	한국은행
수요	가계대출(예금취급기관)	시 단위	2007/12~	한국은행
수요	은행대출 연체율	전국	2005/03~	한국은행
선행지표	S&P/Case-Shiller U.S. National Home Price Index		1987/01~	FRED
공급	서울시 재건축 현황	시	2004~2018	국토교통부

# 분석을 위한 준비과정



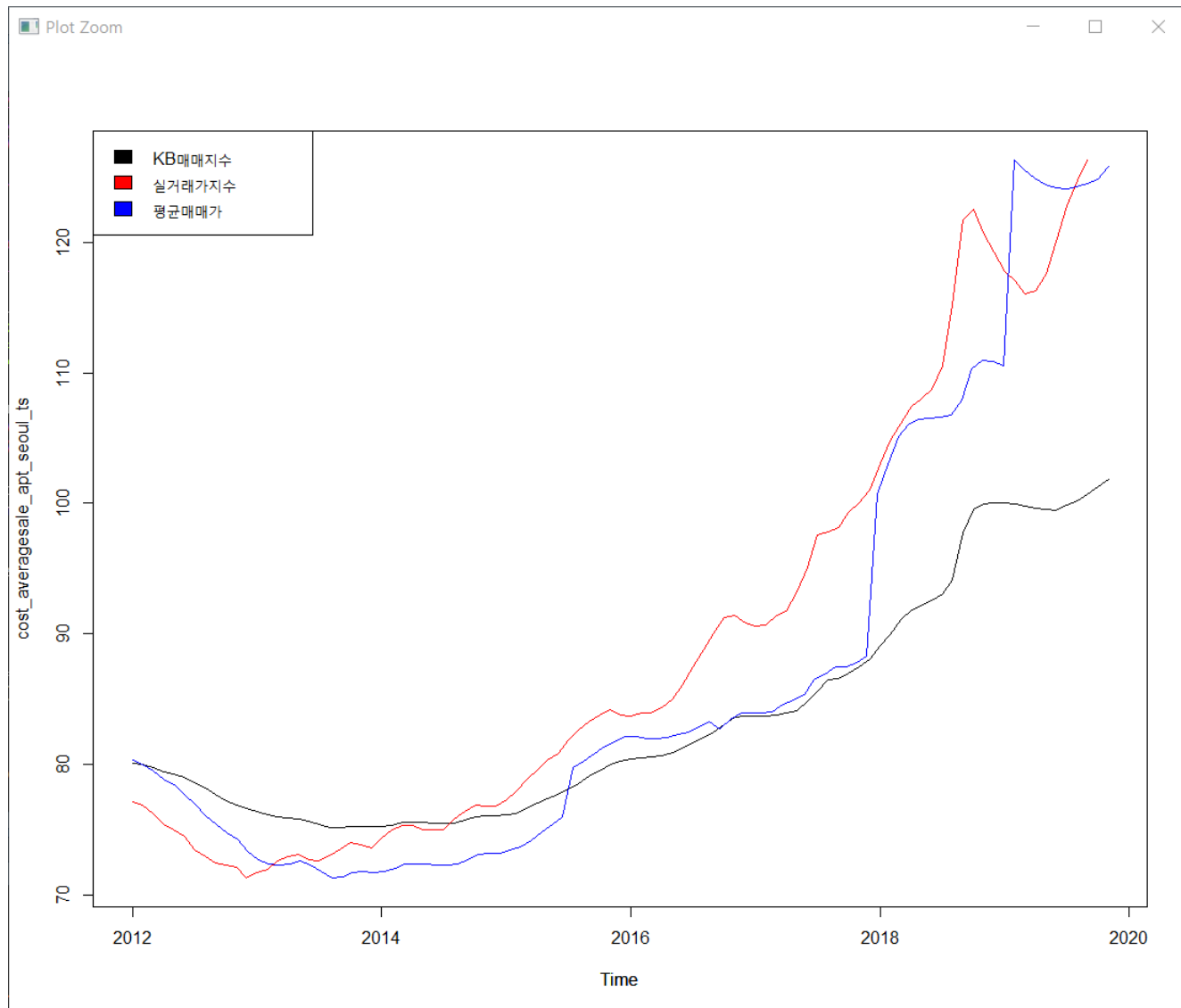
## 최종 변수의 결정



\* 전체적으로 약 30개의 시계열데이터를 수집,전처리하고 사용여부를 검토함

# 자료 살펴보기

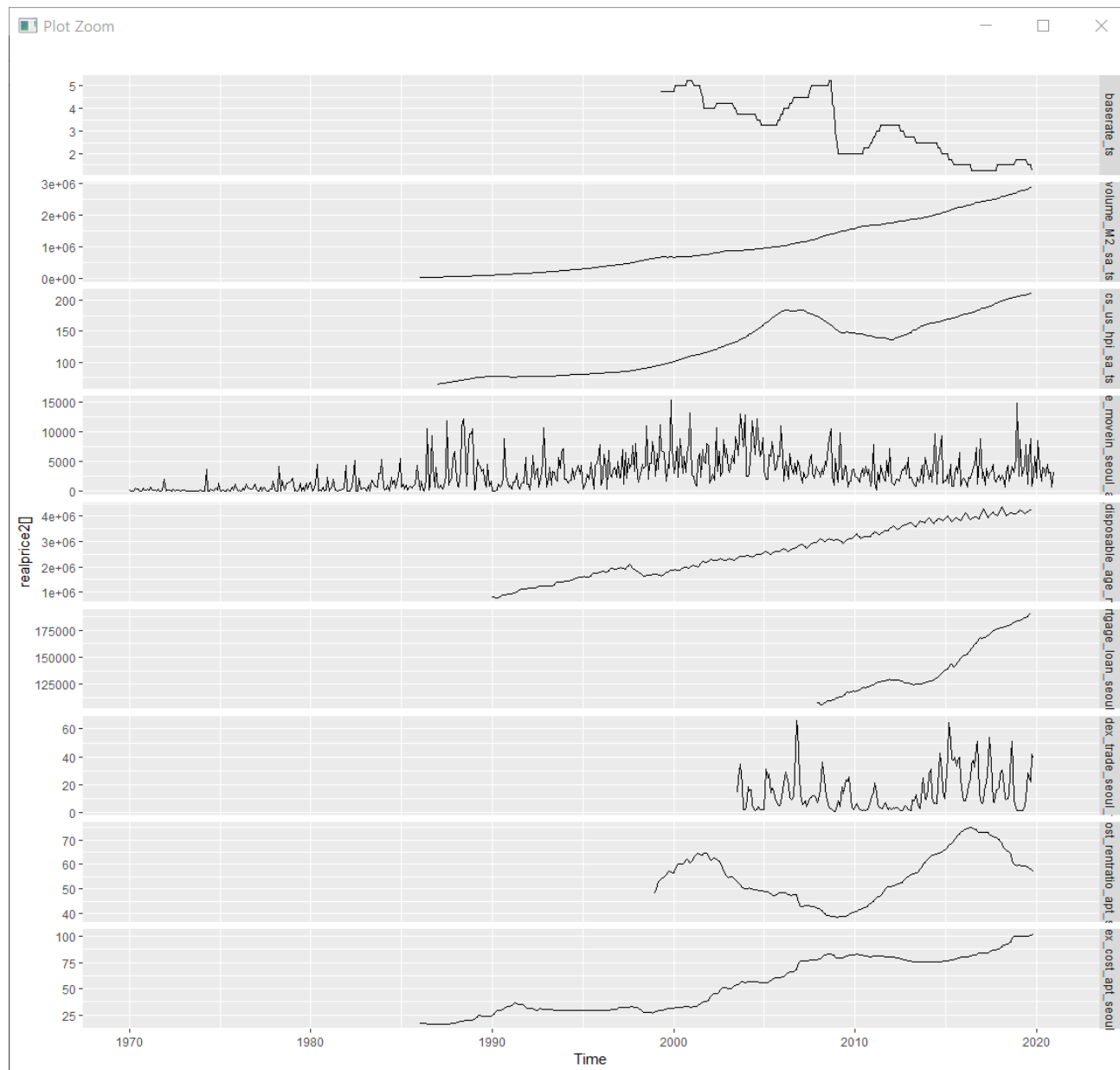
종속변수/ 아파트 매매지수 혹은 매매가격



- ✓ KB매매지수보다 실거래가지수가 지수 산정의 특성 때문에 현실을 더 잘 반영하고, 평균매매가는 그 변동이 급격하여 분석에 사용하기는 부적합하다고 판단함
- ✓ 그러나 실거래가지수는 그 기간이 짧아서 ('12/01~) 분석하기에 부적합하고 KB매매지수는 그 자료가 방대하고('86/01~) 상승추세를 보기에 큰 문제가 없다고 판단하여 KB매매지수를 종속변수로 사용하기로 함

# 자료 살펴보기

모든 변수의 개략 Review



한국은행 기준금리

계절조정 통화량(M2)

미국주택가격지수

S&PCaseShillerUSHomePriceIndex

입주량(민간사이트)서울

가처분소득(가구주,~59)

담보대출(가계대출)서울

매매거래량 지수서울

전세가율(전세/매매)서울

아파트매매지수서울

- ✓ 서울시 데이터
- ✓ 주택이 아닌 아파트 데이터
- ✓ 입주량 및 매매거래량 지수는 계절변동 등에 따라 월별 변동이 매우 심함
- ✓ 통화량 및 미국주택가격 지수는 계절조정데이터
- ✓ KB아파트매매지수는 '18.03이후 계절조정 지수 미제공(사유 파악 불가)

```
library(forecast)
autoplot(realprice2[], facets=T)
GGally::ggpairs(as.data.frame(realprice2[]))
```

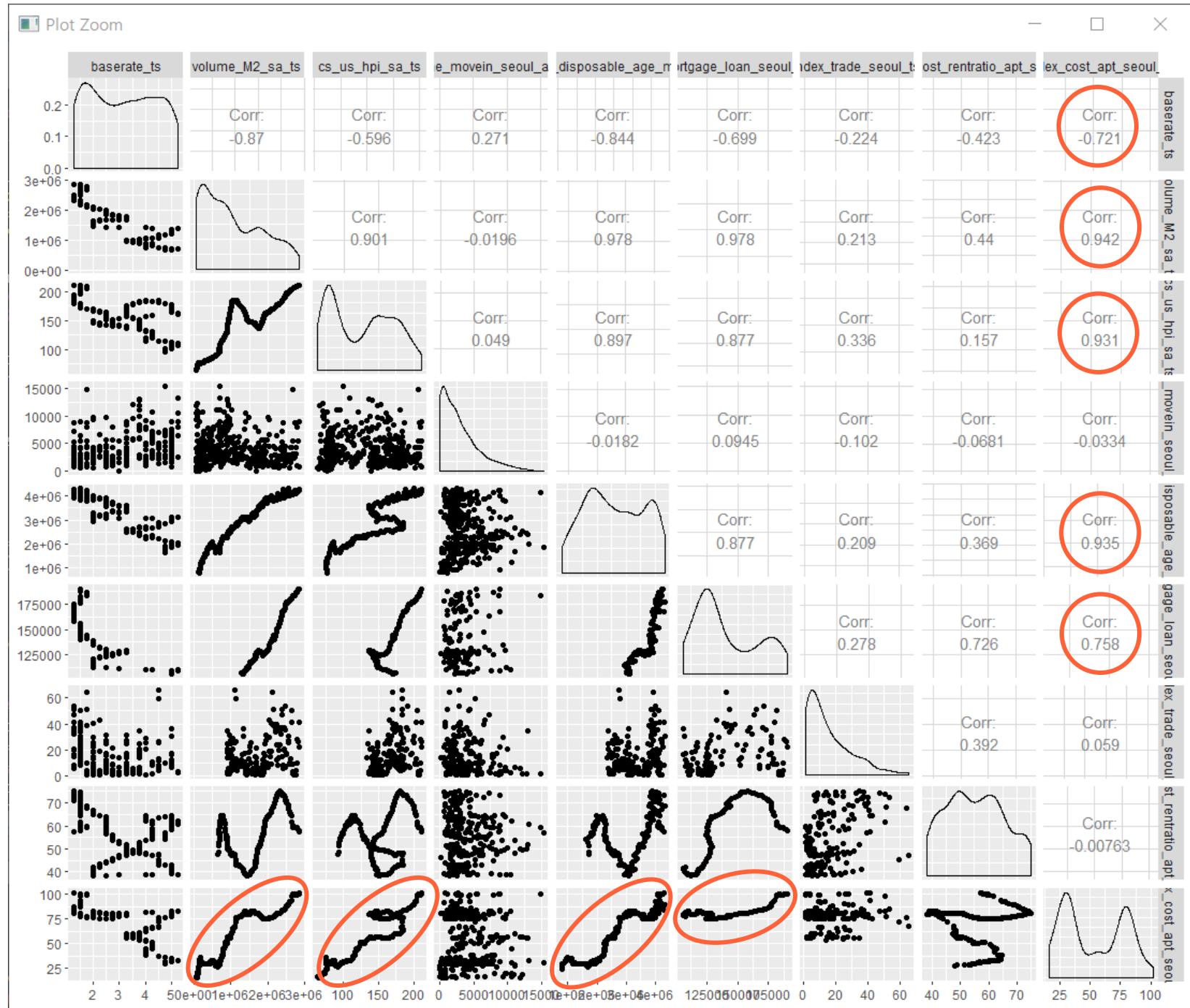
## 자료 살펴보기

## ✓ 산점도의 선형관계

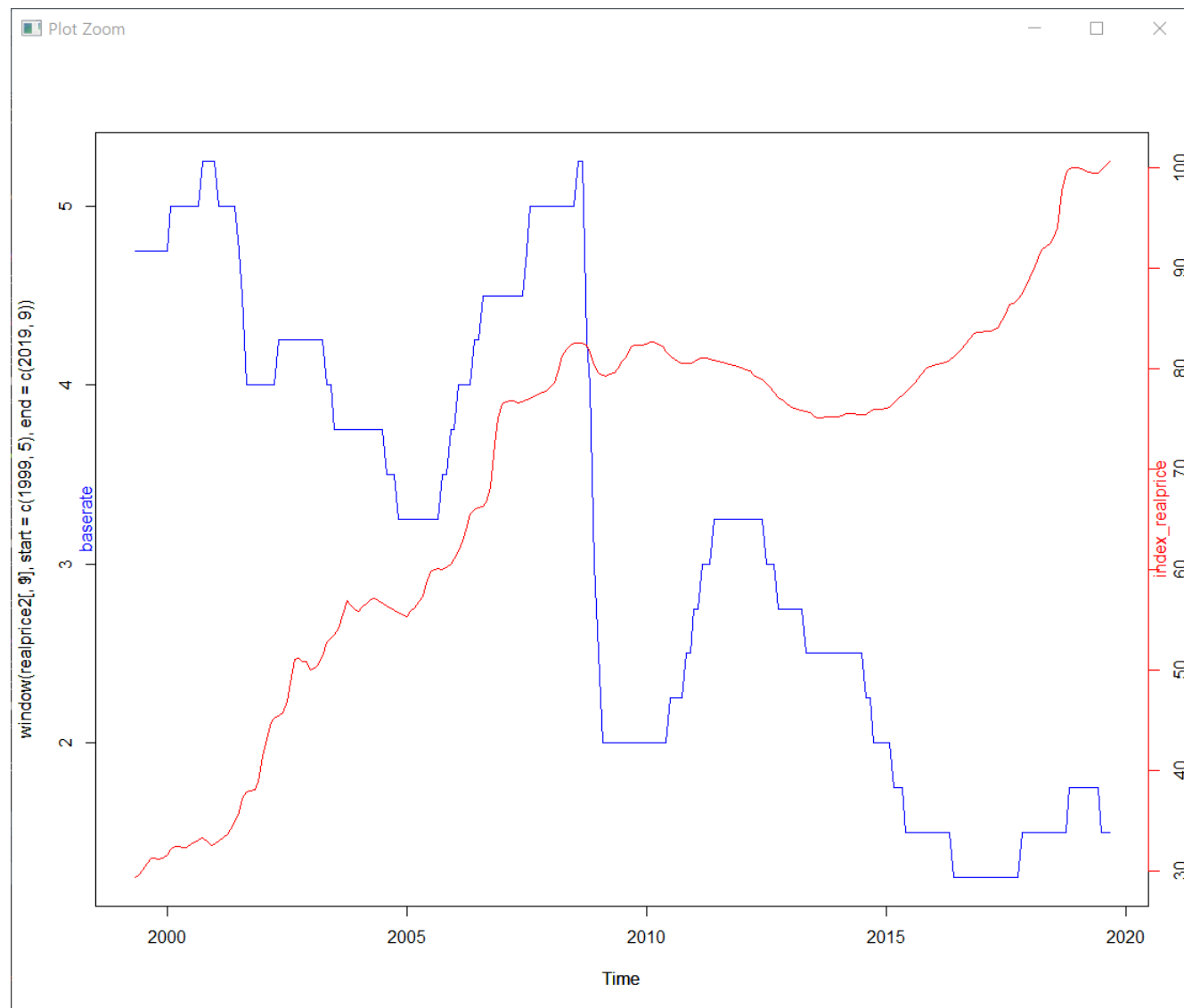
- **매매지수**와 통화량, 미국 주택지수, 가처분소득, 담보대출에 선형관계가 보임

✓ 상관계수 높은관계

- 기준금리: 통화량, 가처분소득, 매매지수
- 통화량: 미국주택지수, 가처분소득, 담보대출, 매매지수
- 미국주택지수: 가처분소득, 담보대출, 매매지수
- 가처분소득: 담보대출, 매매지수
- 담보대출: 전세가율, 매매지수



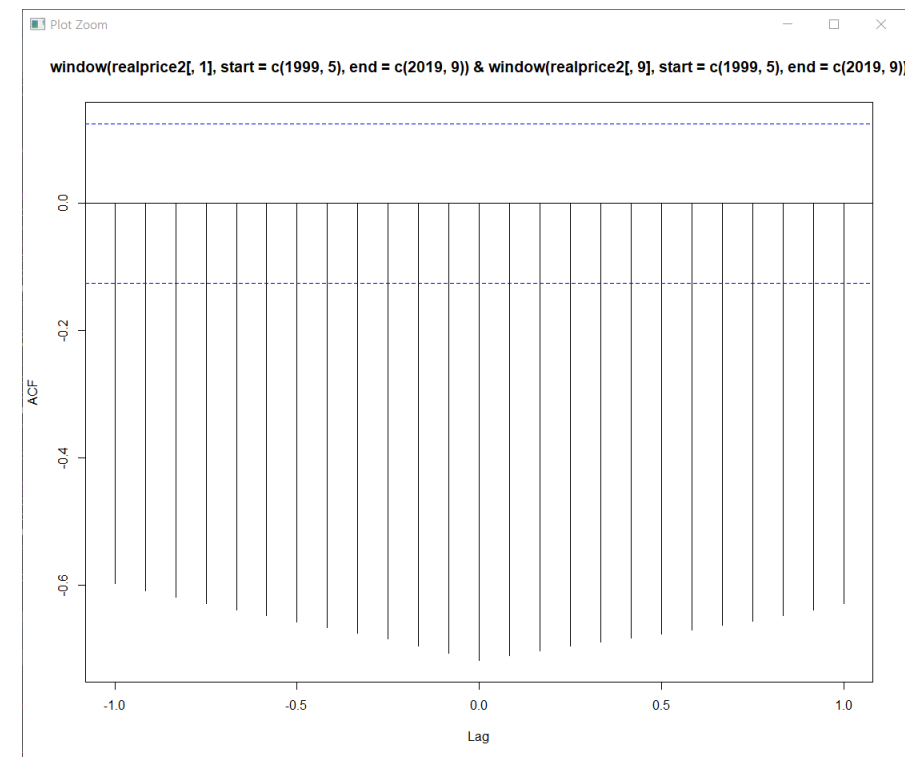
# 자료 살펴보기



\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

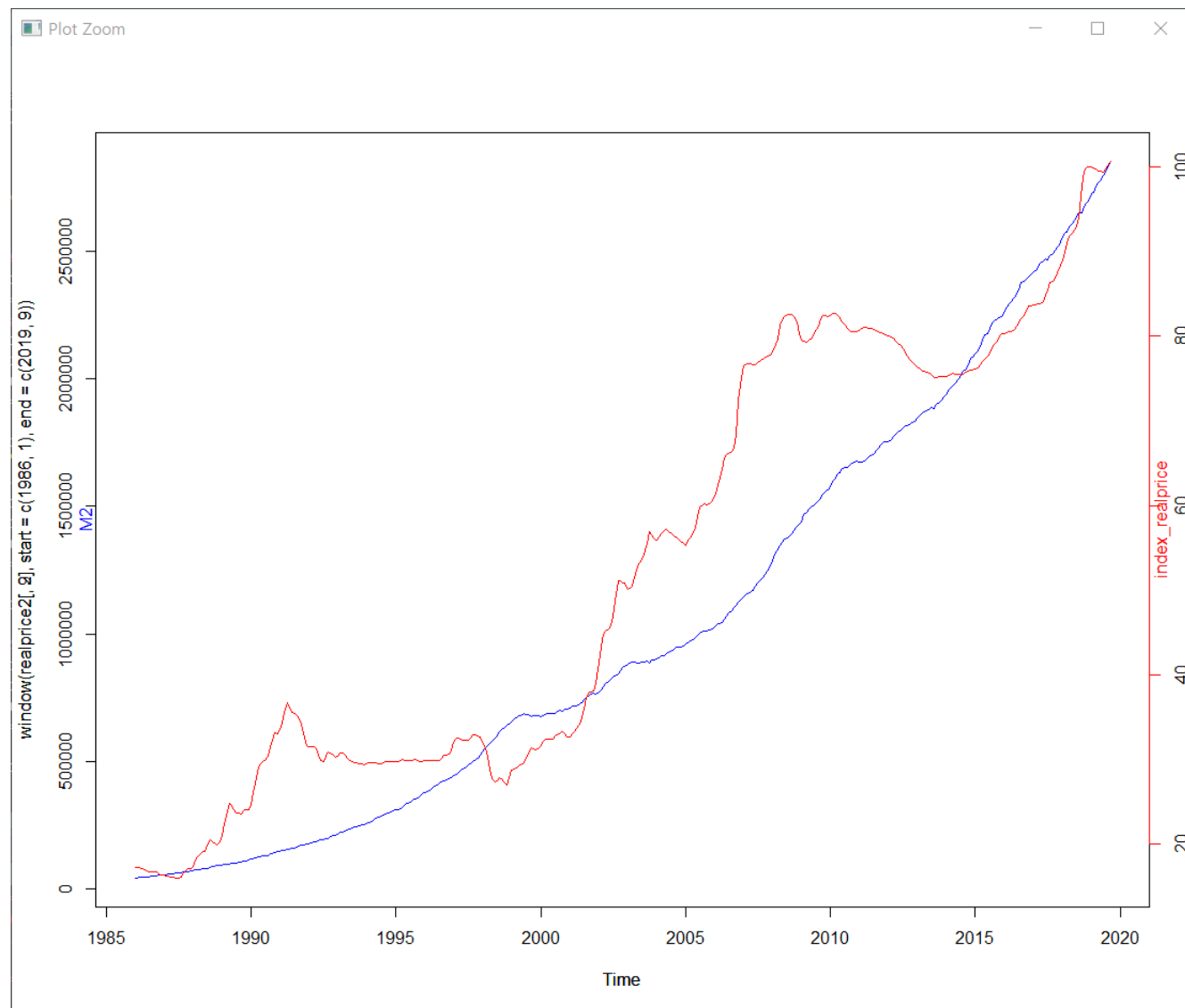
## 그래프 및 CCF (기준금리와 매매지수)

- ✓ 기준금리와 매매지수의 역의 관계는 볼 수 있으나 선후행 관계는 없음



```
plot(window(realprice2[,1], start=c(1999,5), end=c(2019,9)), col="blue")
mtext("baserate", side=2, col="blue")
par(new=T) # 기준금리와 역의 관계
plot(window(realprice2[,9], start=c(1999,5), end=c(2019,9)), col="red", axes=F)
axis(side=4, col="red"); mtext("index_realprice", side=4, col="red")
ccf(window(realprice2[,1], start=c(1999,5), end=c(2019,9)), window(realprice2[,9], start=c(1999,5), end=c(2019,9)), lag=12)
```

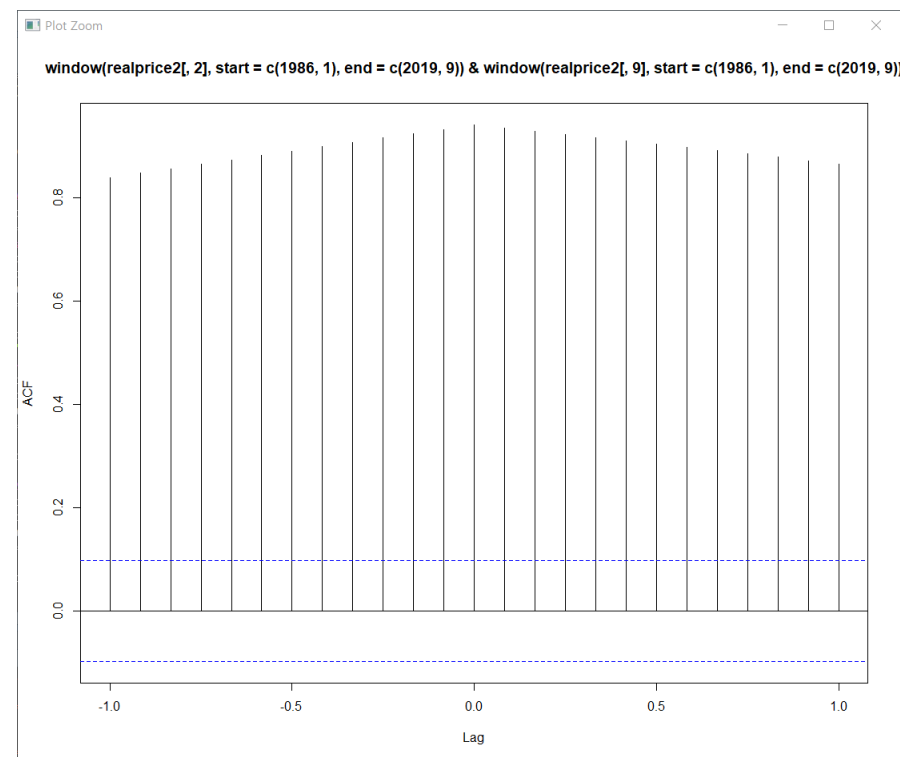
# 자료 살펴보기



\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

## 그래프 및 CCF (통화량과 매매지수)

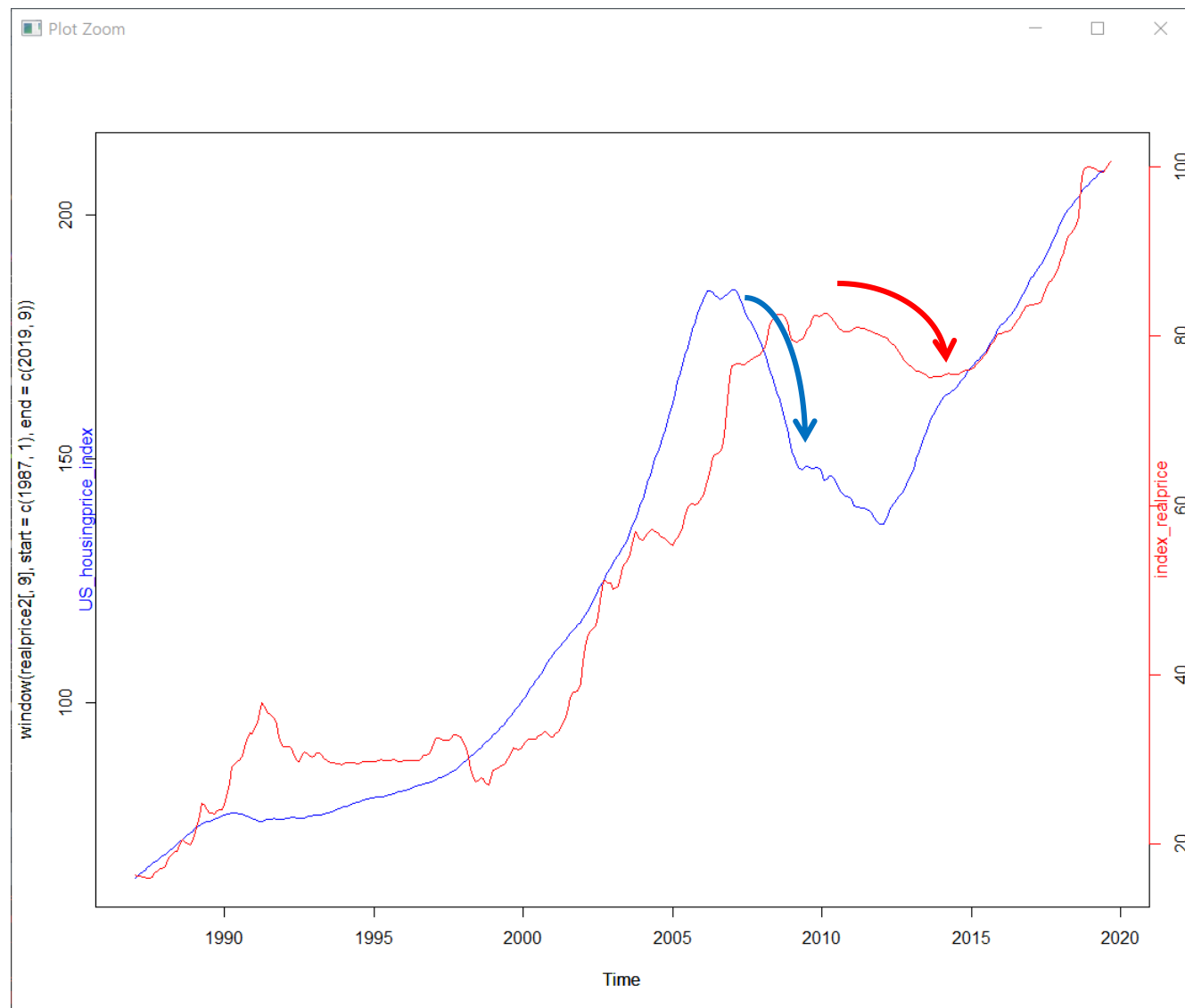
- ✓ 통화량과 매매지수의 정의 관계는 볼 수 있으나 선후행 관계는 없음



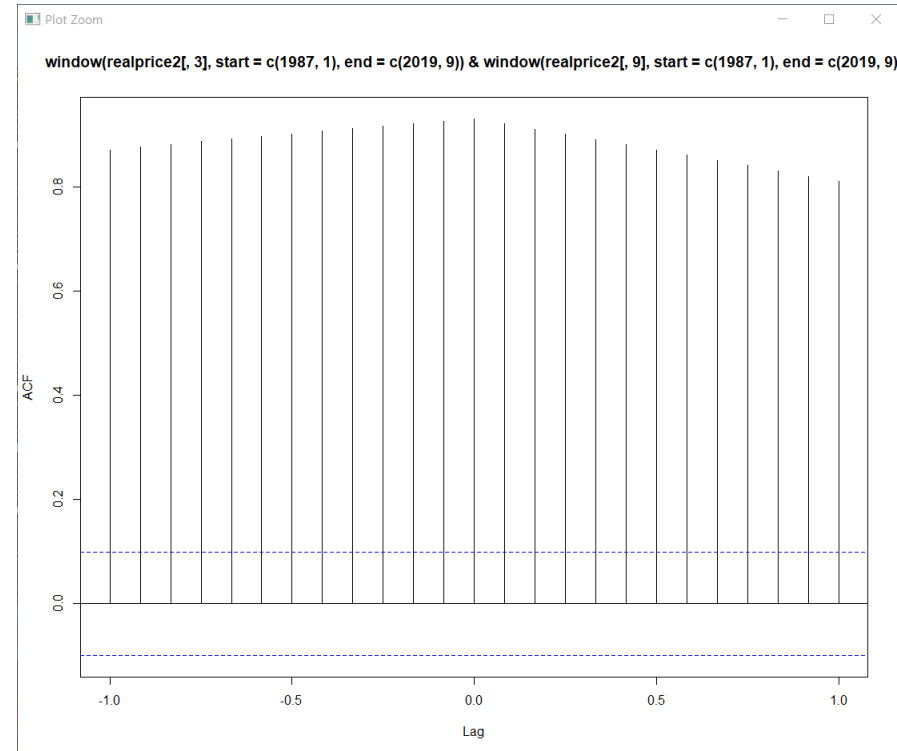


# 자료 살펴보기

## 그래프 및 CCF (미국주택가격지수와 매매지수)

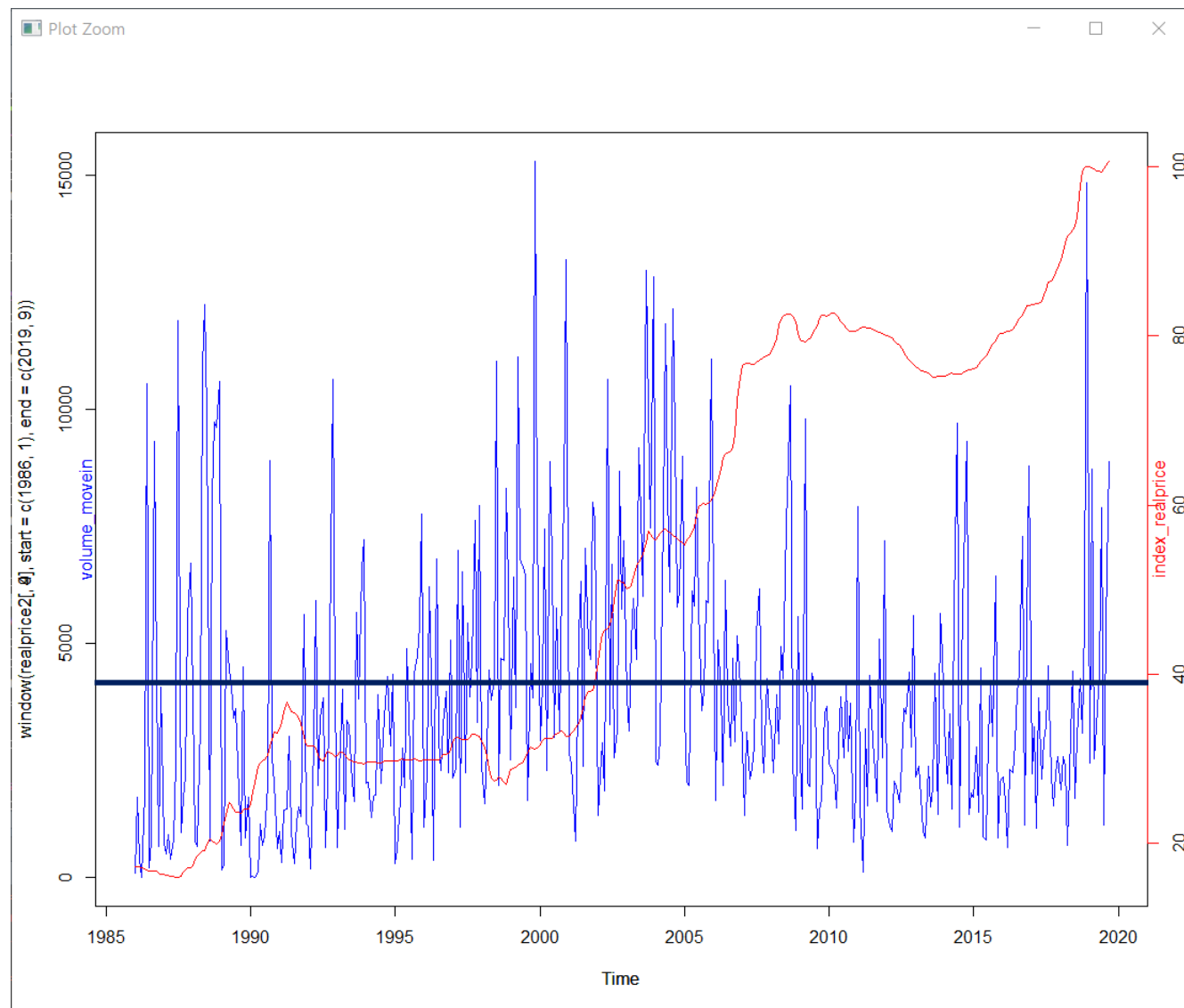


- ✓ 미국주택가격지수와 매매지수의 정의 관계는 볼 수 있으나 선후행 관계는 없음
- ✓ 그래프상으로는 미국주택가격지수가 서울 아파트매매지수보다 선행하는걸로 보임



\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

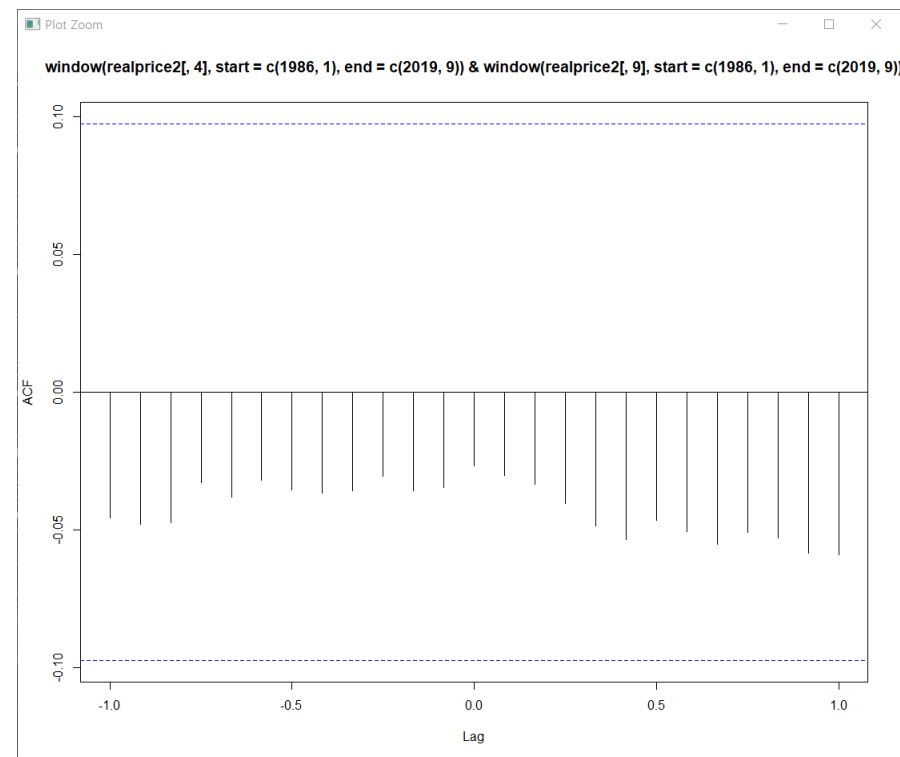
# 자료 살펴보기



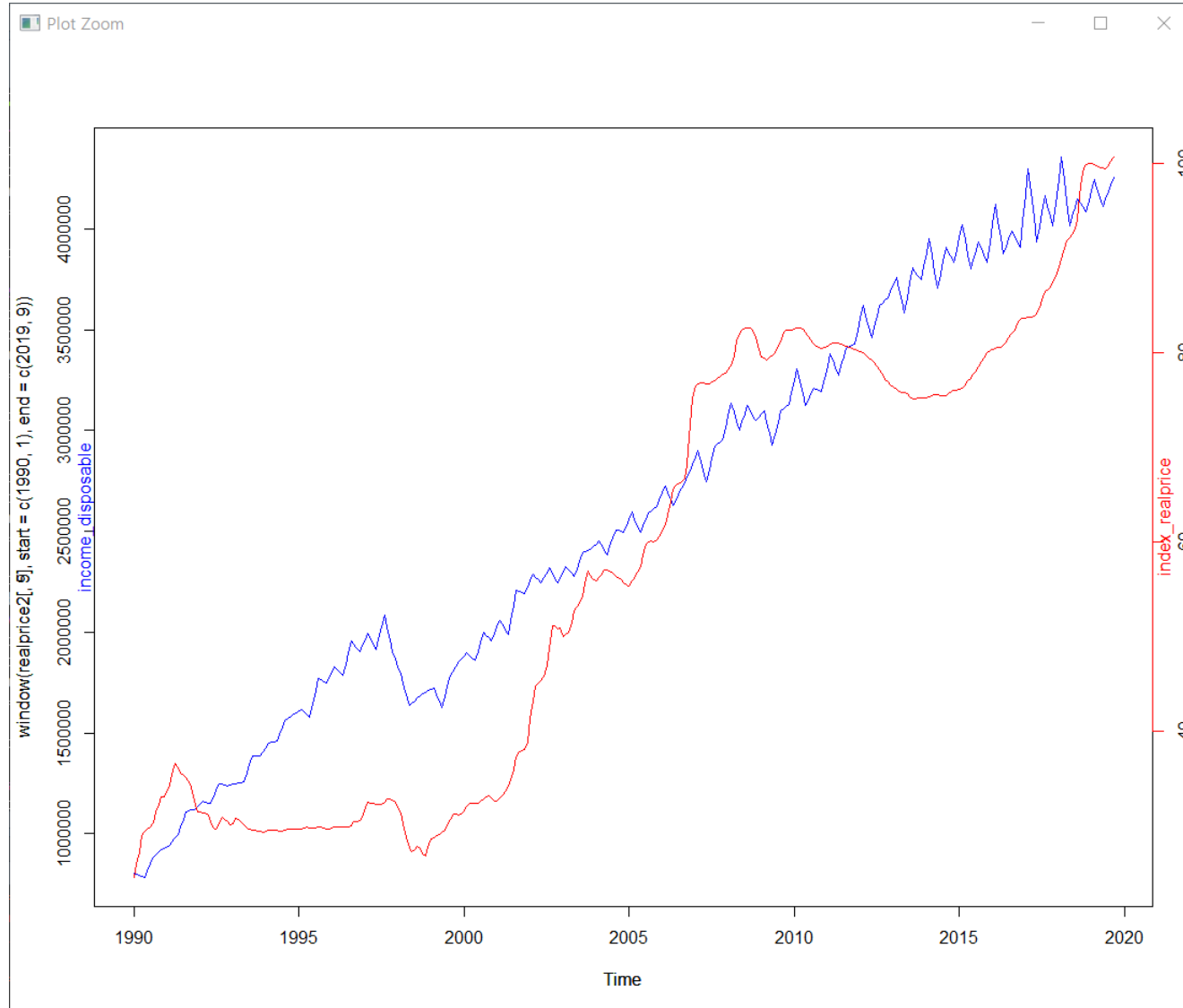
\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

## 그래프 및 CCF (입주량과 매매지수)

- ✓ 입주량은 매월 일정범위에서 변동하는 양으로 장기평균에 따른 방향성을 봐야함
- ✓ 입주량과 매매지수의 역의 관계 및 선후행 관계는 유의미 하다고 볼 수 없음
- ✓ 입주량의 장기평균(3,888세대/월) ↔ 연식별 재고주택(연식별 공급량의 관점, 자료X)



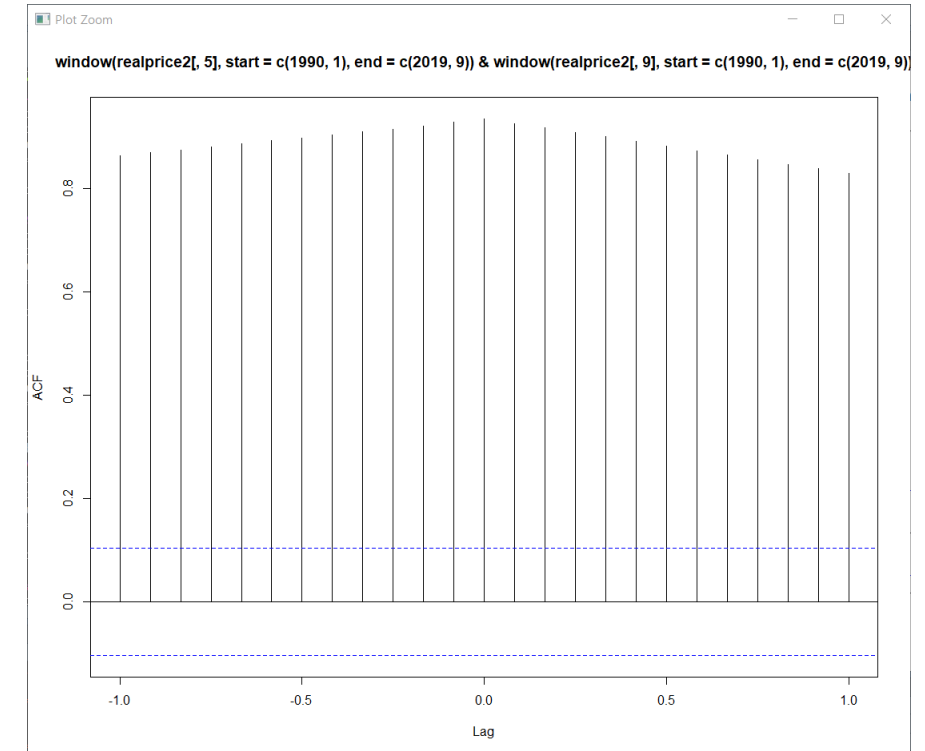
# 자료 살펴보기



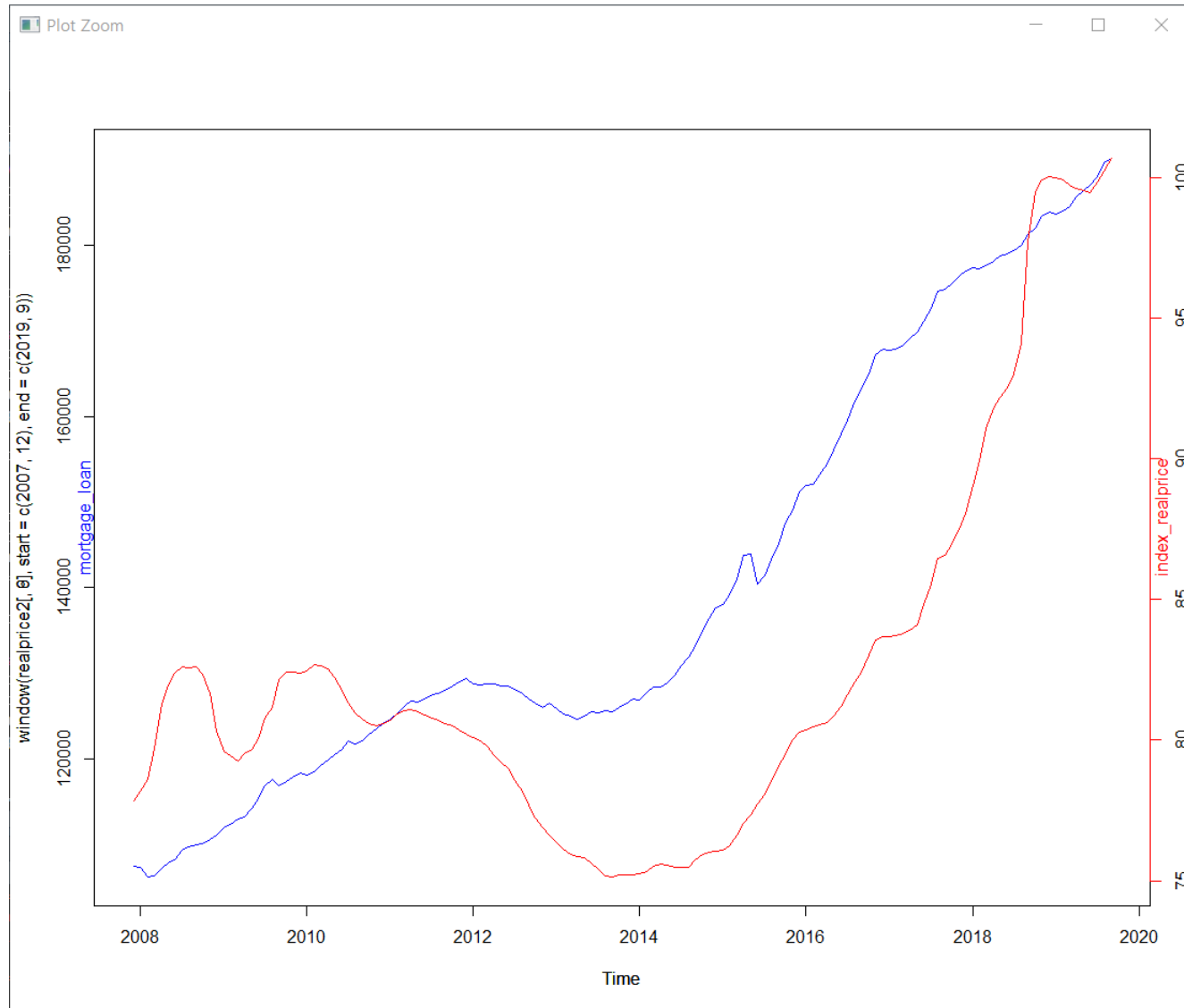
\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

## 그래프 및 CCF (가처분소득과 매매지수)

- ✓ 가처분소득과 매매지수의 정의 관계는 볼 수 있으나 선후행 관계는 없음



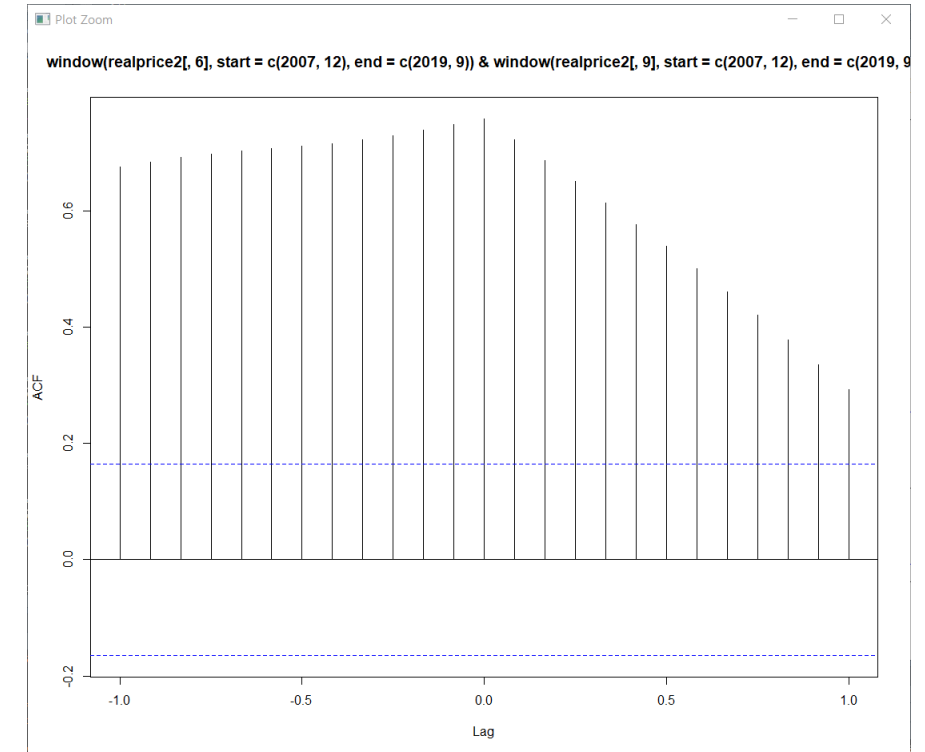
# 자료 살펴보기



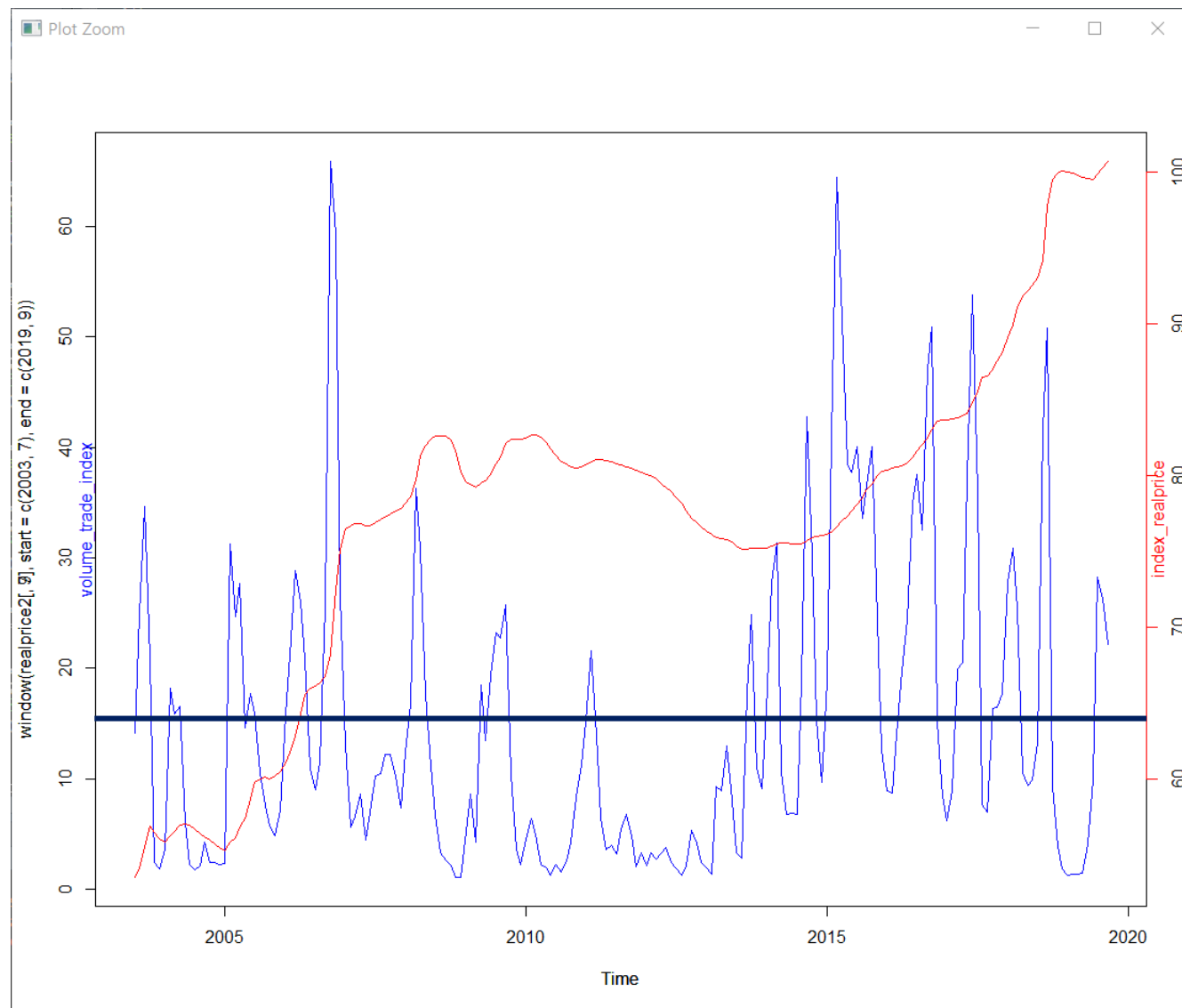
\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

## 그래프 및 CCF (담보대출과 매매지수)

- ✓ 담보대출과 매매지수의 정의 관계는 볼 수 있으나 선후행 관계는 없음



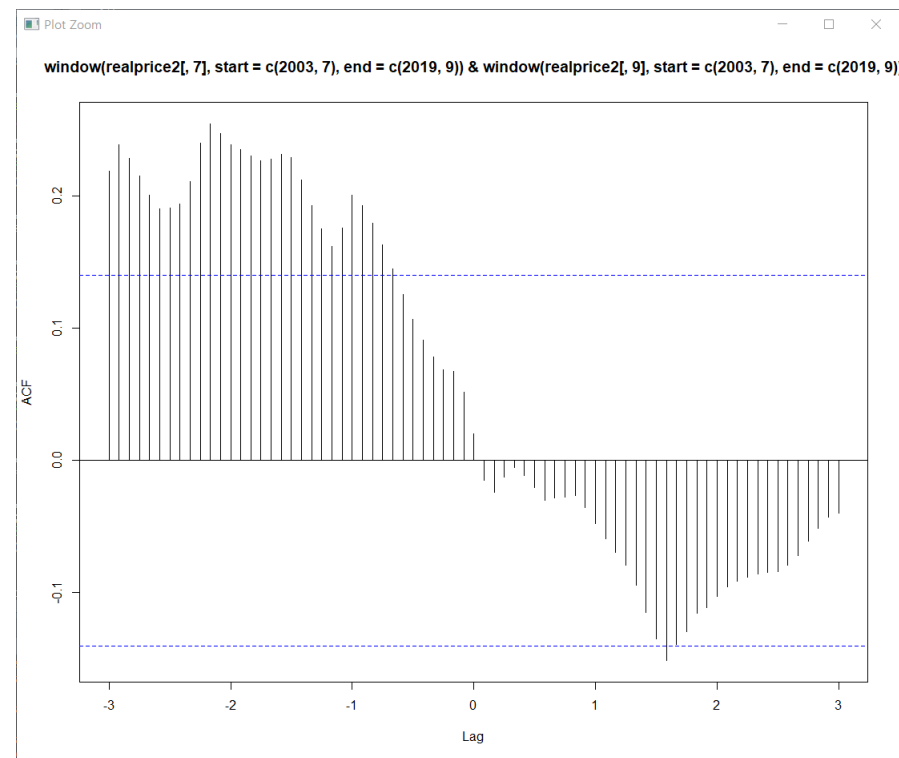
# 자료 살펴보기



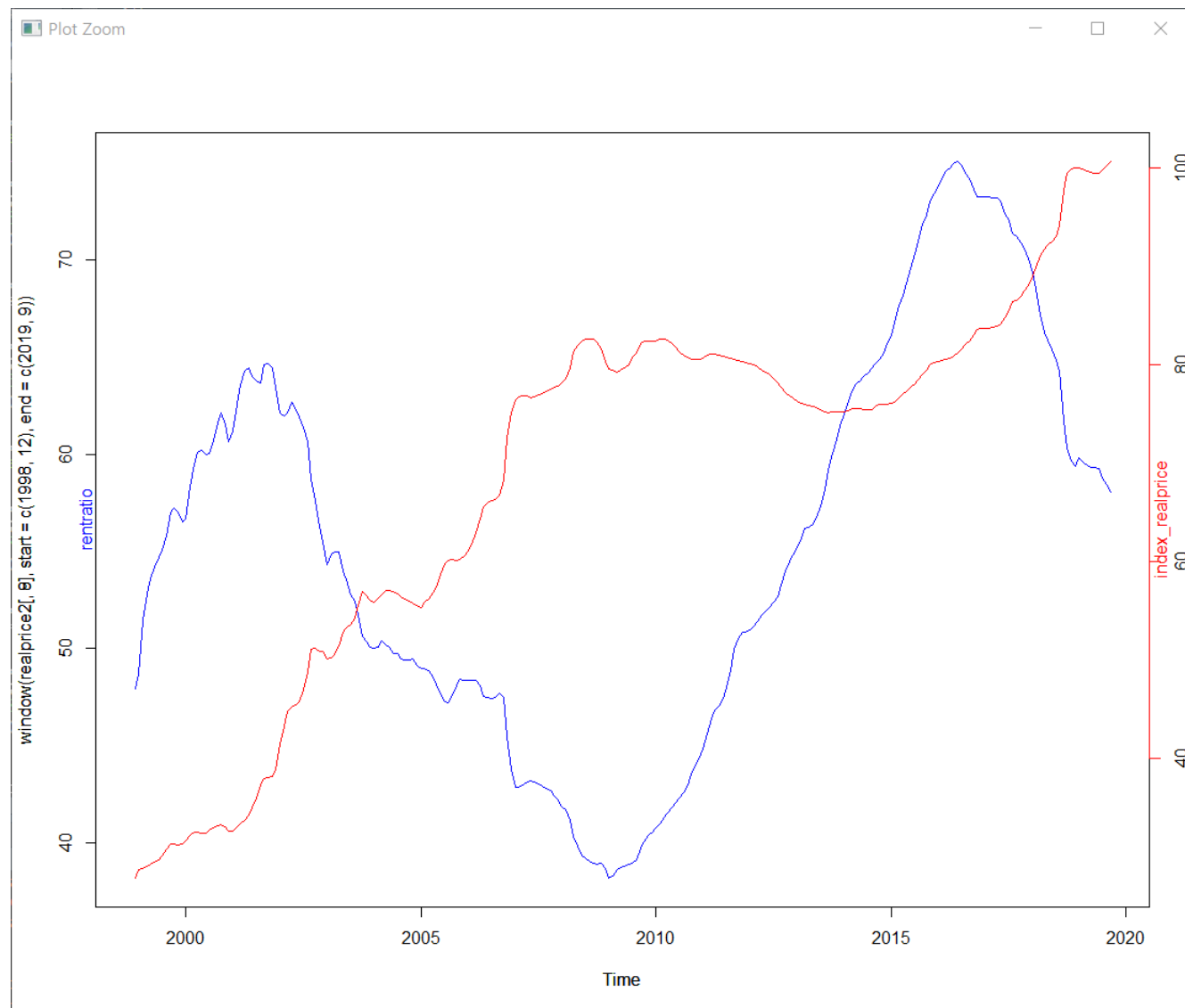
\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

## 그래프 및 CCF (매매거래량지수와 매매지수)

- ✓ 매매거래량지수는 매월 일정범위에서 변동하는 양으로 장기평균(15)에 따른 방향성을 봐야함
- ✓ 매매거래량지수와 매매지수의 정의 관계는 있고, 1~2년 전후로 매매지수에 선행



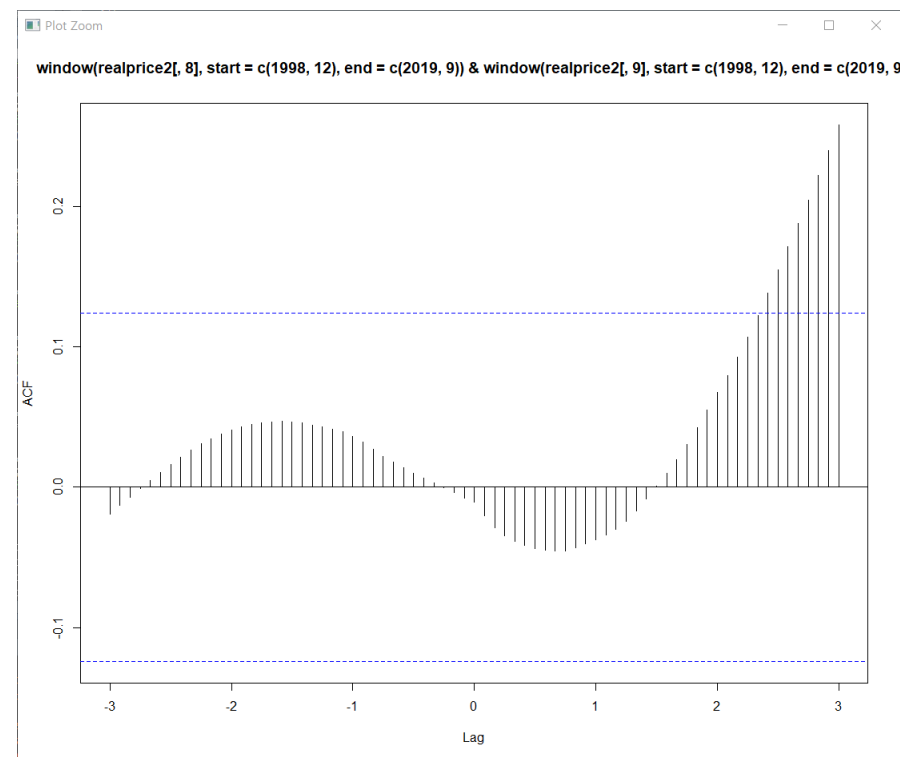
# 자료 살펴보기

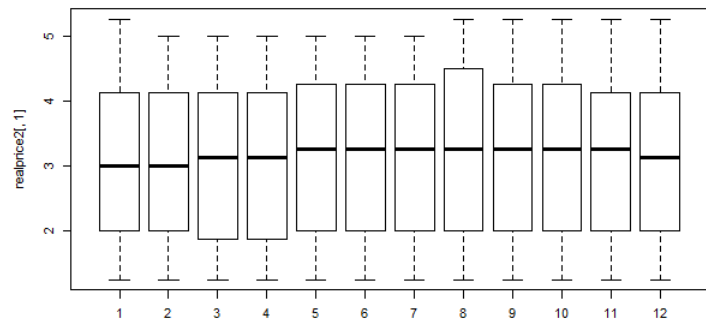


\* 장기적인 관계를 보고자 두 변수간 최대기간으로 살펴봄

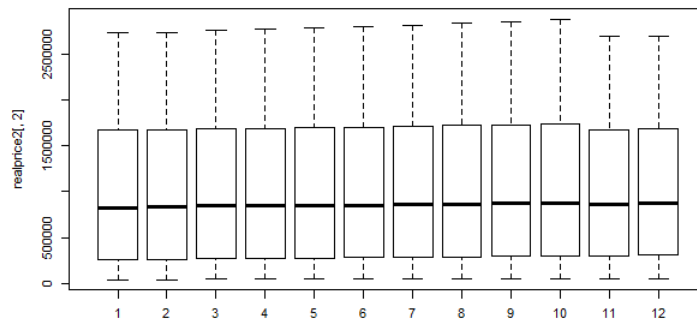
## 그래프 및 CCF (전세가율과 매매지수)

- ✓ 전세가율과 매매지수는 역방향 관계로 보이나 추세때문에 파악하기 힘들
- ✓ 매매지수가 전세가율보다 오히려 2.5년 이상 선행하는 것으로 보임

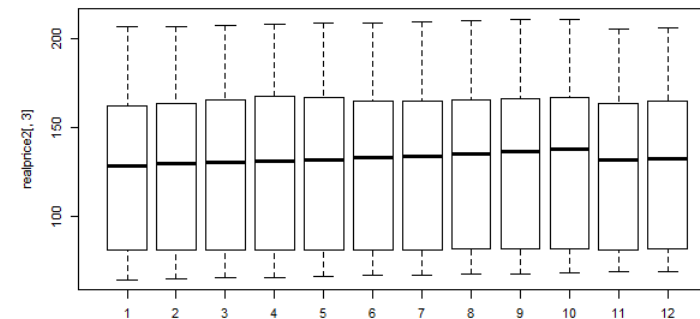




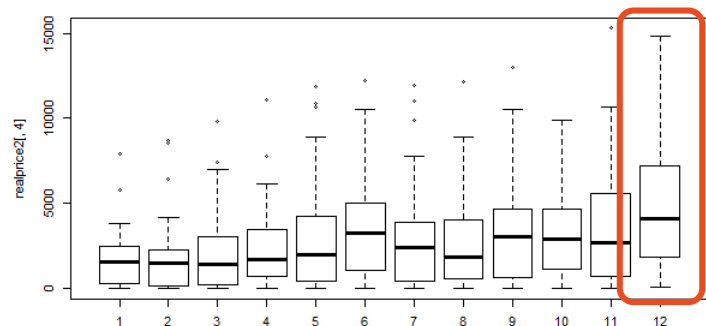
< 기준금리 >



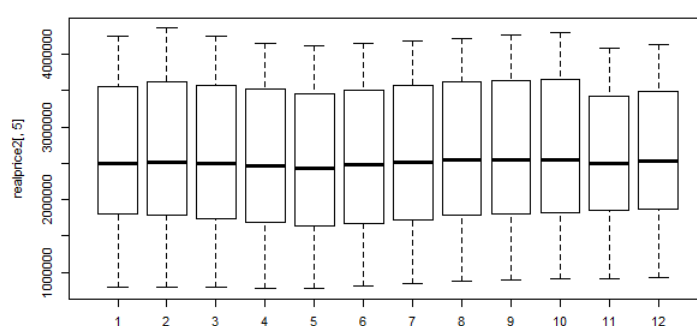
< 통화량 >



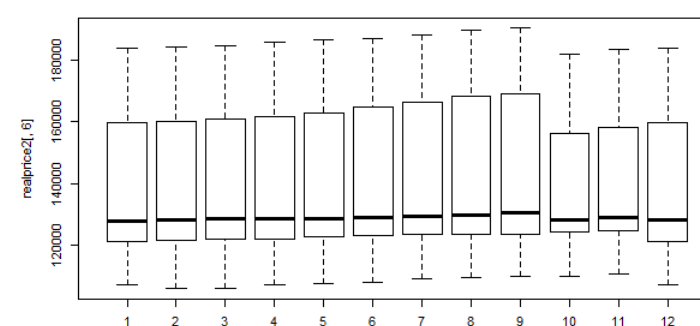
< 미국주택가격지수 >



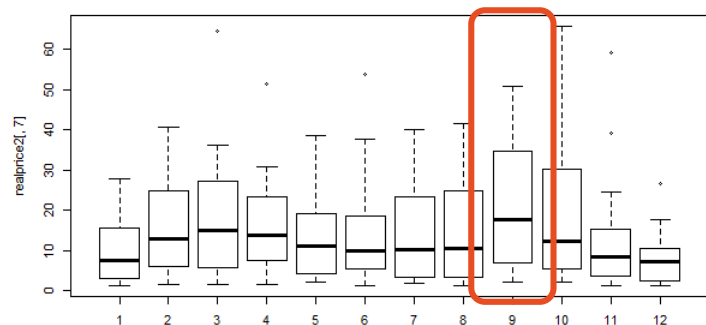
< 입주량 >



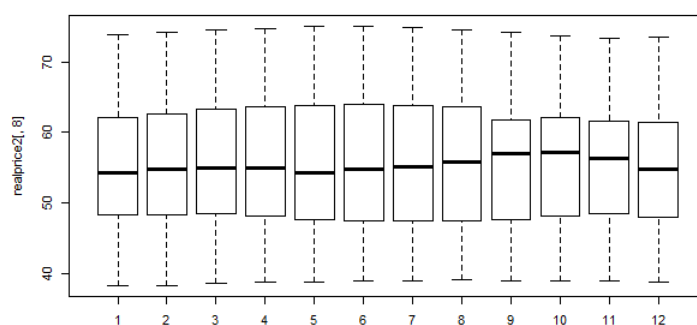
< 가처분소득 >



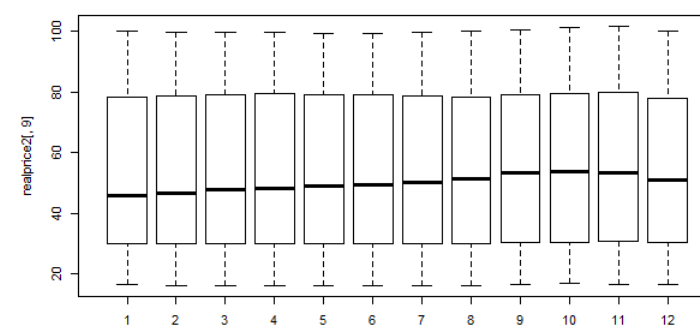
< 담보대출량 >



< 매매거래량지수 >



< 전세가율 >

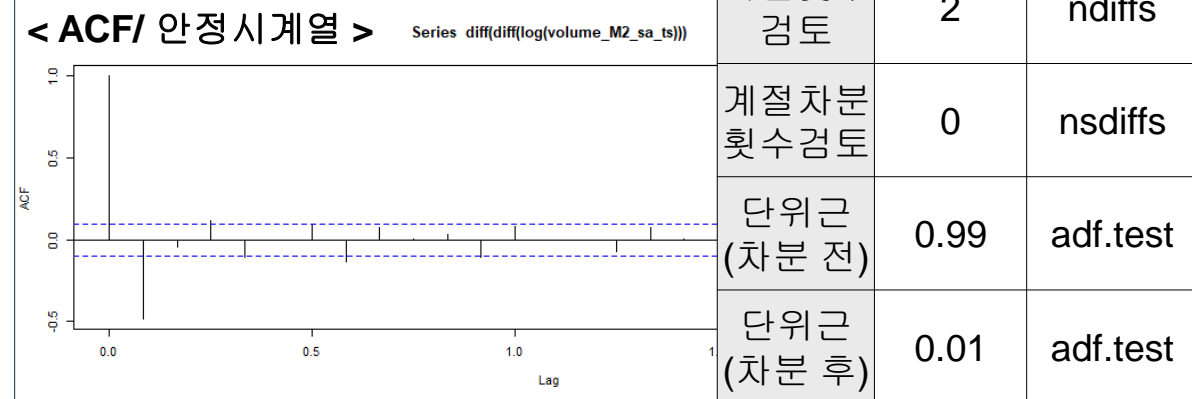
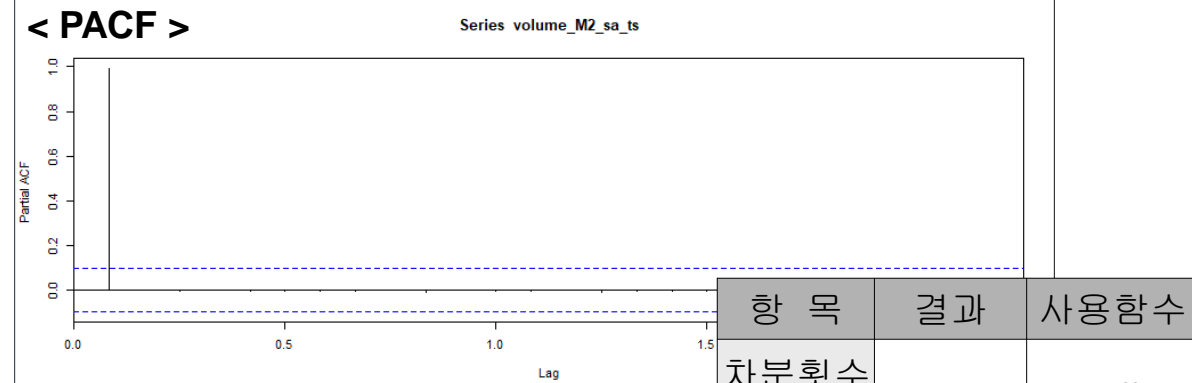
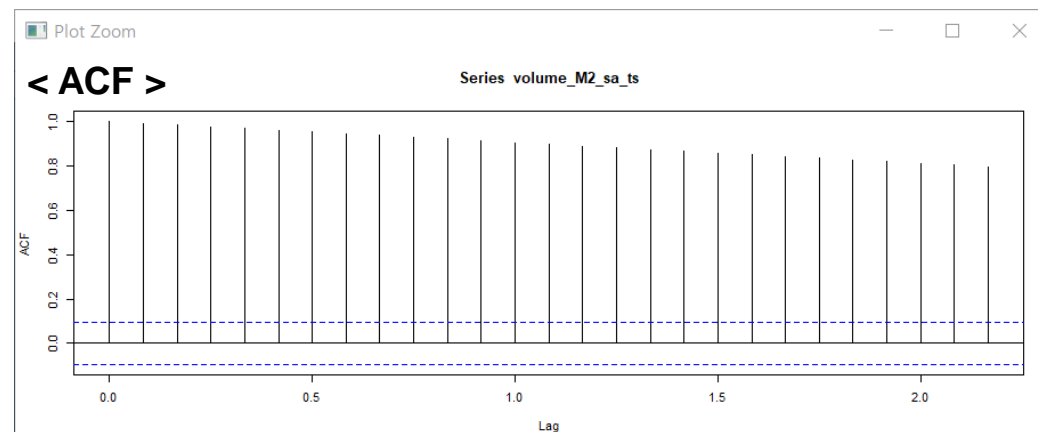
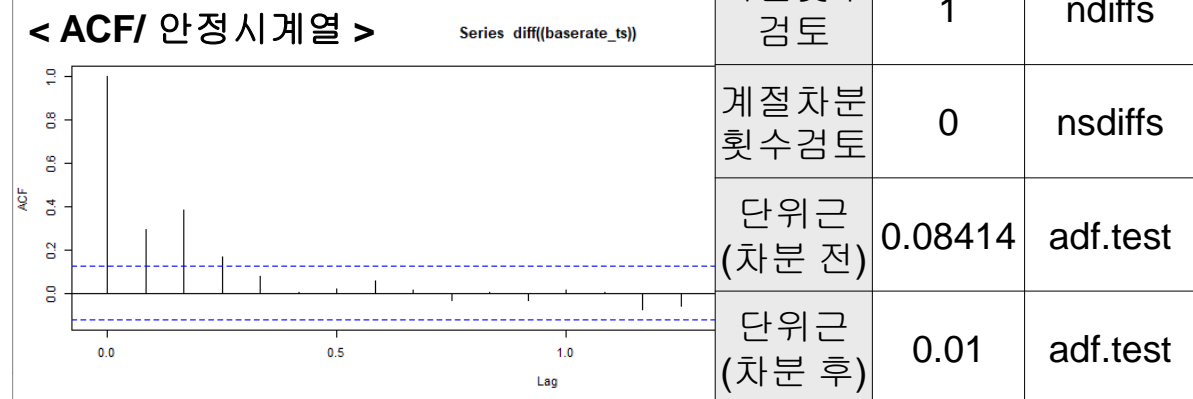
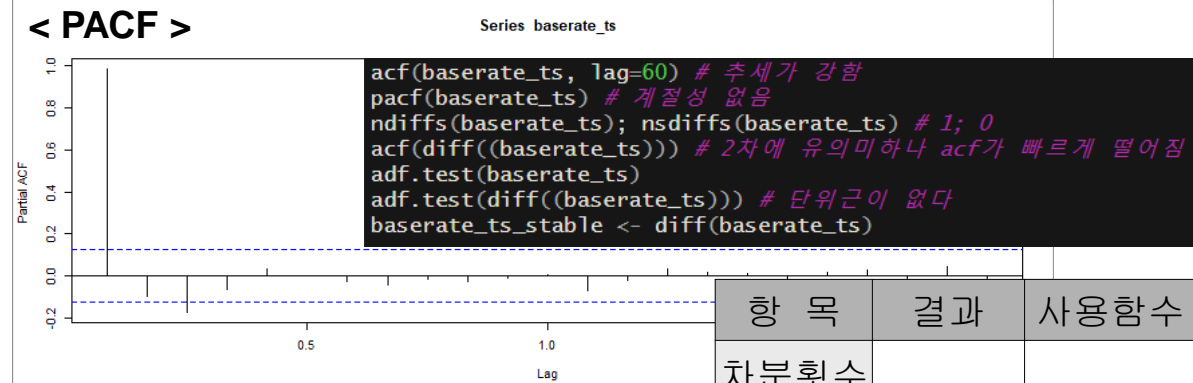
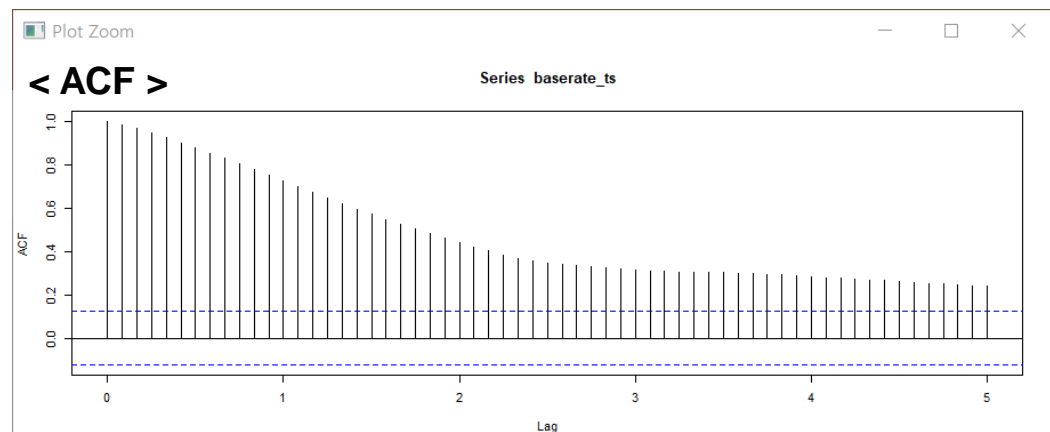


< 매매지수 >



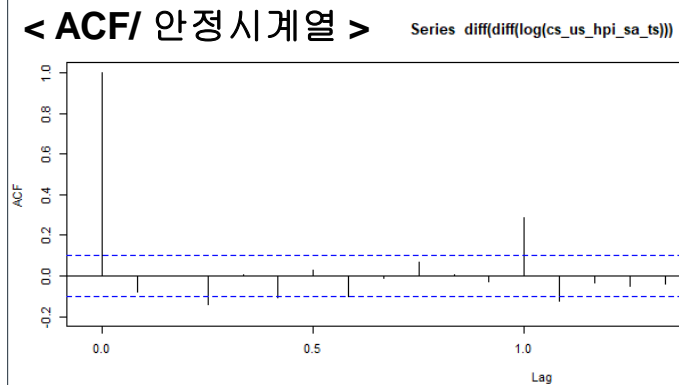
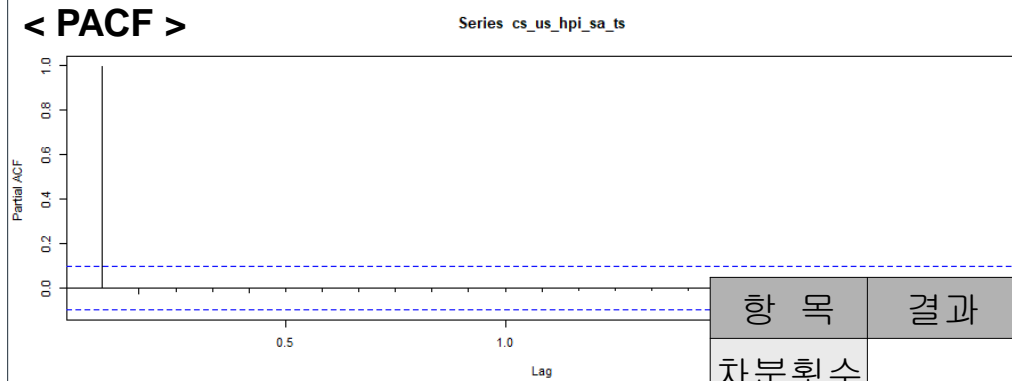
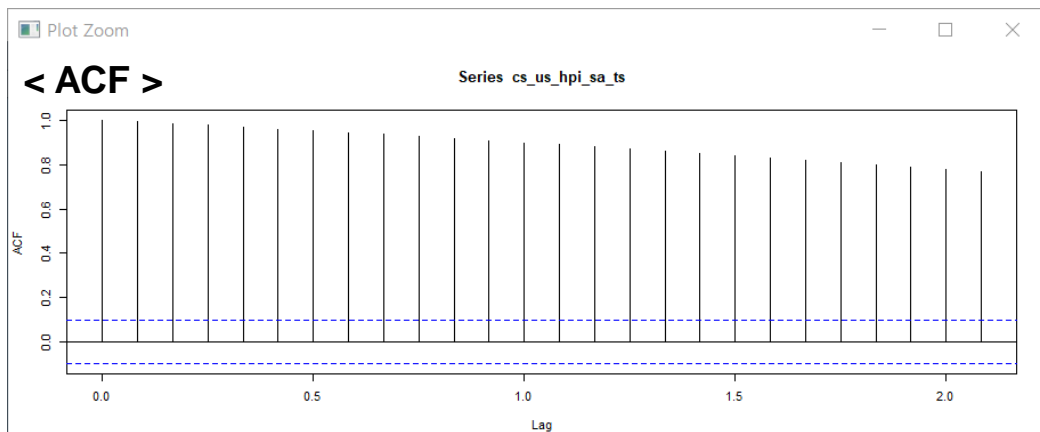
# 안정시계열로 변경

한국은행 기준금리 / 통화량(M2)

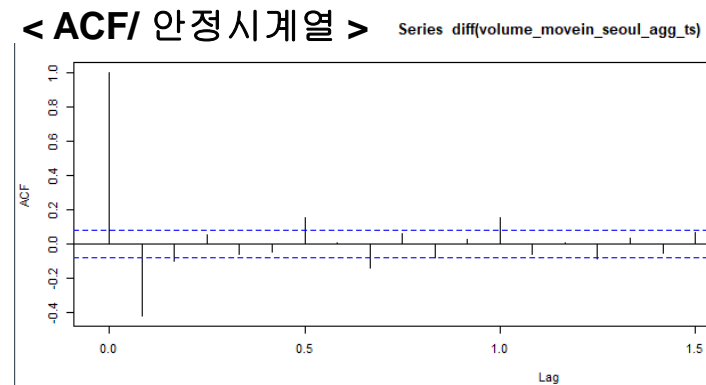
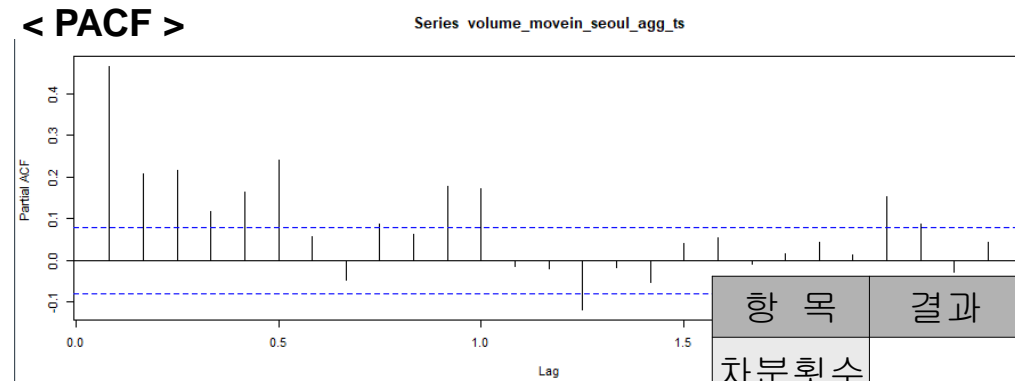
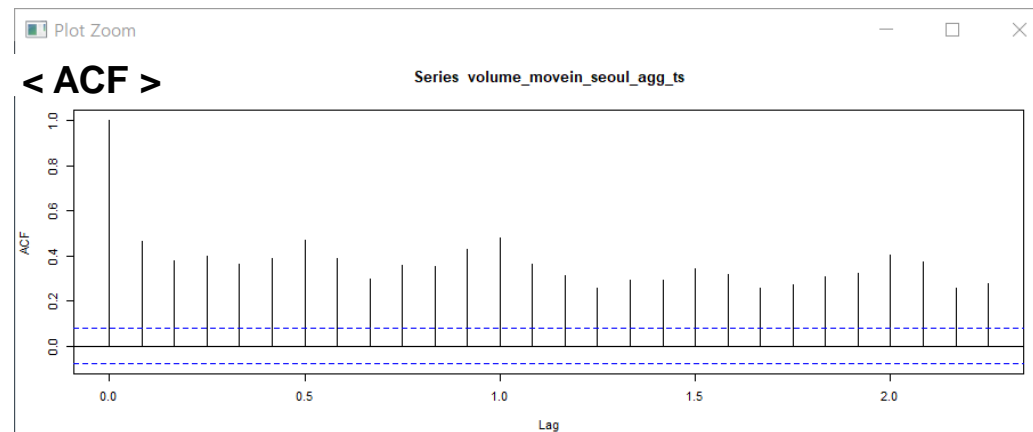


## 안정시계열로 변경

미국 주택가격지수 / 입주량



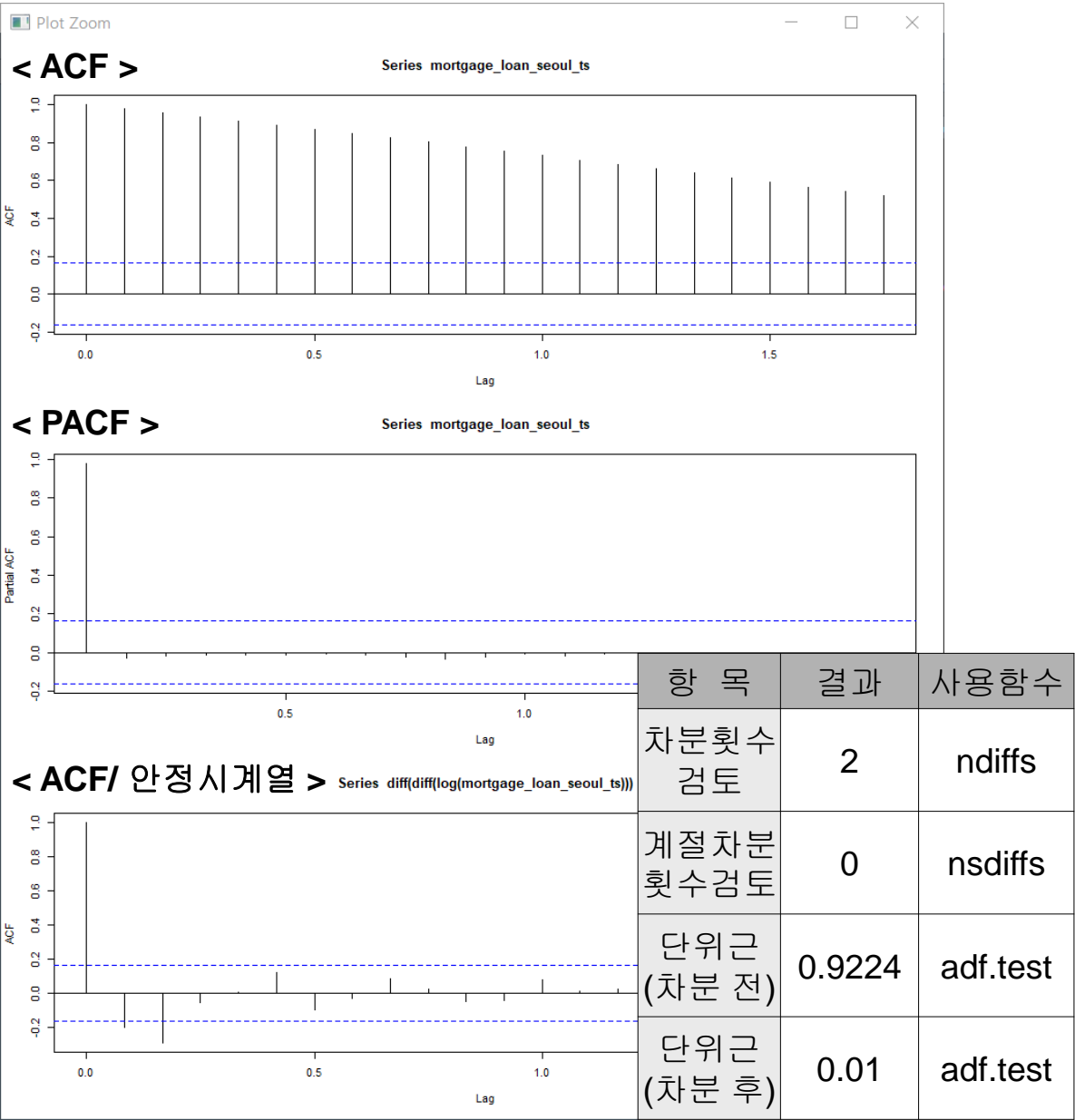
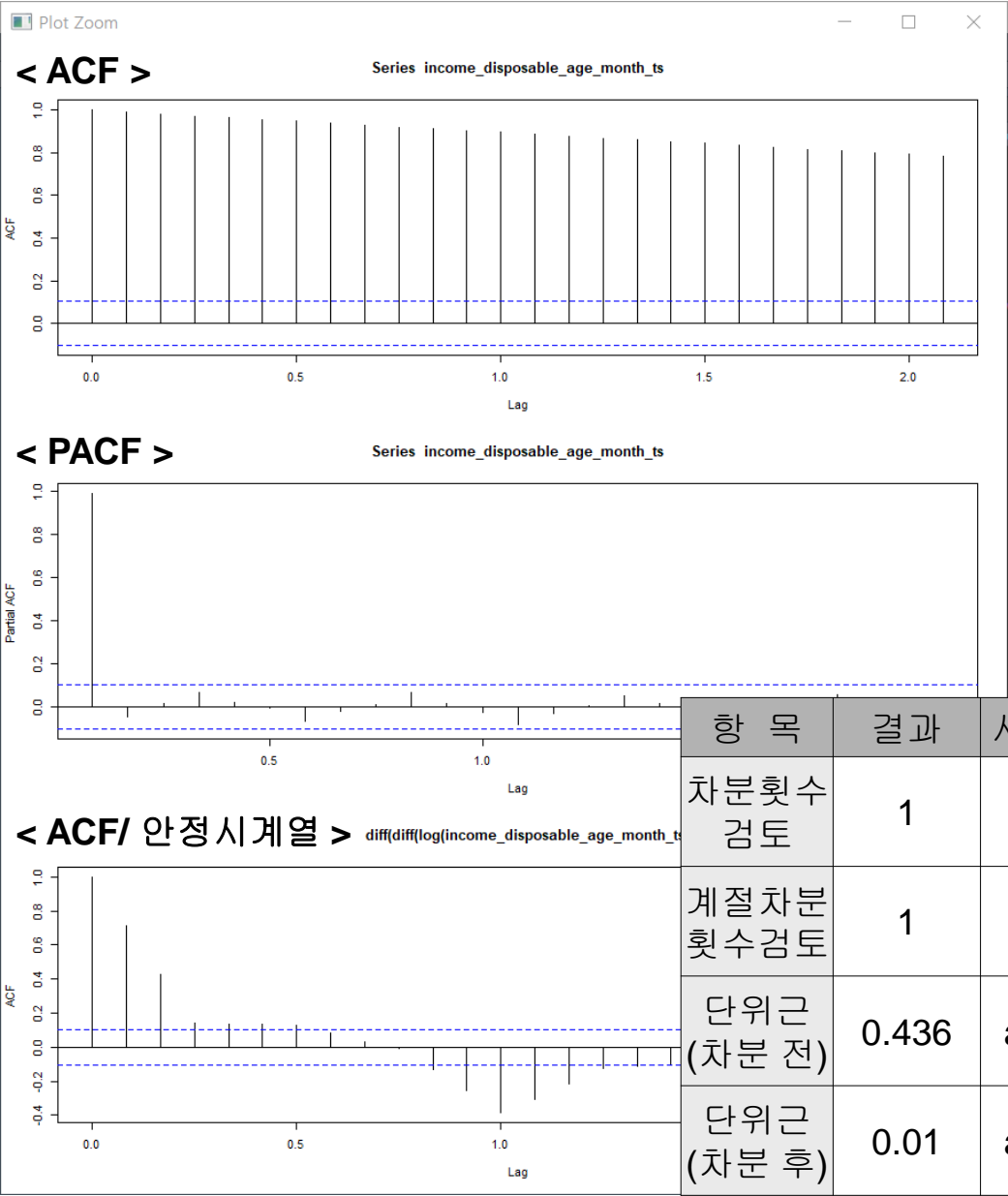
항 목	결과	사용함수
차분횟수 검토	1	ndiffs
계절차분 횟수검토	0	nsdiffs
단위근 (차분 전)	0.07556	adf.test
단위근 (차분 후)	0.01	adf.test



항 목	결과	사용함수
차분 횟수 검토	1	ndiffs
계절 차분 횟수검토	0	nsdiffs
단위근 (차분 전)	0.01	adf.test
단위근 (차분 후)	0.01	adf.test

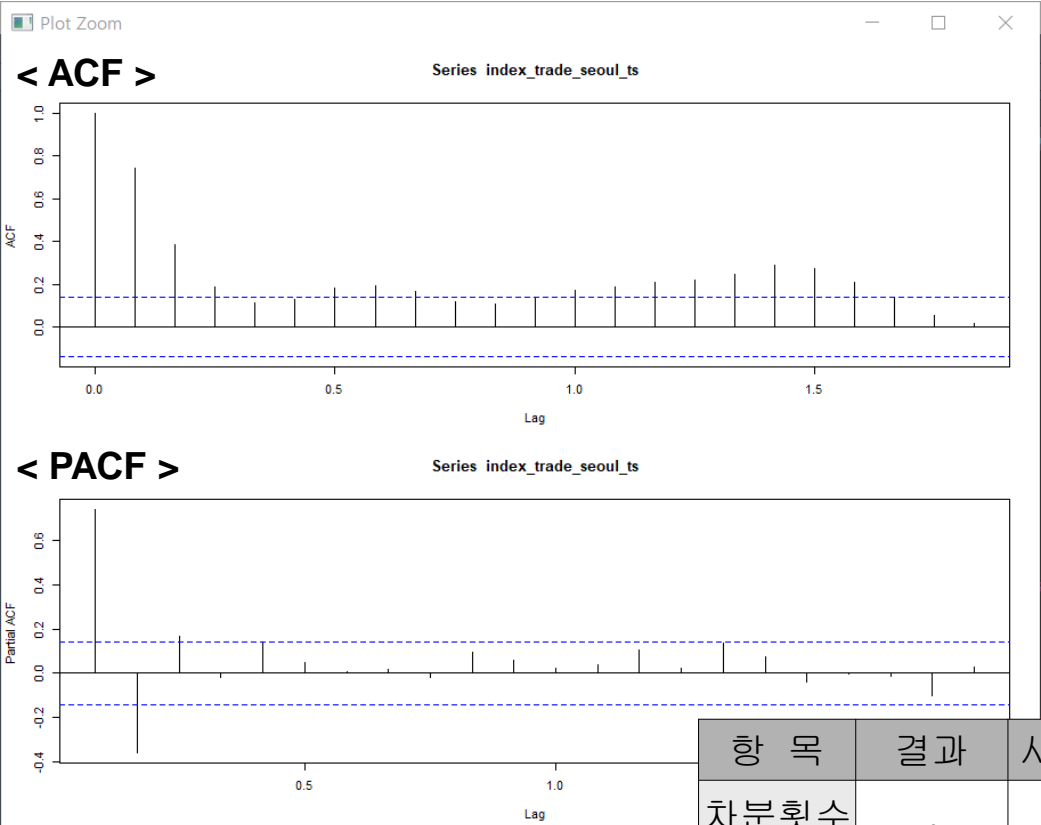
# 안정시계열로 변경

가처분소득 / 담보대출

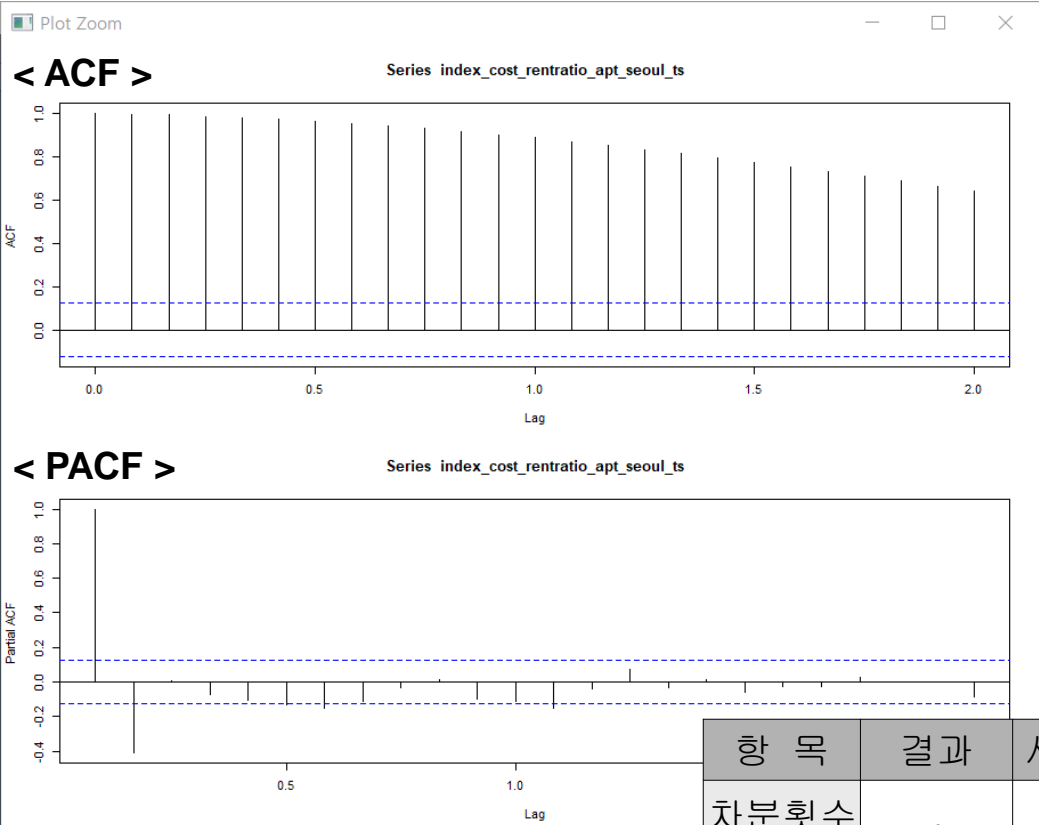


# 안정시계열로 변경

매매거래량지수 / 전세가율



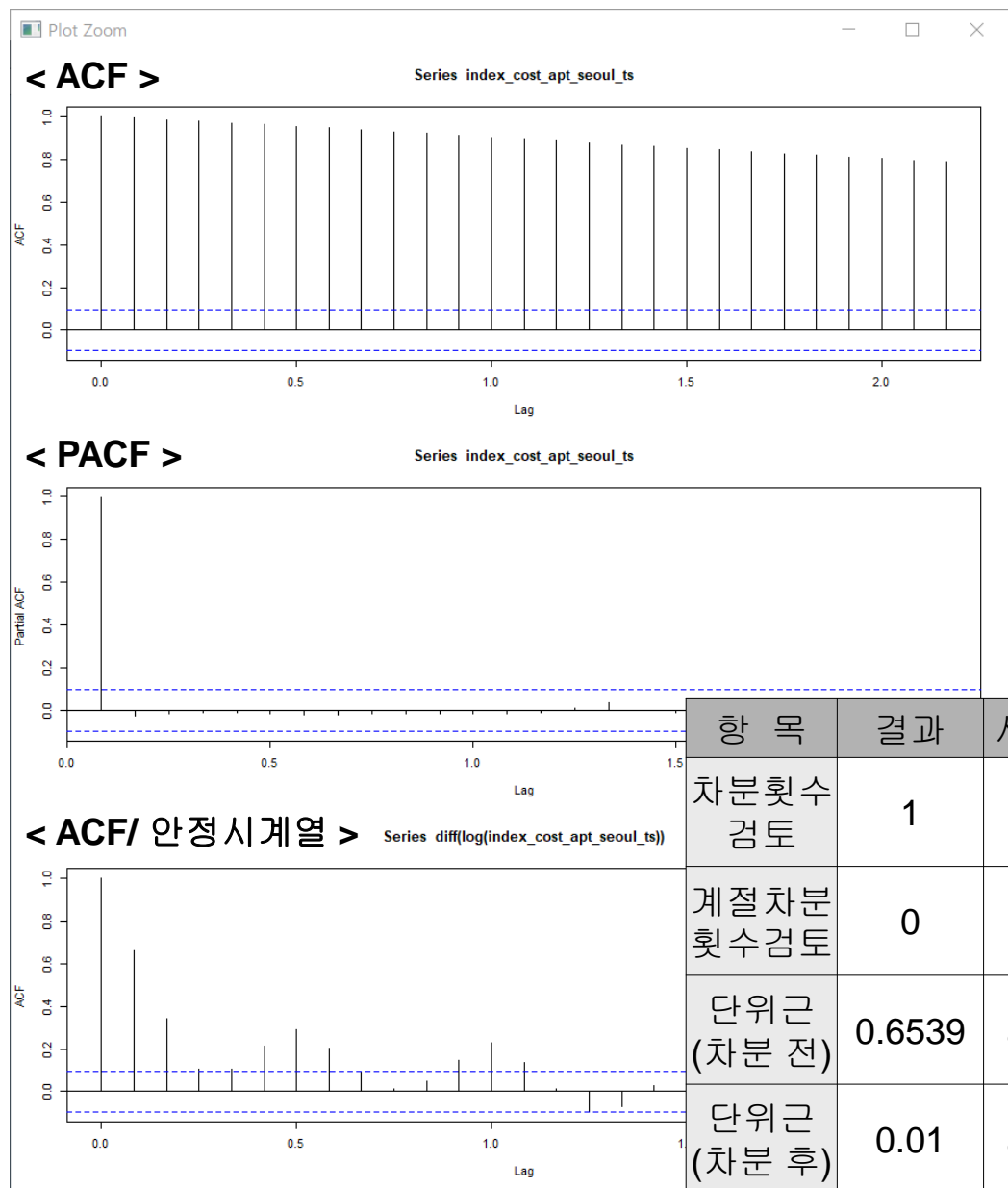
항 목	결과	사용함수
차분횟수 검토	1	ndiffs
계절차분 횟수검토	0	nsdiffs
단위근 (차분 전)	0.015	adf.test
단위근 (차분 후)	0.01	adf.test



항 목	결과	사용함수
차분횟수 검토	1	ndiffs
계절차분 횟수검토	0	nsdiffs
단위근 (차분 전)	0.5413	adf.test
단위근 (차분 후)	0.015	adf.test

# 안정시계열로 변경

매매지수



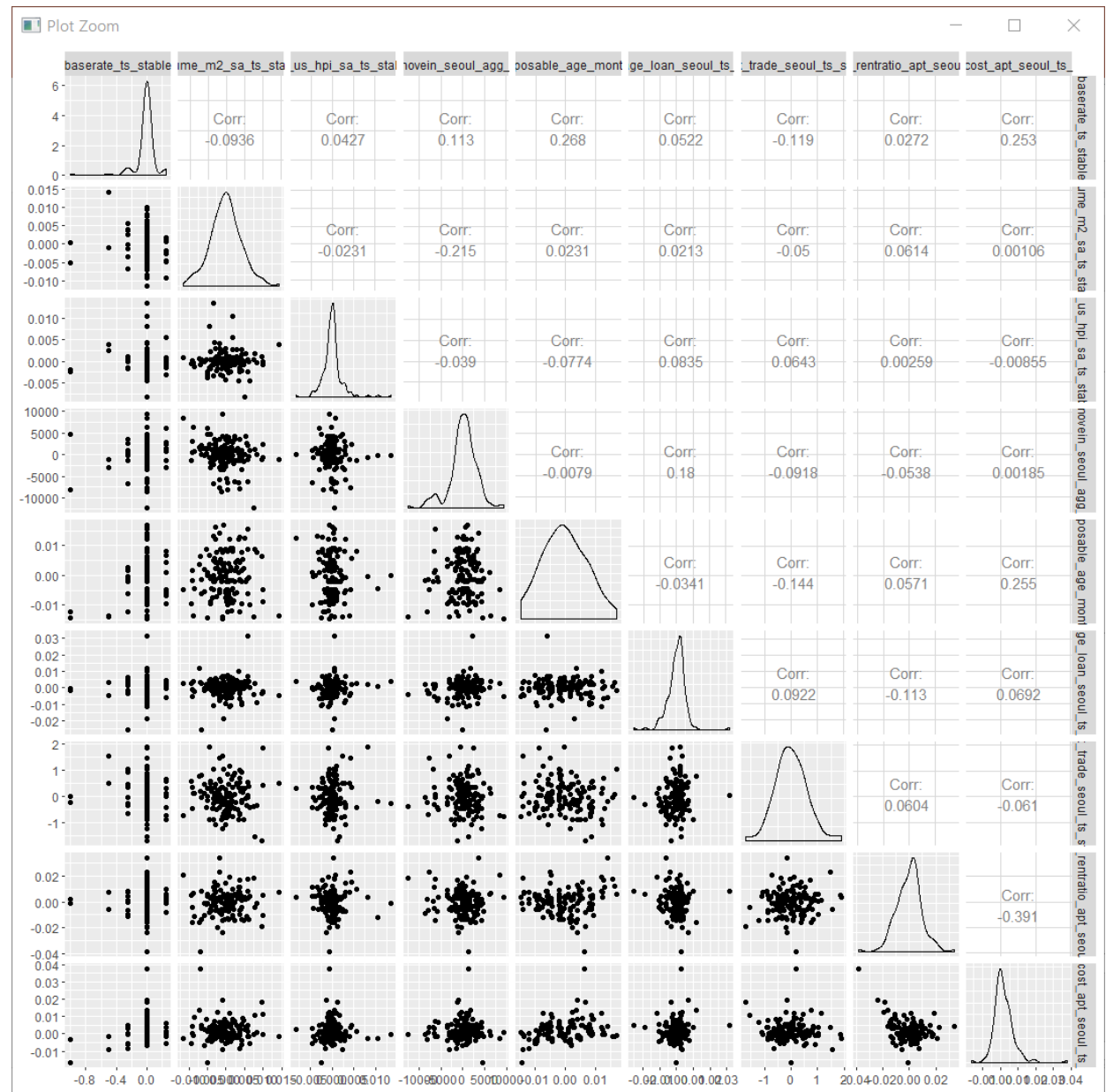
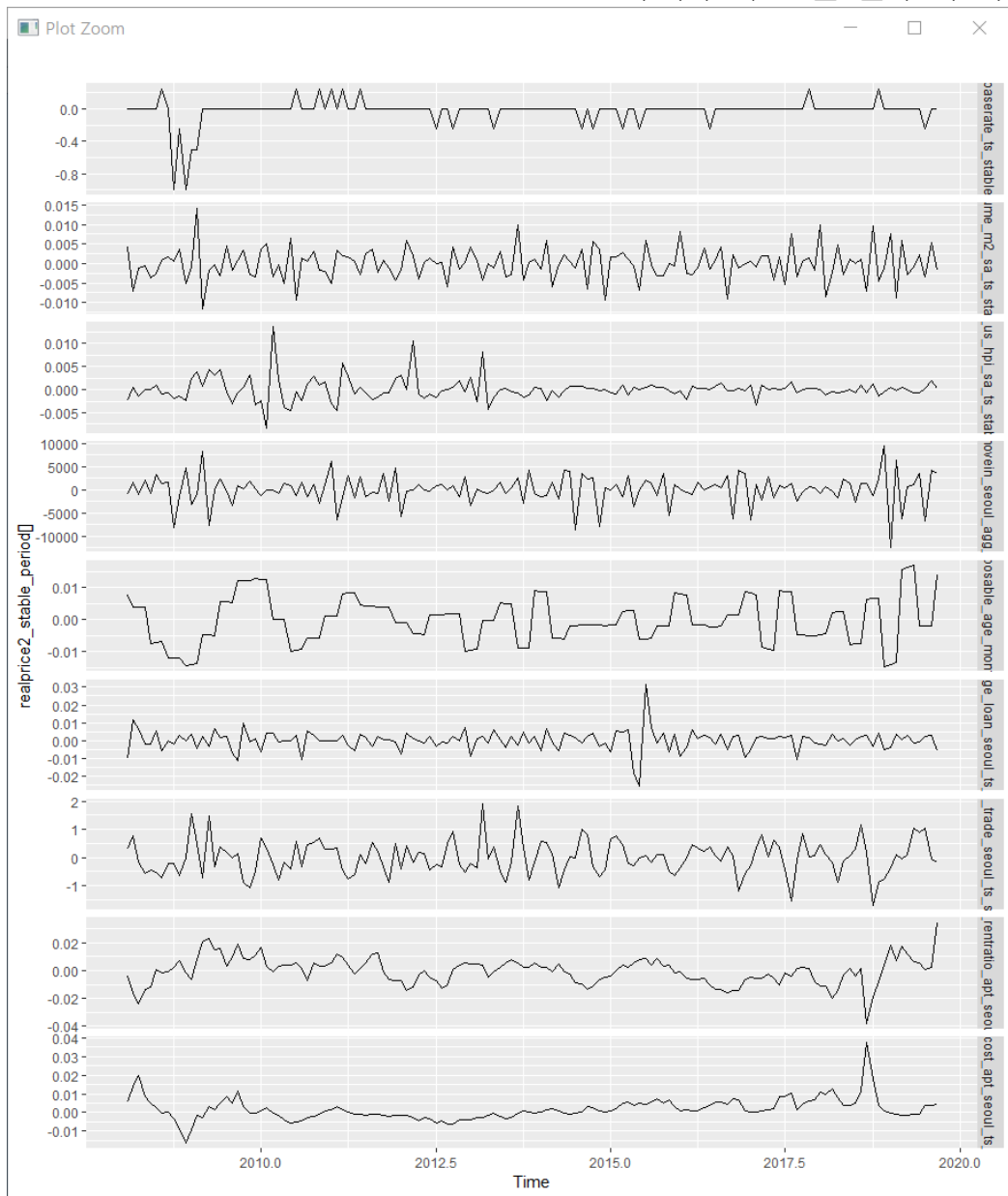
- ✓ 각 변수에 대해서 ndiffs, nsdiffs는 참고로 하면서 acf, pacf 그래프를 여러가지 옵션으로 그려보면서 안정시계열로 변경함
- ✓ 안정시계열로 변경결과 요약

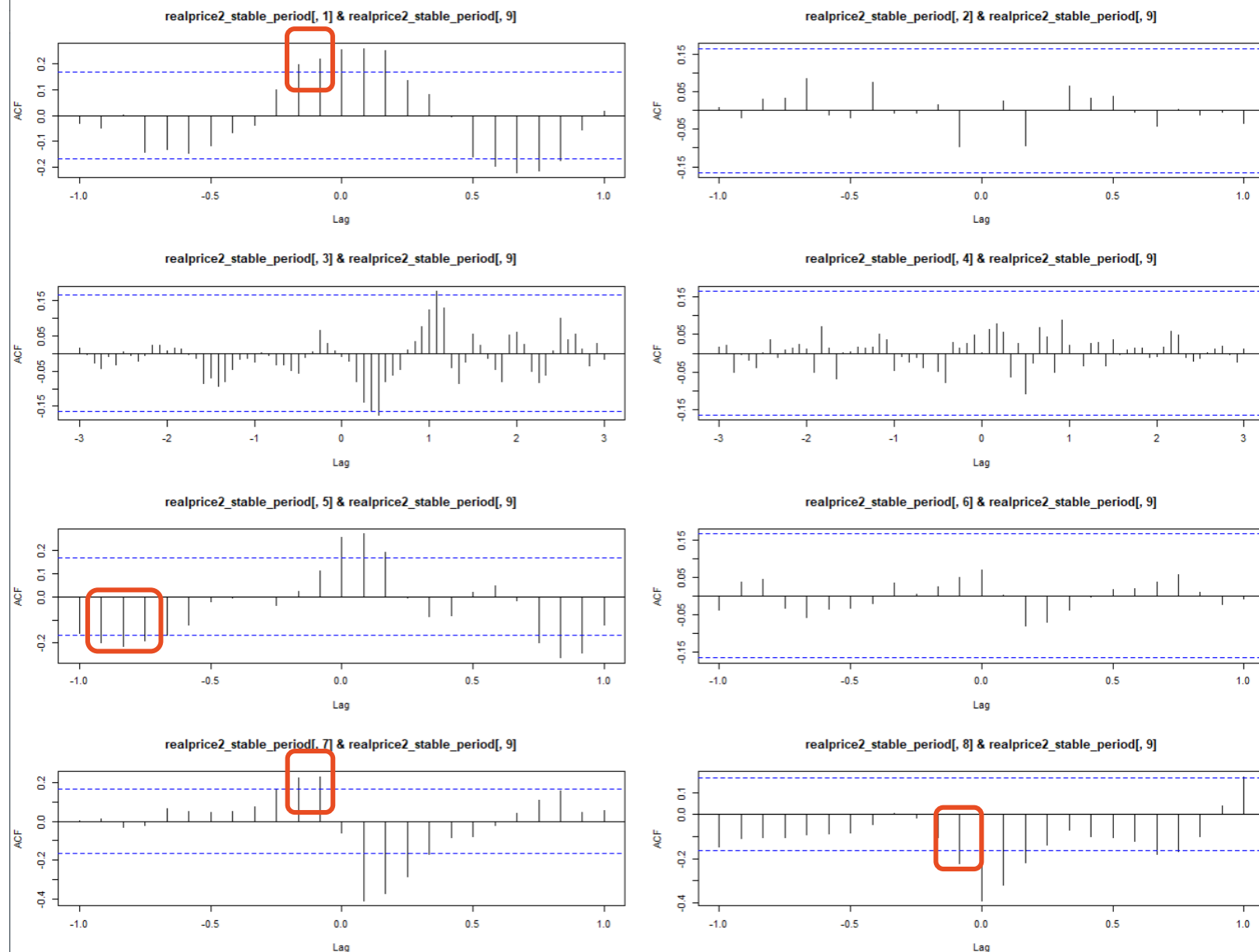
항 목	안정시계열 변경 방법
한국은행 기준금리	1차 차분
통화량(M2)	2차 차분 (log)
미국주택가격지수	2차 차분 (log)
입주량	1차 차분
가처분소득	2차 차분 (log, 12)
담보대출	2차 차분 (log)
매매거래량지수	1차 차분 (log)
전세가율	2차 차분 (log, 12)
매매지수	1차 차분 (log)

## 자료 살펴보기 / 안정시계열

\* 여기부터 모든 변수의 기간통일 : 2008/02 ~ 2019/09

## 모든 변수의 개략 Review





✓ 매매지수와 각 변수간 **CCF** (교차상관도표) 요약 : 각 설명변수의 매매지수에 대한 선행관계 위주로 살펴봄

P38

항 목	선행관계
한국은행 기준금리	1,2개월
통화량(M2)	
미국주택가격지수	
입주량	
가처분소득	9~11개월
담보대출	
매매거래량지수	1~2개월
전세가율	1개월



# 회귀모형

```
225 realprice2_stable_period_lm <- lm(realprice2_stable_period[,9] ~ realprice2_stable_period[,1] + realprice2_stable_period[,2] + realprice2_stable_period[,3] + realprice2_stable_period[,4] + realprice2_stable_period[,5] + realprice2_stable_period[,6] + realprice2_stable_period[,7] + realprice2_stable_period[,8])
226 summary(realprice2_stable_period_lm) # Adjusted R-squared: 0.228
227 library(car)
228 durbinWatsonTest(realprice2_stable_period_lm) # D-W Statistic 0.5728102
```

229:1 # 안정시계열로 재검토 : R Script

Console Terminal Jobs

D:/HousingPrice/

```
> summary(realprice2_stable_period_lm) # Adjusted R-squared: 0.228
```

Call:

```
lm(formula = realprice2_stable_period[, 9] ~ realprice2_stable_period[, 1] + realprice2_stable_period[, 2] + realprice2_stable_period[, 3] + realprice2_stable_period[, 4] + realprice2_stable_period[, 5] + realprice2_stable_period[, 6] + realprice2_stable_period[, 7] + realprice2_stable_period[, 8])
```

Residuals:

Min	1Q	Median	3Q	Max
-0.0110650	-0.0032872	-0.0007467	0.0031544	0.0252416

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.987e-03	4.392e-04	4.525	1.34e-05 ***
realprice2_stable_period[, 1]	7.771e-03	2.880e-03	2.698	0.00789 **
realprice2_stable_period[, 2]	4.576e-02	1.064e-01	0.430	0.66795
realprice2_stable_period[, 3]	-7.455e-03	1.811e-01	-0.041	0.96723
realprice2_stable_period[, 4]	-6.985e-08	1.440e-07	-0.485	0.62848
realprice2_stable_period[, 5]	1.808e-01	6.346e-02	2.849	0.00510 **
realprice2_stable_period[, 6]	2.465e-02	7.969e-02	0.309	0.75758
realprice2_stable_period[, 7]	1.685e-04	7.397e-04	0.228	0.82017
realprice2_stable_period[, 8]	-2.475e-01	4.538e-02	-5.453	2.38e-07 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005124 on 131 degrees of freedom  
Multiple R-squared: 0.2725, Adjusted R-squared: 0.228  
F-statistic: 6.133 on 8 and 131 DF, p-value: 1.047e-06

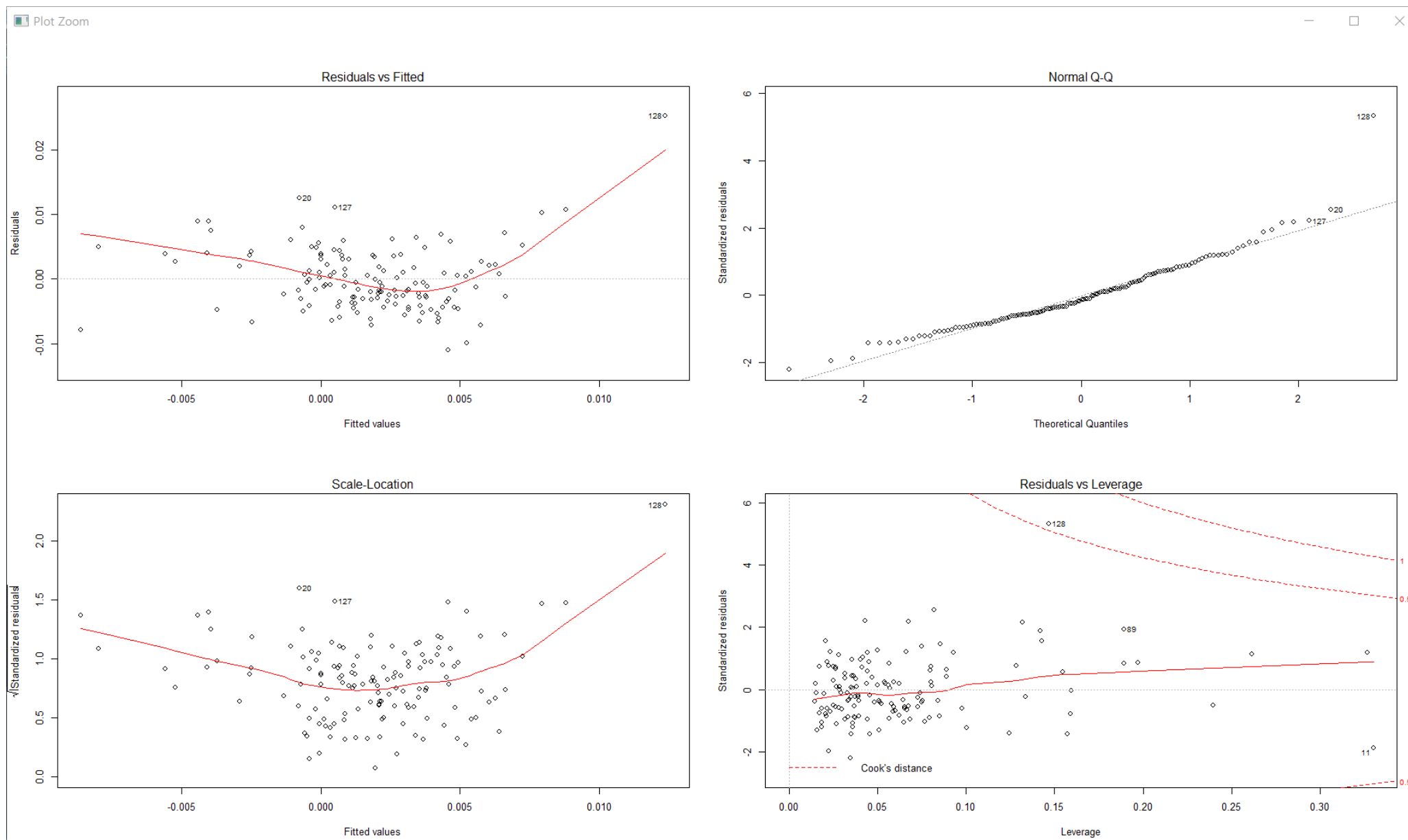
```
> durbinWatsonTest(realprice2_stable_period_lm) # D-W Statistic 0.5728102
```

lag	Autocorrelation	D-W Statistic	p-value
1	0.7019552	0.5728102	0

Alternative hypothesis: rho != 0

```
> |
```

- ✓ 최종 안정시계열의 매매지수를 종속변수로 그 외 8개 변수를 설명변수로 회귀모형 결과
  - 수정 결정계수 : 0.228
  - 기준금리, 가처분소득, 전세가율 유의미
- ✓ 오차의 자기상관 검진 : Durbin-Watson검정
  - 자기상관관계가 없으면 DW값이 2에 가까워짐
  - $0 < DW < 2$  : 양의 자기상관관계
  - $2 < DW < 4$  : 음의 자기상관관계
  - 검정 결과 : DW값(0.573)이 2보다 작고 유의확률값이 거의 0(자기상관관계가 없다는 귀무가설을 기각)이고, 자기상관계수가 0.7로 양의 1차 자기상관이 존재한다.



# 회귀모형

## ALT1) 자기시차변수를 추가

```
241 ##### ALT1. 종속변수에 자기시차변수를 추가/ 오차항에 1차 자기상관을 해결 #####
242 realprice2_stable_addlag <- ts.union(realprice2_stable, index_cost_apt_seoul_ts_stable_add
    lag=lag(index_cost_apt_seoul_ts_stable, -1))
243 realprice2_stable_addlag_period <- window(realprice2_stable_addlag, start=c(2008,2), end=c
    (2019,9))
244 realprice2_stable_addlag_period_lm <- lm(realprice2_stable_addlag_period[,9] ~ realprice2_
    stable_addlag_period[,1] + realprice2_stable_addlag_period[,2] + realprice2_stable_addlag_
    period[,3] + realprice2_stable_addlag_period[,4] + realprice2_stable_addlag_period[,5] +
    realprice2_stable_addlag_period[,6] + realprice2_stable_addlag_period[,7] + realprice2_st
    able_addlag_period[,8] + realprice2_stable_addlag_period[,10]) # 9가 종속변수다!
245 summary(realprice2_stable_addlag_period_lm) # Adjusted R-squared: 0.6379
246 durbinWatsonTest(realprice2_stable_addlag_period_lm) # D-W Statistic 1.919189 !!
247
```

246:81 ALT1. 종속변수에 자기시차변수를 추가/ 오차항에 1차 자기상관을 해결 : R Script

Console Terminal Jobs

D:/HousingPrice/

```
10])
Residuals:
    Min       1Q   Median       3Q      Max
-0.0100670 -0.0014368 -0.0001694  0.0010808  0.0234708

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.690e-04  3.255e-04   1.441  0.15194
realprice2_stable_addlag_period[, 1]  2.769e-03  2.014e-03   1.375  0.17155
realprice2_stable_addlag_period[, 2] -3.693e-03  7.300e-02  -0.051  0.95974
realprice2_stable_addlag_period[, 3] -2.052e-02  1.240e-01  -0.165  0.86887
realprice2_stable_addlag_period[, 4] -8.194e-08  9.864e-08  -0.831  0.40767
realprice2_stable_addlag_period[, 5]  5.868e-02  4.460e-02   1.316  0.19053
realprice2_stable_addlag_period[, 6]  3.735e-02  5.458e-02   0.684  0.49497
realprice2_stable_addlag_period[, 7]  2.629e-03  5.451e-04   4.822  3.91e-06 ***
realprice2_stable_addlag_period[, 8] -9.942e-02  3.336e-02  -2.980  0.00344 **
realprice2_stable_addlag_period[, 10] 7.676e-01  6.282e-02  12.219 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00351 on 130 degrees of freedom
Multiple R-squared:  0.6614,    Adjusted R-squared:  0.6379 
F-statistic: 28.21 on 9 and 130 DF,  p-value: < 2.2e-16

> durbinWatsonTest(realprice2_stable_addlag_period_lm) # D-W Statistic 2.130003 !!
lag Autocorrelation D-W Statistic p-value
1      0.03426501      1.919189 0.57
Alternative hypothesis: rho != 0
```

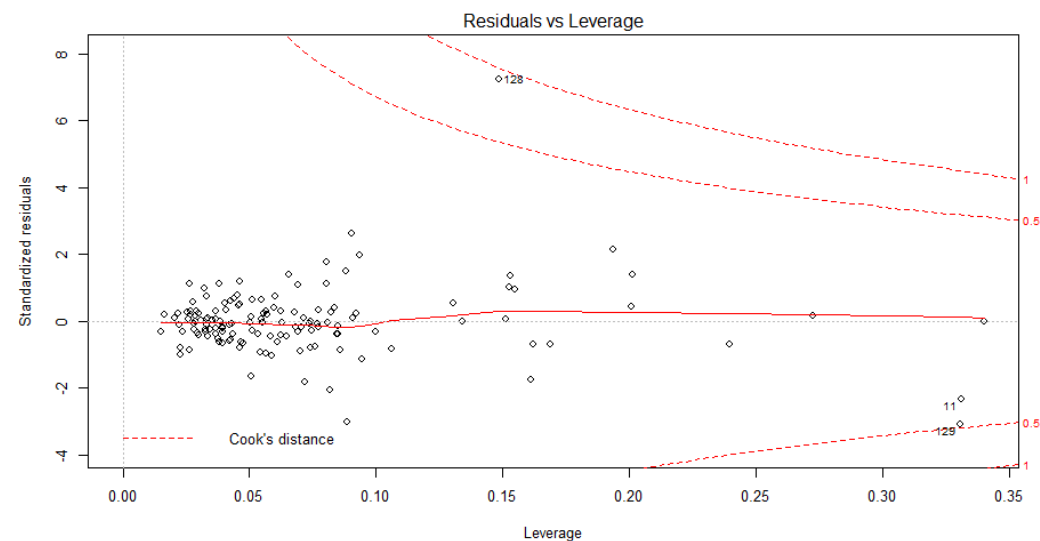
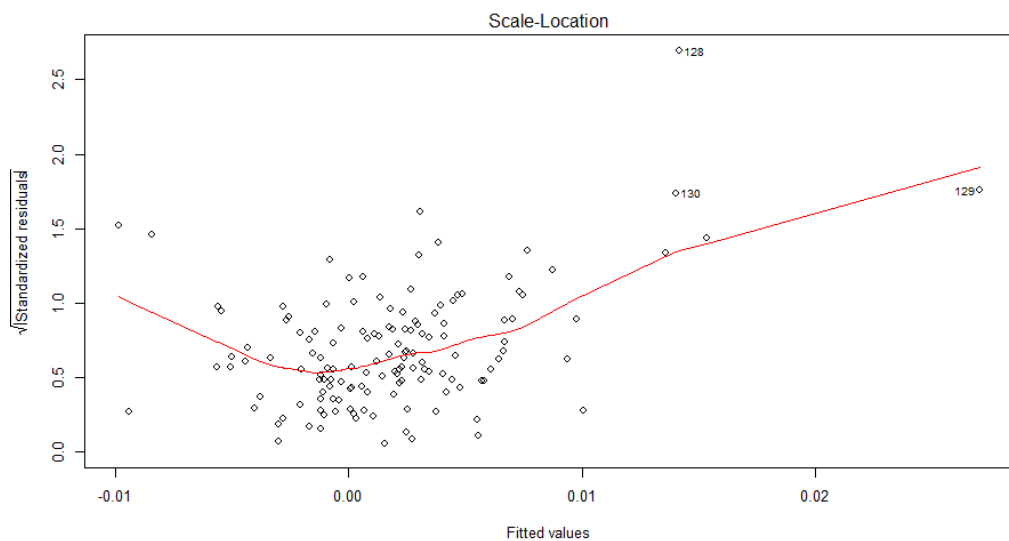
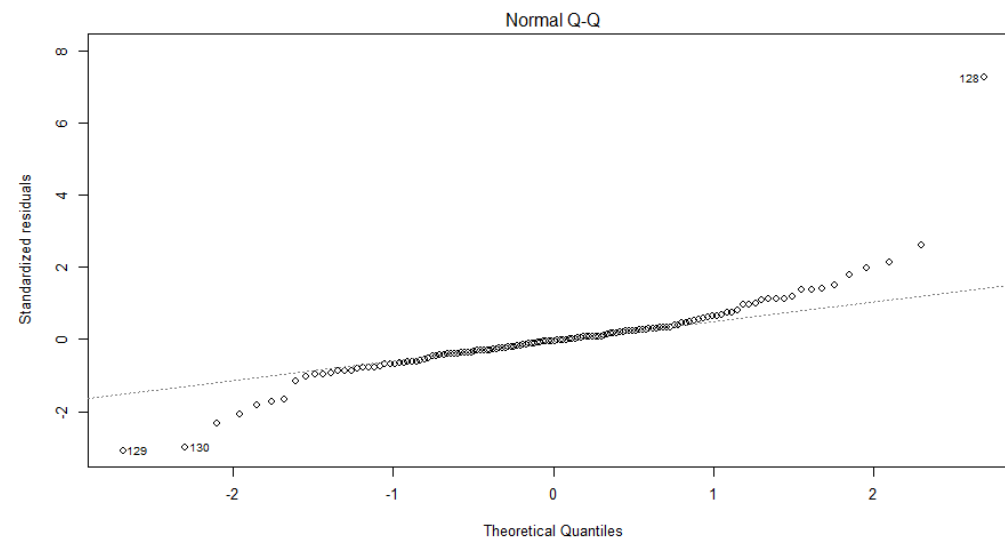
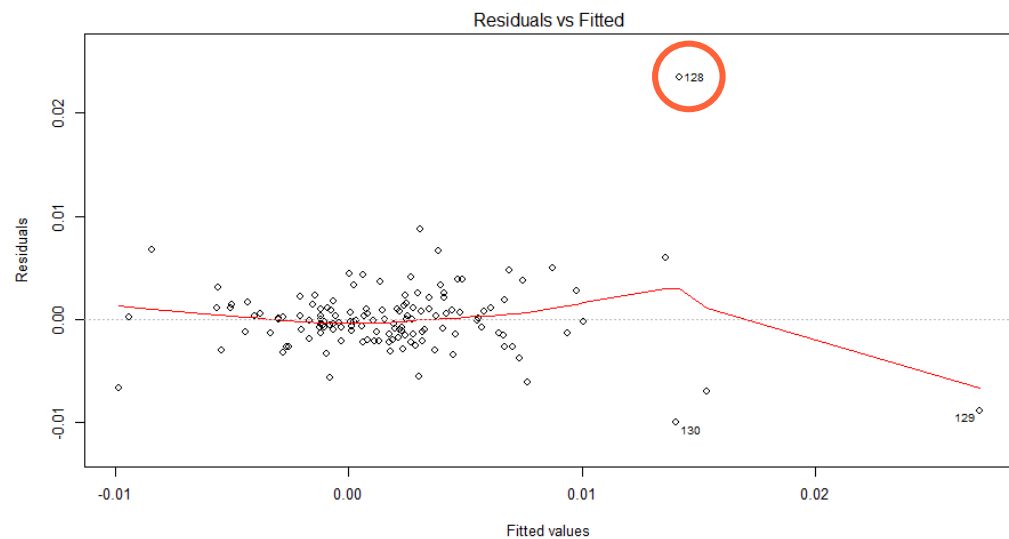
- ✓ 1차 자기상관 문제를 해결하기 위하여 매매지수의 1차 시차변수를 추가하여 모형 작성
  - 수정 결정계수 : 0.6379
  - 매매거래량지수, 전세가율, 1차 시차변수 유의미
- ✓ 오차의 자기상관 검진 : Durbin-Watson검정
  - 검정 결과 : DW값(1.92)이 2에 가깝게 나오고 유의확률값이 0.57로 자기상관관계가 없다는 귀무가설을 기각하지 않고, 자기상관계수가 0.034이다.
- ✓ 시차종속변수가 포함된 회귀모형의 경우, 독립변수와 잔차항이 서로 독립이라는 가정이 위배되어 더빈왓슨의 d통계량의 정확성이 떨어져서 h통계량을 사용해야 한다.

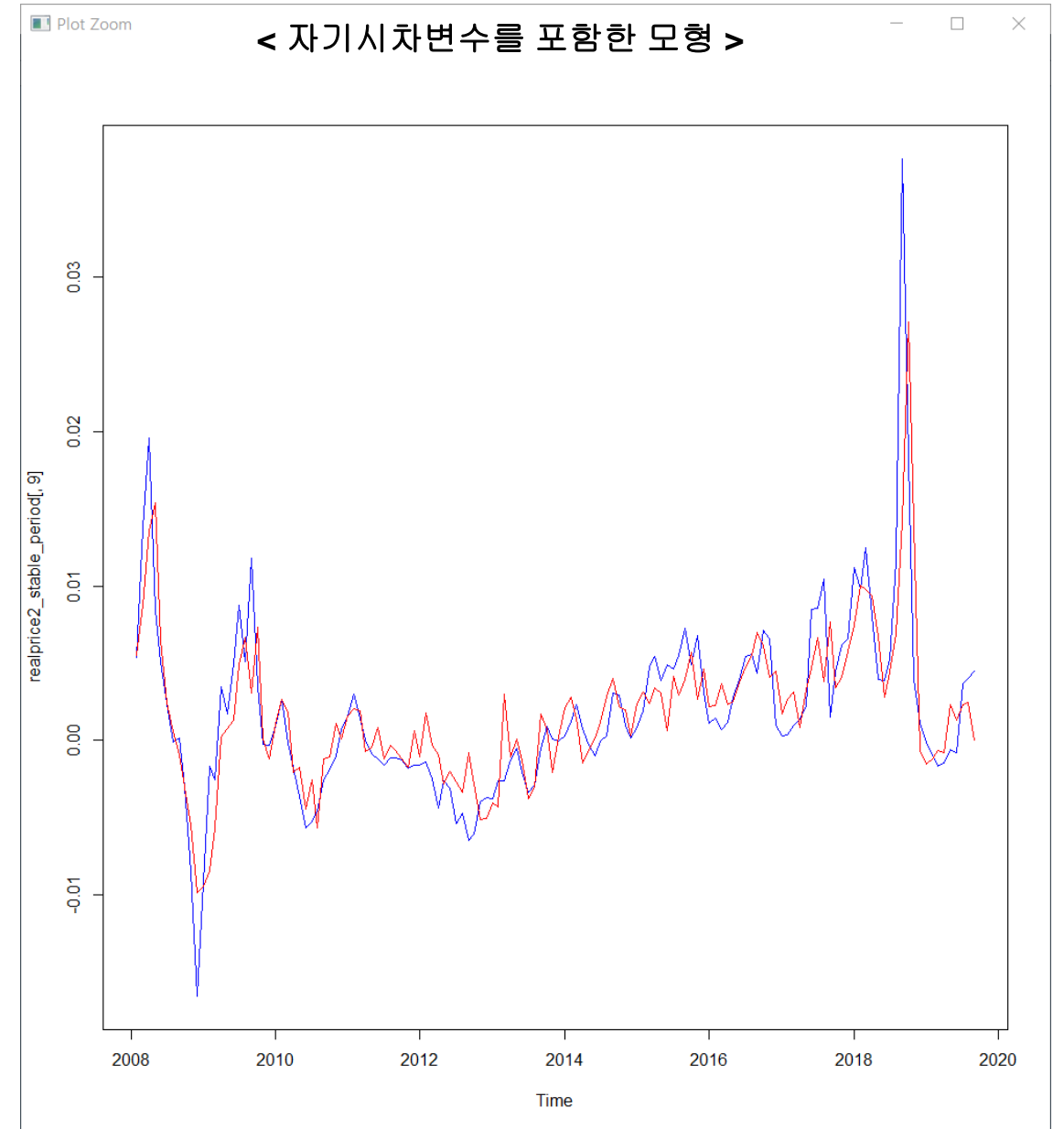
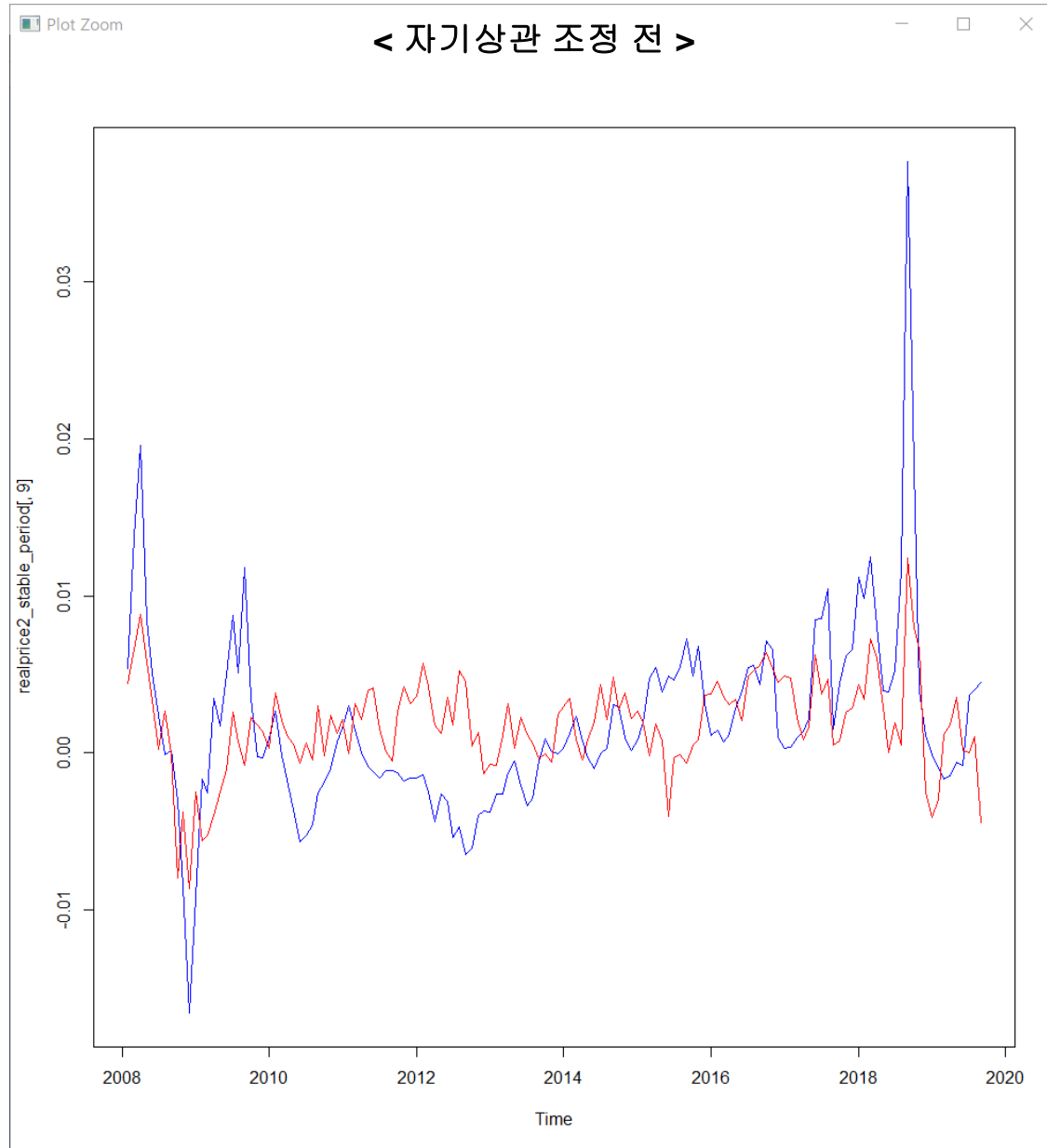
# 회귀모형

## 잔차 검토

Plot Zoom

— □ ×





# 회귀모형

## ALT2) 코크란-오컷 방법

```
254 - ### ALT2. 코크란 오컷 방법/ 오차항에 1차 자기상관을 해결 ###
255 library(orcutt)
256 realprice2_stable_period_lm_coch = cochrane.orcutt(realprice2_stable_pe
257 summary(realprice2_stable_period_lm_coch) # Adjusted R-squared: 0.2264
258 # Durbin-Watson statistic
259 # (original): 0.57281 , p-value: 8.978e-18
260 # (transformed): 1.66358 , p-value: 4.146e-0
261
262 plot(realprice2_stable_period[,9], col="blue")
263 lines(ts(predict(realprice2_stable_period_lm_coch), start=c(2008,2), fr
    ="red"))
264
254:44 ALT2. 코크란 오컷 방법/ 오차항에 1차 자기상관을 해결 :
```

```
> summary(realprice2_stable_period_lm_coch) # Adjusted R-squared: 0.2264
Call:
lm(formula = realprice2_stable_period[, 9] ~ realprice2_stable_period[,
  1] + realprice2_stable_period[, 2] + realprice2_stable_period[,
  3] + realprice2_stable_period[, 4] + realprice2_stable_period[,
  5] + realprice2_stable_period[, 6] + realprice2_stable_period[,
  7] + realprice2_stable_period[, 8])

(Intercept)                2.0387e-03  1.2739e-03   1.600  0.111933
realprice2_stable_period[, 1]  1.8135e-03  1.8062e-03   1.004  0.317245
realprice2_stable_period[, 2]  5.5957e-02  4.7200e-02   1.186  0.237971
realprice2_stable_period[, 3]  2.8298e-03  9.2687e-02   0.031  0.975691
realprice2_stable_period[, 4] -6.8306e-08  6.3386e-08  -1.078  0.283199
realprice2_stable_period[, 5]  1.4251e-01  5.2104e-02   2.735  0.007108 **
realprice2_stable_period[, 6] -4.2835e-03  3.9451e-02  -0.109  0.913705
realprice2_stable_period[, 7]  5.6886e-04  4.4395e-04   1.281  0.202352
realprice2_stable_period[, 8] -2.6406e-01  4.2840e-02  -6.164  8.289e-09 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

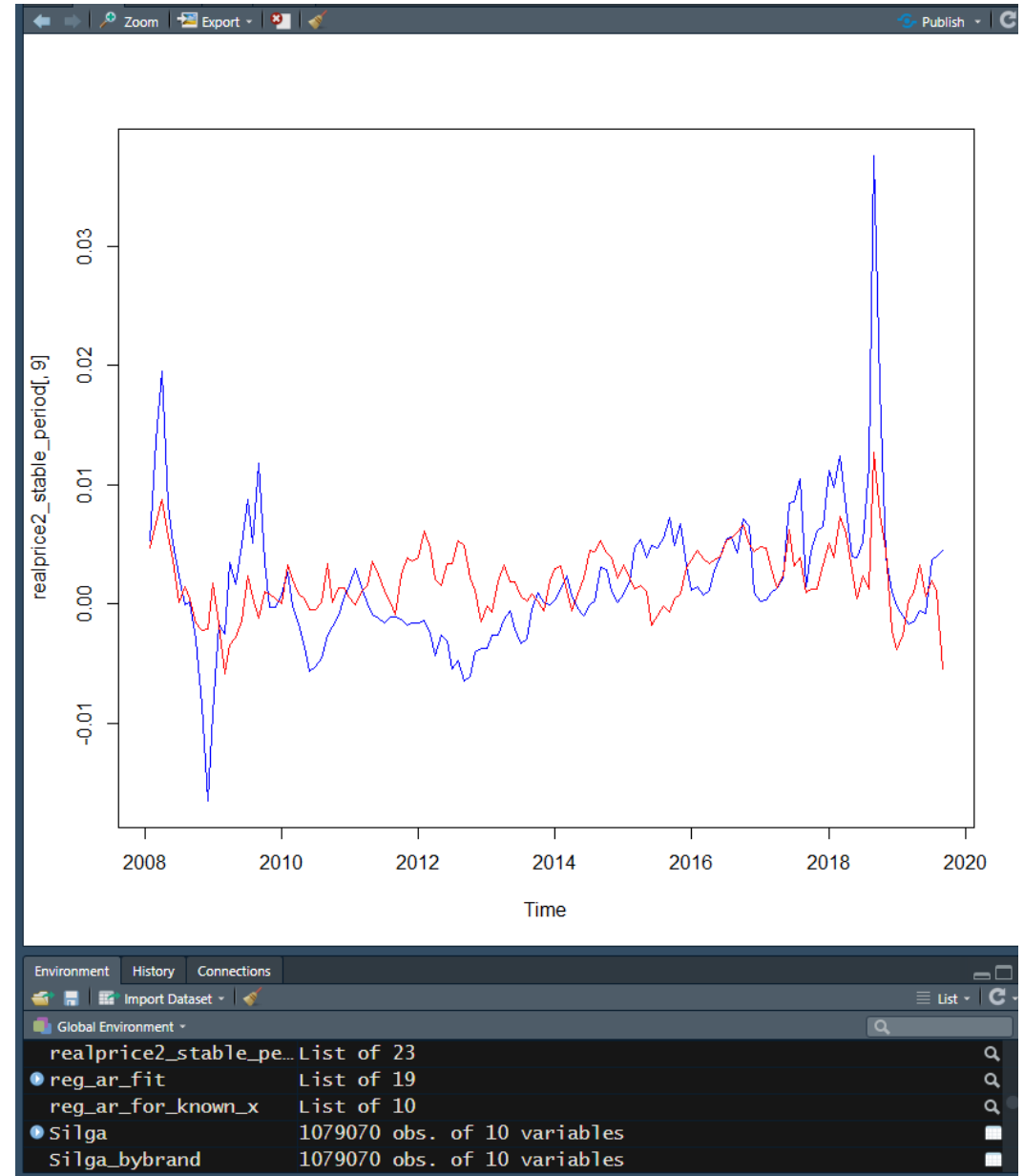
Residual standard error: 0.0034 on 130 degrees of freedom
Multiple R-squared:  0.2712 , Adjusted R-squared:  0.2264
F-statistic: 6 on 8 and 130 DF, p-value: < 1.327e-06

Durbin-Watson statistic
(original): 0.57281 , p-value: 8.978e-18
(transformed): 1.66358 , p-value: 4.146e-02
> plot(realprice2_stable_period[,9], col="blue")
> lines(ts(predict(realprice2_stable_period_lm_coch), start=c(2008,2), frequen
> |
```

✓  $R^2$ : 0.2264

✓ 당초 DW값:  
0.57281

✓ 변경 DW값:  
1.66358



```

276 realprice2_stable_addlag_period_ls_lm <- lm(realprice2_stable_addlag_period_ls[,9] ~
realprice2_stable_addlag_period_ls[,1] + realprice2_stable_addlag_period_ls[,2] +
realprice2_stable_addlag_period_ls[,3] + realprice2_stable_addlag_period_ls[,4] +
realprice2_stable_addlag_period_ls[,5] + realprice2_stable_addlag_period_ls[,6] +
realprice2_stable_addlag_period_ls[,7] + realprice2_stable_addlag_period_ls[,8] +
realprice2_stable_addlag_period_ls[,10])
277 summary(realprice2_stable_addlag_period_ls_lm) # Adjusted R-squared: 0.718
278 durbinWatsonTest(realprice2_stable_addlag_period_ls_lm) # D-W Statistic 2.075291 !!
279
280 ts.plot(realprice2_stable_period[,9], ts(predict(realprice2_stable_addlag_period_ls_lm),
start=c(2018,1), end=c(2019,9), frequency=12), col=c("blue", "red"))
279:1 learning set, test set으로 나누어 예측 / 자기시차변수 추가 :
R Script

```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.0074661	-0.0012289	-0.0000721	0.0012174	0.0072804

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.615e-04	2.427e-04	1.078	0.2836
realprice2_stable_addlag_period_ls[, 1]	2.884e-03	1.472e-03	1.959	0.0527
realprice2_stable_addlag_period_ls[, 2]	8.254e-02	5.921e-02	1.394	0.1662
realprice2_stable_addlag_period_ls[, 3]	1.891e-03	8.940e-02	0.021	0.9832
realprice2_stable_addlag_period_ls[, 4]	-9.808e-08	8.412e-08	-1.166	0.2461
realprice2_stable_addlag_period_ls[, 5]	4.150e-02	3.862e-02	1.074	0.2850
realprice2_stable_addlag_period_ls[, 6]	3.765e-02	4.010e-02	0.939	0.3498
realprice2_stable_addlag_period_ls[, 7]	2.046e-03	4.195e-04	4.877	3.69e-06 ***
realprice2_stable_addlag_period_ls[, 8]	-2.127e-02	2.837e-02	-0.750	0.4550
realprice2_stable_addlag_period_ls[, 10]	8.490e-01	5.770e-02	14.715	< 2e-16 ***

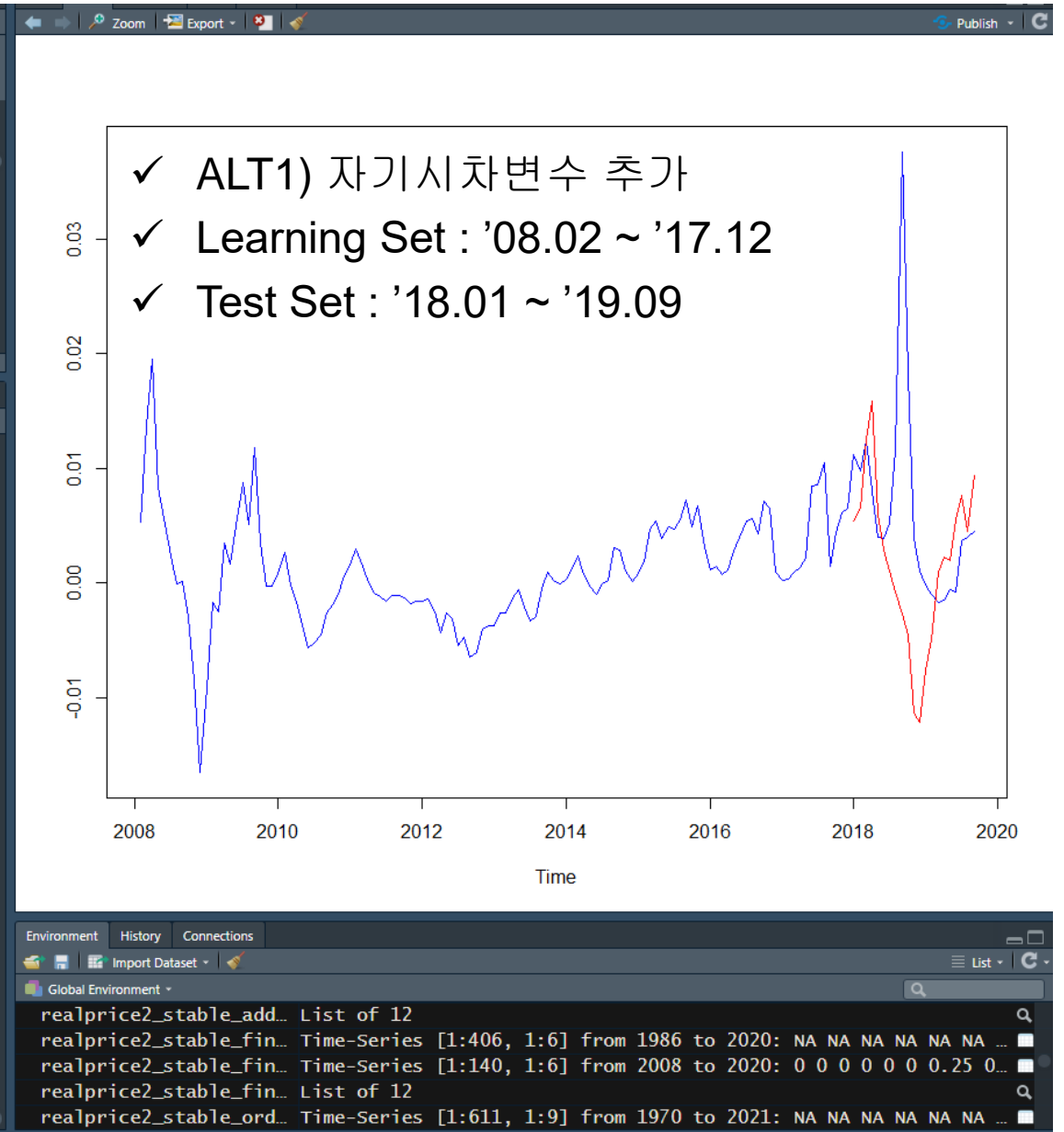
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.002497 on 109 degrees of freedom  
Multiple R-squared: 0.7396, Adjusted R-squared: 0.718  
F-statistic: 34.39 on 9 and 109 DF, p-value: < 2.2e-16

```

> durbinWatsonTest(realprice2_stable_addlag_period_ls_lm) # D-W Statistic 2.075291 !!
lag Autocorrelation D-W Statistic p-value
1 -0.03872911 2.075291 0.778
Alternative hypothesis: rho != 0
> ts.plot(realprice2_stable_period[,9], ts(predict(realprice2_stable_addlag_period_ls_lm),
start=c(2018,1), end=c(2019,9), frequency=12), col=c("blue", "red"))
>

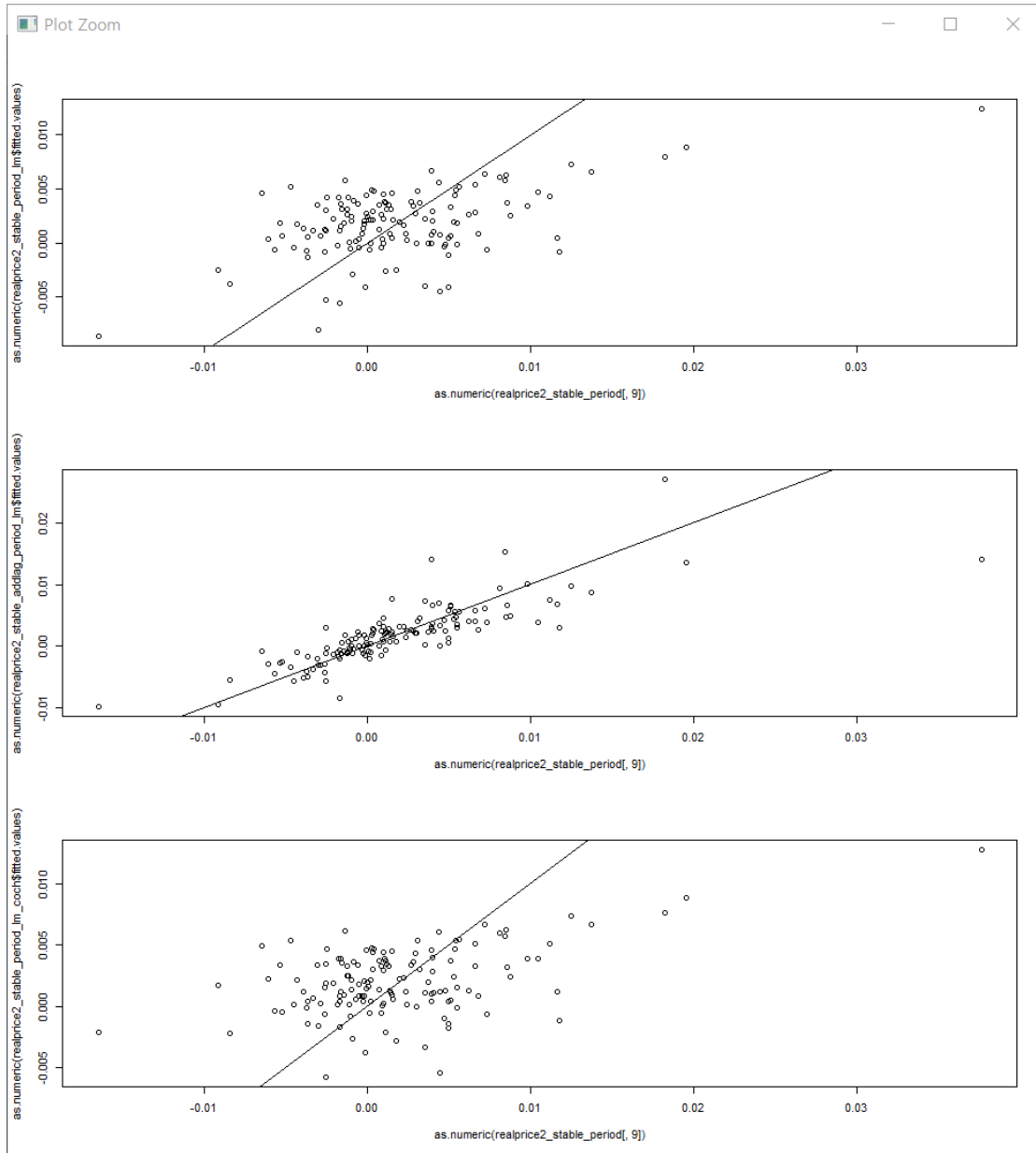
```





# 회귀모형

## 적합도 비교



- ✓ 설정된 3가지 모형에 대한 ‘원래 데이터와 적합 값의 산점도’를 보면 두 번째 자기시차 추가 모형의 적합도가 높은것을 볼 수 있다.
- ✓ RMSE값을 비교해 본 결과 두 번째 자기시차 추가 모형의 값이 낮은것을 알 수 있다.

```
266 # 예측정확도값(RMSE) 비교, 회귀/ 자기시차변수추가/ 코크란-오컷 방법
267 library(forecast)
268 accuracy(realprice2_stable_period_lm$fitted.values, realprice2_stable_addlag_period_lm$fitted.values, realprice2_stable_period_lm_coch$fitted.values[,1], realprice2_stable_period[, 9])
269
```

262:1 # ALT2, 코크란 오컷 방법/ 오차항에 1차 자기상관을 해결

Console Terminal x Jobs x

D:/HousingPrice/ ↗

```
> accuracy(realprice2_stable_period_lm$fitted.values, realprice2_stable_addlag_period_lm$fitted.values, realprice2_stable_period_lm_coch$fitted.values[,1], realprice2_stable_period[, 9])
```

	RMSE	MAE	MPE	MAP
당초) 회귀모형	0.004957027	0.003826808	172.244	386.213
ALT1) 자기시차변수 추가	0.003381854	0.002109862	49.59762	161.381
ALT2) 코크란-오컷방법	0.005092055	0.003813909	145.1551	350.935

# 회귀모형

## 최소한의 설명변수의 선택

### ✓ 설명변수를 선택하는 안 좋은 방법

- 설명변수와 종속변수간 그래프를 그려서 눈에 띄는 관계가 없을때 버리는 것
- 모든 설명변수에 대해 선형회귀 분석을 해서 p-값들이 0.05보다 큰 변수를 무시하는 것 (2개 이상의 설명변수가 서로 관련되어 있을때 p-값들이 오해를 불러일으킴)

### ✓ 설명변수를 선택하는 일반적인 방법

- 조정된  $R^2$ 를 최대화하는 것
- AIC(아카이케의 정보기준) 최소화
- AICc(수정된 아카이케의 정보기준) 최소화
- BIC(SC, 슈바르츠의 베이지안 정보기준) 최소화

→ 후진단계적 회귀를 이용

(BIC와 조정 $R^2$ 을 주로 참고함)

STEP. 1	CV	AIC	AICc	BIC	AdjR2	D-W
당초모형	0.0000154	- 1,571.0	- 1,569.0	- 1,538.7	0.638	1.919
del_1 (기준금리)	0.0000152	- 1,571.0	- 1,569.3	- 1,541.6	0.635	1.932
del_2 (통화량)	0.0000149	- 1,573.0	- 1,571.3	- 1,543.6	0.641	1.919
del_3 (미국주택가격지수)	0.0000152	- 1,573.0	- 1,571.3	- 1,543.6	0.641	1.923
del_4 (입주량)	0.0000154	- 1,572.3	- 1,570.6	- 1,542.9	0.639	1.926
del_5 (가처분소득)	0.0000154	- 1,571.2	- 1,569.5	- 1,541.7	0.636	1.947
del_6 (담보대출)	0.0000152	- 1,572.5	- 1,570.8	- 1,543.1	0.639	1.910
del_7 (매매거래량지수)	0.0000178	- 1,550.0	- 1,548.3	- 1,520.6	0.576	1.627
del_8 (전세가율)	0.0000152	- 1,563.8	- 1,562.1	- 1,534.3	0.616	2.093

STEP. 2	CV	AIC	AICc	BIC	AdjR2	D-W
del_23	0.0000147	- 1,575.0	- 1,573.6	- 1,548.5	0.643	1.922
del_231 (기준금리)	0.0000146	- 1,574.9	- 1,573.9	- 1,551.4	0.641	1.931
del_234 (입주량)	0.0000147	- 1,576.2	- 1,575.1	- 1,552.7	0.644	1.930
del_235 (가처분소득)	0.0000148	- 1,575.1	- 1,574.0	- 1,551.5	0.641	1.953
del_236 (담보대출)	0.0000146	- 1,576.5	- 1,575.4	- 1,553.0	0.645	1.912
del_237 (매매거래량지수)	0.0000171	- 1,553.9	- 1,552.8	- 1,530.3	0.582	1.623
del_238 (전세가율)	0.0000145	- 1,567.6	- 1,566.5	- 1,544.1	0.622	2.096

STEP. 3	CV	AIC	AICc	BIC	AdjR2	D-W
del_236	0.0000146	- 1,576.5	- 1,575.4	- 1,553.0	0.645	1.912
del_2361 (기준금리)	0.0000145	- 1,576.4	- 1,575.5	- 1,555.8	0.642	1.924
del_2364 (입주량)	0.0000145	- 1,577.9	- 1,577.1	- 1,557.3	0.646	1.921
del_2365 (가처분소득)	0.0000146	- 1,576.6	- 1,575.8	- 1,556.0	0.643	1.944
del_2367 (매매거래량지수)	0.0000170	- 1,554.6	- 1,553.7	- 1,534.0	0.582	1.594
del_2368 (전세가율)	0.0000144	- 1,568.6	- 1,567.8	- 1,548.0	0.622	2.092

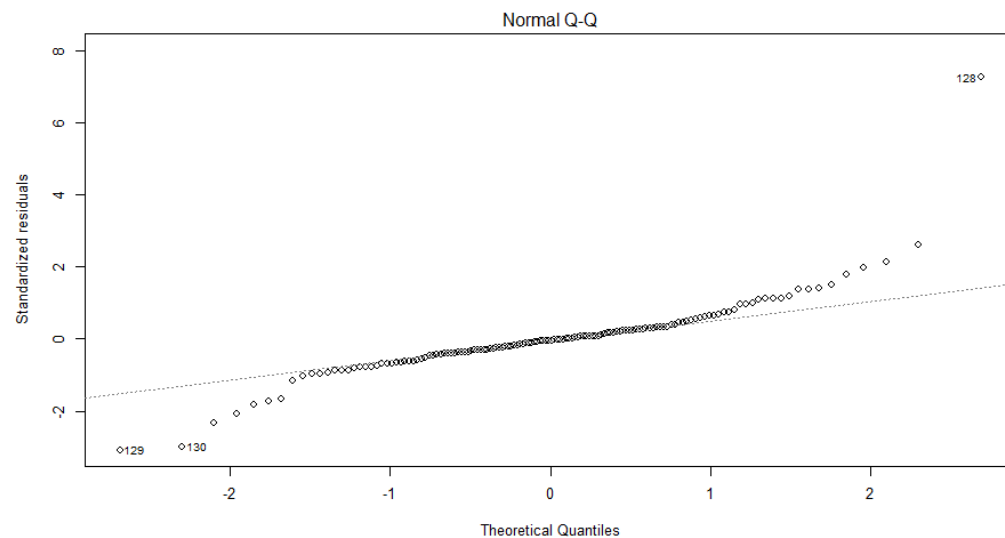
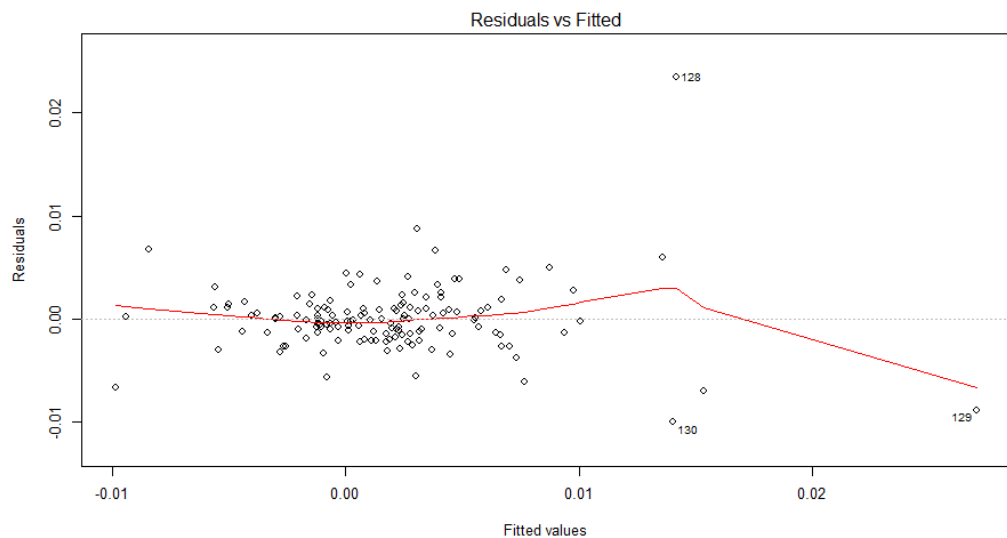
STEP. 4	CV	AIC	AICc	BIC	AdjR2	D-W
del_2364	0.0000145	- 1,577.9	- 1,577.1	- 1,557.3	0.646	1.921
del_23641 (기준금리)	0.0000144	- 1,578.0	- 1,577.4	- 1,560.4	0.644	1.930
del_23645 (가처분소득)	0.0000146	- 1,578.0	- 1,577.4	- 1,560.3	0.644	1.956
del_23647 (매매거래량지수)	0.0000170	- 1,555.5	- 1,554.9	- 1,537.8	0.582	1.602
del_23648 (전세가율)	0.0000143	- 1,570.3	- 1,569.7	- 1,552.6	0.624	2.096

\* 자기시차변수 추가 모형 기준

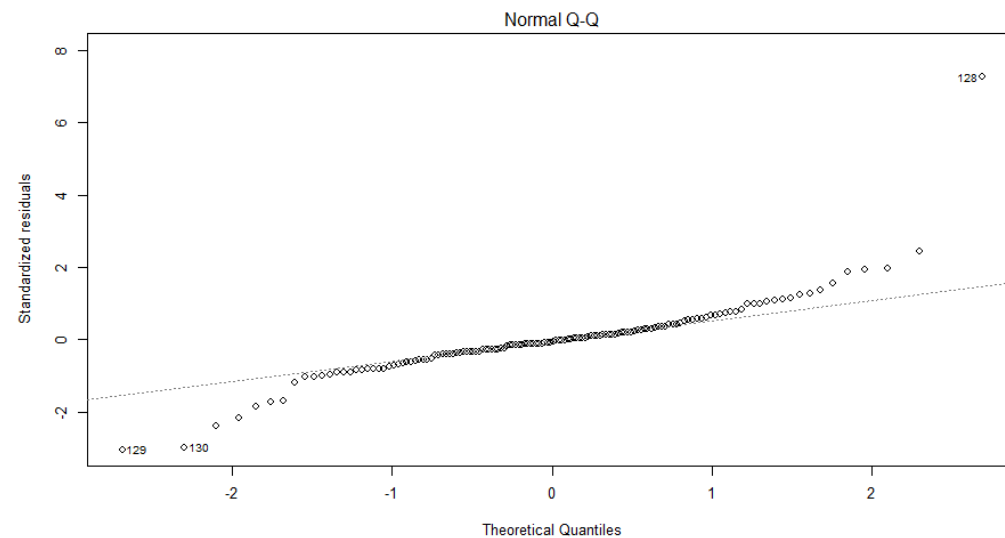
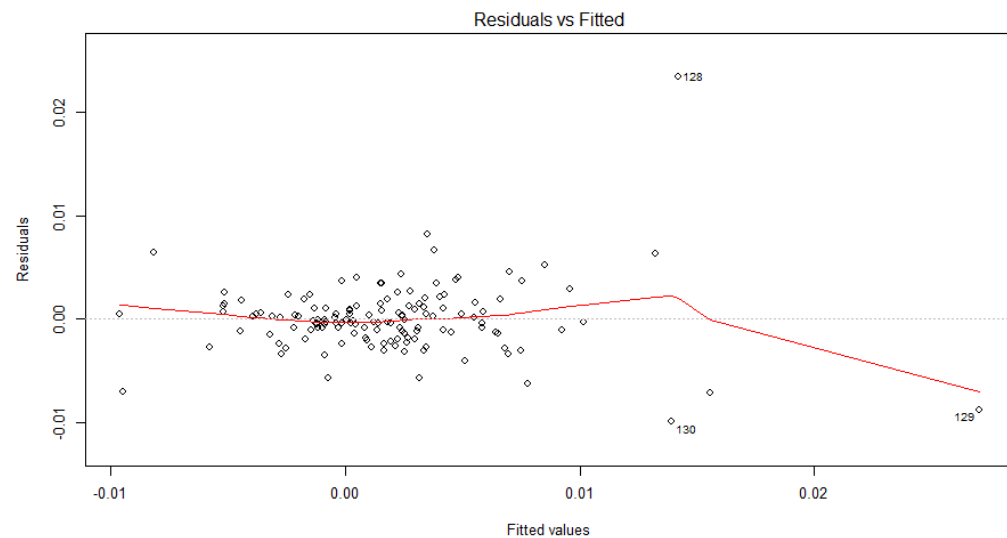
# 회귀모형

## 최소한의 설명변수의 선택

모든 설명변수  
(8개) 포함 모형  
(자기시차변수  
포함)

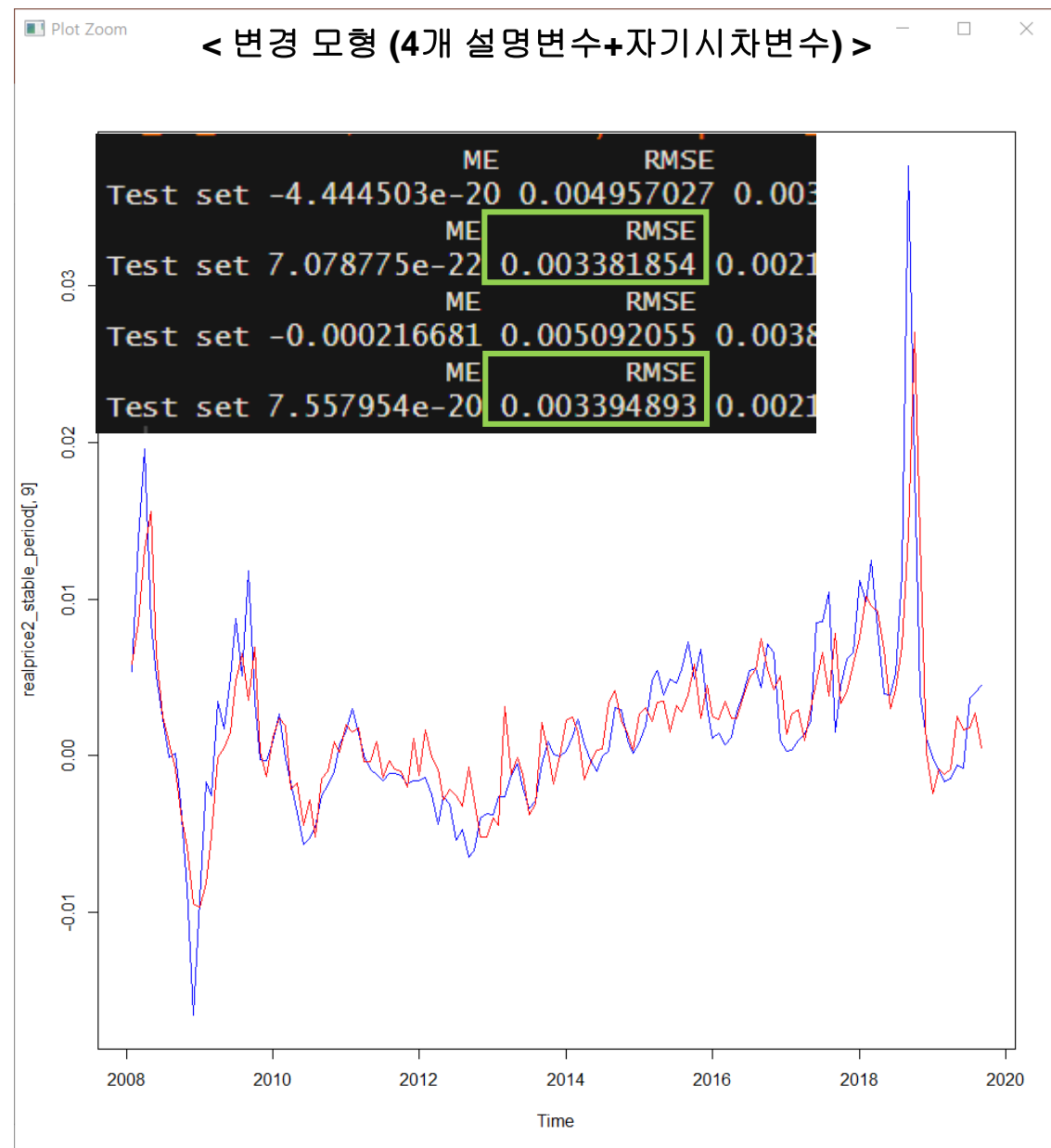
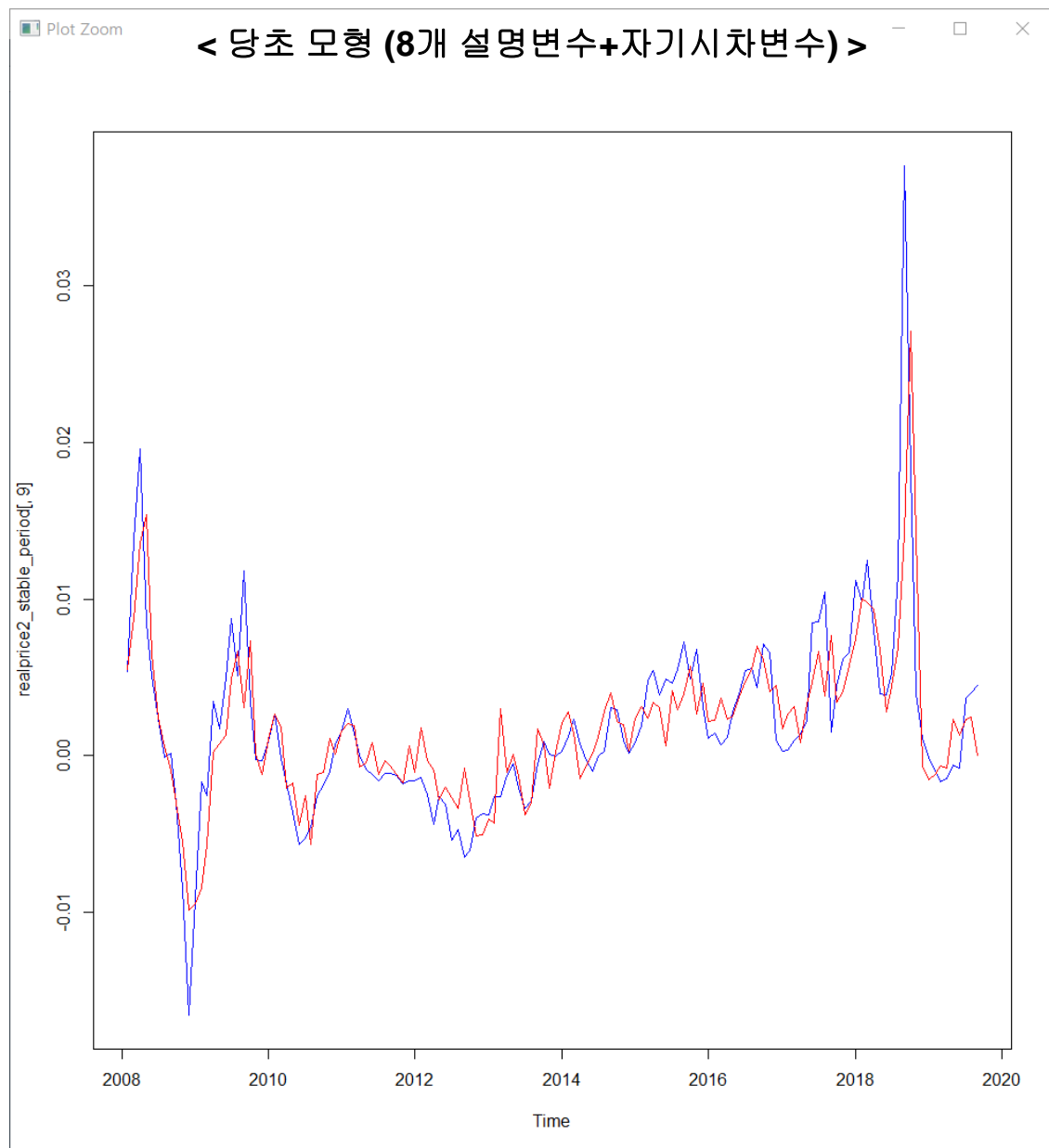


일부 설명변수  
(4개) 포함 모형  
(자기시차변수  
포함)



# 회귀모형

최소한의 설명변수의 선택/ 잔차 검토



- ✓ ARIMA 모형 같은 단일 시계열모형은 한 변수 자체의 특성만을 이용한다. 이를 해결하기 위해 **벡터자기회귀모형(Vector Autoregressive, VAR)**과 같이 시계열 분석의 특징과 회귀분석의 특징을 결합하여 변수 사이에 나타나는 상관계수와 인과관계를 추정할 수 있는 다변수 시계열모형이 사용된다.
- ✓ VAR모형은 **내생변수 및 외생변수를 구분하지 않고**, 계수값에 대한 제약을 고려하지 않으면서 **시계열이 나타내는 정보만을 이용하는 모형**으로, 이용 가능한 모든 시계열 사이의 관계를 이론적 제약 없이 추정함으로써 경제변수 사이에 나타나는 동태적인 관계를 분석하는 기법
- ✓ VAR모형을 이용하여 예측값을 추정하거나 충격반응분석 등을 시행하는 경우에는 VAR모형의 구성시계열이 안정적이 아니어도 무방하다.
- ✓ 그러나 인과관계분석 등 계수의 유의성 검증에서 전통적인 회귀분석 이론을 적용하려면 불안정시계열을 차분하여 VAR모형을 작성하거나 벡터 오차수정모형(VECM)을 이용해야 한다.
- ✓ 변수의 선정, 배열순서, 시차길이 등에 따라 변수 사이의 파급효과에 대한 분석결과가 달라진다. 또한 경제이론에 근거하지 않는다는 비판을 받는다. → **BVAR**모형과 구조적 VAR모형 사용

# VAR모형

## VAR모형 및 인과관계분석

```
library(vars)
VARselect(realprice2_stable_period[,1:9], lag.max=24, type="const")$["selection"]
# AIC(n) HQ(n) SC(n) FPE(n) → 시차 p=13를 고른다.
# 13 13 13 13
var13 <- VAR(realprice2_stable_period[,1:9], p=13, type="const")
serial.test(var13, lags.pt=18, type="PT.asymptotic") # 기각
var6 <- VAR(realprice2_stable_period[,1:9], p=6, type="const")
serial.test(var6, lags.pt=18, type="PT.asymptotic") # 자기상관이 없다는 귀무가설을
# 기각하지 않는 p의 최댓값 (6개월차)

realprice2_stable_period
causality(var6, cause="base_rate_ts_stable") # p-value = 0.0004432
causality(var6, cause="volume_m2_sa_ts_stable") # p-value = 0.1964
causality(var6, cause="cs_us_hpi_sa_ts_stable") # p-value = 0.2664
causality(var6, cause="volume_movein_seoul_agg_ts_stable") # p-value = 0.8913
causality(var6, cause="income_disposable_age_month_ts_stable") # p-value = 0.8421
causality(var6, cause="mortgage_loan_seoul_ts_stable") # p-value = 0.2558
causality(var6, cause="index_trade_seoul_ts_stable") # p-value = 0.03441
causality(var6, cause="index_cost_rent_ratio_apartment_seoul_ts_stable") # p-value = 0.2758
causality(var6, cause="index_cost_apartment_seoul_ts_stable") # p-value = 0.1839

grangertest(realprice2_stable_period[,1], realprice2_stable_period[,2], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,2], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,2], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,3], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,3], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,3], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,4], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,4], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,4], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,5], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,5], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,5], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,6], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,6], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,6], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,7], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,7], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,7], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,8], order=3)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,8], order=6)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,8], order=12)
grangertest(realprice2_stable_period[,1], realprice2_stable_period[,9], order=3)
```

- ✓ 적정시차값의 산정(**VARselect**) : BIC(SC)값이 최소인 13(월)을 기준
- ✓ 포트맨트검정을 이용하여 잔차에 상관관계가 없는지 확인(**serial.test**) : 13차시에서 귀무가설을 기각하여 시차를 변경해가면서 확인해본 결과 6차시에서 기각하지 않음
- ✓ 인과관계 검정을 위해서 각 변수별로 원인(cause)으로 두고 해보았으나 2개 초과인 다변수의 그랜저인과관계검정 결과의 해석 어려움

# VAR모형

## VAR모형 및 인과관계분석

- ✓ 모든 변수간의 인과관계 검정(grangertest) 실시 :  $9 \times 9 - 9 = 72$ 개
- ✓ lag는 3, 6, 12 에 대해서 검토 후 6으로 결정
- ✓ 유의확률 0.05 이내에서 p값의 순서를 따져서 배열순서를 결정함

lag=6, 아래(행)가 cause		1	2	3	4	5	6	7	8	9
1	baserate_ts_stable		0.02111	0.03302	0.17300	0.19250	0.06212	0.08200	0.40360	0.19950
2	volume_m2_sa_ts_stable	0.81020		0.77180	0.32740	0.96630	0.07055	0.40630	0.92270	0.51850
3	cs_us_hpi_sa_ts_stable	0.86440	0.27570		0.91570	0.20670	0.82080	0.01554	0.90710	0.87960
4	volume_movein_seoul_agg_ts_stable	0.70640	0.90850	0.95360		0.41840	0.74780	0.29210	0.42810	0.31950
5	income_disposable_age_month_ts_stable	0.30680	0.74190	0.84130	0.64310		0.46090	0.75220	0.97580	0.59450
6	mortgage_loan_seoul_ts_stable	0.58510	0.62520	0.02827	0.42430	0.97690		0.27580	0.93900	0.89730
7	index_trade_seoul_ts_stable	0.18030	0.60060	0.33430	0.25680	0.25970	0.05487		0.36400	0.00003
8	index_cost_rentratio_apr_seoul_ts_stable	0.22140	0.33540	0.27080	0.01097	0.02555	0.93350	0.78030		0.30780
9	index_cost_apr_seoul_ts_stable	0.04854	0.52580	0.16800	0.02449	0.22840	0.66280	0.00285	0.86230	

- ✓ 최종 배열순서 : 담보대출 → 기준금리 → 미국주택가격지수 → 통화량 → 매매거래량지수 → 전세가율 → 매매지수 → 입주량 → 가처분소득
  - VAR모형은 모든 변수가 내생변수으로써 모든 시계열 사이의 관계를 (경제적인) 이론의 제약 없이 추정하기 때문에 알고 있는 선후관계와 다른 모습을 보인다.



# VAR모형

## VAR모형(순서조정) 및 충격반응함수 결과

- ✓ 적정시차값 : 13(월)
- ✓ 포트맨트검정: 6차시에서 기각하지 않음
- ✓ 충격반응함수(irf) 결과 아래와 같이 9개 변수 전부 입주량에만 영향을 미침

```
590 library(vars)
591 VARselect(realprice2_stable_order_period[,1:9], lag.max=24, type="const")["selection"]
592 # AIC(n) HQ(n) SC(n) FPE(n) → 시차 p=13를 고른다.
593 # 13 13 13 13
594 var_order_13 <- VAR(realprice2_stable_order_period[,1:9], p=13, type="const")
595 serial.test(var_order_13, lags.pt=18, type="PT.asymptotic") # 기각
596 var_order_6 <- VAR(realprice2_stable_order_period[,1:9], p=6, type="const")
597 serial.test(var_order_6, lags.pt=18, type="PT.asymptotic") # 포트맨토(portmanteau) 검정을
# 이용하여 잔차(residual)에 상관관계(correlation)가 없는지 확인 (6개월차)
598
599 var_order_6_irf <- irf(var_order_6, n.ahead=12, ortho=T, boot=F)
600 plot(var_order_6_irf) # 전반적으로 서로간의 영향력이 안보임, 통화량만 영향을 받고 있음
601 # from mortgage 입주량에 지속적으로 영향
602 # from baserate 입주량에 지속적으로 영향
603 # from us_hpi 입주량에 지속적으로 영향
604 # from M2 입주량에 지속적으로 영향
605 # from trade 입주량에 지속적으로 영향
606 # from reentratio 입주량에 지속적으로 영향
607 # from cost index 입주량에 지속적으로 영향
608 # from disposable income 입주량에 지속적으로 영향
609 var_order_6_fevd <- fevd(var_order_6, n.ahead=12, ortho=TRUE, boot=FALSE)
610 plot(var_order_6_fevd)
611 var_order_6_pred <- predict(var_order_6, n.ahead=12, ci=0.95)
612 plot(var_order_6_pred)
```

611:3 # learning set, test set으로 나누어 예측 / 자기시차변수 추가

R Script

Console Terminal Jobs

D:/HousingPrice/

```
AIC(n) HQ(n) SC(n) FPE(n)
13 13 13 13
> serial.test(var_order_13, lags.pt=18, type="PT.asymptotic") # 기각

Portmanteau Test (asymptotic)

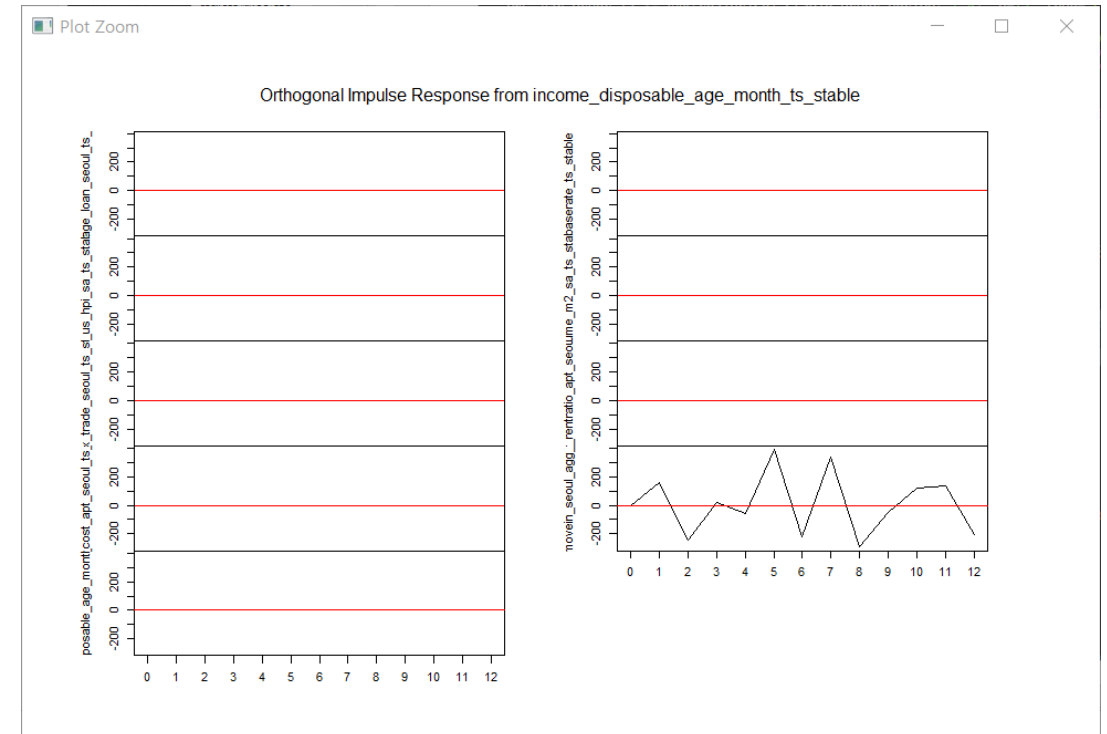
data: Residuals of VAR object var_order_13
Chi-squared = 2910.1, df = 405, p-value < 2.2e-16

> serial.test(var_order_6, lags.pt=18, type="PT.asymptotic") # 포트맨토(portmanteau) 검정을 이용하여
잔차(residual)에 상관관계(correlation)가 없는지 확인 (6개월차)

Portmanteau Test (asymptotic)

data: Residuals of VAR object var_order_6
Chi-squared = 988.71, df = 972, p-value = 0.3476

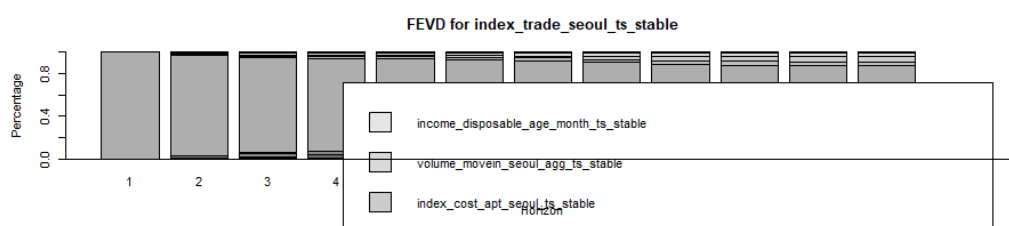
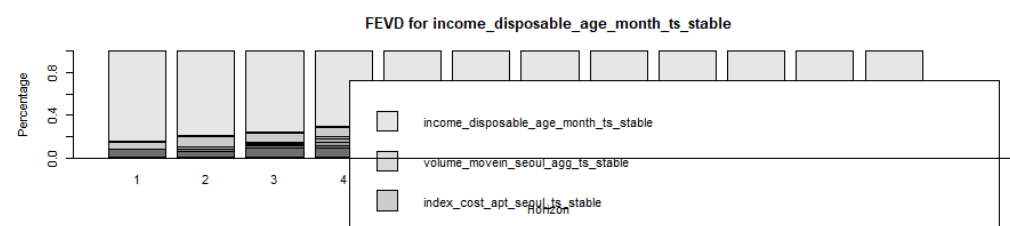
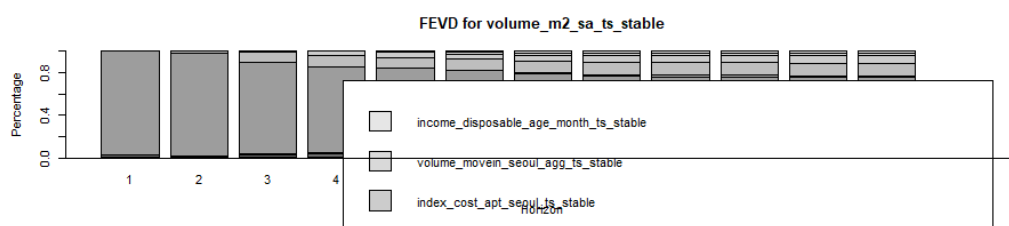
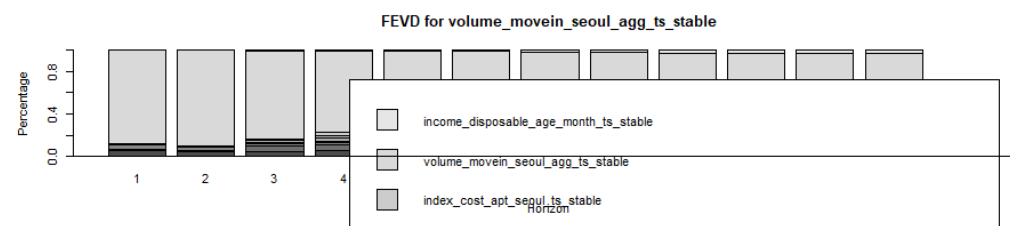
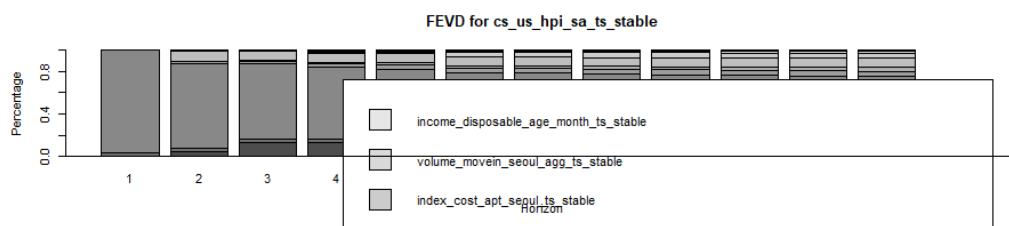
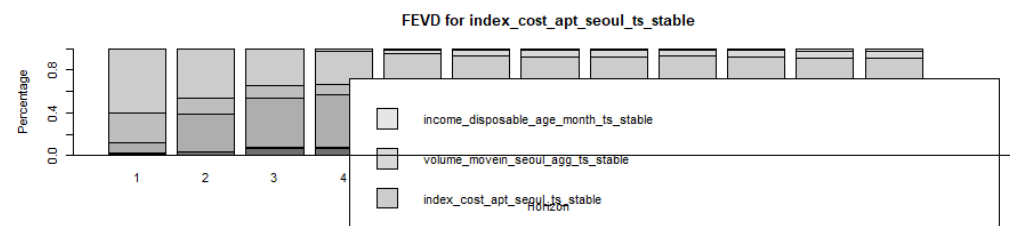
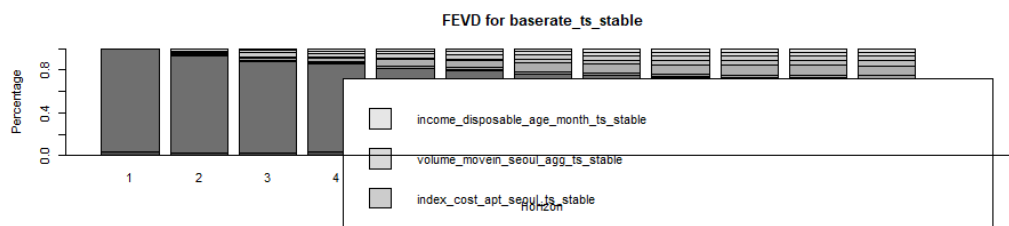
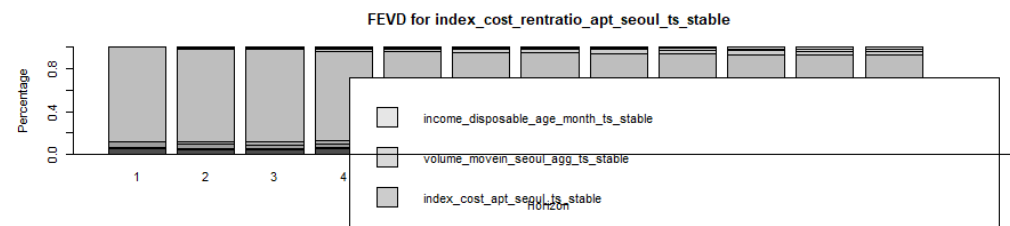
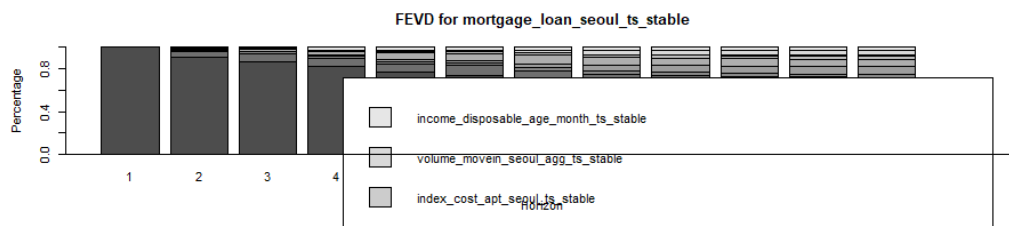
> |
```

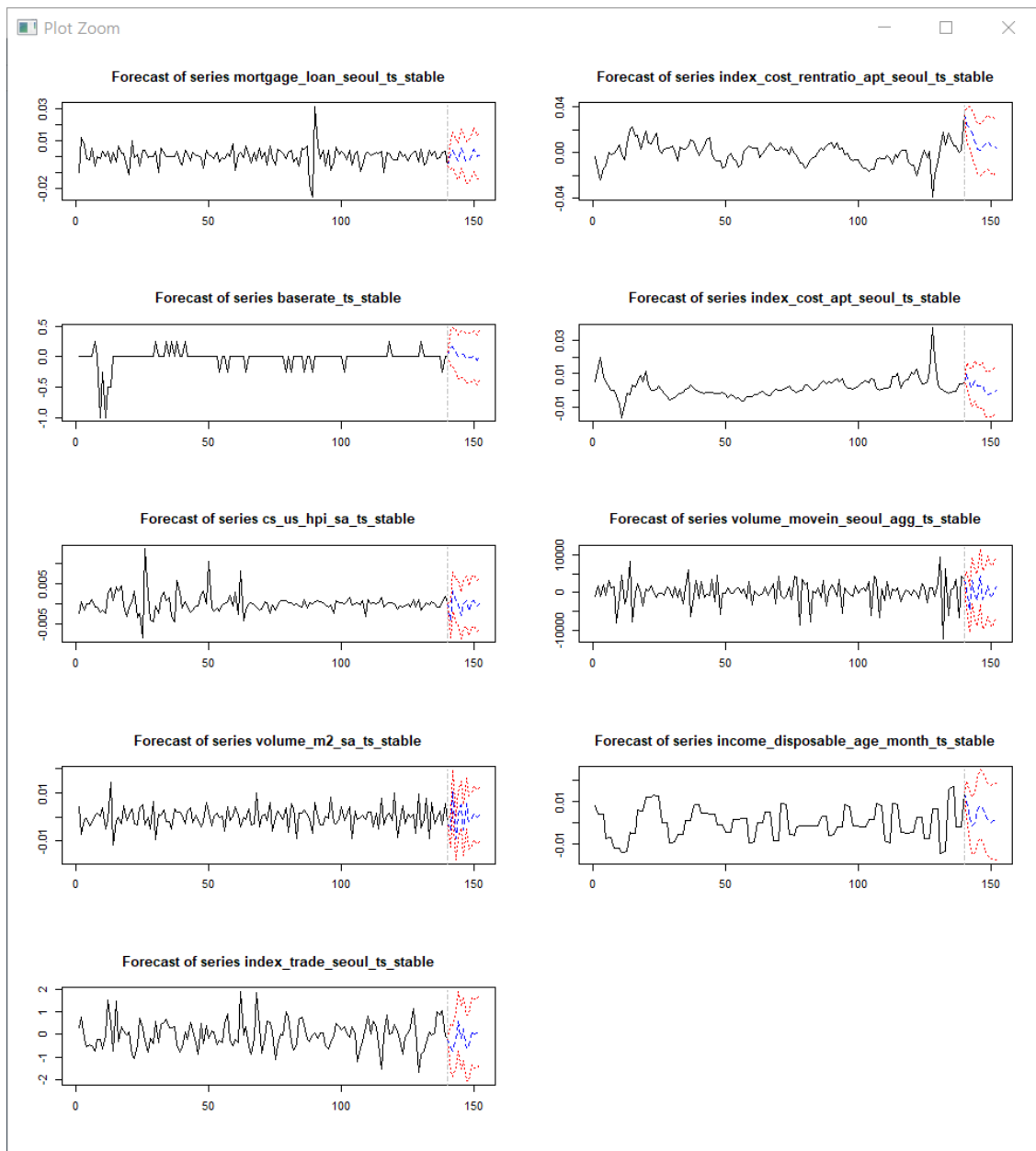




## VAR모형

## 예측오차 분산분해

 Plot Zoom



- ✓ 9개 모든 변수에 대해서 각 변수를 종속변수로 하여 6차시에 대해서 VAR회귀모형이 만들어졌고 회귀모형에서 종속변수로 선정했던 매매지수의 모형만 살펴봄
- ✓ 변수들의 단위 및 크기가 상이하여 각 변수별로 어떤 차수에서 큰 값을 가지는지 살펴봄
- ✓ 매매지수의 VAR모형

변수(순서대로)	1차	2차	3차	4차	5차	6차
담보대출	0.0524	(0.0184)	0.0103	0.0883	(0.0180)	0.0439
기준금리	0.0032	0.0041	(0.0050)	(0.0058)	0.0055	(0.0017)
미국주택가격지수	(0.1668)	0.2149	(0.1423)	(0.0486)	(0.0892)	0.0910
통화량	(0.0579)	0.1263	(0.0817)	0.0035	0.0018	(0.0036)
매매거래량지수	0.0034	0.0013	0.0008	0.0010	0.0005	(0.0002)
전세가율	0.1236	(0.2236)	0.2479	(0.0364)	(0.0479)	(0.0553)
매매지수	0.7676	(0.2493)	0.3365	(0.0749)	0.2837	(0.1245)
입주량	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
가처분소득	(0.0647)	0.1353	(0.2020)	0.1145	0.0341	(0.0096)

\* 상수항 0.0001

- ✓ 가성적 회귀 현상 : 불안정시계열을 포함하는 모형에 대하여 회귀분석을 시행하는 경우, 실제로는 상관관계가 없는데도 불구하고, 외견상 회귀모형의 유의성이 높은 것처럼 보이는 경우가 많다. 이 때 결정계수  $R^2$  값은 매우 높으나 더빈-왓슨(DW) 통계량은 매우 낮게 추정되어 실제로는 별 의미가 없는 현상을 말함 (다음 페이지 참조)
- ✓ 어떤 시계열이 d번의 차분을 통하여 안정적인 계열이 되는 경우, 이런 형태의 불안정시계열을 적분계열이라 하고  $Y_t \sim I(d)$ 로 표기
- ✓ 불안정시계열에 대해서도 회귀분석을 비롯한 전통적 계량이론을 적용할 수 있는 특별한 경우가 나타날 수 있는데, 불안정시계열 사이에 안정시계열을 생성하는 선형결합이 존재하는 경우가 이에 해당함
- ✓ 이 경우 이 시계열들에 나타나는 적분계열의 특징이 공통적인 추세를 공유한다는 의미에서 공적분(cointegration) 관계에 있다고 한다.
- ✓ 공적분의 존재는 일련의 경제변수가 **단기적으로는 상호 괴리**를 보이지만 **장기적으로는 일정한 관계를 유지함**을 의미한다.
- ✓ 공적분검정 : 회귀모형을 추정하고 단위근검정법을 적용하여 이 회귀식의 잔차항이 안정시계열을 이루는가를 분석 (공적분 관계가 존재), 요한슨의 최우도 검정법
- ✓ 오차수정모형 : 장기적 균형관계에 대한 정보와 함께 단기적 움직임도 동시에 파악

# 공적분분석

## 가성적 회귀현상

```
659 # 증감률이 아닌 원지수 그대로 회귀를 돌려보자
660 realprice2_period_lm <- lm(realprice2_period[,9] ~ realprice2_period[,1] + realprice2_per
  id[,2] + realprice2_period[,3] + realprice2_period[,4] + realprice2_period[,5] +
  realprice2_period[,6] + realprice2_period[,7] + realprice2_period[,8])
661 summary(realprice2_period_lm)
662 library(car)
663 durbinWatsonTest(realprice2_period_lm)
```

661:20 안정시계열 중 선별하여 재검토 R Script

```
> summary(realprice2_period_lm)
```

Call:  
lm(formula = realprice2\_period[, 9] ~ realprice2\_period[, 1] +  
 realprice2\_period[, 2] + realprice2\_period[, 3] + realprice2\_period[,  
 4] + realprice2\_period[, 5] + realprice2\_period[, 6] + realprice2\_period[,  
 7] + realprice2\_period[, 8])

Residuals:

Min	1Q	Median	3Q	Max
-1.7788	-0.7941	-0.2882	0.8001	2.4793

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.800e+01	2.262e+00	25.638	< 2e-16 ***
realprice2_period[, 1]	7.841e-01	1.645e-01	4.767	4.83e-06 ***
realprice2_period[, 2]	9.285e-06	2.047e-06	4.537	1.26e-05 ***
realprice2_period[, 3]	6.605e-02	9.506e-03	6.948	1.50e-10 ***
realprice2_period[, 4]	2.150e-05	3.906e-05	0.550	0.582929
realprice2_period[, 5]	-3.996e-06	1.130e-06	-3.537	0.000557 ***
realprice2_period[, 6]	2.420e-04	2.357e-05	10.267	< 2e-16 ***
realprice2_period[, 7]	-6.663e-03	7.886e-03	-0.845	0.399678
realprice2_period[, 8]	-4.641e-01	1.982e-02	-23.411	< 2e-16 ***

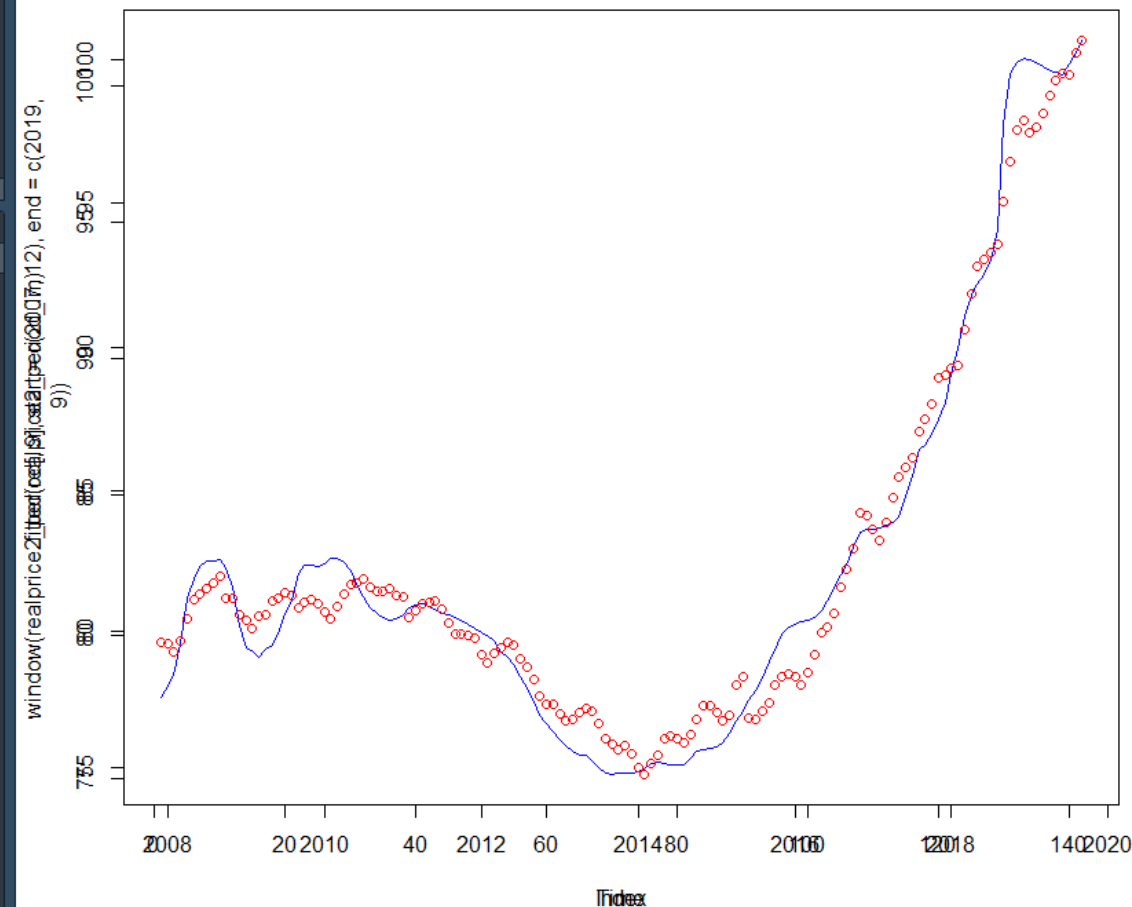
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.046 on 133 degrees of freedom  
Multiple R-squared: 0.9776, Adjusted R-squared: 0.9763  
F-statistic: 725.5 on 8 and 133 DF, p-value: < 2.2e-16

```
> durbinWatsonTest(realprice2_period_lm)
```

lag	Autocorrelation	D-W Statistic	p-value
1	0.8747347	0.2219909	0

Alternative hypothesis: rho != 0



Environment	History	Connections
Global Environment		
var2_1	List of 10	
var2_2	List of 10	
var2_3	List of 10	
var3	List of 10	

- ✓ 9개 변수 전체에 대해서 공적분모형 함수가 작동하지 않아(5개 변수 이상 안됨) 유의성이 높은 4개 변수(담보대출, 매매거래량, 전세가율, 매매지수)에 대해서만 검토함

- ✓ 요한슨 공적분검정 : 시차항12, 상수항(const) 포함, 'trace'는 우도비검정통계량을 도출  
변수 사이에 공적분 관계가 없다는 귀무가설은  $r=0$  에서 1% 유의수준의 임계치보다 커서 기각되고,  $r \leq 1$  의 가설은 10% 수준에서도 채택되는 것으로 나타나 4개의 변수 사이에 한 개의 공적분 관계가 존재하는 것으로 분석됨

```
681 ##### 원지수로 공적분검정을 실시 #####
682 library(vars)
683 str(realprice2_period)
684 vecm <- ca.jo(realprice2_period[,c(6:9)], type="trace", ecdet="const", K=12)
685 summary(vecm)

709:1 원지수로 공적분검정을 실시

> summary(vecm)

#####
# Johansen-Procedure #
#####

Test type: trace statistic , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1] 2.511674e-01 1.364894e-01 5.060143e-02 3.715164e-02 -2.501738e-16

Values of teststatistic and critical values of test:

      test 10pct  5pct  1pct
r <= 3 |  4.92  7.52  9.24 12.97
r <= 2 | 11.67 17.85 19.96 24.60
r <= 1 | 30.75 32.00 34.91 41.07 ←
r = 0  | 68.35 49.65 53.12 60.16

Eigenvectors, normalised to first column:
(These are the cointegration relations)
```

# 공적분분석

## 벡터 오차수정모형 및 예측

```
> var_vecm <- vec2var(vecm, r=1)
> var_vecm
```

Coefficient matrix of lagged endogenous variables:

A1:

	mortgage_loan_seoul_ts.11	index_trade_seoul_ts.11
mortgage_loan_seoul_ts	1.402417e+00	30.00714164
index_trade_seoul_ts	3.826362e-05	1.14175066
index_cost_renratio_apt_seoul_ts	1.633354e-05	-0.02102221
index_cost_apt_seoul_ts	2.637925e-05	0.02904359

A2:

	index_cost_renratio_apt_seoul_ts.11	index_cost_apt_seoul_ts.11
mortgage_loan_seoul_ts	508.0611872	170.2987427
index_trade_seoul_ts	1.7948341	-4.0873553
index_cost_renratio_apt_seoul_ts	1.7894370	0.2439521
index_cost_apt_seoul_ts	-0.0251375	1.4398938

A3:

	mortgage_loan_seoul_ts.12	index_trade_seoul_ts.12
mortgage_loan_seoul_ts	-5.071893e-01	-20.03527487
index_trade_seoul_ts	-2.207316e-04	-0.91971876
index_cost_renratio_apt_seoul_ts	-1.778715e-05	0.01893503
index_cost_apt_seoul_ts	4.389521e-05	-0.02090241

A4:

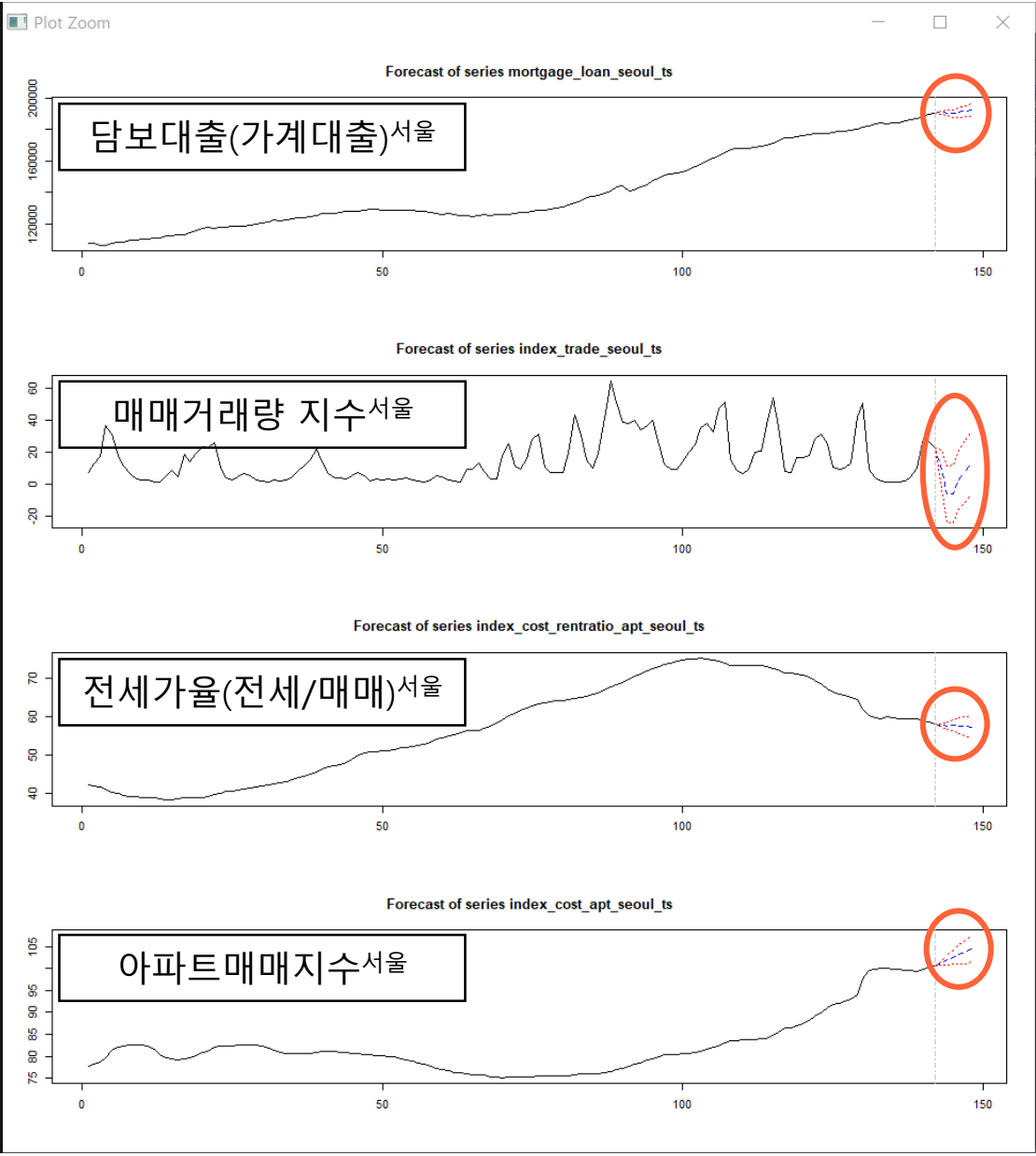
	index_cost_renratio_apt_seoul_ts.12	index_cost_apt_seoul_ts.12
mortgage_loan_seoul_ts	-1.060519e+03	170.0560046
index_trade_seoul_ts	-1.924101e+00	15.0674465
index_cost_renratio_apt_seoul_ts	-7.431004e-01	-0.5069659
index_cost_apt_seoul_ts	-5.969361e-02	-0.3698349

A5:

	mortgage_loan_seoul_ts.13	index_trade_seoul_ts.13
mortgage_loan_seoul_ts	2.843813e-01	-9.16755327
index_trade_seoul_ts	2.155249e-03	0.43101383
index_cost_renratio_apt_seoul_ts	-1.138323e-05	-0.01033526
index_cost_apt_seoul_ts	-1.294816e-04	0.01114591

A6:

	index_cost_renratio_apt_seoul_ts.13	index_cost_apt_seoul_ts.13
mortgage_loan_seoul_ts	864.76003645	-206.8263672
index_trade_seoul_ts	6.73013911	-15.7183986
index_cost_renratio_apt_seoul_ts	-0.01499636	0.4859875
index_cost_apt_seoul_ts	-0.06847585	-0.3973471

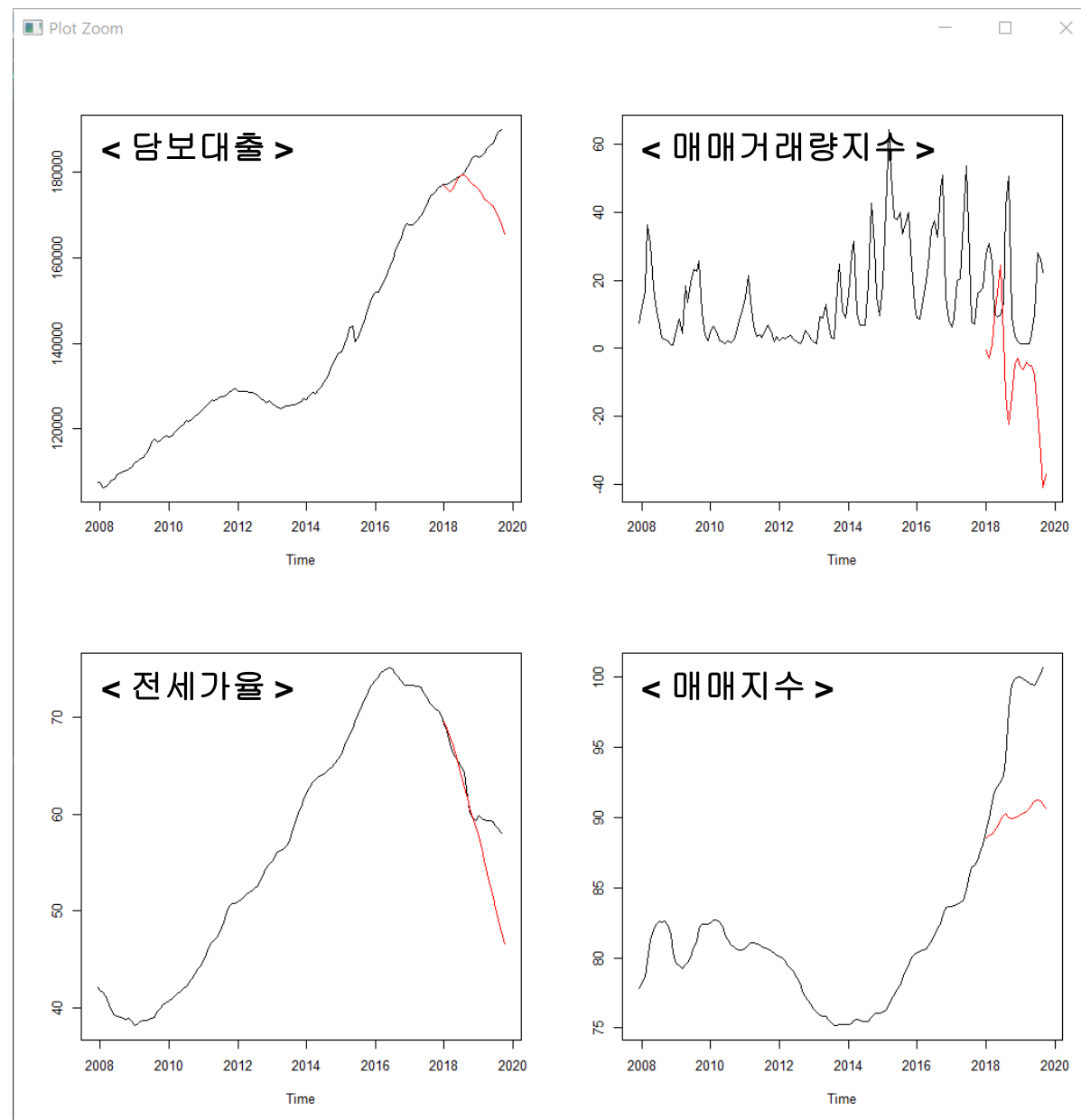


# 공적분분석

- ✓ Learning Set : '07.12 ~ '17.12  
Test Set : '18.01 ~ '19.09
- ✓ 요한슨공적분 검정 결과 :  $r \leq 3$

```
summary(vecm_1s)
# Values of teststatistic and critical values of test:
#      test 10pct  5pct  1pct
# r <= 3 |  9.50  7.52  9.24 12.97
# r <= 2 | 27.41 17.85 19.96 24.60
# r <= 1 | 57.73 32.00 34.91 41.07
# r = 0  | 95.33 49.65 53.12 60.16
```

## Learning Set & Test Set



# 분석하면서 생각 혹은 고민한 사항

- ✓ 정책변수(가변수, dummy)를 고려한 모형
- ✓ 교통호재에 따른 영향을 고려한 모형(시간에 따라 가격에 영향을 끼치는 강도가 변함, 지역별 비교시 차별화된 변수)
- ✓ 서울시 뿐만이 아닌 각 구별로 차별화된 영향요인을 발굴하여 예측
- ✓ 공급의 지표로 지역별, 연령별(~5년, ~10년...) 재고주택량 활용 방안 (통계자료 X)
- ✓ 시계열 데이터를 일정한 간격으로 분리하여 교차검증하는 방법
- ✓ VAR모형에서는 과거의 데이터를 고려하므로 상관이 없겠지만 회귀모형에서는 CCF를 고려한 lag변수를 반영하면 정확도가 개선될까?
- ✓ 회귀모형에서는 예측을 하기 위해서 설명변수의 미래값이 필요한데 각 설명변수의 값을 어떻게 예측 혹은 정할것인가?
- ✓ 종속변수를 포함해서 모든 변수가 차분이 되어 있는데 어떻게 해석할 것인가?
- ✓ 자기시차변수 활용 모형은 1차시 후의 모형만 예측가능한 한계



감사합니다!