

개입분석 2019221010 방태모

목차

- 개입분석 내용 정리

- 경부선 역별 10년 수송량 예측
 1. 개입분석을 이용할 역별 수송량 추세
 2. 역별 개입지점 결정
 - 2.1 서울역
 - 2.2 석수역
 - 2.3 광명역
 3. 수송량 예측 및 기존 ARIMA모형과 비교평가
 - 3.1 서울역
 - 3.2 석수역
 - 3.3 광명역

개입분석 내용 정리

1. 기본 개념

1.1 개입

시계열에 영향을 미치는 외부사건

1.2 개입분석

외부의 개입에 의해 시계열의 형태가 변했을 때 이를 시계열 모형에 반영한 분석

1.3 기본가정

(1) 외부사건 개입시점을 안다.

(2) 개입의 속성과 크기를 알고 있다.

(3) 개입시점 전과 후에 동일한 모형을 따른다(개입효과 제외).

예를들어 외환위기때 소비가 많이 감소했는데, 그 시기를 빼고 시계열을 연결하면 동일한 Arima 모형을 따른다.

2. 개입모형

$$Z_t = \frac{w(B)B^b}{\delta(B)}I_t^{(T)} + \frac{\theta(B)}{\Phi(B)}\epsilon_t$$
$$w(B) = w_0 - w_1B - \dots - w_sB^s$$
$$\delta(B) = \delta_0 - \delta_1B - \dots - \delta_rB^r$$

b : 개입효과에 대한 지연시간
 $I_t^{(T)}$: 개입에 관한 변수

위처럼 모형이 표현될 수 있다. $w(B), \delta(B)$ 는 개입의 형태를 나타낸다. 우변의 첫 번째 항은 개입에 관한 변수로 표현이 되어있다. 두 번째 항은 시계열 자체의 움직임을 나타내며, 즉 기존의 ARIMA 모형 부분이다. 그래서 개입모형은 개입에 의한 부분과 ARIMA 모형으로 구분된다. b 는 개입의 지연시간으로 예를 들어 전쟁이 일어나면 즉시 개입효과가 일어나지만, 금리 정책과 같은 것은 일정한 시차를 두고 개입이 일어난다.

2.1 개입의 기본적 형태

$$S_t^{(T)} = \begin{cases} 1, & t \geq T \\ 0, & t < T \end{cases} \dots (a)$$

$$P_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \dots (b)$$

$$\eta_t^{(T)} = \begin{cases} 1, & T_1 \leq t \leq T_2 \\ 0, & \text{그밖에서} \end{cases} \dots (c)$$

T : 개입이 일어난시점

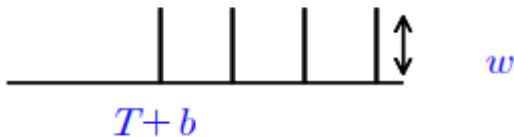
- (a) $S_t^{(T)}$: 개입이 나타난 시점을 T라고 한다면, 그 시점 전까지는 효과가 없다가 그 시점 이후에는 효과가 일정하게 존재하는 것. 계단함수(Step function)라고 한다.
- (b) $P_t^{(T)}$: T 시점에만 효과가 갑자기 나고, 나머지 시점에는 효과가 나지 않음. 지시함수(Indicator function) 또는 펄스함수(Pulse function)라고 한다.
- (c) $\eta_t^{(T)}$: 일정기간($T_1 \sim T_2$)에는 개입효과가 있고, 나머지 시점에는 효과가 없는 경우.

-> 위 세 가지를 조합해서 다양한 패턴을 만들 수 있다.

2.2 개입함수의 형태에 따른 반응형태

(1) 모형 1(Step function)

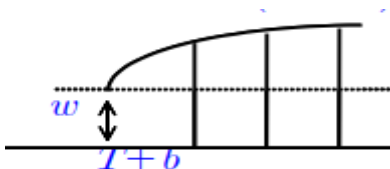
$$f_t = wB^b S_t^{(T)}$$



어떤 시계열의 개입이 있으면, b(지연시간)라는 시점 이후에 w라는 크기로 일정하게 나타나는 개입모형 형태(아주 단순한 형태).

(2) 모형 2(Delayed step function)

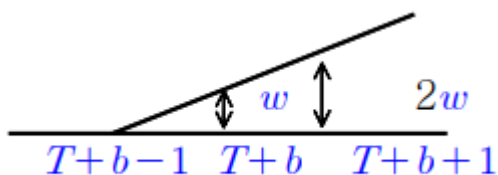
$$f(t) = \frac{w}{1-\delta B} B^b S_t^{(T)}$$



개입효과가 발생해 서서히 증가하며 빠르게 상승하다가 일정한 수준에 수렴하는 개입모형 형태. 예를 들어, 제품에 대해 광고를 한 경우, 급격히 광고 효과가 나타나다가 나중엔 일정 수준으로 수렴하는 경우.

(3) 모형 3

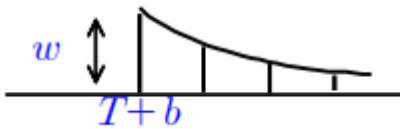
$$f(t) = \frac{w}{1-B} B^b S_t^{(T)}$$



이 모형은 어떤 광고를 했을 때, 그 효과가 선형적으로 증가하는 경우다. 모형 2와 차이점은 $\frac{w}{1-B}$ 에 모수 δ 가 포함되어 있지 않음.

(4) 모형 4(Decayed pulse function)

$$f(t) = \frac{w}{1-\delta B} B^b P_t^{(T)}$$

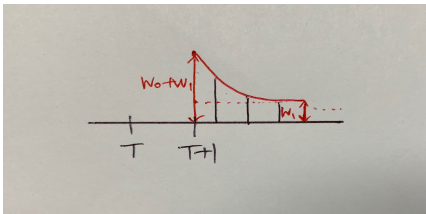


예를 들어, 전쟁 발발해서 유가가 급락한 효과가 발생했을 때, 전쟁이 끝나면서 그 효과가 서서히 없어지는 경우. 즉, 일정 시점이 지나서 개입효과가 서서히 사라지는 경우.

-> 통상적인 경우, 증가하다가 일정 수준에 수렴하는 모형 2와 개입효과가 서서히 사라지는 모형 4를 많이 쓴다. 위에서 본 모형들 외에도 여러 모형을 혼합한 다양한 패턴을 생각할 수 있다. 개입모형의 기본 가정으로 개입 전후(개입제외하면)에 같은 모형 따른다는 점을 꼭 기억하자.

(5) 모형 5

$$f(t) = \left[\frac{w_0}{1-\delta B} + \frac{w_1}{1-B} \right] B P_t^{(T)}$$



$T+1$ 시점에 Pulse function 형태로 두 효과 w_0, w_1 가 나타나고, w_0 의 효과는 서서히 사라지는 패턴의 개입효과를 나타낸 개입함수.

Example) 모형식 해석

$$\begin{aligned} & (1-B)(1-B^4) \left(Z_t - \frac{w}{1-\delta_1 B} B S_t^{(67)} \right) \\ &= \theta_0 + \frac{1}{(1-\Phi B^4)} \epsilon_t \end{aligned}$$

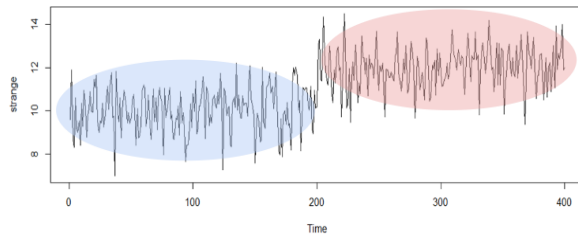
$(1-B)$ 는 1차 차분을 해준 것이고, $(1-B^4)$ 은 계절 차분을 해준 것이다. Z_t 는 원 계열 또는 로그변환을 한 계열을 의미한다. 항 $\frac{w}{1-\delta_1 B} B S_t^{(67)}$ 은 개입의 형태이며, Step function 형태의 개입이 있었는데, 그 시점이 67이다(예를 들면, 외환위기가 발생한 시기). 그리고 이때 앞

쪽에 B 가 존재하므로 한 개의 시차를 두고 나타났음을 (여기서는 분기 자료 이므로 한 분기) 의미한다. 그 앞쪽의 $\frac{w}{1-\delta_1 B}$ 항은 개입의 형태를 의미한다. 앞서 배운 모형 2에 해당하며 즉, 개입효과가 나타나고 일정 수준으로 수렴하는 형태를 나타낸다. 따라서 원계열에 이러한 개입형태를 빼고 모형식을 정리하면 $\theta_0 + \frac{1}{(1-\phi B^d)} \epsilon_t$ 와 같이 계절형 ARIMA 모형을 가지는 것을 알 수 있다.

3. arimax(){TSA}

(출처 - <https://slides.com/tonyg/intervention-analysis#/>)

Examine separately at the parts before and after the level shift. There are in total **400** time-points. Select the first **190** and the last **190** observations.



Let:

$$I_t = \begin{cases} 1 & t \geq 200 \\ 0 & t < 200 \end{cases}$$

Our model is:

$$Y_t = \bar{Y} + \delta_0 I_t + \frac{1-\theta_1 B}{1-\phi_1 B} \epsilon_t$$

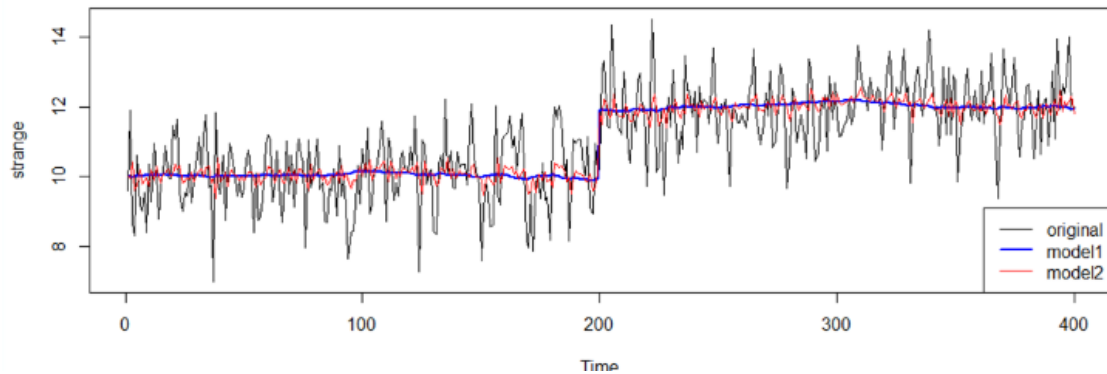
```
strange.model <-arimax(strange,order=c(1,0,1),
                        xtransf=data.frame(step200=1*(seq(strange)>=200)),
                        transfer=list(c(0,0)))
```

1. The arimax(){TSA} command works like the arima command, but allows inclusion of co-variates.
2. The argument xtransf is followed by a data frame.
3. in which each column correspond to a co-variate time series (same number of observations as Y_t).
4. Data frame : 1*(seq(strange)>=200)
5. Transfer is followed by a list comprising one two-dimensional vector (comprise v.구성되다)
6. for each co-variate specified by xtransf.
7. list(c(0,0)) implies that the co-variate shall be included as it stands (no lagging, no filtering).

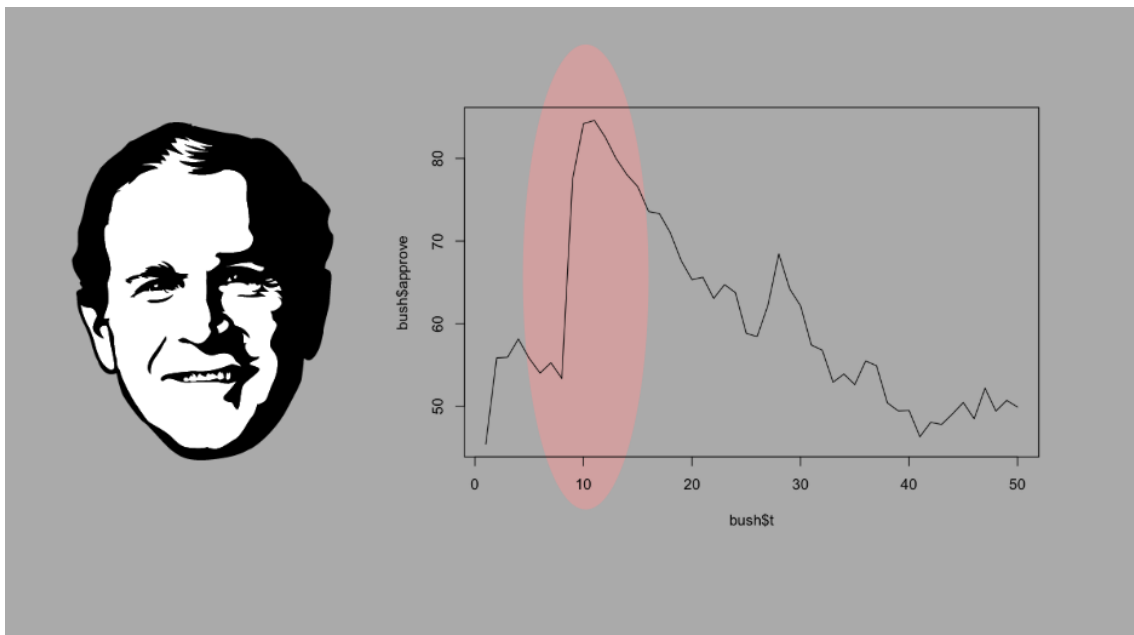
$$\delta_0 \frac{1-w_1B-w_2B^2 \dots -w_sB^s}{1-\sigma_1B-\sigma_2B^2 \dots -\sigma_rB^r} I_t$$

8. $c(r,s)$ where both r and s are > 0 will enter the term

So, $c(0,0)$ gives $\delta_0 I_t$, which means basic step function like the below plot.



Fun examples)



it's like decayed pulse function. That is, $Z_t - \bar{Z} = \frac{\delta_0}{1-w_1B} P_t + \frac{\theta(B)}{\Phi(B)} e_t$

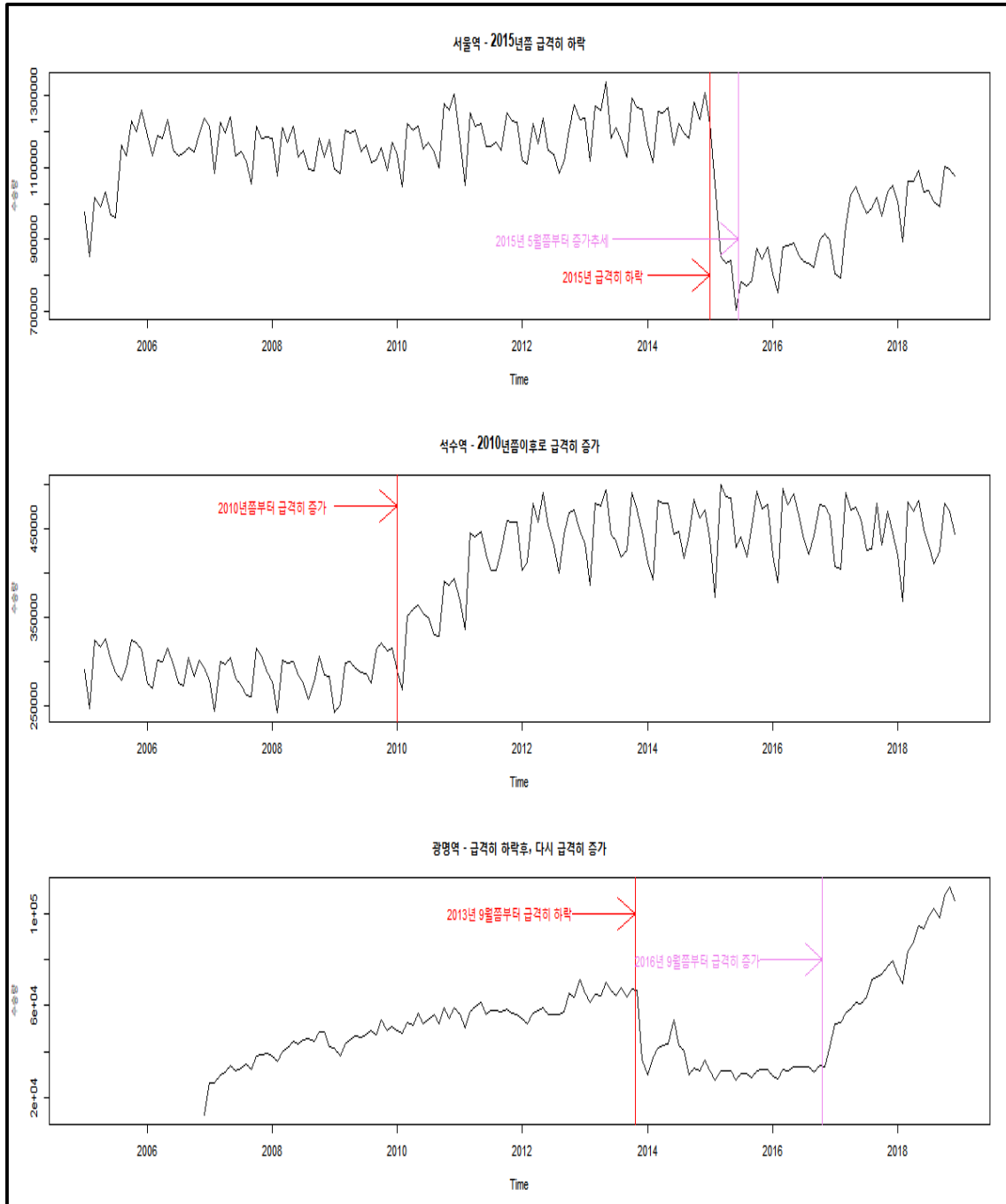
-> transfer=list(c(1,0))

```
#Looks like I(1)
#estimate intervention analysis
mod.2 <- arimax(bush$approve, order=c(0,1,0), xtransf=bush$s11, transfer=list(c(1,0)))
mod.2
```

경부선 역별 10년 수송량 예측

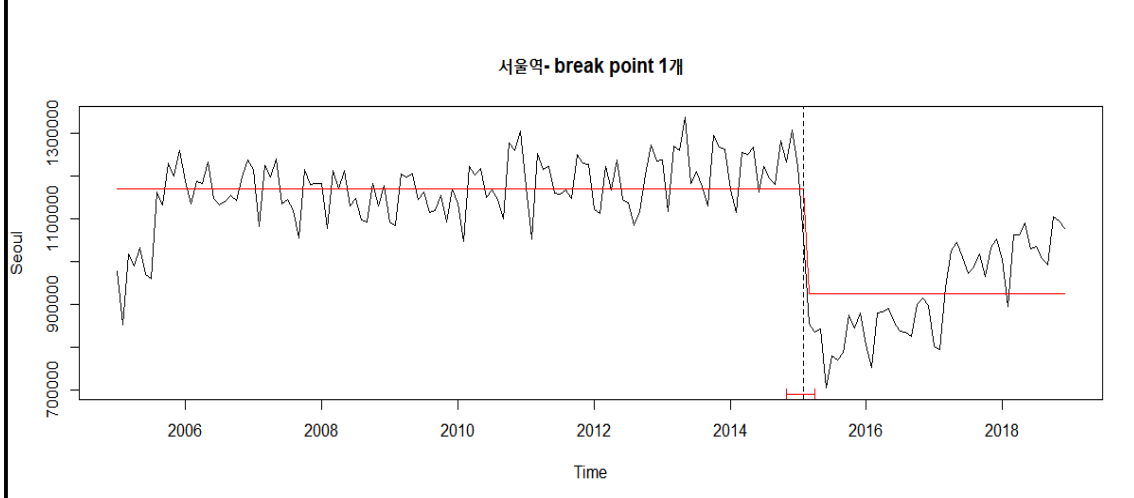
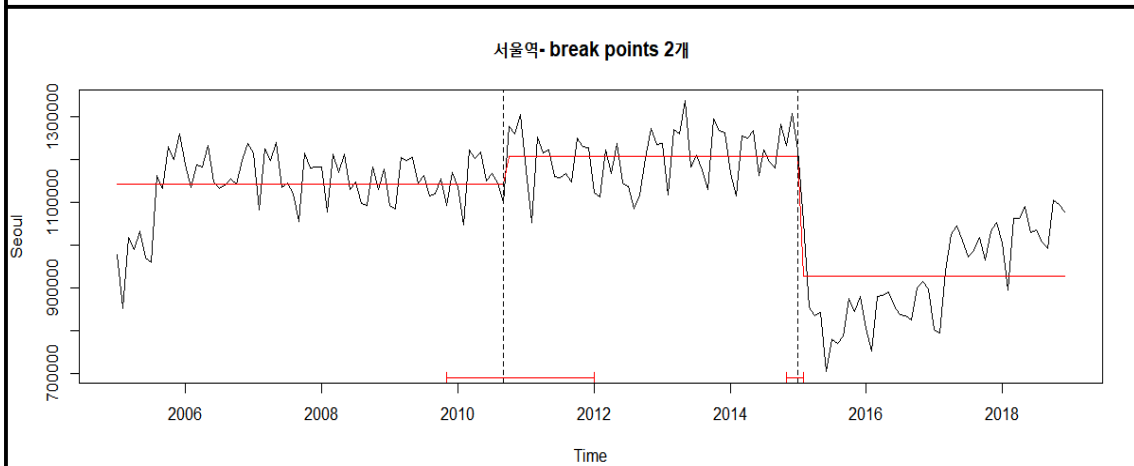
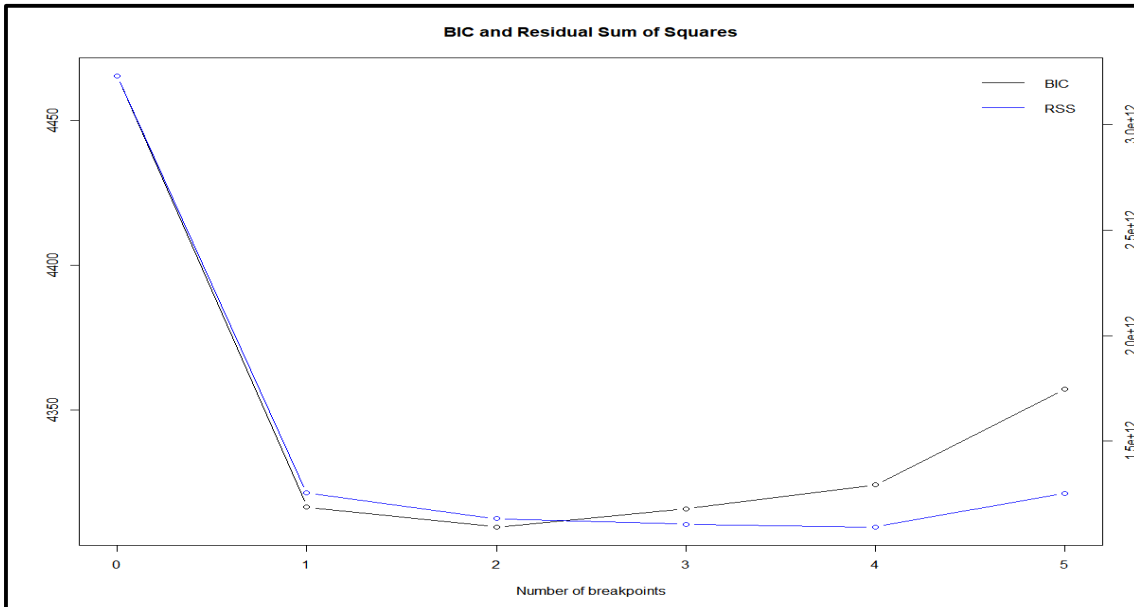
1. 개입분석을 이용할 역별 수송량 추세

: 서울역, 석수역, 광명역(2006년 12월 개통)

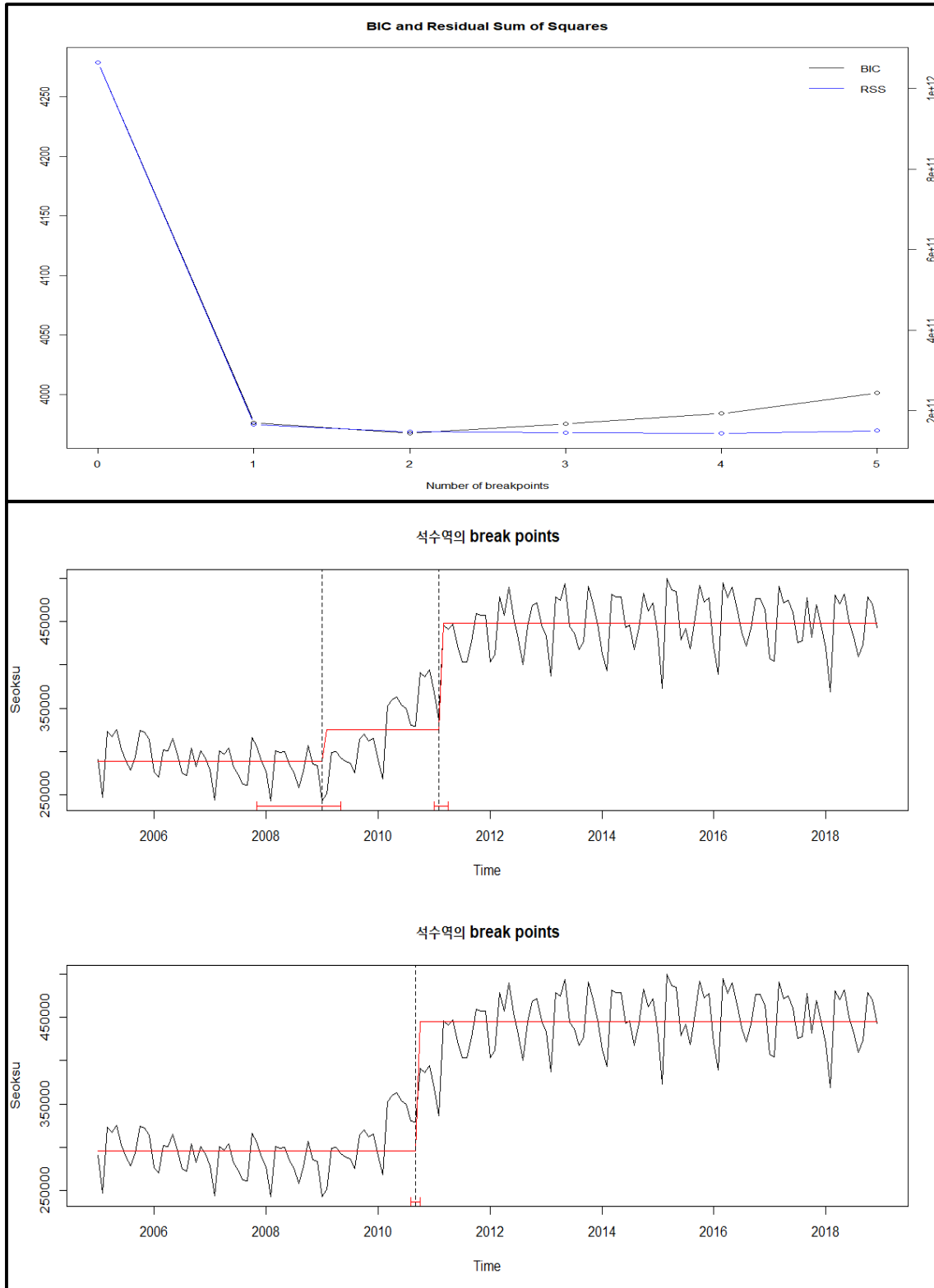


2. 역별 개입지점 결정

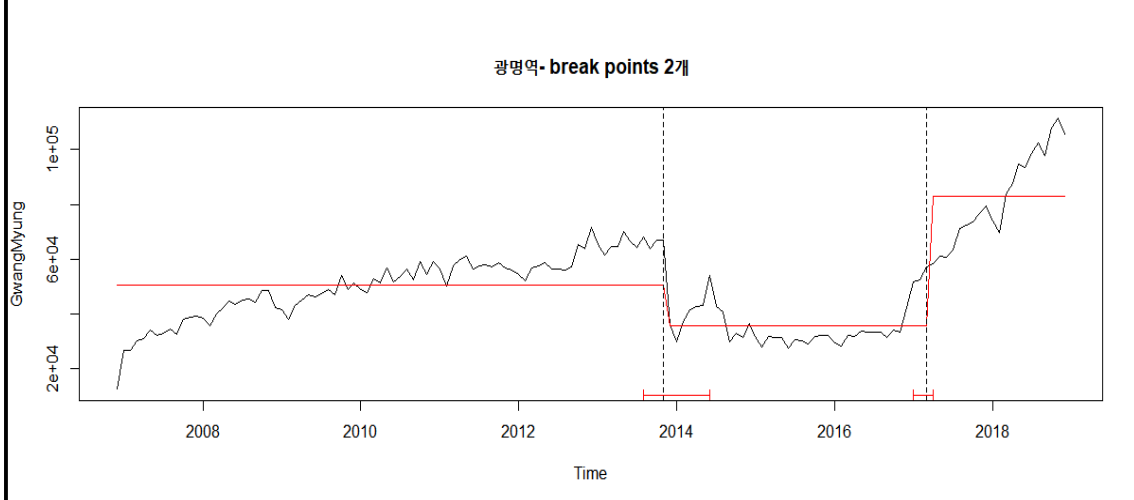
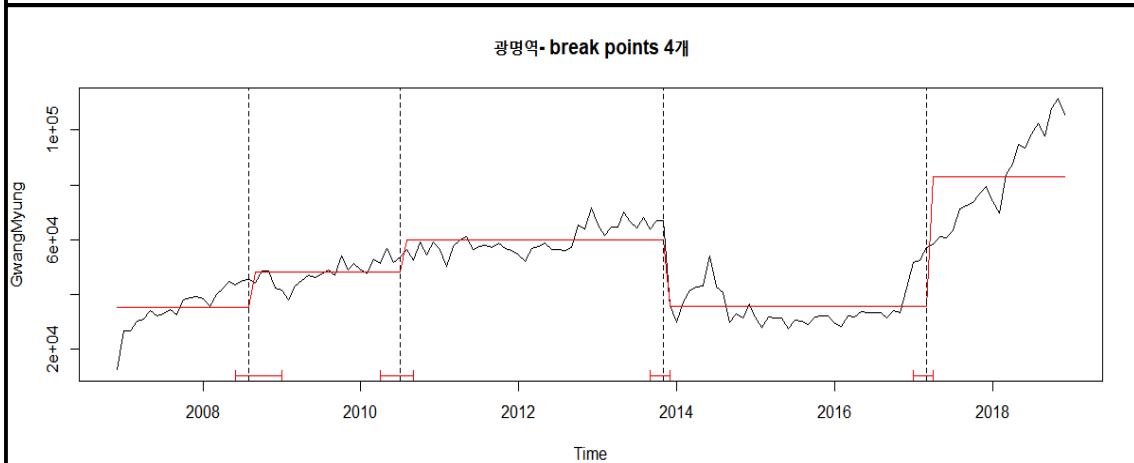
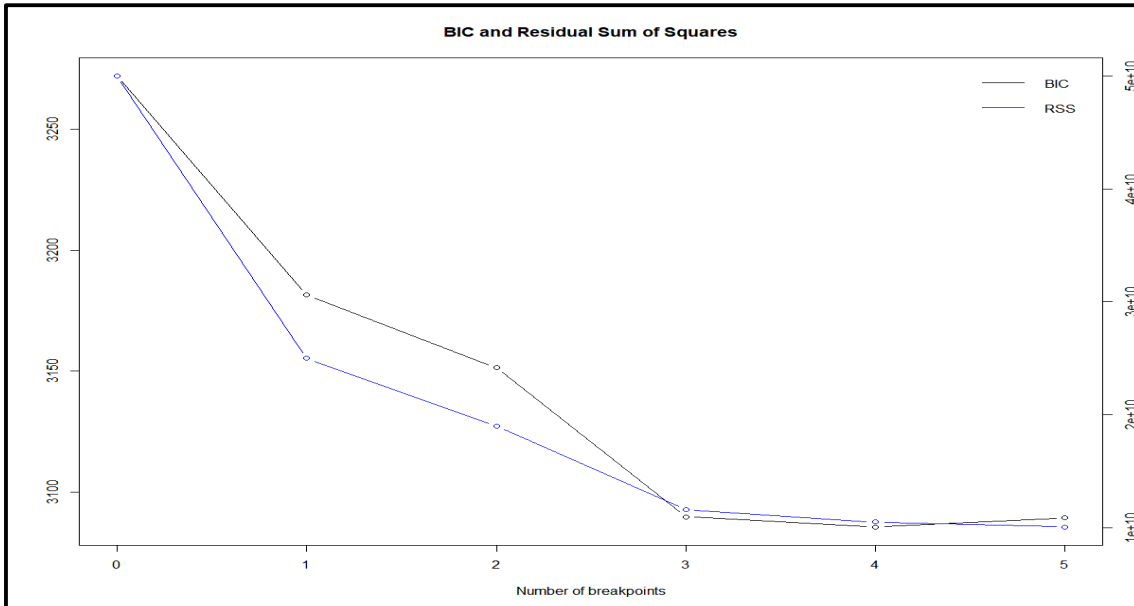
2.1 서울역 : 개입지점(break point) 1개로 결정함.



2.2 석수역 : 개입지점(break point) 1개로 결정함.



2.3 광명역 : 개입지점(break points) 2개로 결정함.

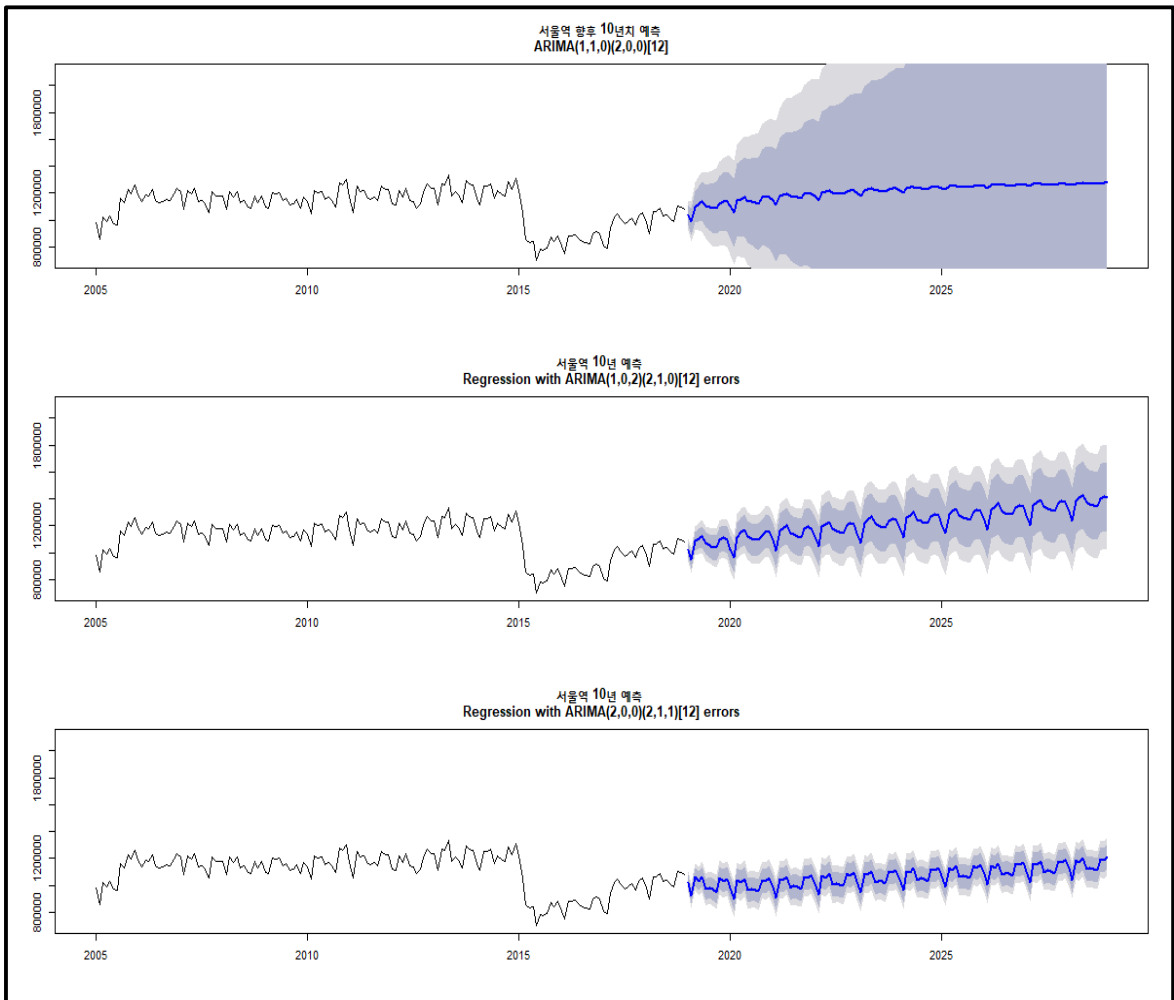


3. 수송량 예측 및 기존 ARIMA모형과 비교평가

3.1 서울역

○ 모형 평가 및 예측

| | RMSE | MAE | MAPE | ACF1 | 비고 |
|--|----------|----------|--------|---------|---------------------------------|
| ARIMA (1,1,0)(2,0,0)[12] | 57642.81 | 41483.77 | 3.8978 | 0.0049 | 모든 자료 이용 |
| Regression with ARIMA(1,0,2)(2,1,0)[12] errors | 41374.5 | 30818.69 | 2.8506 | -0.0095 | auto.arima() 로 차수 결정 |
| Regression with ARIMA(2,0,0)(2,1,1)[12] errors | 36409.35 | 26952.37 | 2.4852 | -0.0322 | 개입이전 시점 자료만 이용하여 차수 결정 |



○ 개입효과 유의성 검정

① auto.arima()로 차수 결정한 모형

(Regression with ARIMA(1,0,2)(2,1,0)[12] errors)

| | Estimate | Std. Error | z-value | Pr(> z) |
|-------|----------|------------|---------|----------|
| ar1 | 0.929 | 0.043 | 21.498 | <2e-16 |
| ma1 | -0.430 | 0.094 | -4.592 | <2e-16 |
| ma2 | 0.091 | 0.085 | 1.064 | 0.287 |
| sar1 | -0.598 | 0.085 | -7.066 | <2e-16 |
| sar2 | -0.382 | 0.086 | -4.430 | <2e-16 |
| drift | 2717.976 | 1422.078 | 1.911 | 0.056 |
| xreg | 1.606 | 0.146 | 10.981 | <2e-16 |

∴ 개입효과(xreg)는 유의했다.

② 개입 이전시점 자료만 이용하여 차수 결정한 모형

(Regression with ARIMA(2,0,0)(2,1,1)[12] errors)

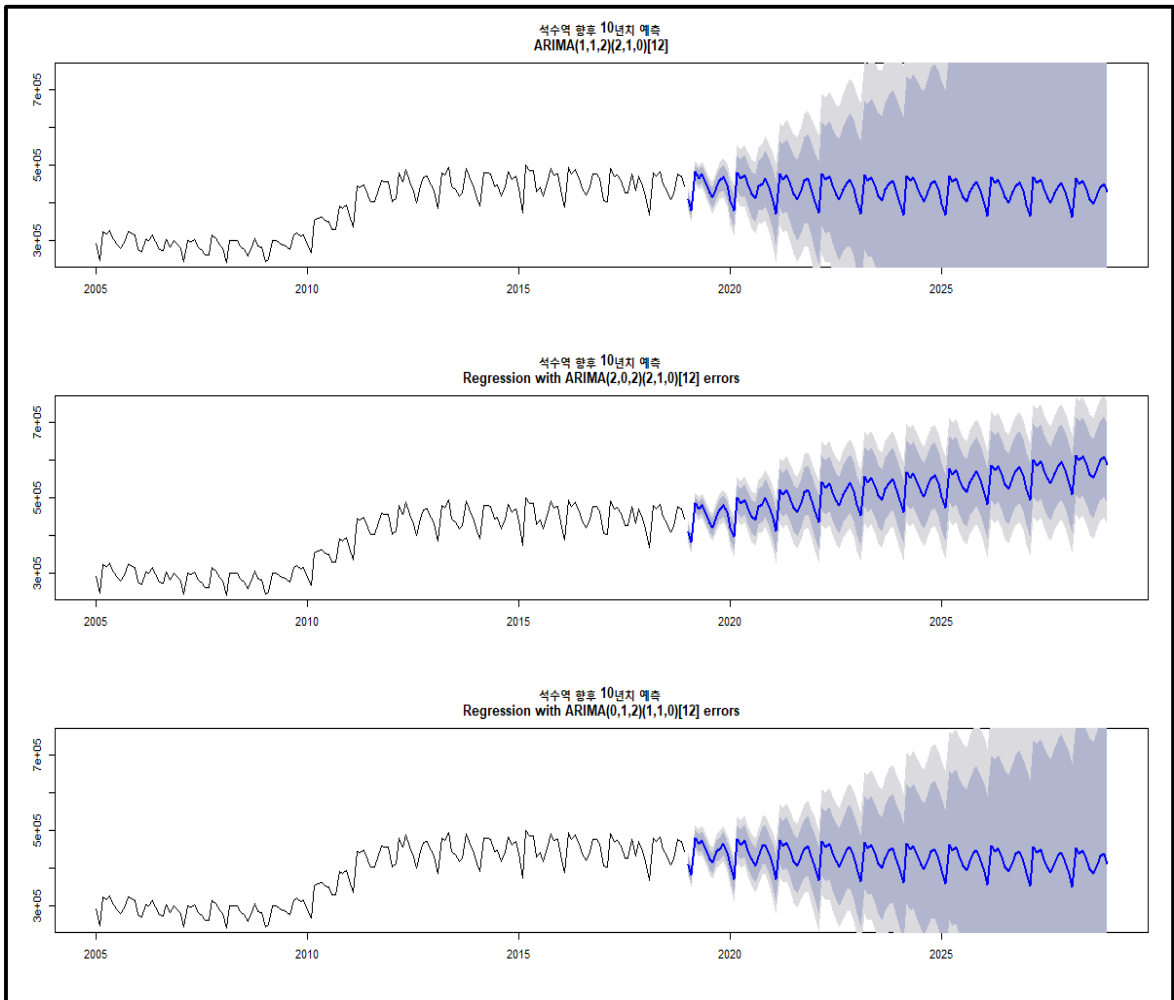
| | Estimate | Std. Error | z-value | Pr(> z) |
|-------|----------|------------|---------|----------|
| ar1 | 0.541 | 0.077 | 7.018 | <2e-16 |
| ar2 | 0.313 | 0.080 | 3.914 | <2e-16 |
| sar1 | 0.084 | 0.093 | 0.899 | 0.369 |
| sar2 | -0.118 | 0.109 | -1.080 | 0.280 |
| sma1 | -1.000 | 0.206 | -4.863 | <2e-16 |
| drift | 1590.000 | 472.946 | 3.362 | 0.001 |
| xreg | 1.495 | 0.146 | 10.250 | <2e-16 |

∴ 개입효과(xreg)는 유의했다.

3.2 석수역

○ 모형 평가 및 예측

| | RMSE | MAE | MAPE | ACF1 | 비고 |
|--|----------|----------|--------|---------|------------------------|
| ARIMA(1,1,2)(2,1,0)[12] | 14027.39 | 10434.68 | 2.755 | -0.0075 | 모든 자료 이용 |
| Regression with ARIMA(2,0,2)(2,1,0)[12] errors | 13704.79 | 10228.74 | 2.706 | -0.0396 | auto.arima()로 차수 결정 |
| Regression with ARIMA(0,1,2)(1,1,0)[12] errors | 15458.44 | 11675.21 | 3.0862 | 0.0195 | 개입이전 시점 자료만 이용하여 차수 결정 |



○ 개입효과 유의성 검정

① auto.arima()로 차수 결정한 모형

(Regression with ARIMA(2,0,2)(2,1,0)[12] errors)

| | Estimate | Std. Error | z-value | Pr(> z) |
|-------|----------|------------|---------|----------|
| ar1 | 1.954 | 0.028 | 70.867 | <2e-16 |
| ar2 | -0.960 | 0.027 | -34.898 | <2e-16 |
| ma1 | -1.788 | 0.047 | -38.020 | <2e-16 |
| ma2 | 0.847 | 0.043 | 19.570 | <2e-16 |
| sar1 | -0.704 | 0.080 | -8.844 | <2e-16 |
| sar2 | -0.381 | 0.085 | -4.499 | <2e-16 |
| drift | 1074.336 | 547.831 | 1.961 | 0.050 |
| xreg | 0.017 | 0.051 | 0.326 | 0.744 |

∴ 개입효과(xreg)는 유의하지 않았다.

② 개입 이전시점 자료만 이용하여 차수 결정한 모형

(Regression with ARIMA(0,1,2)(1,1,0)[12] errors)

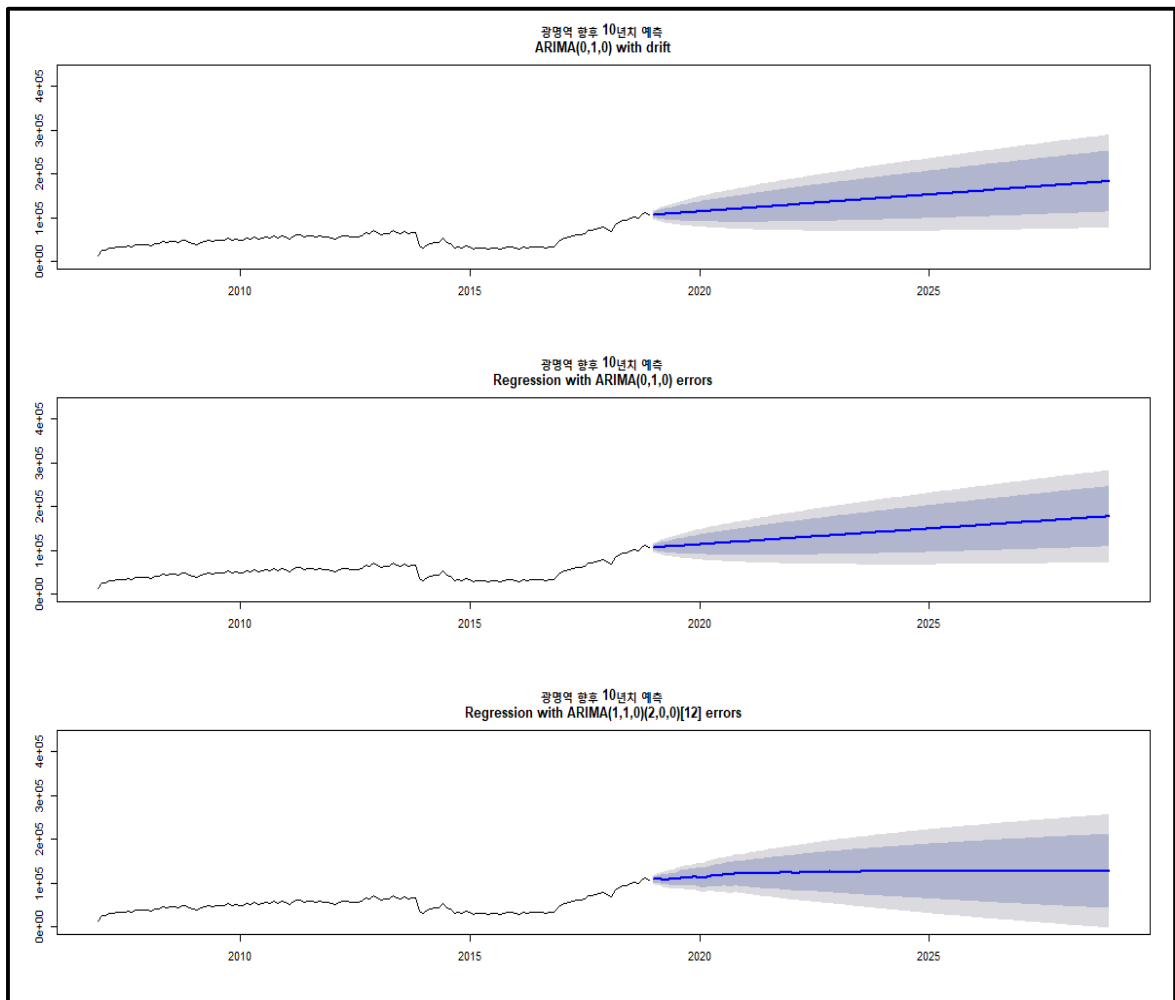
| | Estimate | Std. Error | z-value | Pr(> z) |
|------|----------|------------|---------|----------|
| ma1 | -0.842 | 0.086 | -9.760 | <2e-16 |
| ma2 | 0.202 | 0.084 | 2.397 | 0.017 |
| sar1 | -0.489 | 0.077 | -6.318 | <2e-16 |
| xreg | 0.175 | 0.068 | 2.568 | 0.010 |

∴ 개입효과(xreg)는 유의했다.

3.3 광명역 : 2006년 12월 개통

○ 모형 평가 및 예측

| | RMSE | MAE | MAPE | ACF1 | 비고 |
|--|----------|----------|--------|---------|------------------------|
| ARIMA(0,1,0) with drift | 4917.496 | 3377.123 | 7.2409 | -0.075 | 모든 자료 이용 |
| Regression with ARIMA(0,1,0) errors | 4843.802 | 3411.361 | 7.2695 | -0.1012 | auto.arima()로 차수 결정 |
| Regression with ARIMA(1,1,0)(2,0,0)[12] errors | 4701.725 | 3221.206 | 6.8906 | -0.0069 | 개입이전 시점 자료만 이용하여 차수 결정 |



○ 개입효과 유의성 검정

① auto.arima()로 차수 결정한 모형

(Regression with ARIMA(0,1,0) errors)

| | Estimate | Std. Error | z-value | Pr(> z) |
|-------|----------|------------|---------|----------|
| drift | 599.395 | 405.656 | 1.478 | 0.140 |
| xreg | 0.206 | 0.098 | 2.101 | 0.036 |

∴ 개입효과(xreg)는 유의했다.

② 개입 이전시점 자료만 이용하여 차수 결정한 모형

(Regression with ARIMA(1,1,0)(2,0,0)[12] errors)

| | Estimate | Std. Error | z-value | Pr(> z) |
|------|----------|------------|---------|----------|
| ar1 | -0.063 | 0.087 | -0.721 | 0.471 |
| sar1 | 0.062 | 0.081 | 0.760 | 0.447 |
| sar2 | 0.245 | 0.085 | 2.900 | 0.004 |
| xreg | 0.230 | 0.095 | 2.421 | 0.015 |

∴ 개입효과(xreg)는 유의했다.