

서울시 오존 농도 시계열 예측

2022711778윤탤희

목차

- 1.서론
- 2.데이터 설명
- 3.데이터 처리
- 4.데이터 모델링(SARIMA, VAR)
- 5.모델 SUMMARY
- 6.결론
- 7.READER FEEDBACK

1.서론

오존에 대해서 성층권의 오존은 해로운 자외선을 차단 해주는 이로인 효과가 있지만 대류권에 있는 오존은 사람의 호흡계, 눈, 농작물에 악영향을 끼친다. 미세먼지가 다양한 예보등급이 있는 것처럼 오존 또한 좋음, 보통, 민감군, 나쁨, 매우나쁨, 위험이라는 다양한 예보 등급이 있다. 그래서 본 보고서에서는 서울시 오존 데이터를 이용하여 오존 예보 등급을 예측하는 몇가지의 시계열 모델을 만들어 볼 것이다. 다양한 시계열 모델들의 지표 및 특징을 살펴보고 지표가 가장 좋은 모델을 최종 모델로 선택할 것이다. 최종으로 선택된 모형을 토대로 향후 2년간 서울시의 오존 농도에 대한 특성을 살펴보고 이것이 현재 오존의 예보 등급을 얼마나 잘 예측하는지 확인할 것이다.

2. 데이터 설명

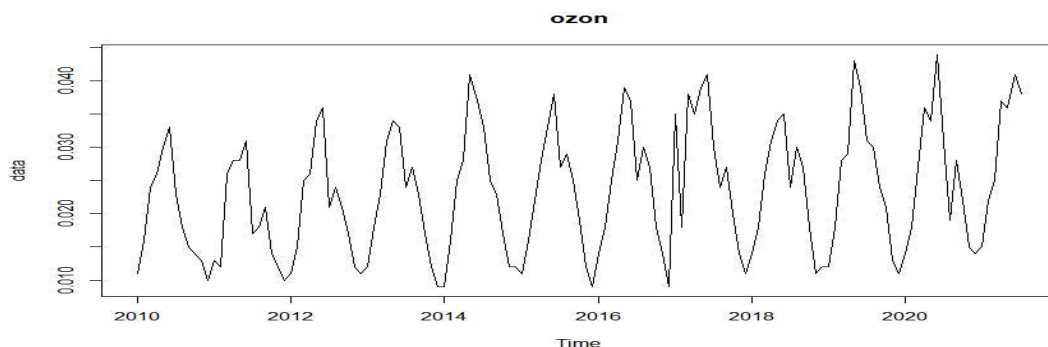
데이터는 환경부에서 제공하는 서울시 월별 오존 데이터를 이용했다. 시점은 2010년도 1월부터 2021년 7월까지이며 총 139개이다. 오존 농도의 단위는 ppm이다.

<표1>

구분	예보등급					
예측농도	좋음	보통	민감군	나쁨	매우나쁨	위험
	0~0.04	0.041~0.08	0.081~0.12	0.121~0.30	0.301~0.5	0.501
		0		0		

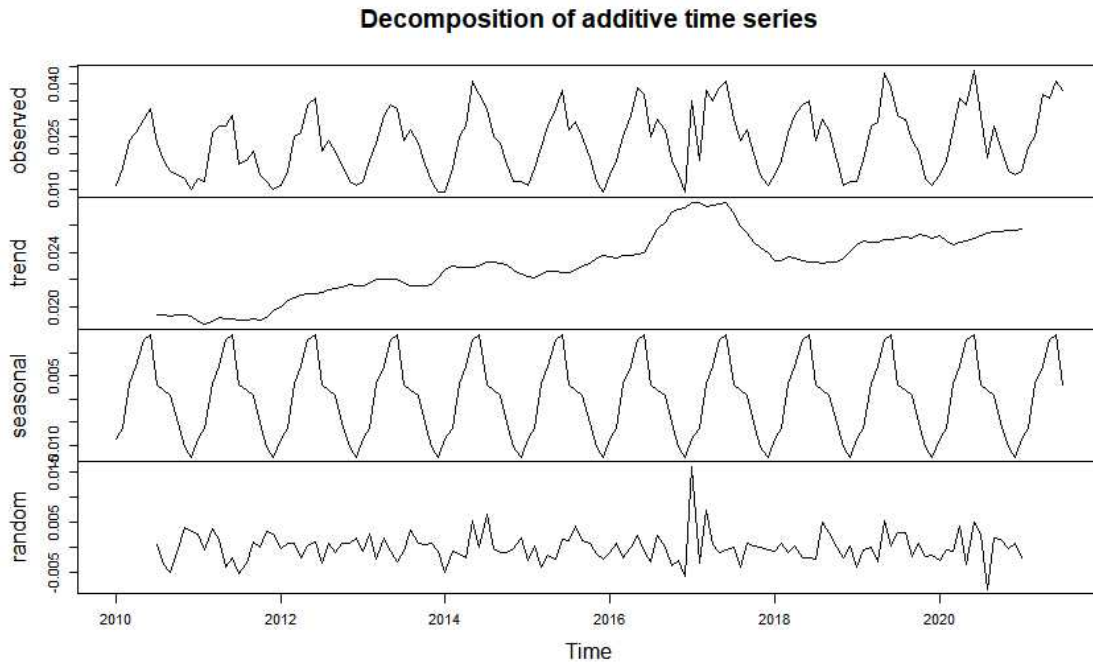
<표1>은 실제 예보등급을 나타낸다. 수집한 데이터를 살펴보면 월별 오존 농도의 최대값은 0.044이고 최소값은 0.009인 것을 확인할 수 있었다. 이 수치에 따르면 2010년 1월부터 2021년 7월까지의 오존 농도의 예보 등급은 좋음과 보통 총 2가지 등급에 국한 되어 있는 것을 확인할 수 있다.

<그래프1>



<그래프1>은 2010년 1월부터 2021년 7월까지의 서울시 오존 농도의 월별 데이터이다. 그래프를 살펴보면 확실히 1년 단위로 seasonality가 존재하는 것을 확인할 수 있다. 또한 엄밀히 따져보았을 때 그래프가 약간 완만하게 우상향하는 특징을 보인다. trend와 seasonality가 확실히 존재하는지 확인해 보기 위하여 시계열 분해를 실시한 결과 그 결과는 아래와 같다.

<그래프2>

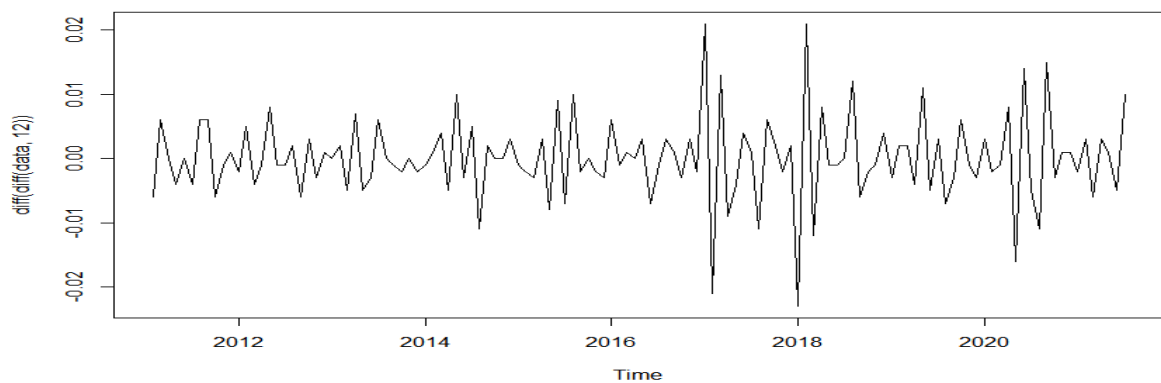


<그래프2>를 살펴보면 우선 trend 부분은 확실히 우상향하는 trend가 있는 것을 확인할 수 있었다. 또한 seasonal 부분을 보면 seasonality가 대략 1년 주기로 존재하는 것을 확인할 수 있었다. 이를 통해서 데이터 처리를 할 때 trend와 seasonality를 없애기 위해 일반 차분과 계절 차분을 진행하는 것이 데이터를 stationary하게 만드는 조건으로 보았다.

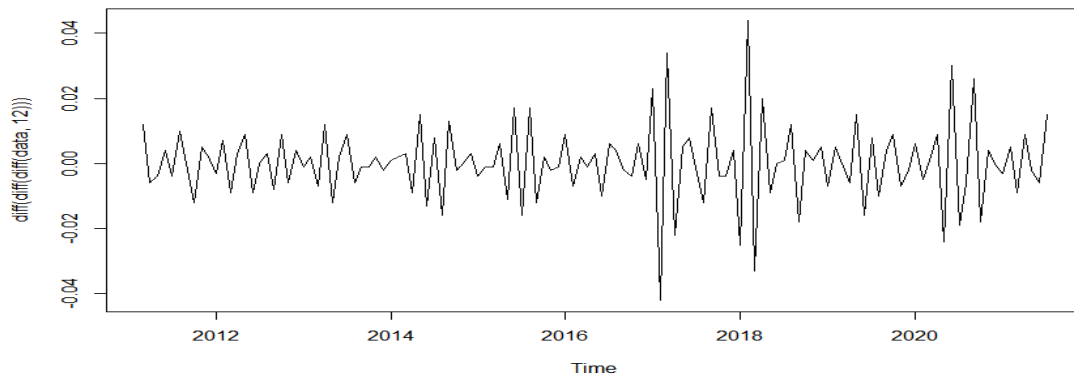
3.데이터 처리

데이터를 stationary하게 만들기 위해서 계절 차분을 1회 실시한 다음 일반 차분한 결과와 이후에 한번 더 일반 차분한 결과는 다음과 같다.

<그래프3>



<그래프4>



<그래프3>은 주기를 12로 정하고 계절 차분을 한 상태에서 일반 차분을 1번 한 데이터의 그래프이다. <그래프4>는 <그래프3>에서 일반 차분을 1번 더 한 데이터의 그래프이다. 2개의 data plot을 살펴보면 그렇게 큰 차이가 없는 것을 확인할 수 있다. 그런데 두 그래프 모두 2017년과 2018년 근방에서 분산이 증가해 있는 것을 확인할 수 있다. 그래서 log 변환을 실시해 보았지만 그렇게 큰 변화가 탐지 되지 않아 log 변환은 실시하지 않았다. 결과적으로 <그래프3>과 <그래프4>를 비교한 결과 그렇게 큰 차이가 없는 것을 확인할 수 있다. 그래서 데이터의 정제는 계절 차분을 한 후에 일반 차분을 1번까지만 실시하는 것으로 결정하였다. 혹시나 이렇게 차분한 결과의 데이터가 stationarity를 가지는지 확인하기 위해 단위근 검정을 실시하였다. 결과는 아래와 같다.

<표2>

Augmented Dickey-Fuller Test
data: diff(diff(data,12))
Dickey-Fuller=-23.567, Lag order=0, p-value=0.01
alternative hypothesis: stationary

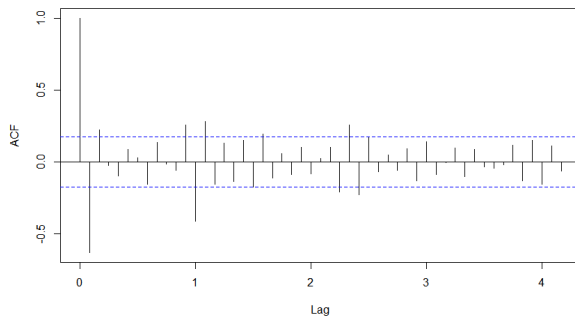
검정에 있어 귀무가설은 ‘데이터는 stationary하지 않다’이고 대립가설은 ‘데이터는 stationary하다’라고 설정할 수 있다. p-value 값을 보면 유의수준 5%에서 굉장히 작은 값이 나왔기 때문에 귀무가설을 기각하고 대립가설을 채택할 수 있다. 그래서 계절차분을 실시하고 일반 차분을 1번 한 데이터는 stationarity를 가진다고 할 수 있다. 모델링할 종류의 모델은 SARIMA 모형인데 어떤 order를 선택할 것인지에 대해서는 최종적으로 차분된 데이터의 ACF와 PACF를 보고 모델 결정을 하기로 하였다.

4. 데이터 모델링

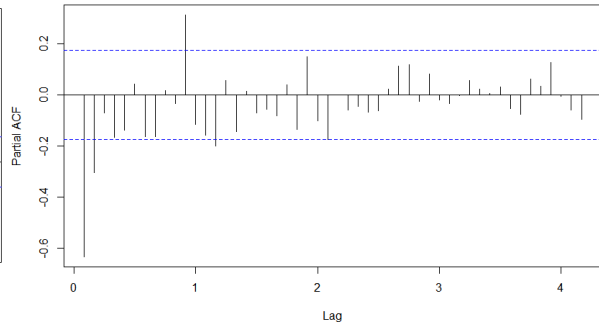
-SARIMA 모형

데이터를 어떤 방식으로 모델링 할지 결정하기 위해 우선 최종적으로 차분된 데이터의 ACF와 PACF를 살펴보았다. 그래프의 결과는 아래와 같다.

<그래프5>



<그래프6>



<그래프5>는 최종 차분된 데이터의 ACF <그래프6>은 PACF를 나타낸다. 우선 ACF를 살펴보면 lag이 12인 부분에 크게 boundary를 벗어나 있는 것을 확인할 수 있다. 또한 ACF와 PACF 두 그래프 모두 0 근방에서 boundary를 벗어나 있는 것을 확인할 수 있다. 두 그래프의 결과만 가지고 SARIMA모형의 order 값을 어떤 값으로 확정할지에 대해서 확신할 수 없었다. 그래서 여러 가지의 order 값들을 산정한 다음에 여러 가지 지표값들을 비교함으로써 지표가 가장 좋은 모델을 최종모형으로 선택하고자 하였다. 우선 임의로 상정한 모델은 $ARIMA(2,1,0)(0,1,1)$, $ARIMA(1,1,0)(1,1,0)$, $ARIMA(0,1,0)(1,1,1)$, $ARIMA(0,1,1)(2,1,1)$, $ARIMA(0,0,0)(0,1,1)$ 총 5가지이다. 여기서 추가적으로 R에서 `auto.arima` 함수를 사용한 결과로 나온 모형은 $ARIMA(0,0,0)(2,1,0)[12]$ 였고 총 6가지의 모델을 FITTING하여 여러지표들을 비교해 보기로 하였다. 모델링을 한 이후에 각 모델들의 지표값들은 아래와 같다.

<표3>

	ARIMA(2,1,0)(0,1,1)	ARIMA(1,1,0)(1,1,0)	ARIMA(0,1,0)(1,1,1)	ARIMA(1,1,1)(0,1,0)	ARIMA(0,0,0)(1,1,0)	ARIMA(0,0,0)(2,1,0)
Ljung-Box Q	0.8539	0.0494	0	0.0044	0.0989	0.3474
McLeod-Li Q	1	0.6517	7e-04	0.0105	0.1838	0.7857
Turning points T	0.0069	0.2802	0.0103	0.2802	0.9462	0.7872
Diff signs S	0.5582	0.3798	0.1432	0.0404	0.0192	0.0192
Rank P	0.877	0.439	0.4742	0.6217	0.8713	0.682

<표3>은 총 6개의 모델로 모델링을 한 후에 잔차 검정을 실시한 결과의 지표를 나타낸다. 여기서 2~4번째 모델들을 보면 Ljung-Box Q값의 p-value 값들이 유의 수준 5% 보다 미만인 값을 나온 것을 확인할 수 있다. 이를 통해 2~4번째 모델들의 잔차가 iid process를 따르지 않는다고 할 수 있다. 이로 인해 2~4번째 모델은 최종 모델 선택에 있어 배제 하였다. 그래서 1번째, 5번째, 6번째 모델 3가지의 모델들 중에 선택을 하기로 결정하였다.

다음으로 각 fitting된 모델들에 대해서 여러 가지 지표들을 살펴보겠다. 결과는 아래와 같다.

<표4>

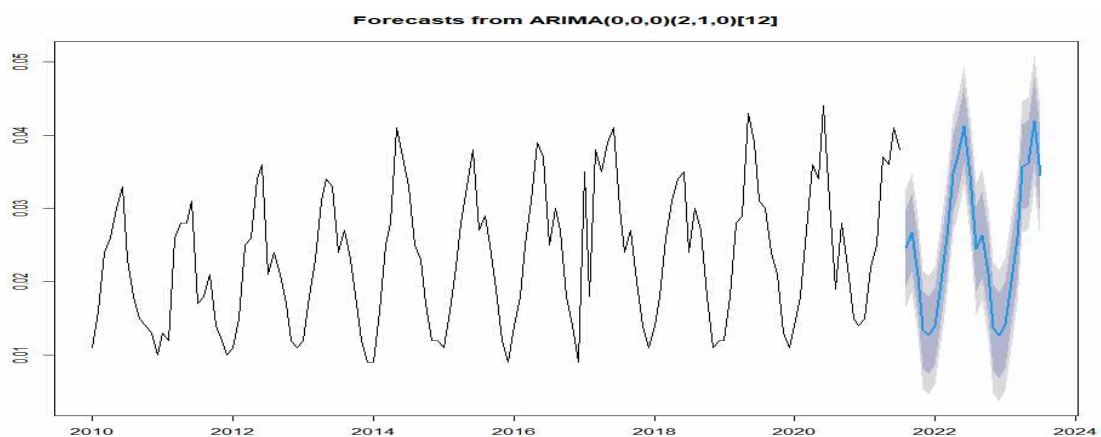
	ARIMA(2,1,0)(0,1,1)	ARIMA(1,1,0)(1,1,0)	ARIMA(0,1,0)(1,1,1)	ARIMA(1,1,1)(0,1,1)	ARIMA(0,0,0)(1,1,0)	ARIMA(0,0,0)(2,1,0)
RMSE	0.0034918 1	0.0043223 27	0.0044306 95	0.0044456 68	0.0040715 58	0.0039252 99
MAE	0.0024102 27	0.0028828 14	0.0029271 82	0.0030029 25	0.0027009 55	0.0025669 55
MAPE	-2.763059	13.37884	12.91164	13.47914	11.777263	11.07753
AIC	-1025.64	-992.95	-960.37	--987.84	-1019.99	-1025.64
BIC	-1014.297	-984.4431	-951.8654	-979.3326	-1014.304	-1017.107

<표4>를 보면 RMSE와 MAE만 보았을 때는 첫 번째 모델이 가장 수치가 적기 때문에 가장 좋은 모델이라고 생각할 수 있다. 그러나 AIC와 BIC를 보면 마지막 모델의 수치가 가장 적기 때문에 마지막 모델이 가장 좋은 모델이라고 생각할 수도 있다. 결국 모델 선택 기준을 RMSE, MAE, BIC 중에 어떤 것으로 잡느냐에 따라서 최종 선택 모델이 달라질 수 있는데 모델 선택 기준을 BIC로 정했기 때문에 BIC가 가장 작은 모델인 마지막 모델인 SARIMA(0,0,0)(2,1,0)[12]를 선택하였다.

SARIMA(0,0,0)(2,1,0)[12]로 fitting 된 모델식과 향후 2년간의 예측 값들과 80%, 95% 신뢰 구간을 그래프로 살펴보면 아래와 같다.

$$X_t = -0.5090X_{t-12} - 0.2499X_{t-24} + Z_t$$

<그래프7>



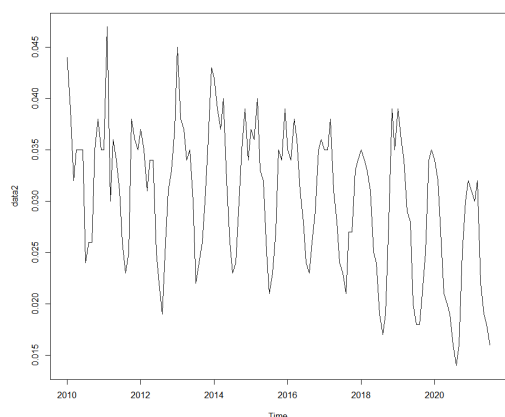
SARIMA모형 이외에 다른 시계열 모델을 고려해 본 후 SARIMA모형과 비교를 실시하겠다.

-VAR모형(이산화질소, 오존)-

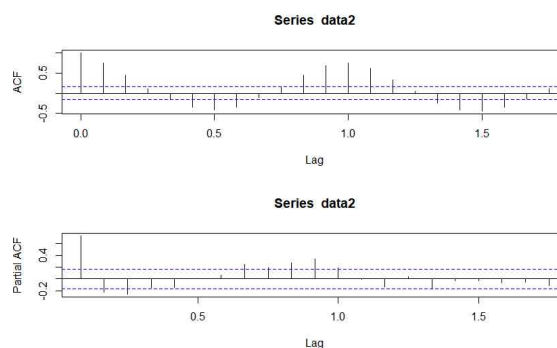
오존은 질소산화물과 휘발성 유기 화합물이 자외선을 받아 생성 된다. 이 때 오존 생성의 가장 영향을 미치는 요인은 질소산화물이라 할 수 있다. 이 질소산화물은 자동차나, 기차 등 교통수단에서 나오는 배기가스에서 가장 많이 검출 된다. 그래서 질소산화물의 대표적인 원소인 이산화질소와 오존이 연관이 있다고 생각했다. 이에 따라 오존과 이산화질소 총 두 변수를 이용해 VAR모형을 만들어 보고자 하였다.

이산화질소도 오존과 마찬가지로 2010년 1월부터 2021년 7월까지의 총 139개 데이터를 환경부에서 수집했다. 데이터의 특성을 살펴보면 아래와 같다.

<그래프8>



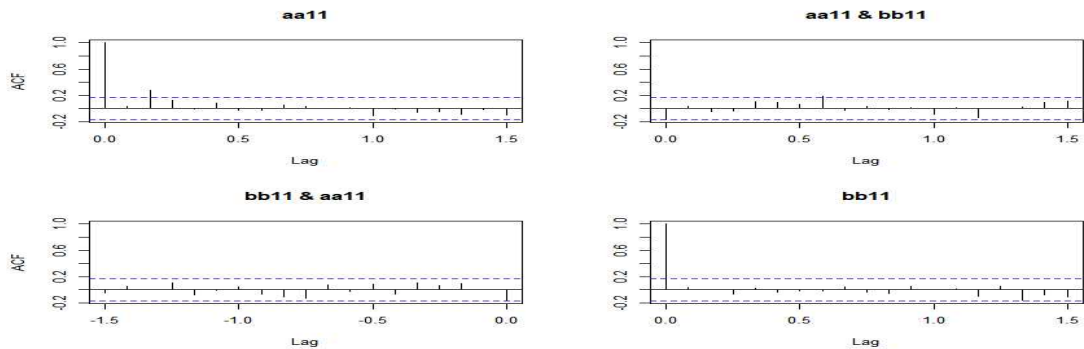
<그래프9>



<그래프8>은 이산화질소의 data plot을 나타내고 <그래프9>는 이산화질소 data의 ACF와 PACF를 나타낸다. data plot을 보면 seasonality가 존재하는 것을 확인할 수 있다. 또한 완만하게 우상향 했던 오존과는 다르게 완만하게 우하향하는 것을 확인할 수 있었다. 어떤 모델로 모델링할지에 대해서는 이산화질소는 오존과 마찬가지로 SARIMA모형으로 모델 fitting하기로 결정하였다. 따라서 R에서 auto.arima 함수를 사용하여 모델링한 결과 SARIMA(1,0,1)(0,1,1)[12] 모형이 나왔다.

이후 SARIMA(0,0,0)(2,1,0)[12]로 fitting한 오존과 SARIMA(1,0,1)(0,1,1)[12]로 fitting한 이산화질소가 실제로 연관이 있는지 확인하기 위해 CCF를 확인해 보고자 하였다. free whitening을 거쳐서 CCF를 확인한 결과는 아래 그래프와 같다.

<그래프10>



<그래프10>은 free whitening을 거친 오존과 이산화질소의 CCF를 나타낸다. 두 번째 그래프를 보면 대부분의 lag이 boundary안에 들어가 있지만 lag 7쯤에 boundary를 조금 벗어나 있는 것을 확인할 수 있다. 이로 인해 오존과 이산화질소는 어느 정도 연관성을 가지고 있다고 판단하였다. 이를 통해 오존과 이산화질소를 이용한 VAR모형을 만들고자 하였다.

우선 오존과 이산화질소 data를 stationary하게 만들기 위해서 앞에서는 계절차분 후에 1번 일반 차분까지 하였다. 이번 VAR모형을 만들 때는 1번 일반 차분을 한 상태에서 각각 stationarity를 가지는지 단위근 검정을 실시하였다. 단위근 검정을 실시하기 위해 R에서 adf.test를 실시하였다.

<표5>

Augmented Dickey-Fuller Test	Augmented Dickey-Fuller Test
data: ozon	data: no2
Dickey-Fuller=-11.6, Lag order= 0,	Dickey-Fuller=-10.526, Lag order= 0,
p-value=0.01	p-value=0.01
alternative hypothesis: stationary	alternative hypothesis: stationary

<표5>는 adf.test를 실시한 결과를 나타낸다. 오존 데이터, 이산화질소 데이터 둘 다 모두 p-value의 값이 유의 수준 5%보다 미만이므로 '데이터는 stationary하지 않다'는 귀무가설은 기각되고 '데이터는 stationary하다'는 대립가설이 채택된다. <표2>를 보면 계절차분을 한 후에 일반 차분을 1번 한 데이터도 stationary하다는 것을 확인할 수 있다. 그렇지만 일반 차분 1회만 실시한 데이터도 stationary하기 때문에 계절차분을 진행하지 않고 VAR모형을 만들어도 문제 없다고 판단했다. 그래서 일반 차분을 1회 실시한 채로 VAR모델 fitting을 실시하기로 결정하였다.

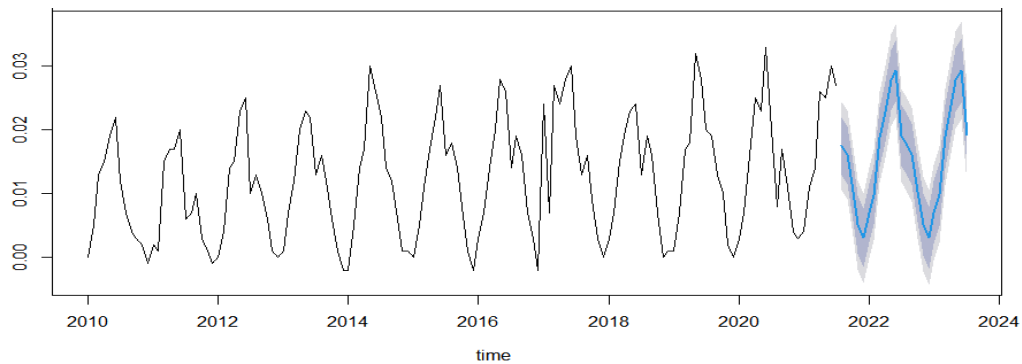
R에서 VARselect 함수를 사용한 결과 BIC 기준으로 적합하다고 판단된 p값은 8이었다. 그래서 p값을 8로하여 var함수를 사용한 결과 다음과 같은 모형이 fitting 되었다.

$$\begin{aligned}
X_t = & \begin{pmatrix} -0.56255027 & -0.0135059 \\ -0.03172114 & -0.81905983 \end{pmatrix} X_{t-1} + \begin{pmatrix} -0.53920679 & -0.1259378 \\ -0.07354236 & -0.5608200 \end{pmatrix} X_{t-2} \\
& + \begin{pmatrix} -49580913 & -0.3573690 \\ -0.08713002 & -0.3854174 \end{pmatrix} X_{t-3} + \begin{pmatrix} -0.3710931 & -0.3812348 \\ 0.3676391 & -0.2990041 \end{pmatrix} X_{t-4} \\
& + \begin{pmatrix} -0.4012263 & -0.3297817 \\ 0.6652812 & -0.2070856 \end{pmatrix} X_{t-5} + \begin{pmatrix} -0.2640142 & -0.1315353 \\ 0.6367449 & -0.3526087 \end{pmatrix} X_{t-6} \\
& + \begin{pmatrix} -0.2561217 & -0.01145089 \\ 0.7888435 & -0.31152398 \end{pmatrix} X_{t-7} + \begin{pmatrix} -0.02958839 & 0.1234639 \\ 0.38259764 & -0.1790676 \end{pmatrix} X_{t-8} \\
& + \begin{pmatrix} -0.0002364509 \\ 0.0004573635 \end{pmatrix} + Z_t
\end{aligned}$$

모형의 계수의 유의성을 확인해 본 결과 p-value의 값들은 2개를 제외하고는 모두 유의수준 5% 미만인 것을 확인할 수 있었다.

VAR(8)모형의 향후 2년동안의 예측값을 plot으로 나타낸 것은 아래와 같다.

<그래프11>

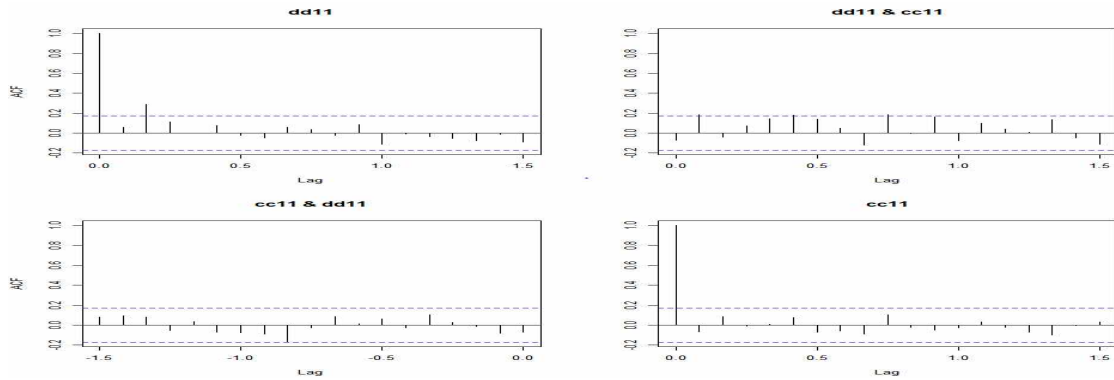


<그래프11>은 VAR(8)모형의 향후 2년 동안의 예측값들과 예측값들의 80%, 95%신뢰 구간을 나타낸 그래프이다. 이 밖에도 오존과 연관이 있는 다른 변수를 이용하여 VAR모형을 fitting 하겠다.

-VAR모형(기온,오존)-

다음은 오존과 기온이 영향이 있다고 판단하여 2010년도 1월부터 2020년 12월까지의 서울시 월 평균 기온데이터를 구하여 VAR모형을 만들고자 하였다. 서울시 월평균 온도 데이터를 auto.arima 함수를 사용한 결과 ARIMA(1,0,0)(0,1,1)이 나온 것을 확인할 수 있었다. 따라서 두 데이터를 SARIMA모형으로 fitting하고 freewhitening을 거쳐서 CCF를 확인한 결과 다음과 같다.

<그래프12>



<그래프12>의 두 번째 그래프를 보면 boundary를 조금 벗어난 lag들이 몇 개 보여서 서울시 월 평균 기온과 오존의 농도는 어느 정도 연관이 있을 수 있다고 판단하였다. 이후 두 데이터를 stationary하게 만들기 위하여 1번 차분을 실시한 다음 단위근 검정을 실시하였다

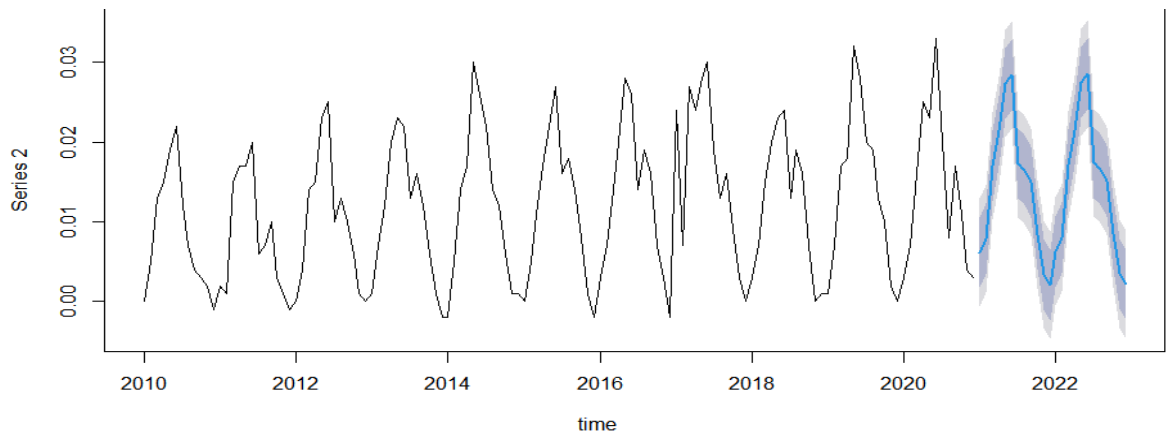
<표6>

<p>Augmented Dickey-Fuller Test data: ozonb Dickey-Fuller=-11.36, Lag order= 0, p-value=0.01 alternative hypothesis: stationary</p>	<p>Augmented Dickey-Fuller Test data: weat2 Dickey-Fuller=-4.3833, Lag order= 0, p-value=0.01 alternative hypothesis: stationary</p>
---	--

<표6>은 2010년도 1월부터 2020년 12월까지의 서울시 월별 오존 데이터와 기온 데이터를 1번 차분한 상태에서 단위근 검정을 실시한 결과이다. 두 데이터 모두 p-value 값이 작은 값을 나왔으므로 stationary하다고 가정하였다.

이후 VARselect 함수를 사용한 결과를 토대로 BIC 기준으로 p값을 8로 VAR모형을 fitting한 모형과 향후 2년 치의 예측치를 plot으로 나타낸 결과는 아래와 같다.

$$\begin{aligned}
 X_t = & \begin{pmatrix} -0.192146517 & -34.2175862 \\ 0.000799926 & -0.8746044 \end{pmatrix} X_{t-1} + \begin{pmatrix} -0.0210686133 & 22.7761139 \\ 0.0005866759 & -0.6364082 \end{pmatrix} X_{t-2} \\
 & + \begin{pmatrix} -0.2784880969 & 41.076872 \\ 0.0001704008 & -0.463793 \end{pmatrix} X_{t-3} + \begin{pmatrix} -0.4609055533 & 110.2753691 \\ -0.0001613865 & -0.4124004 \end{pmatrix} X_{t-4} \\
 & + \begin{pmatrix} -0.2510173411 & 218.2047489 \\ -0.0002843802 & -0.3447201 \end{pmatrix} X_{t-5} + \begin{pmatrix} -0.2967364735 & 153.1773506 \\ -0.0002559201 & -0.5211265 \end{pmatrix} X_{t-6} \\
 & + \begin{pmatrix} -0.3552029944 & 22.0691677 \\ 0.0004463897 & 0.4561972 \end{pmatrix} X_{t-7} + \begin{pmatrix} -4.012144e-01 & -25.5787916 \\ -1.123244e-05 & -0.1836099 \end{pmatrix} X_{t-8} \\
 & + \begin{pmatrix} -0.0252452173 \\ 0.0003183136 \end{pmatrix} + Z_t
 \end{aligned}$$



이후 2개의 VAR모형이 타당하지 R에서 serial.test 함수를 사용하여 검정하였다.

<표7>

Portmanteau Test(asymptotic) data: Residuals of VAR object var1 Chi-squared=53.578, df=32, p-value=0.009782	Portmanteau Test(asymptotic) data: Residuals of VAR object var1 Chi-squared=47.446, df=32, p-value=0.03865
--	---

<표7>은 2개의 VAR모형을 serial test를 실시한 결과인데 p-value 값이 모두 유의수준 5% 미만이므로 타당한 모형이라 볼 수 없다.

5. 모델 SUMMARY

<표8>

	SARIMA(0,0,0)(2,1,0)	VAR(8)(no2, 오존)	VAR(8)(기온, 오존)
MAX & MIN(미래 2년치 평균)	6월:0.041481705 12월:0.012673045	6월:0.029302249 12월:0.00308323	6월:0.028507685 12월:0.002091264
예보 등급	6월:보통 12월:좋음	6월:좋음 12월: 좋음	6월:좋음 12월:좋음
MAX증가폭	3월~4월	2월~3월	2월~3월
MAX감소폭	7월~8월	6월~7월	6월~7월
MSPE for 1-step ahead prediction covariance maxtrix of residuals:	1.686e-05	1.781e-05	1.677e-05

<표7>을 확인해 본 결과 SARIMA모형, VAR모형 다 공통적으로 향 후 2년 동안의 평균 월

별 오존 중 6월에서 가장 높은 농도를 보이고 12월에서 가장 낮은 농도를 보였다. SARIMA 모형의 향후 예보 등급은 보통과 좋음 2가지에 한정 되어 있었고 2개의 VAR모형은 좋음 1 개에 한정 되어 있었다. 증가폭이 가장 큰 월은 SARIMA 모형은 3월에서 4월 VAR 모형들은 2월에서 3월이었다. 감소폭이 가장 큰 월은 SARIMA 모형은 7월에서 8월, VAR모형들은 6월에서 7월이었다. 3가지의 모형 중 covariance matrix of residuals는 기온과 오존 두 변수를 통해 만든 VAR모형이 가장 작아서 최적의 모델로 VAR(8)(기온,오존)이 가장 적합해 보인다. 그렇지만 앞에서 실시한 serial test에서 p-value 값이 작게 나와서 VAR모형들은 적합하지 않았다고 판단하여 결국 SARIMA(0,0,0)(2,1,0)[12] 모형이 가장 적합한 모형이라고 보았다.

6. 결론

본 분석에서는 서울시의 향후 2년 동안의 월 평균 오존 농도 예보 등급을 예측하는 시계열 모델을 만들고자 하였다. 계절 차분 후 일반 차분을 1번 실시한 데이터를 이용하여 SARIMA 모형들 중 BIC가 가장 작은 모델로 SARIMA(0,0,0)(2,1,0)[12]이 선택되었다. 이 밖에도 다른 모형을 만들기 위해 오존과 연관성이 있다고 판단한 이산화질소와 기온을 이용하여 두 가지의 VAR(8) 모형을 fitting 하였다. 세 모델 모두 향후 2년 동안 예측해 보았을 때 월 평균 농도는 6월이 가장 높고 12월이 가장 낮았다. 증가폭에 대해서는 주로 2~4월 봄 쯤이 가장 많이 증가했고 감소폭은 6~8월 여름 쯤이 가장 컸다. covariance matrix of residuals은 기온과 오존 데이터를 활용한 VAR(8) 모형이 가장 작은 값을 보여 이 모형이 가장 최적의 모델로 보였으나 VAR(8) 모형들은 serial test를 한 결과 p-value 값들이 작게 나와 적합한 모형이라고 볼 수 없다고 판단하였다. 따라서 최적의 모델로 SARIMA(0,0,0)(2,1,0)[12]를 선택하였다. 최근에 오존 농도와 비교하기 위해 2022년 5월 23일 오후 3시쯤에 서울시 오존 농도를 확인했었다. 그러나 서울시의 오존 농도의 예보 등급은 전부 '나쁨'이었다. 오존은 오전 농도와 오후 농도로 나뉘는데 오전 농도와 오후 농도의 편차는 굉장히 컸다. 오전 농도의 예보 등급은 대부분 좋음이었지만 오후 농도는 대부분 나쁨에 속했다. 아마 데이터 자체가 월평균 오존 농도라서 현재의 오존 농도를 예측하는 면에 있어서는 분명한 차이를 확인할 수 있었다. 이렇게 월 평균 데이터를 이용하다보니 오후 농도와 차이가 크다는 점이 이 모델의 한계점이라 할 수 있다. 따라서 월 평균 데이터가 아닌 일 평균 데이터를 이용한다면 미래의 오존 농도를 더 잘 예측하는 모델을 fitting 할 수 있다고 생각한다.

7. READER FEEDBACK

- 권용태: 1.VAR에서 계절차분 한 데이터에 대해 adf test가 필요한 것 같습니다.
- 2.Summary에서 MSPE를 비교해주셨으면 좋겠습니다.(표에 sigma만 나와있습니다!)
- 3.VAR에서 기온변수도 고려하면 더 나은 모형이 나타날 것 같습니다!(오존에 영향을 많이 미칠 것으로 예상)
- 4.VAR모형에 대해서 모형이 타당한지에 대한 검정도 필요한 것 같습니다.

-문세인: auto.arima를 포함해서 6가지 모델을 선정하셨는데, 혹시 이때 p,q,P,Q의 조합을 골랐던 기준이나 이유가 있으면 좋을 것 같습니다.