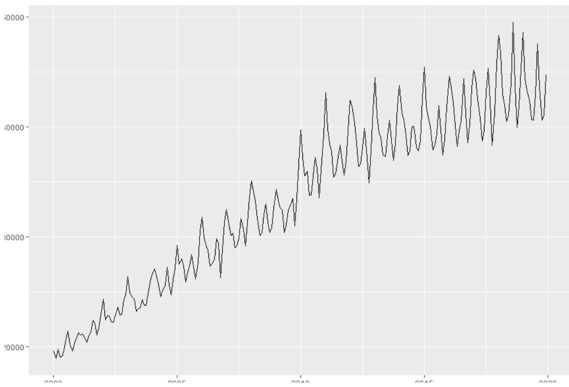
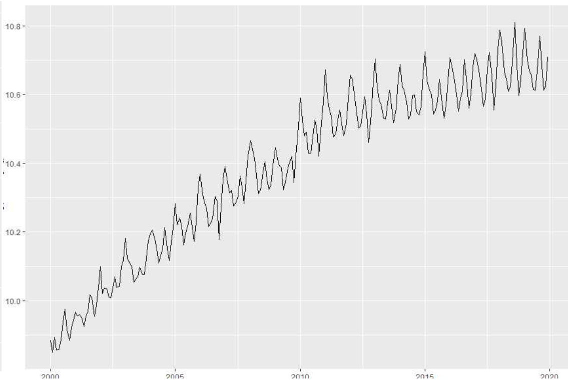
시계열 분석 기말고사
201552001 유승우

```
# 라이브러리 불러오기
library(tidyverse)
library(forecast)

# 데이터 로드 및 2000년 1월부터의 월별 ts객체로 변환
Final <- read.csv("C:/Data/Final.csv")
Final.ts <- ts(Final[,3:4], start=c(2000,1),freq=12)

# 데이터 분포 확인
autoplot(Final.ts,facet=T)

# Elec에 변화폭이 점점 커지니 로그변환을 실시
autoplot((Final.ts[,1]))
autoplot(log(Final.ts[,1]))
```

(변환전)



(변환후)

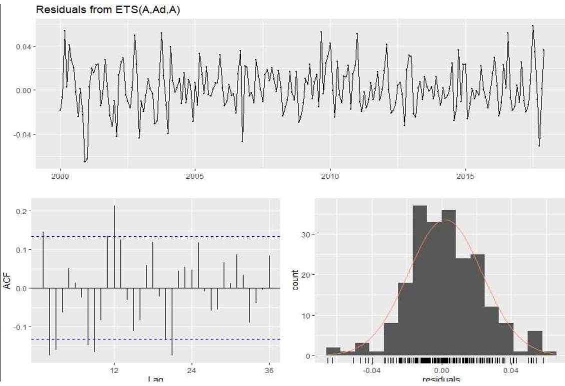


로그변환 후 변화폭이 좀 더 줄어들었다.

```
# 데이터 분할
train <- window(Final.ts, end=c(2017,12))
test <- window(Final.ts, start=c(2018,1))
```

# ets모형
fit_ets <- ets(train[,1],lambda=0)
fit_ets
checkresiduals(fit_ets)

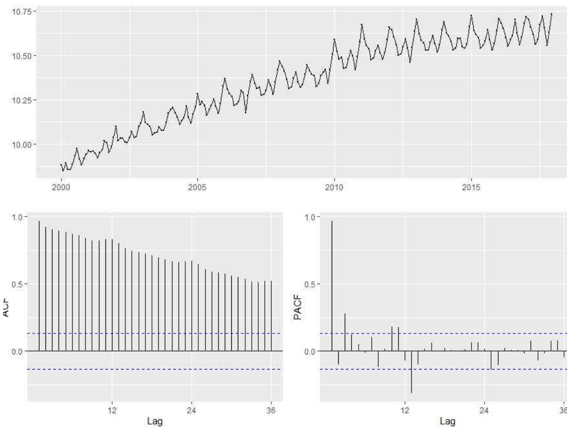
```
> fit_ets
ETS(A,Ad,A)

Call:
 ets(y = train[, 1], lambda = 0)

  Box-Cox transformation: lambda= 0

  Smoothing parameters:
    alpha = 0.5669
    beta  = 0.0041
    gamma = 1e-04
    phi   = 0.98

  Initial states:
    l = 9.7985
    b = 0.0115
    s = 0.0438 -0.0233 -0.073 -0.0208 0.0253 -0.0132
           -0.0438 -0.0523 -0.0041 0.0226 0.0455 0.0934

  sigma:  0.0221
```




```
> checkresiduals(fit_ets)

        Ljung-Box test

data:  Residuals from ETS(A,Ad,A)
Q* = 73.004, df = 7, p-value = 3.643e-13

Model df: 17.   Total lags used: 24
```

ETS(A,Ad,A)가 적합되었고, 레벨에는 변동이 있었고, 기울기와 계절성은 변동이 거의 없는 것으로 나타났다.
또한 잔차에 대한 독립성 검정결과, 독립이 아니라고 나타났지만, 예측값에는 큰 문제가 없다.

# arima
# 정상성 & 차분 확인
ggtsdisplay(log(train[,1]))



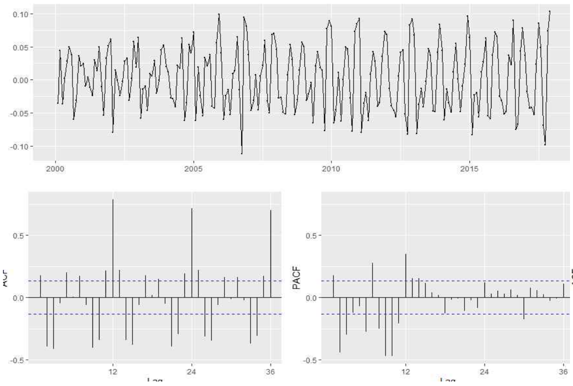
ACF가 감소하는 추세를 보아, 정상성을 나타내지 못하는 것으로 보인다
따라서 차분이 필요하다

train_l <- log(train[,1])
ndiffs(train_l)
nsdiffs(train_l)

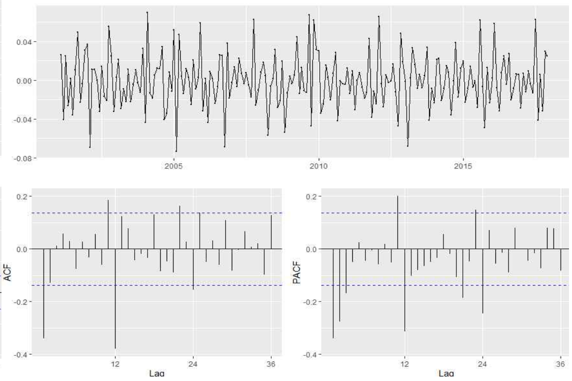
d_train <- diff(train_l)
ggtsdisplay(d_train)
ggtsdisplay(diff(d_train,lag=12))
(단위근검정)

```
> train_l <- log(train[,1])
> ndiffs(train_l)
[1] 1
> nsdiffs(train_l)
[1] 1
>
```

(일반차분)



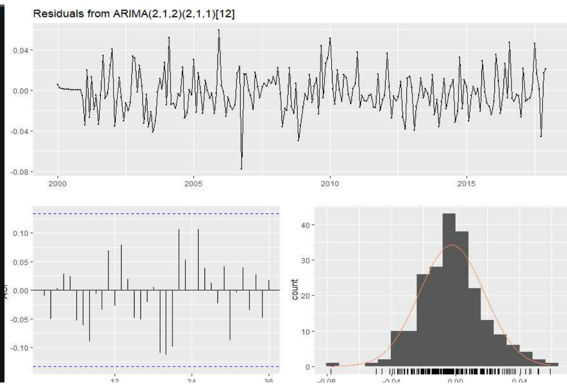
(계절차분)



일반차분결과, 12, 24, 36에서 acf가 높아지는 것으로 보아 계절변동이있으므로, 계절차분을 실시한다. 일반차분과 계절차분 모두 실행한 결과 어느정도 정상성을 만족한다고 나타났으며, 단위근검정결과 역시 각각 1회씩 차분하라고 나타났다.

```
# arima모형적합
fit_arima <- auto.arima(train[,1],lambda=0)
fit_arima
checkresiduals(fit_arima)
```

```
> fit_arima
Series: train[, 1]
ARIMA(2,1,2)(2,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
         ar1      ar2      ma1     ma2      sar1     sar2     sma1
      1.0352  -0.2185  -1.4303  0.4930  -0.1491  -0.2156  -0.4867
s.e.  0.5565   0.2565   0.5513  0.4406   0.1411   0.0993   0.1386

sigma^2 estimated as 0.0004571:  log likelihood=493.58
AIC=-971.16   AICc=-970.42   BIC=-944.66
> checkresiduals(fit_arima)

        Ljung-Box test

data:  Residuals from ARIMA(2,1,2)(2,1,1)[12]
Q* = 20.372, df = 17, p-value = 0.2557

Model df: 7.   Total lags used: 24

>
```



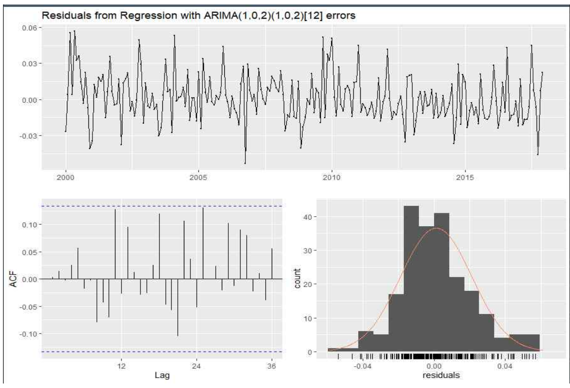
arima모형적합결과 ARIMA(2,1,2)(2,1,1)이 적합되었고, 잔차의 독립성 검정결과 잔차는 백색잡음이라고 할 수 있다.

모형식은 (1-1.0352B +0.2185B^2)(1+0.1491B^12,+0.2156B^24)(1-B^12)(1-B)Yt = (1-1.4303B +0.4930B^2)(1-0.4867B^12)e이다.

```
# arma 오차회귀모형
# 설명변수 생성
TIME <- time(train[,1])
MONTH <- seasonaldummy(train[,1])
```

train의 시점에 따라 Time변수를 생성하고, MONTH에는 계절더미변수를 할당했다

```
# Dummy 오차 회귀
new1 <- cbind(TIME,MONTH)
fit_arma1 <- auto.arima(train[,1],
                    xreg=new1,lambda=0)
summary(fit_arma1)
checkresiduals(fit_arma1)
```




```
Series: train[, 1]
Regression with ARIMA(1,0,2)(1,0,2)[12] errors
Box Cox transformation: lambda= 0

Coefficients:
         ar1      ma1      ma2    sar1     sma1     sma2  intercept    TIME  MONTH.Jan  MONTH.Feb  MONTH.Mar
      0.9834  -0.3351  -0.2091  0.6684  -0.4280  -0.0454   -83.6451  0.0468     0.0472    -0.0013    -0.0222
s.e.  0.0206   0.0758   0.0762  0.1901   0.2037   0.0934    11.3340  0.0056     0.0078     0.0094     0.0098
      MONTH.Apr  MONTH.May  MONTH.Jun  MONTH.Jul  MONTH.Aug  MONTH.Sep  MONTH.Oct  MONTH.Nov
        -0.0487    -0.0972    -0.0857    -0.0528    -0.0129    -0.0584    -0.1130    -0.0648
s.e.     0.0100     0.0101     0.0102     0.0102     0.0101     0.0098     0.0094     0.0077

sigma^2 estimated as 0.0004392:  log likelihood=536.42
AIC=-1032.84   AICc=-1028.54   BIC=-965.34

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 20.48209 645.8355 497.7789 0.1003524 1.548384 0.3487665 -0.01729065
> checkresiduals(fit_arma1)
```

dummy로 만든 오차회귀 모형은 ARMA(1,0,2)(1,0,2)[12] error 모형이 적합되었다.

```
> checkresiduals(fit_arma1)

        Ljung-Box test

data:  Residuals from Regression with ARIMA(1,0,2)(1,0,2)[12] errors
Q* = 21.797, df = 5, p-value = 0.0005722

Model df: 19.   Total lags used: 24

> |
```

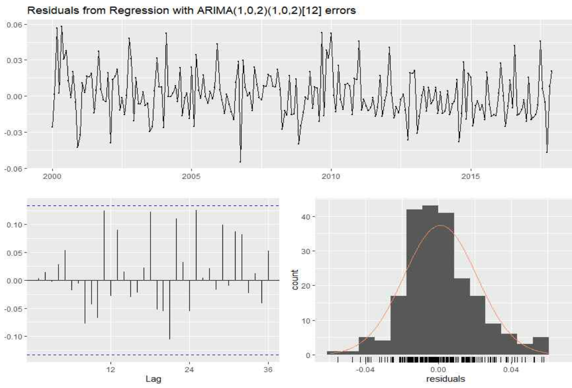
독립성 검정결과 잔차는 독립이 아니라고 나타났다.

```
# k별 AICc 값
Time <- time(train[,1])
res <- vector("numeric",6)
for (i in seq(res)){
  xreg <- cbind(Time, fourier(train[,1],K=i))
  fit <- auto.arima(train[,1],lambda=0 ,xreg=xreg)
  res[i] <- fit$aicc
}
which.min(res)
```

```
+ }
> which.min(res)
[1] 5
> res
[1]  -977.3623 -1005.9512 -1025.7411 -1022.3532 -1030.0900 -1028.5353
>
```

푸리에 변수를 사용하여 주기별 모형의 aicc를 확인한 결과 5주기에서 가장 낮게 나타났다. 따라서 더미변수를 사용한 모델보다 변수의 개수를 줄여서 분석할 수 있으므로, 푸리에 변수를 사용하여 분석한다.

```
# Fourier 오차 회귀
Time <- time(train[,1])
Fourier <- fourier(train[,1], K=5)
fit_arma2 <- auto.arima(train[,1],xreg=cbind(Time,Fourier),
                        lambda=0)
summary(fit_arma2)
checkresiduals(fit_arma2)
```




```
> fit_arma2
Series: train[, 1]
Regression with ARIMA(1,0,2)(1,0,2)[12] errors
Box Cox transformation: lambda= 0

Coefficients:
         ar1      ma1      ma2    sar1     sma1     sma2  intercept    Time  Fourier.S1-12  Fourier.C1-12
      0.9830  -0.3428  -0.2029  0.6884  -0.4420  -0.0421   -83.6511  0.0468         0.0224         0.0289
s.e.  0.0205   0.0750   0.0758  0.1726   0.1887   0.0921    11.4412  0.0057         0.0042         0.0041
      Fourier.S2-12  Fourier.C2-12  Fourier.S3-12  Fourier.C3-12  Fourier.S4-12  Fourier.C4-12  Fourier.S5-12
             0.0439        -0.0002         0.0052         0.0231         0.0012         0.0009         0.0007
s.e.         0.0032         0.0032         0.0027         0.0027         0.0022         0.0022         0.0019
      Fourier.C5-12
            -0.0092
s.e.         0.0019

sigma^2 estimated as 0.0004384:  log likelihood=535.98
AIC=-1033.97   AICc=-1030.09   BIC=-969.84
```

```
	Ljung-Box test

data:  Residuals from Regression with ARIMA(1,0,2)(1,0,2)[12] errors
Q* = 21.612, df = 6, p-value = 0.001423

Model df: 18.   Total lags used: 24
```

푸리에변수를 사용한 모형은 ARIMA(1,0,2)(1,0,2)[12] errors가 적합되었다.
독립성 검정결과 잔차는 독립이 아니라고 나타났다.

```
# Dynamic 회귀
fit_dy <- auto.arima(train[,1], lambda=0,
                     xreg=train[,2])
checkresiduals(fit_dy)
```

```
> fit_dy
Series: train[, 1]
Regression with ARIMA(2,1,2)(2,1,1)[12] errors
Box Cox transformation: lambda= 0

Coefficients:
         ar1      ar2      ma1     ma2     sar1     sar2     sma1    xreg
      1.0257  -0.2193  -1.4191  0.4850  -0.1520  -0.2195  -0.4836  -3e-04
s.e.  0.6027   0.2728   0.5986  0.4784   0.1419   0.0999   0.1402   5e-04

sigma^2 estimated as 0.0004585:  log likelihood=493.78
AIC=-969.57   AICc=-968.63   BIC=-939.75
>
```

다이나믹 회귀 모형 적합결과 ARIMA(2,1,2)(2,1,1)[12] errors 모형이 적합되었다.

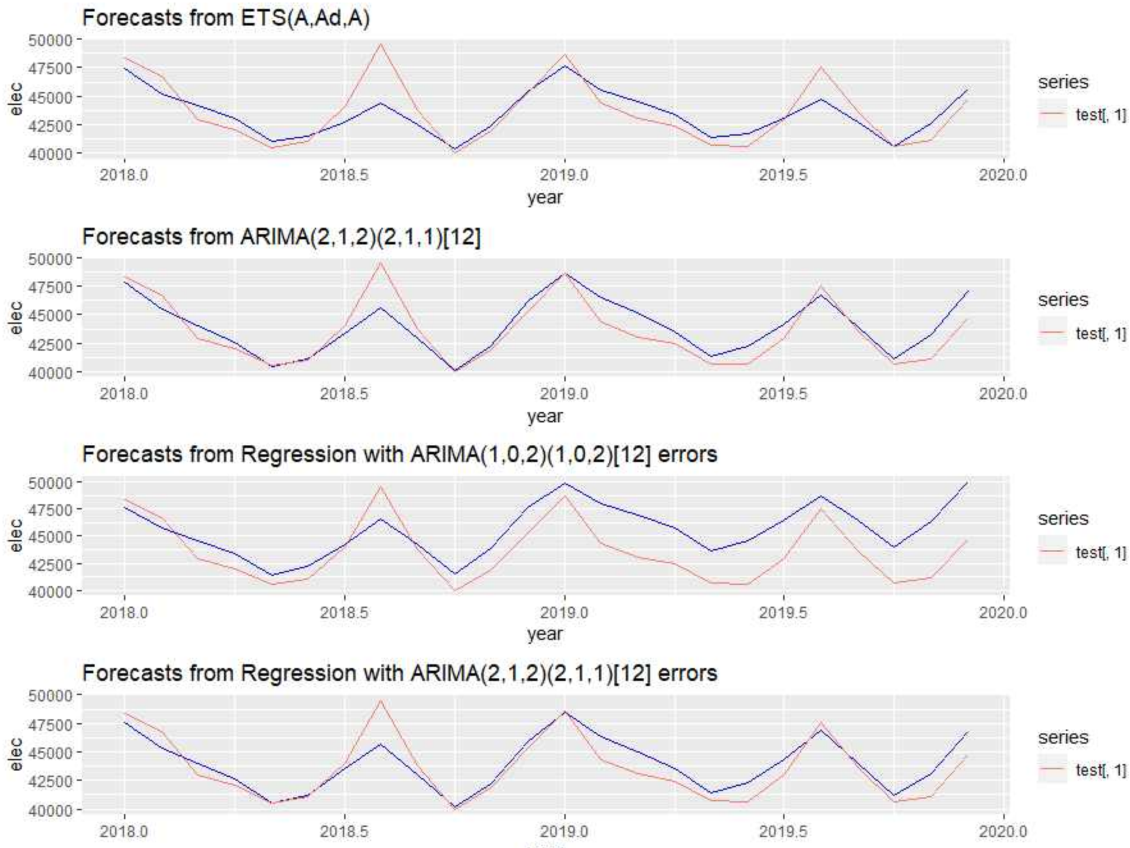
다이나믹 회귀모형의 RMSE가 가장작게 나타났고, MASE도 1보다 작으며, MAPE에도 문제가 없어보인다.

```
# 예측
# ets 예측
fc_ets <- forecast(fit_ets,h=length(test[,1]))


# arima예측
fc_arima <- forecast(fit_arima, h=length(test[,1]))

# ARMA 오차 회귀 예측
new3 <- cbind(Time = time(test[,1]),
              Fourier=fourier(test[,1], K=5))

fc_arma <- forecast(fit_arma2, xreg=new3)


# dynamic 예측
test[,2] <- mean(train[,2])
test

fc_dy <- forecast(fit_dy, xreg=test[,2])
fc_dy

# 예측비교
accuracy(fc_ets, test[,1])
accuracy(fc_arima, test[,1])
accuracy(fc_arma, test[,1])
accuracy(fc_dy, test[,1])
```

```
> # 예측비교
> accuracy(fc_ets, test[,1])
                    ME      RMSE       MAE       MPE     MAPE      MASE      ACF1 Theil's U
Training set  77.20854  688.4691  520.5089 0.1848550 1.626660 0.3646921 0.1201071        NA
Test set     127.96571 1505.5998 1111.3788 0.1460883 2.476144 0.7786823 0.3739542 0.5003199
> accuracy(fc_arima, test[,1])
                    ME      RMSE       MAE        MPE     MAPE      MASE        ACF1 Theil's U
Training set  -66.83785  677.7817  509.1954 -0.2483005 1.558187 0.3567654 -0.00191566        NA
Test set     -372.31443 1385.0482 1045.1850 -0.9499143 2.364654 0.7323040  0.42522560 0.4670685
> accuracy(fc_arma, test[,1])
                     ME      RMSE       MAE         MPE     MAPE      MASE        ACF1 Theil's U
Training set    20.26253  646.6326  495.6957  0.09911415 1.540575 0.3473068 -0.01458536        NA
Test set     -1946.26676 2736.2255 2342.9135 -4.62228469 5.436259 1.6415513  0.60492173 0.9449626
> accuracy(fc_dy, test[,1])
                    ME      RMSE      MAE        MPE     MAPE      MASE         ACF1 Theil's U
Training set  -67.23645  674.7951  508.777 -0.2492254 1.559951 0.3564722 -0.002879673        NA
Test set     -358.39128 1344.3493 1033.185 -0.9251010 2.338288 0.7238964  0.406696543 0.4516764
>
```

다이나믹 회귀모형의 RMSE가 가장작게 나타났고, MASE도 1보다 작으며, MAPE에도 문제가 없어보인다.

```
plot_ets <- autoplot(fc_ets,include=0) +
  autolayer(test[,1]) +
  labs(x="year",y="elec")
plot_arima <- autoplot(fc_arima,include=0) +
  autolayer(test[,1]) +
  labs(x="year",y="elec")
plot_arma <- autoplot(fc_arma,include=0) +
  autolayer(test[,1]) +
  labs(x="year",y="elec")
plot_dy <- autoplot(fc_dy,include=0) +
  autolayer(test[,1]) +
  labs(x="year",y="elec")

Rmisc::multiplot(plot_ets,plot_arima,plot_arma,plot_dy)
```



그림보았을 때 다이나믹 회귀모형이 가장 가까워보인다