pre-assignment

Jakub Skrajny

27 04 2021

Libraries

```
library(openxlsx)
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(forecast)
library(EnvStats)
```

```
##
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

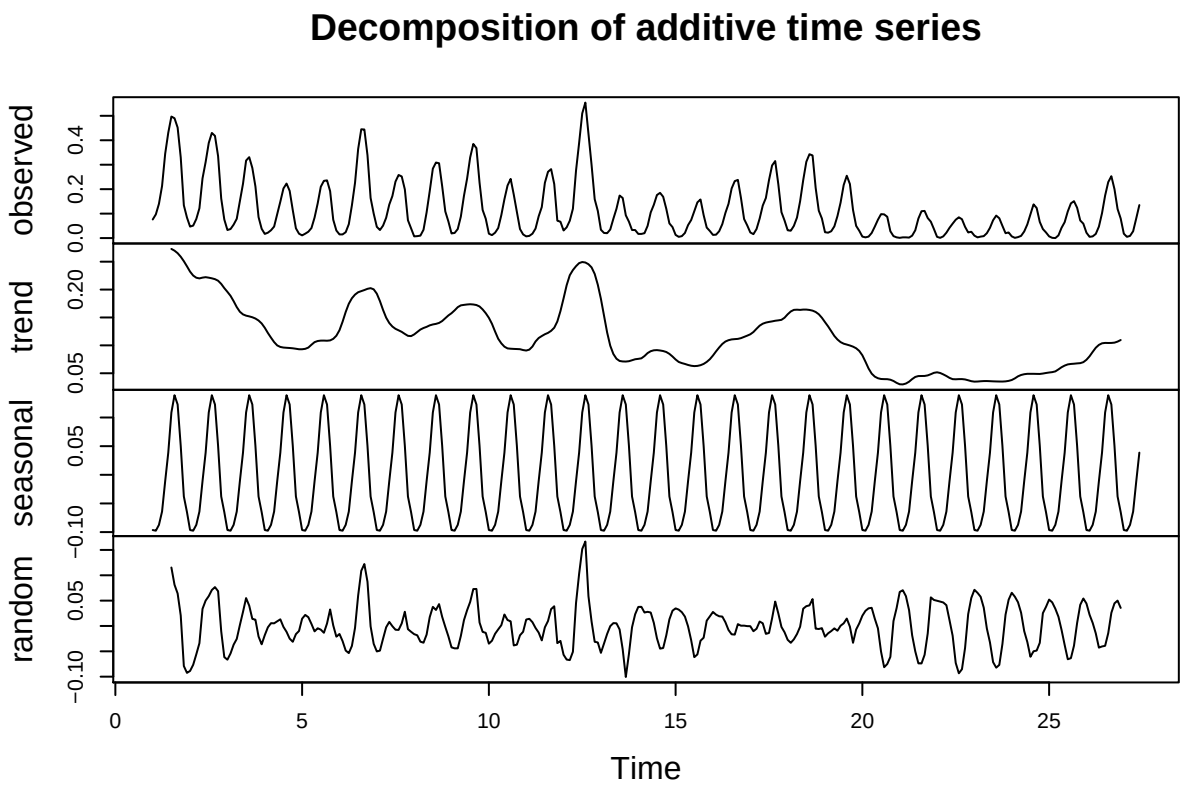
```
## The following object is masked from 'package:base':
##
##     print.default
```

Read data (date = number of days from 1900)

```
data <- read.xlsx("Lions_Den_data.xlsx")
timeseries <- unlist(data[2])
dates <- unlist(data[1])
```

Decomposition

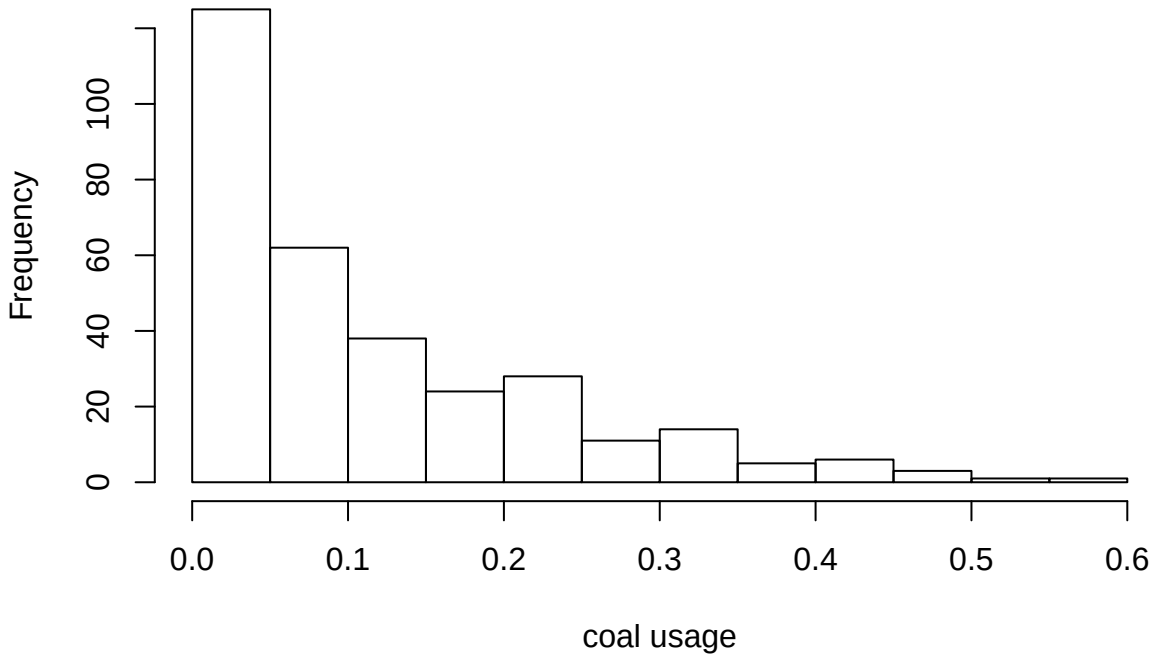
```r
ts <- ts(timeseries, frequency = 12)
decompose <- decompose(ts, "additive")
plot(decompose)
```



Outliers

Rosner's test suggest that there is only one outlier.

```r
hist(timeseries,
  xlab = "coal usage",
)
```




```
rosnerTest(timeseries, k = 3)
```

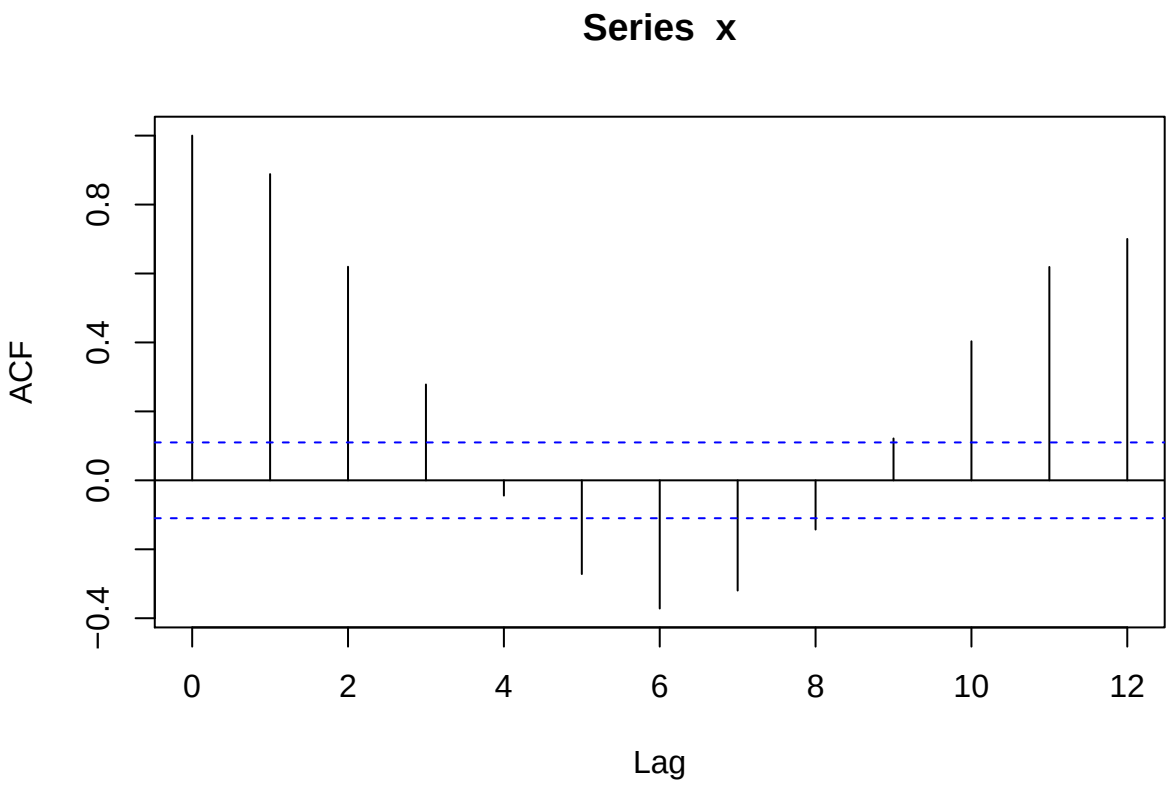
```
##
## Results of Outlier Test
## -------------------------
##
## Test Method:                     Rosner's Test for Outliers
##
## Hypothesized Distribution:       Normal
##
## Data:                            timeseries
##
## Sample Size:                     318
##
## Test Statistics:                 R.1 = 3.782021
##                                  R.2 = 3.487550
##                                  R.3 = 3.445420
##
## Test Statistic Parameter:        k = 3
##
## Alternative Hypothesis:          Up to 3 observations are not
##                                  from the same Distribution.
##
## Type I Error:                    5%
##
## Number of Outliers Detected:     1
##
##   i    Mean.i      SD.i Value Obs.Num    R.i+1 lambda.i+1 Outlier
## 1 0 0.1170692 0.1155284 0.554     140 3.782021   3.739949    TRUE
```

```
## 2 1 0.1156909 0.1130619 0.510     139 3.487550   3.739067   FALSE
## 3 2 0.1144430 0.1110335 0.497       7 3.445420   3.738181   FALSE
```

Autocorrelation and stationarity analysis

We can see that time-series is already stationary.

```
x <- timeseries
#autocorrelation
acf(x, lag.max = 12)
```



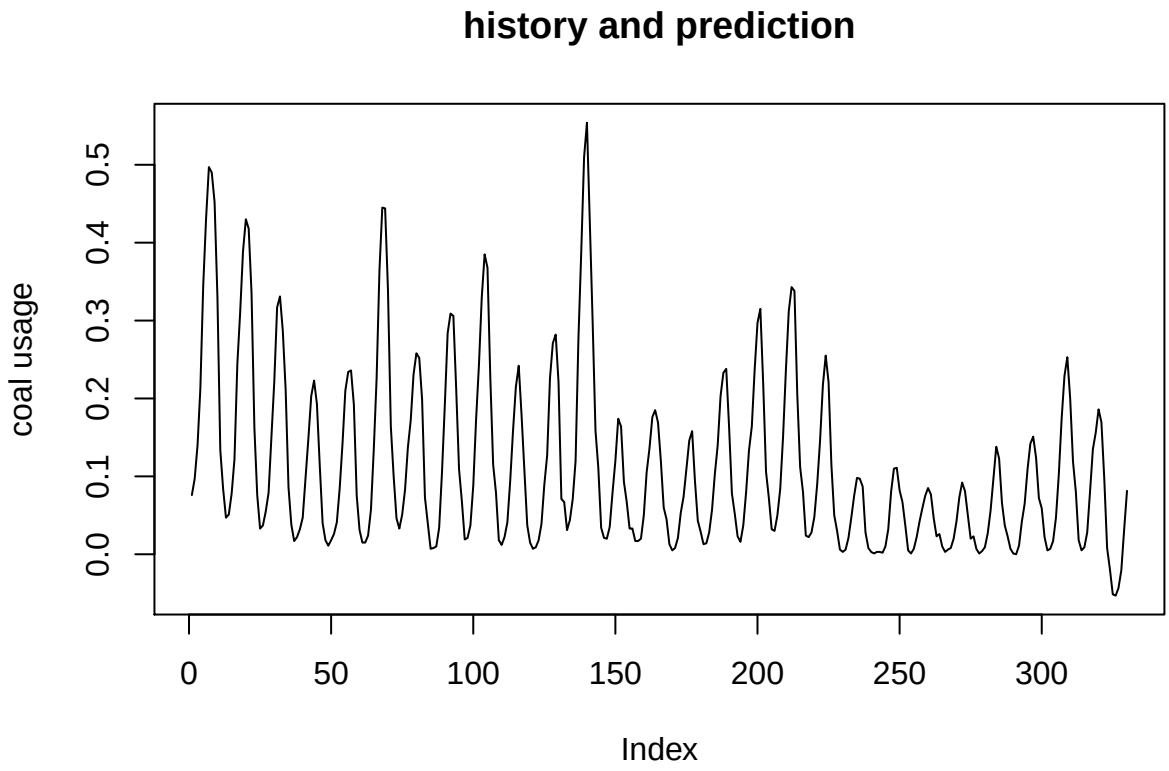

```
#stationary test
adf.test(x)
```

```
## Warning in adf.test(x): p-value smaller than printed p-value
```

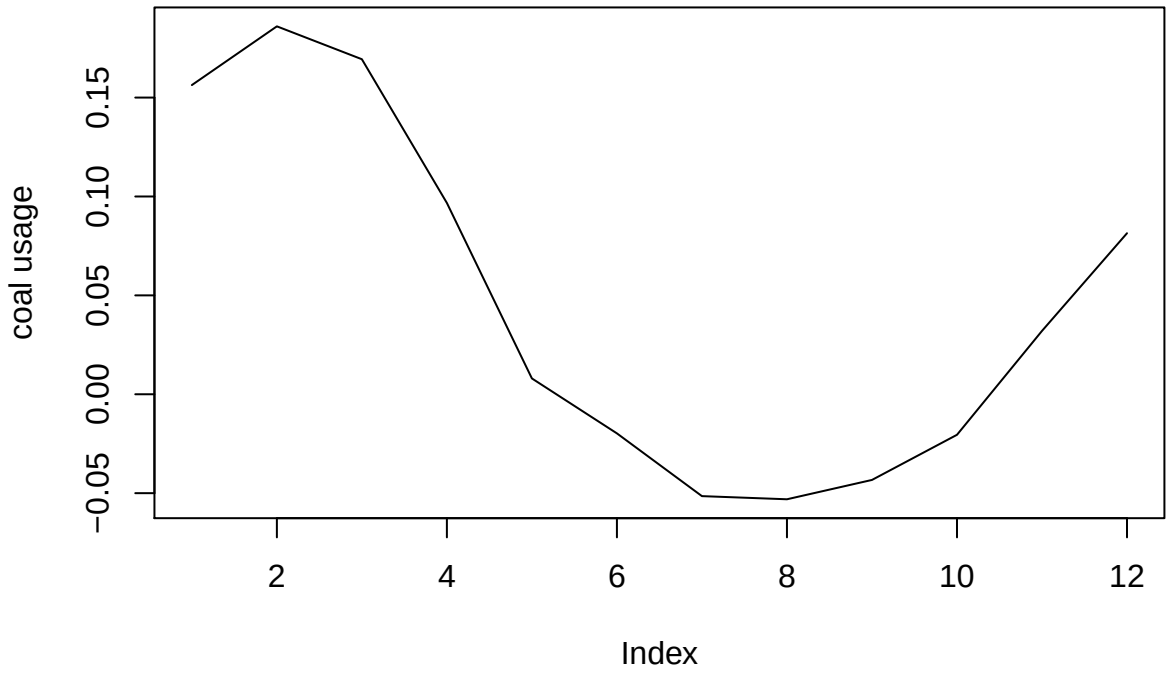
```
##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -4.9571, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Prediction using decomposition and linear regression

```
#trend
x <- as.numeric(decompose$trend)
x <- x[!is.na(x)]
time <- seq(1, length(x))
df <- data.frame(x, time)
model <- lm(x ~ time, data = df)
p <- as.data.frame(seq(length(x)+1, length(x)+12))
colnames(p) <- "time"
trend_prediction <- as.numeric(predict(model, newdata=p))
concat_trend <- c(x, trend_prediction)
#seasonal
seasonal_prediction <- as.numeric(tail(decompose$seasonal, n=12))
concat_seasonal <- c(
  as.numeric(decompose$seasonal),
  seasonal_prediction
)
# overall predicion
overall_prediction <- trend_prediction + seasonal_prediction
plot(c(decompose$x, overall_prediction), type="l", ylab="coal usage",
     main="history and prediction")
```




```
plot(overall_prediction, type="l", main="prediction for next year",
     ylab="coal usage")
```




```
overall_prediction
```

```
##  [1]  0.156327689  0.186011968  0.169390158  0.096725078  0.008003908
##  [6] -0.019770146 -0.051498174 -0.053085561 -0.043319615 -0.020492003
## [11]  0.032028943  0.081419889
```