Group Project

Peter Ye

2024-04-15

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tseries)
```

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

```r
library(ggplot2)
library(TSA)
```

```
##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar
```

```r
library(fpp)
```

```
## Loading required package: forecast
```

```
## Registered S3 methods overwritten by 'forecast':
##   method       from
##   fitted.Arima TSA
##   plot.Arima   TSA

## Loading required package: fma

## Loading required package: expsmooth

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(forecast)
library(lmtest)
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
library(readxl)
library(writexl)
```

Data Import

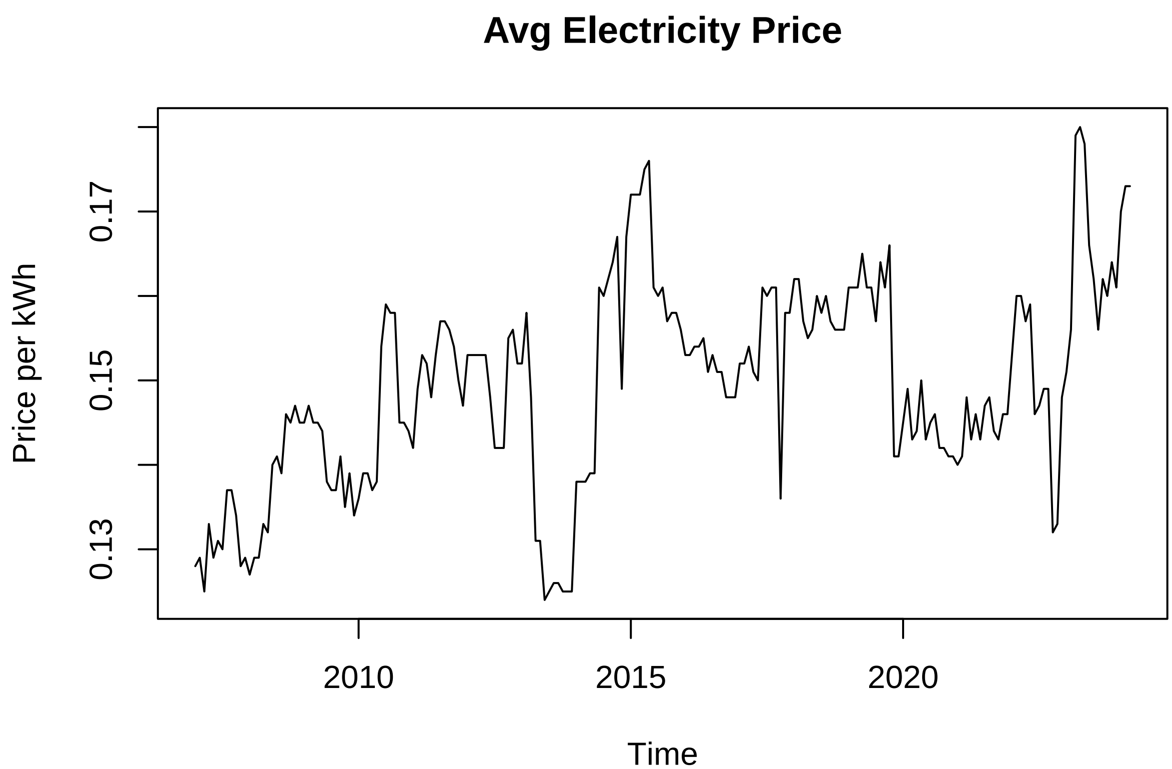
```r
# Read csv
electricity <- read.csv("APUS23A72610.csv")
```

```r
# Convert the DATE column to a Date type & Rename price column
electricity <- electricity %>%
  mutate(DATE = as.Date(DATE)) %>%
  rename(price = APUS23A72610)
```

```r
# Filter the data from 2007-01-01 to the most recent
electricity <- electricity %>%
  filter(DATE >= as.Date("2007-01-01"))
```

```r
# Filter the price data and create a time series object
electricity$price <- as.numeric(as.character(electricity$price))
electricity_price = electricity$price
electricity_price_ts = ts(electricity_price, frequency = 12, start=c(2007, 1), end=c(2024,3))
```

```r
# Plot the graph
plot(electricity_price_ts, xlab="Time", ylab="Price per kWh", main="Avg Electricity Price")
```

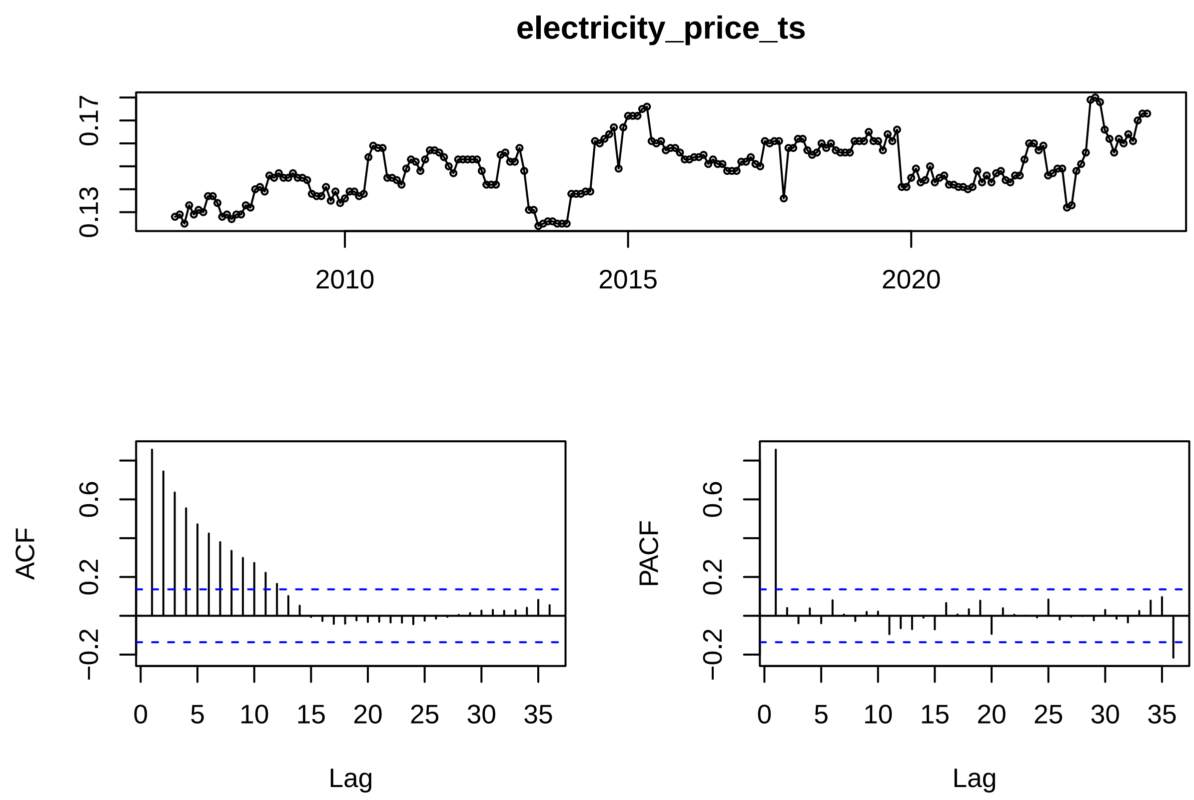


Analysis

```r
# Perform the Augmented Dickey-Fuller test
adf_test_result <- adf.test(electricity_price_ts, alternative = "stationary")
adf_test_result
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  electricity_price_ts
## Dickey-Fuller = -3.3743, Lag order = 5, p-value = 0.06027
## alternative hypothesis: stationary
```

```r
tsdisplay(electricity_price_ts)
```



- There's no clear long-term trend. The variability does appear to have some periodicity, with a noticeable pattern of ups and downs. However, the fluctuations don't seem to be strictly regular, and there may be outlier peaks that could be due to specific events or anomalies.
- The ACF plot shows a sharp drop after the first lag, indicating that the value at one time point is most strongly related to its immediate predecessor. However, there are several other points outside the confidence bounds at later lags, suggesting some degree of seasonal or cyclic behavior
- The PACF shows a significant autocorrelation at the first lag and perhaps at the second, but subsequent lags fall within the confidence bounds. This suggests that an AR(1) or AR(2) model might be a good fit for the non-seasonal part of the model.

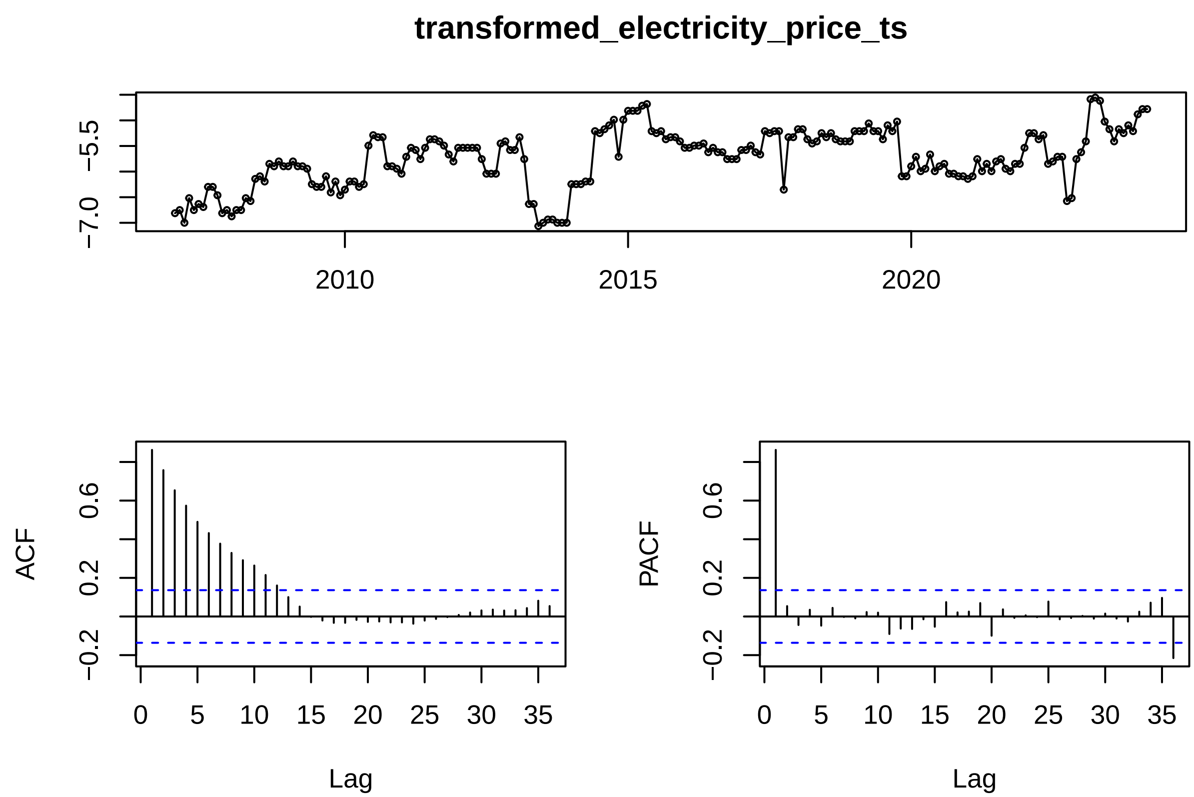
```r
# Find lambda
lambda <- BoxCox.lambda(electricity_price_ts)

# Apply the Box-Cox transformation
transformed_electricity_price_ts <- BoxCox(electricity_price_ts, lambda)

lambda
```

```
## [1] -0.9999242
```

```r
tsdisplay(transformed_electricity_price_ts)
```



Model

Auto-arima without lambda

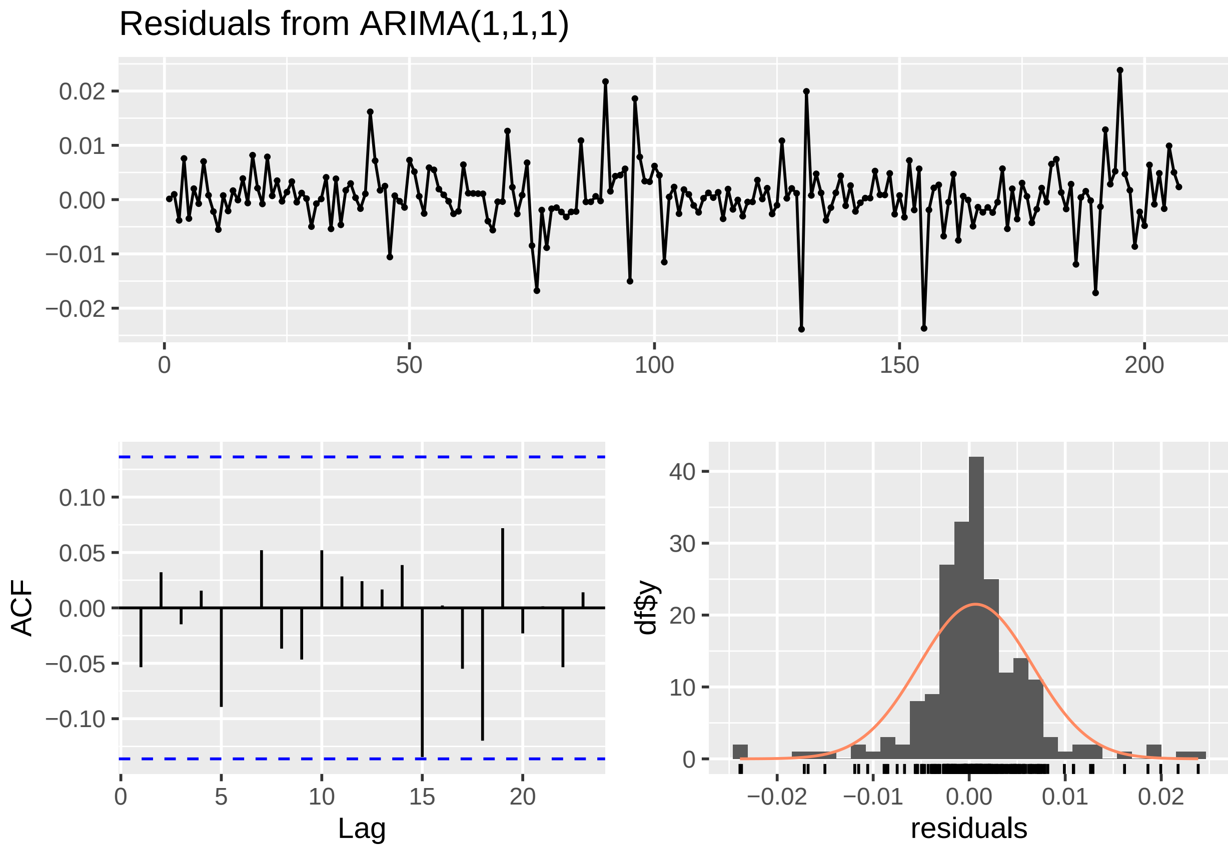
```r
# Create auto arima
arima_model <- auto.arima(electricity_price, seasonal=TRUE, stepwise = FALSE)
```

```r
# Summary of the fit
summary(arima_model)
```

```
## Series: electricity_price
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1      ma1
##       0.8432  -0.9691
## s.e.  0.0526   0.0265
##
## sigma^2 = 3.573e-05:  log likelihood = 763.05
## AIC=-1520.1   AICc=-1519.98   BIC=-1510.11
```

```
##
## Training set error measures:
##                        ME        RMSE         MAE       MPE     MAPE     MASE
## Training set 0.0006682183 0.005934147 0.003823606 0.3236928 2.551607 1.037764
##                     ACF1
## Training set -0.05353383
```

```
checkresiduals(arima_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)
## Q* = 4.5772, df = 8, p-value = 0.8017
##
## Model df: 2.   Total lags used: 10
```

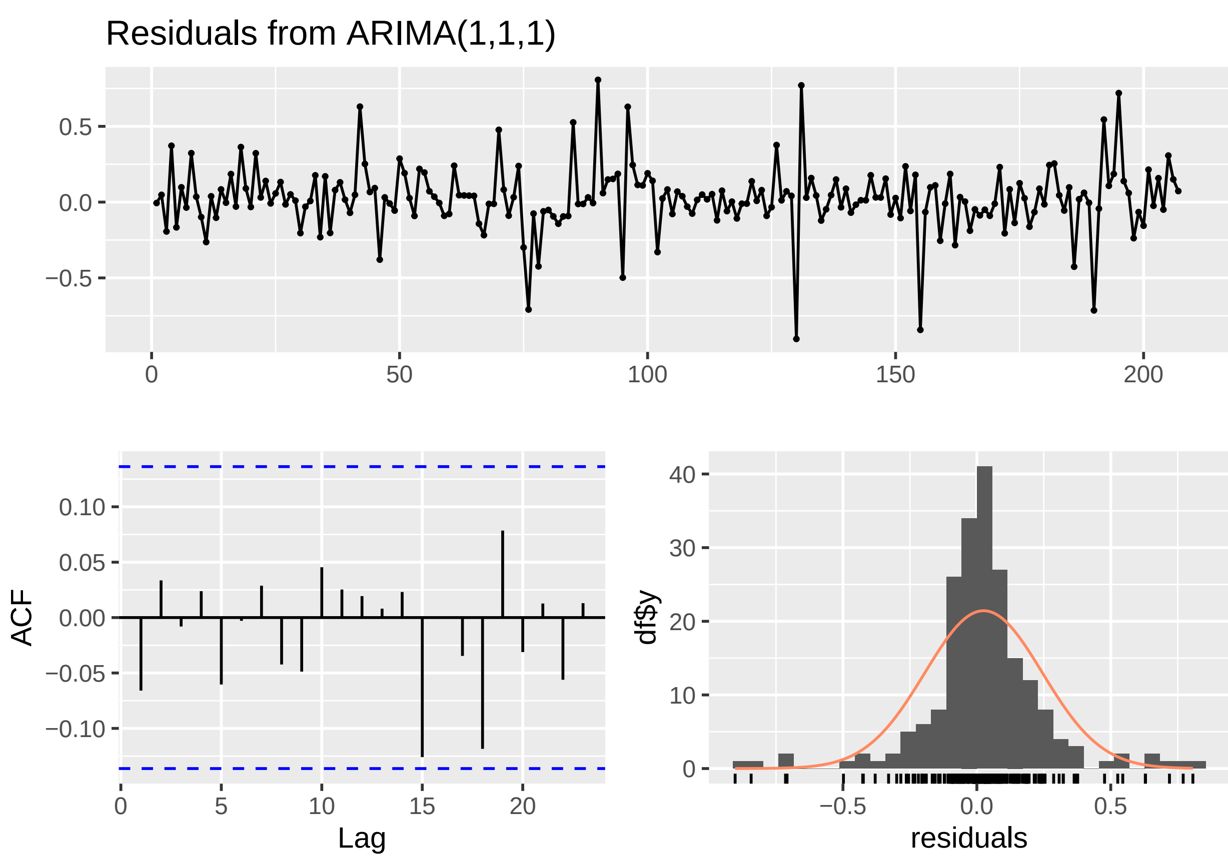
```
# Create auto arima
arima_lambda_model <- auto.arima(electricity_price, seasonal=TRUE, stepwise = FALSE, lambda = "auto")
```

```
# Summary of the fit
summary(arima_lambda_model)
```

```
## Series: electricity_price
```

```
## ARIMA(1,1,1)
## Box Cox transformation: lambda= -0.8999268
##
## Coefficients:
##          ar1      ma1
##       0.8494  -0.9695
## s.e.  0.0514   0.0259
##
## sigma^2 = 0.04933:  log likelihood = 18.34
## AIC=-30.67   AICc=-30.56   BIC=-20.69
##
## Training set error measures:
##                        ME        RMSE         MAE     MPE     MAPE     MASE
## Training set 0.0008088525 0.005951075 0.003838664 0.41703 2.556409 1.041851
##                     ACF1
## Training set -0.05434927
```

```
checkresiduals(arima_lambda_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)
## Q* = 3.6088, df = 8, p-value = 0.8906
##
## Model df: 2.   Total lags used: 10
```

- These residuals do not show any obvious patterns or trends
- All the autocorrelations for the lags are within the bounds, suggesting that there is no significant autocorrelation left in the residuals.
- The residuals here appear to be fairly normally distributed, with perhaps a slight skew to the left.