NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

R-Programming and Application in Finance
Take-home Exam

Energy curves seasonal patterns: determining an accurate model

Elizaveta Kurenina
1st year Master Student
Strategic Corporate Finance Master Program
Faculty of Economic Sciences
evkurenina@edu.hse.ru

Moscow
2021

Introduction

Despite a rapid development of alternative sources of energy, oil still remains one of the most valuable energy sources. Being also one of the underlying assets, determining market movements, it does not show the stable price dynamics and clearly observable fluctuation patterns. Thus, complex approach for modeling oil price-curves is needed for accurate and reliable forecasting, vital both for individual investors and for global market stability.

The proposed practical research aims to fit an appropriate model, which would handle with Brent Crude Oil Price predictions, clearly demonstrating seasonal patterns involving market anomalies.

Data description

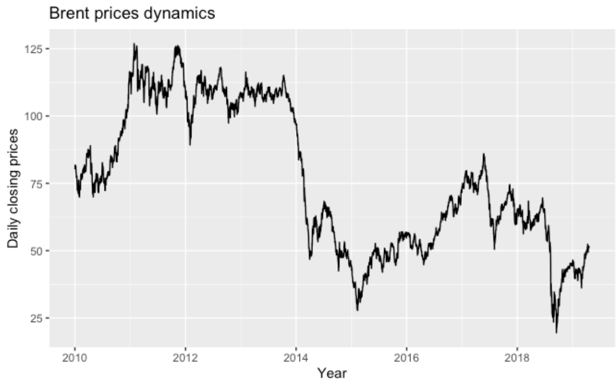
For research purposes an open data from https://www.finam.ru was employed. The initial dataset contains 3396 observations of **daily** Brent Crude Oil Prices from 2010 to 2019, including open, close, high and low prices, as well as trading volume. In the presented paper closing price and volume are investigated.

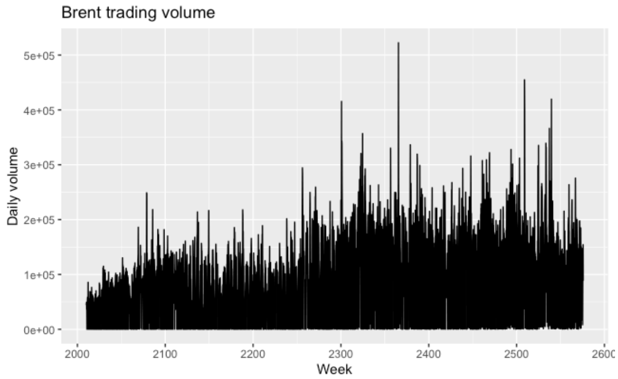
The initial empirical analysis of the data reveals a clear **weekly seasonal pattern for the trading volume** variable [Table 1].

However, the behavior price curves derived from daily observations is not transparent in such away. As seasonality may occur over the big period of time, plotting the data over many time-periods only could capture assumed seasonality of oil prices.

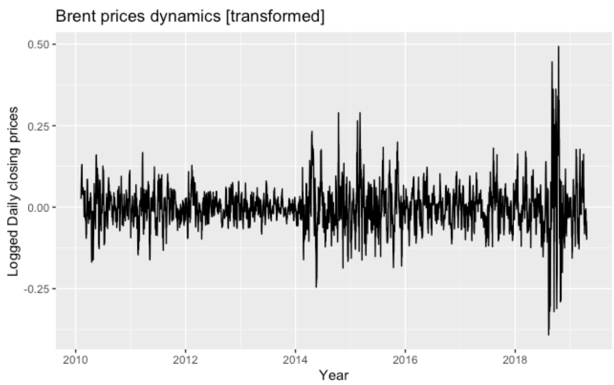
Seasonality validation

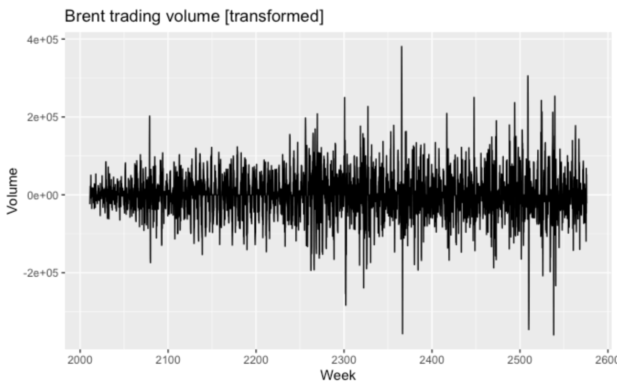
To trace potential seasonality the initial data was plotted.





Clearly, if volume observations do produce a weekly pattern, they first need to be transformed applying differences throughout the appropriate lags. For daily prices demanded difference predicted clearly by correlograms' patterns – a continuous decline for ACF and a spike with the following fluctuating waves for PACF [Picture 1]. In addition, daily prices' series would be more accurate with log-transformation, smoothing minimal-step price changes differences as well. As supposed, after applied transformation the series look more stationary:

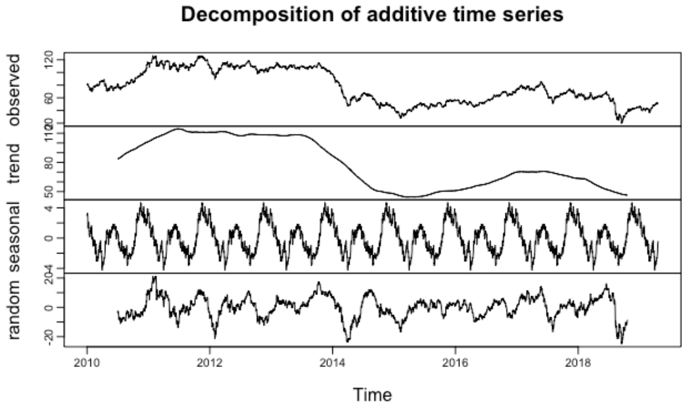




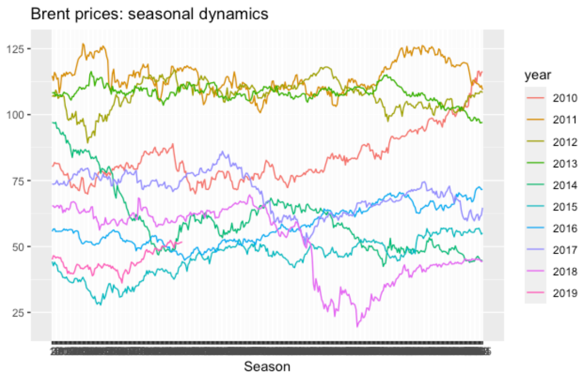
In the following analysis it worth to examine prices and trading volume behavior in separate sections because of different series structure.

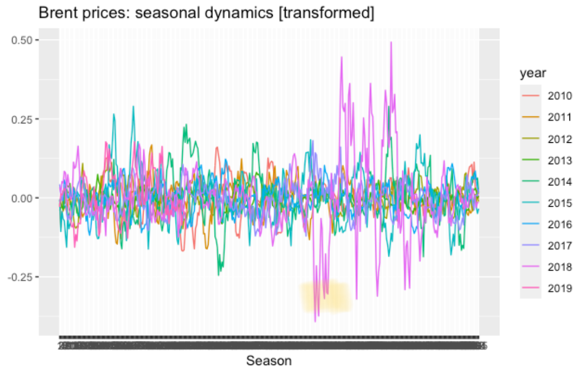
Price dynamics modeling

Let us start with the price dynamics, which is the series with frequency 365.25 – the average length of the year in days. Basic decomposition outlines no stable trend, but clearly observable seasonality.



Despite the fact that no uniform behavior within the seasons through different years is observed, seasonal visualization of transformed data reveals an interesting insight on market environment: in 2018 […]





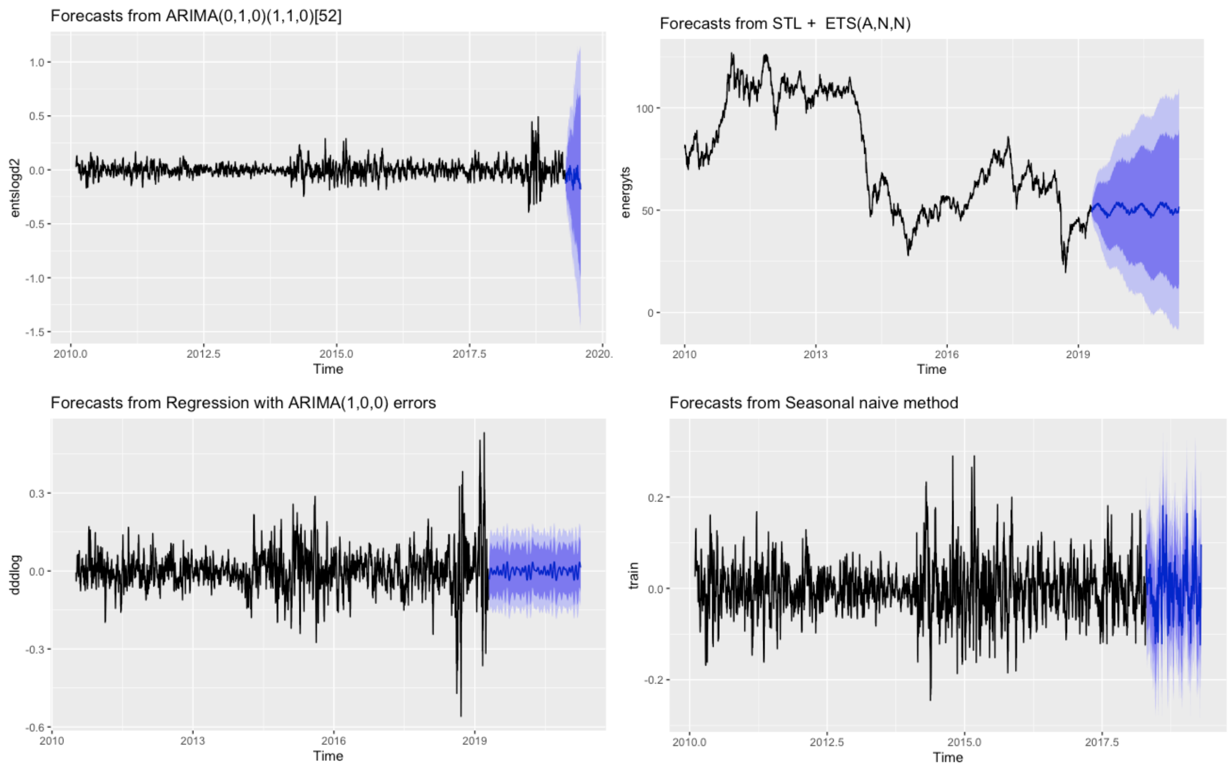
Applying different lags in the Box-Ljung test allows to check the empirical hypothesis about the seasonal period. Plugging lag = 1, 6, 30 and 180 for daily data, the maximum p-value derived from lag-6 and lag-30 manual differencing [Picture 2, 3]. This outcome allows us to conclude that both weekly and monthly seasonality could be observed for Brent prices dynamics.

After observing clear seasonal patterns, a set of seasonal-specific models is tested to derive the relevant current data explanation and more or less accurate forecasting.

Concerning seasonal models, the following were compared:

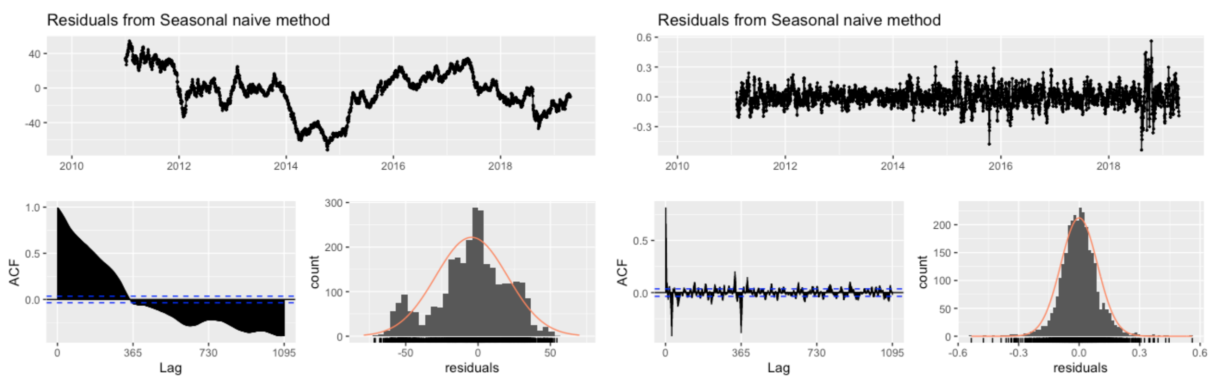
- Seasonal ARIMA with specification (0-1-0)(1-1-0) – applied to the differenced data with AR in seasonal part; the model is chosen after comparison with the same class model of another specification by RMSE and AIC [Picture 4];
- Complex STL + ETS with non-stationary data;

- Fourier series with K=15 (chosen by AIC comparison) – which is in fact ARIMA (1-0-0);
- Seasonal Naive Forecast applying training set method.



Comparing the set of models, simple **Seasonal Naïve Forecast** appears to be the most relevant (as it was proposed in the beginning of the course, often the easiest model is the most accurate one ☺). Even taking into account the overlapping seasonality derived at the first stage (both weekly and monthly), complex approach with Fourier series failed to forecast the date with the accuracy as Seasonal Naive did.

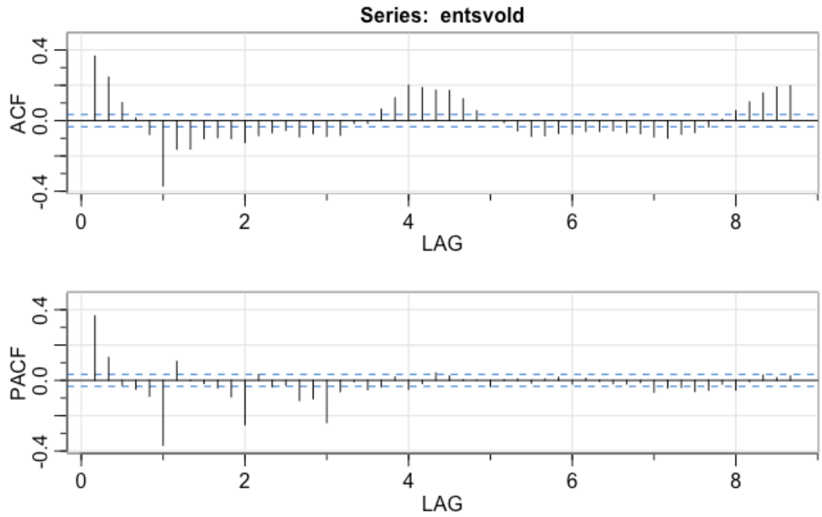
Taking a closer look for the best fitted model, let us check the accuracy by residuals test.



The residuals fail the white noise tests, but their autocorrelations are tiny, even though they are significant. This is because of the length of the series, but the effect of the remaining correlations on the forecasts supposed to be negligible [Rob J Hyndman, 'Forecasting: Principles & Practice'].

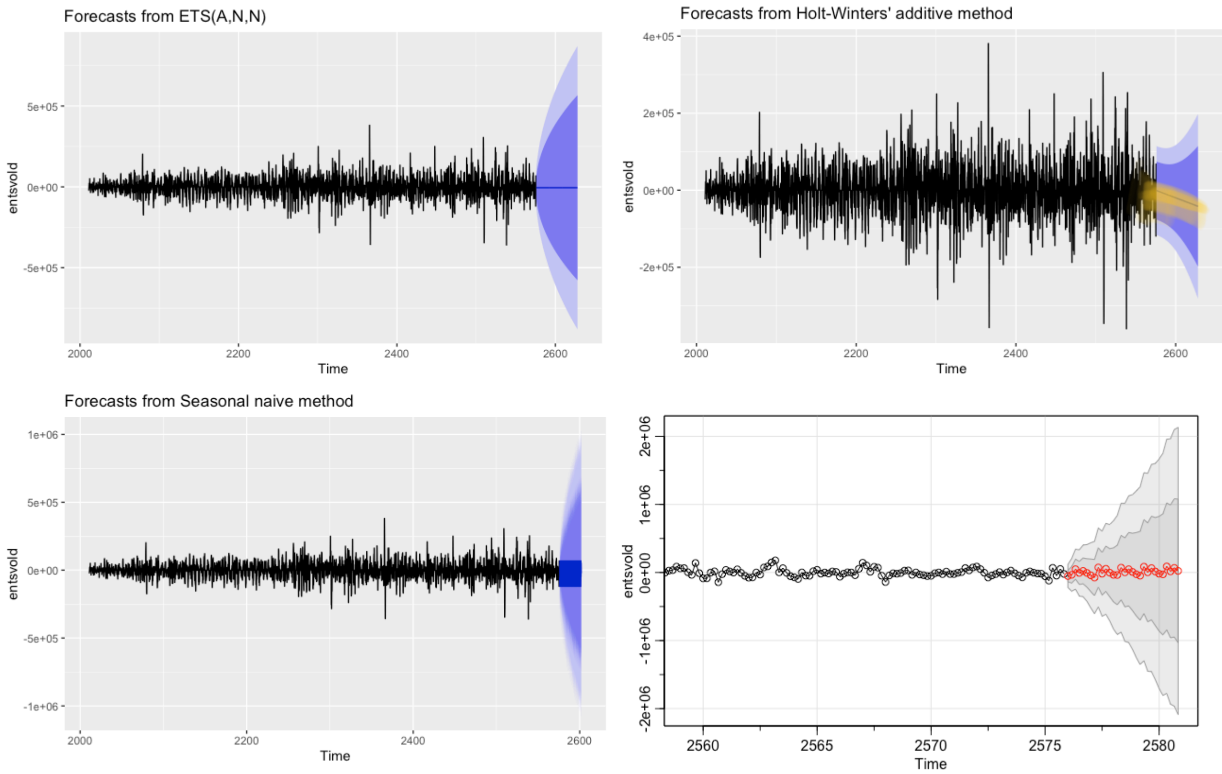
Volume fluctuations modeling

Modeling trading volume with observed lower frequency allows us to employ a class of models, which could not handle with the data of higher frequency level as it was on the first stage. Recall that from a glimpse observation of initial data table, volume frequency is set as 6. A neat weekly pattern could also be observed from the correlograms:



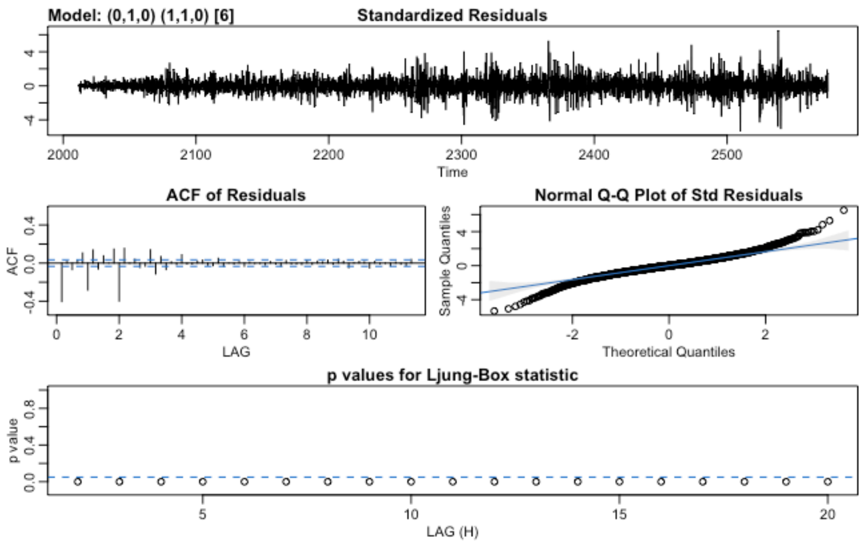
As on the previous stage, non-seasonal models are not discussed, as they fail to capture seasonal fluctuations. In special case of low-frequency data the following models are explored:

- ETS model with auto-matched parameters;
- Holt-Winters model with additive seasonal parametr;
- Seasonal Naïve model – recalling it's latter success;
- Seasonal ARIMA model in (0-1-0), (1-1-0) configuration.



In contrast with the high frequency price prediction, Seasonal ARIMA model overperformed Seasonal Naïve, producing a smoother data representation in the forecast period, while ETS and HW models both failed to capture seasonality.

Examining closer model's accuracy, we again face with the failed white noise test for the residuals, explaining by the same reason of long data frame. At the same time, qqplot shapes rather accurate predictions with minimal number of outliers.



Conclusion

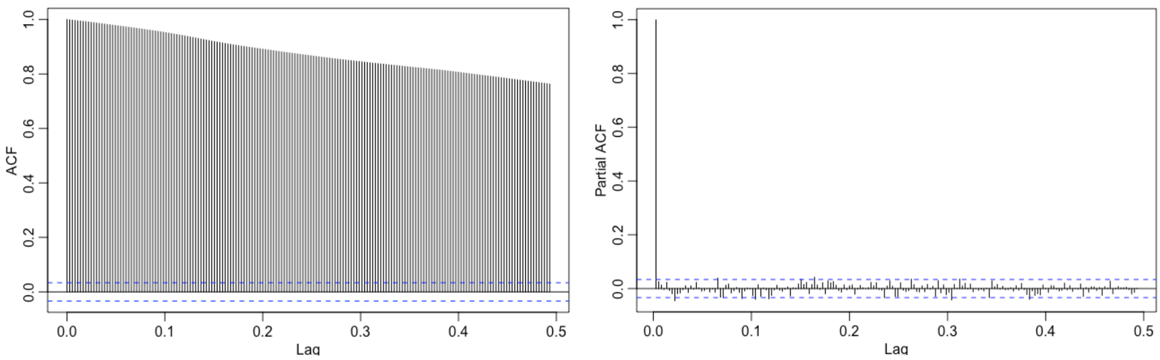
Obviously, for accurate assessing and forecasting of seasonal data one needs to employ a complex approach. It would be relevant for further research to investigate the macro-factors, affecting the fluctuations of energy curves, as well as account for irrational anomalies, empirically observed on the market.

Appendix

Table 1. Initial Dataset: observing weekly pattern

<DATE>	<OPEN>	<HIGH>	<LOW>	<CLOSE>	<VOL>
20100104	78.4100000	80.3400000	78.3400000	80.2400000	47103
20100105	80.2200000	80.8400000	79.7500000	80.7100000	44830
20100106	80.7500000	82.2000000	79.7800000	81.8200000	49453
20100107	81.7800000	82.0600000	81.0500000	81.4700000	36568
20100108	81.4600000	82.0400000	80.5900000	81.4600000	42240
20100109	81.5400000	81.5500000	81.4300000	81.4300000	**77**
20100111	82.1700000	82.9500000	81.0900000	81.2900000	23094
20100112	81.0600000	81.4000000	79.2800000	79.6300000	35263
20100113	79.6000000	79.6900000	77.5700000	78.8000000	54102
20100114	78.8000000	79.4500000	77.9400000	78.4200000	58472
20100115	78.3100000	78.4300000	76.8100000	77.1500000	50603
20100116	77.1000000	77.2000000	77.1000000	77.2000000	**177**
20100118	76.9000000	77.5900000	76.3900000	77.0800000	33423
20100119	77.1200000	77.8500000	75.3800000	77.7200000	85722
20100120	77.7500000	77.7600000	75.7500000	75.8600000	53632
20100121	75.8800000	76.7500000	74.2300000	74.6200000	50216
20100122	74.6200000	75.1500000	72.3900000	72.4200000	55550
20100123	72.3900000	72.5000000	72.3300000	72.3800000	**102**

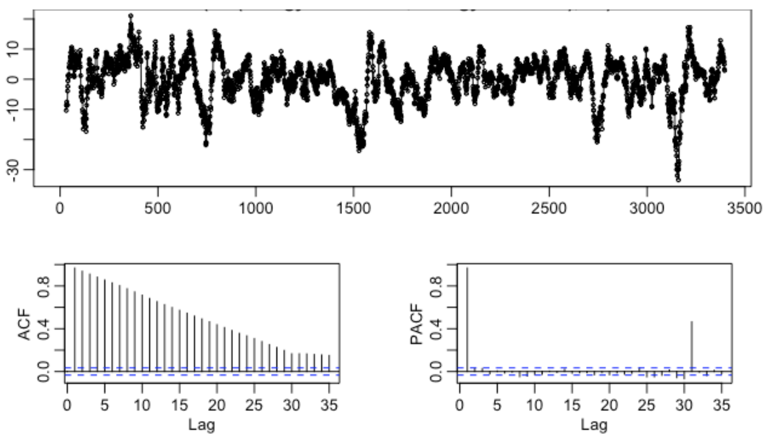
Picture 1. ACF and PACF: seasonal AR for daily prices



Picture 2. Box-Ljung test for seasonality patterns: max p-value for lag 6 and lag 30

```
        Box-Ljung test

data:  energydif
X-squared = 8.5035, df = 6, p-value = 0.2035

> Box.test(energydif, lag = 1, type = "Ljung")

        Box-Ljung test

data:  energydif
X-squared = 3.1184, df = 1, p-value = 0.07741

> Box.test(energydif, lag = 6, type = "Ljung")

        Box-Ljung test

data:  energydif
X-squared = 8.5035, df = 6, p-value = 0.2035

> Box.test(energydif, lag = 30, type = "Ljung")

        Box-Ljung test

data:  energydif
X-squared = 37.947, df = 30, p-value = 0.1511

> Box.test(energydif, lag = 180, type = "Ljung")

        Box-Ljung test

data:  energydif
X-squared = 230.62, df = 180, p-value = 0.006446
```

Picture 3. Lag 30 is clearly observed through the PACF analysis



Picture 4. sARIMA models comparison: choosing the minimal RMSE and AIC

```
> summary(marimass)

Call:
arima(x = entslogd2, order = c(0, 1, 0), seasonal = list(order = c(0, 1, 0),
    period = 52))


sigma^2 estimated as 0.003356:  log likelihood = 4727.56,  aic = -9453.12

Training set error measures:
                      ME       RMSE        MAE      MPE    MAPE     MASE        ACF1
Training set 2.99709e-06 0.05747239 0.04001413 142.5614 910.495 1.445612 -0.05455507
> summary(marimas)

Call:
arima(x = entslogd2, order = c(0, 1, 0), seasonal = list(order = c(1, 1, 0),
    period = 52))

Coefficients:
         sar1
      -0.4872
s.e.   0.0152

sigma^2 estimated as 0.00256:  log likelihood = 5168.33,  aic = -10332.66

Training set error measures:
                        ME       RMSE        MAE      MPE     MAPE    MASE       ACF1
Training set -5.689318e-06 0.05019377 0.03513995 183.6175 727.4451 1.26952 -0.0629033
>
```