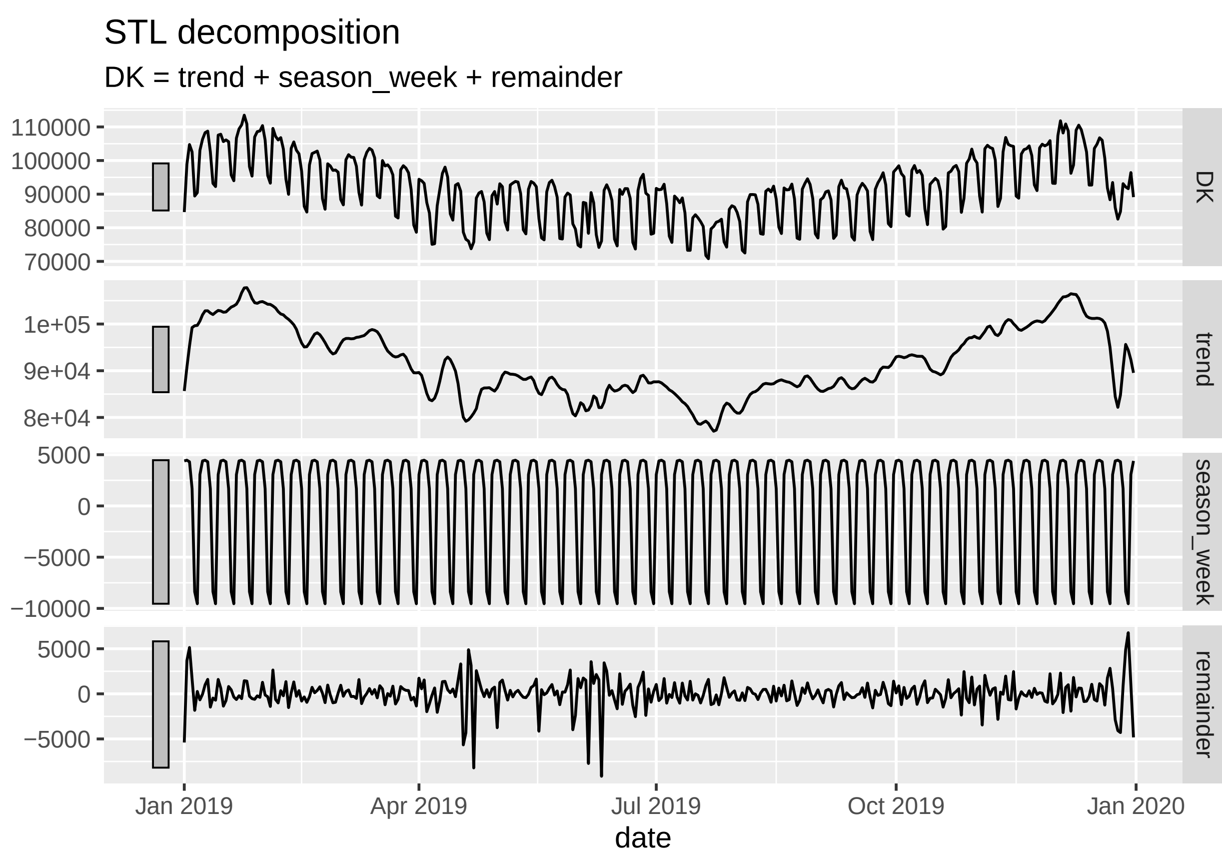
Assignment 5

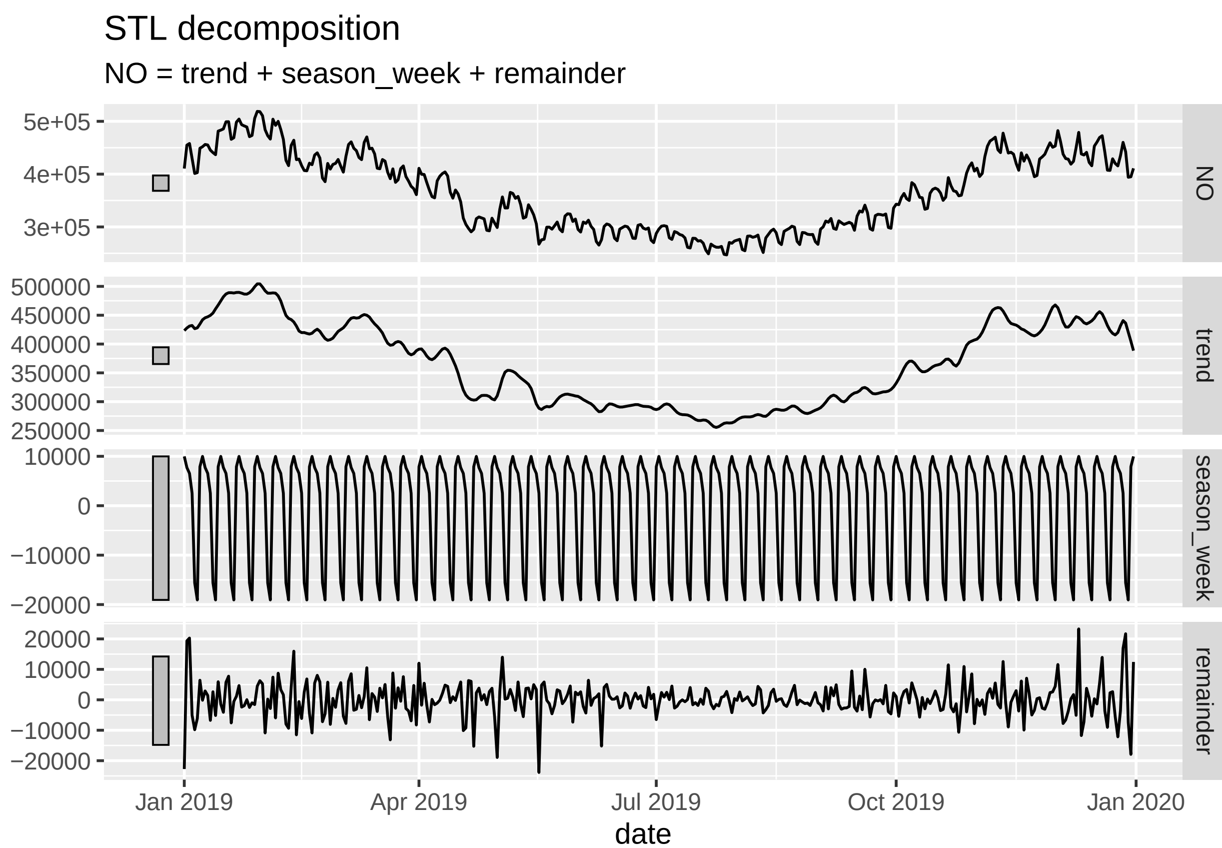
Task 1: Comparison of consumption data in Norway and Denmark

We start our comparison of norwegian and danish power consumption by looking at the STL decompositon plots.

```
## Seasonal decompositon of Danish electricity consumption

cons_comp_dk = cons_ts %>% model(
  STL(DK ~ trend(window=7) + season(window="periodic"))
) %>% components

cons_comp_no = cons_ts %>% model(
  STL(NO ~ trend(window=7) + season(window="periodic"))
) %>% components


cons_comp_dk %>%  autoplot()
```



```
cons_comp_no %>%  autoplot()
```

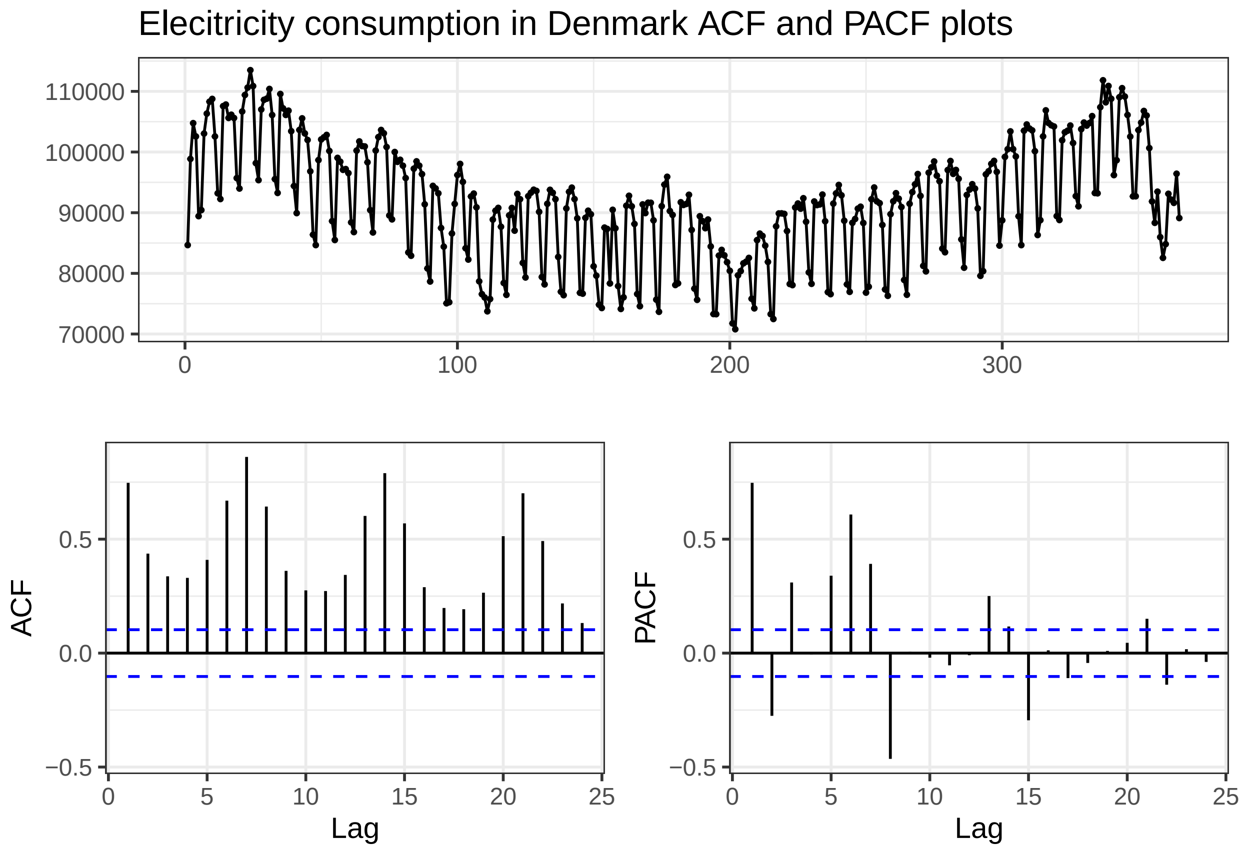


From the components plot we can see the same general trend of winter increase, and the corresponding decrease in power consumption in the summer months. This difference in summer/winter months however, appears to be larger in Norway than Denmark, with higher consumption in the winter relative to Denmark. Another difference is the magnitude of the weekly-sesaonal component: In Denmark the weekend effect of reduced power consumption is more noticable.

Forecasting using a seasonal ARIMA model

A clear assumption in an ARIMA forecasting model is the that the data is stationary in terms of its variance and mean. We plot the time series containing electricity consumption data, as well its autocorrelation and partial autocorrelation plots. We see clear signs of non-stationarity and perform unit root tests confirming the need for differencing. As there appears to be a strong seasonal autocorrelation, we will first conduct a seasonal differencing, and see if this solves our non-stationarity issue. If not, further differencing will be needed. Some information contained in the data is lost by performing a differencing, but we conform the the assumption of stationarity of the data.

```
forecast::ggtsdisplay(cons$DK, plot_type='partial',
                      lag.max = 24,
                      theme = theme_bw(),
                      main = "Elecitricity consumption in Denmark ACF and PACF plots ")
```




```r
unitroot_kpss(cons$DK)
```

```
##   kpss_stat kpss_pvalue
##    1.219224    0.010000
```

```r
adf.test(cons$DK)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  cons$DK
## Dickey-Fuller = -2.17, Lag order = 7, p-value = 0.5052
## alternative hypothesis: stationary
```

We perform a first order differencing and perform the same stationarity analysis.

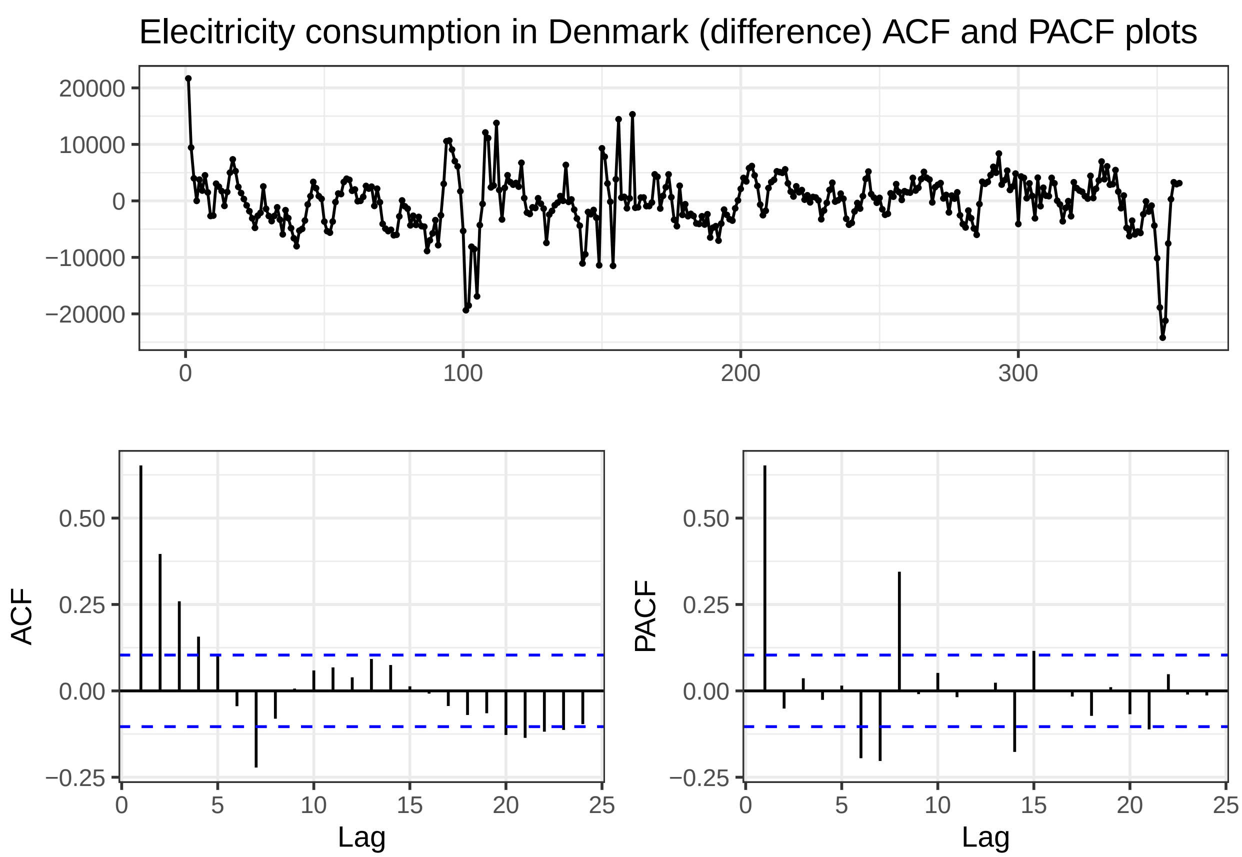
```r
# Perform differencing

cons_diff_dk <- cons %>% mutate(DK = difference(DK,7)) %>% dplyr::filter(!is.na(DK)) #Take first order

unitroot_kpss(cons_diff_dk$DK)
```

```
##   kpss_stat kpss_pvalue
##   0.1060349   0.1000000
```

```r
adf.test(cons_diff_dk$DK) #Stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  cons_diff_dk$DK
## Dickey-Fuller = -5.5696, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```r
# New ACF and PACF plots

forecast::ggtsdisplay(cons_diff_dk$DK, plot_type='partial',
                      lag.max = 24,
                      theme = theme_bw(),
                      main = "Elecitricity consumption in Denmark (difference) ACF and PACF plots ")
```



We note that there are significant autocorrelations at the weekly lag (i.e 7, 14). Luckily, the fable package correctly identified the seasonality as weekly, e.g an ARIMA PDQ pdq[7], regardless of the specific terms.

We will now perform two forecasts, a manually specified ARIMA model and an automatically determined ARIMA model made by the fable ARIMA() function. In the PACF plot we can see a clear autocorrelation in seasonal lag terms, in a decreasing fashion. This calls for an MA(1) term to applied to the seasonal component of the ARIMA model. There is a significant but decreasing correlation at lag 1 as shown in the PACF plot, and an AR(1) term applied to the non-seasonal component is appropiate. As the lags are decreasing after 1, and MA(1) term might also be necessary.

A such we might reason that our model might look something like: ARIMA pdq(1,0,1) pdq(1,1,1)

We will compare this model to an AIC_c optimized model made by the fable package.

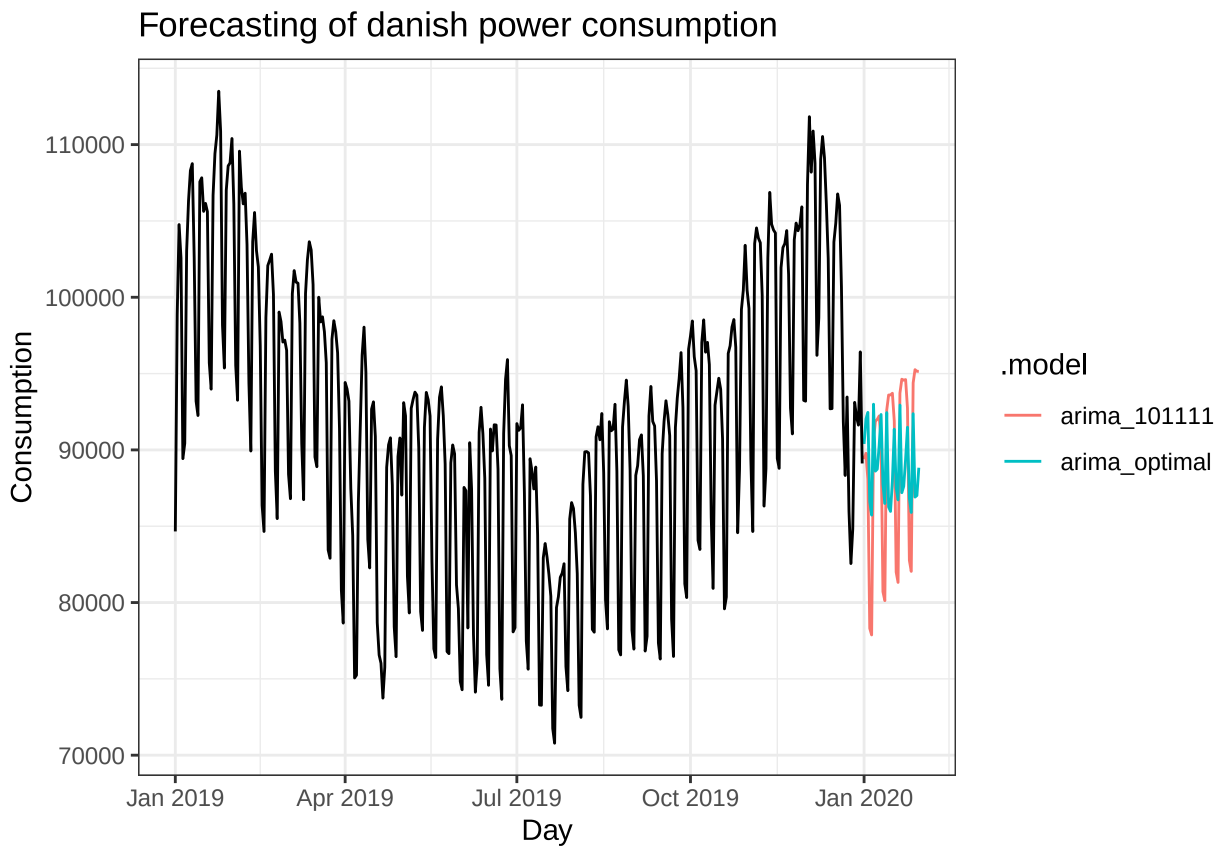
```
fit_arima_optimal_cons_dk <-
  cons %>%
  as_tsibble(index = date) %>%
  model(arima_optimal =  ARIMA(DK, stepwise = FALSE, approximation = FALSE))


fit_arima_manual_cons_dk <- cons %>% as_tsibble(index = date) %>%
  model(arima_101111       = ARIMA(DK ~ 0 + pdq(1,0,1) + PDQ(1,1,1)))

fit_cons_dk  <-  fit_arima_manual_cons_dk  %>% bind_cols(fit_arima_optimal_cons_dk)
```

In the plot below we have used our manually selected model to forecast danish power consumption for a time horizon of 30 days.

```
#Forecast for h = 30
fc_cons_dk <-  fit_cons_dk %>% forecast(h = 30)

# Plot forecast
fc_cons_dk  %>% ggplot() +
  geom_line(aes(x = date, y = .mean, color = .model)) +
  geom_line(aes(x  = date, y = DK), data = cons) +
  theme_bw()   +
   labs(title = "Forecasting of danish power consumption",
         y = "Consumption",
         x = "Day")
```

Table 1: Training data performance metrics: Danish power consumption

Model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	AC
arima_101111	Training	-48.38959	2860.688	1853.999	-0.1144917	2.052392	0.5298189	0.5781630	-0.03580
arima_optimal	Training	-50.35888	3070.275	2111.181	-0.1114627	2.331917	0.6033140	0.6205218	-0.02976



As we can see our manually selected model slightly outperforms the optimally selected model based on AIC. The Ljung Box test reveals that there is no residual autocorrelation and as such much of the error term is explained in our model.

```
# Forecast evaluation

fit_cons_dk %>%  accuracy()  %>%
  rename("Model" = .model) %>%
  kbl(caption = "Training data performance metrics: Danish power consumption") %>%
  kable_classic(full_width = F, html_font = "Times new roman")
```

```
# Autocorrelation tests


feasts::ljung_box(
  (fit_cons_dk  %>%
     augment()  %>% dplyr::filter(.model == "arima_101111"))$.innov) # Passing
```

```
##   lb_stat lb_pvalue
```

```
## 0.4718201 0.4921510
```

Perhaps the most notable weakness of such a model is its weekly sesaonality. It would be interesting to tweak the model with longer seasonality, as a week may be too small a period to capture the winter/summer differences in consumption.

Another weakness is that some information useful for forecasting may be found in other variables, and as a result a multivariate model should also be tested.

Task 2:

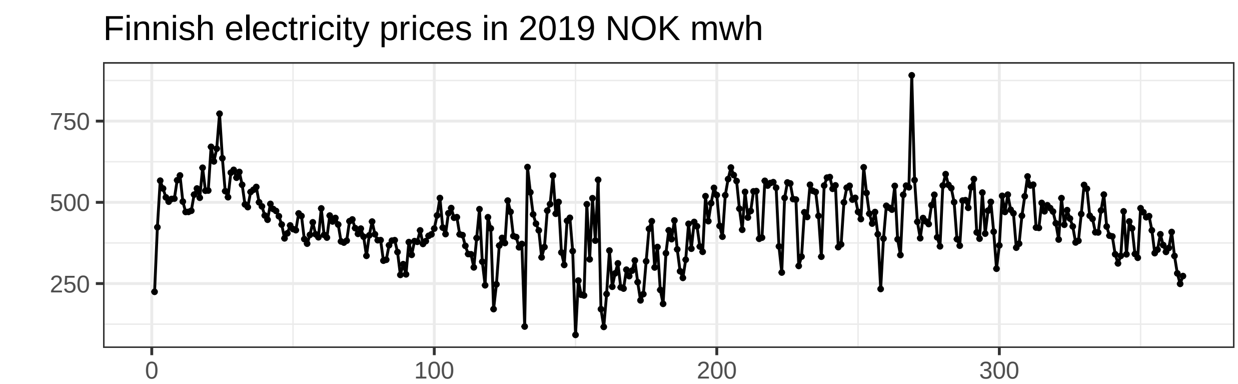
We retrieve finnish electricity price data from Nordpool and clean them.

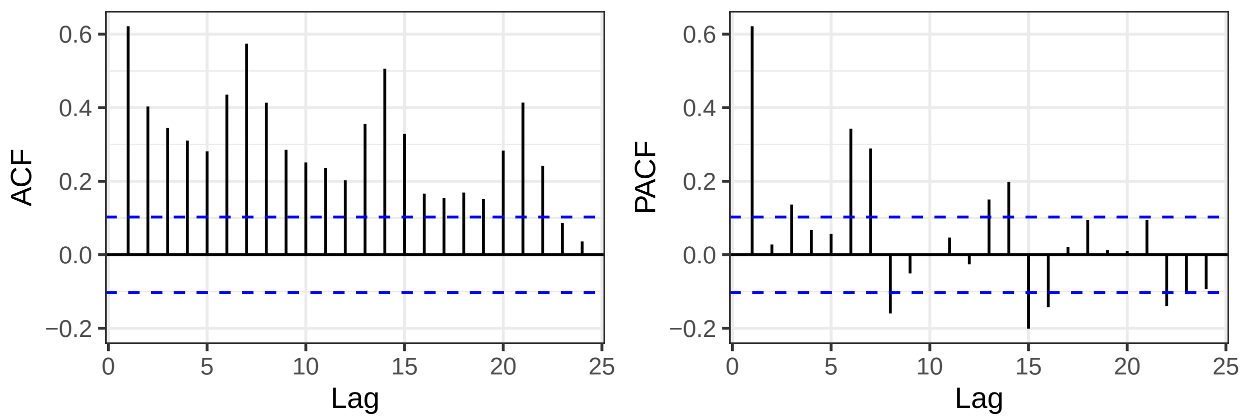
```
elspot_data <- read_csv("elspot-prices_2019_daily_nok.csv") # Load data
```

```
## Parsed with column specification:
## cols(
##   .default = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
colnames(elspot_data) <- elspot_data[2,]
elspot_data <- elspot_data[3:nrow(elspot_data),]
colnames(elspot_data)[1] <- "date"

elspot_data  %<>%
    mutate(date = lubridate::dmy(date),
           FI = as.numeric(gsub(",", ".", FI)))  %>%  #Substitute commas
    select(date, FI)  %>%
    as_tsibble(index = date)

#Plot of finnish electricity prices and ACF/PACF plots
forecast::ggtsdisplay(elspot_data$FI, plot_type='partial',
                      lag.max = 24,
                      theme = theme_bw(),
                      main = "Finnish electricity prices in 2019 NOK mwh")
```





We find the conditional variance of the series by calculating the residuals after an ARIMA model (in our case an ARIMA(1,1) model). By subtracting the original values with with the fitted values of this ARIMA model, we end up with the "return residuals".

```
fi_garch <- garchFit(~arma(1,1) + garch(1,1), data = elspot_data$FI, trace = F)
summary(fi_garch)
```

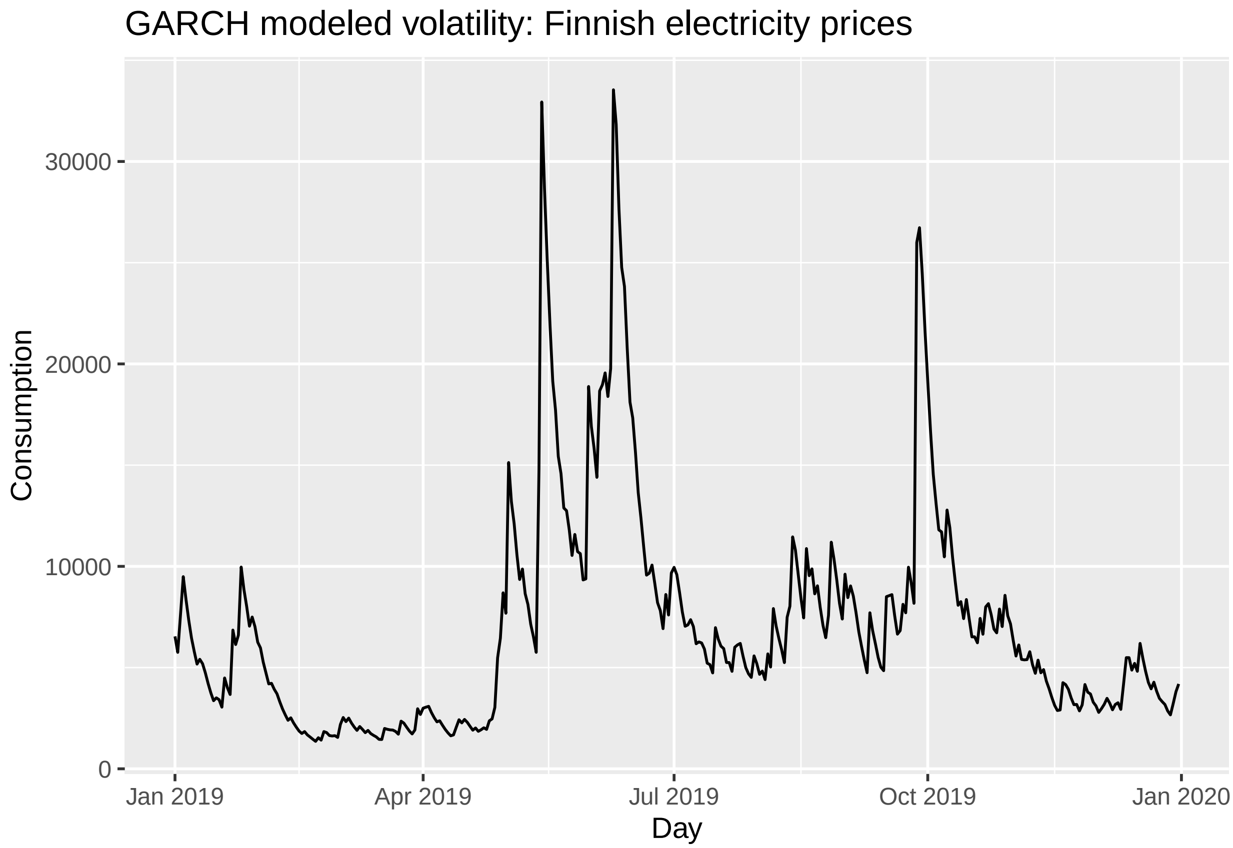
```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(1, 1) + garch(1, 1), data = elspot_data$FI,
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(1, 1) + garch(1, 1)
## <environment: 0x000000002cd34488>
##  [data = elspot_data$FI]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##         mu        ar1        ma1      omega     alpha1      beta1
## 123.56463    0.71407   -0.03790   81.57607    0.13143    0.86680
```

```
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     123.56463    33.04677    3.739 0.000185 ***
## ar1      0.71407     0.07565    9.439  < 2e-16 ***
## ma1     -0.03790     0.12114   -0.313 0.754387
## omega   81.57607    63.61225    1.282 0.199704
## alpha1   0.13143     0.03585    3.666 0.000246 ***
## beta1    0.86680     0.03105   27.918  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  -2088.787    normalized:  -5.722703
##
## Description:
##  Tue Apr 13 09:59:17 2021 by user: sondr
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  25.00763  3.712463e-06
##  Shapiro-Wilk Test  R    W      0.9874976 0.00312319
##  Ljung-Box Test     R    Q(10)  97.34201  2.220446e-16
##  Ljung-Box Test     R    Q(15)  189.0253  0
##  Ljung-Box Test     R    Q(20)  213.1669  0
##  Ljung-Box Test     R^2  Q(10)  11.40592  0.3267794
##  Ljung-Box Test     R^2  Q(15)  18.17036  0.2538154
##  Ljung-Box Test     R^2  Q(20)  23.56921  0.2617165
##  LM Arch Test       R    TR^2   14.24749  0.285184
##
## Information Criterion Statistics:
##      AIC      BIC      SIC     HQIC
## 11.47828 11.54239 11.47775 11.50376
```

We note the signifiance of the alpha and gamma parameters, and our time series variance appears to be conditional on previous variance.

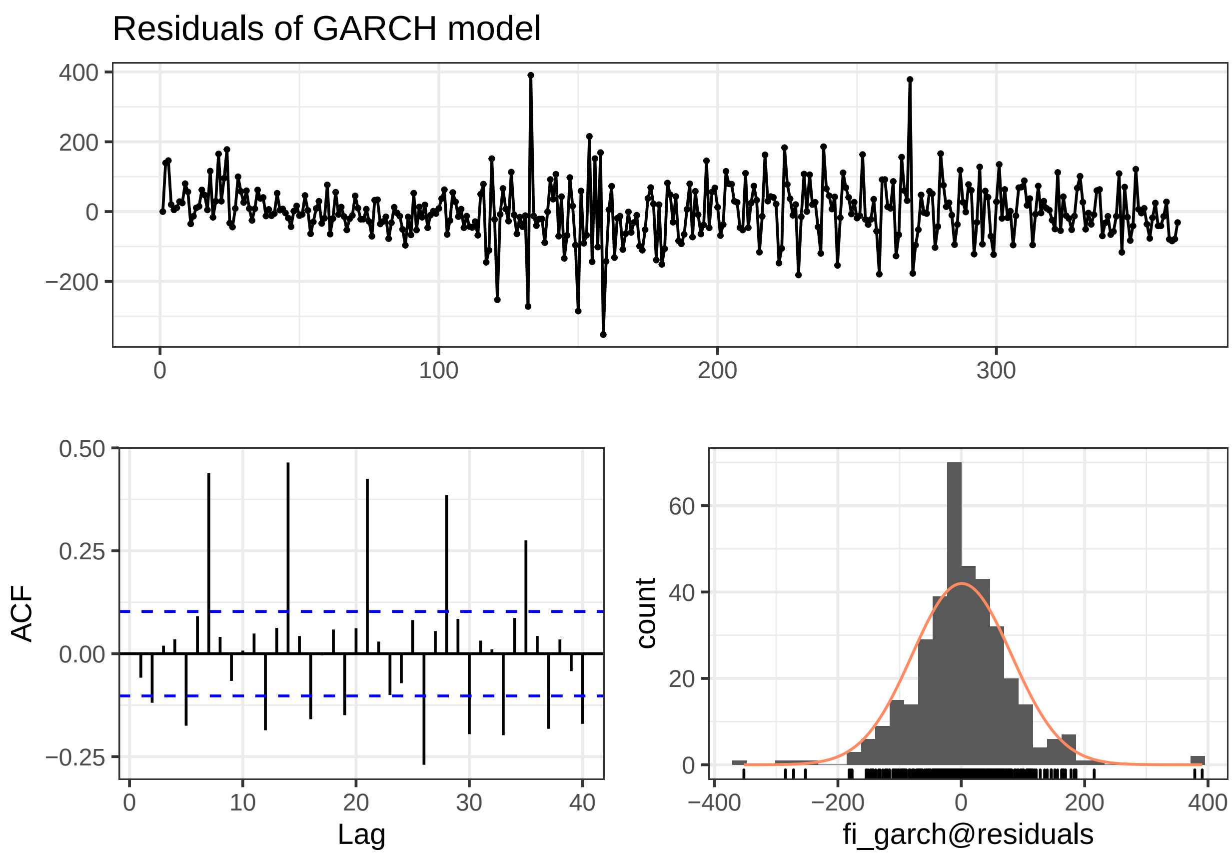
We model the conditonal volatility of the time series, and note that there are several significant spikes in volatility of electricity prices. For instance around the months May, June, and October.

```
elspot_data  %<>% mutate(volatility = fi_garch@h.t)
elspot_data  %>%
    ggplot() +
    geom_line(aes(y = volatility, x = date))  +
    labs(title = "GARCH modeled volatility: Finnish electricity prices",
         y = "Consumption",
         x = "Day") +
     scale_colour_manual(values = c("black", "orange"))
```



We note that the residuals show autocorrelation but not enough to reject the ljung box null hypothesis of white noise residuals. Furthermore the residuals appears to be normally distributed.

```
forecast::ggtsdisplay(fi_garch@residuals,
                      plot.type="histogram",
                      lag.max = 40,
                      theme = theme_bw(),
                      main = "Residuals of GARCH model")
```

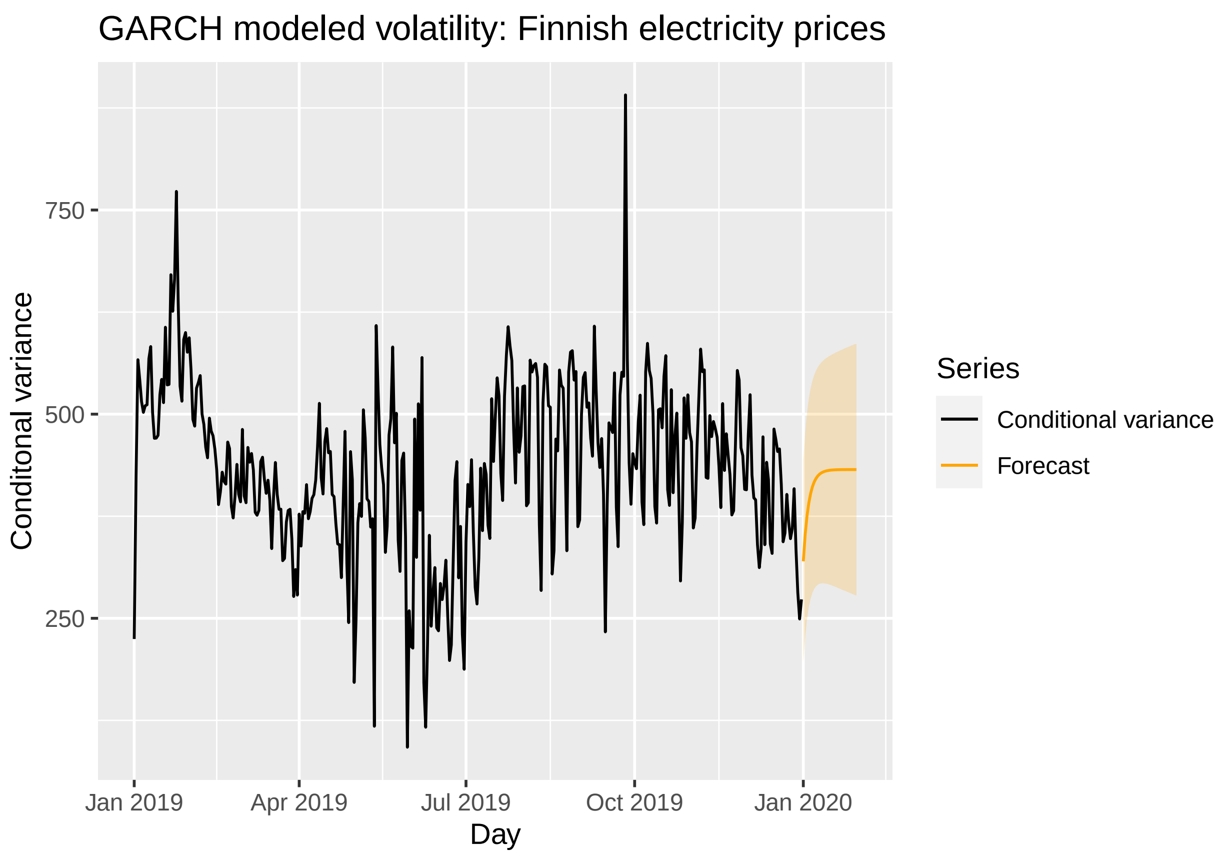



```
ljung_box(fi_garch@residuals)
```

```
##   lb_stat lb_pvalue
## 1.2493384 0.2636789
```

Using our GARCH model we forecast 30 days ahead with prediction intervals.

```
fc_garch = predict(fi_garch, n.ahead = 30)
fc_garch["date"] = seq(as.Date("2020-01-01"), as.Date("2020-01-01")+29, by = "days")

elspot_data  %>%
  ggplot() +
  geom_line(aes(x = date, y = FI, color = "Conditional variance")) +
  geom_line(aes(x = date, y = meanForecast, color = "Forecast"), data =  fc_garch) +
  geom_ribbon(aes(x=date, ymin=(meanForecast-2*standardDeviation), ymax=(meanForecast+2*standardDeviati
  scale_colour_manual(values = c("black", "orange")) +
  labs(title = "GARCH modeled volatility: Finnish electricity prices",
         y = "Conditonal variance",
         x = "Day",
         colour = "Series")
```



An ARCH model assumes symmetric shocks on volatility, and this may not be the case for the electricity market. Furthermore volatility, by its nature has sudden increases and decreases and this makes forecasting difficult. This may lead the forecast to "underfit" the actual future observations.

The advantage of such a model, is that given no residuals error (such as a non-normal distribution) we can trust the prediction intervals to at least give us a pinpoint of future volatility.

Task 3

The inclusion of additional independent variables in a forecasting model may be helpful if they provide additional causal information to the model. The amount of rain has a significant and causal relation to power prices in Norway, as the availability of rain increases the supply of hydroelectric power. As such, it is plausible that such a multivariate model of rain amount and previous power prices may yield a better result than a univariate model based on power prices. Causality ensures future stability, as previous correlations may disappear going forward in time. As causality may be very hard to determine, including a variable may contribute more noise than information for a forecasting model.

The future is uncertain, and adding independent variables may only be prudent if it contributes significant information.