Forecasting Electricity Consumption in Sweden

Caglar Altunel

6/14/2021

Forecasting Electricity Consumption in Sweden. A comparison of univariate and multivariate models

Introduction

Forecasting electricity consumption has been a hot topic in literature as not only it provides important information for various needs of businesses, such as revenue forecasting and profit optimization, but also due to the debatable nature of its causality, for instance, whether it can be explained by economic growth, or the direction of the causality is other way around. Literature is very rich when it comes to the direction of causality. Behera, J. (2015) gathered these causality types in a very concise way in his article. According to the article (Behera, 2015), there are mainly four types of causalities that are used in literature to explain the relationship between energy consumption and economic growth. These are conservation hypothesis, growth hypothesis, feedback hypothesis and neutrality hypothesis. The conservation hypothesis refers to a unidirectional causality that holds economic growth impacts energy consumption, while the growth hypothesis explains this unidirectional causality in an opposite way, meaning that energy consumption impacts economic growth. The feedback hypothesis suggests a bidirectional relationship between energy consumption and economic growth, as the neutrality hypothesis refers to no relationship between those.

To give examples from the literature, Cowan et al. (2014) investigated the causality between electricity consumption and economic growth and found evidence on the feedback hypothesis for Russia and the conservation hypothesis for South Africa, while they found no relationship between electricity consumption and economic growth for Brazil, India and China. A similar study was conducted by Sarwar et al. (2017) and they concluded that there is a bidirectional relationship between electricity consumption and GDP. In attempts to explain the causal relationship between those, it can be observed that Granger causality is widely used in the literature. According to Seth (2007), Granger causality is a statistical concept based on prediction and used to determine whether one time series can be used to forecast another time series or not. To give an example for this, Magazzino (2014) tried to find an answer to a similar question and found evidence on a bidirectional Granger causality flow between electricity demand and real GDP per capita. Tursoy & Resatoglu (2016) also used a revised form of Granger causality and concluded that there is a bidirectional causality between electricity consumption and GDP. In addition to this, Karanfil & Li (2015) also found evidence on the feedback hypothesis, in other words, bidirectional relationship, between electricity consumption and GDP in their study relying on panel data that consist information about 160 countries between the years 1980 - 2010. Bélaïd & Abderrahmani (2013) also found out the same bidirectional relationship between those for Algeria. There are also studies that conclude a unidirectional relationship from energy consumption to economic growth in the literature. For instance, Cheng et al. (2013) based their study on an econometric analysis of the annual growth data for China’s GDP and electricity generation and found evidence on the growth hypothesis between those. Also, Mohanty & Chaturvedi (2015) concluded that electricity consumption fuels economic growth both in short run and long run. That being said, studies conducted by Chen et al. (2007) and Ghosh (2002) concluded that there is a unidirectional relationship from economic growth to electricity consumption, meaning that economic growth is a significant indicator of electricity consumption.

Another striking observation rather than strong focus on the direction of causality is that the vast majority of studies are conducted by relying on the datasets that cover information regarding developing countries. However, within the last decades, it can be observed that economic growth continues in developed countries, while governments try to reduce energy consumption in order to guarantee a more sustainable future. According to the article published in The Guardian (Harvey, 2015), thanks to the efforts taken by the EU countries and businesses to reduce energy consumption and increase efficiency, energy consumption fell by more than 9% from 2006 to 2013 within member states. EU also has a clear goal of reducing greenhouse gas emissions by at least 55% by 2030, compared to 1990 (ec.europa.eu, 2020), along with 32.5% improvement in the energy efficiency (ec.europa.eu, 2014). In addition to these, EU countries have a target to make all new buildings to be nearly-zero energy buildings by the end of 2020 according to its Energy performance of buildings directive (ec.europa.eu, 2019) and has a committed itself to the goal of reducing the energy consumption by 20% compared to baseline (2007) projections.

Unlike the abovementioned studies, this project will not focus on the direction of causality between electricity consumption and economic growth. Rather, it will assume an existence of conservation hypothesis between electricity consumption and the economic growth, and test whether it remains or not by making use of the relevant data gathered from an EU country, namely Sweden, relying on the idea that successful efforts taken by the member states to reduce energy consumption would potentially disconnect the hypothetical relationship between economic growth and energy consumption. To that end, I will first build a univariate ARIMA model to forecast electricity consumption in Sweden. Then an alternative model where economic growth (change in GDP) is included as an exogenous variable will be built. Initially, the entire dataset will be used for modelling purposes due to scarcity of observations, but if forecast figures calculated by each model will differ significantly, then I will consider train-test split and run the procedure from scratch to find out the best fit. If adding GDP as an exogenous variable will not make a significant contribution on the forecast figures of the univariate ARIMA model, it will be possible to argue that economic growth and energy consumption no longer have a significant relationship for Sweden as a consequence of the efforts taken to reduce energy consumption.

Methodology (ARIMA)

Autoregression refers to such models that the dependent variable is regressed on its own lagged values, while integration represents differencing of raw values of the dataset and moving average implies the level of dependency between observations and residuals (Hayes, 2021). Combining these, Autoregressive Integrated Moving Average (ARIMA) is a statistical model that is built to forecast future values of a dependent variable by regressing it on its lagged values and lagged error terms. A typical ARIMA model is simply denoted as ARIMA(p,d,q)(P,D,Q) where p is the lag order, d stands for the number of times that raw values are differenced and q is the order of moving average, while the upper case ones are the same for the seasonal component. An exogenous variable can be implemented to the model. If the dependent variable is differenced, then the exogenous variable should be differenced as the same way it is done for the dependent variable. As I briefly discussed under Introduction section, I will build both with and without exogenous variable to see whether the exogenous one will make a contribution to forecast accuracy or not.

Data Gathering & Cleaning Process

In this project, I will use two kinds of dataset, one is monthly electricity consumption data of Sweden whereas the other one is monthly GDP data. Both data covers the period January 2013 to March 2021. Electricity consumption data was downloaded from Nordpool's historical market data page (Main source: https://www.nordpoolgroup.com/historical-market-data/ (https://www.nordpoolgroup.com/historical-market-data/)). It was downloaded as multiple Excel files covering daily electricity consumption (for all Nordpool countries, from 2013 to 2021. Example file name for the year 2021 is, "Consumption per country_2021_Daily". All of these from 2013 to 2021 were downloaded separately.) in MWh terms and before uploading it to R environment, those files merged under one Excel file while other country columns apart from Sweden were deleted. Later, as I am more familiar with Python programming, I uploaded the merged and cleaned daily dataset to Python environment and convert it to monthly data by making use of pandas library. Then I downloaded the transformed dataset as csv file. Meanwhile, monthly GDP indicator data for Sweden was gathered from Statistics Sweden (Main source: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/national-accounts/ovrigt/national-accounts-other/pong/tables-and-graphs/gdp-indicator-monthly-time-series--revisions-at-each-release/ (https://www.scb.se/en/finding-statistics/statistics-by-subject-area/national-accounts/ovrigt/national-accounts-other/pong/tables-and-graphs/gdp-indicator-monthly-time-series--revisions-at-each-release/)). Later, the dataset was cleaned up in such a way that only monthly GDP indicator left as a column and rows covering the same period as electricity consumption data does. Then the dataset were converted into csv and merged with electricity consumption data, and the final version was uploaded to Google Drive to further analyze it in R environment.

In addition to the main dataset that will be used for ARIMA modelling, an additional dataset was created by transforming the existing dataset in such a way that both electricity consumption and monthly GDP indicator was fixed to 100 for January 2013 to visually inspect the trend that each of them posted over entire period being subject to analysis. Later, 12-month moving averages of each were calculated as well in order to enrich the analysis.

Now I will load relevant libraries and the dataset to R environment.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## -- Attaching packages --------------------------------------------------- tidyverse
1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'readr' was built under R version 3.6.3
```

```
## Warning: package 'purrr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

```
## -- Conflicts ---------------------------------------------------- tidyverse_confli
cts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

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

```
library(fpp3)
```

```
## Warning: package 'fpp3' was built under R version 3.6.3
```

```
## -- Attaching packages ------------------------------------------------------ fpp3
0.4.0 --
```

```
## v lubridate   1.7.10     v feasts      0.2.1
## v tsibble     1.0.0      v fable       0.3.0
## v tsibbledata 0.3.0
```

```
## Warning: package 'lubridate' was built under R version 3.6.3
```

```
## Warning: package 'tsibble' was built under R version 3.6.3
```

```
## Warning: package 'tsibbledata' was built under R version 3.6.3
```

```
## Warning: package 'feasts' was built under R version 3.6.3
```

```
## Warning: package 'fabletools' was built under R version 3.6.3
```

```
## Warning: package 'fable' was built under R version 3.6.3
```

```
## -- Conflicts ----------------------------------------------------------- fpp3_conf
licts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::index()     masks zoo::index()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

```
library(fpp2)
```

```
## Warning: package 'fpp2' was built under R version 3.6.3
```

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

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

```
## Loading required package: fma
```

```
## Warning: package 'fma' was built under R version 3.6.3
```

```
## Loading required package: expsmooth
```

```
## Warning: package 'expsmooth' was built under R version 3.6.3
```

```
##
## Attaching package: 'fpp2'
```

```
## The following object is masked from 'package:fpp3':
##
##     insurance
```

```
library(lubridate)
library(dynlm)
```

```
## Warning: package 'dynlm' was built under R version 3.6.3
```

```
library(fpp2)
library(seasonal)
```

```
## Warning: package 'seasonal' was built under R version 3.6.3
```

```
##
## Attaching package: 'seasonal'
```

```
## The following object is masked from 'package:tibble':
##
##     view
```

```
library(curl)
```

```
##
## Attaching package: 'curl'
```

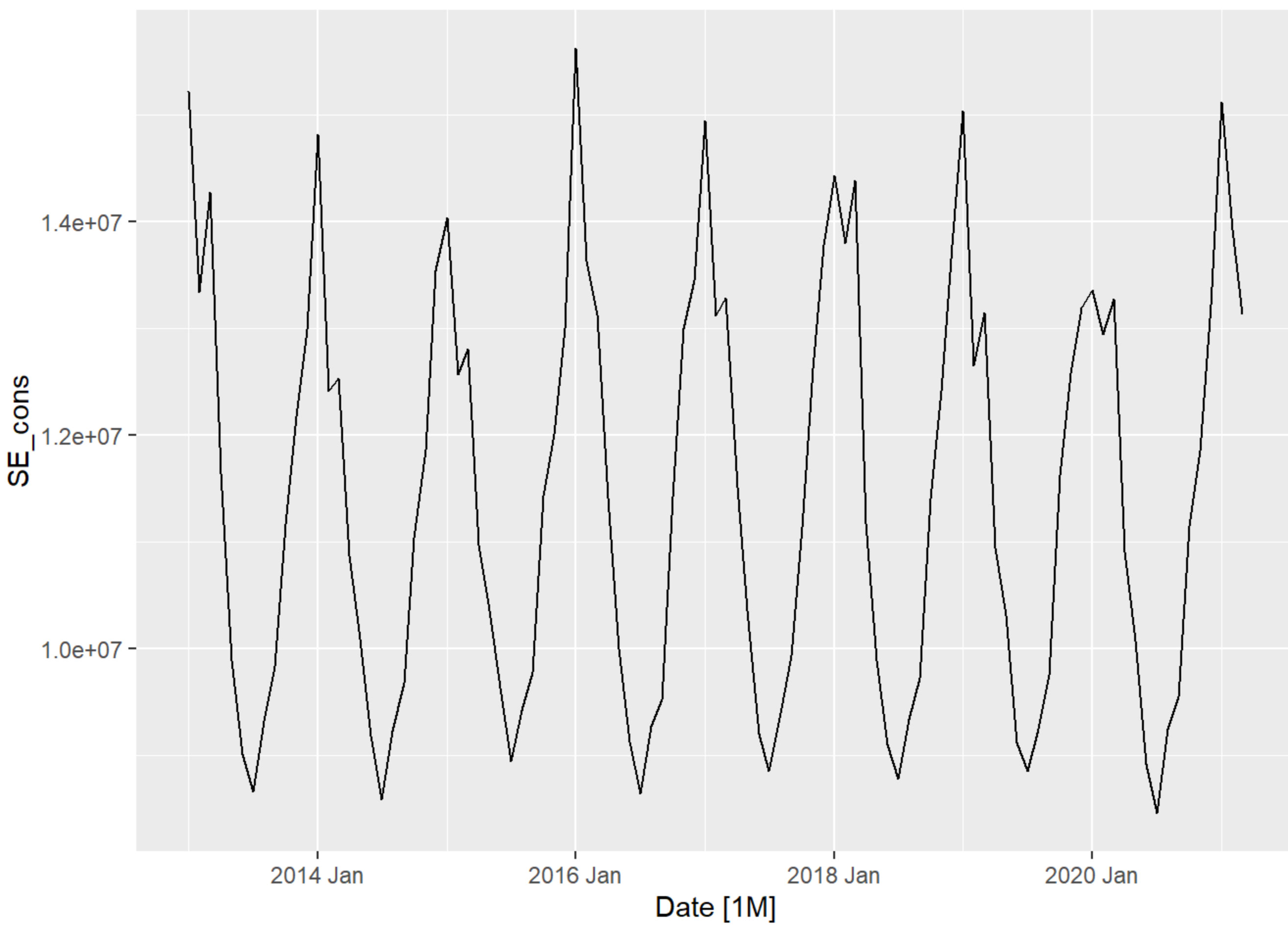
```
## The following object is masked from 'package:readr':
##
##     parse_date
```

```
library(tseries) ## I may not use some of those, but I think it is nice to upload all
potentially relevant ones in one go for convenience.

#Uploading datasets
id1 = "1PE2xN6HnhwElkECIyGB48gyiEuCU9a4E" #Google Drive id of the csv file.
df_fixed = read.csv(sprintf("https://docs.google.com/uc?id=%s&export=download", id1),
fileEncoding = "UTF-8-BOM")

id2 = "1l7WiuxLDHfJcNo7tjPd6NSrxy8qdpz1B" #Google Drive id of the csv file.
df = read.csv(sprintf("https://docs.google.com/uc?id=%s&export=download", id2), fileE
ncoding = "UTF-8-BOM")
```

Descriptive Statistics

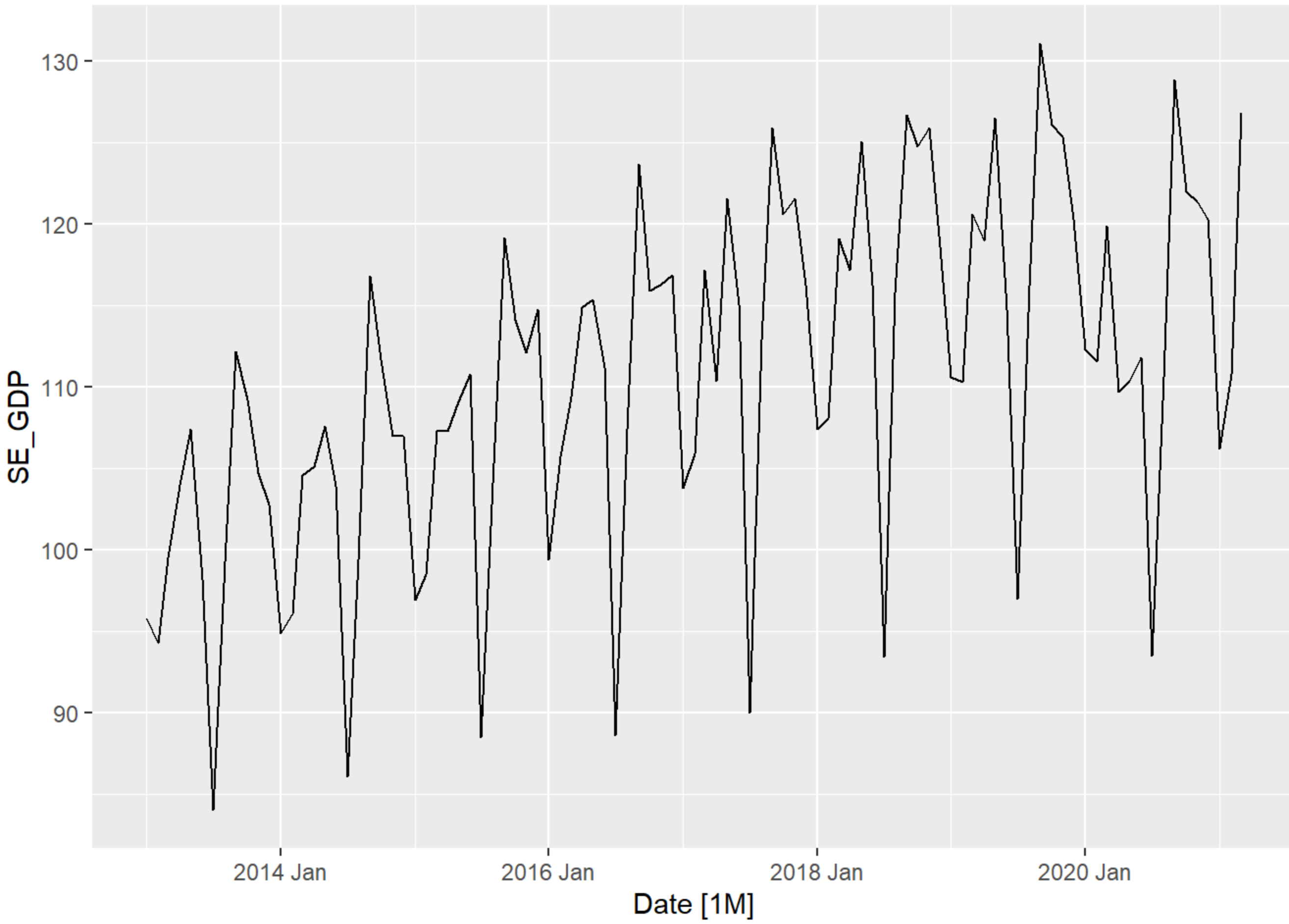
I will start summary statistics with converting datasets into time series and plotting monthly electricity consumption in MWh terms in Sweden.

```
##Converting them into time series
df$Date = as.Date(df$Date, "%m/%d/%Y")
df$Date = yearmonth(df$Date)
df_ts = tsibble(df, index=Date)

df_fixed$Date = as.Date(df_fixed$Date, "%m/%d/%Y")
df_fixed$Date = yearmonth(df_fixed$Date)
df_fixed_ts = tsibble(df_fixed, index=Date)

df_ts %>% autoplot(SE_cons)
```



Very strong seasonality and flat trend can be visually inspected. It also seems that electricity consumption in Sweden was almost not affected from Covid-19 outbreak at all, as no noticeable change in its seasonal trend from March 2020 onwards. As I explained at Introduction, I will eventually build a multivariate forecasting model where GDP indicator will be used as an exogenous variable. Therefore, it makes sense to plot monthly GDP indicator of Sweden for the same period in order to see whether there are similarities with electricity consumption in terms of trend and seasonality.

```
df_ts %>% autoplot(SE_GDP)
```

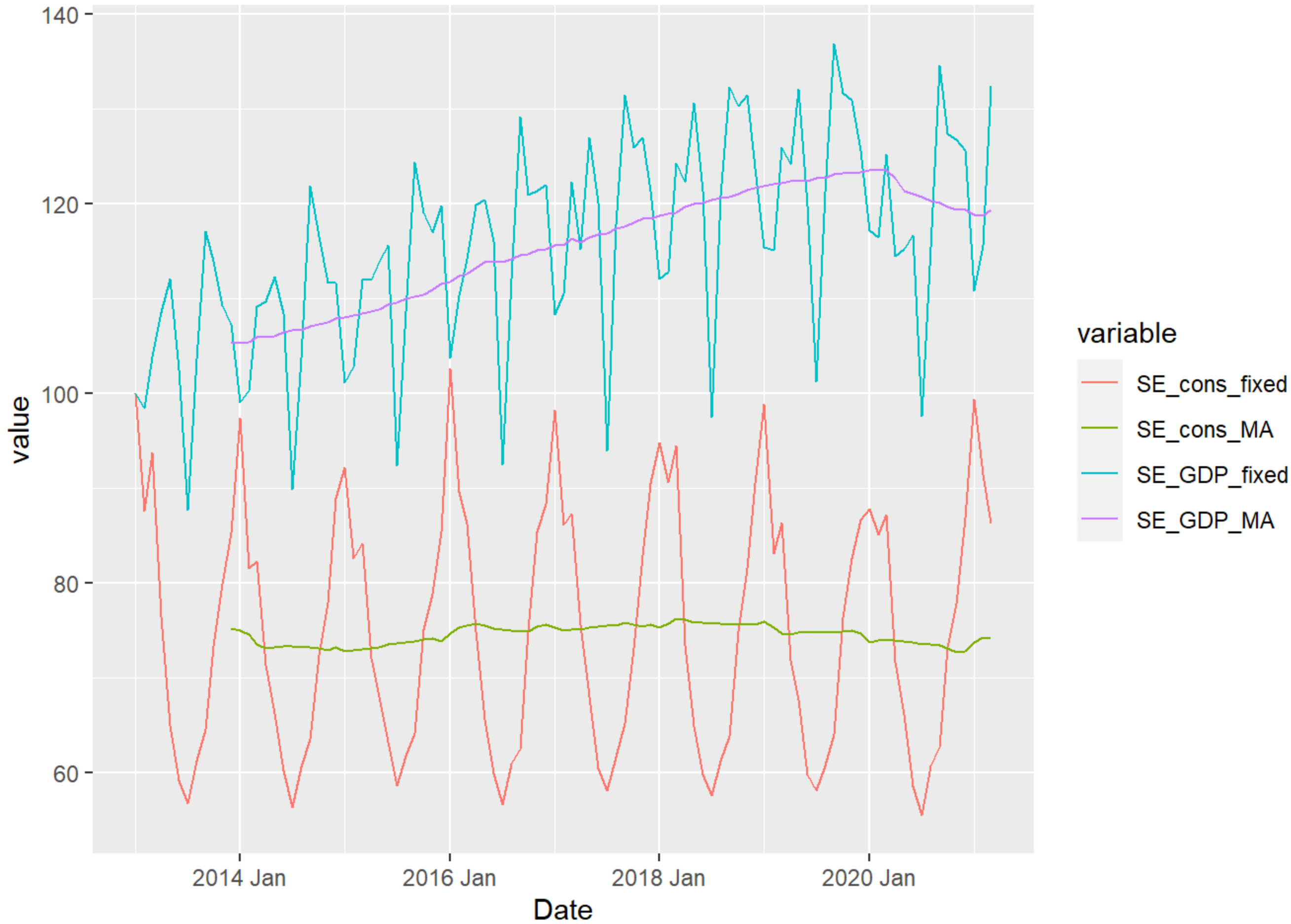


It can clearly be seen that monthly GDP indiator has much higher volatility compared to electricity consumption. More importantly, it has a clear upward trend over the period 2013 - 2020, unlike electricity consumption.

In order to have better look of each of those trend and seasonality in one go, I opted to fix both of their January 2013 value to 100 and kept the growth rate constant for the rest. I did this as electricity consumption data and GDP indicator have siginificantly different magnitudes, given that one of them is in MWh terms while the other is an indexed value.

```
## Jan 2013 fixed to 100 and plotted along with 12-month moving average
df_fixed_ts %>% dplyr::select(SE_GDP_fixed, SE_cons_fixed, SE_GDP_MA, SE_cons_MA, Dat
e) %>%
  pivot_longer(-Date, names_to="variable", values_to="value") %>%
  ggplot(aes(x=Date, y=value, color=variable)) +
  geom_line()
```

```
## Warning: Removed 22 row(s) containing missing values (geom_path).
```

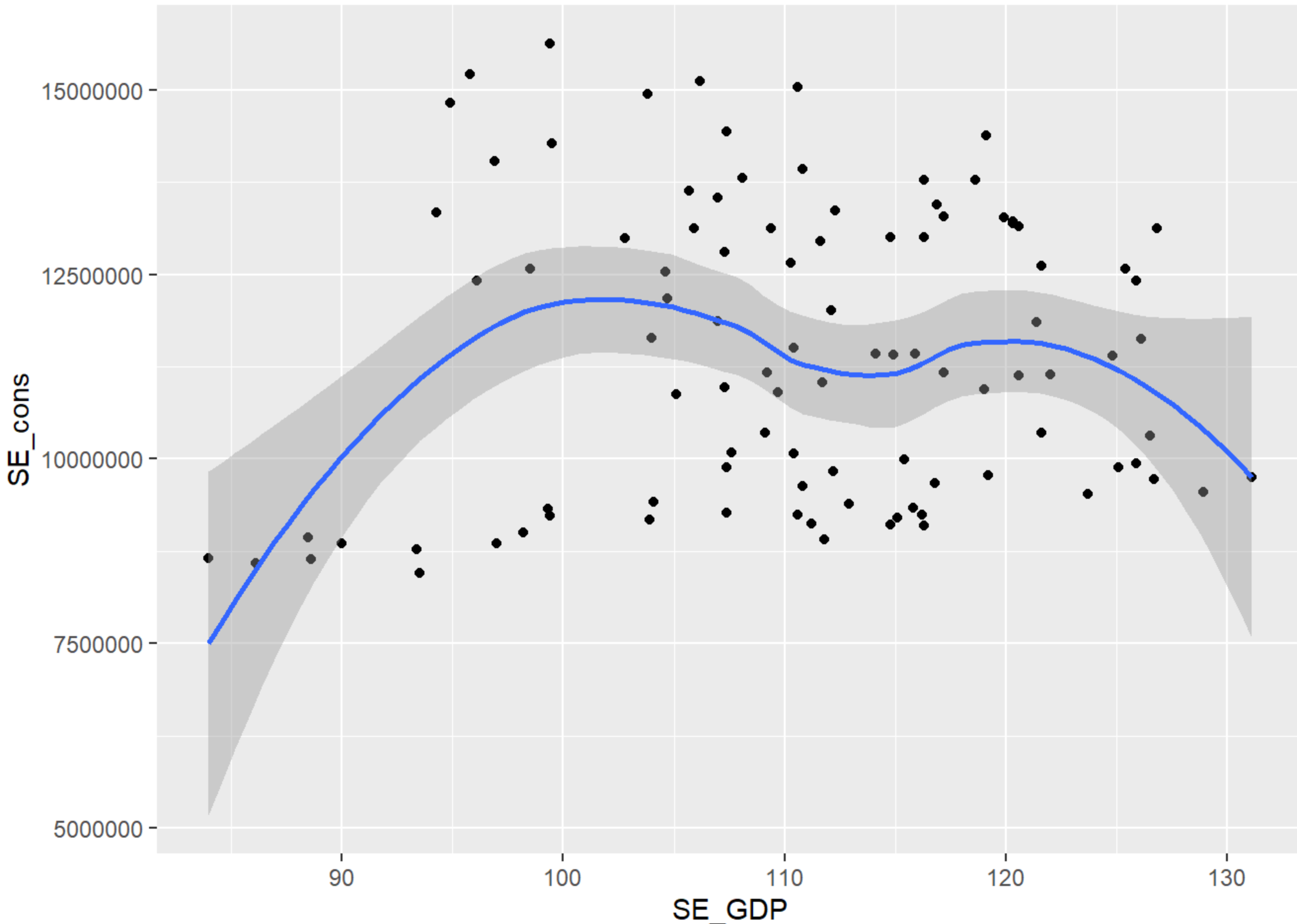



When plotted this way, the difference in the overall trend and volatility has become more apparent. One thing that can be also be noticed thanks to 12-month moving average trend is that GDP has started to go downwards from Q1 2020 onwards, most probably (almost 100% surely) due to Covid-19 outbreak.

In order to see the relationship between GDP indicator and electricity consumption, we can plot them to X-Y axes:

```
## The relationship between GDP and electricity consumption
df_ts %>% ggplot(aes(x=SE_GDP, y=SE_cons)) +
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



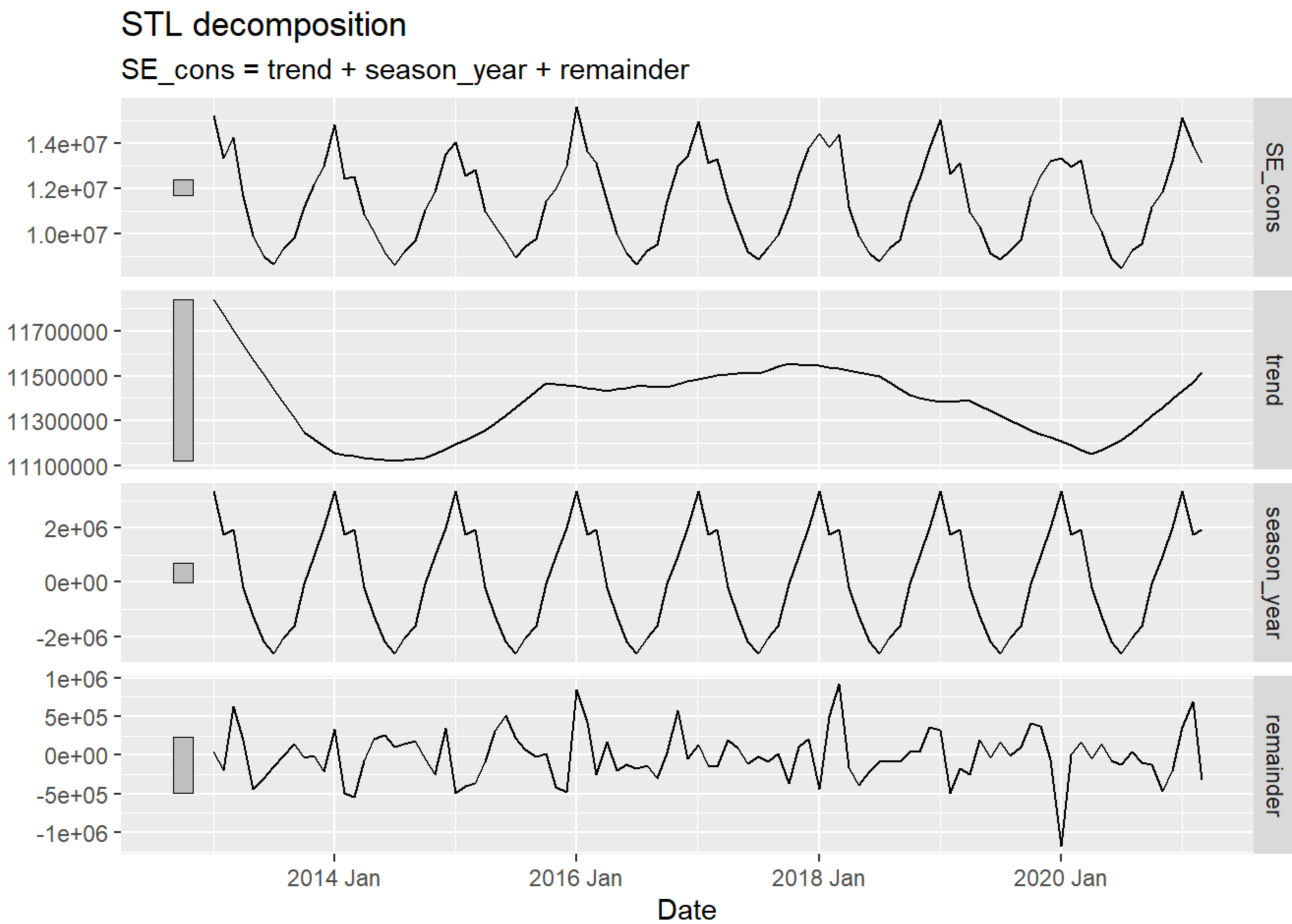
Local polynomial regression shows a floating and unclear pattern between them. Quite high number of outliers also give an impression that GDP indicator and electricity consumption in Sweden between 2013 and March 2021 are not that correlated.

Modelling and Forecasting

After having a glance at descriptive statistics of electricity consumption and GDP indicator, it is time to make use of the information being gathered for modelling purposes.

Along with summary statistics, STL (Seasonal Trend decomposition using Loess) decomposition can also be useful to obtain information before modelling phase. STL decomposition of electricity consumption in Sweden can be seen in the below:

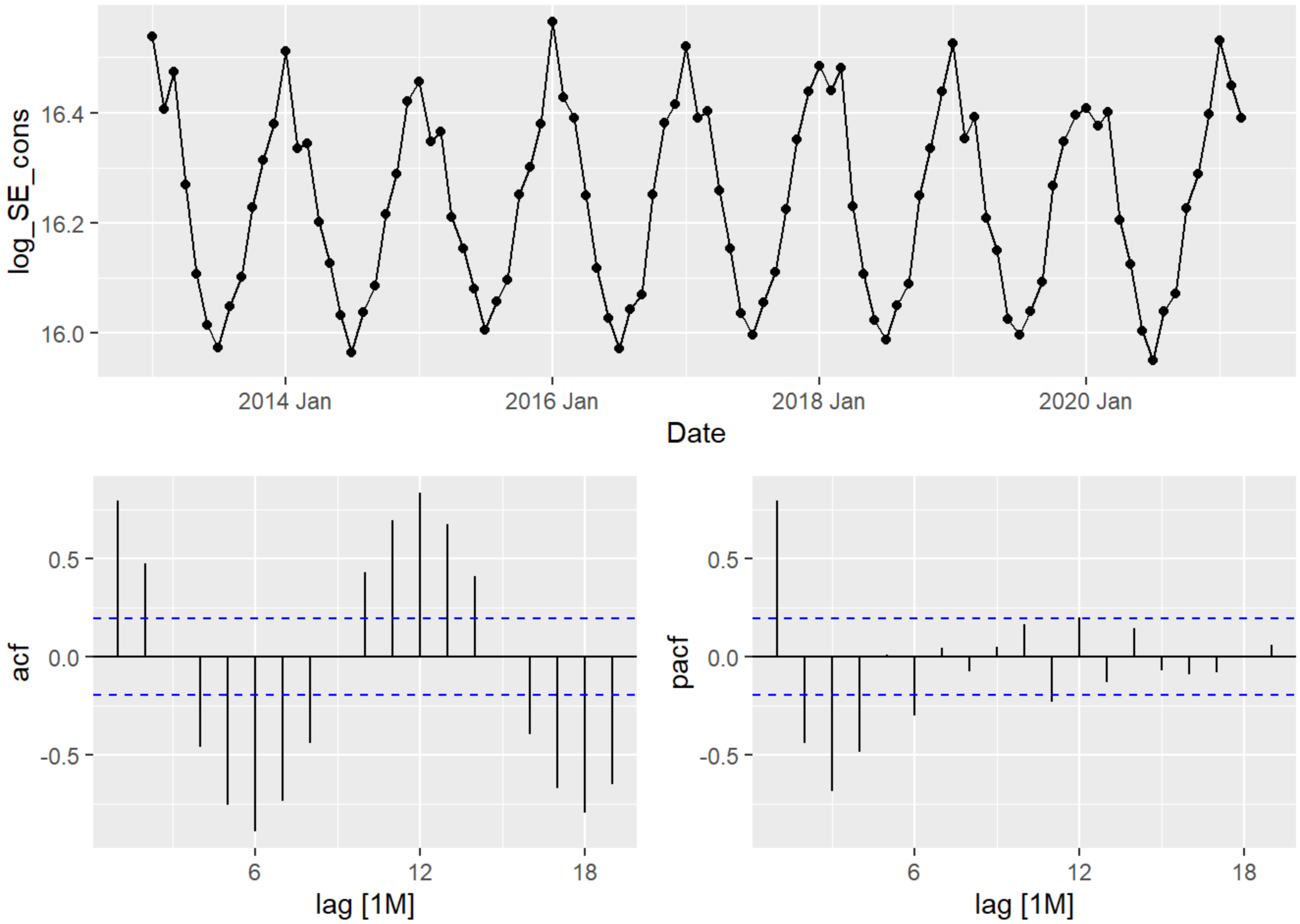
```
cons_comp = df_ts %>% model(
  STL(SE_cons ~ trend(window=21) + season(window="periodic"))
) %>% components

cons_comp %>% autoplot()
```



Similar with what was found at Descriptive Statistics, STL decomposition also shows that seasonal trend is particularly strong. Here trend window is set to 21 as advised by Hyndman & Athanasopoulos (2021). STL decomposition of GDP indicator could have also been looked at but since I will eventually forecast electricity consumption and use GDP indicator as an exogenous variable, I conclude what I have found out for GDP indicator at Descriptive Statistics satisfactory.

As mentioned before, so far I have found out that both variables demonstrate a strong seasonal pattern, along with an upward trend in GDP indicator. Besides, it is also obvious that both variables differ significantly in terms of magnitude as one of it is in MWh terms while the other one has indexed values. Due to these reasons, I opted to log-scale both variables as an initial step of modelling, relying on that it will decrease the magnitude difference and help stationerizing the dataset, since it is crucially important when it comes to time series modelling. Later, I will start with having a glance at ACF and pACF (Partial Autocorrelation) of log-scaled electricity consumption.

```
df_ts = df_ts %>% mutate(
  log_SE_cons = log(SE_cons),
  log_SE_GDP = log(SE_GDP))


df_ts %>% fill_gaps() %>% gg_tsdisplay(log_SE_cons, plot_type='partial')
```

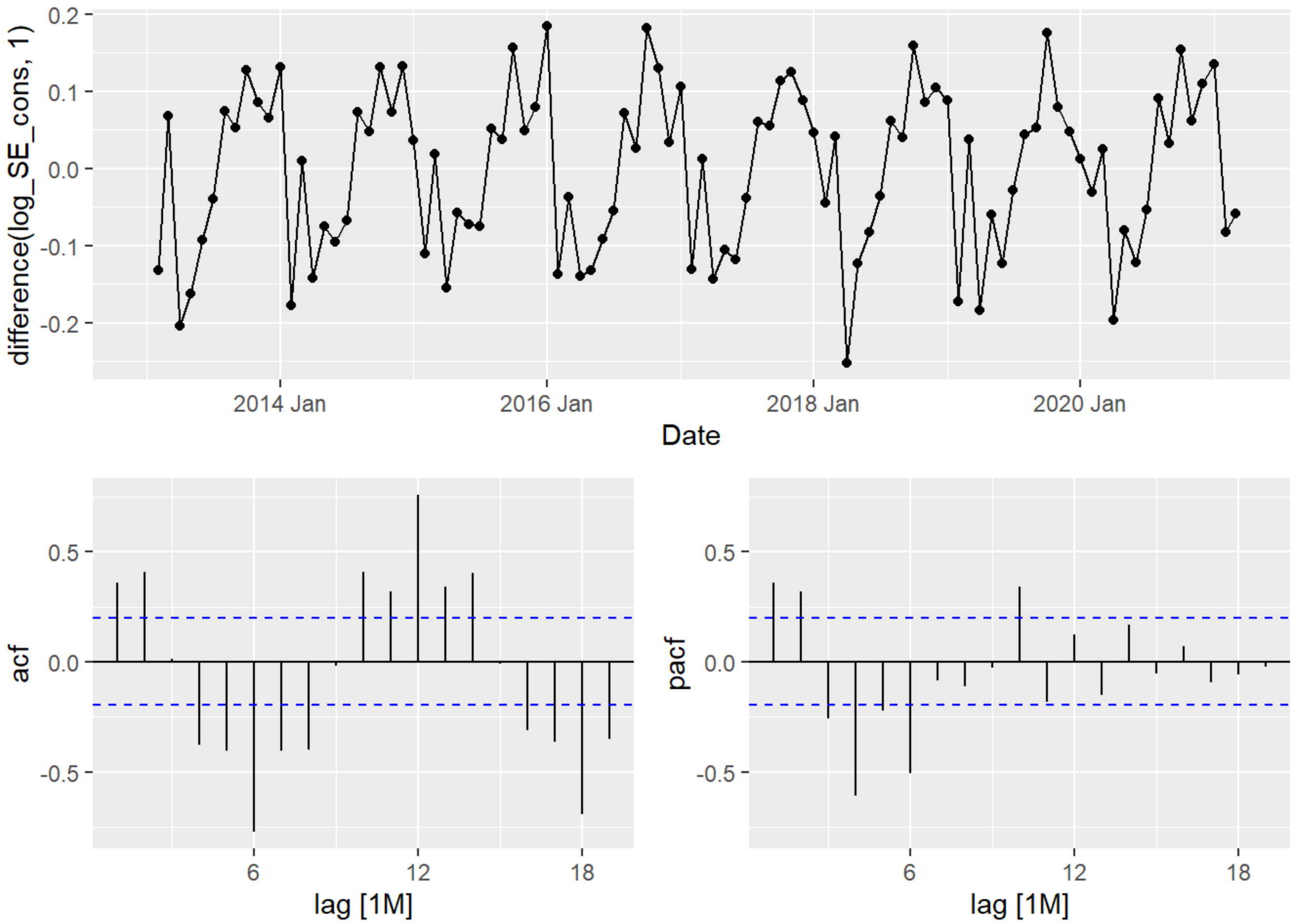


A quite strong temporal pattern can easily be detected when having a look at both ACF and pACF plots. Differentiation comes in handy when attempting to stationerize such series posting very strong temporal pattern. Therefore, I opted to plot the ACF and pACF of the first-differenced series as well.

```
df_ts %>% fill_gaps() %>% gg_tsdisplay(difference(log_SE_cons, 1), plot_type='partial
')
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

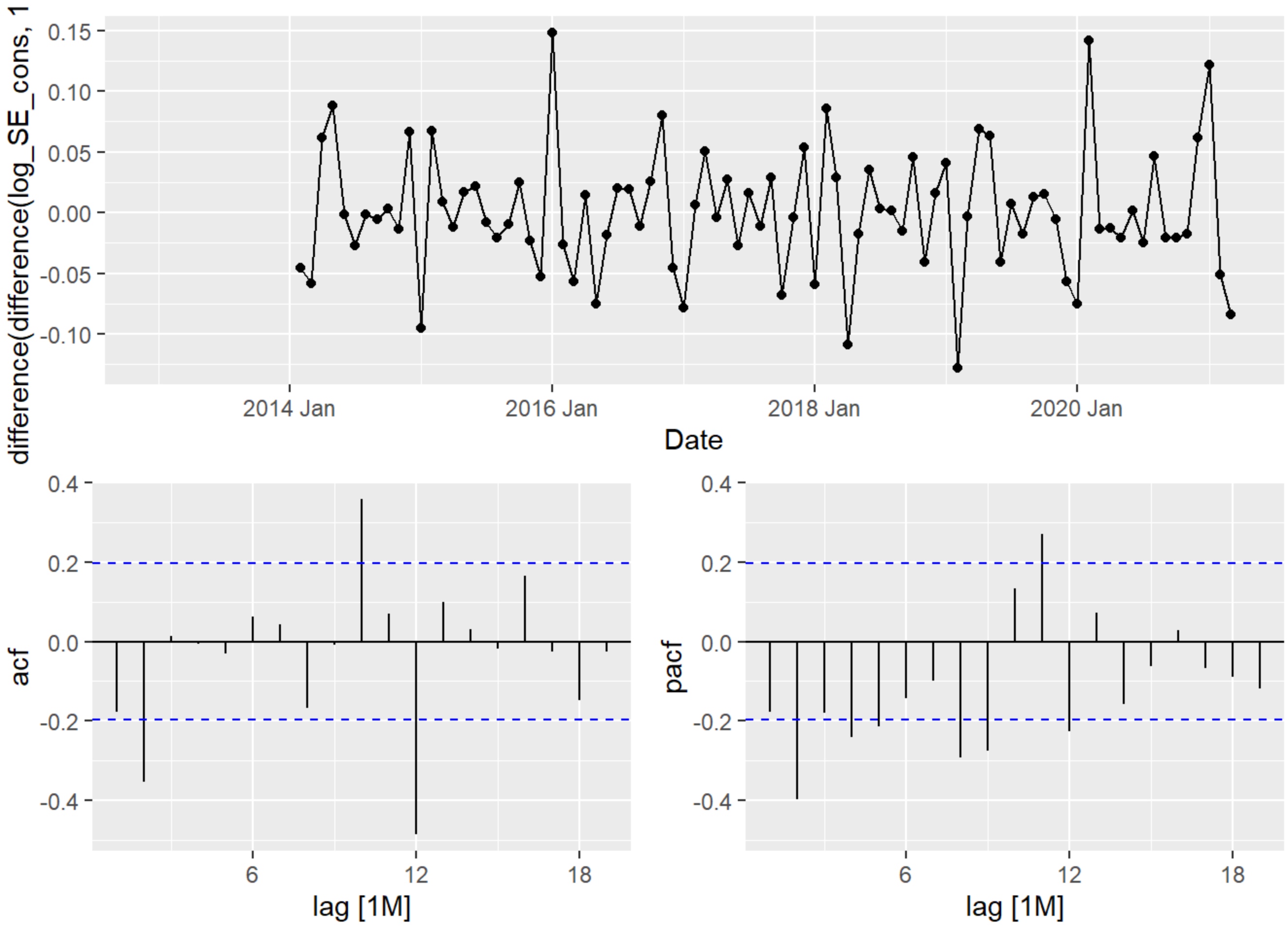


pACF plot shows first differencing helped out in terms of stationerizing the series but initial visual inspection suggests addition of AR(1) and AR(2) terms to the model. However, significant spikes at lags 6, 12, 18 suggests there is a seasonal pattern in series as well. Which is why, I want to continue experimenting by taking seasonal difference of the first-differenced series.

```
df_ts %>% fill_gaps() %>% gg_tsdisplay(difference(difference(log_SE_cons, 1), 12), pl
ot_type='partial')
```

```
## Warning: Removed 13 row(s) containing missing values (geom_path).
```

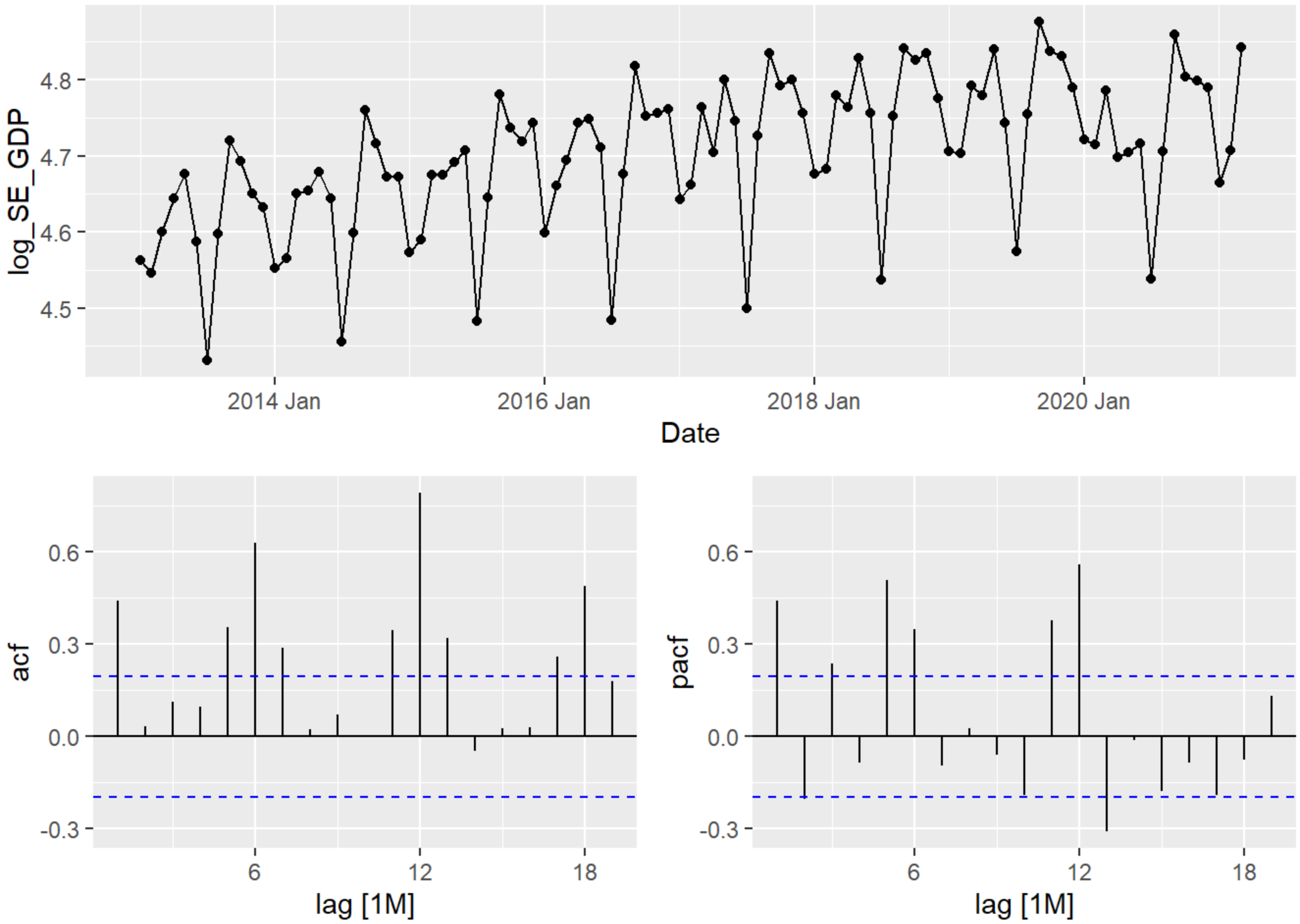
```
## Warning: Removed 13 rows containing missing values (geom_point).
```



ACF and pACF suggests along with both first and seasonal differentiation, AR(1) and possibly AR(2) plus one seasonal AR term can be introduced to the model. However, it must be noted that differentitaion increases the complexity of the model and therefore increases the variance, which may potentially make forecasts impractical. Therefore, during modelling phase, along with what I have found during experimentation, I will also let ARIMA() function to find a good fit automatically.

But beforehand, since I eventually inspect whether GDP indicator will make an impact on electricity consumption or not, I opted to follow the same procedure for GDP indicator in order to both see whether it makes sense and stationerize it.

```
df_ts %>% fill_gaps() %>% gg_tsdisplay(log_SE_GDP, plot_type='partial')
```

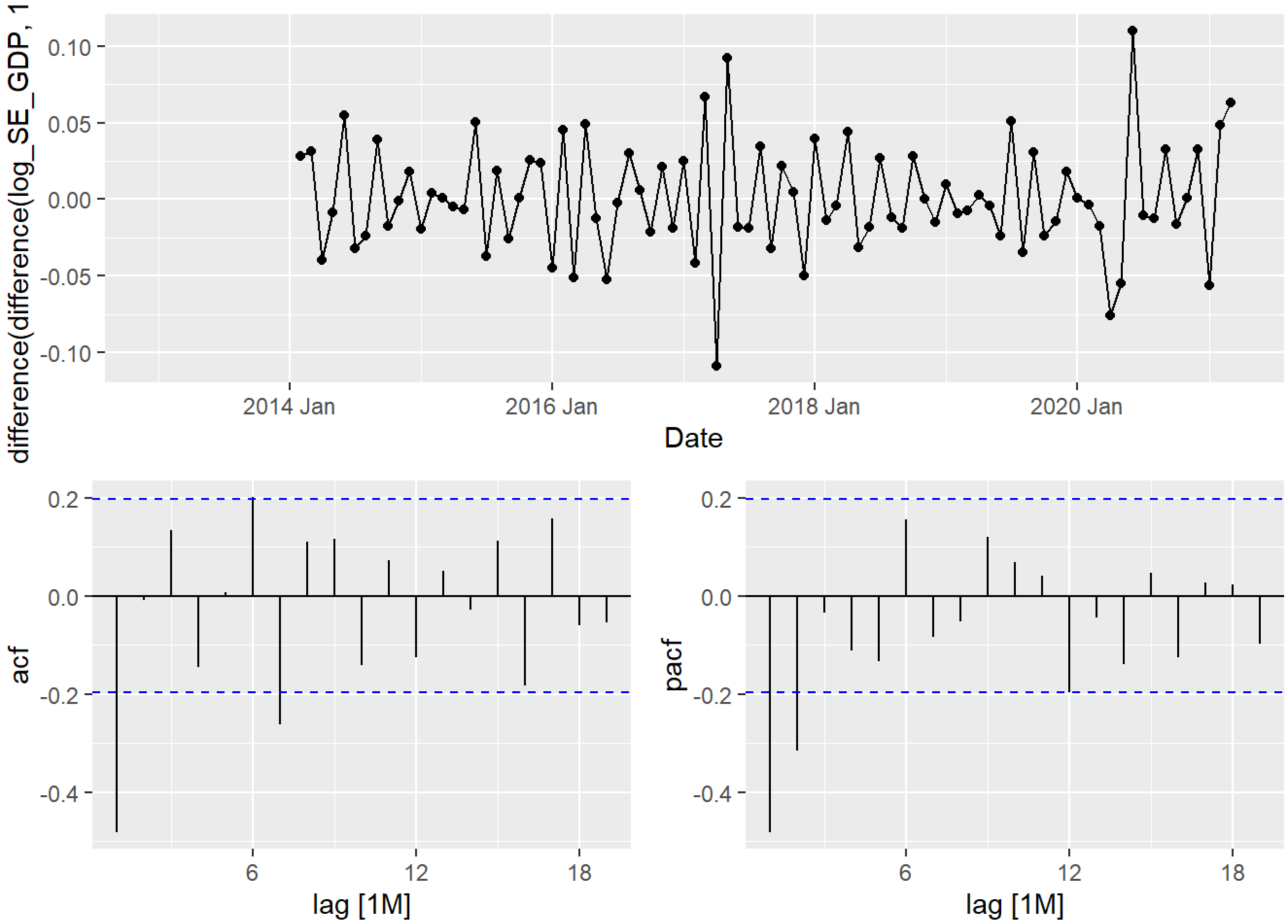


Both AR(1) and seasonal components can be seen significant in ACF and pACF plots and therefore I can conclude that along with electricity consumption, I can take both first and seasonal differences of GDP indicator as well.

```
df_ts %>% fill_gaps() %>% gg_tsdisplay(difference(difference(log_SE_GDP, 1), 12), plo
t_type='partial')
```

```
## Warning: Removed 13 row(s) containing missing values (geom_path).
```

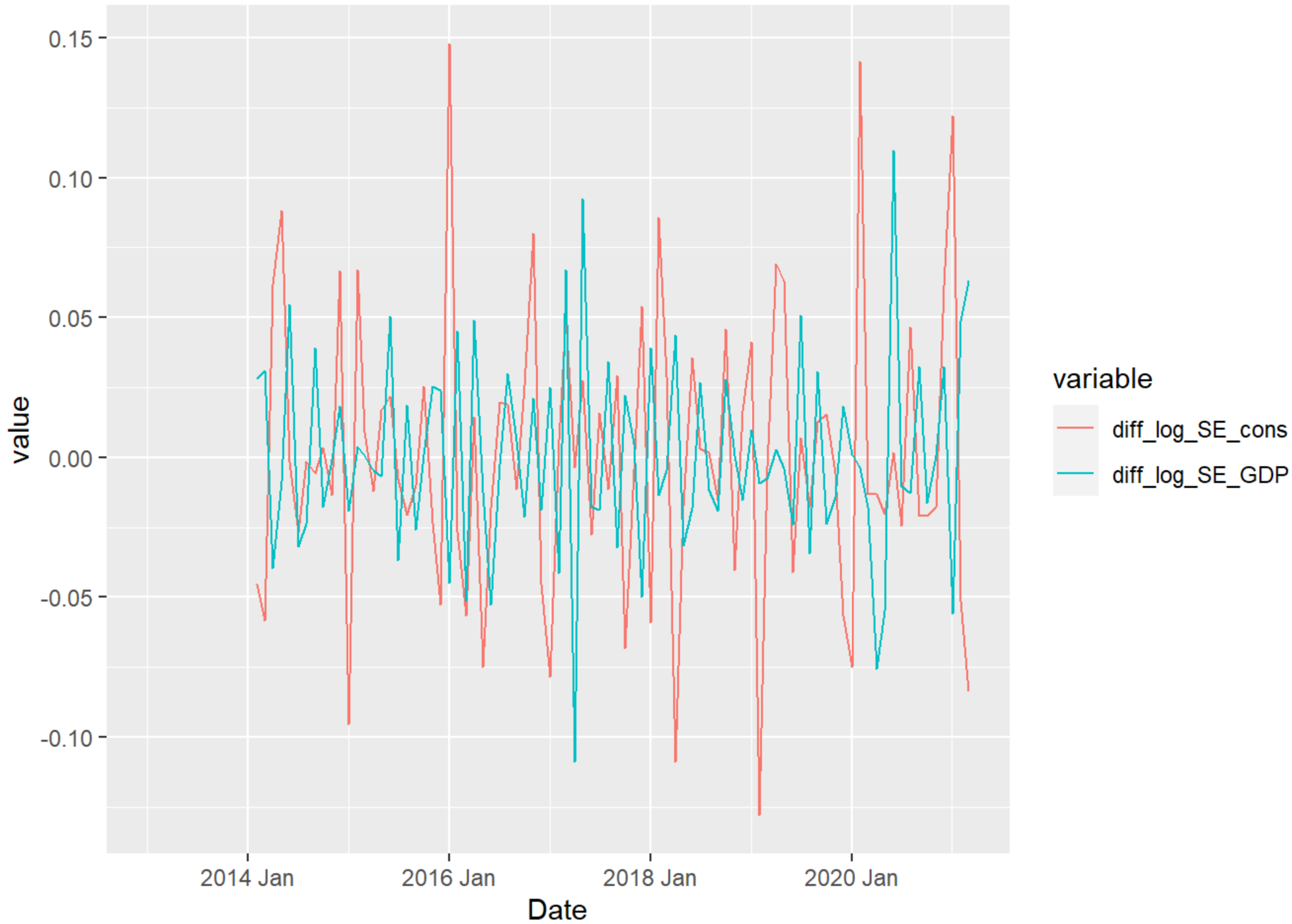
```
## Warning: Removed 13 rows containing missing values (geom_point).
```



Differentiations make GDP indicator series as stationerized as possible. So now, before getting into the modelling phase, I want to initiate another visual inspection in order to attempt to detect a relationship between electricity consumption and GDP indicator.

```
df_ts = df_ts %>% mutate(
  diff_log_SE_cons = difference(difference(log_SE_cons, 1), 12),
  diff_log_SE_GDP = difference(difference(log_SE_GDP, 1), 12))

df_ts %>% dplyr::select(diff_log_SE_cons, diff_log_SE_GDP, Date) %>%
  pivot_longer(-Date, names_to="variable", values_to="value") %>%
  ggplot(aes(x=Date, y=value, color=variable)) +
  geom_line()
```

```
## Warning: Removed 26 row(s) containing missing values (geom_path).
```



It can be observed that spikes are much more apparent for electricity consumption, and when one have a detailed look at the volatilities of both series, it can be said that the patterns each of them follow are quite different. Therefore, I can say that initial visual inspection suggests no evidence on the conservation hypothesis.

We can do the inspection in a statistically more fancier and perhaps more proper way by forming a linear regression between those.

```
reg = lm(diff_log_SE_cons ~ diff_log_SE_GDP, data=df_ts)
summary(reg)
```

```
##
## Call:
## lm(formula = diff_log_SE_cons ~ diff_log_SE_GDP, data = df_ts)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.129988 -0.027703 -0.005268  0.024843  0.140480
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0003186  0.0055102   0.058    0.954
## diff_log_SE_GDP -0.1822371  0.1534142  -1.188    0.238
##
## Residual standard error: 0.05109 on 84 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.01652,    Adjusted R-squared:  0.004813
## F-statistic: 1.411 on 1 and 84 DF,  p-value: 0.2382
```

When regressing first and seasonally differenced consumption series on the first and seasonally differenced GDP indicator, high p-value (0.238) suggests the impact of economic growth on energy consumption is insignificant. It seems there is a negative relationship between change in GDP indicator and change in electricity consumption (with an estimated coefficient -0.1822371.) but deriving causal interpretation out of this regression would be impractical since the relationship between those is statistically insignificant. It must be noted that a substantial amount of observations were lost during differencing procedures (13 observations, to be more precise.) and the model was initially run on 84 observations. Therefore, lack of sample size may also be a reason behind high p-value.

So far, I have experimented on the electricity consumption data and concluded that both first and seasonal differentiation may be helpful during modelling phase. In addition, no evidence were found on the existance of conservation hypothesis between electricity consumption and GDP indicator. Given these, I can start modelling electricity consumption in Sweden with a univariate model, which will make use of solely the past values of electricity consumption as explanatory variables.

As I briefly explained above, I will run both a model that I conclude while experimenting and a model that will be determined automatically by ARIMA() function. The model that I select manually will be ARIMA(2,1,0)(1,1,0).

```
fit <- df_ts %>%
  model(
    arima210110 = ARIMA(log_SE_cons ~ 0 + pdq(2,1,0) + PDQ(1,1,0)),
    auto = ARIMA(log_SE_cons)
  )
```

```
## Warning in sqrt(diag(best$var.coef)): NaNs produced
```

```
fit %>% select(arima210110)%>% report()
```

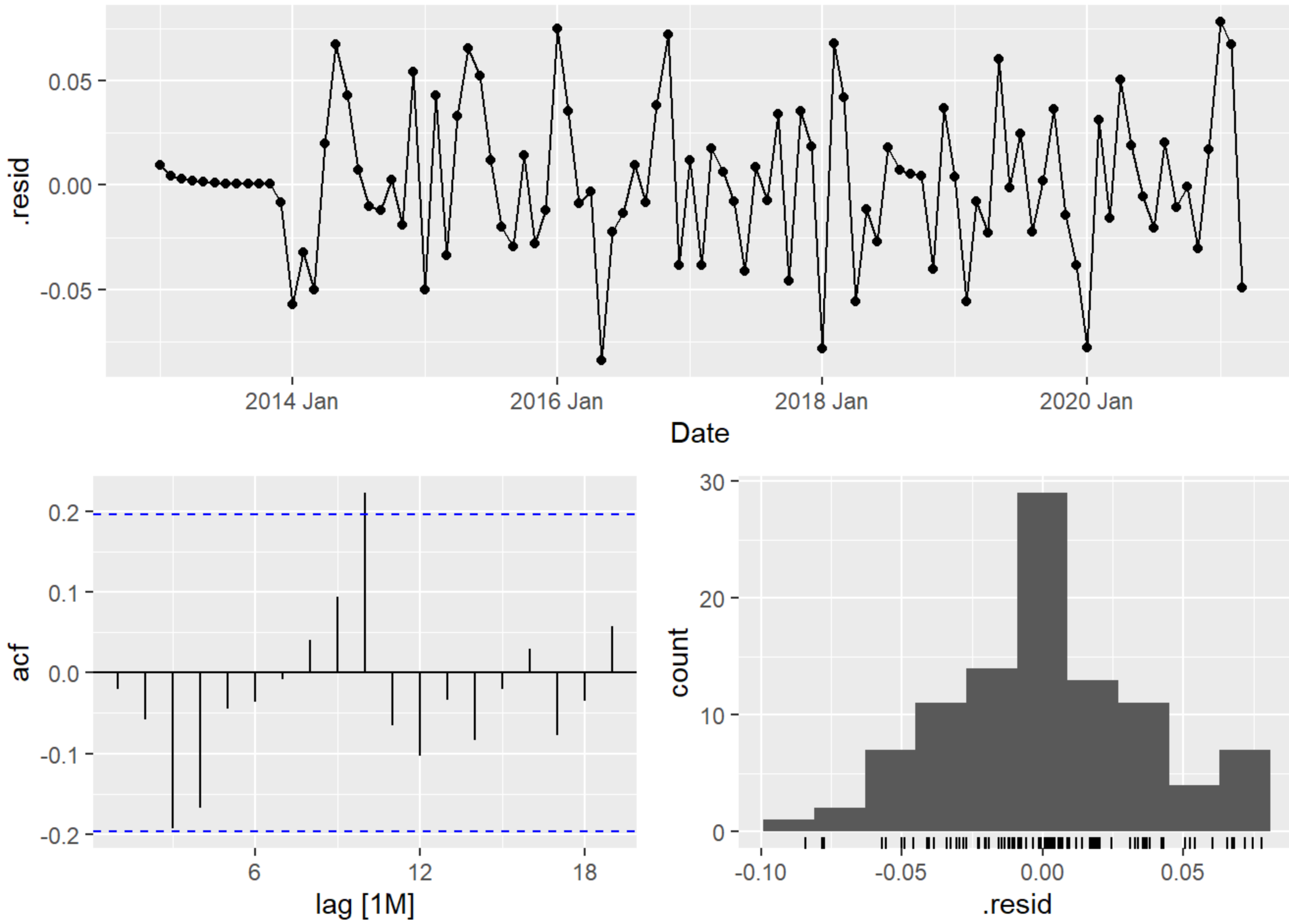
```
## Series: log_SE_cons
## Model: ARIMA(2,1,0)(1,1,0)[12]
##
## Coefficients:
##           ar1      ar2     sar1
##       -0.3114  -0.4179  -0.5851
## s.e.   0.1012   0.1001   0.0909
##
## sigma^2 estimated as 0.001472:  log likelihood=158.4
## AIC=-308.79   AICc=-308.3   BIC=-298.98
```

```
fit %>% select(auto)%>% report()
```

```
## Series: log_SE_cons
## Model: ARIMA(1,0,1)(2,1,1)[12]
##
## Coefficients:
##           ar1     ma1     sar1     sar2     sma1
##       -0.0415  0.4764  -0.1021  -0.0578  -0.8409
## s.e.   0.0033  0.1330      NaN   0.0048      NaN
##
## sigma^2 estimated as 0.0009392:  log likelihood=175.57
## AIC=-339.13   AICc=-338.08   BIC=-324.34
```

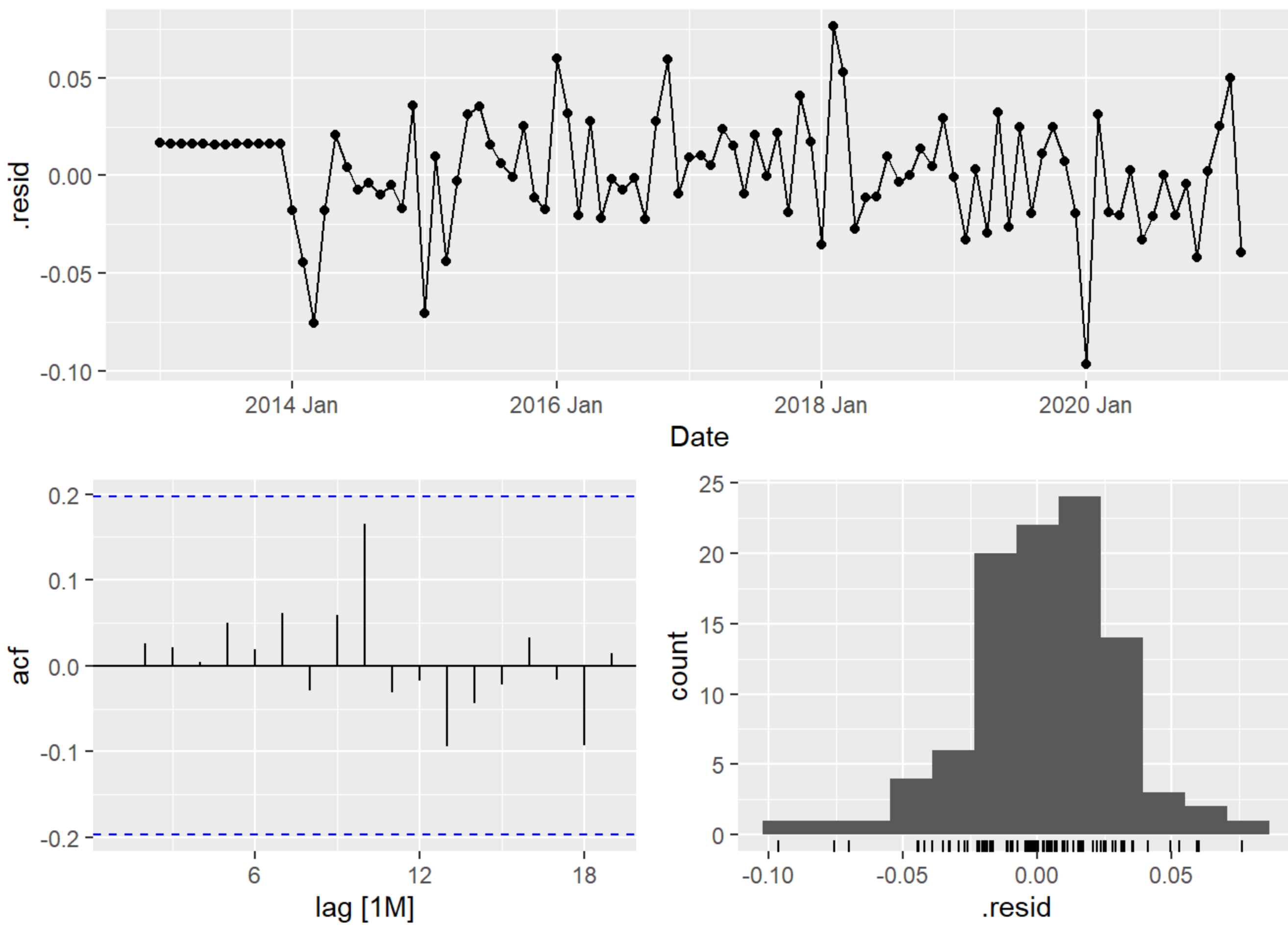
The results suggest that the model that was automatically detected by ARIMA() function, which is ARIMA(1,0,1)(2,1,1), has lower AICc, therefore I can now consider ARIMA(1,0,1)(2,1,1) as the main model. However, residuals should also be checked to determine which model is more appropriate to use in forecasting electricity consumption.

```
fit %>% select(arima210110) %>%
  gg_tsresiduals()
```



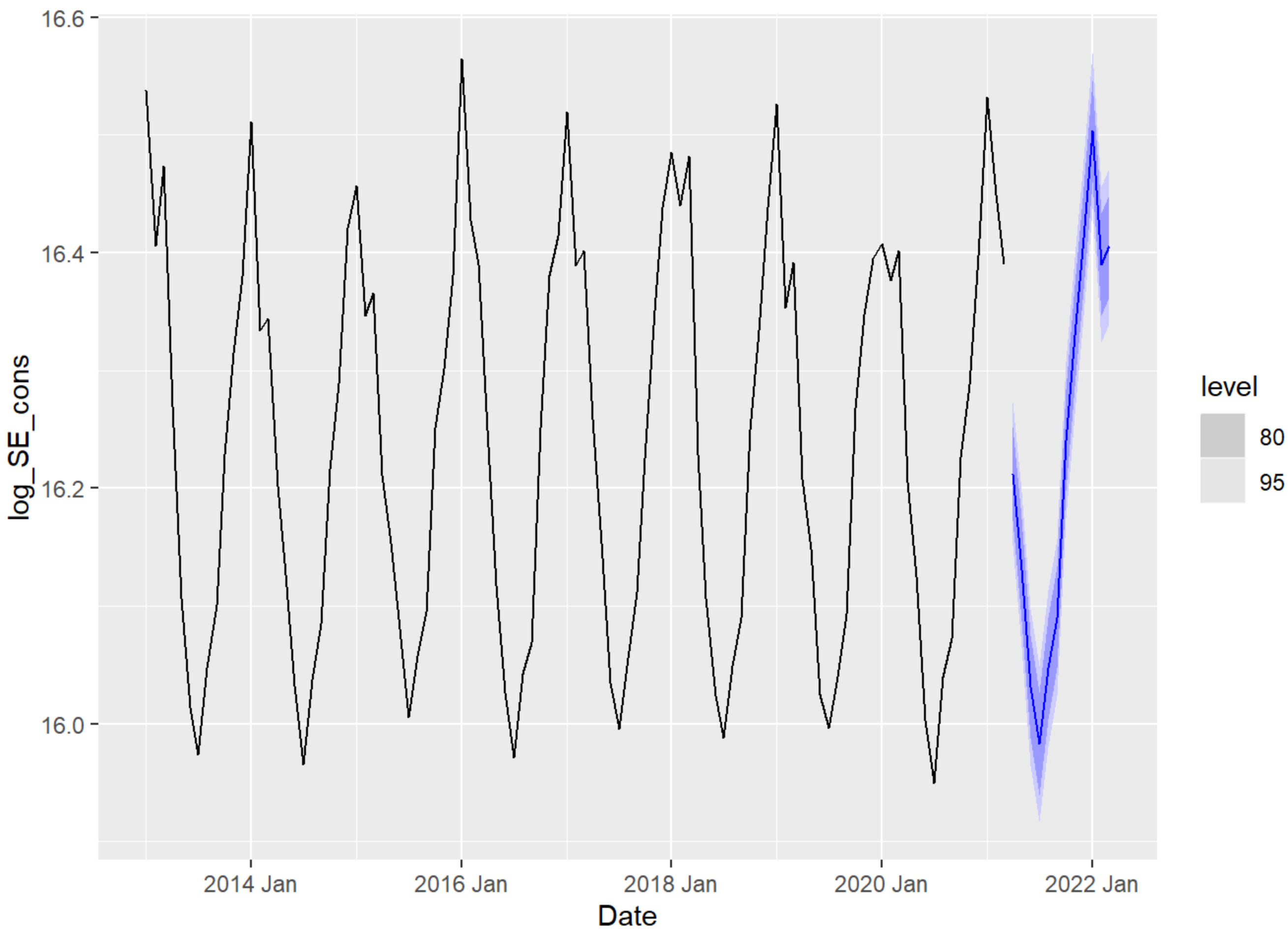
Lag 10 is quite significant in ARIMA(2,1,0)(1,1,0) along with a not-very-normally distributed residuals. It can also be argued after visually inspecting the volatility in residuals that residuals still demonstrate some kind of a certain temporal pattern. Same procedure should be run on the model ARIMA(1,0,1)(2,1,1) as well.

```
fit %>% select(auto) %>%
  gg_tsresiduals()
```



Here it can be seen obviously that residuals are much randomly distributed through time along with a quite normal distribution. No significant spikes left as well. Therefore, I can conclude ARIMA(1,0,1)(2,1,1) would be the best-fit to forecast electricity consumption in Sweden. Now, I will forecast the following year and see whether it makes sense or not.

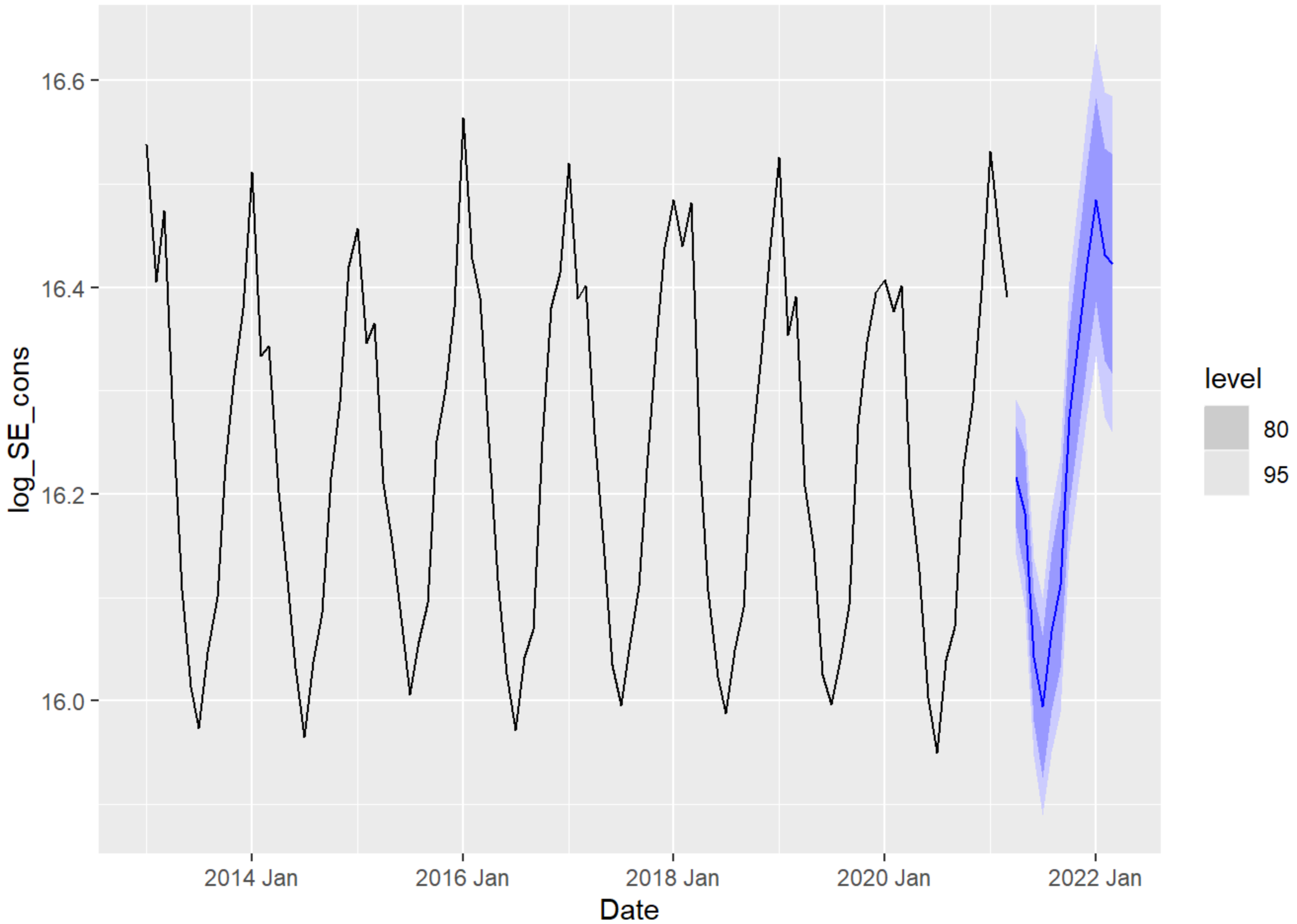
```
fit %>% select(auto) %>%forecast(h=12) %>% autoplot(df_ts)
```



As I discussed above, only seasonal differencing make the model somehow less complex which led the model to end up with a quite practical forecast interval. The temporal pattern that forecasted figured demonstrate also makes sense, given the overall flat trend that electricity consumption demonstrates throughout 2014 and March 2021.

Regarding forecast intervals, it can be demonstrated how impractical it can go by making the model uselessly more complex by forecasting the next 12 months with the model that I initially eliminated, in other words, the one that I conclude with some experimentation. Perhaps, I should have started the experimentation with a seasonal difference, instead of first difference, but luckily fpp3 package enables one to select a good fit with automatic ARIMA() function.

```
fit %>% select(arima210110) %>%forecast(h=12) %>% autoplot(df_ts)
```



One can easily inspect visually how larger the forecast interval is, compared to ARIMA(1,0,1)(2,1,1), which makes this one eventually impractical, even though the volatility it suggests throughout 12 month horizon somehow makes sense. While point forecast somehow makes sense, the forecast interval - for instance for the summer of 2021 - suggests either an almost all-time low or all-time record, which eventually makes it so impractical for potential decision makers.

After undoubtfully concluding that ARIMA(1,0,1)(2,1,1) is superior compared to ARIMA (2,1,0)(1,1,0), I can now continue with building a multivariate forecasting model by making use of GDP indicator as an exogenous variable.

Initially, I found evidence on the insignificant relationship between electricity consumption and GDP indicator for Sweden. In order to get the best out of modelling, I will leave ARIMA() function to automatically find out the best fit.

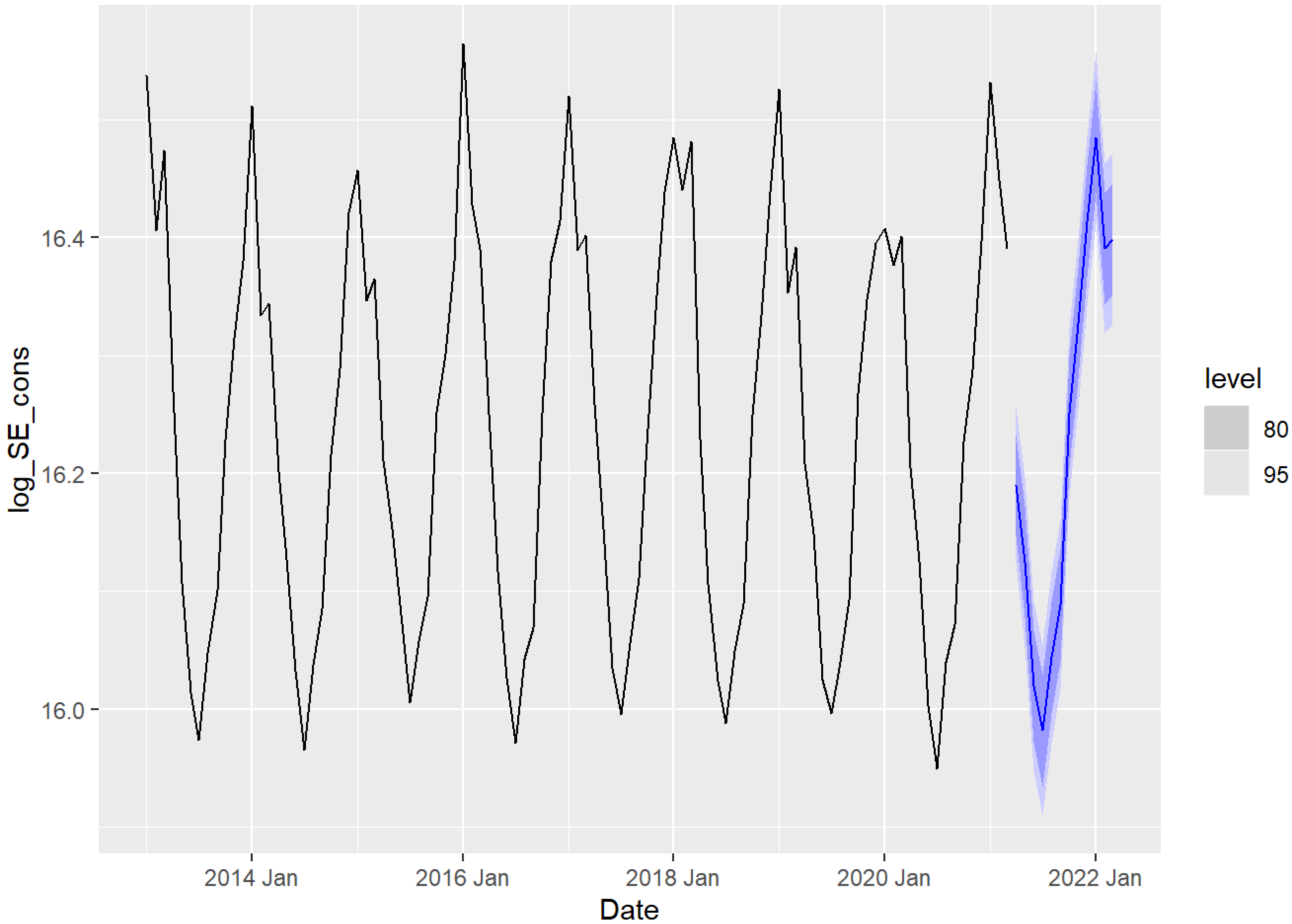
```
fit_GDP <- df_ts %>%
  model(
    arima_GDP = ARIMA(log_SE_cons ~ log_SE_GDP)
  )

fit_GDP %>% select(arima_GDP)%>% report()
```

```
## Series: log_SE_cons
## Model: LM w/ ARIMA(0,0,1)(2,1,0)[12] errors
##
## Coefficients:
##          ma1     sar1     sar2  log_SE_GDP
##       0.4630  -0.7217  -0.3571      0.0494
## s.e.  0.1004   0.1121   0.1142      0.0814
##
## sigma^2 estimated as 0.001117:  log likelihood=171.98
## AIC=-333.97   AICc=-333.22   BIC=-321.64
```

Automatic procedure of ARIMA() function suggests a model of log_SE_GDP (log-scaled monthly GDP indicator) + ARIMA (0,0,1)(2,1,0) as the best-fit. Similar with the univariate model that I initially concluded as the best-fit, this model also has one-time seasonally differenced variables. The similar complexity of this model compared to ARIMA (1,0,1)(2,1,1) and AICc, I have an initial idea that it will provide a similar forecast figures.

To run a forecast on the values of exogenous variables that have not yet been observed, I initially have to come up with a scenario, meaning that I need to build GDP indicator values for the period April 2021 and March 2022. Given the ongoing recovery along with mass vaccination programme throughout Europe against Covid-19, I expect a 5% annual GDP growth for each month. In other words, each observation will be 5% higher than the one that was observed on the corresponding month of the preceding year. Below code demonstrates how those values are generated and used in the forecasting phase.

```
scen = new_data(df_ts, 12) %>% mutate(
  log_SE_GDP = log(df_ts$SE_GDP[88:99]*1.05)
) ## Scenario is based on a 5% GDP growth for each month compared to the correspondin
g month of the preceding year.

forecast1 = forecast(fit_GDP, new_data=scen)
forecast1 %>% autoplot(df_ts)
```



As expected, I have reached a very similar forecast with multivariate forecast model, compared to ARIMA (1,0,1)(2,1,1).

Synthesis of Findings and Conclusions

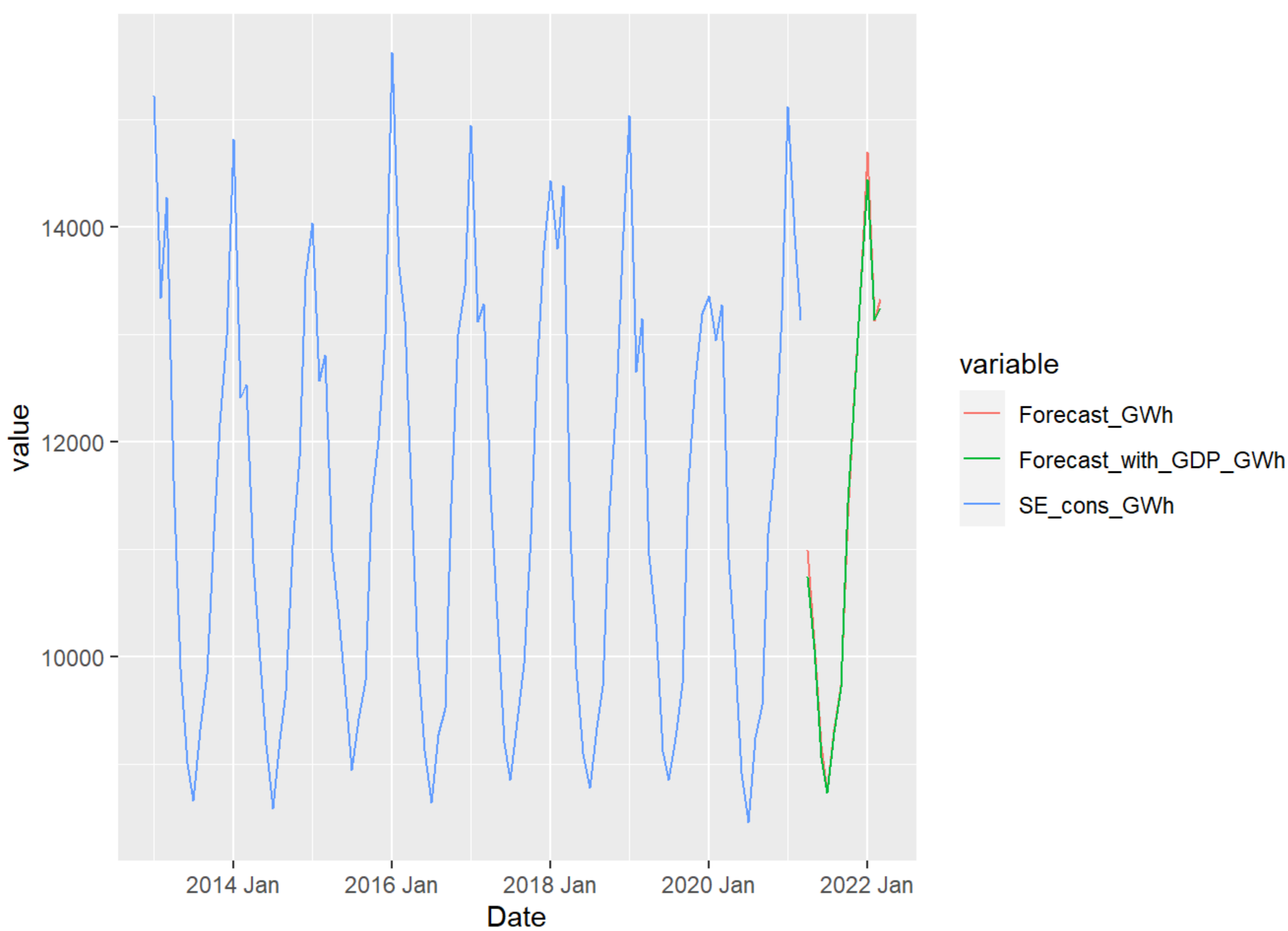
At Descriptive Statistics, I have found out some evidence that there might be no significant relationship, which suggests neutrality hypothesis, between electricity consumption and GDP indicator, when monthly figures from January 2013 to March 2021 taken into consideration. Strong temporal patterns were observed in both electricity consumption and GDP indicator. For stationerizing purposes, both variables log-scaled at the first place. ACF and pACF plots suggested both lagged and seasonally lagged patterns for electricity consumption. One univeriate model was determined after some experimentation over ACF and pACF plots, while an alternative univariate model was also built by relying on the automatic procedures of ARIMA() function. It was concluded that automatic procedure found a better fit, which is ARIMA(1,0,1)(2,1,1).

In the meantime, first and seasonally differenced log-scaled electricity consumption and GDP indicator were regressed, and found out clear evidence that changes in GDP indicator has no significant impact on the changes in electricity consumption. Along with this information, a multivariate model with GDP indicator as an exogenous variable were built and ARIMA() function's automatic procedure found a best-fit that derives very similar forecast compared to ARIM(1,0,1)(2,1,1), relying on the scenario of an annual GDP growth of 5% compared to the corresponding month of the previous year.

These made me to conclude that there is a substantial evidence suggesting neutrality hypothesis for the relationship between electricity consumption and economic growth (by making use of GDP indicator).

```
forecast = fit %>% select(auto) %>%forecast(h=12)
forecast["Forecast"] = exp(forecast$.mean)
forecast1["Forecast_with_GDP"] = exp(forecast1$.mean)

forecast_merged = forecast1 %>% full_join(forecast, by=c("Date"))
forecast_merged = forecast_merged %>% select(Date, Forecast, Forecast_with_GDP)

df_ts_merged = df_ts %>% full_join(forecast_merged, by=c("Date"))

## Converting to GWh
df_ts_merged = df_ts_merged %>% mutate(
  SE_cons_GWh = SE_cons/1000,
  Forecast_GWh = Forecast/1000,
  Forecast_with_GDP_GWh = Forecast_with_GDP/1000)

df_ts_merged %>% dplyr::select(SE_cons_GWh, Forecast_GWh, Forecast_with_GDP_GWh, Dat
e) %>%
  pivot_longer(-Date, names_to="variable", values_to="value") %>%
  ggplot(aes(x=Date, y=value, color=variable)) +
  geom_line()
```

```
## Warning: Removed 210 row(s) containing missing values (geom_path).
```



Above, log-scaled forecasted figures -by both ARIMA(1,0,1)(2,1,1) and multivariate model with GDP indicator- were inversed by taking their exponents and divided by 1000 to convert it to GWh to provide a simpler inspection of the visualization. It can be seen that addition of GDP indicator to a model made a marginal change in the forecast.

The project can be further improved by making use of a richer dataset that enables the researcher to make train and test split without worrying about losing substantial amount of observations for model-training purposes. In addition to this, a similar study can be conducted over a developing country to see whether GDP can still be used as an exogenous variable while forecasting electricity consumption, as this project relies on and suggests the fact that successful policies over increasing energy efficiency and reducing energy consumptions in the EU disconnected relationship between electricity consumption and economy growth.

References

Cowan, W. N., Chang, T., Inglesi-Lotz, R., & Gupta, R. (2014). The nexus of electricity consumption, economic growth and CO2 emissions in the BRICS countries. Energy Policy, 66, 359-368.

Cheng, Y. S., Wong, W. K., & Woo, C. K. (2013). How much have electricity shortages hampered China's GDP growth?. Energy Policy, 55, 369-373.

Sarwar, S., Chen, W., & Waheed, R. (2017). Electricity consumption, oil price and economic growth: Global perspective. Renewable and Sustainable Energy Reviews, 76, 9-18.

Magazzino, C. (2014). Electricity demand, GDP and employment: evidence from Italy. Frontiers in Energy, 8(1), 31-40.

Tursoy, T., & Resatoglu, N. G. (2016). Energy consumption, electricity, and GDP causality; the case of Russia, 1990-2011. Procedia Economics and Finance, 39, 653-659.

Karanfil, F., & Li, Y. (2015). Electricity consumption and economic growth: exploring panel-specific differences. Energy policy, 82, 264-277.

Bélaïd, F., & Abderrahmani, F. (2013). Electricity consumption and economic growth in Algeria: A multivariate causality analysis in the presence of structural change. Energy Policy, 55, 286-295.

Mohanty, A., & Chaturvedi, D. (2015). Relationship between electricity energy consumption and GDP: evidence from India. International Journal of Economics and Finance, 7(2), 186-202.

Behera, J. (2015). Examined the energy-led growth hypothesis in India: Evidence from time series analysis. Energy Economics Letters, 2(4), 46-56.

Seth, A. (2007). Granger causality. Scholarpedia, 2(7), 1667.

Chen, S. T., Kuo, H. I., & Chen, C. C. (2007). The relationship between GDP and electricity consumption in 10 Asian countries. Energy policy, 35(4), 2611-2621.

Ghosh, S. (2002). Electricity consumption and economic growth in India. Energy policy, 30(2), 125-129.

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 5/25/2021

Harvey, F. (2015, February 9). EU energy consumption level falls to 20-year low. The Guardian Newspaper

Hayes, A. (2021). Autoregressive Integrated Moving Average (ARIMA). Retrieved from: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

(https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp)

ec.europa.eu (2020). State of the Union: Commission raises climate ambition and proposes 55% cut in emissions by 2030. Retrieved from: https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1599 (https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1599)

ec.europa.eu (2014). 2030 climate & energy framework. Retrieved from: https://ec.europa.eu/clima/policies/strategies/2030_en (https://ec.europa.eu/clima/policies/strategies/2030_en)

ec.europa.eu (2019). Energy performance of buildings directive. Retrieved from: https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en)

ec.europa.eu (2021). Energy saving statistics. Retrieved from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_saving_statistics#cite_note-1 (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_saving_statistics#cite_note-1)