US Monthly Energy Sources Consumption Outlook

Final Project

Authors: Fulvio Zamengo
Federico Zanotti
Juan Carlos Arganaraz

Course : Business, Financial and Economy Data
Teacher : Prof. Mariangela Guidolin

Agenda

- Overview
- Main goals
- Gathering information
- Renewables
- Petroleum
- Coal
- Natural Gas
- Competition Model
- Results and Conclusions

Overview

US citizens use a lot of energy in homes, in businesses, and in industry, and to travel and transport goods. There are **four end-use sectors** that purchase or produce energy for their own **consumption** and not for resale:

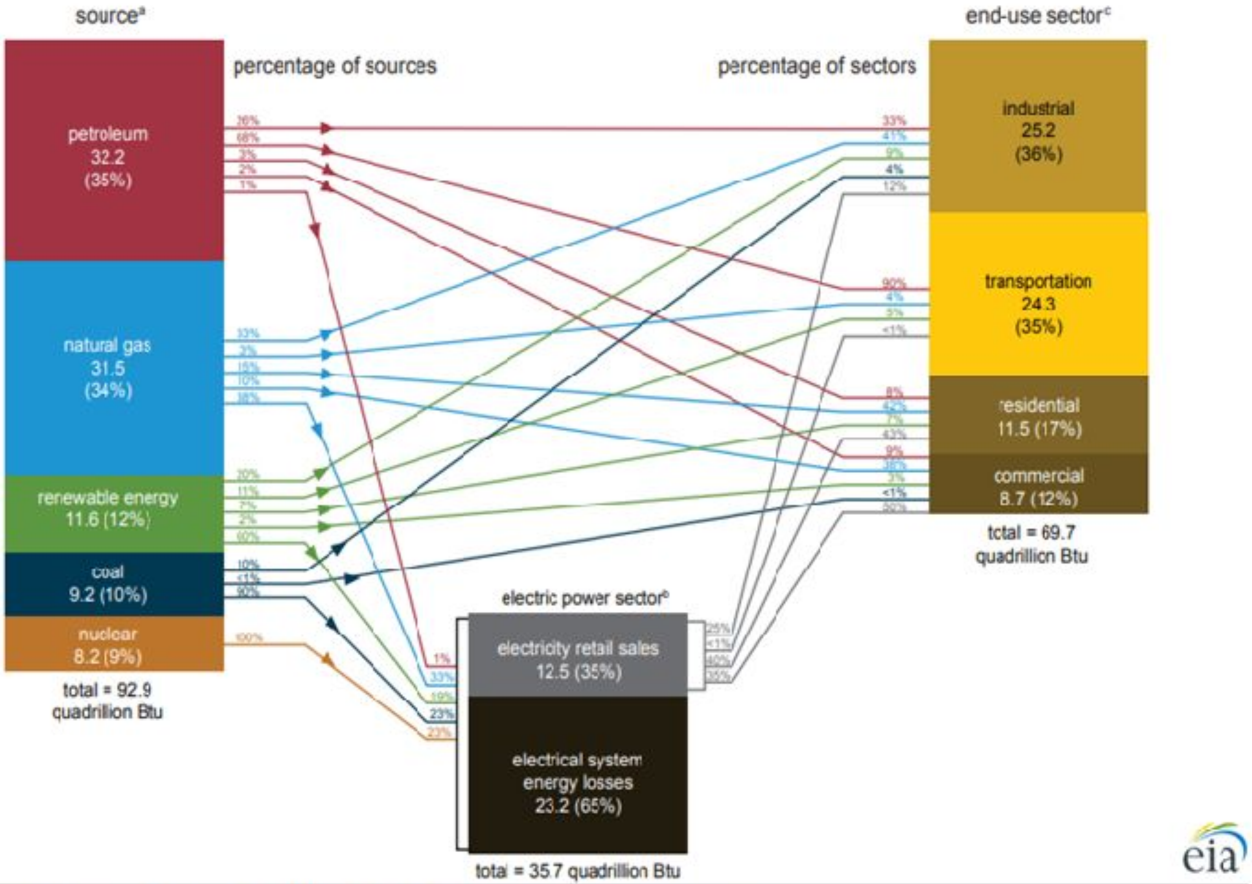
- The **residential** sector includes homes and apartments.
- The **commercial** sector includes offices, malls, stores, schools, hospitals, hotels, warehouses, restaurants, and places of worship and public assembly.
- The **industrial** sector includes facilities and equipment used for manufacturing, agriculture, mining, and construction.
- The **transportation** sector includes vehicles that transport people or goods, such as cars, trucks, buses, motorcycles, trains, aircraft, boats, barges, and ships.

These end-use sectors consume **primary energy** and also purchase and use most of the electricity (a secondary energy source) the electric power sector produces and sells.

In the present work, we will evaluate the consumption of four specific primary energy sources: **petroleum**, **coal**, **natural gas** and **renewables**.

U.S. energy consumption by source and sector, 2020

quadrillion British thermal units (Btu)



Sources: U.S. Energy Information Administration (EIA), *Monthly Energy Review* (April 2021), Tables 1.3 and 2.1-2.6.

Main Goals

1. Presents specific modeled projections of future energy sources consumptions in the United States.
2. Analyze the role of the natural gas as transition energy source.
3. Evaluate the process of decarbonization.
4. Impact of COVID 19 on trends

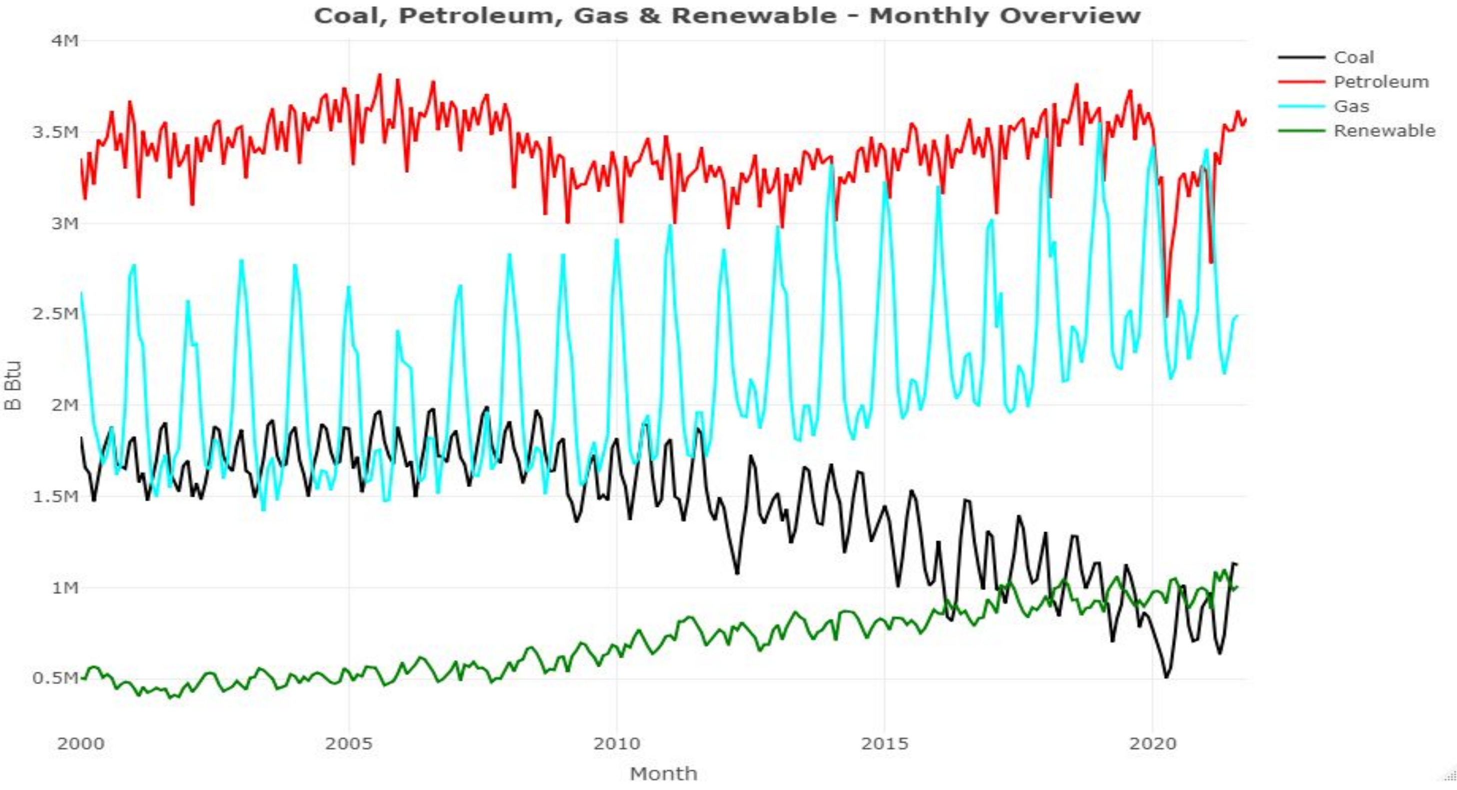
Gathering information

The datasets of each energy sources were collected from the U.S. Energy Information Administration (EIA) website. This Office collects, analyzes, and disseminates independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment.

The links to the website is the following:
https://www.eia.gov/totalenergy/data/monthly/index.php

The time-series data was collected in a monthly basis, the observation period start on January 1st 1973 till to July 2021. All the series have been truncated starting at 2000 because it’s when the renewables begins to increase.

Primary energy consumption comparison

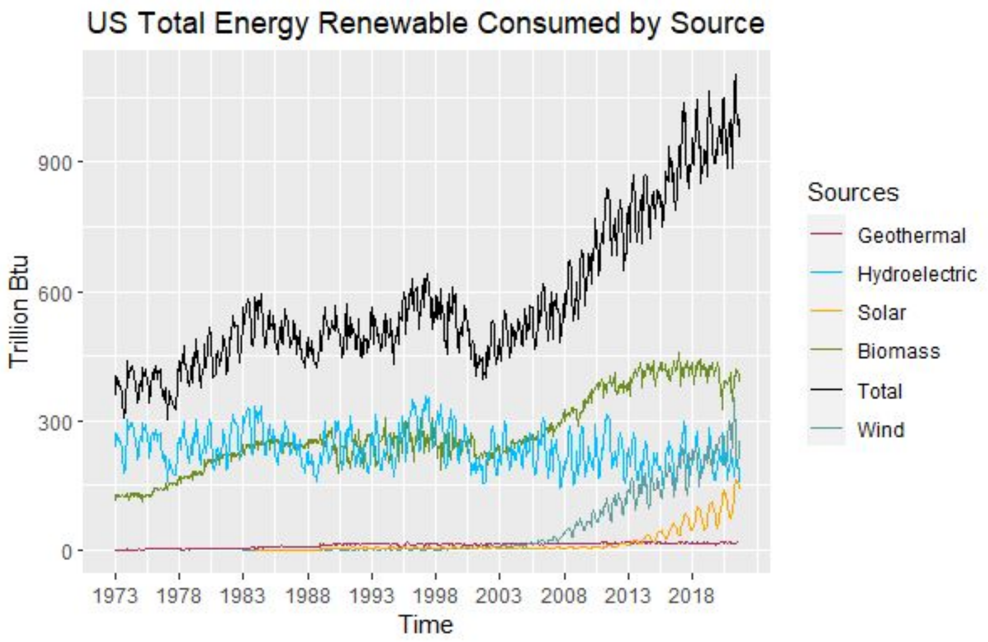


US Monthly Renewable Energy Consumption

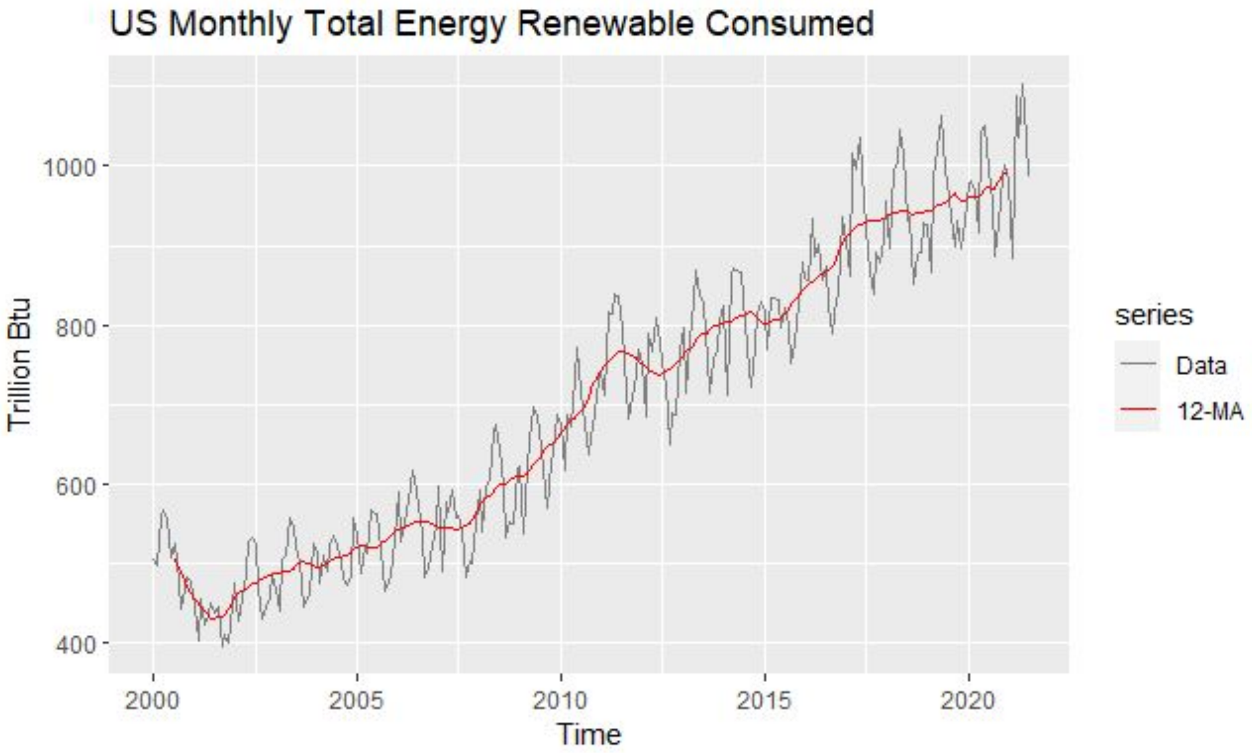
Renewable energy consumption consists of: conventional **hydroelectricity** net generation; **geothermal electricity** net generation , and **geothermal heat pump** and **geothermal direct** use energy; **solar thermal** and **photovoltaic electricity** net generation, and **solar thermal direct** use energy; **wind** electricity net generation.

Biomass sources for **energy** include: **wood and wood-derived** fuels consumption; biomass **waste** (i.e.: municipal solid waste from biogenic sources) consumption; **biofuels** (i.e. biodiesel, fuel ethanol) consumption.

All values were converted to Btu by multiplying by the total fossil fuels heat rate factors.

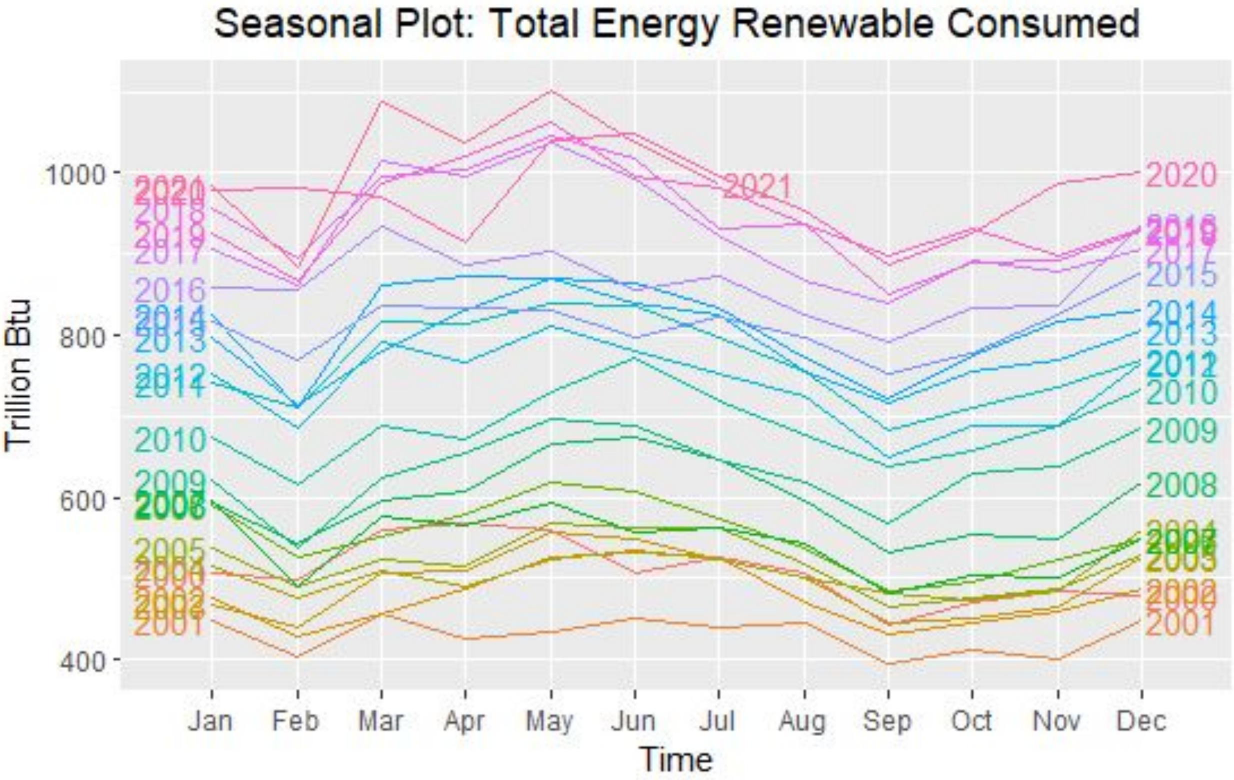


Renewable Energy Consumption time-series plot



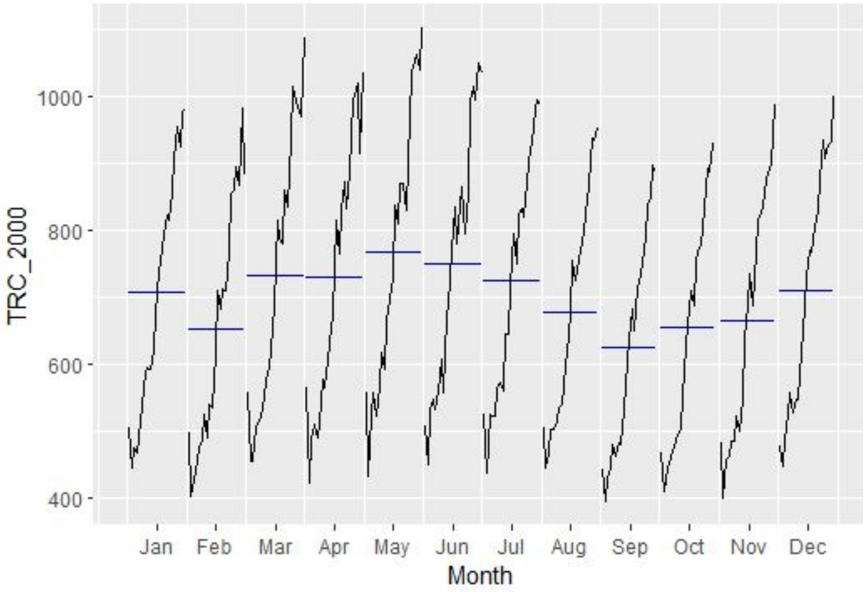
- ❏ There was an initial period, from Jan 2000 till to Nov 2001 when the consumption decrease.
- ❏ After the previous mentioned above, there is a clear long-term increase trend in the data. It was applied a 12-MA to better visualize the trend.
- ❏ There is also an strong seasonal pattern that increase in size every five years approximately.

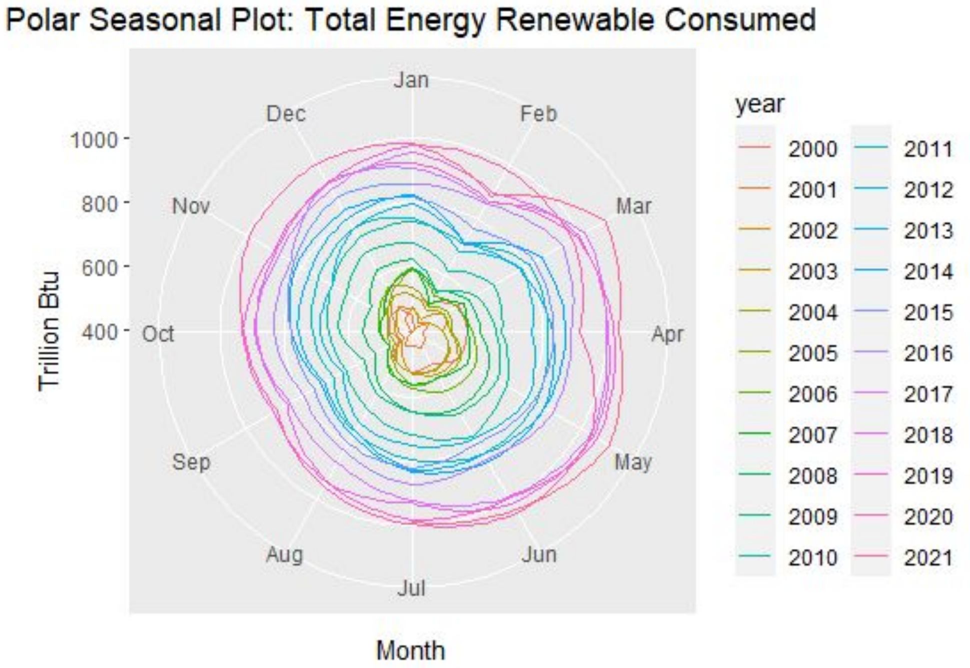
Seasonal plot



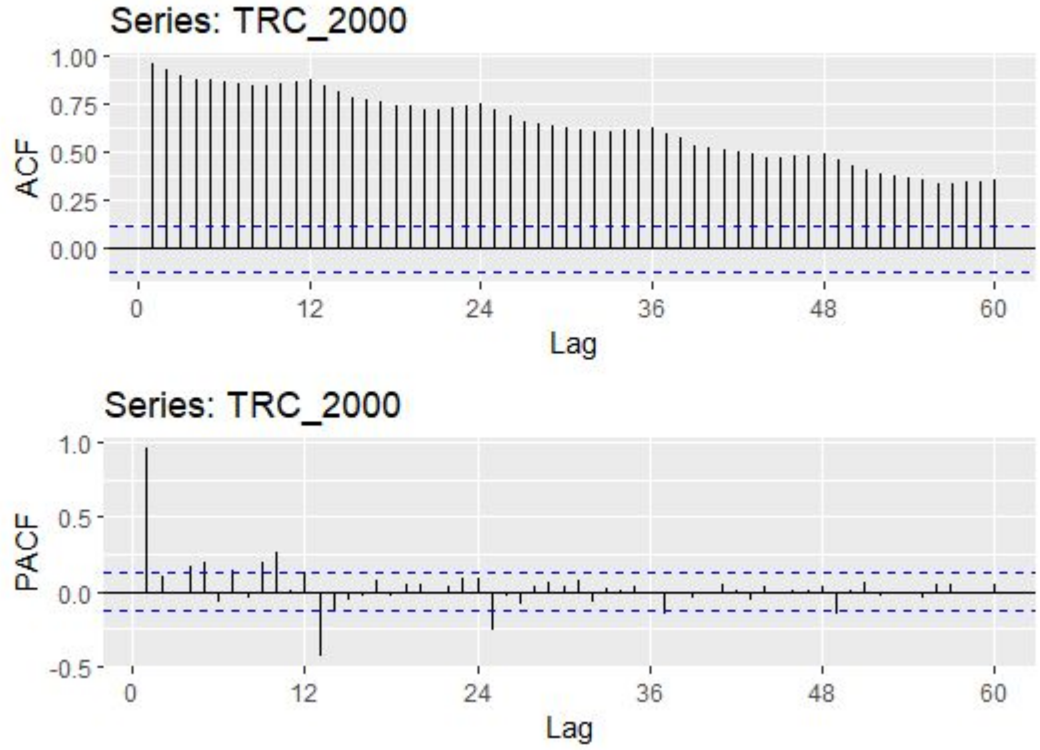
- ❏ It is clear that consumption is higher in May every year.
- ❏ From the other side, the energy consumption drops off in September and February.
- ❏ The behaviour explained above is due to the relation with the US weather: May-June is summer and southwest states experiences very hot temperatures. September (fall) and February (winter) are cold specially in northern states.

More seasonal plots





Check autocorrelations

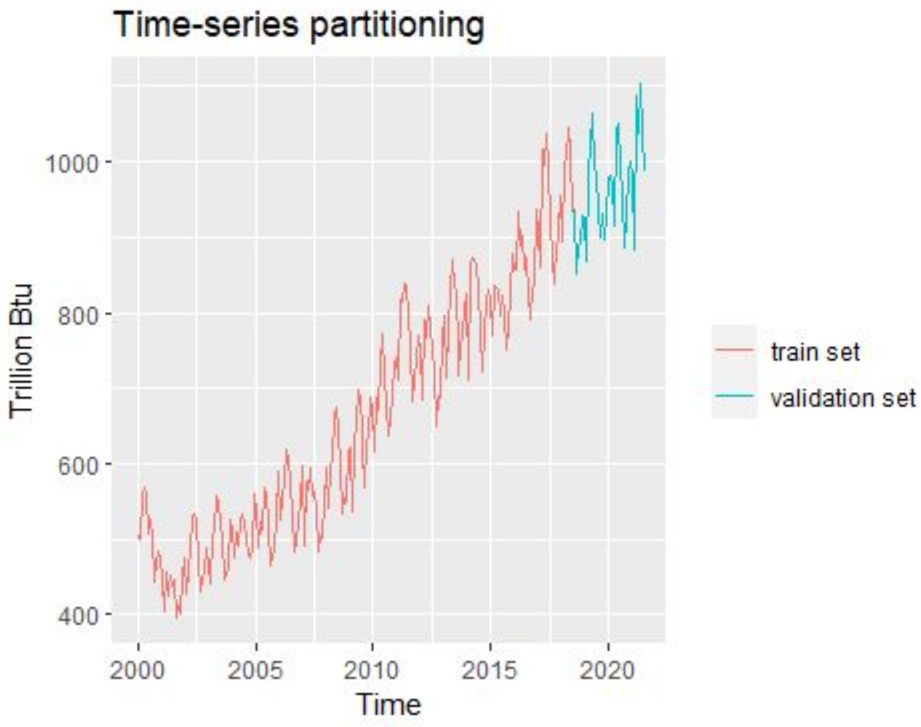


- ❑ From the correlogram chart (ACF), we have clear evidence that the time-series is autocorrelated.
- ❑ Absence of white noise consent us to perform forecasting.
- ❑ The time-series are seasonal, autocorrelations are larger for the seasonal lags.
- ❑ ACF decrease slowly as the lag increase due to the trend while the “pattern” shape is due to seasonality

Modelling: testes models

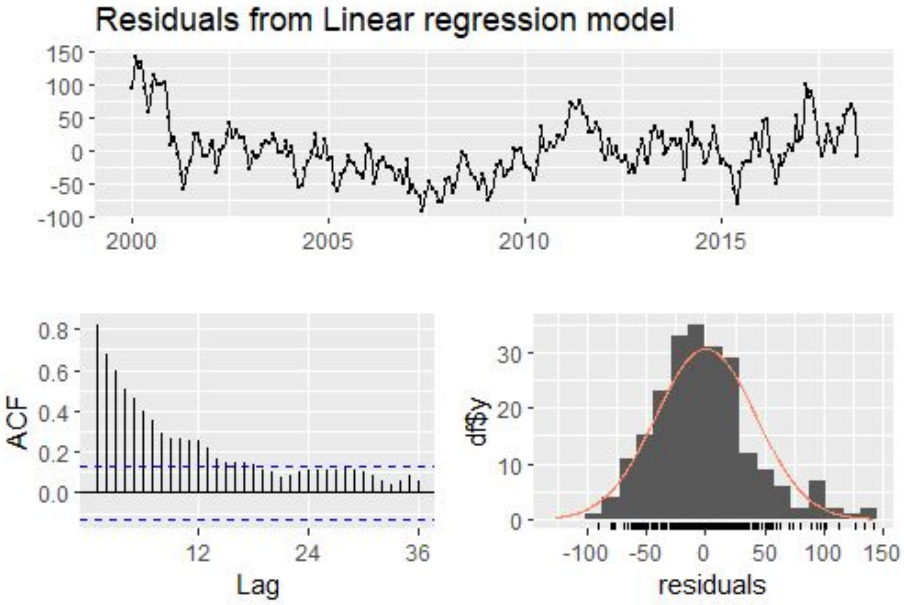
1. Linear Regression (trend + seasonality)
2. Holt Winters' exponential smoothing.
3. Linear regression model (y ~ t + (t^2) + season)
4. ARIMA
5. GAM (y ~ s(t) + s)
6. GAM(y ~ lo(t) + s)

Time-series data was splitted into two periods. Forecast models will be developed using only the train set our models. After we define the models, we try it out on validation set and see how it performs.



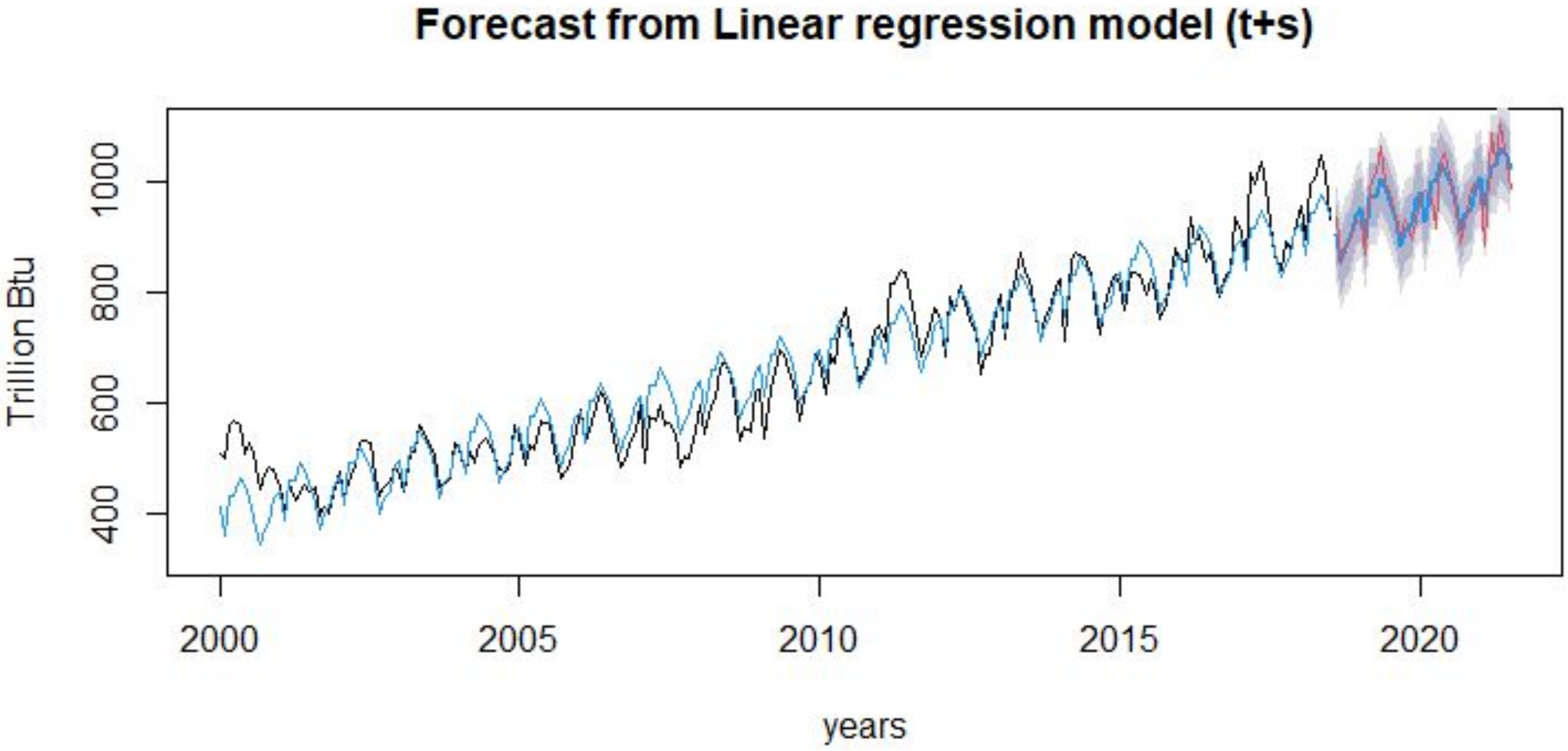
Linear regression model (y ~ t + s)

	train.ts		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	409.03	387.10 – 430.97	**<0.001**
trend	2.37	2.28 – 2.46	**<0.001**
season [2]	-56.20	-84.00 – -28.39	**<0.001**
season [3]	16.54	-11.26 – 44.34	0.242
season [4]	14.05	-13.76 – 41.85	0.320
season [5]	43.35	15.54 – 71.15	**0.002**
season [6]	27.18	-0.62 – 54.99	0.055
season [7]	2.80	-25.01 – 30.61	0.843
season [8]	-35.48	-63.67 – -7.30	**0.014**
season [9]	-87.61	-115.80 – -59.42	**<0.001**
season [10]	-62.81	-91.00 – -34.62	**<0.001**
season [11]	-52.82	-81.01 – -24.63	**<0.001**
season [12]	-8.81	-37.00 – 19.38	0.538
Observations	223		
R^2 / R^2 adjusted	0.933 / 0.930		



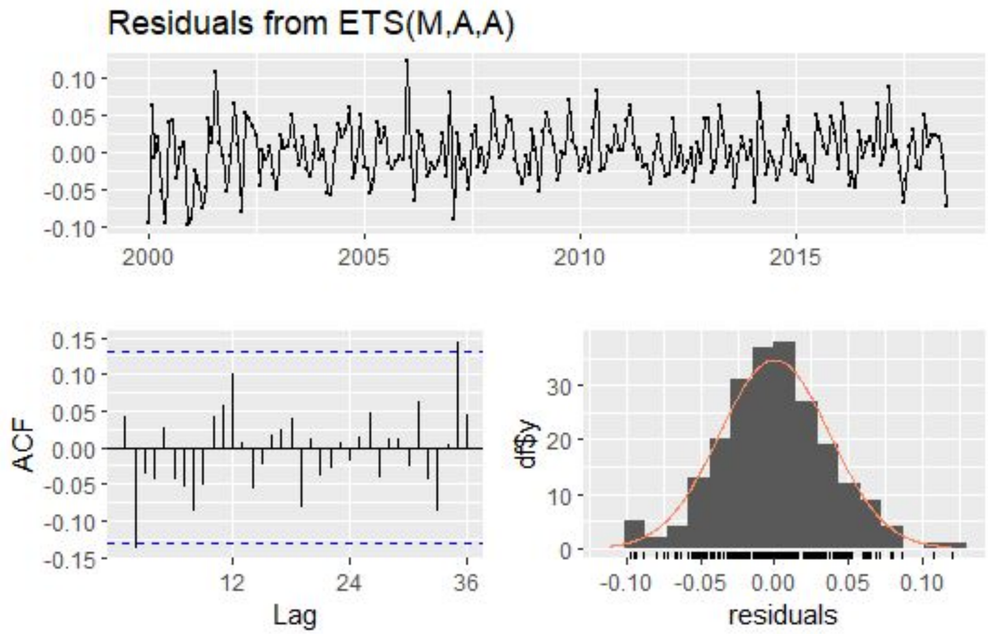
- We observed that residuals are highly correlated, so there is information left in the residuals.
- Residuals have zero mean, then the forecasts are not biased.

Linear regression model (y ~ t + s)

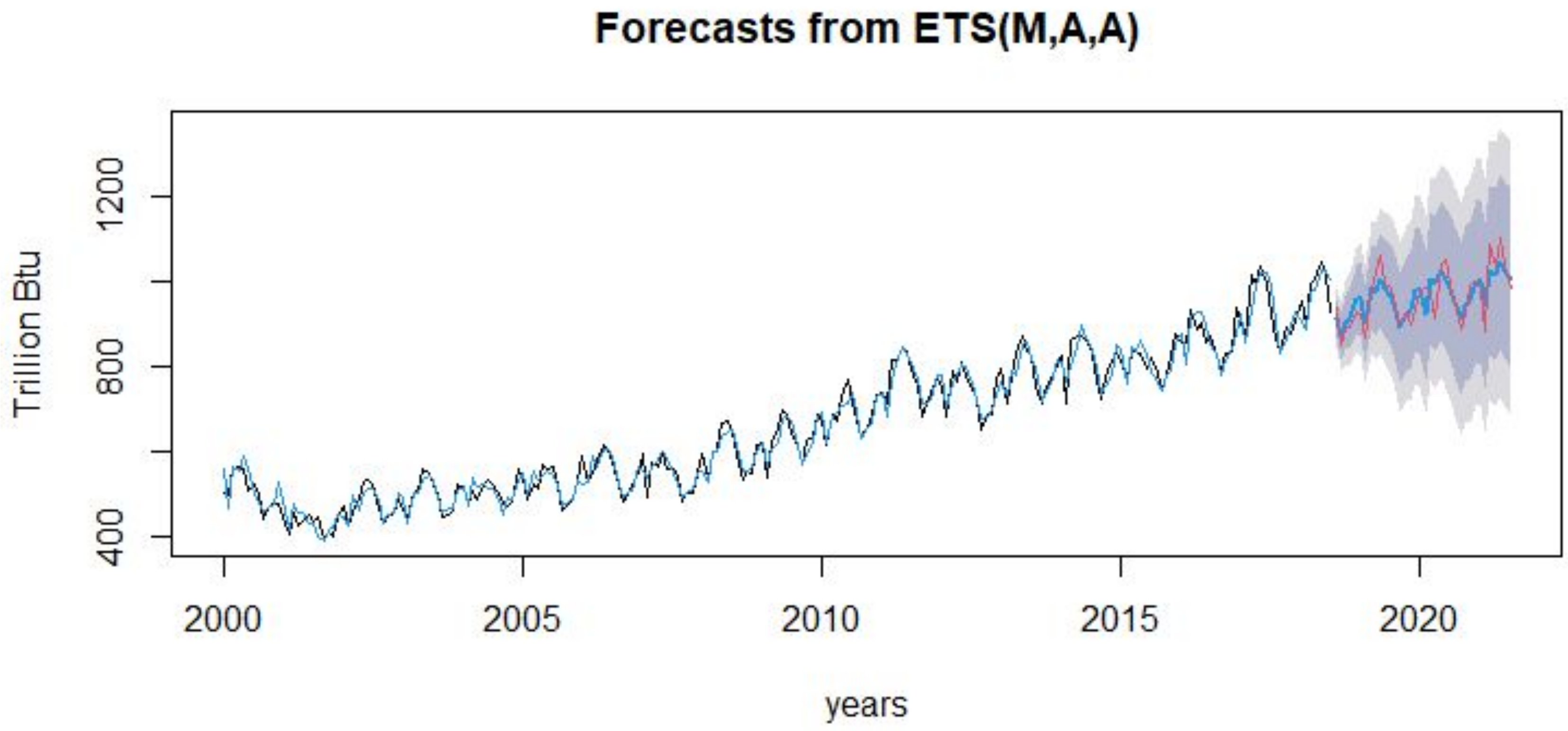


Holt Winters’ exponential smoothing model

```
ETS(M,A,A)

Call:
 ets(y = train.ts, model = "MAA")

  Smoothing parameters:
    alpha = 0.7163
    beta  = 1e-04
    gamma = 0.0724

  Initial states:
    l = 544.1413
    b = 1.7033
    s = 13.4787 -33.1345 -41.174 -57.9796 -16.3354 19.9941
           39.2971 48.3513 24.5769 30.2661 -39.7602 12.4195

  sigma:  0.0387

     AIC     AICc      BIC
2653.878 2656.863 2711.799

Training set error measures:
                     ME     RMSE      MAE        MPE     MAPE
Training set 0.07605769 24.23109 18.80124 -0.1369043 2.913783
                  MASE       ACF1
Training set 0.4491048 0.05156983
```

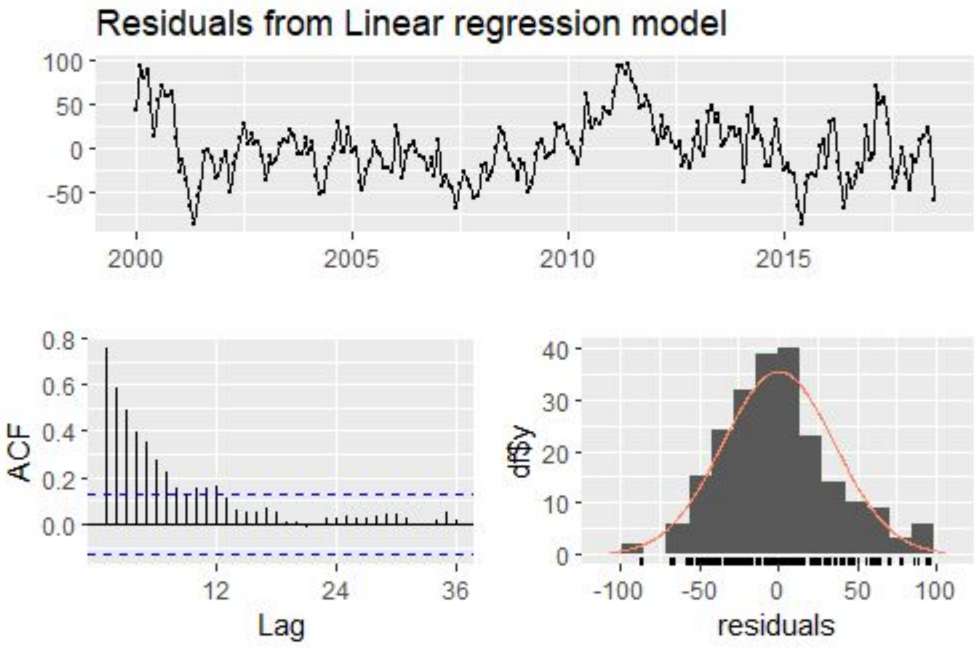


Holt Winters' exponential smoothing model

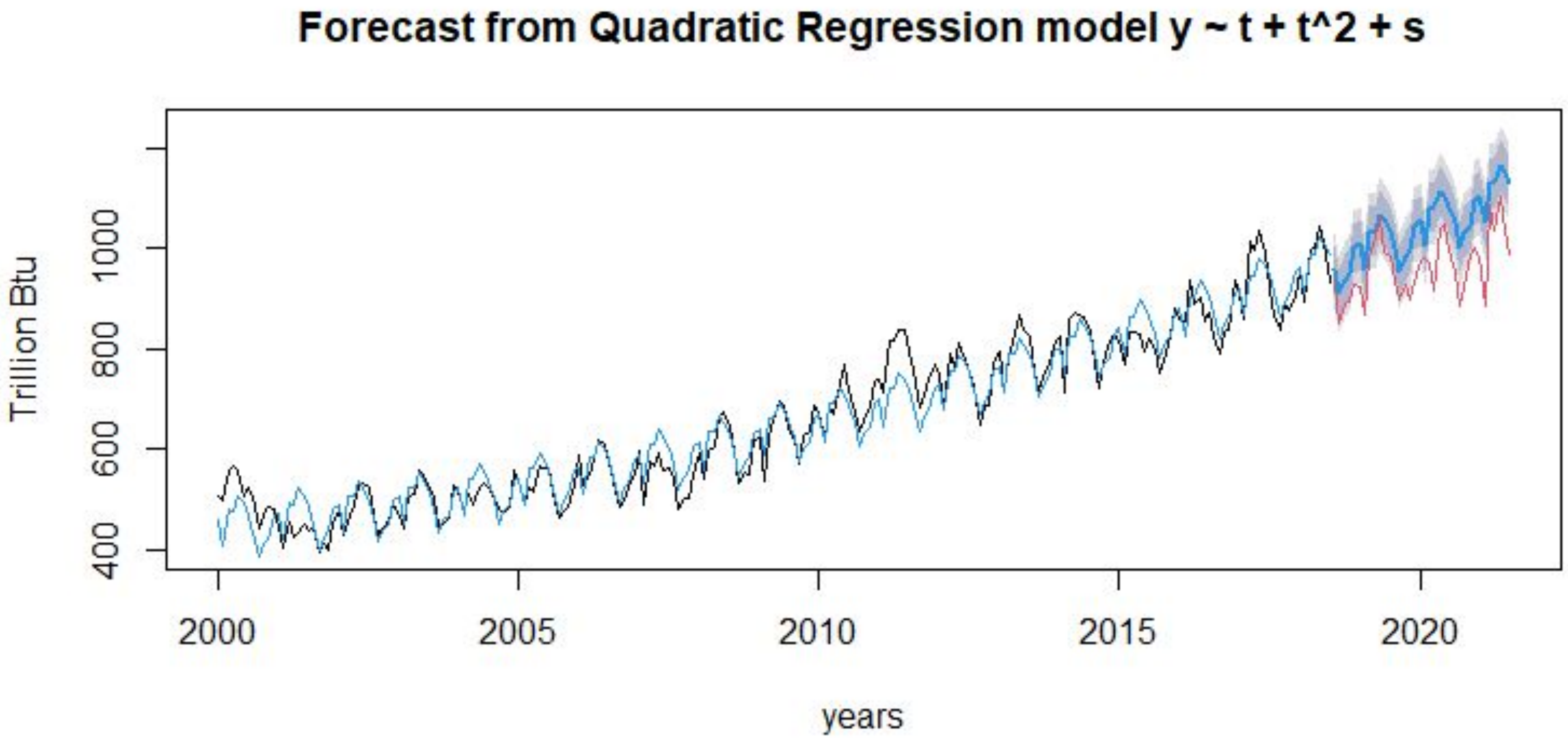


Linear regression model (y ~ t + (t^2) + season)

	train.ts		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	459.91	438.57 – 481.26	**<0.001**
trend	0.98	0.68 – 1.28	**<0.001**
trend^2	0.01	0.00 – 0.01	**<0.001**
season [2]	-56.17	-79.57 – -32.77	**<0.001**
season [3]	16.59	-6.81 – 39.99	0.164
season [4]	14.10	-9.30 – 37.50	0.236
season [5]	43.40	19.99 – 66.80	**<0.001**
season [6]	27.21	3.81 – 50.62	**0.023**
season [7]	2.80	-20.61 – 26.20	0.814
season [8]	-32.70	-56.43 – -8.97	**0.007**
season [9]	-84.81	-108.54 – -61.08	**<0.001**
season [10]	-60.00	-83.73 – -36.27	**<0.001**
season [11]	-50.02	-73.75 – -26.29	**<0.001**
season [12]	-6.03	-29.77 – 17.70	0.617
Observations	223		
R^2 / R^2 adjusted	0.953 / 0.950		



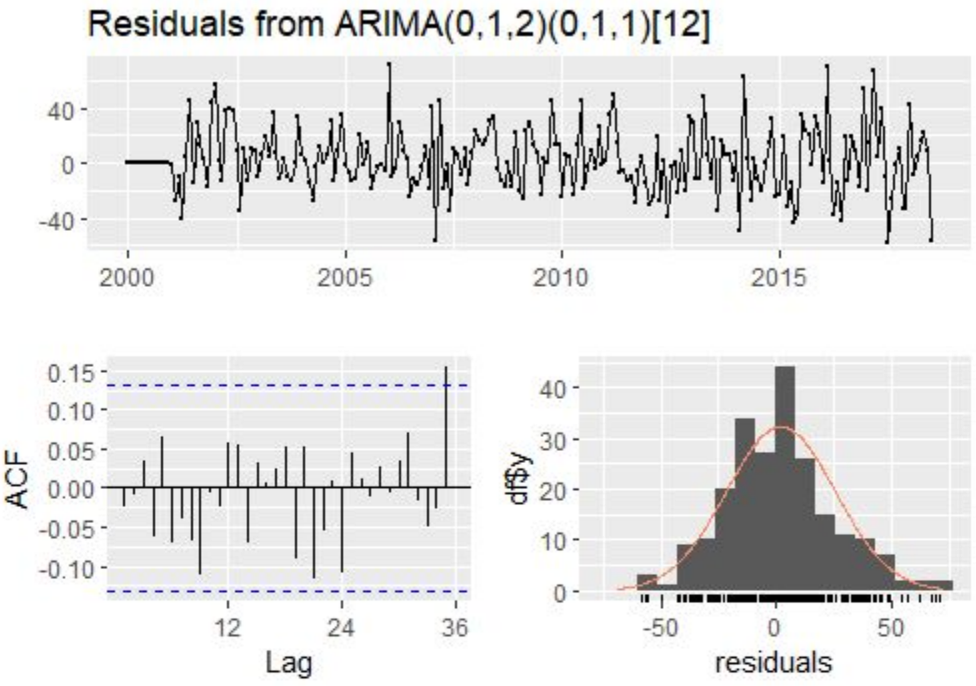
Linear regression model (y ~ t + (t^2) + season)



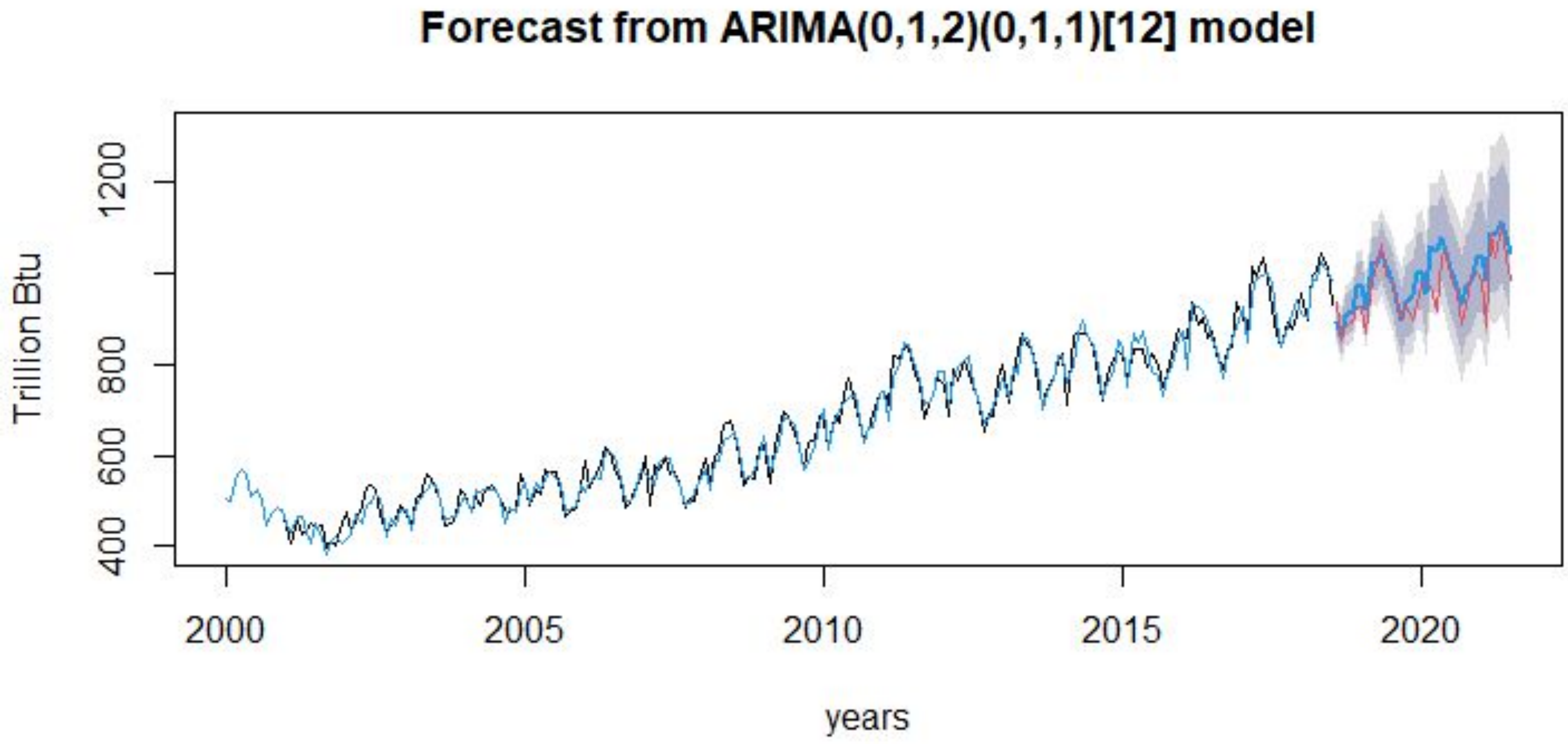
ARIMA(0,1,2)(0,1,1)[12] model

	Dependent variable		
Predictors	*Estimates*	*CI*	*p*
ma1	-0.22	-0.35 – -0.08	**0.002**
ma2	-0.24	-0.40 – -0.09	**0.002**
sma1	-0.74	-0.85 – -0.62	**<0.001**
Observations	210		
R^2	0.979		

```
sigma^2 estimated as 603.9:  log likelihood=-973.64
AIC=1955.28   AICc=1955.48   BIC=1968.67

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE
Training set 2.016994 23.67728 17.89548 0.2958543 2.748728 0.427469
                    ACF1
Training set -0.02344139
```

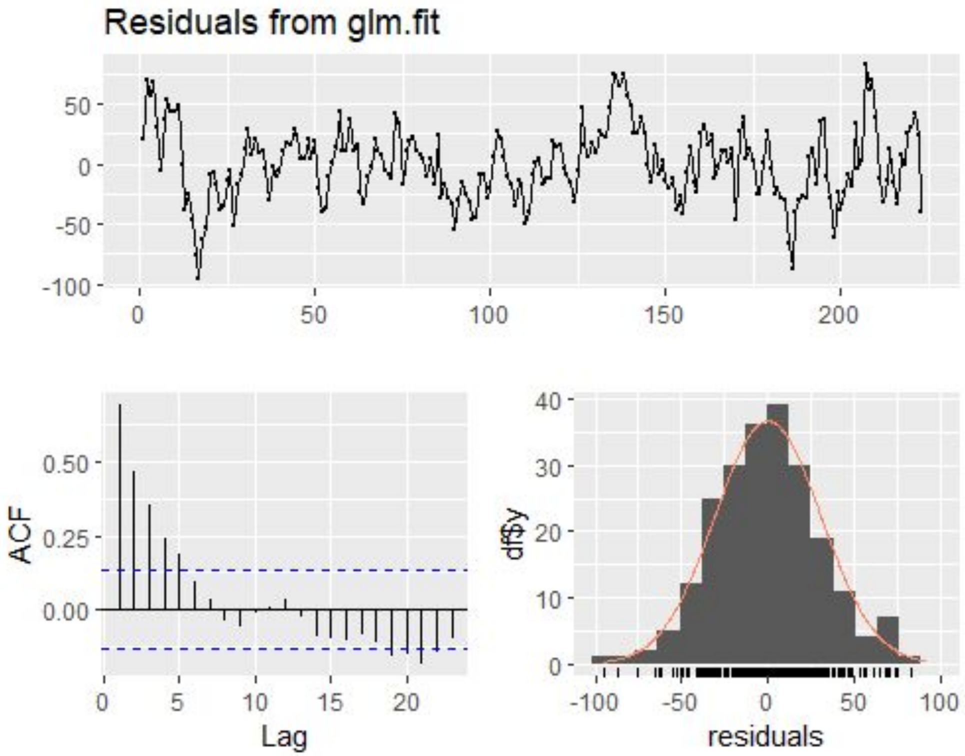


ARIMA(0,1,2)(0,1,1)[12] model

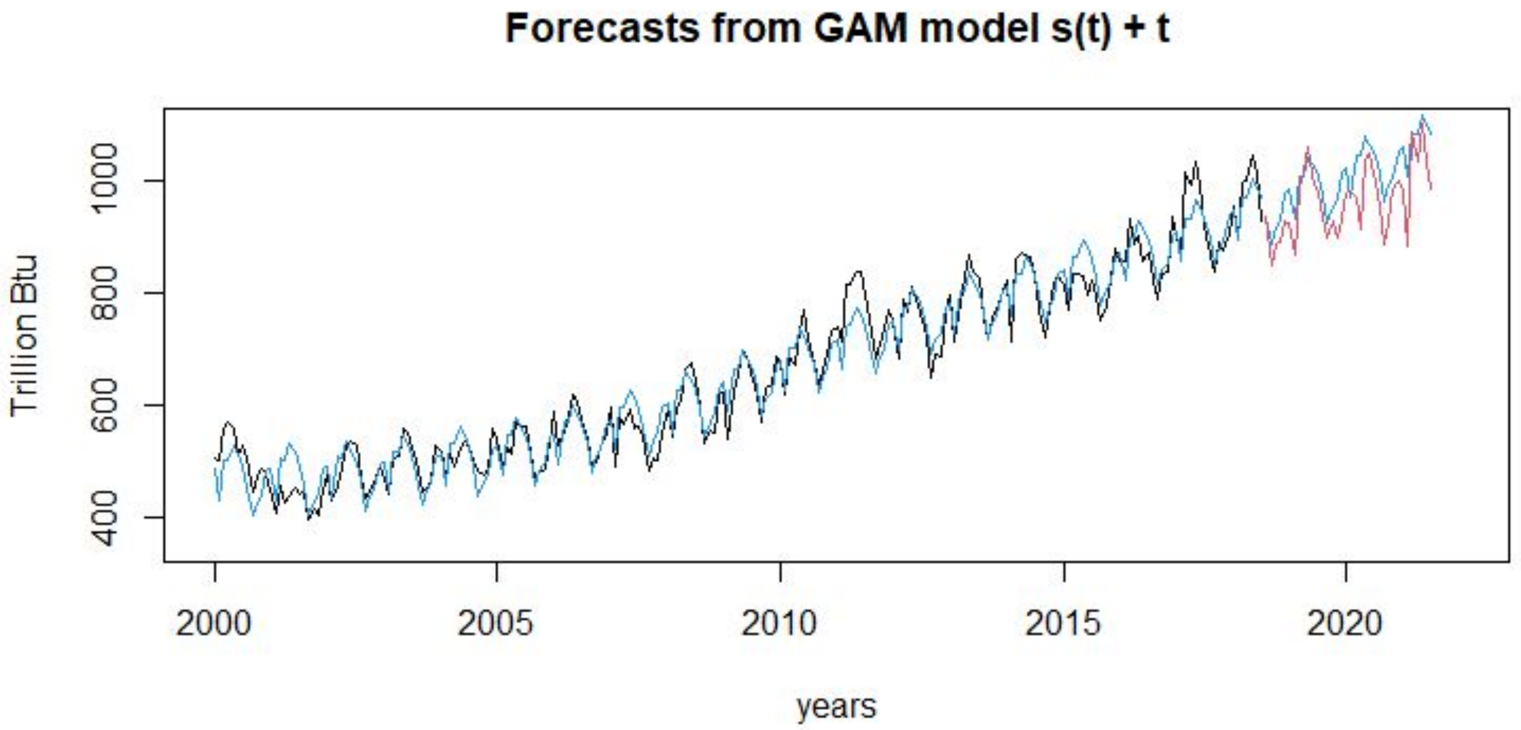


Generalized Additive Model (GAM) (y ~ s(t) + s)

```
Call: gam(formula = train.ts ~ s(t) + seas)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-95.394 -19.563  -1.086  20.223  82.535

(Dispersion Parameter for gaussian family taken to be 1020.872)

    Null Deviance: 5952257 on 222 degrees of freedom
Residual Deviance: 211320.6 on 207 degrees of freedom
AIC: 2195.28

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
           Df  Sum Sq Mean Sq  F value    Pr(>F)
s(t)        1 5209625 5209625 5103.112 < 2.2e-16 ***
seas       11  330904   30082   29.467 < 2.2e-16 ***
Residuals 207  211321    1021
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
            Npar Df Npar F     Pr(F)
(Intercept)
s(t)              3 60.583 < 2.2e-16 ***
seas
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

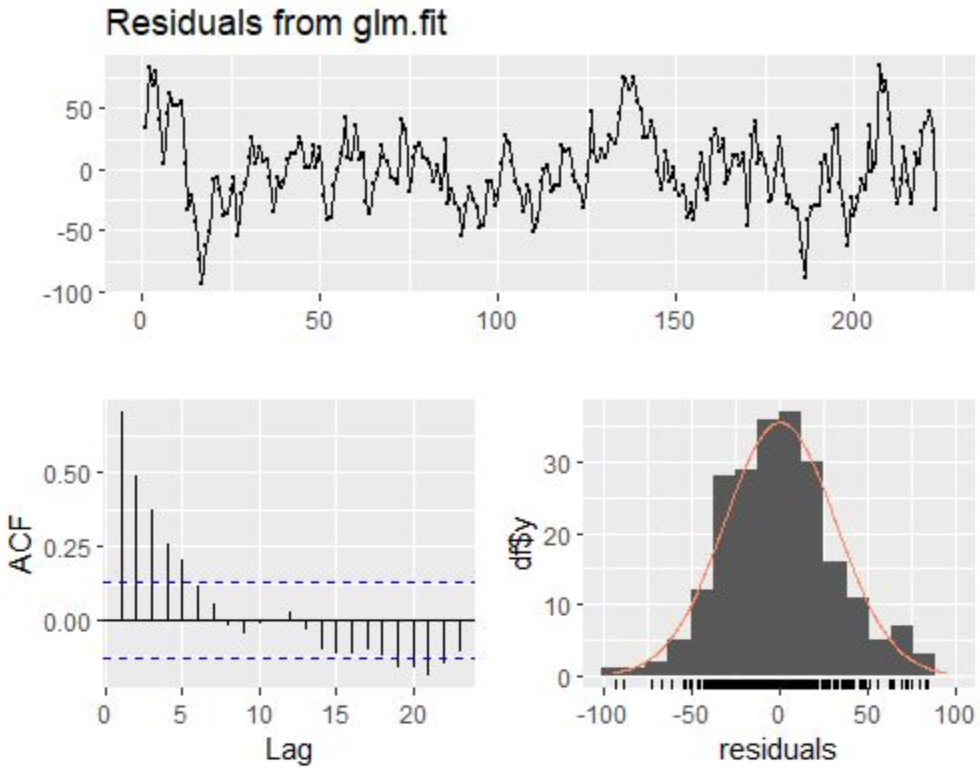


Generalized Additive Model (GAM) (y ~ s(t) + s)

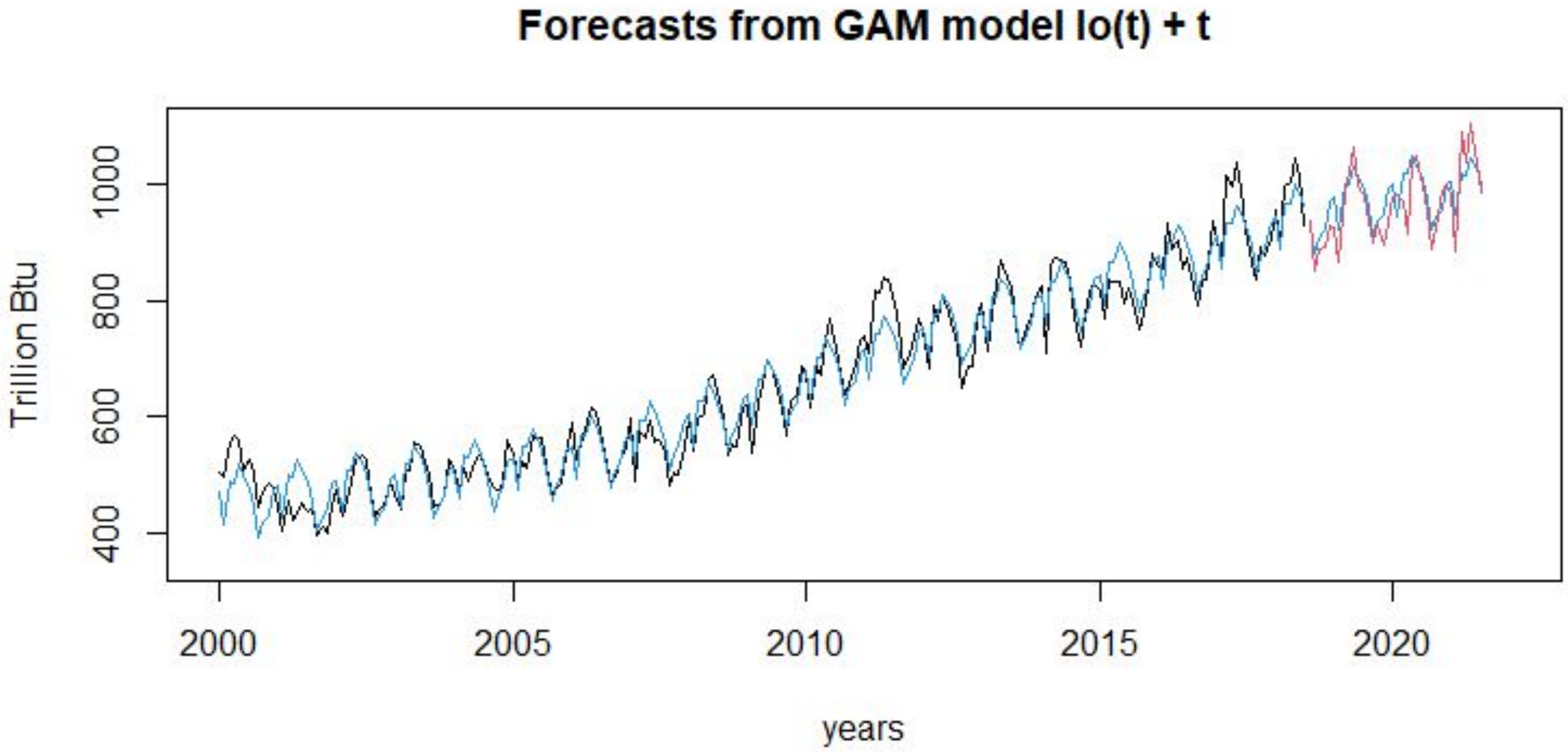


Generalized Additive Model (GAM) (y ~ lo(t) + s)

```
Call: gam(formula = train.ts ~ lo(t) + seas)
Deviance Residuals:
     Min       1Q   Median       3Q      Max
-93.4266 -21.1622  -0.8443  18.1895  84.3126

(Dispersion Parameter for gaussian family taken to be 1075.419)

    Null Deviance: 5952257 on 222 degrees of freedom
Residual Deviance: 223343.6 on 207.6804 degrees of freedom
AIC: 2206.258

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
              Df  Sum Sq Mean Sq  F value    Pr(>F)
lo(t)       1.00 5209625 5209625 4844.272 < 2.2e-16 ***
seas       11.00  334004   30364   28.235 < 2.2e-16 ***
Residuals 207.68  223344    1075
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
            Npar Df Npar F     Pr(F)
(Intercept)
lo(t)           2.3 69.559 < 2.2e-16 ***
seas
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Generalized Additive Model (GAM) (y ~ lo(t) + s)



Performance of forecast models

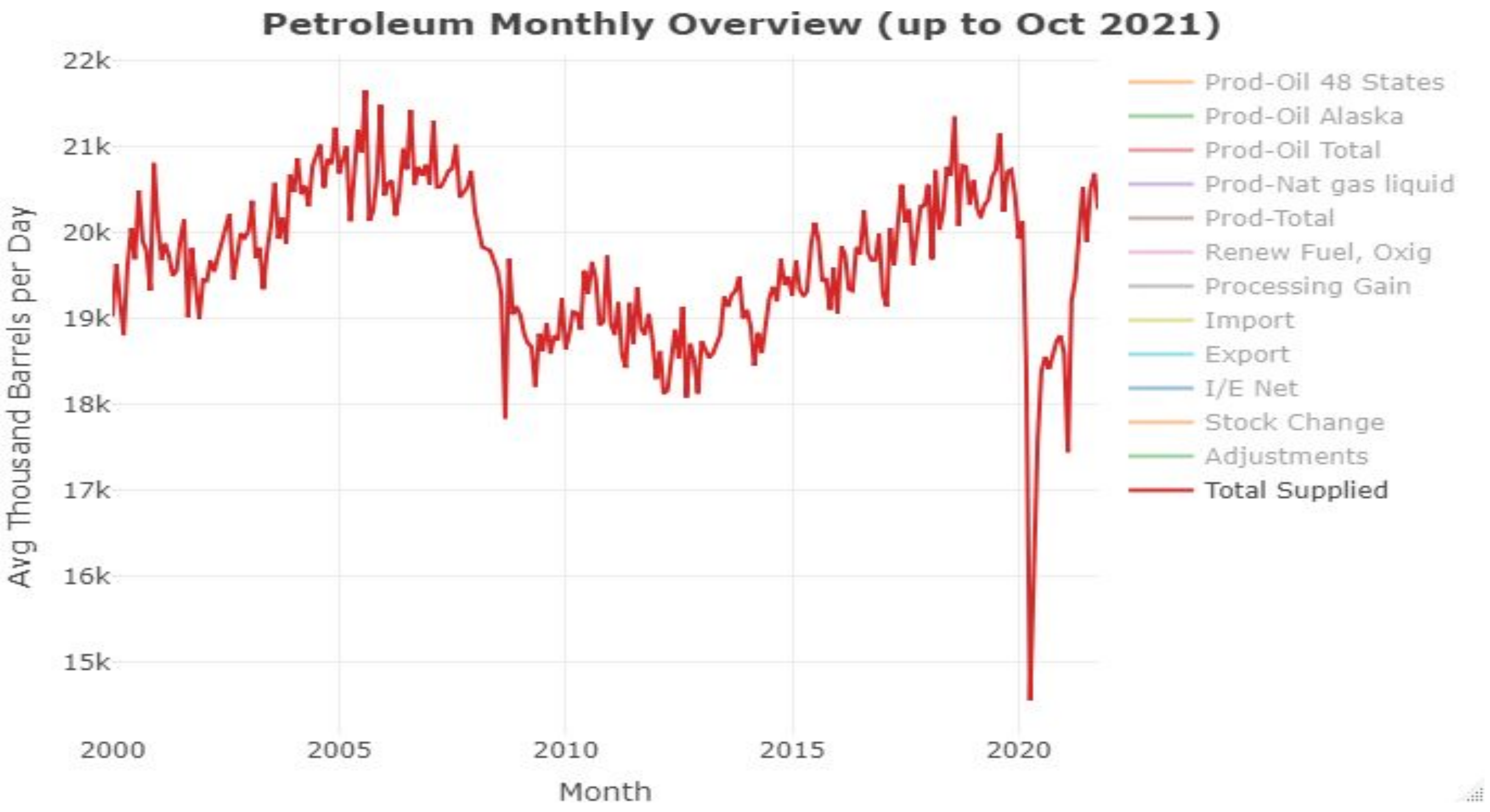
Model	Predictors	Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	AIC
LMTS	t + s	Training set	-1.51E-15	42.186	32.119	-0.305	5.253	0.767	2329.816
		Test set	0.734	32.820	25.114	-0.048	2.595	0.600	NA
LMQTS	MAA	Training set	1.02E-15	35.419	27.706	-0.316	4.396	0.662	2253.843
		Test set	-78.171	87.113	78.171	-8.208	8.208	1.867	NA
Holt Winters	t + t^2 + s	Training set	0.076	24.231	18.801	-0.137	2.914	0.449	2653.878
		Test set	-2.399	33.816	26.888	-0.395	2.819	0.642	NA
ARIMA	ARIMA(0,1,2)(0,1,1)[12]	Training set	2.017	23.677	17.895	0.296	2.749	0.427	1955.284
		Test set	-29.647	46.049	34.966	-3.152	3.699	0.835	NA
GAM	s(t) + s	Training set	3.14E-14	30.784	24.381	-0.239	3.895	0.170	2195.28
		Test set	-45.096	56.874	47.747	-4.795	5.051	NA	NA
GAM	lo(t) + s	Training set	-2.92E-13	31.647	24.895	-0.251	3.977	0.173	2206.258
		Test set	-13.387	37.604	30.524	-1.567	3.203	NA	NA

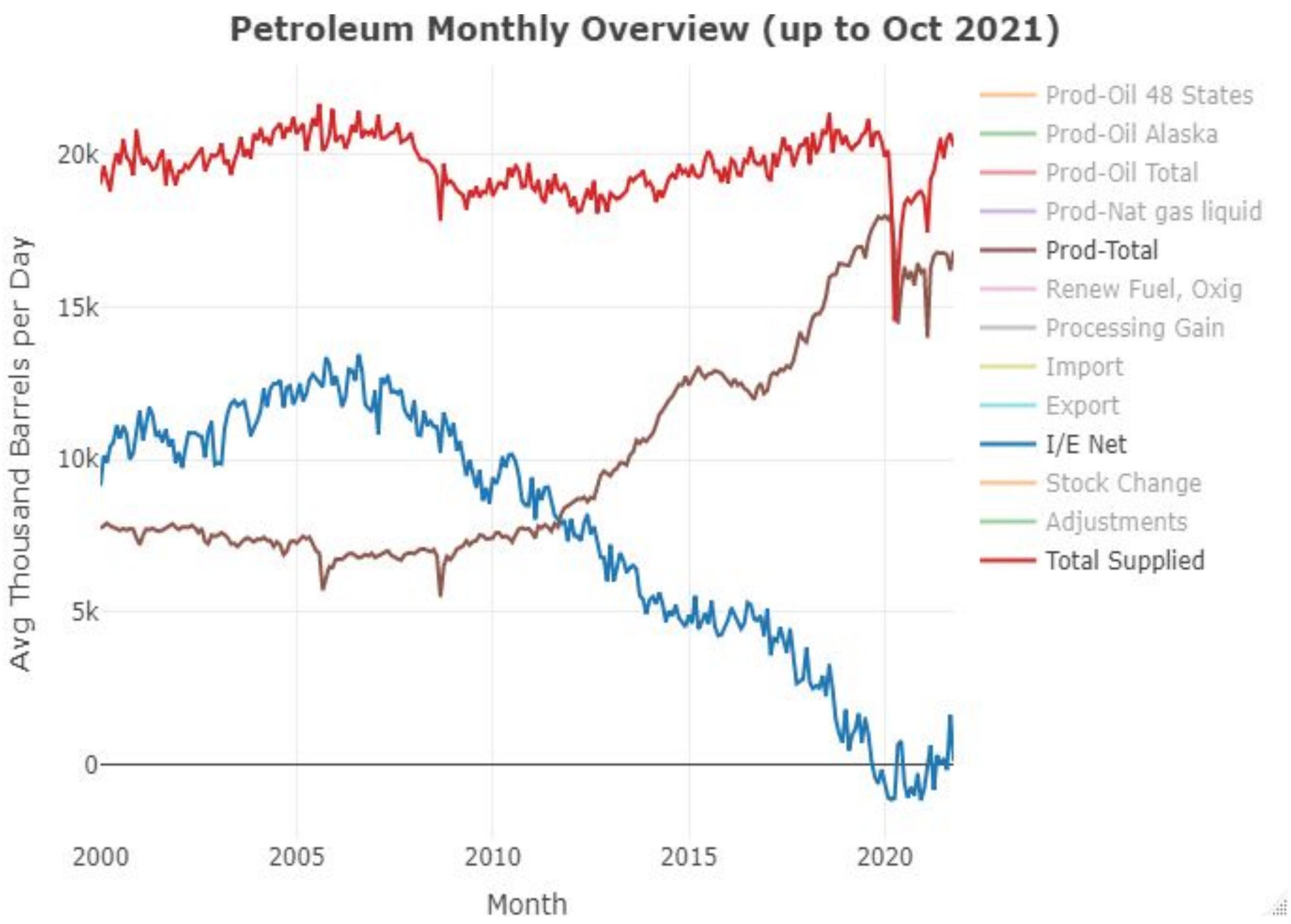
- Considering the RMSE and MAPE values, all of the results points to Holt Winter’s and ARIMA models. Their performance on training set are similar, but ARIMA model perform at last 1% worst than Holt Winter’s on validation set.
- However, considering the AIC accuracy measure we noted that ARIMA model has a lower value than the HW.
- WE adopt the ARIMA model as the best model to forecast the monthly renewable energy consumption.

Renewables Energy Consumption models - Conclusions

➢ Both the **Holt-Winter** and the **ARIMA** model are **performing well**, we decide to choose the **ARIMA** model due to the fact that has the lower value of **AIC measure**.

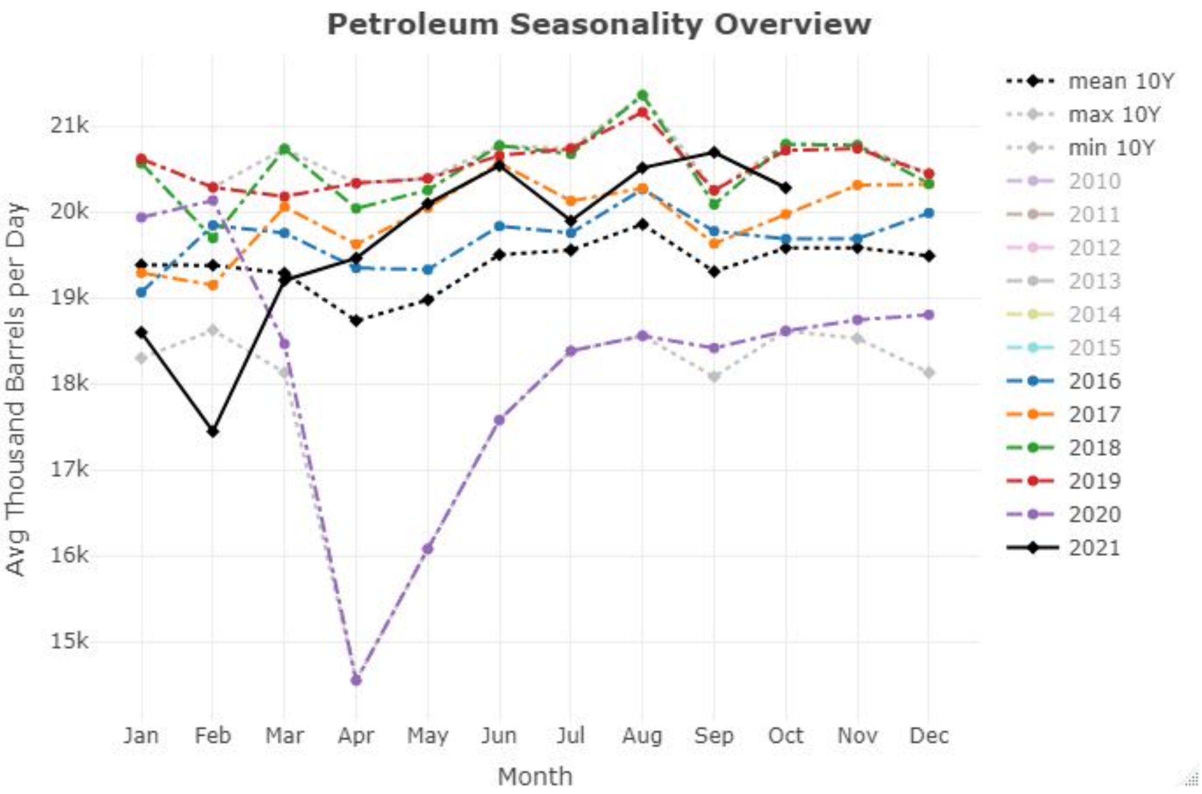
US Monthly Petroleum Consumption





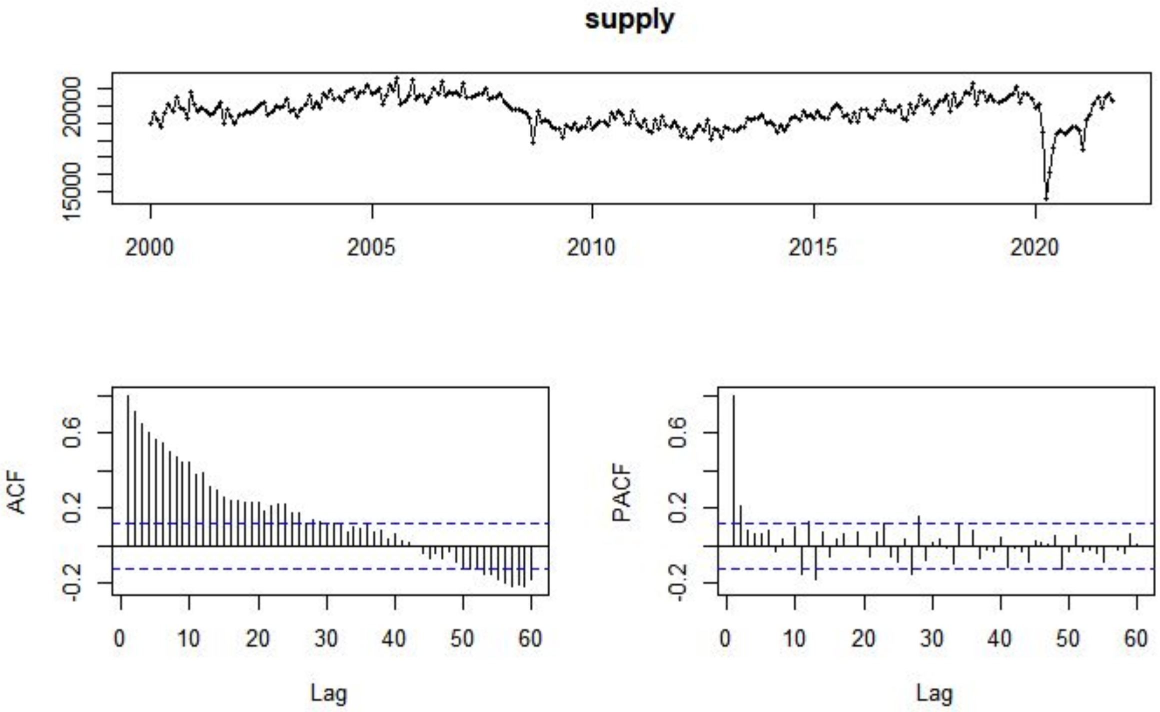
- **Consumption ≈** Production + I/E Net
- Slightly increasing **trend** except for the year 2008 and the shock in 2020
- Note: the **order** is of **Million barrels** per day!! (1 barrel ~ 150 l)

Seasonal overview



- Consumption is **stable** throughout the **year** period.
- The **overall value** was **increasing** in the last years.
- **2020** has seen a **big drop** due to the pandemic.
- The **shock** seems to have already been **absorbed**.

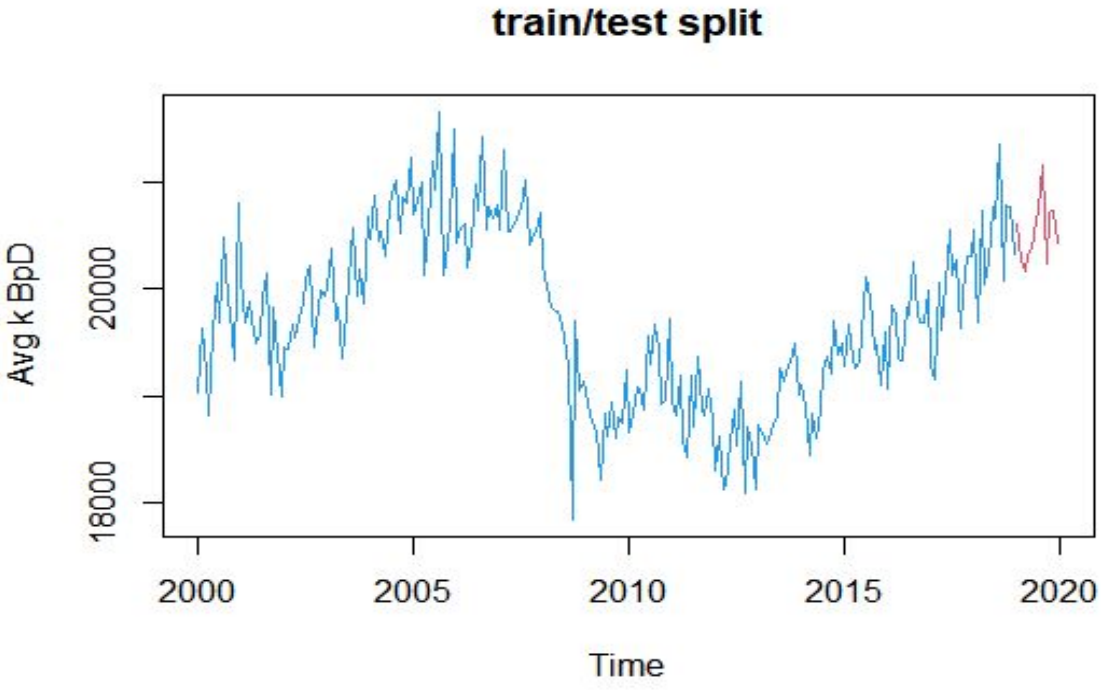
Modelling - warm up



- ➢ **Past** data can be exploited.
- ➢ There is a **trend** component.
- ➢ The **seasonal** component, seems non significant.
- ➢ The first **lag** data is the most relevant.

Modelling - tested models

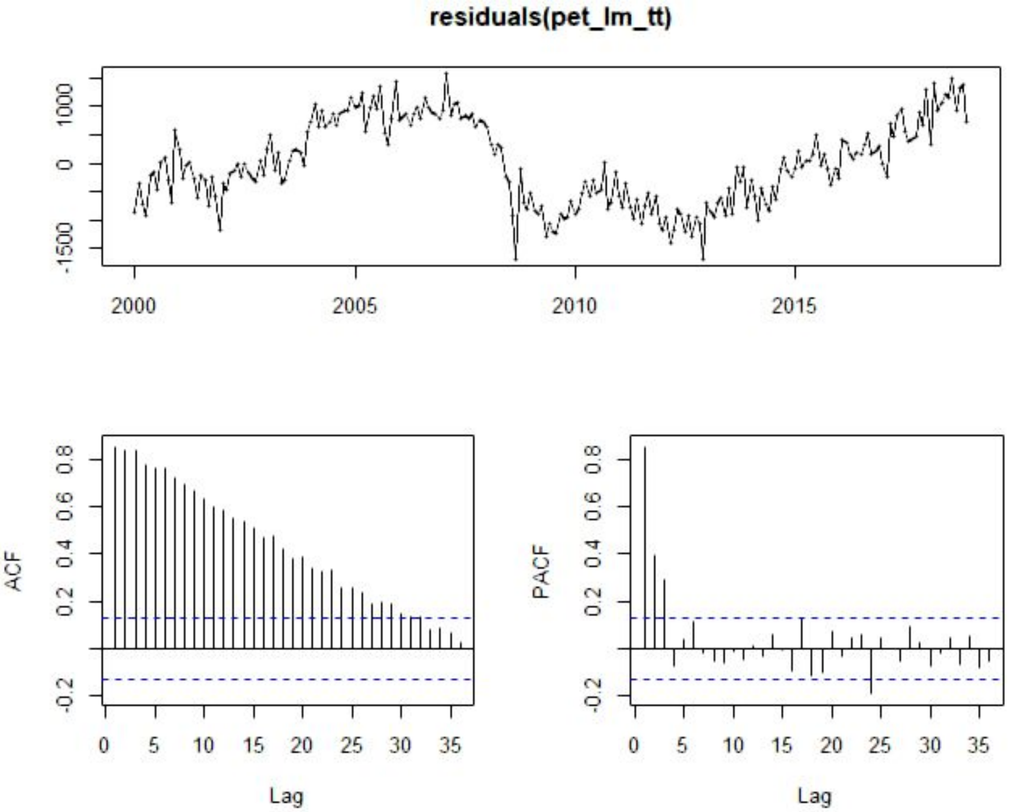
1. Linear model
2. GAM
3. Holt-Winter
4. ARIMA



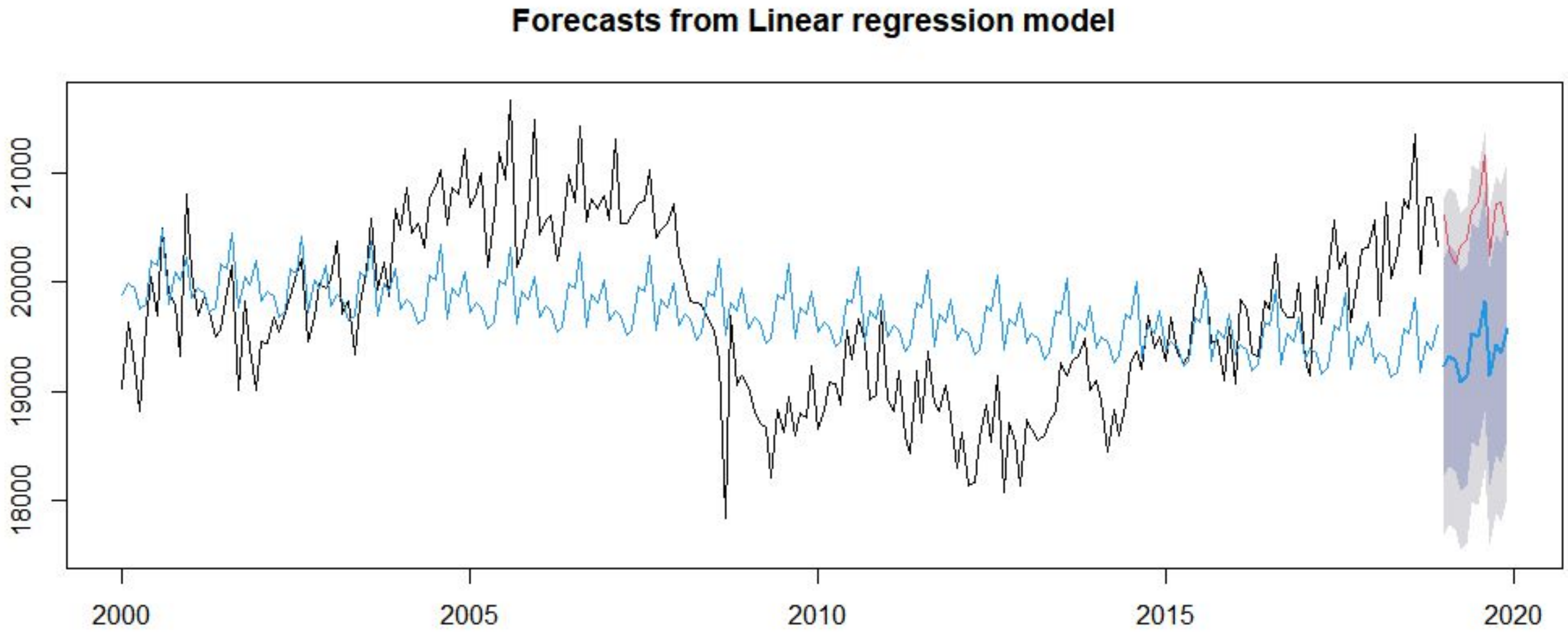
Modelling - Linear Model

```
tslm(formula = pet_train_set ~ trend + season)

Residuals:
     Min       1Q   Median       3Q      Max
-1682.05  -606.30   -51.81   663.07  1567.43

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 19893.1622   192.9390 103.106  < 2e-16 ***
trend          -2.9013     0.7643  -3.796 0.000191 ***
season2        97.8929   246.1095   0.398 0.691201
season3        56.3781   246.1131   0.229 0.819029
season4      -129.2881   246.1190  -0.525 0.599911
season5       -73.6402   246.1273  -0.299 0.765080
season6       321.9119   246.1380   1.308 0.192320
season7       282.1644   246.1510   1.146 0.252943
season8       618.5090   246.1665   2.513 0.012720 *
season9       -68.0939   246.1843  -0.277 0.782355
season10      224.4044   246.2044   0.911 0.363076
season11      153.9395   246.2270   0.625 0.532507
season12      371.7228   246.2519   1.510 0.132633
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 758.6 on 215 degrees of freedom
Multiple R-squared:  0.1267,    Adjusted R-squared:  0.07795
F-statistic: 2.599 on 12 and 215 DF,  p-value: 0.003002
```

	RMSE	MAE	MAPE
Training set	736.6139	624.4088	3.171246
Test set	1185.4490	1172.7531	5.703449



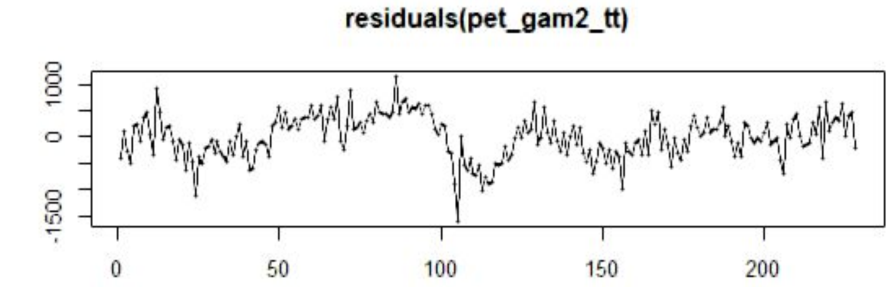
Linear Model - test forecast

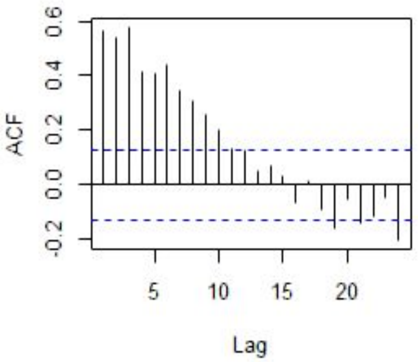


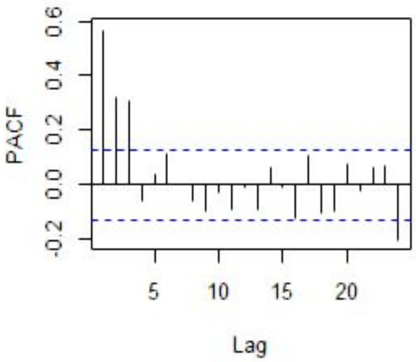
Modelling - GAM

```
Call: gam(formula = pet_train_set ~ lo(t) + seas, data = df_gam_train)
AIC: 3424.097

Anova for Parametric Effects
              Df   Sum Sq Mean Sq F value    Pr(>F)    
lo(t)       1.00  7872432 7872432 43.4773 3.334e-10 ***
seas       11.00  9185781  835071  4.6119 2.704e-06 ***
Residuals 212.72 38516466  181070                      
---
Anova for Nonparametric Effects
            Npar Df Npar F     Pr(F)    
(Intercept)                             
lo(t)           2.3 206.02 < 2.2e-16 ***
seas                                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

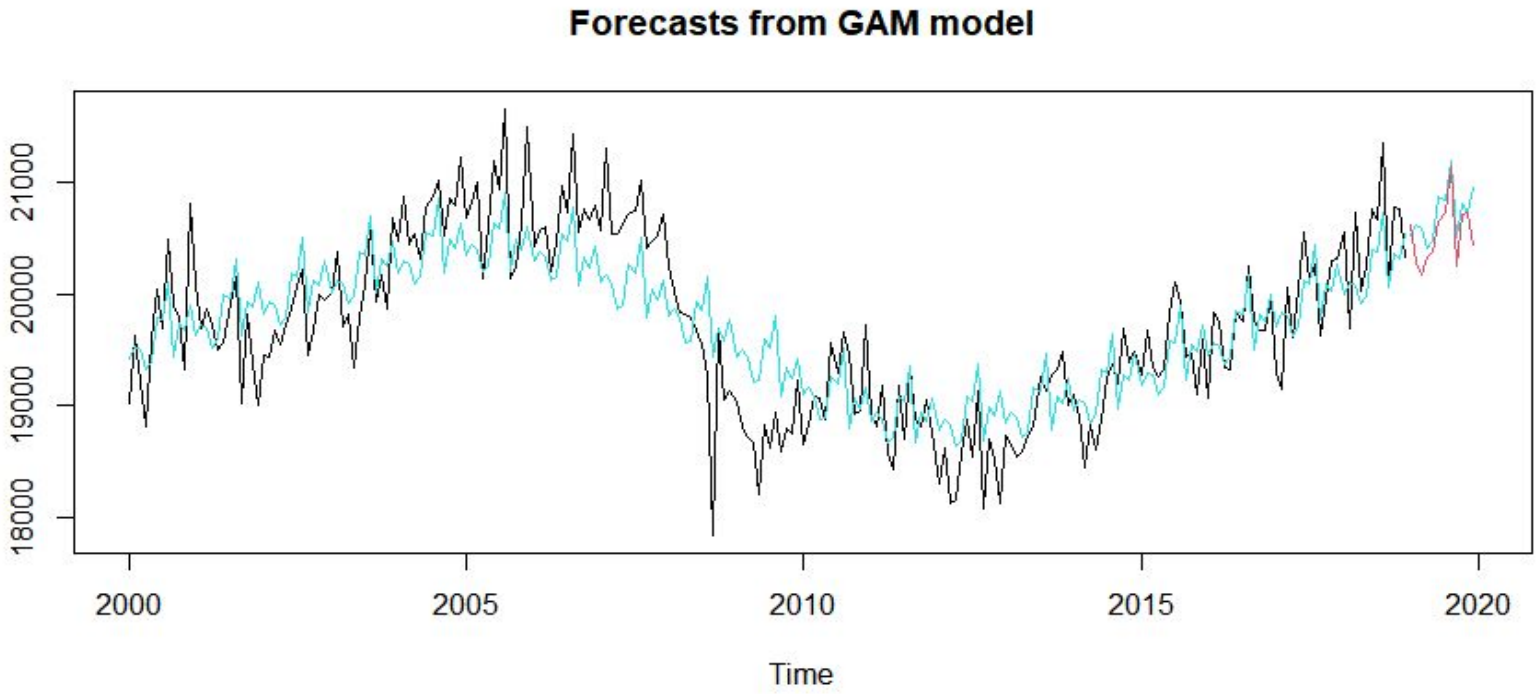
```
                 RMSE      MAE     MAPE
Training set 411.0132 330.0797 1.677766
Test set     246.8964 191.7166 0.939528
```







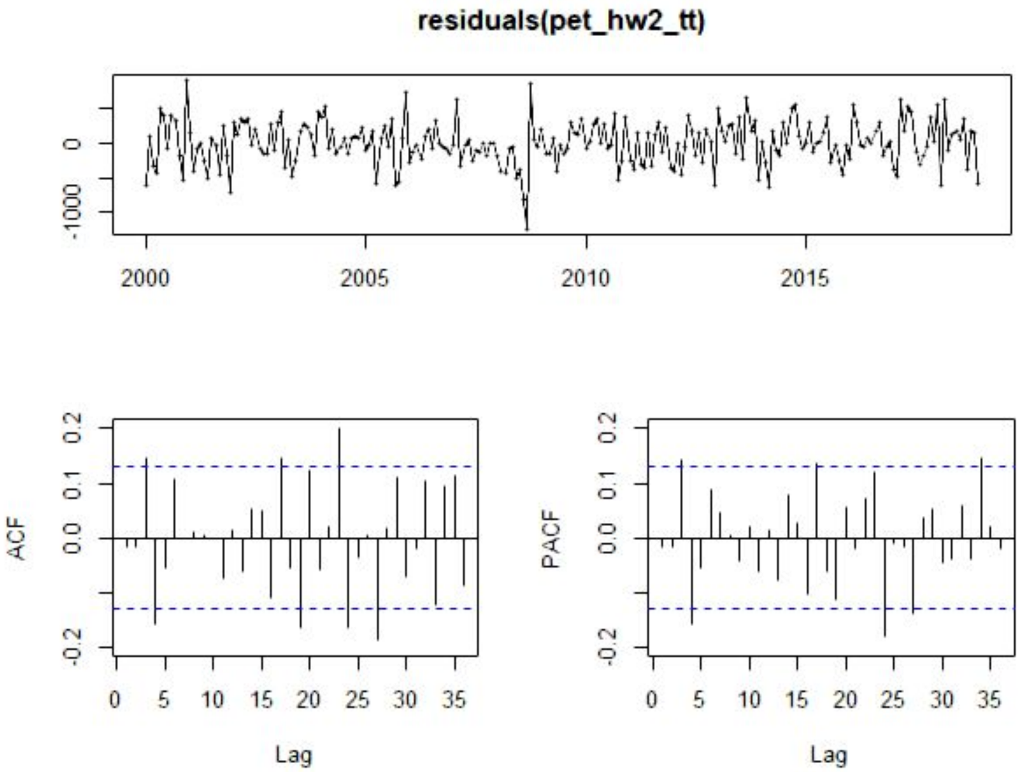
GAM - test forecast



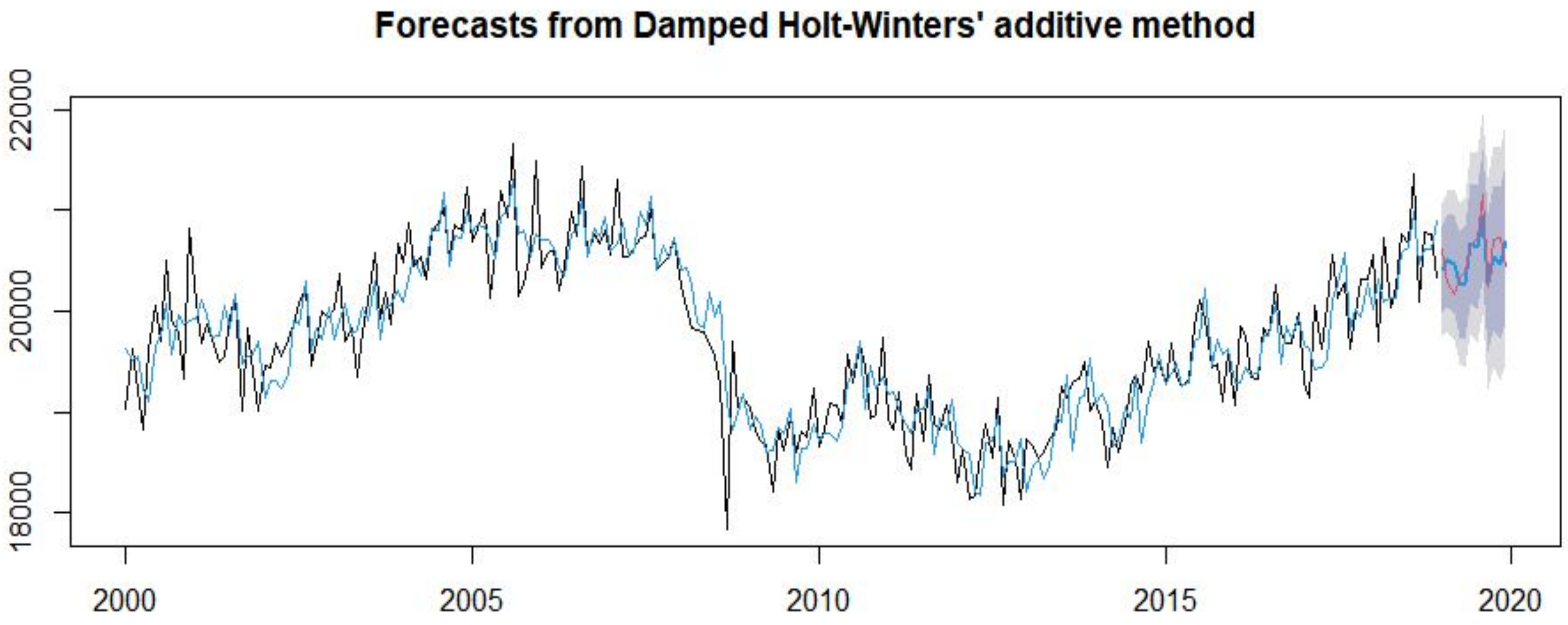
Modelling - Holt-Winter

```
Forecast method: Damped Holt-Winters' additive method

Model Information:
Damped Holt-Winters' additive method

Call:
 hw(y = pet_train_set, damped = TRUE)

  Smoothing parameters:
    alpha = 0.3228
    beta  = 0.0466
    gamma = 2e-04
    phi   = 0.8247

 AIC  3907.865
```

	RMSE	MAE	MAPE
Training set	322.5503	250.6503	1.2773898
Test set	182.5849	159.2170	0.7748818



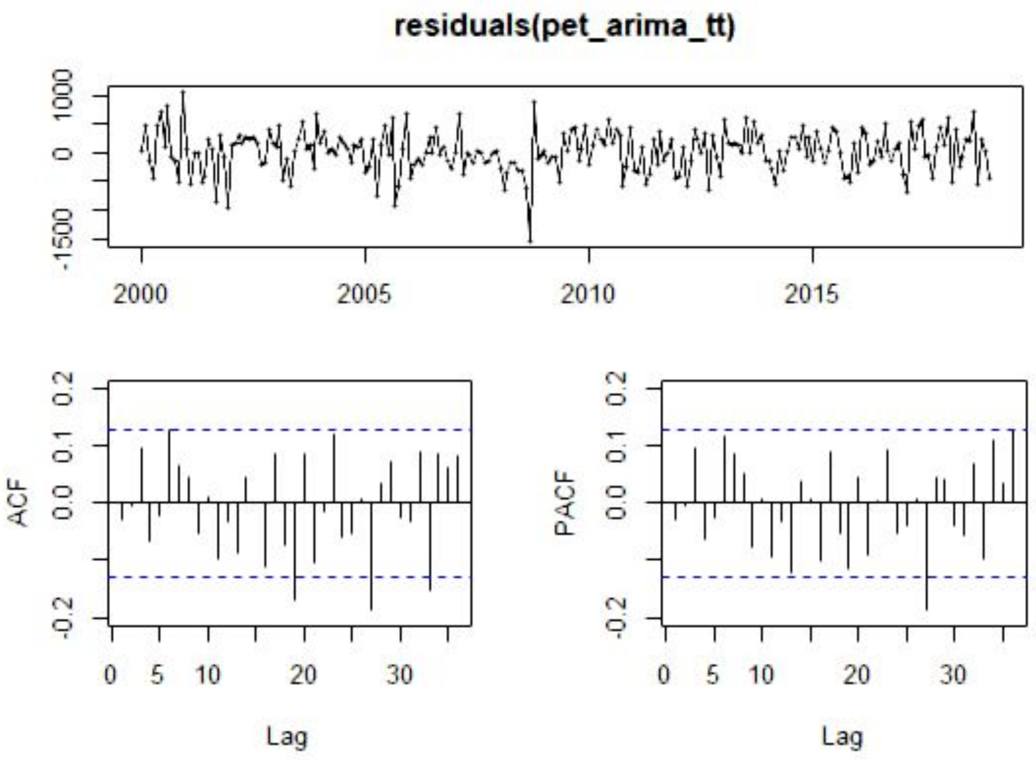
Holt-Winter - test forecast



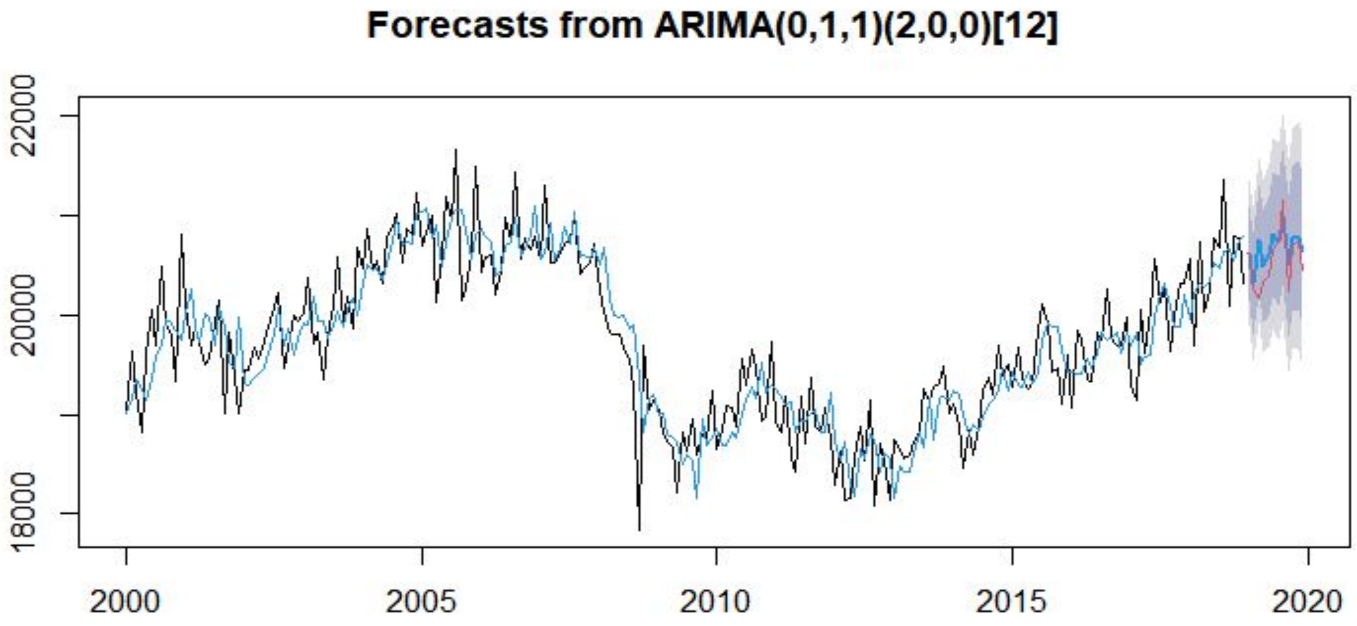
Modelling - ARIMA

```
Series: pet_train_set
ARIMA(0,1,1)(2,0,0)[12]

Coefficients:
          ma1    sar1    sar2
      -0.6166  0.3334  0.0966
s.e.   0.0494  0.0718  0.0744

AIC=3338.96
```

	RMSE	MAE	MAPE
Training set	368.8413	285.6619	1.4541835
Test set	218.6210	158.8158	0.7779088



ARIMA - test forecast



Test Results Table

Model	AIC	RMSE	MAE	MAPE
Linear	-	1185	1173	5.70
GAM	3424	247	1192	0.94
H-W	3908	**183**	**159**	**0.77**
ARIMA	3339	219	**159**	0.78

Modelling - Conclusion

- The **linear** model is **poor** from all points of view.

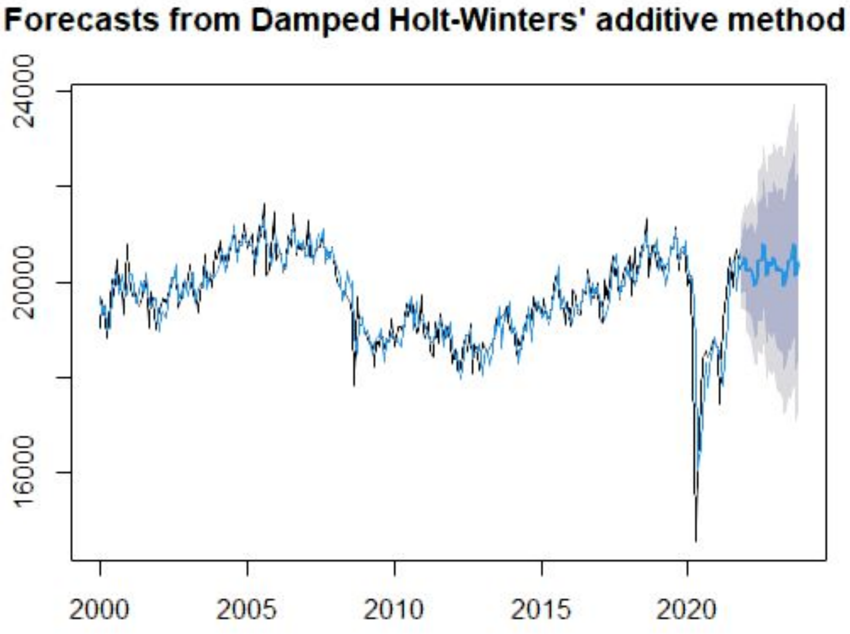
  Not surprising because forcing linearity is a too strong hp.

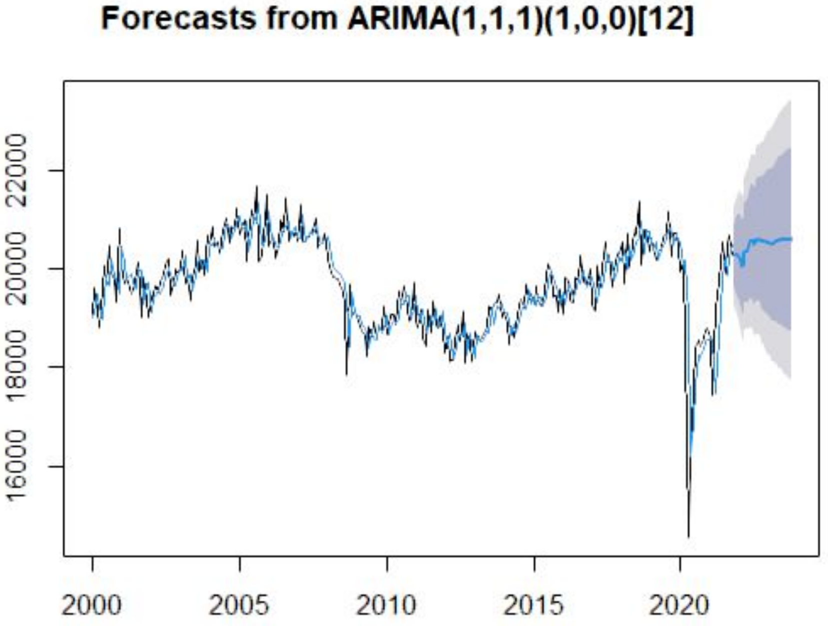
- The **GAM** model is **decent** but still with some error.

  Moreover the residual inspection suggest it's missing something.

- Both the **Holt-Winter** and the **ARIMA** model are **good**.

Holt-Winter and ARIMA - future forecast

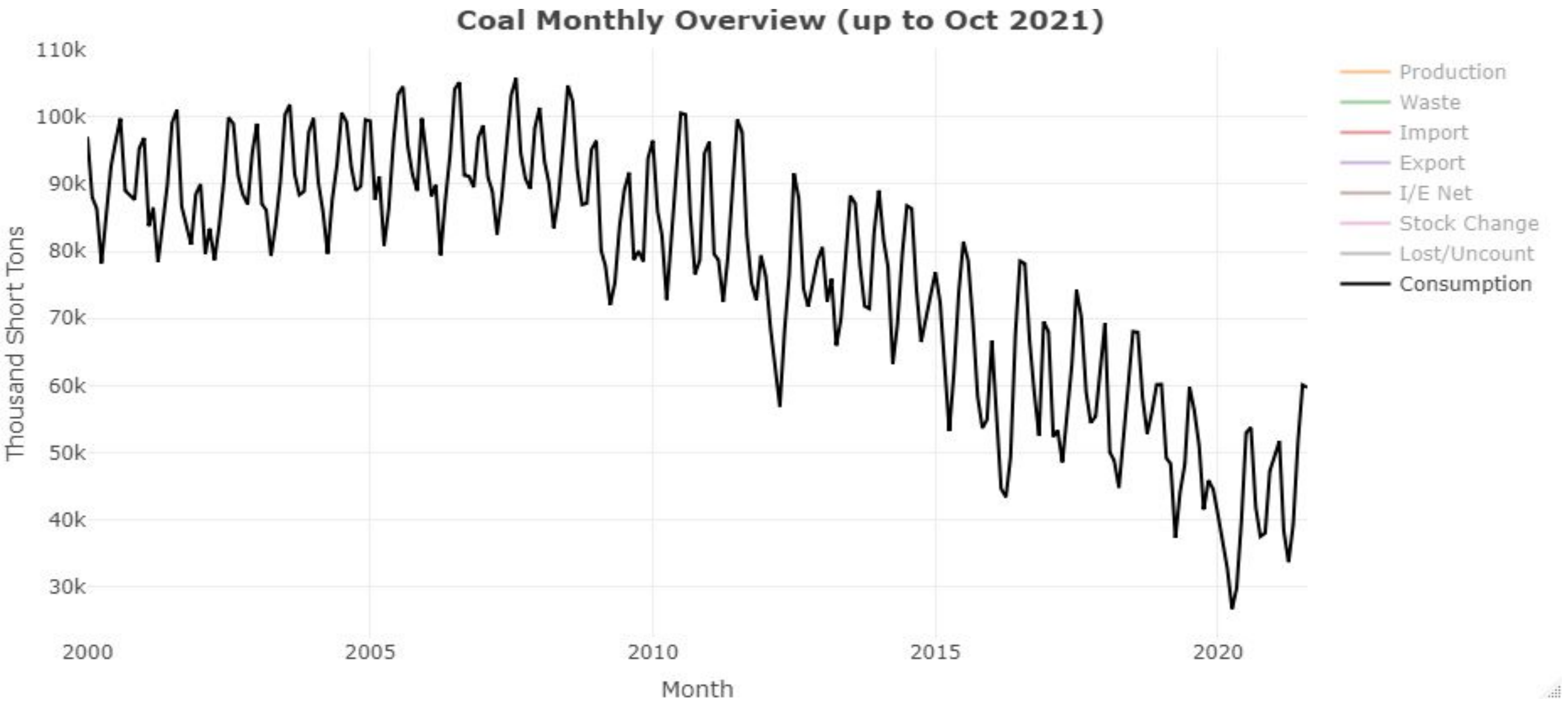


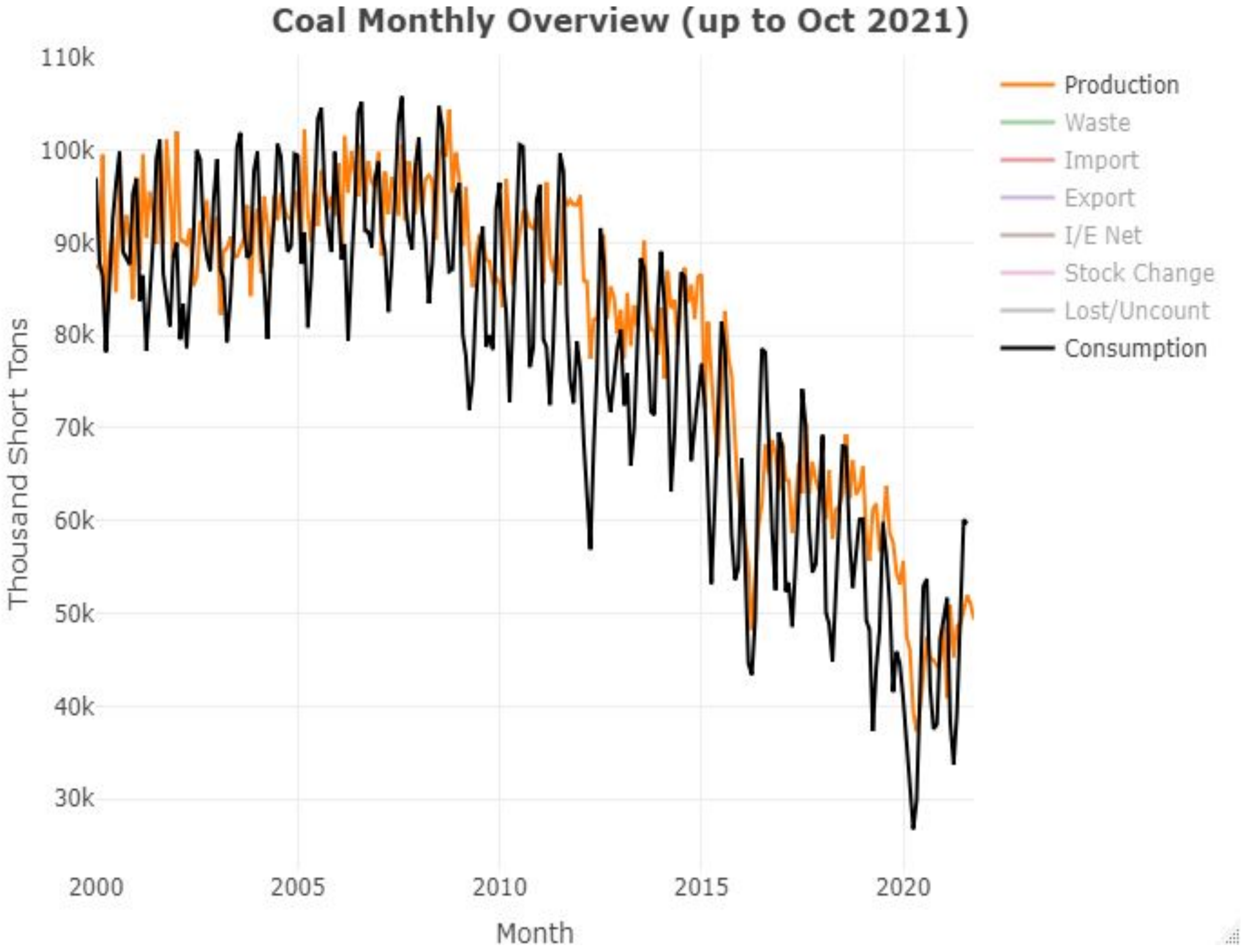


Petroleum conclusions - general insights

- Petroleum consumption is dominant in the transport and quite relevant in the industrial sector.
- The 2008 crisis and above all the 2020 pandemic had huge impact on those sectors. As a result the petroleum consumption saw a negative trend in 2008 and a shock in 2020.
- Omitting those two years the general trend is slightly increasing.
- The models results suggest that, at least in the next few years, this trend isn’t going to change.

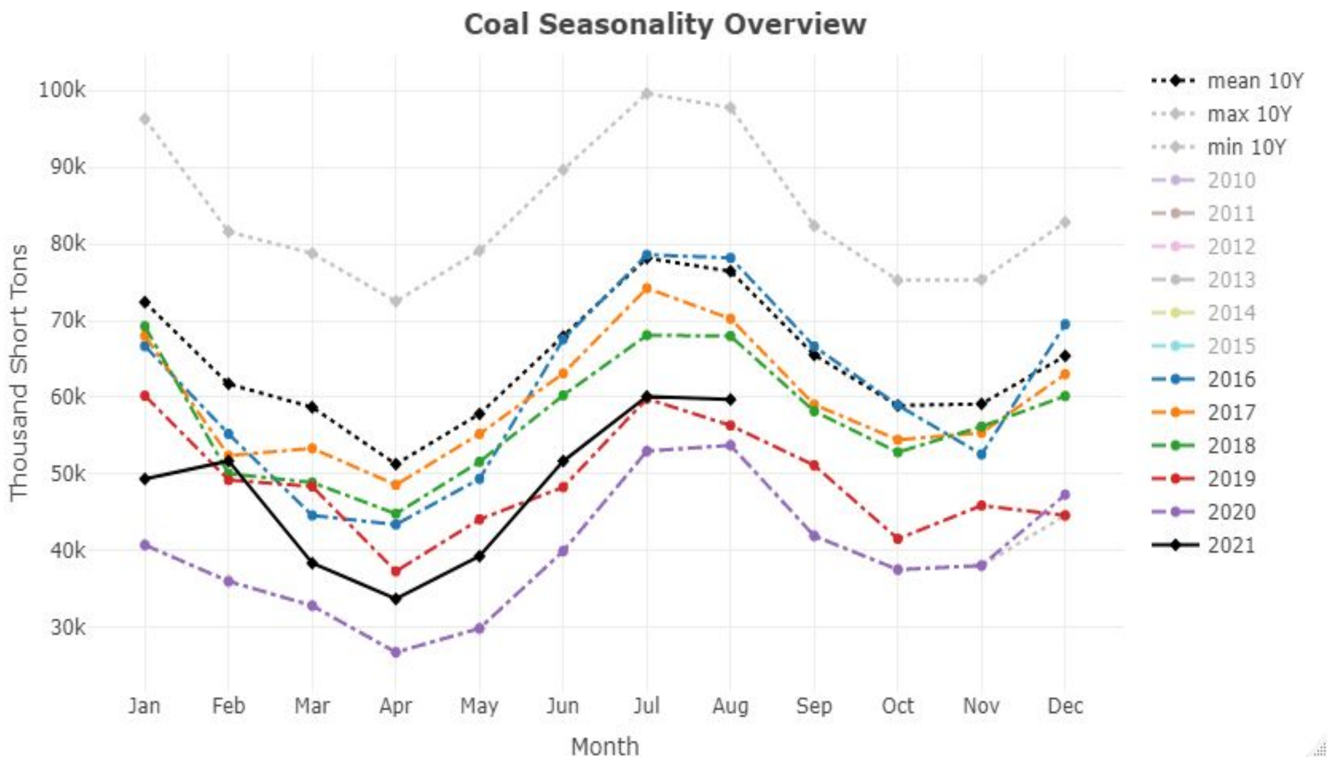
US Monthly Coal Consumption





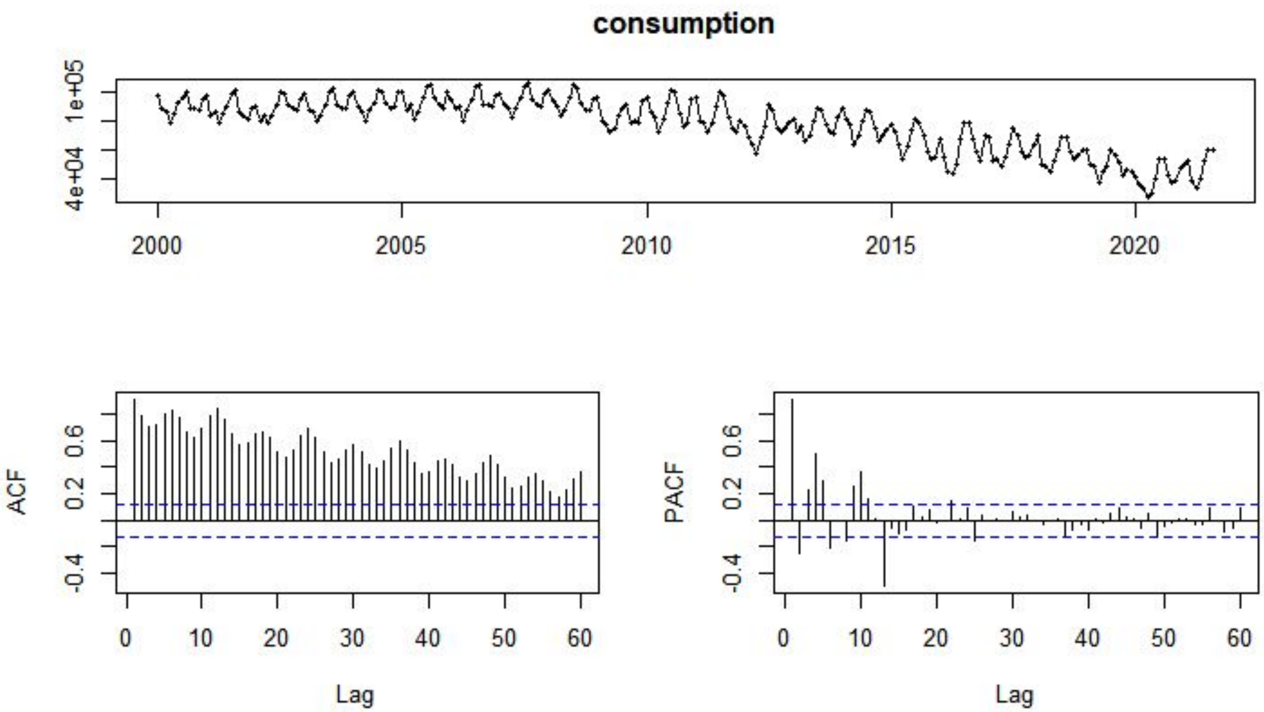
- **Consumption** ≈ Production
- Decreasing **trend**
- **Seasonality**
- Note: the **order** is of **Million Short Tons** per month!! (1 ST ~ 900 Kg)

Seasonal overview



- ➢ Consumption has a **sinusoid** behaviour throughout the **year** period.
- ➢ The **overall value** was **decreasing** in the last years.
- ➢ **2020** has seen a **slight extra drop** due to the pandemic.
- ➢ The effect seems to have already been **absorbed**.

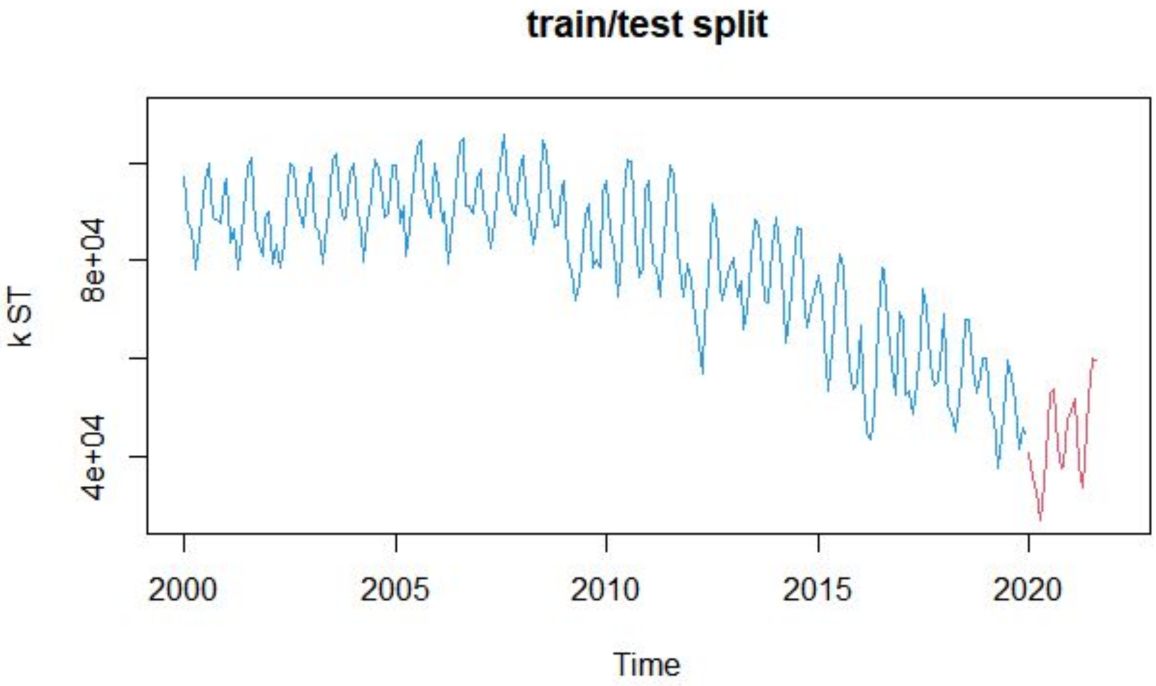
Modelling - warm up



- ➢ **Past** data can be exploited.
- ➢ There is a **trend** component.
- ➢ There is a **seasonal** component (of lag 6).
- ➢ The first and the seasonal **lag** data are the most relevant.

Modelling - tested models

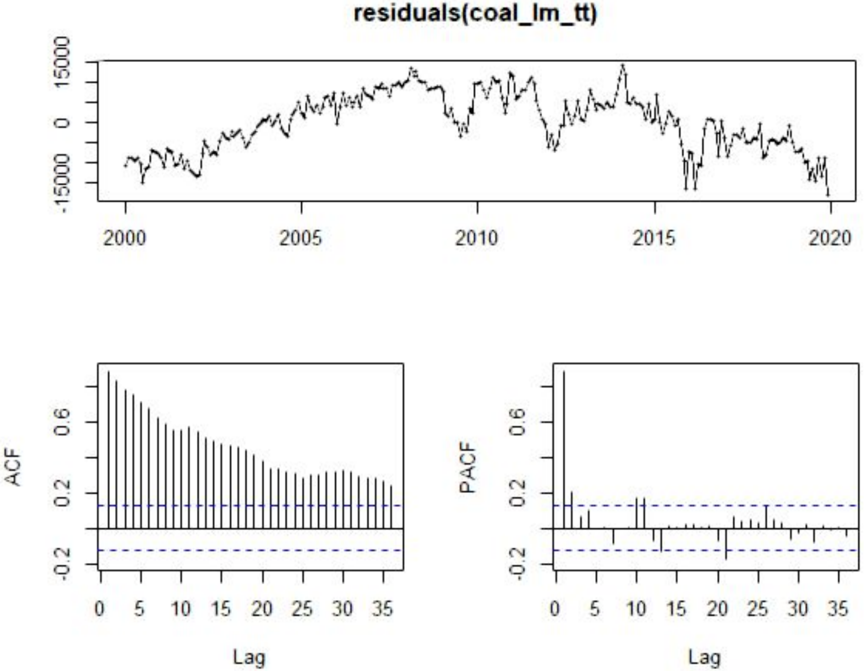
1. Linear model
2. GAM
3. Holt-Winter
4. ARIMA



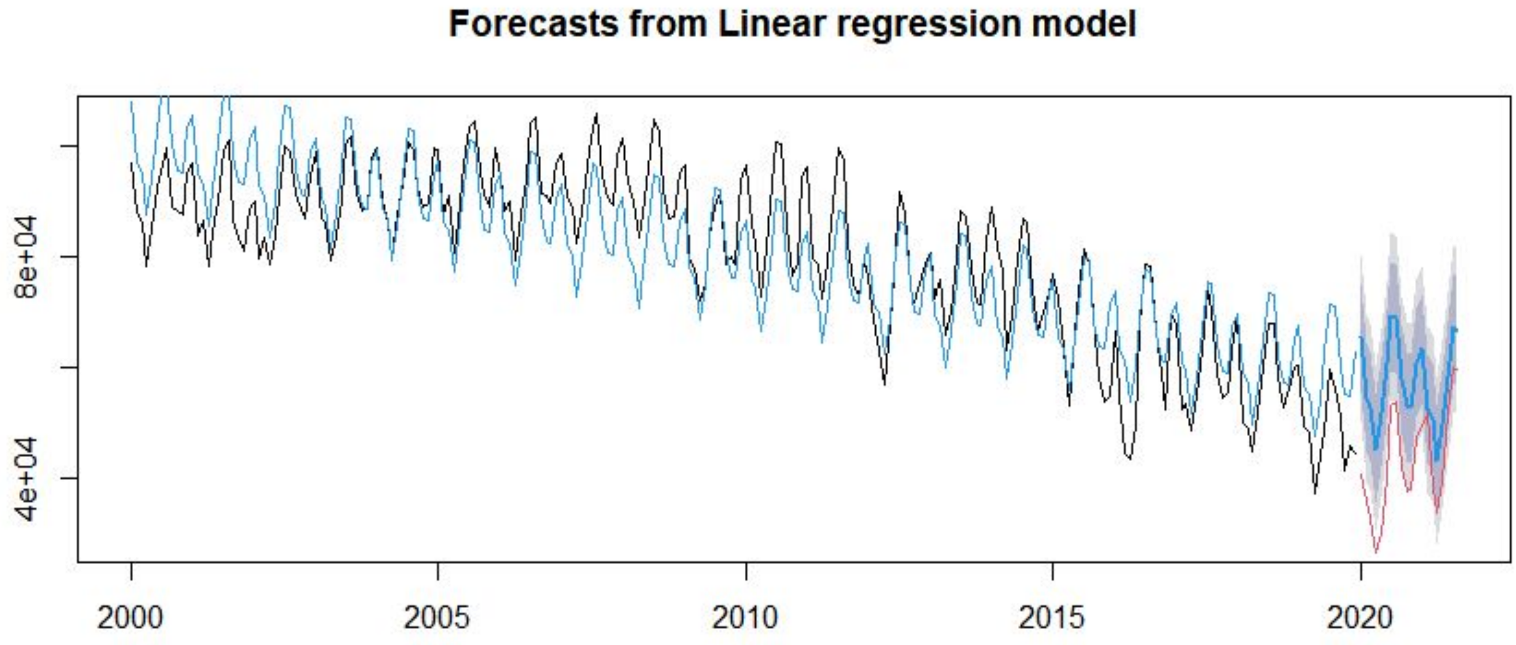
Modelling - Linear Model

```
tslm(formula = coal_train_set ~ trend + season)

Residuals:
     Min       1Q   Median       3Q      Max
-18337.0  -5274.3    389.9   5852.5  14302.8

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 107973.340   1840.095  58.678  < 2e-16 ***
trend         -176.668      6.921 -25.525  < 2e-16 ***
season2     -10661.648   2346.246  -4.544 8.97e-06 ***
season3     -12147.700   2346.276  -5.177 4.95e-07 ***
season4     -19570.691   2346.327  -8.341 7.15e-15 ***
season5     -12954.010   2346.399  -5.521 9.19e-08 ***
season6      -3980.500   2346.491  -1.696  0.09119 .
season7       4919.344   2346.603   2.096  0.03716 *
season8       4660.504   2346.735   1.986  0.04824 *
season9      -6138.271   2346.889  -2.615  0.00951 **
season10    -10755.231   2347.062  -4.582 7.59e-06 ***
season11    -10903.030   2347.256  -4.645 5.76e-06 ***
season12     -2661.253   2347.470  -1.134  0.25813
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7419 on 227 degrees of freedom
Multiple R-squared:  0.7932,    Adjusted R-squared:  0.7822
F-statistic: 72.54 on 12 and 227 DF,  p-value: < 2.2e-16
```

```
                  RMSE       MAE      MAPE
Training set  7215.707  6079.207  8.095656
Test set     15307.296 14165.607 36.344068
```



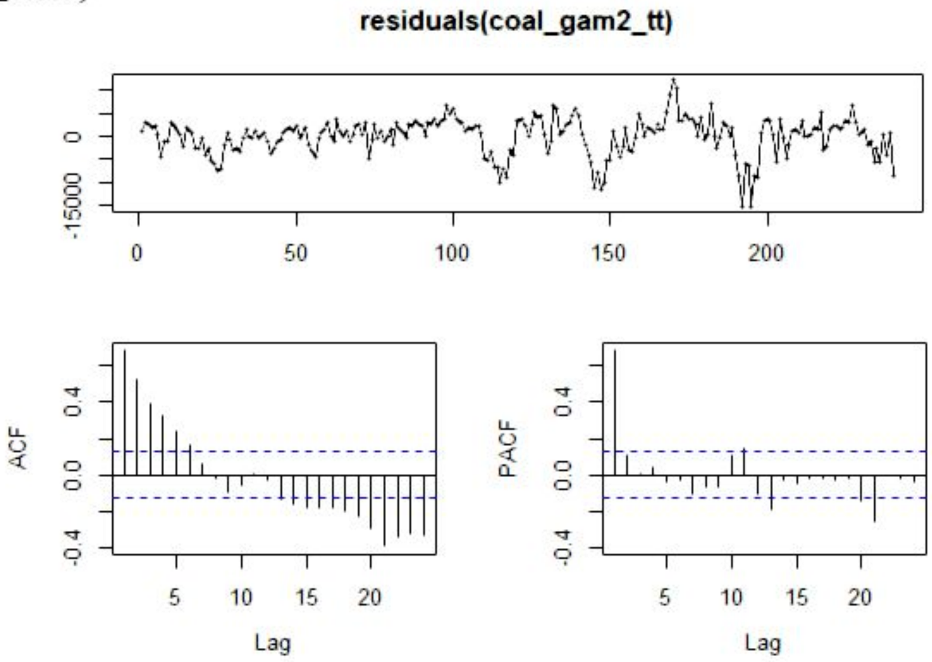
Linear Model - test forecast



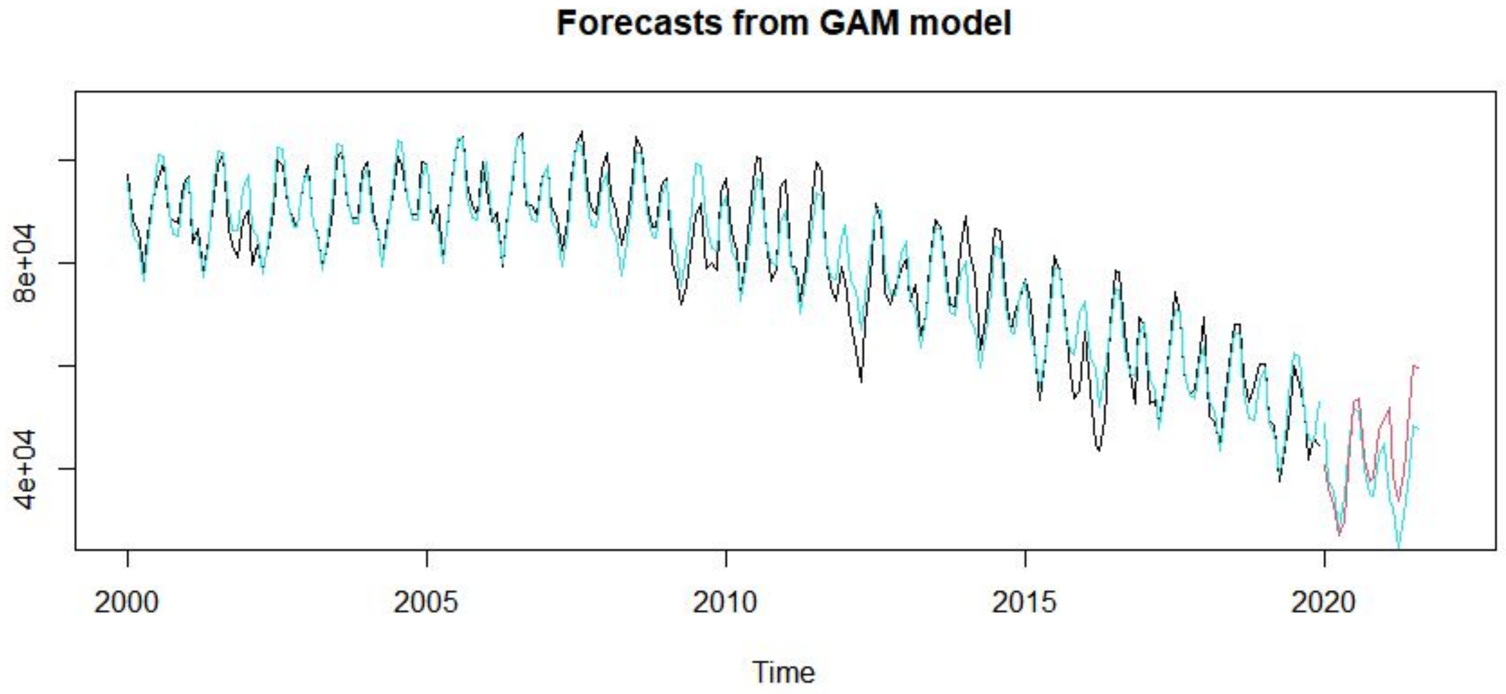
Modelling - GAM

```
Call: gam(formula = coal_train_set ~ lo(t) + seas, data = df_gam_train)

AIC: 4690.346

Anova for Parametric Effects
              Df     Sum Sq    Mean Sq  F value    Pr(>F)
lo(t)       1.00 3.5577e+10 3.5577e+10 2120.987 < 2.2e-16 ***
seas       11.00 1.2323e+10 1.1202e+09   66.784 < 2.2e-16 ***
Residuals 224.72 3.7694e+09 1.6774e+07
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects
            Npar Df Npar F     Pr(F)
(Intercept)
lo(t)           2.3 227.86 < 2.2e-16 ***
seas
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
                 RMSE      MAE      MAPE
Training set 3963.051 2954.402  4.049468
Test set     7450.581 5984.618 13.45717
```



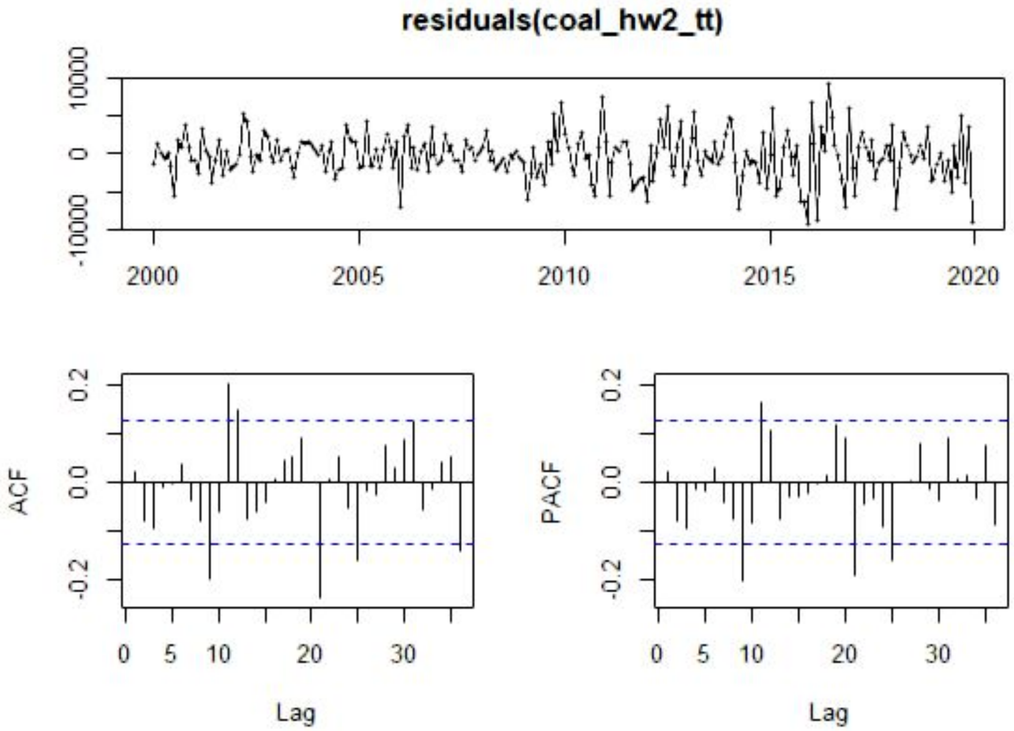
GAM - test forecast



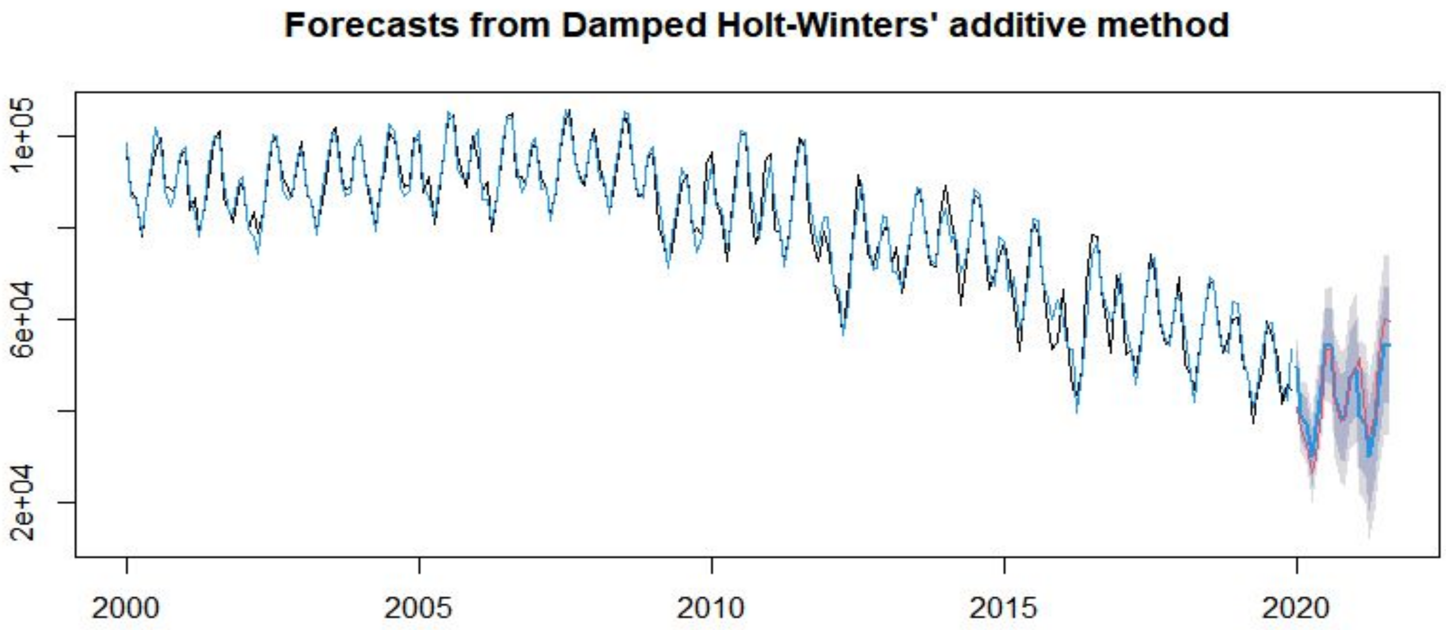
Modelling - Holt-Winter

```
Forecast method: Damped Holt-Winters' additive method

Model Information:
Damped Holt-Winters' additive method

Call:
 hw(y = coal_train_set, damped = TRUE)

  Smoothing parameters:
    alpha = 0.7033
    beta  = 1e-04
    gamma = 1e-04
    phi   = 0.9657

 AIC  5198.440
```

	RMSE	MAE	MAPE
Training set	3025.294	2284.820	3.194272
Test set	4928.855	3730.715	8.991099



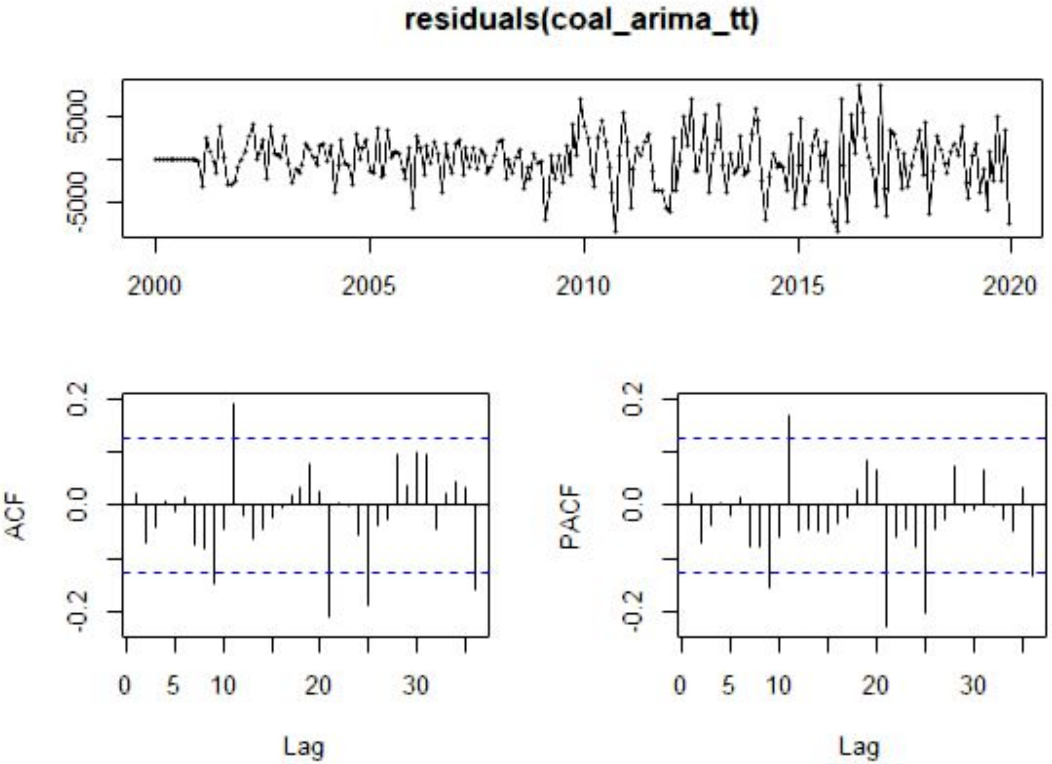
Holt-Winter - test forecast



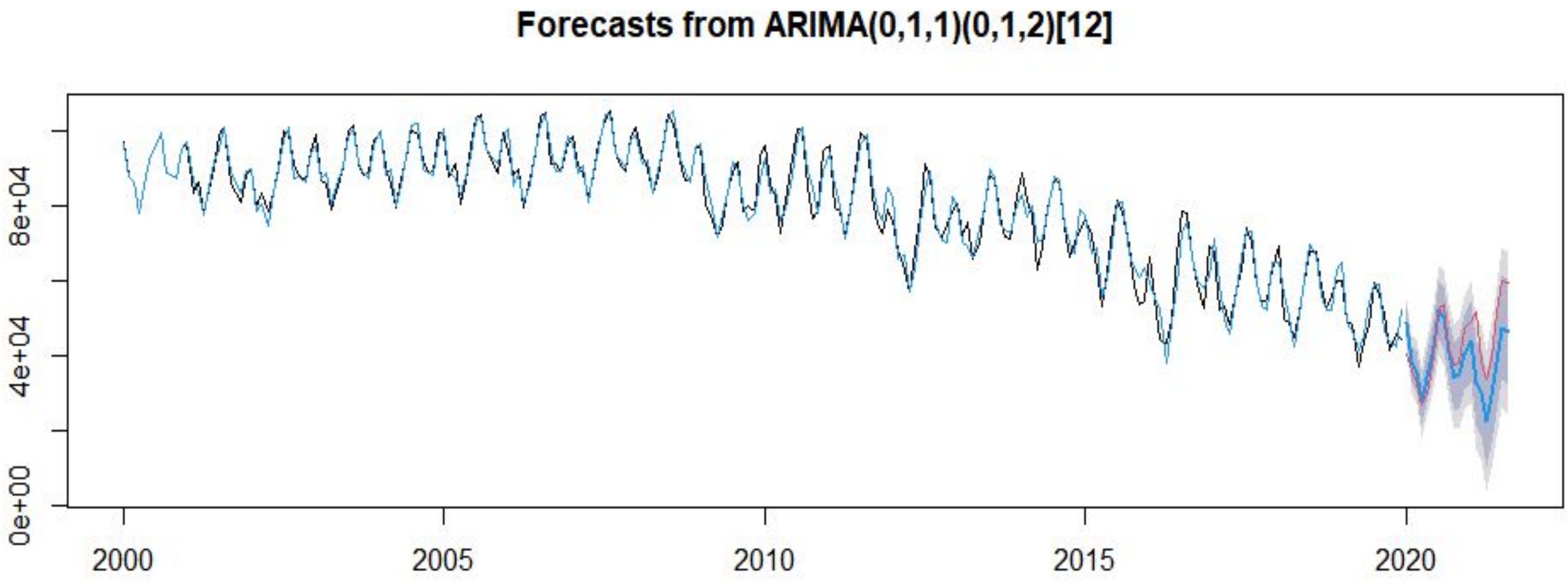
Modelling - ARIMA

```
Series: coal_train_set
ARIMA(0,1,1)(0,1,2)[12]

Coefficients:
          ma1     sma1     sma2
      -0.3266  -0.7066  -0.1503
s.e.   0.0699   0.0717   0.0682

AIC=4324.65
```

```
                 RMSE      MAE      MAPE
Training set 3068.285 2323.557  3.260595
Test set     8323.527 6593.639 14.801376
```



ARIMA - test forecast



Test Results Table

Model	AIC	RMSE	MAE	MAPE
Linear	-	15307	14166	36.34
GAM	4690	7451	5985	13.46
H-W	5198	**4929**	**3731**	**8.99**
ARIMA	4325	8324	6594	14.80

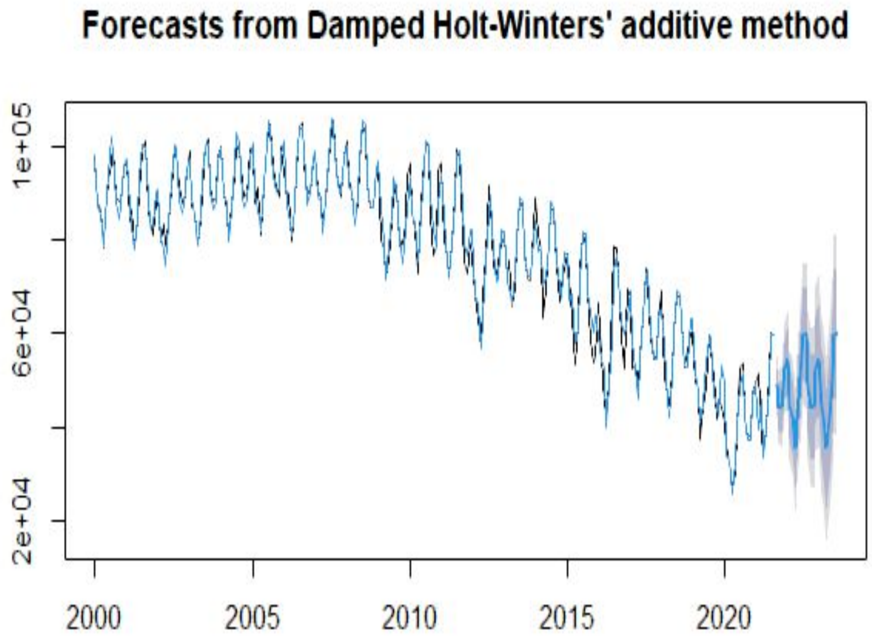
Modelling - Conclusion

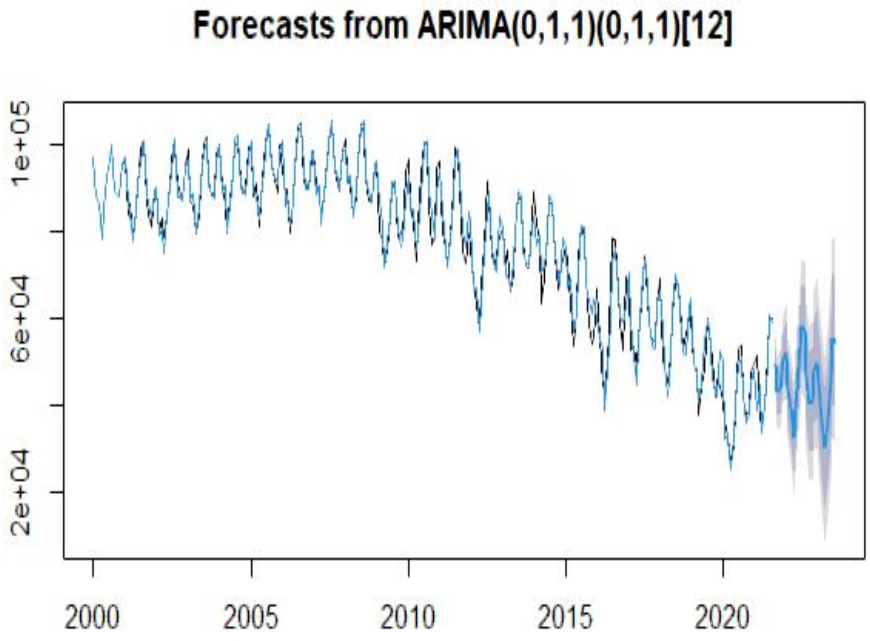
- The **linear** model is **poor** from all points of view.
  Not surprising because forcing linearity is a bit too strong hp.

- The **GAM** and **ARIMA** model are **good** but underestimate a bit in the test.
  Moreover the residual inspection suggest the GAM is missing something.

- The **Holt-Winter** is **good**.

Holt-Winter and ARIMA - future forecast





Coal conclusions - general insights

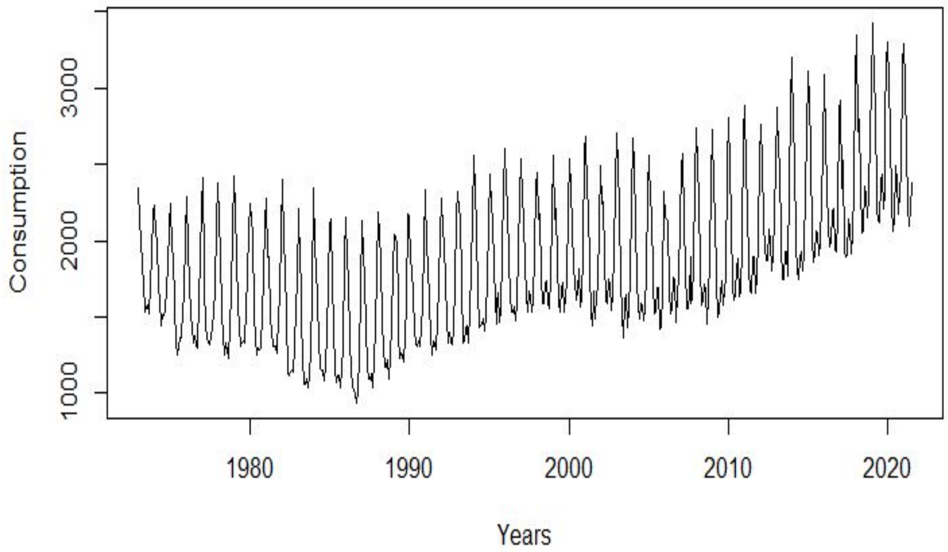
- ➢ Coal consumption is mainly reportable for the electricity production.
- ➢ The general coal trend is decreasing except for the current year that seems to settle at the 2019 level.
- ➢ This suggests that the coal is being replaced by gas and renewables.
- ➢ The models results suggest that the coal consumption will remain stable at the actual levels in the next few years.
- ➢ So indeed there is an ongoing decarbonization process but it will take a long time.

Natural Gas

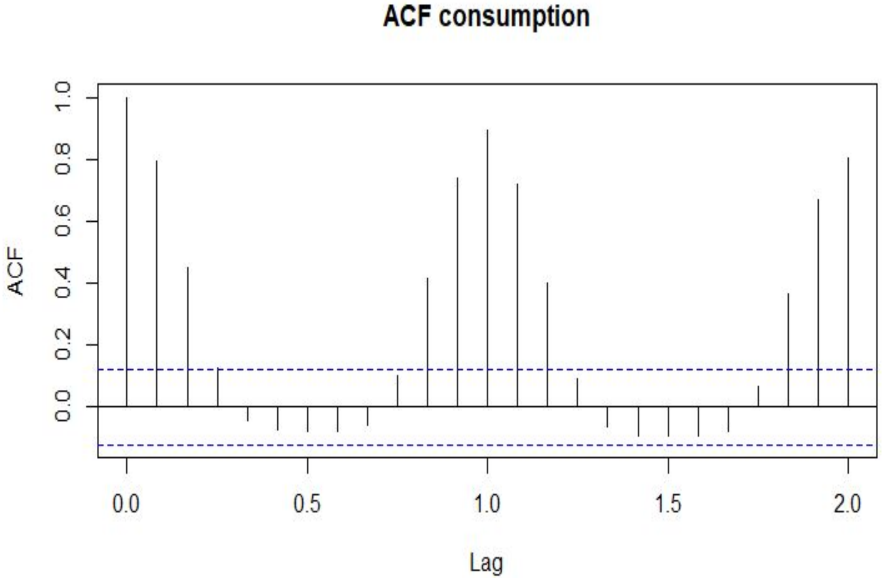
- **Natural gas production will continue to grow,** mainly to support increasing energy consumption in developing Asian economies.
- Driven by increasing populations and fast-growing economies, consumption of liquid fuels **will grow the most in non-OECD Asia**, where total consumption nearly doubles by 2050 from 2020 levels in the Reference case. Because these countries will consume more liquid fuels than they produce in the Reference case, we project that **non-OECD Asia will supplement local production with increased imports of crude oil and finished petroleum products**. The increased imports will primarily be supported by increased production in the Middle East. In the Reference case, by 2050, **non-OECD Asia will become the largest importer of natural gas**, and Russia will become the largest net exporter of natural gas.
- But what happens in United States?

Source: **U.S. Energy Information Administration.** *EIA projects accelerating renewable consumption and steady liquid fuels growth to 2050*

EDA

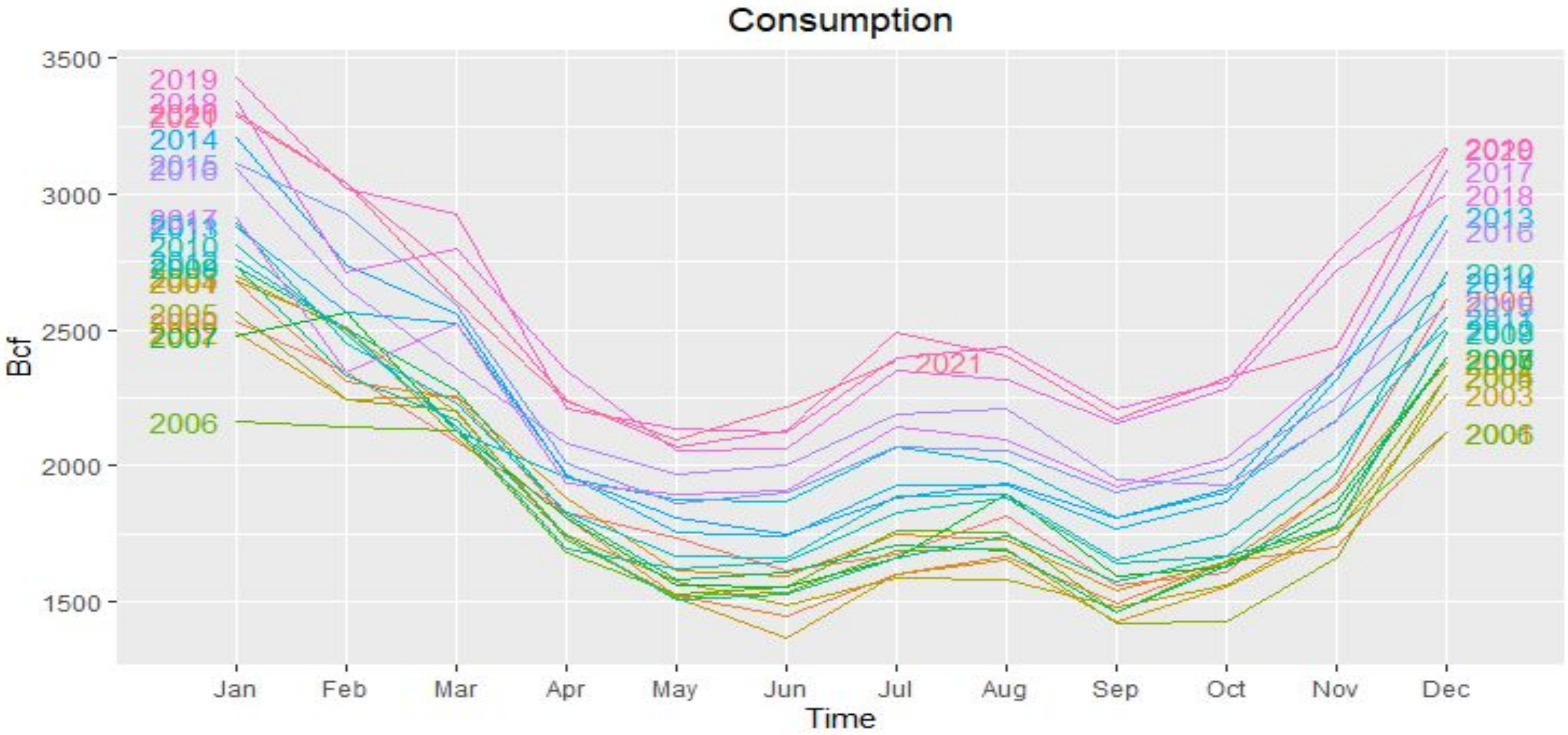


Consumption
3000
2000
1000
1980
1990
2000
2010
2020
Years



ACF consumption
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
Lag

EDA



Consumption
Bcf
Time
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
3500
3000
2500
2000
1500
2019
2014
2006
2021
2003

Model Overview

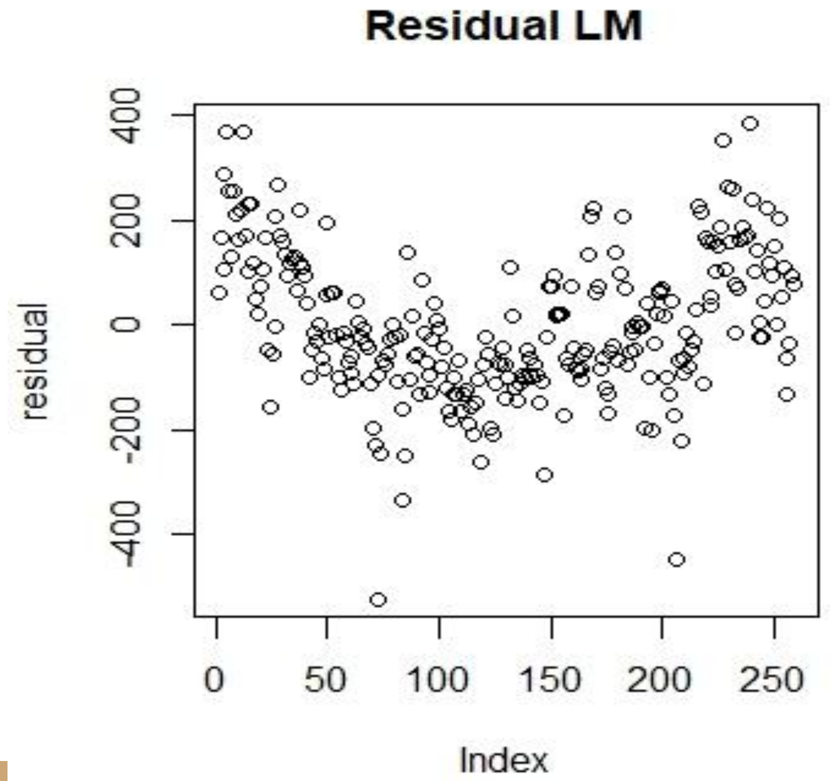
- Linear Regression
- Bass Model
- GBM
- GGM
- TSLM
- ARIMA

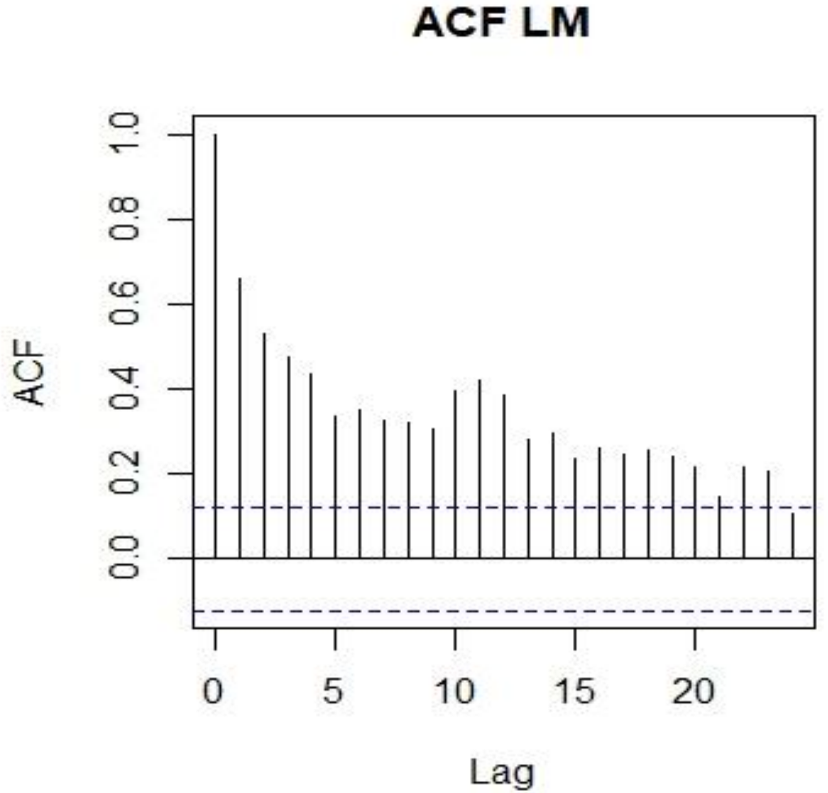
Linear regression (1)

Coefficients:

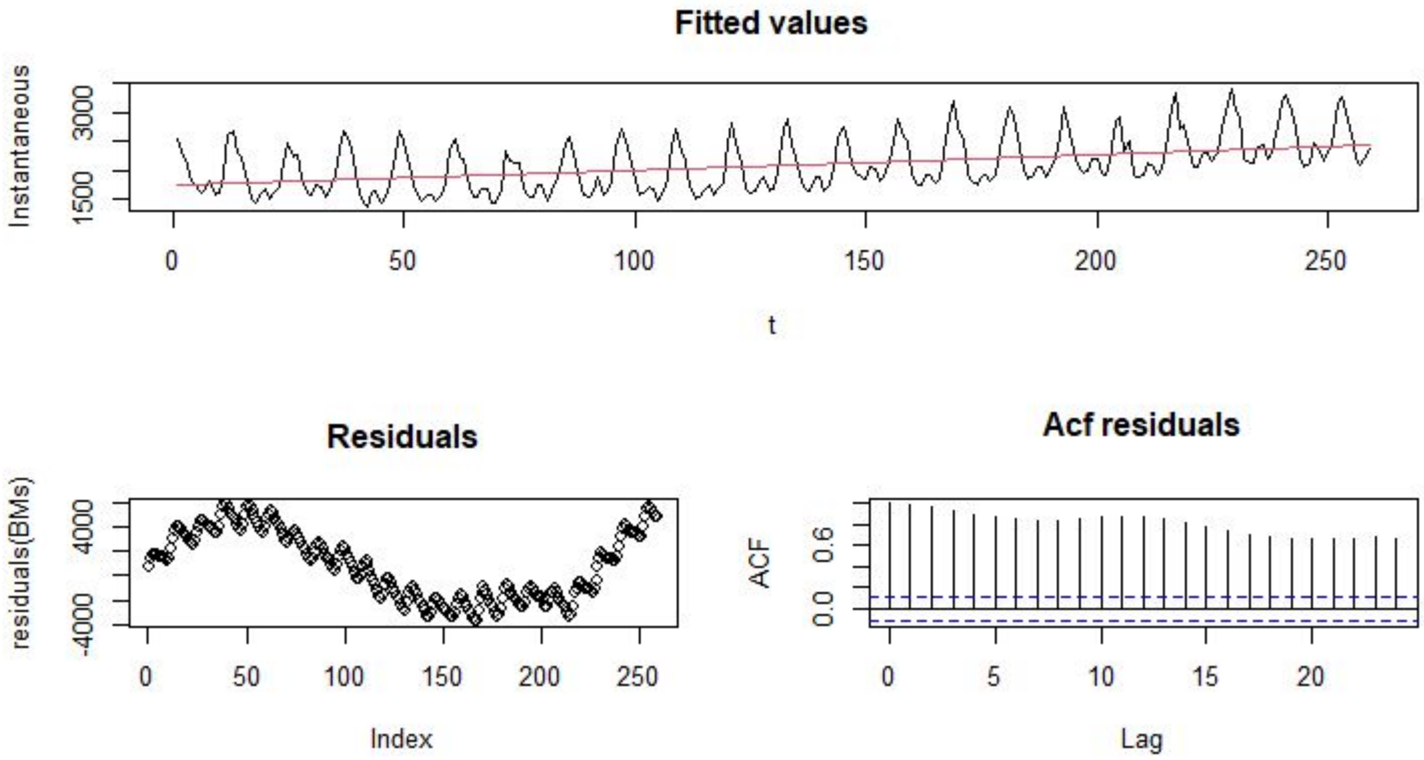
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2468.7263	33.6147	73.442	< 2e-16	***
tt	3.0195	0.1175	25.707	< 2e-16	***
seas2	-302.2000	42.6012	-7.094	1.38e-11	***
seas3	-499.6368	42.6017	-11.728	< 2e-16	***
seas4	-940.8376	42.6025	-22.084	< 2e-16	***
seas5	-1117.1498	42.6037	-26.222	< 2e-16	***
seas6	-1127.3224	42.6051	-26.460	< 2e-16	***
seas7	-948.3176	42.6069	-22.257	< 2e-16	***
seas8	-933.9443	43.1054	-21.667	< 2e-16	***
seas9	-1143.9282	43.1059	-26.538	< 2e-16	***
seas10	-1053.4761	43.1067	-24.439	< 2e-16	***
seas11	-791.1781	43.1078	-18.353	< 2e-16	***
seas12	-263.7551	43.1092	-6.118	3.70e-09	***

Linear Regression (2)



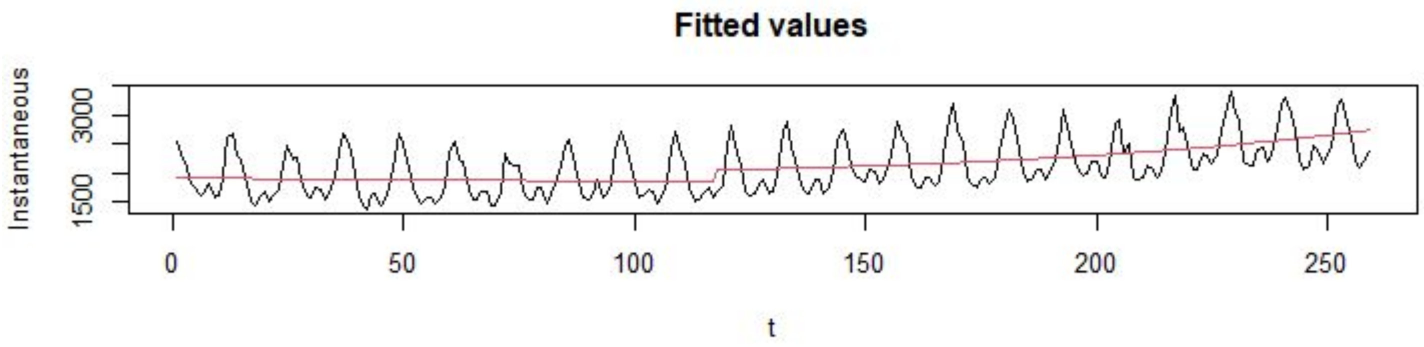


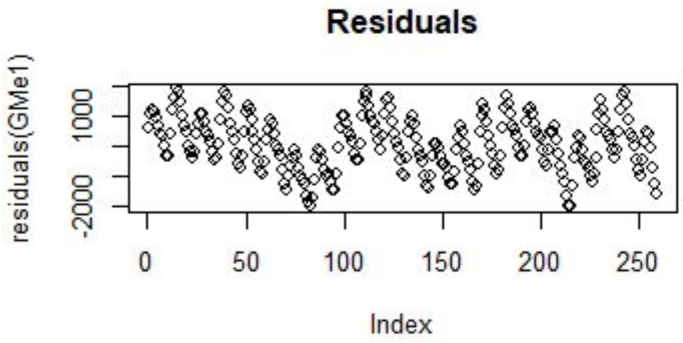
Standard Bass Model

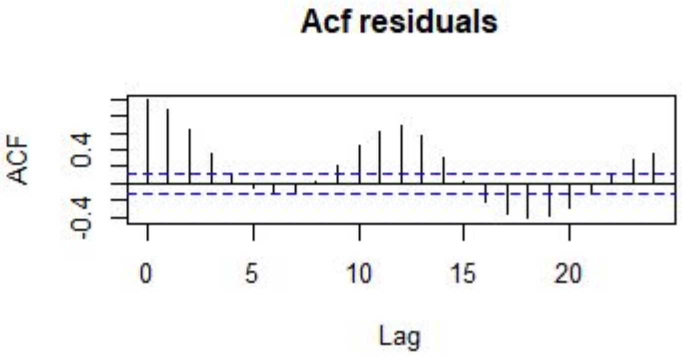


- ➢ ***m*** market potential relevant
- ➢ ***p*** innovation relevant
- ➢ ***q*** not relevant

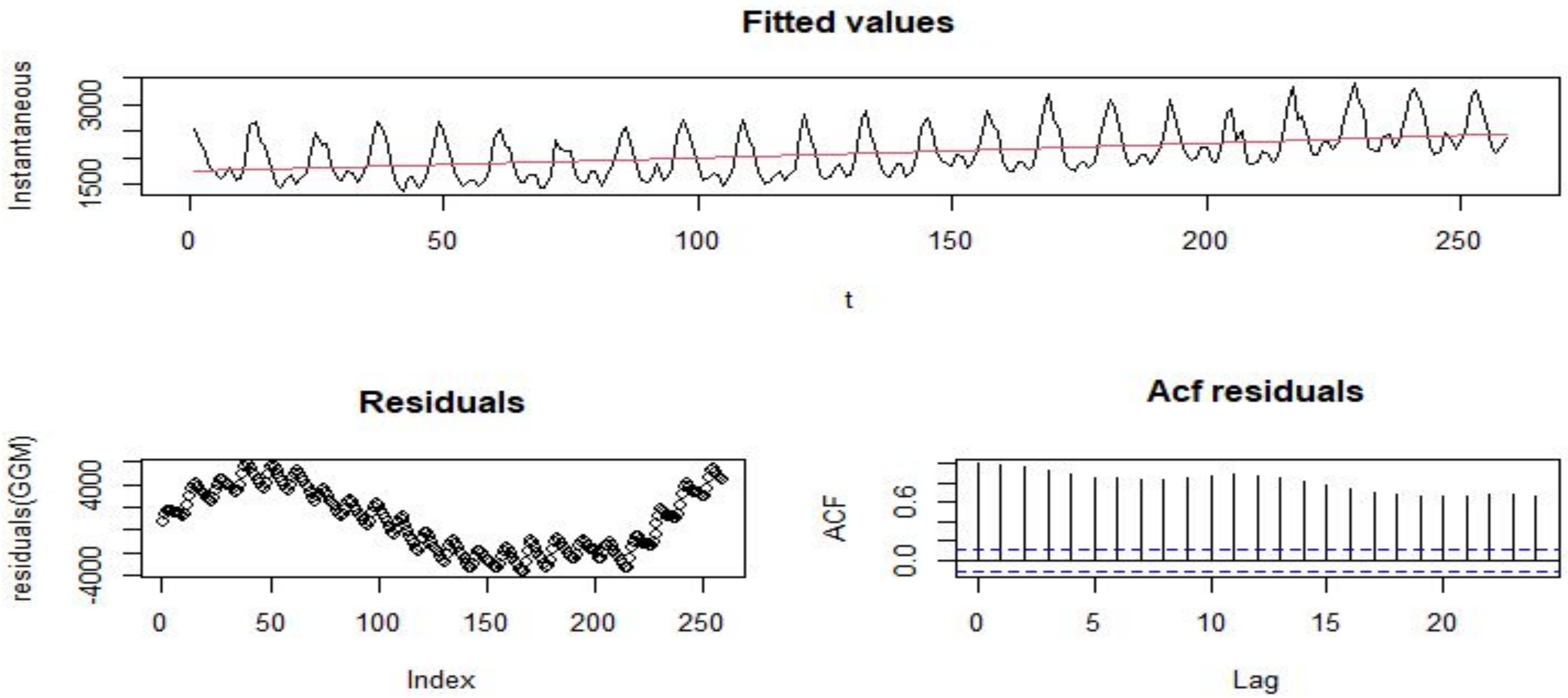
Generalized Bass Model







GGM model


Fitted values
Instantaneous
1500
3000
0
50
100
150
200
250
t
Residuals
residuals(GGM)
-4000
4000
0
50
100
150
200
250
Index
Acf residuals
ACF
0.0
0.6
0
5
10
15
20
Lag

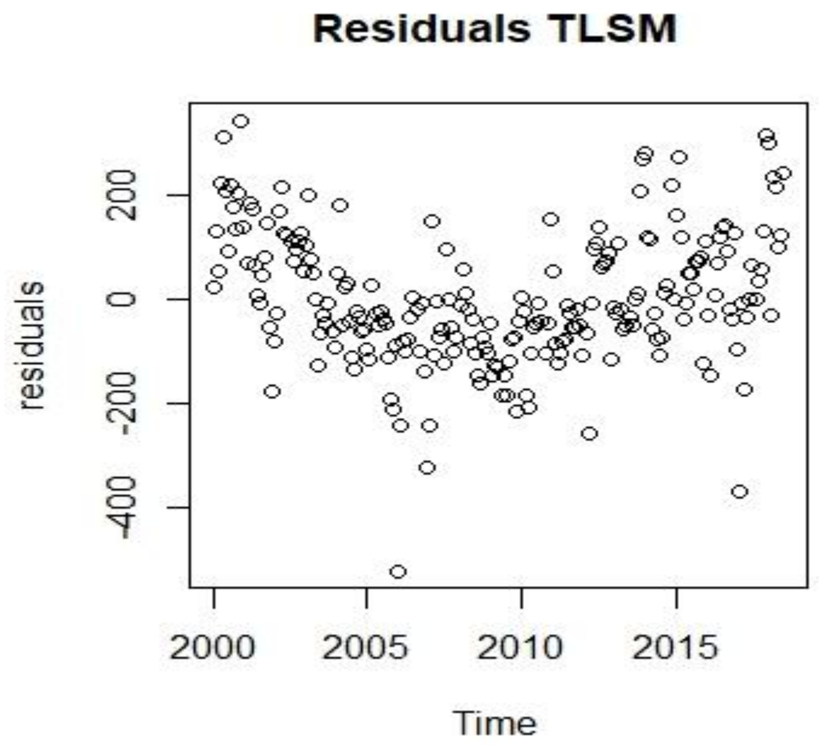
TLSM (1)

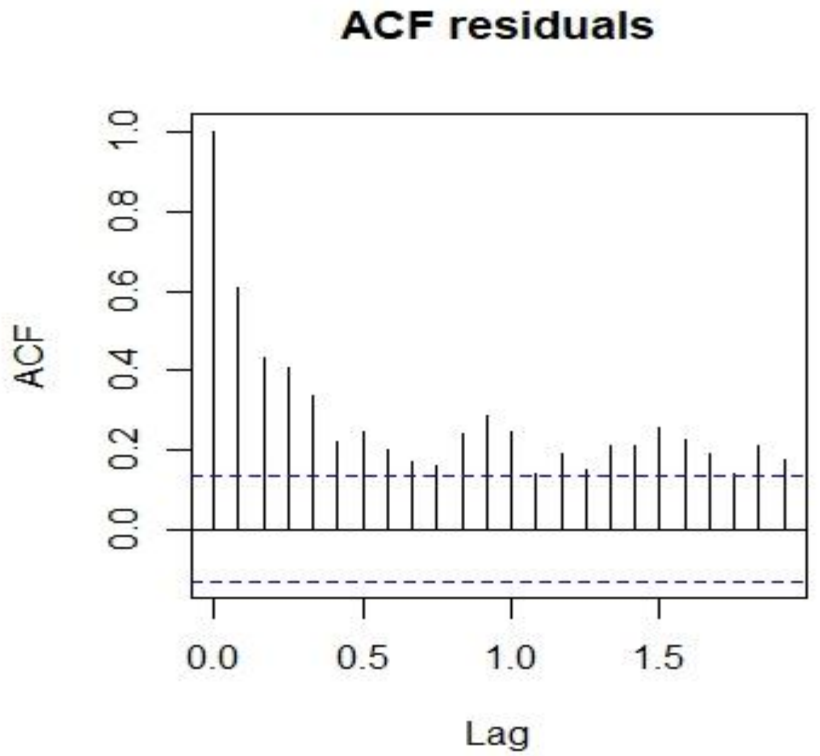
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2508.1167	33.3940	75.107	< 2e-16	***
trend	2.4551	0.1358	18.080	< 2e-16	***
season2	-301.1488	42.3345	-7.114	1.75e-11	***
season3	-483.4219	42.3351	-11.419	< 2e-16	***
season4	-911.5534	42.3362	-21.531	< 2e-16	***
season5	-1093.8609	42.3377	-25.837	< 2e-16	***
season6	-1113.4970	42.3397	-26.299	< 2e-16	***
season7	-947.3462	42.3421	-22.374	< 2e-16	***
season8	-933.8674	42.9184	-21.759	< 2e-16	***
season9	-1143.6169	42.9191	-26.646	< 2e-16	***
season10	-1058.0375	42.9201	-24.651	< 2e-16	***
season11	-807.7408	42.9216	-18.819	< 2e-16	***
season12	-269.3178	42.9236	-6.274	1.98e-09	***

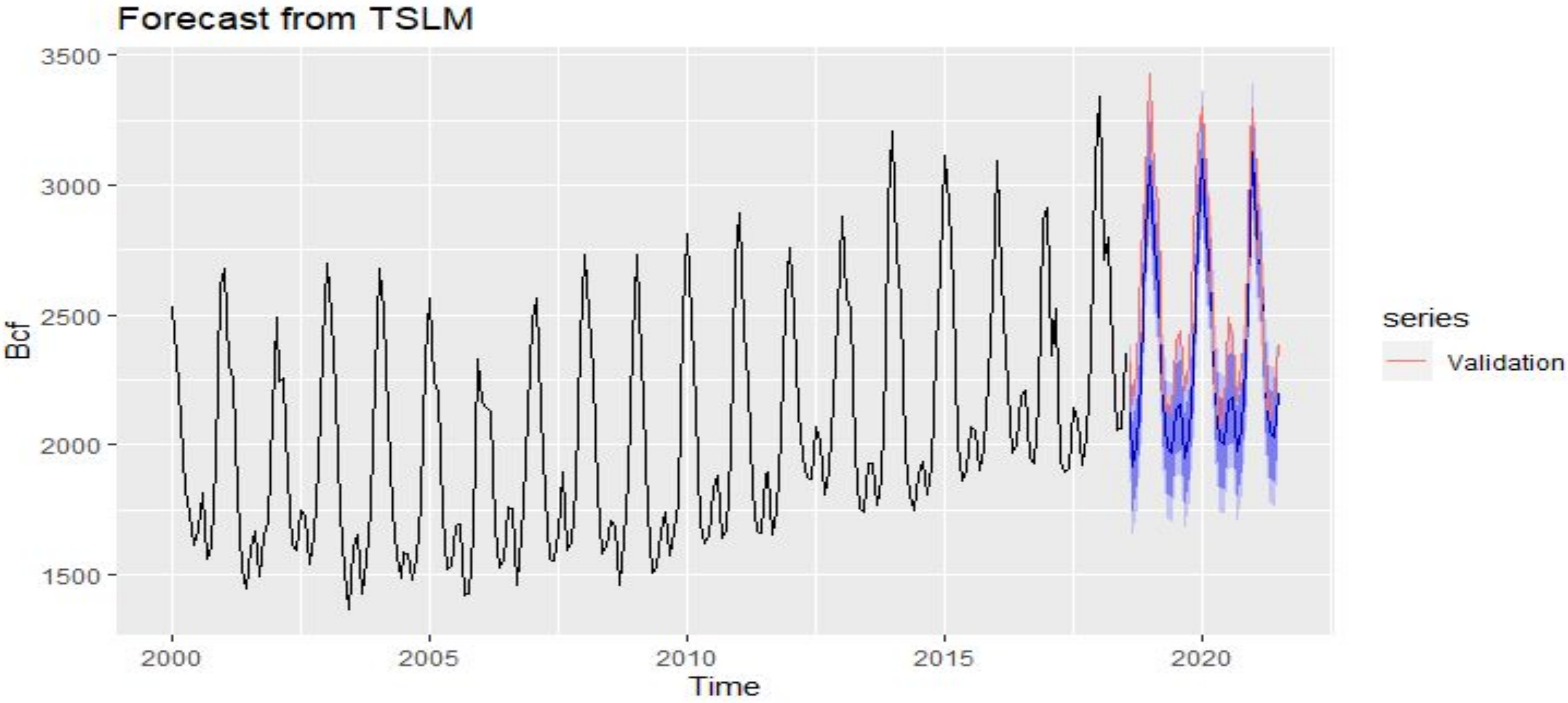
- Train-test split

TLSM(2)


Residuals TLSM
residuals
200
0
-200
-400
2000
2005
2010
2015
Time

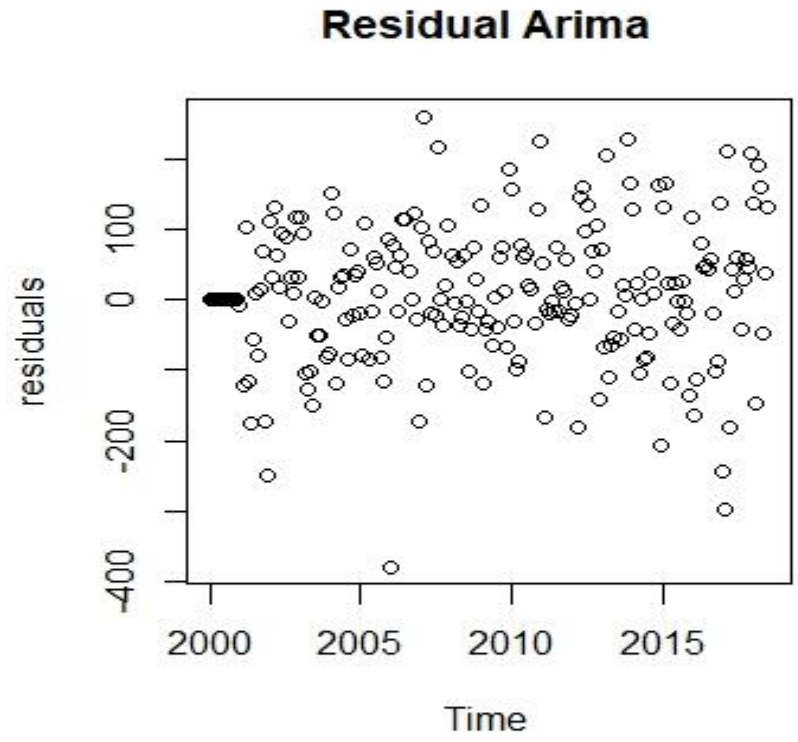

ACF residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
Lag

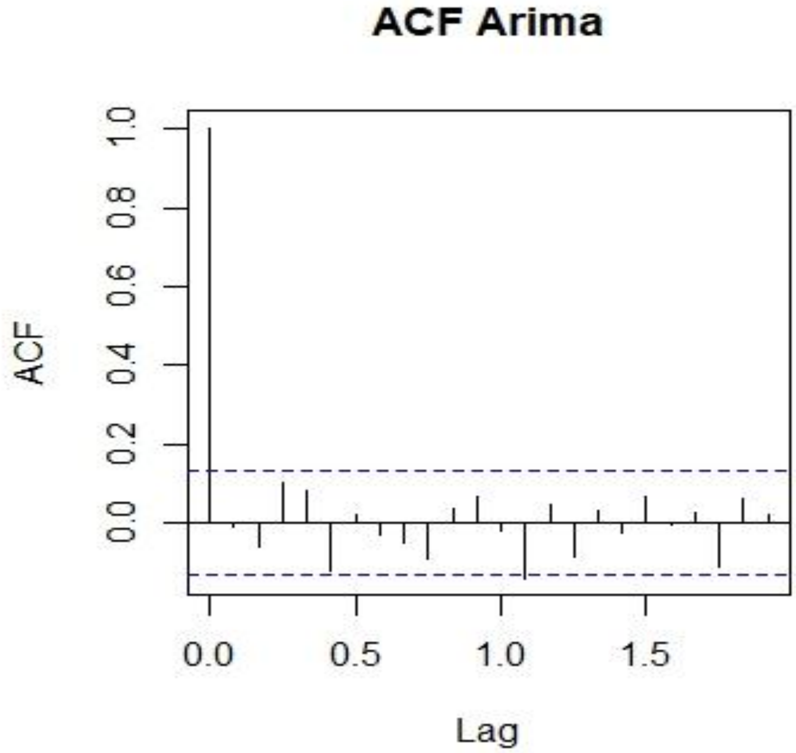
TLSM (3)



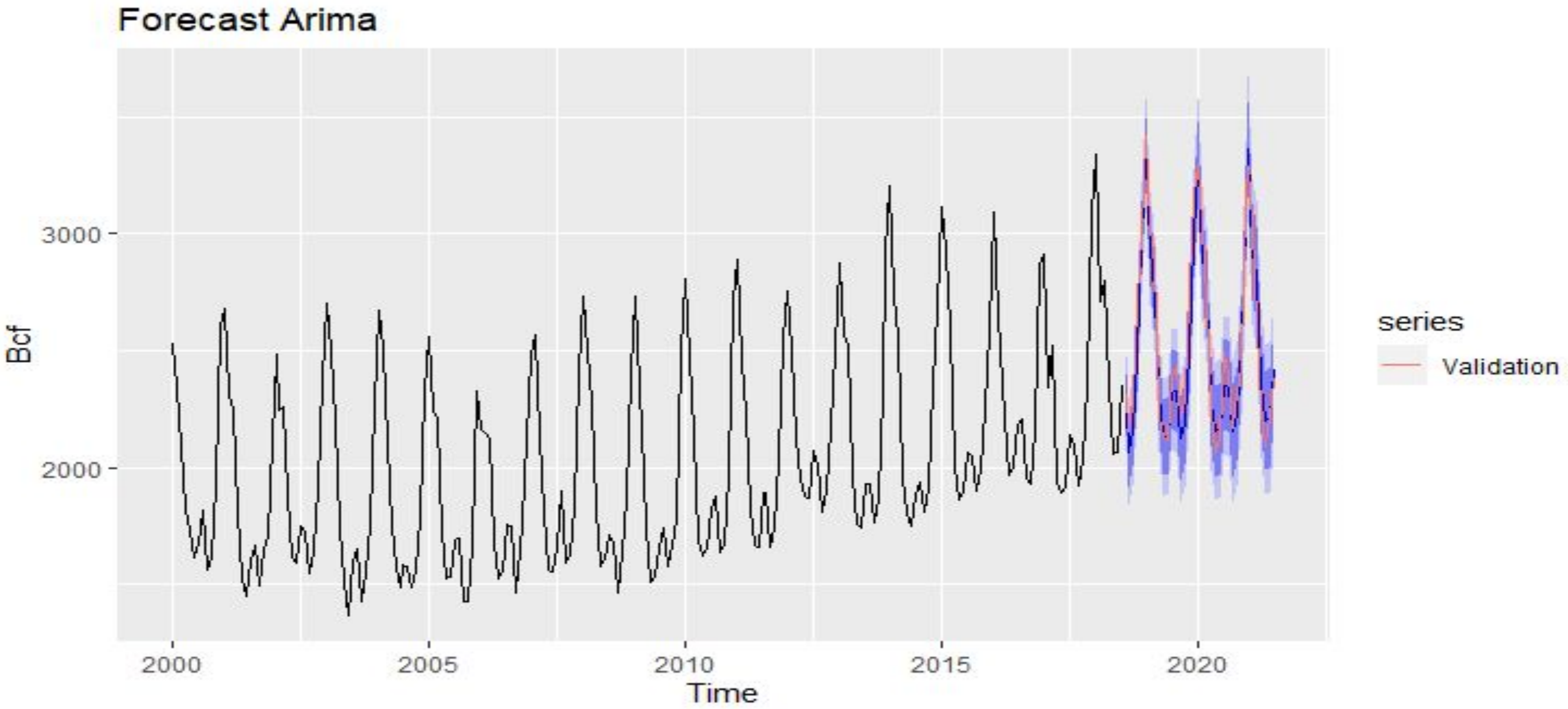
Forecast from TSLM
3500
3000
2500
2000
1500
Bcf
2000
2005
2010
2015
2020
Time
series
Validation

ARIMA ARIMA(2,1,1)(2,1,1)





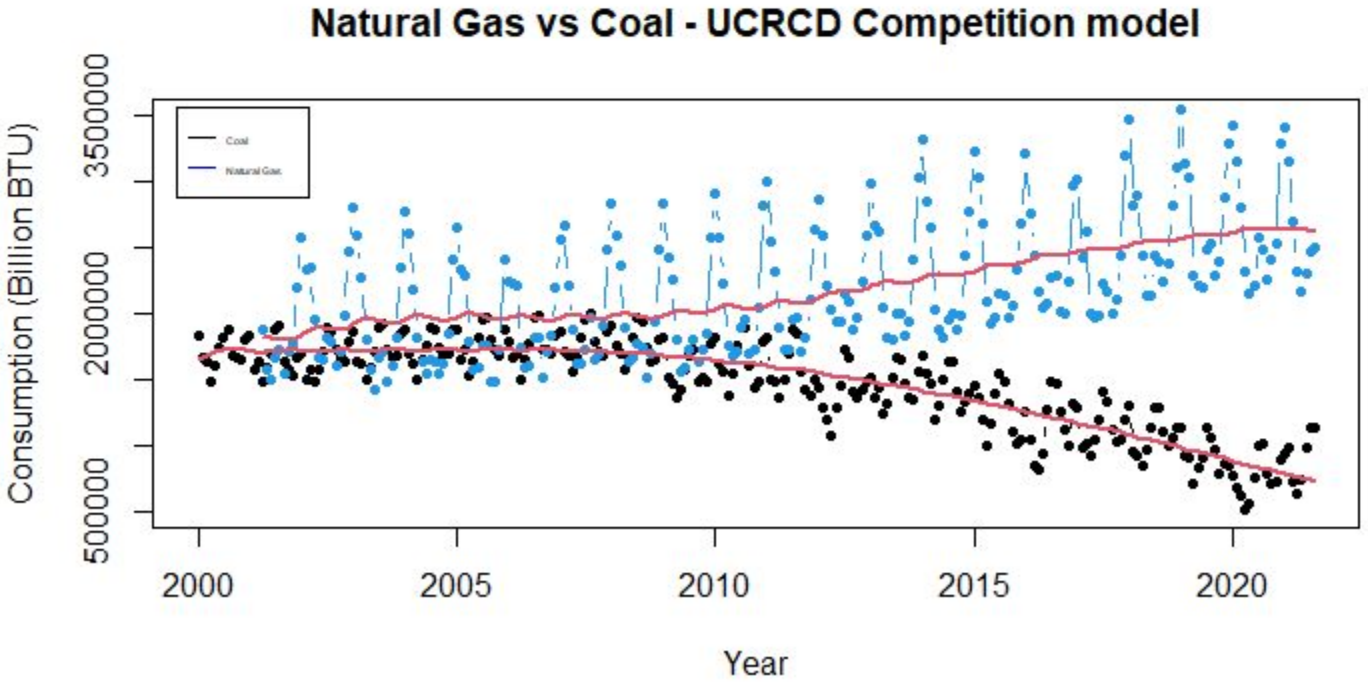
ARIMA


Forecast Arima
3000
2000
Bcf
2000
2005
2010
2015
2020
Time
series
Validation

Final results Natural Gas

Model	Dataset	RMSE	MAE	MPE	MAPE	AIC
LMTS	Training Set	126,62	98,22	-0,34	4,86	2.820,03
	Test Set	238,88	210,53	8,04	8,13	
ARIMA	Training Set	98,85	74,53	0,22	3,57	2567,85
	Test Set	119,50	95,91	24,74	3,71	

Competition model: Natural gas against Coal



Competition model: Natural gas against Coal

Coefficient	Estimate	Description	p-value	
m1	81748176.84664	Market Potential 1 stand-alone	2.92e-02	*
p1a	0.01999	Innovation coefficient 1 stand-alone	2.33e-02	*
q1a	0.03209	Imitation coefficient 1 stand-alone	6.63e-02	.
mc	1329264478.57830	Market Potential 1 competition	2.41e-23	***
p1c	0.00128	Innovation coefficient 1 competition	2.55e-23	***
q1c+delta	0.00915	Within-product wom 1 competition	2.31e-20	***
q1c	-0.00511	Cross-product wom 1 competition	3.44e-02	*
p2c	0.00136	Innovation coefficient 2 competition	5.42e-04	***
q2c	0.02968	Within-product wom 2 competition	4.47e-03	**
q2c-gamma	-0.02779	Cross-product wom 2 competition	1.32e-03	**

Final Conclusion

- For the forecast the best models are: *Holt-Winters* and *ARIMA*
- Natural gas is slowly replacing the coal Resource (US domain)
- Renewable resources consumption is gradually growing but, up to now, the gap to fill is still big