

A Time Series Practice Focusing on Coal-Generated Electricity in Germany - PROJECT2

YITE ZHU, YUANJIAN ZHOU, ZIYING CHENG

2019/12/12

I. Introduction

1. Background

Energy has always been a major concern. It is closely related to development and national security, and is associated with all aspects of people's daily life. Despite that for decades, diversified energy sources have been introduced to generate electricity, coal still remains to be one of the most economical sources, accounting for a great proportion of electricity generation. The pollution and environmental externalities of coal usage in electricity generation, however, have been one major political issue, especially for European Union. Meanwhile, that the resource reserve is limited also makes it not as economical as before. Led by countries like Germany, the world is struggling to make a trade-off between economic growth and environmental protection. And Germany has been doing a great job in withdrawing the coal for fuel gradually. Being one of the most important economy and electricity & coal consumer, Germany is a good case for analytical problems relating to electricity generation and environment. This report takes a closer look at the electricity generation by coal in Germany, builds time series models, tries to develop a comprehensive understanding of its trend and volatility, and make predictions, to help provide a glimpse into broader perspective of global energy market.

2.data description

The report utilizes two serial monthly data of German coal from Jan 2002 to Aug 2019, provided by Federal Statistics Office Germany. One series is monthly data of electricity generated by crude lignite, referred to as brown coal. With a relatively low carbon content around 60-70, it is considered the lowest rank of coal, also most harmful to health, while the cost of it remains cheap. The other series describes the electricity generated by hard coal, which is a usual source adopted by economies around the world. Blessed with the highest carbon content, it has the highest energy density, ranked the highest of all coals. From the basic knowledge, the two raw materials for electricity generation have positive relationships to the whole coal industry production development and fluctuations. In addition, they may be negatively correlated with each other because they are both exclusively used to generate electricity, acting as the substitutional products to some extent.

Thus, it is interesting to explore their respective innovations through time and the relationship between them from the view of time series. The center of the study is to fit time series model for them and make predictions. Then, vector autoregression (VAR) are applied to figure out the interdependence between them. To sum up, we generate a full model to optimize the performance of model and conduct forecast, bringing a more accurate prospect of the Germany electricity generation by coal. The Unit of monthly data is megawatt-hours(MWH). To narrow down the scale of data and transform data to a similar scale, as well as easing potential heteroskedasticity, we exert log transformation to data. Then we use different models and plots to make analysis as follows.

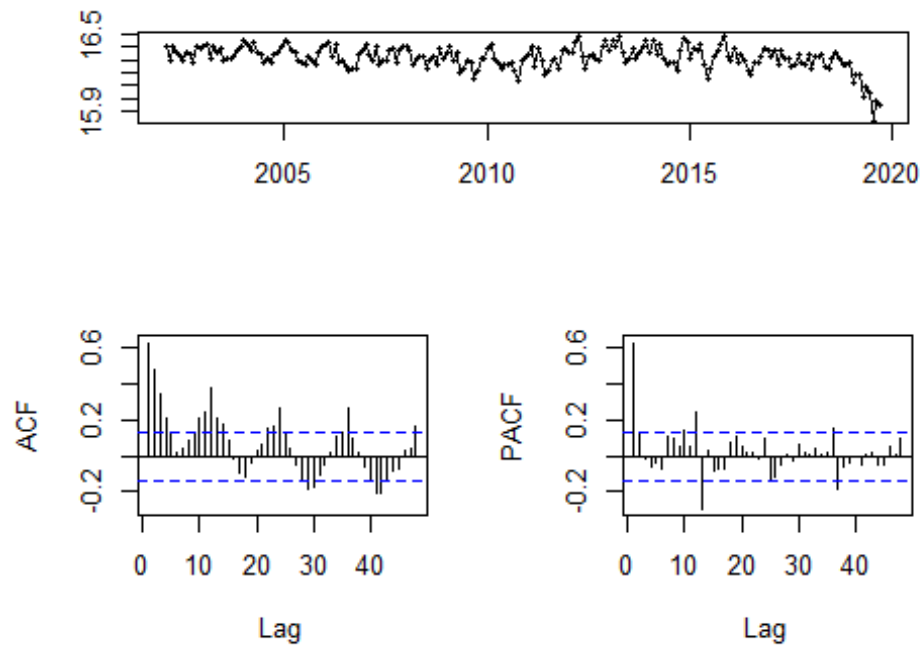
In practice, since we do not know the future yet, we can not evaluate the prediction of our models in a normal way. Most often, we will separate our data into estimation set and prediction set, and used the model fitted in estimation set to predict data in prediction set. Since data in prediction set comes from the past, we usually call this process backtest.

Here in our analysis, we have 212 observations for each variable from Jan 2002 to Aug 2019. We separate them into 140 estimation ones and 72 prediction ones.

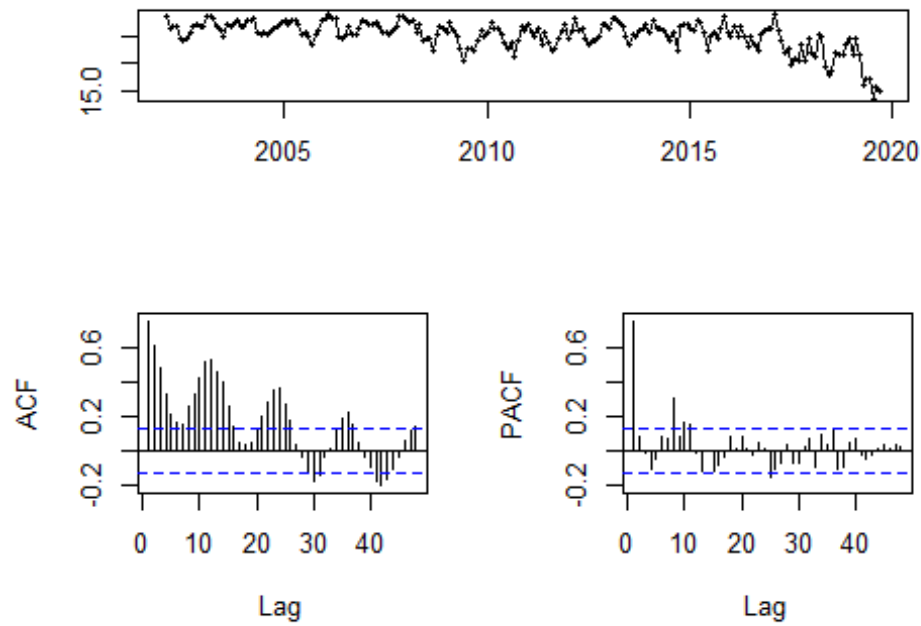
II. Results

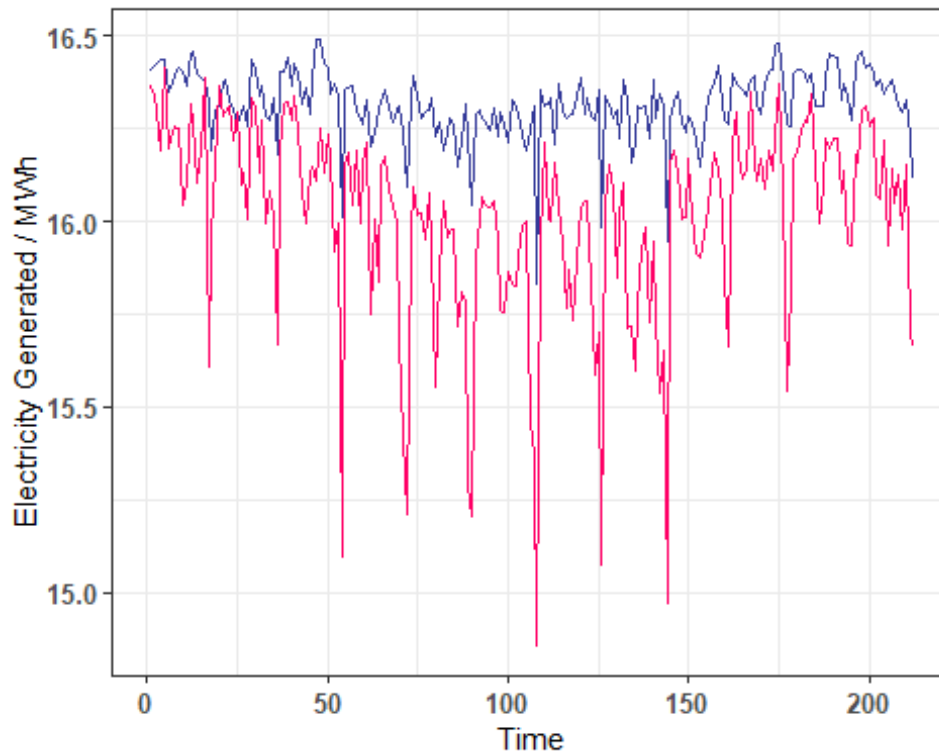
1

Crude Lignite



Coal





The plot of two series have revealed the characteristics of the two log series. We can observe a vague drift in the log series. Cyclical and seasonality are present considering the stochastic fluctuation in the overall series. It is more prudent to include all three components in the model.

Then we turn to ACF and PACF for more implications. The seasonality is apparent judging from the periodic repetition of spikes in acf. ACF and PACF both displays decaying characteristic but with periodic repetitions and fluctuations. This means that the two series requires more complicated models than simple MA(q) or AR(p).

To understand the interdependence, we plot the two series in one plot. The resulting graph shows the correlation between two series, which points to the possible lead-lag relationship of the two series and stresses the necessity of VAR model.

2

```
## Series: nt
## Regression with ARIMA(0,1,3)(2,0,0)[12] errors
##
## Coefficients:
##          ma1      ma2      ma3      sar1      sar2      xreg
##      -0.4012 -0.2682  0.1172  0.4380  0.2314 -0.0161
## s.e.   0.0687   0.0751  0.0728  0.0727  0.0736  0.0097
##
## sigma^2 estimated as 0.003069:  log likelihood=310.92
## AIC=-607.85   AICc=-607.29   BIC=-584.38
##
```

```

## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE
## Training set -0.003758202 0.05447996 0.04294953 -0.0241282 0.2635829
##              MASE      ACF1
## Training set 0.774937 -0.01232398

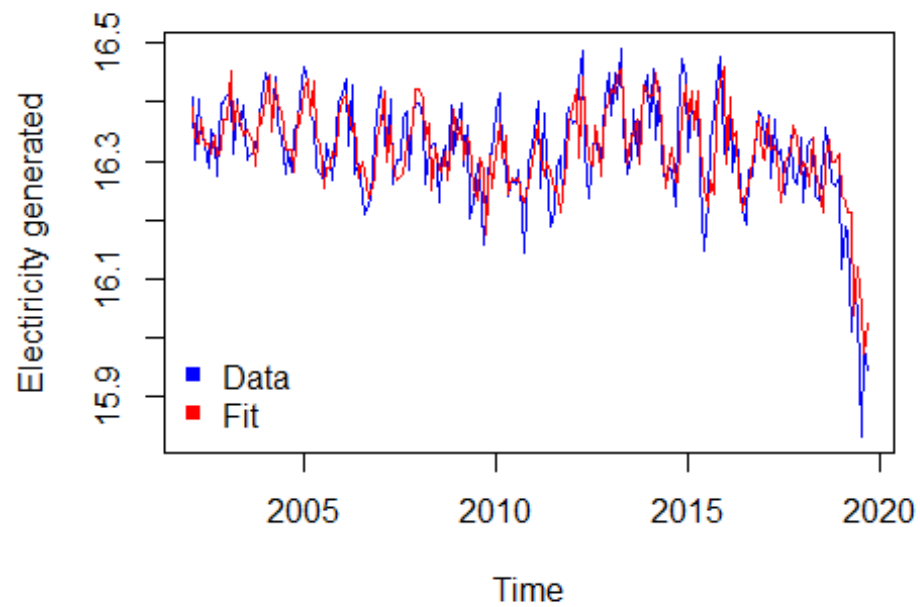
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1  -0.4011843  0.0686787 -5.8415 5.174e-09 ***
## ma2  -0.2682102  0.0751326 -3.5698 0.0003572 ***
## ma3   0.1171701  0.0727791  1.6099 0.1074107
## sar1  0.4379744  0.0726670  6.0271 1.669e-09 ***
## sar2  0.2314070  0.0735738  3.1452 0.0016595 **
## xreg -0.0160939  0.0096564 -1.6667 0.0955819 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Series: ct
## Regression with ARIMA(1,1,1)(0,0,2)[12] errors
##
## Coefficients:
##      ar1      ma1      sma1      sma2      xreg
##      0.4465 -0.8709  0.1746  0.2361 -0.0414
## s.e.  0.0911  0.0468  0.0797  0.0854  0.0152
##
## sigma^2 estimated as 0.02246: log likelihood=102.53
## AIC=-193.06 AICc=-192.65 BIC=-172.95
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE
## Training set -0.01400655 0.1477244 0.1135015 -0.09684077 0.7133019
##              MASE      ACF1
## Training set 0.8079134 -0.03376065

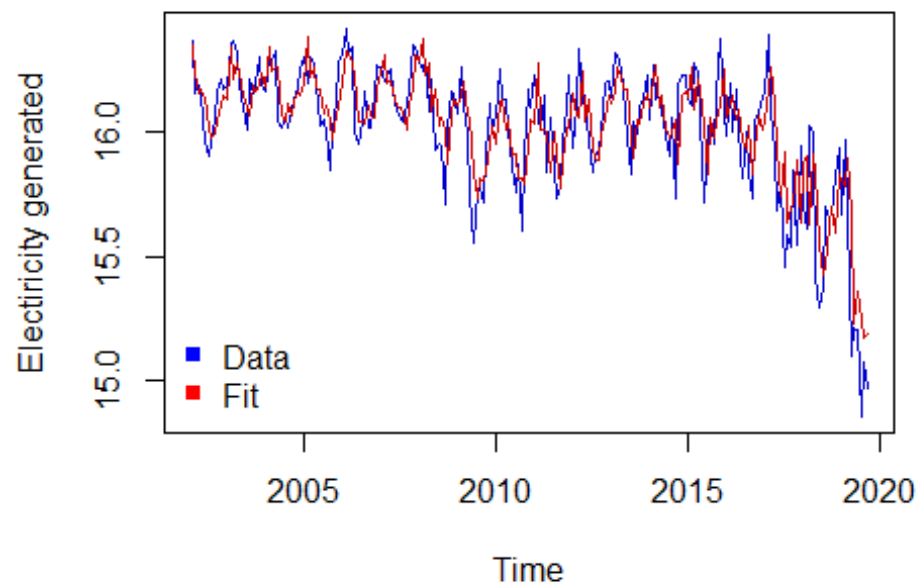
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.446529  0.091051  4.9042 9.383e-07 ***
## ma1  -0.870860  0.046840 -18.5921 < 2.2e-16 ***
## sma1  0.174570  0.079693  2.1905 0.028487 *
## sma2  0.236112  0.085373  2.7657 0.005681 **
## xreg -0.041354  0.015173 -2.7255 0.006421 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Fit for Crude Lignite



Fit for Coal



```
##  
## Box-Pierce test  
##
```

```
## data:  n1$residuals
## X-squared = 0.032199, df = 1, p-value = 0.8576

##
## Box-Pierce test
##
## data:  c1$residuals
## X-squared = 0.24163, df = 1, p-value = 0.623
```

Using Algorithms we obtain baseline models as follows:

Crude Lignite model has ARIMA(0,1,3)(2,0,0)12 with log trend.

Coal has ARIMA(1,1,1)(0,0,2)12 with log trend.

The results suggest that the model will include stochastic seasonality, trend, and cycles.

On graph, the fitted line captures the stochastic seasonality and cycles in a good manner. The residuals from both models are tested to be normal. The results of Box-Pierce test suggest their residuals are both white noise, demonstrated that all the patterns in original process have been detected.

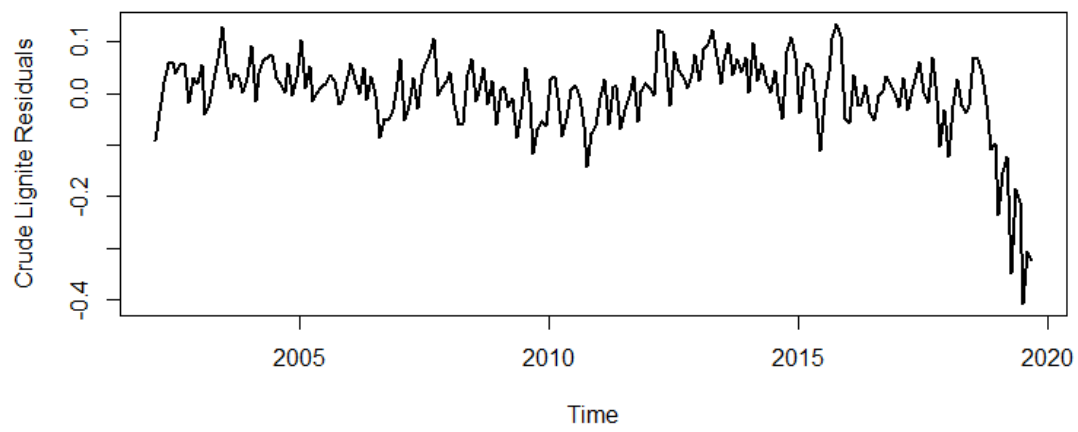
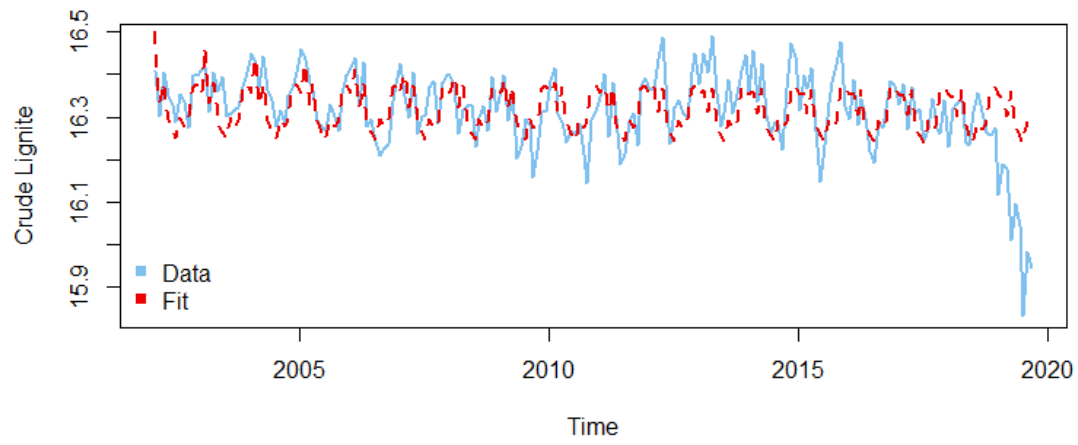
3

To produce a valid model for crude lignite, we try different components step by step, based on the ARIMA results from the baseline in part b. We started by digging into a combination of different trends and season, then trend and cycle, then all three components. AIC and BIC pointed to a best choice of all trials.

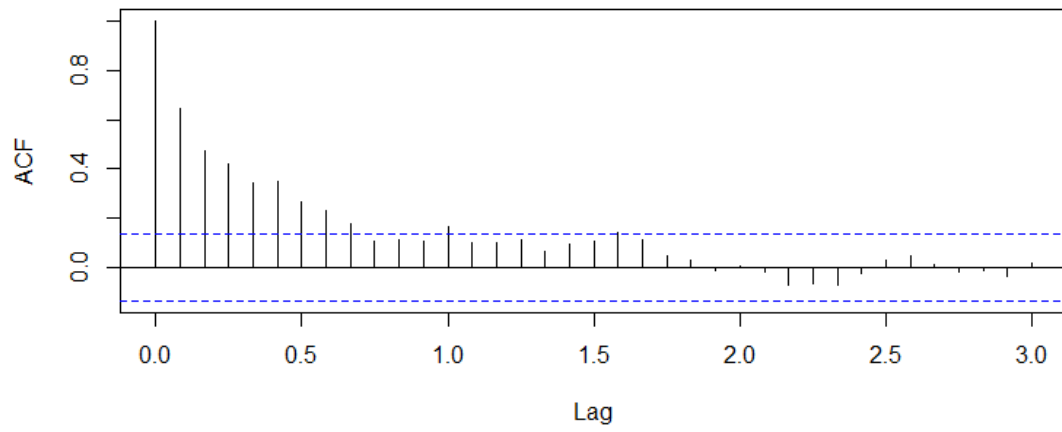
Overall, trend + season is too simple and is weak in explaining the arbitrary fluctuations in the series. While trend + cycle appear desirable, a lack of seasonality in the model still fails to replicate periodic repetitions, especially those that goes with calenders. It might be acceptable for other series, but considering our data is about energy generation, seasonality is expected to be of some significance due to the link between produciton activities and seasons.

The result is not so surprising that a model with all three components performs the best.

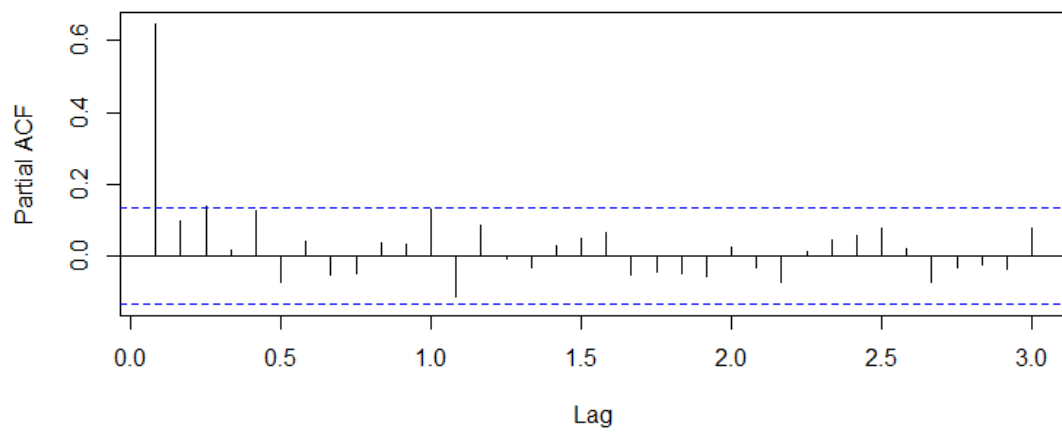
seasonal+trend1 model for Lignite



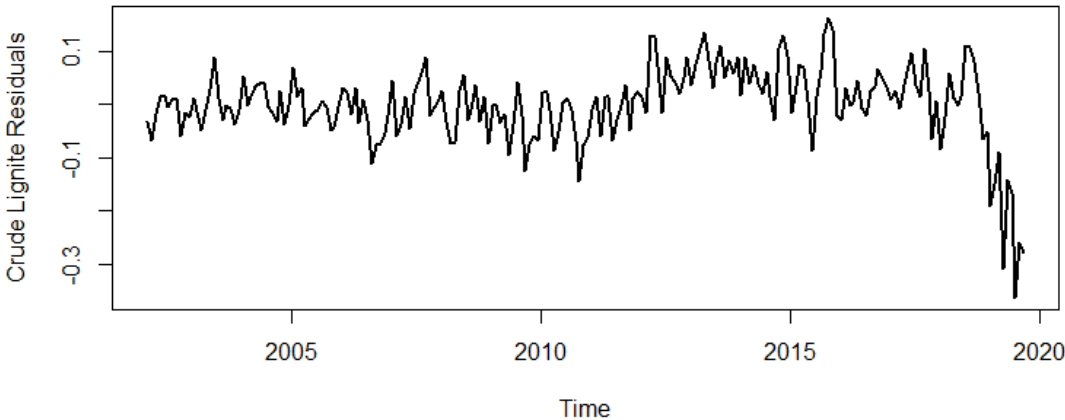
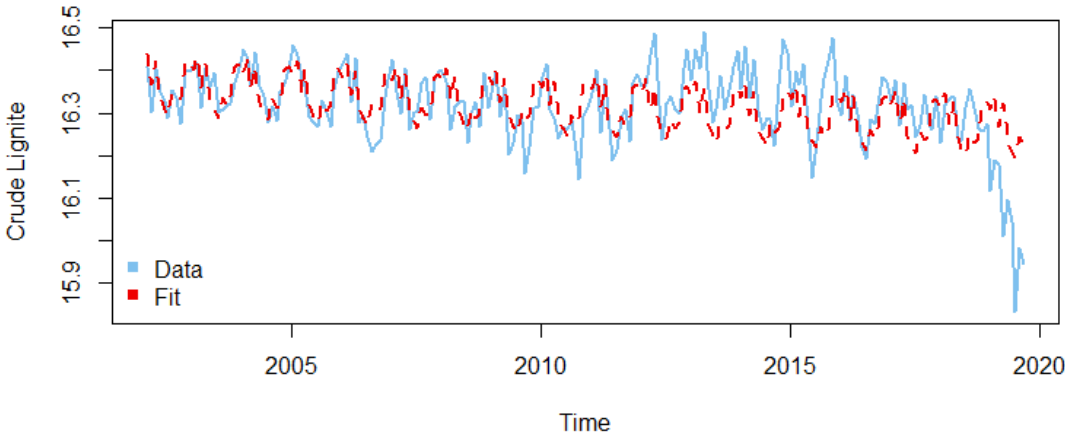
ACF-Crude Lignite Residuals



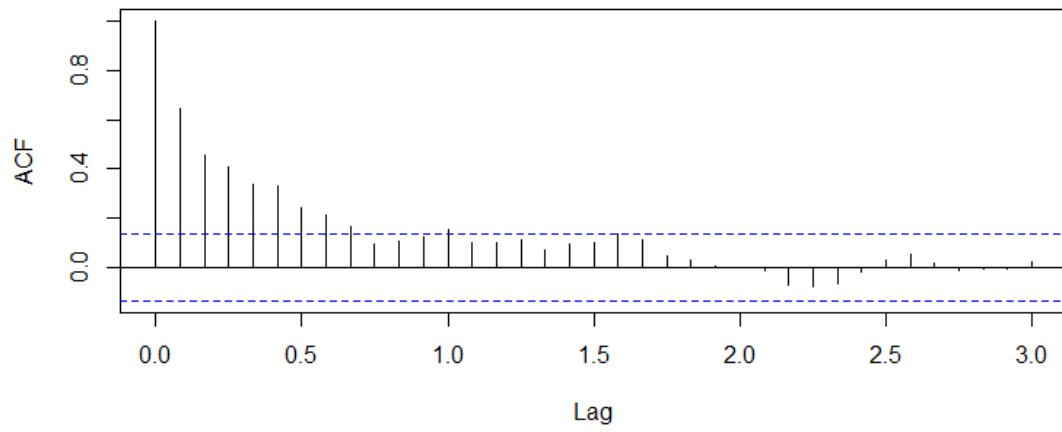
PACF-Crude Lignite Residuals



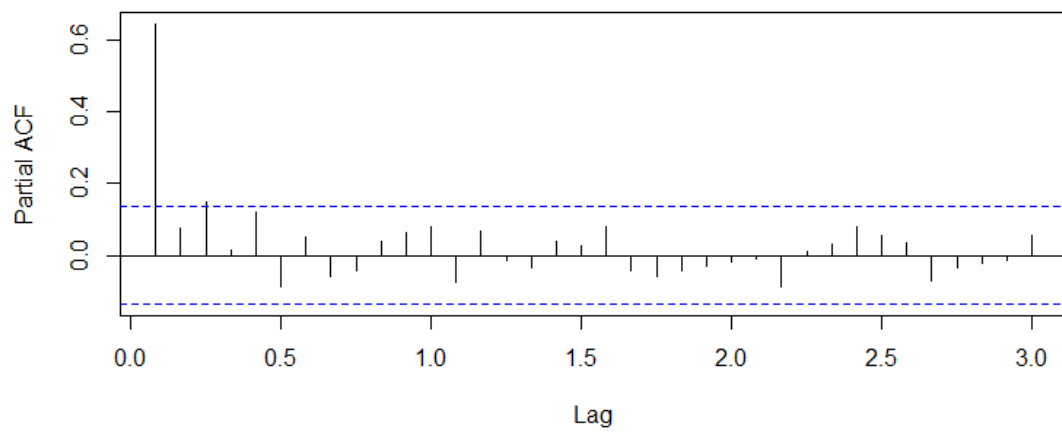
seasonal+trend2 model for Lignite



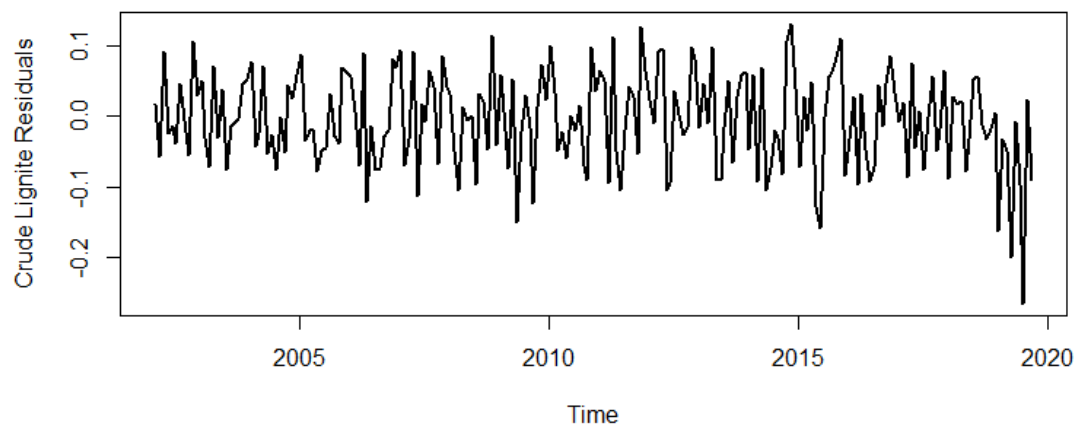
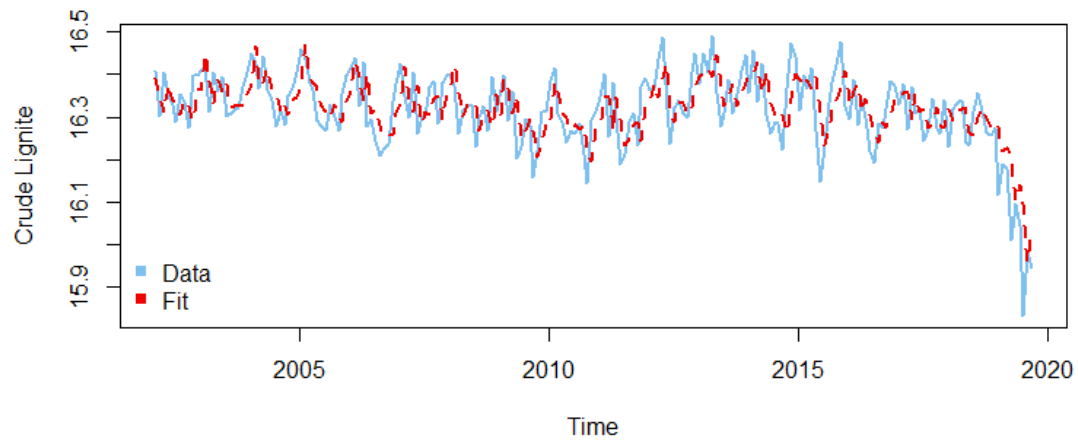
ACF-Crude Lignite Residuals



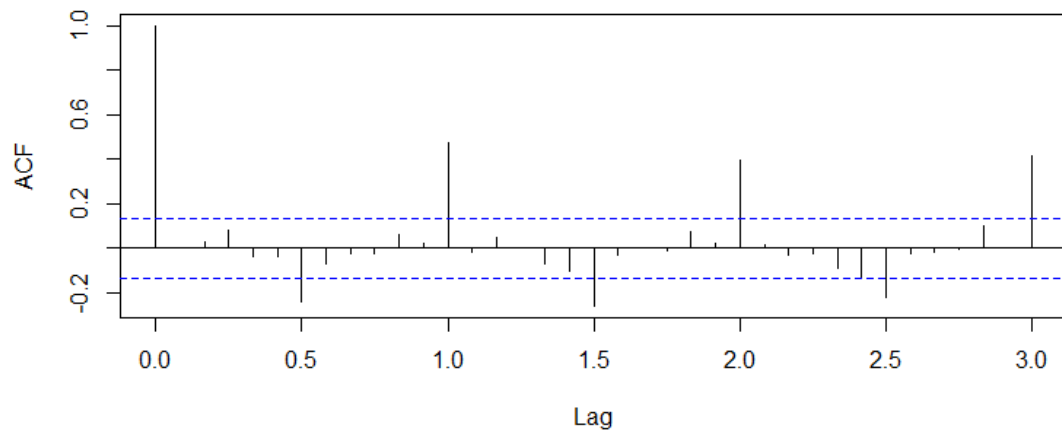
PACF-Crude Lignite Residuals



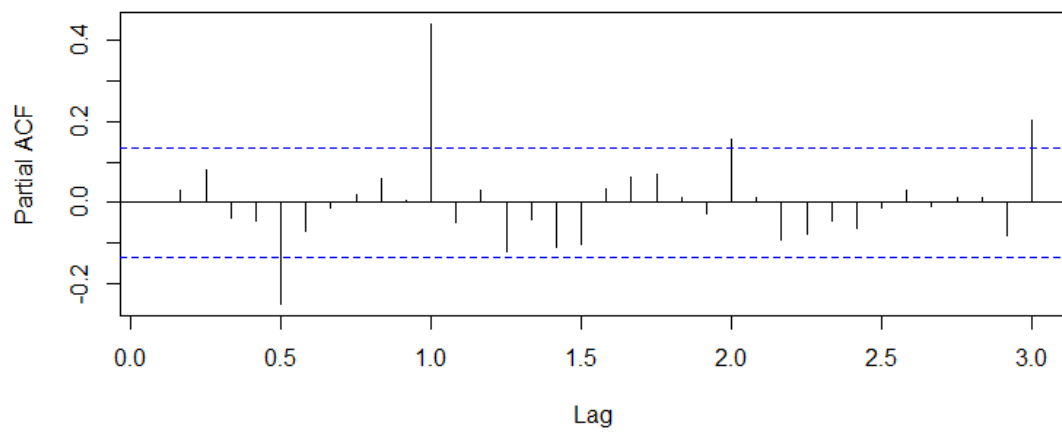
trend + cycle model

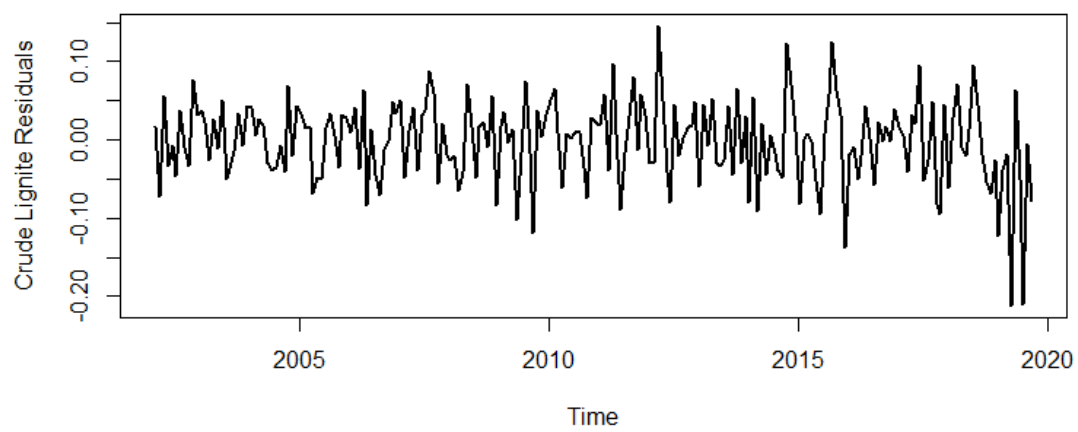
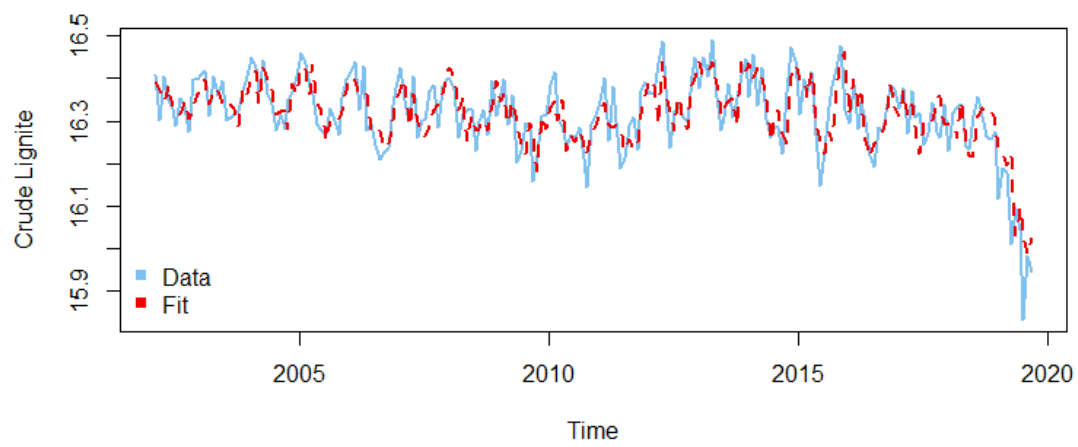


ACF-Crude Lignite Residuals

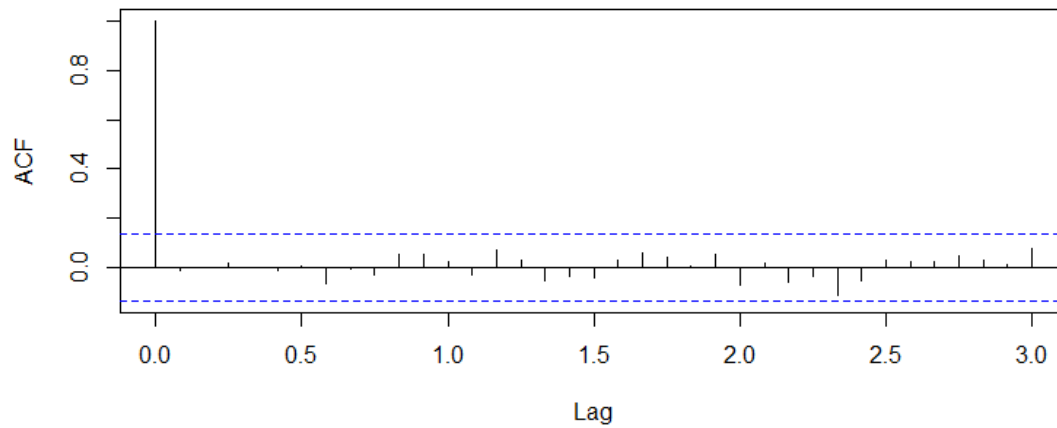


PACF-Crude Lignite Residuals

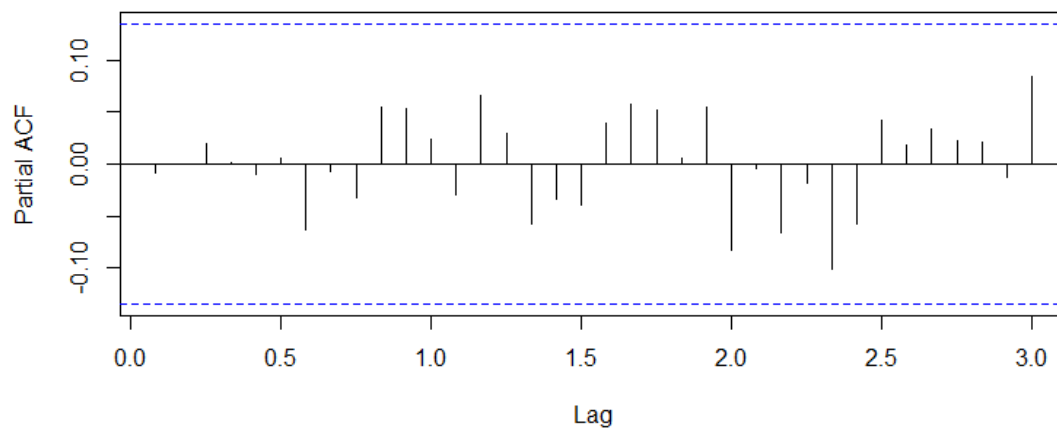




ACF-Crude Lignite Residuals



PACF-Crude Lignite Residuals

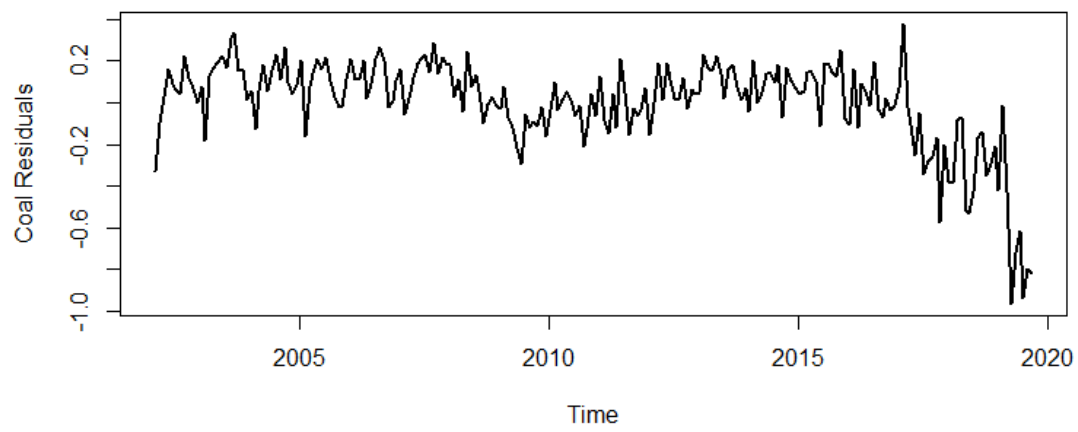
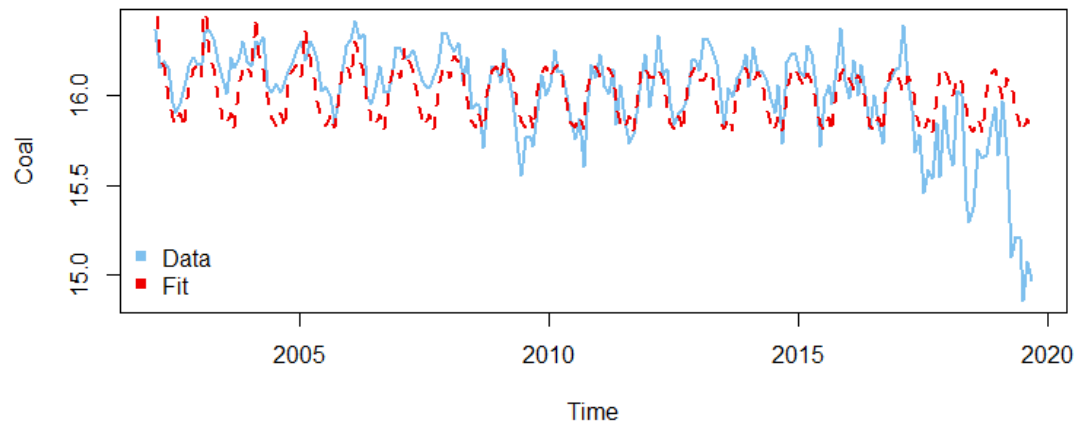


```
##      df      AIC
## n3   14 -457.9274
## n32  14 -485.3904
## n4    5 -531.3607
## n2    7 -617.8267

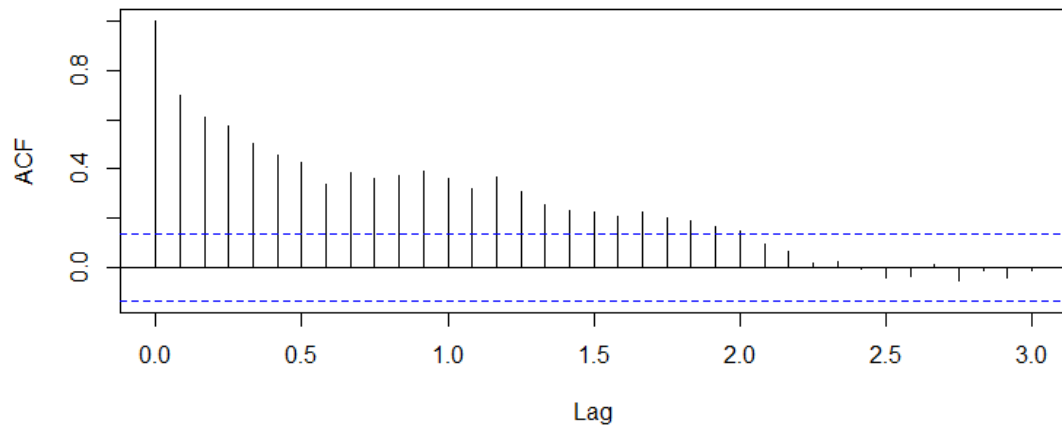
##      df      BIC
## n3   14 -410.9352
## n32  14 -438.3982
## n4    5 -514.6015
## n2    7 -594.3637

##
## Box-Pierce test
##
## data:  n2$residuals
## X-squared = 0.015775, df = 1, p-value = 0.9
```

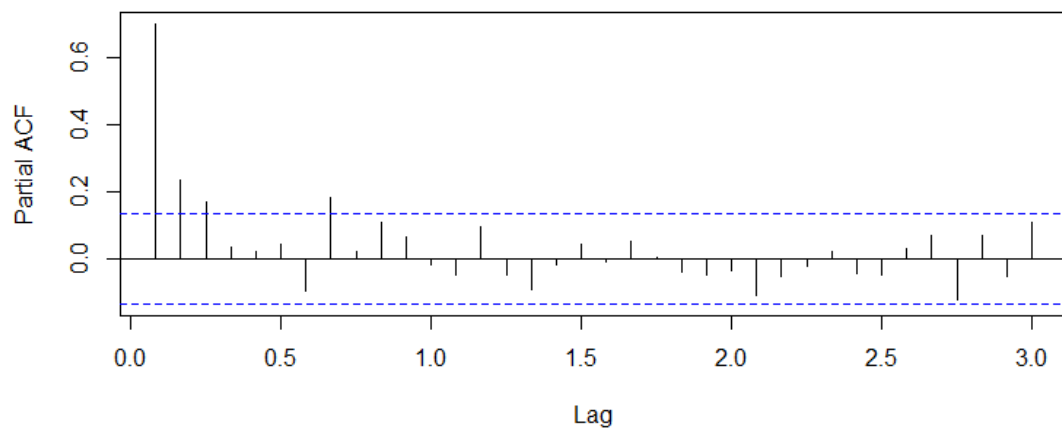

Similarly, we produce a valid model for coal. As before, the models that capture all three components tend to outperform the others.

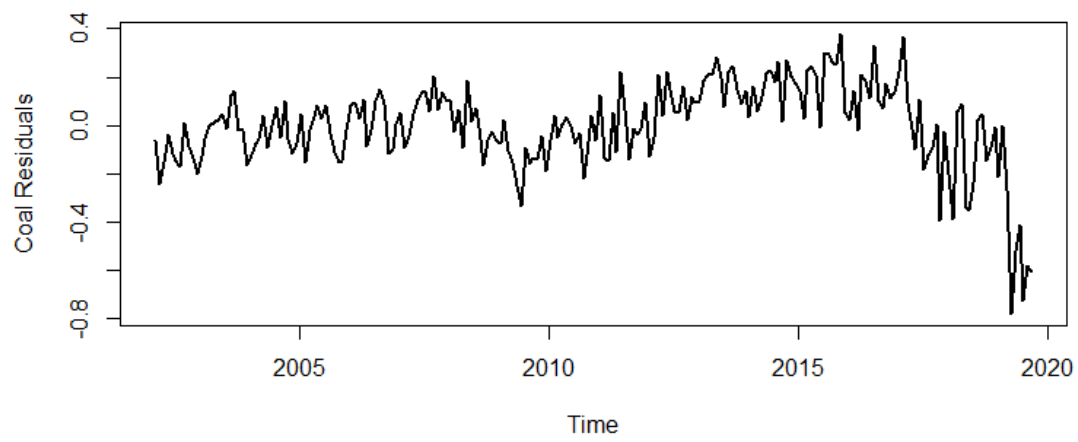
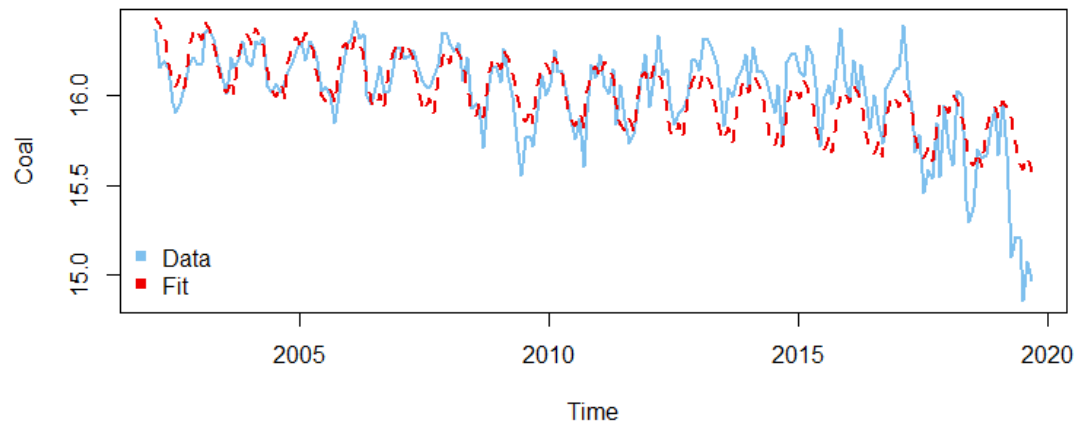


ACF-Coal Residuals

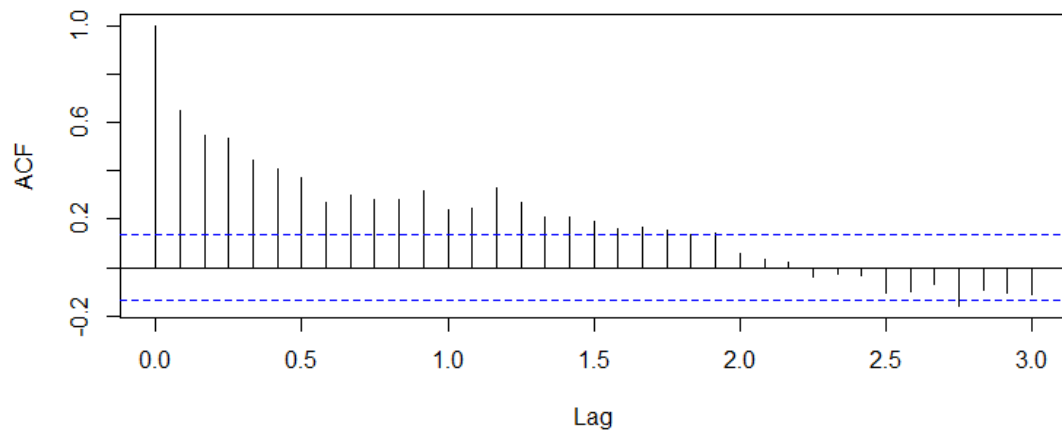


PACF-Coal Residuals

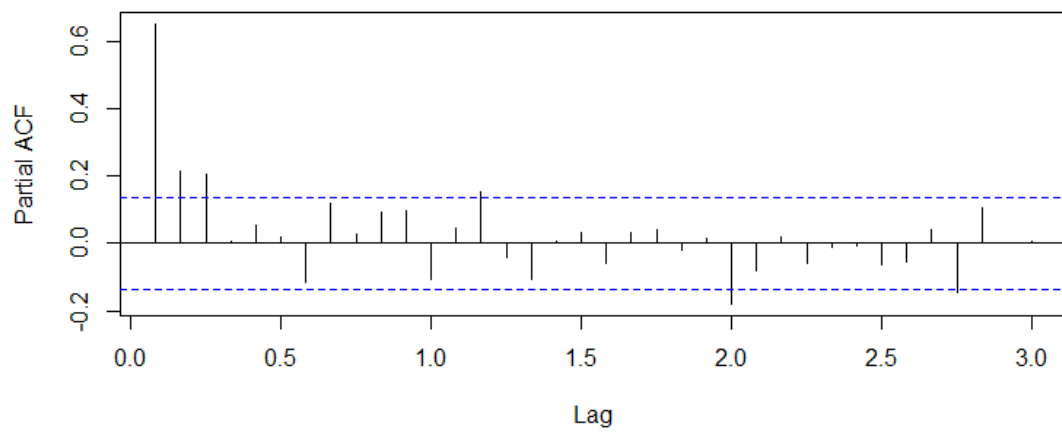


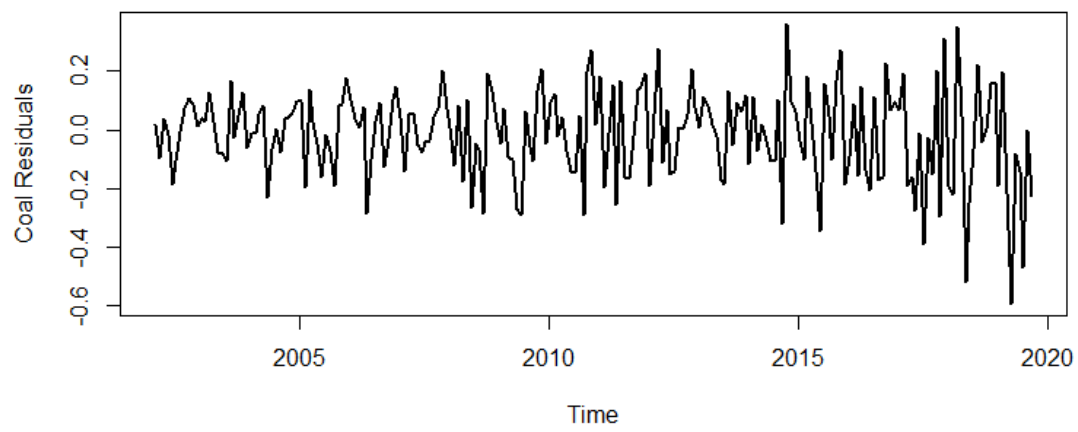
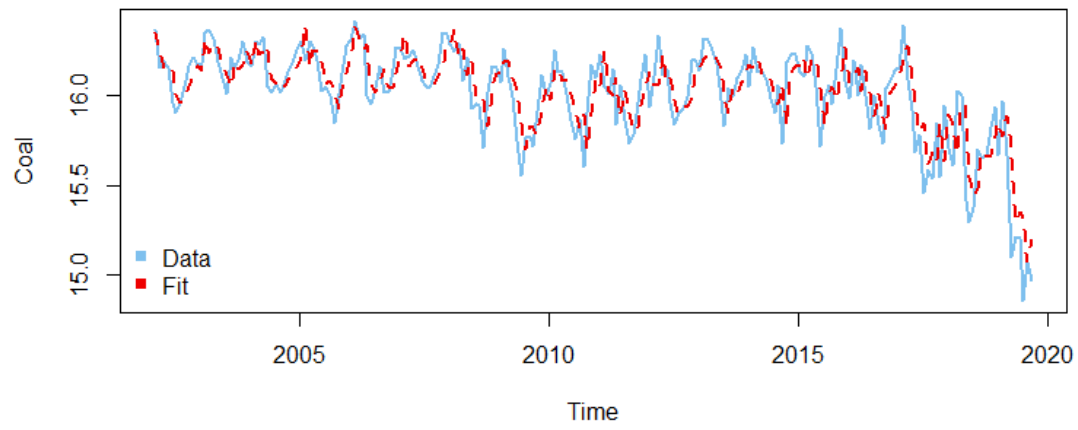


ACF-Coal Residuals

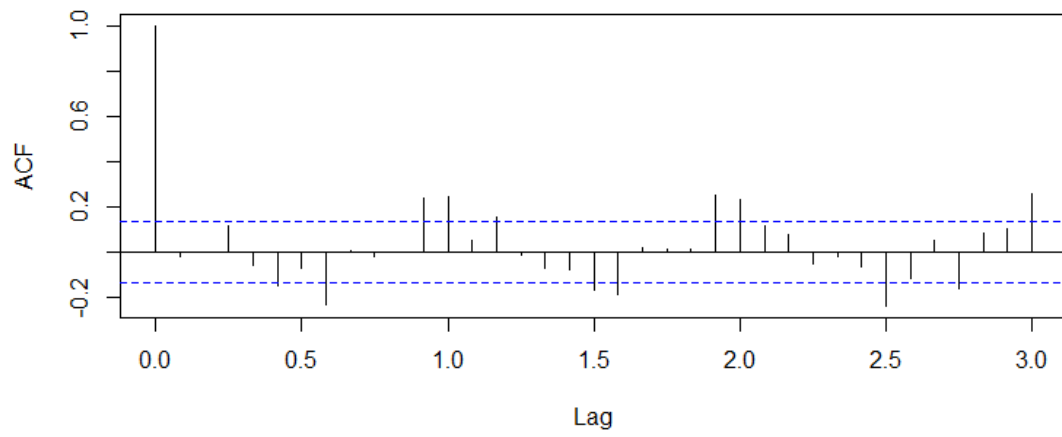


PACF-Coal Residuals

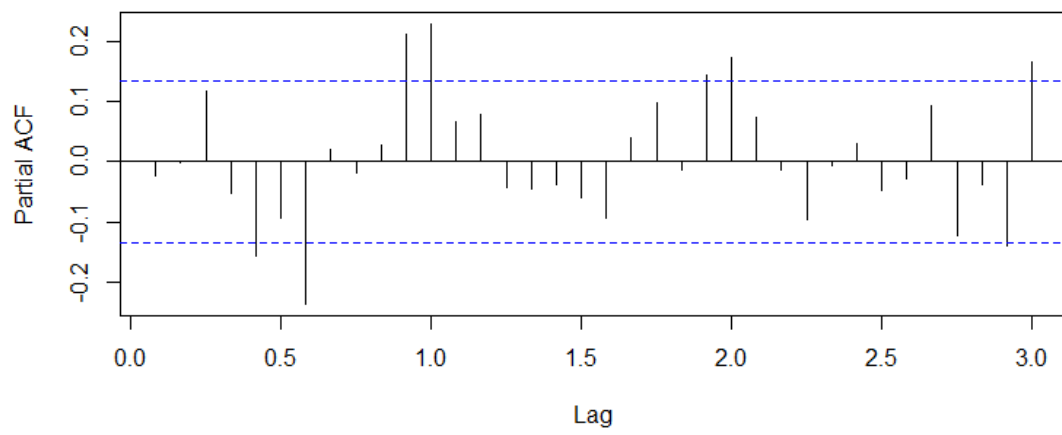


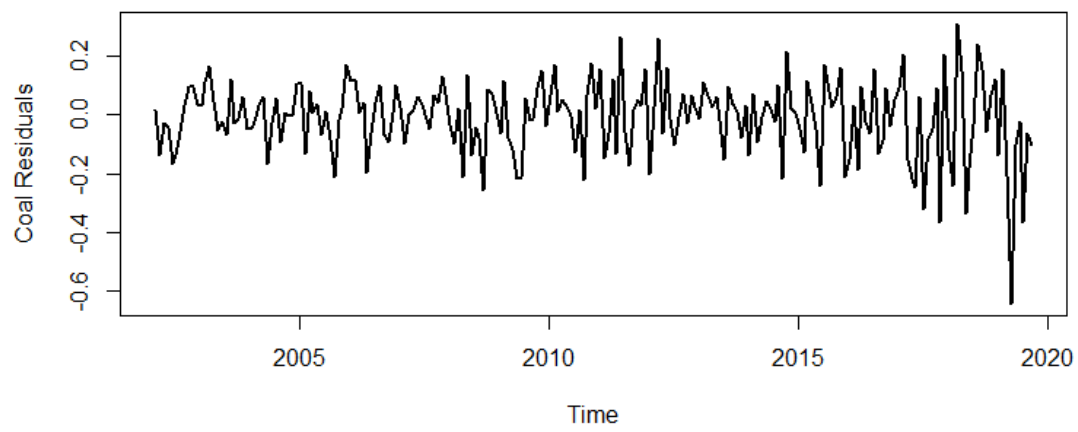
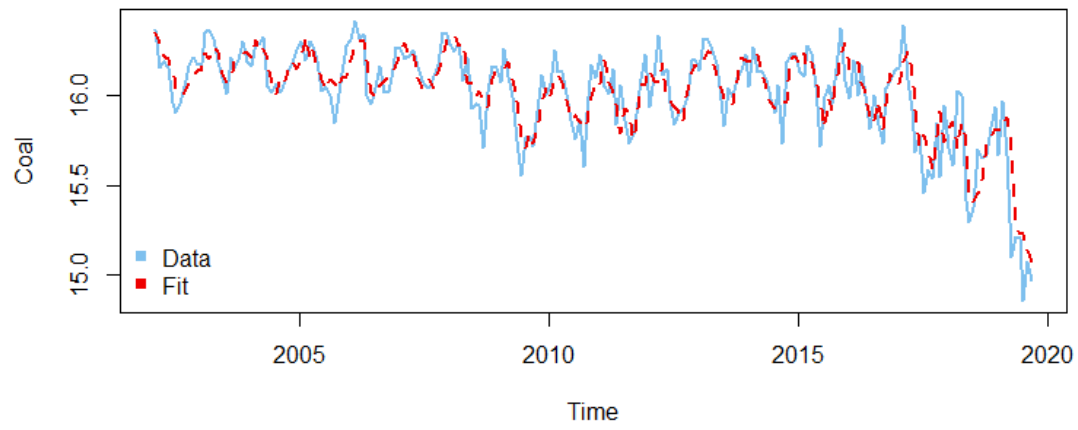


ACF-Coal Residuals

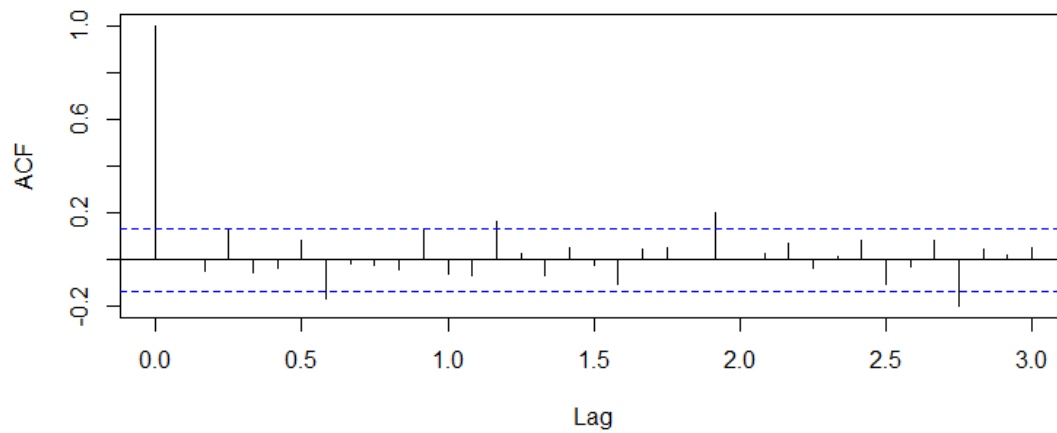


PACF-Coal Residuals

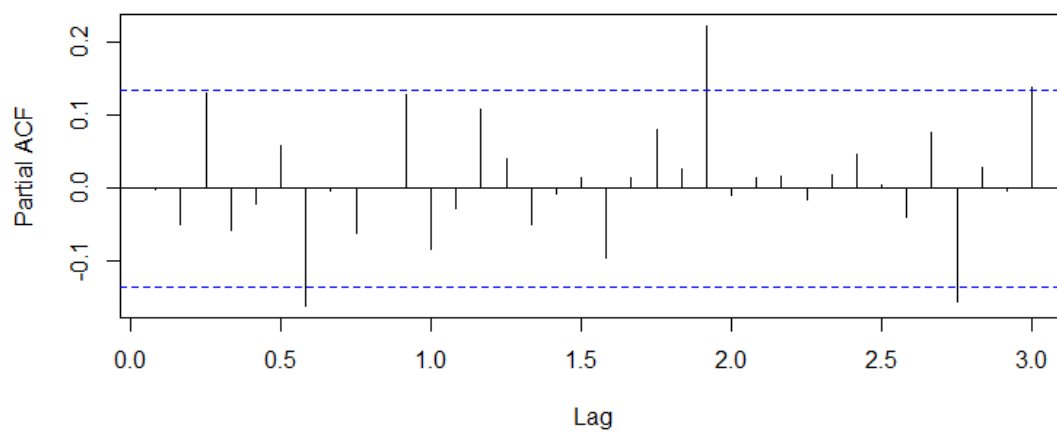




ACF-Coal Residuals



PACF-Coal Residuals



```
##      df      AIC
## c3  14  -19.39572
## c32 14 -103.82749
## c4   4 -177.37396
## c2   6 -232.10709

##      df      BIC
## c3  14   27.59649
## c32 14  -56.83528
## c4   4 -163.96653
## c2   6 -211.99594

##
## Box-Pierce test
##
## data:  c2$residuals
## X-squared = 0.00024759, df = 1, p-value = 0.9874
```

In the end, we conclude our model results as the following:

```
## Series: nt
## ARIMA(1,1,2)(1,0,1)[12] with drift
##
## Coefficients:
##          ar1          ma1          ma2          sar1          sma1          drift
##        -0.3935  -0.0138  -0.4272   0.9286  -0.6258  -0.0037
## s.e.    0.1918   0.1780   0.0887   0.0366   0.0927   0.0051
##
## sigma^2 estimated as 0.002888:  log likelihood=315.91
## AIC=-617.83  AICc=-617.27  BIC=-594.36
##
## Training set error measures:
##              ME              RMSE              MAE              MPE              MAP
E
## Training set -0.001950813  0.05284739  0.04132444  -0.01304626  0.253586
4
##              MASE              ACF1
## Training set 0.7456156 -0.008626221
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1   -0.3935192   0.1917602  -2.0521   0.04016 *
## ma1   -0.0138427   0.1779806  -0.0778   0.93801
## ma2   -0.4271666   0.0886836  -4.8167  1.459e-06 ***
## sar1    0.9285681   0.0366094  25.3642 < 2.2e-16 ***
## sma1   -0.6258217   0.0926608  -6.7539  1.439e-11 ***
## drift -0.0036651   0.0051042  -0.7180   0.47273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Series: ct
## ARIMA(1,1,1)(1,0,1)[12] with drift
##
## Coefficients:
##          ar1          ma1          sar1          sma1          drift
##          0.1812  -0.7406   0.9954  -0.9151  -0.0047
## s.e.    0.1197   0.0867   0.0094   0.0866   0.0108
##
## sigma^2 estimated as 0.01713:  log likelihood=122.05
## AIC=-232.11  AICc=-231.7  BIC=-212
##
## Training set error measures:
##              ME              RMSE              MAE              MPE              MAPE
## Training set -0.01011434  0.1290303  0.09725765  -0.07090187  0.6111034
##              MASE              ACF1
## Training set 0.6922885 -0.001080694
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error  z value Pr(>|z|)
## ar1    0.1811989  0.1197375   1.5133  0.1302
## ma1   -0.7405770  0.0866707  -8.5447 <2e-16 ***
## sar1    0.9954415  0.0093617 106.3308 <2e-16 ***
## sma1   -0.9150989  0.0865797 -10.5694 <2e-16 ***
## drift -0.0047497  0.0107617  -0.4414  0.6590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

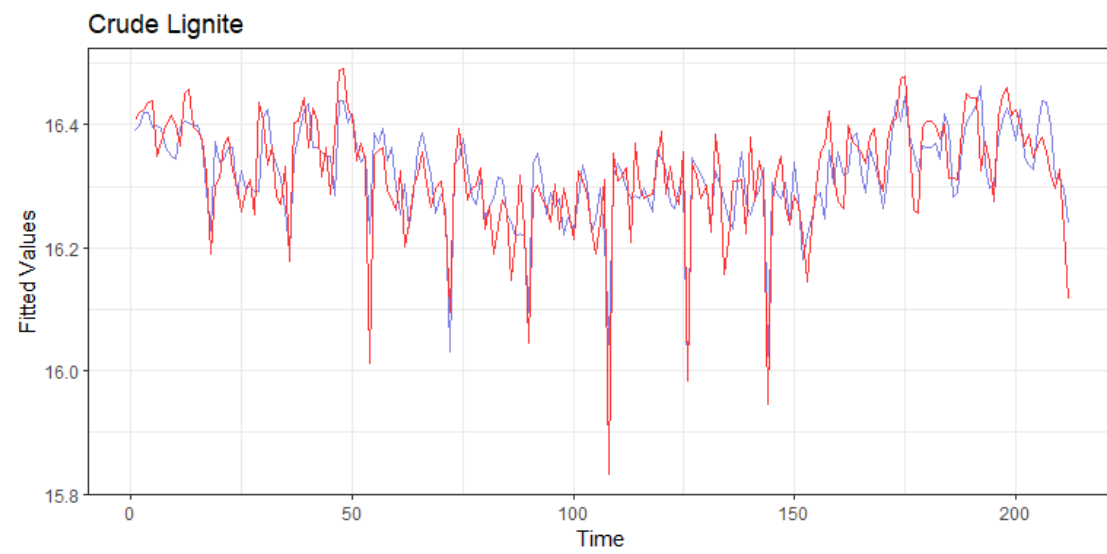
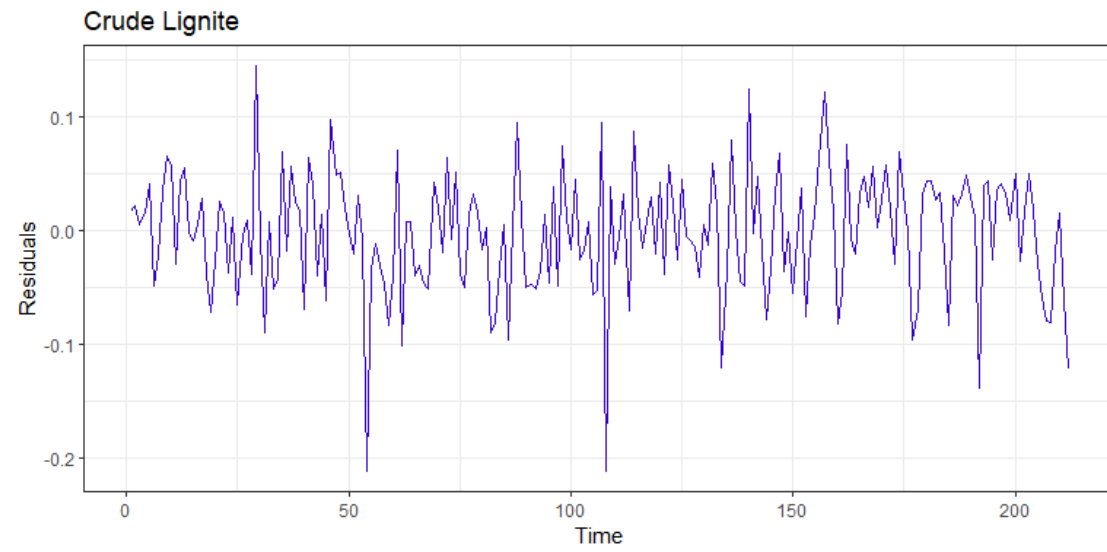
For coal lignite: After doing first difference and removing drift from the log process, we get a stationary process and using ARMA to fit the remaining pattern, detecting from acf,pacf plot.Finally, we acquire the full model composed of cycle, drift and seasonality. Among the terms in model,the drift is negative, showing the log process is decreasing slowly, which is an essential part of the model. The estimated coefficient of AR(1) is statistically significant at 0.1% level and the terms AR(2),SAR(1) SMA(1) are significant under 1% level.Despite the insignificance to some extent of MA(1), the model still turns out to be good as a whole.The RMSE is 0.0528 and BIC is -594.36, suggesting the model performs pretty desirable for good of fitness.

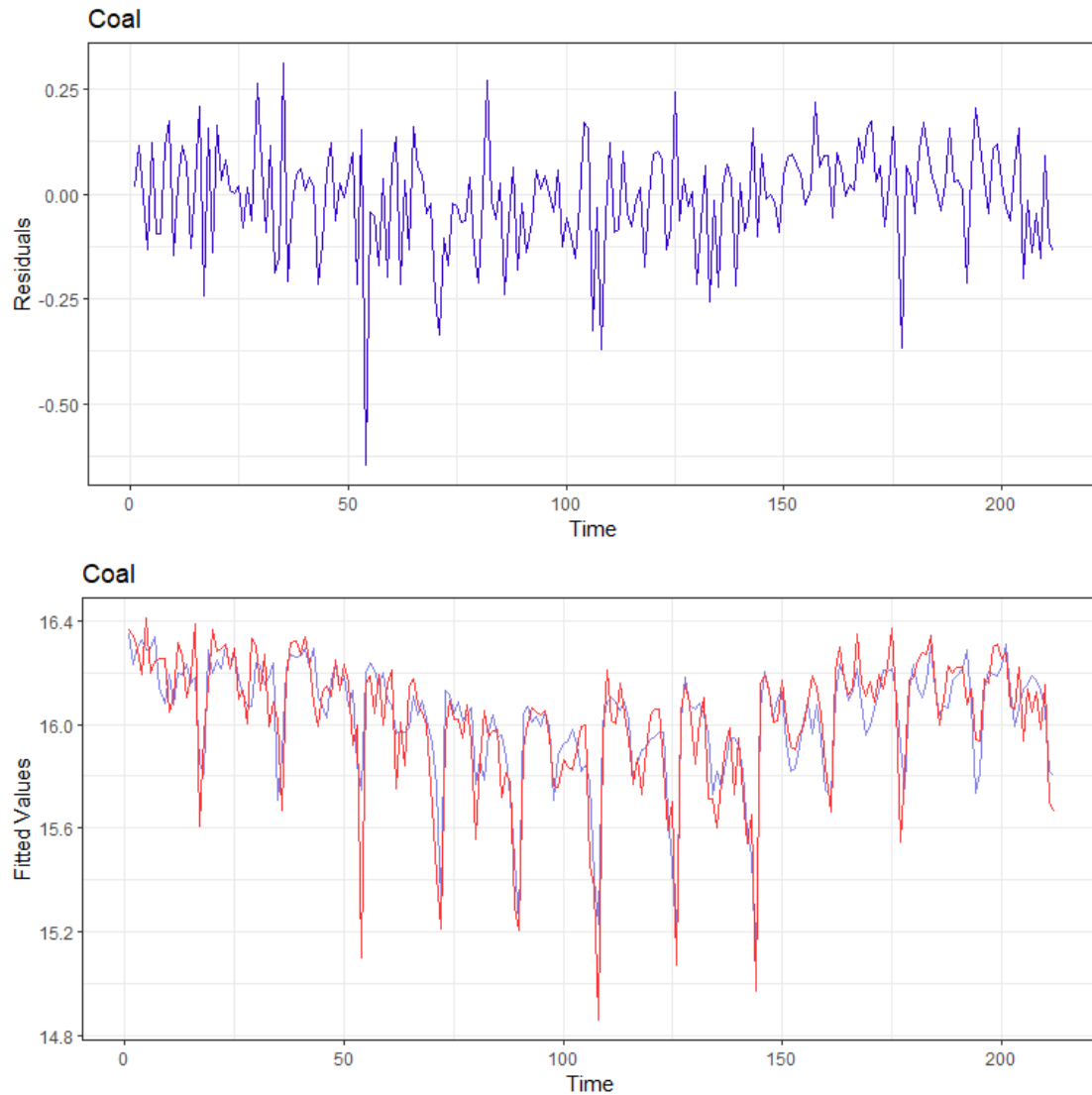
For hard coal:Similarly, the log process is also stationary after first difference, drift removal, and including seasonal autoregression.Accordingly, we utilize cycle model to fit the remaining pattern in the stationary process, leading to the full model.Specificlly, the drift of model is negative, indicating a decreasing trend of log process, playing a nonnegligible role in our model. As for seasonality, estimated coefficients of SAR(1) and SMA(2) are both statistically significant under the confidence level of 99.9%. For the cycle part,AR(1) is not statistically significant while MA(1) is significant under 0.1%. Although not all independent variables in our SARMA(1,1,1)(1,0,1)12 with drift, all the compositions still indispensable, which preform well in forecast cooperatively.RMSE is 0.1290 and BIC is -212, suggesting the model is pretty good.

4

To check the goodness of fit for our model, we check the residuals and fitted values of our models.

As can be seen from graph, our model fits are good in reflecting the zig-zags of real data for either of the series. The residuals from both series seemed to be zero mean series, the autoregressiveness are not unacceptably apparent, and the variance seemed stable most of the time. The two models seem to have some resemblance of white noise residuals, but for future work, introducing GARCH models to thoroughly account for a changing variance would be a step up to improve the performance of our models.

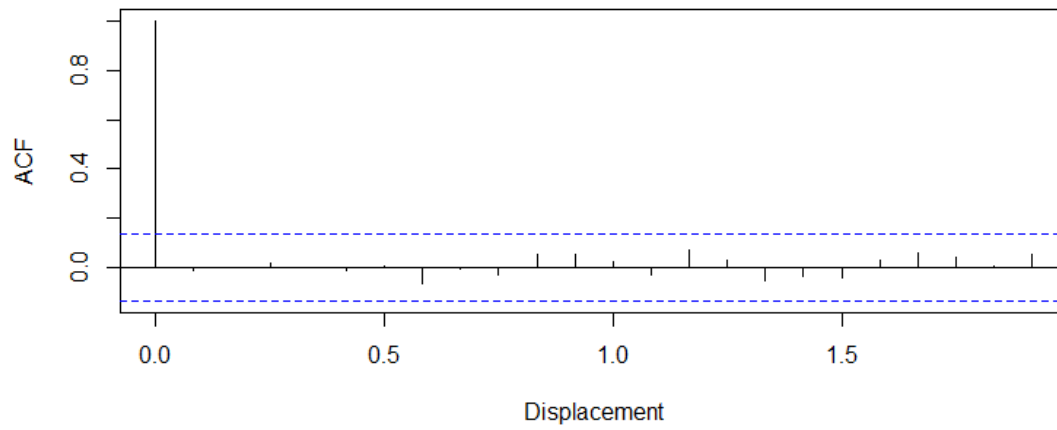




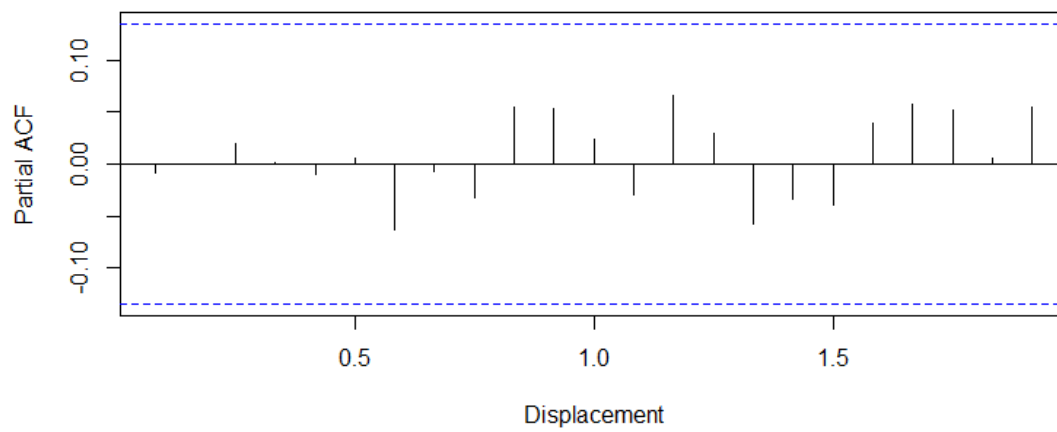
5

To validate our observation about residuals, we examine the ACF and PACF of both residual series. By and large, the two series have ACF and PACF that have virtually no spikes. No major time dependence is observed. These suggest that the models have almost explained a large proportion of the variation in the series. These well-behaved patterns suggest that we have two white noise residual series.

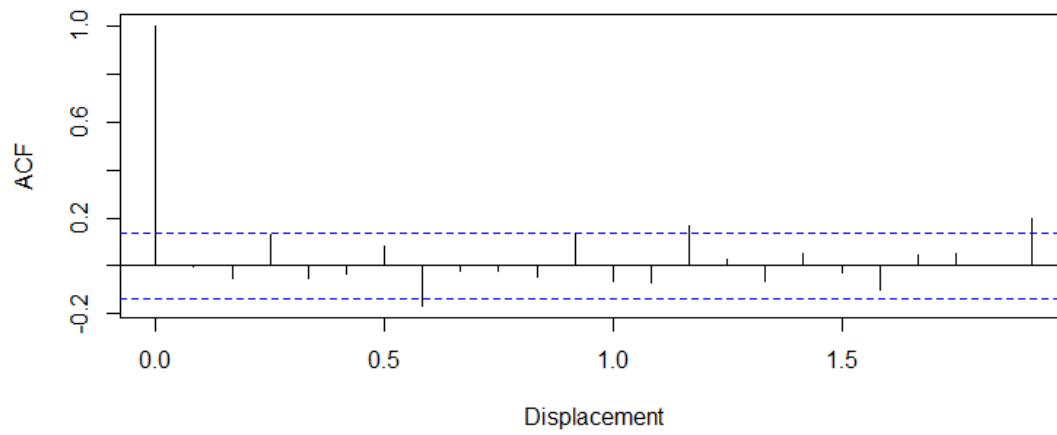
ACF - Residual from Crude-Lignite-Generated Electricity



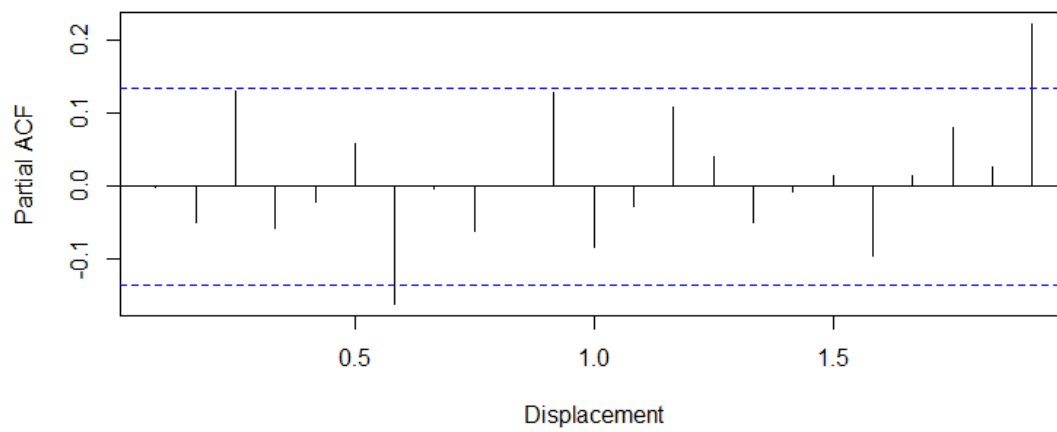
PACF - Residual from Crude-Lignite-Generated Electricity

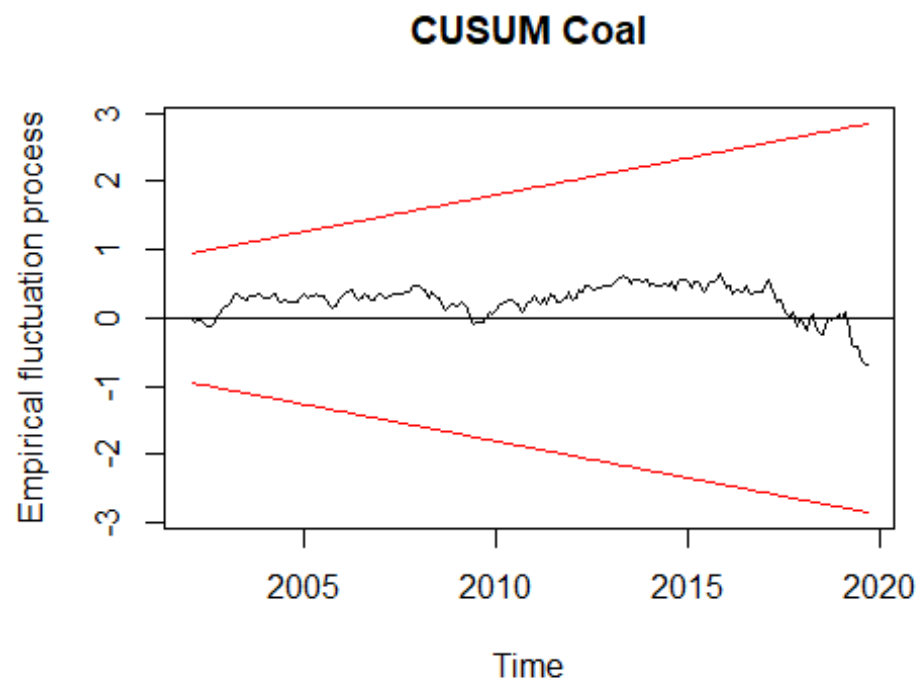
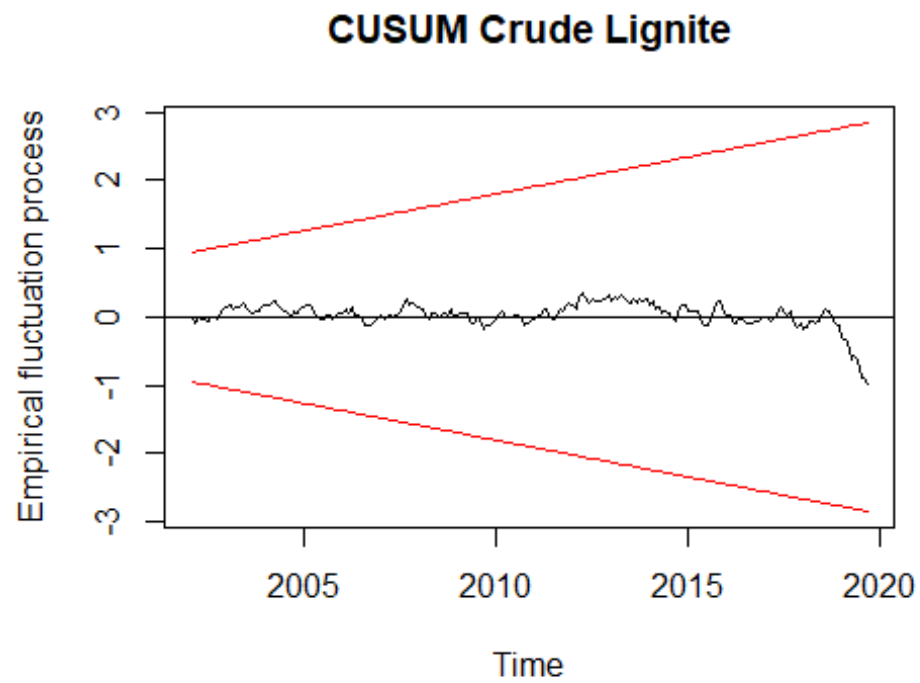


ACF - Residual from Coal-Generated Electricity



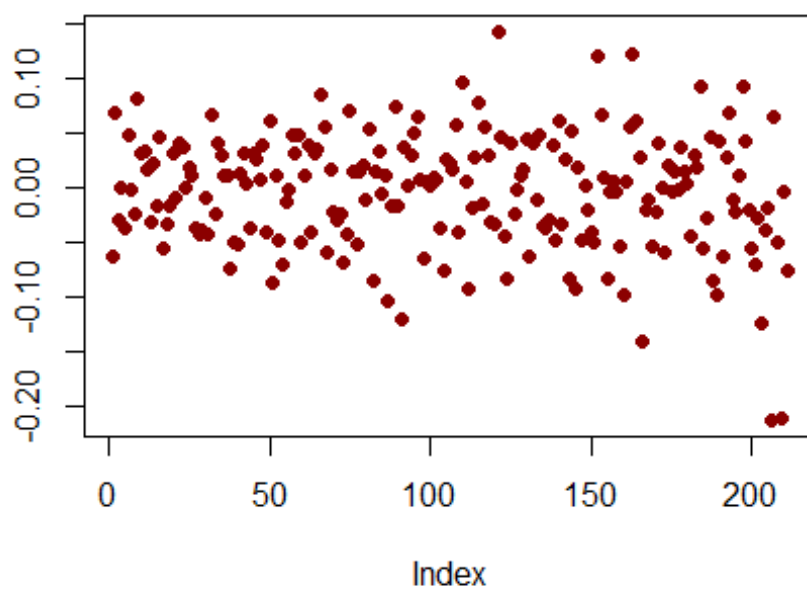
PACF - Residual from Coal-Generated Electricity



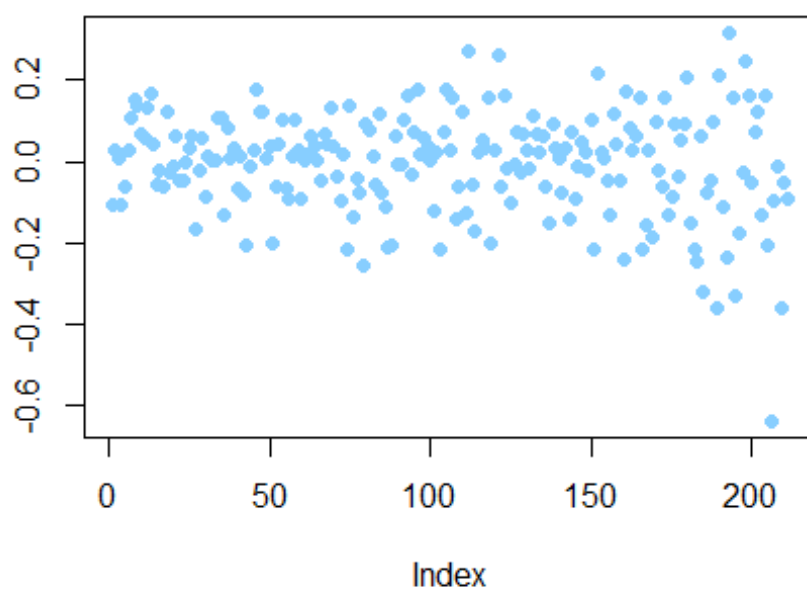


CUSUM does not go over the boundaries in either model. It fluctuates closely along the 0 baseline, Meaning that the models for the two series both works.

Recursive Residuals for Crude Lignite



Recursive Residuals for Coal



The recursive residuals seem to fluctuate around zero-mean, meaning that the model has basically accounted for most of the variation in the series, and the coefficient estimates are relatively robust, especially when doing forecast.

8

```
##      df      AIC
## n3   14 -457.9274
## n32  14 -485.3904
## n4    5 -531.3607
## n2    7 -617.8267

##      df      BIC
## n3   14 -410.9352
## n32  14 -438.3982
## n4    5 -514.6015
## n2    7 -594.3637

##      df      AIC
## n3   14 -457.9274
## n32  14 -485.3904
## n4    5 -531.3607
## n2    7 -617.8267

##      df      BIC
## n3   14 -410.9352
## n32  14 -438.3982
## n4    5 -514.6015
## n2    7 -594.3637

##                                     ME      RMSE      MAE      MPE      M
APE
## Training set -1.005462e-16 0.07691967 0.05232345 -0.002247834 0.3216
488
##                                     MASE      ACF1
## Training set 0.9440704 0.646932

##                                     ME      RMSE      MAE      MPE      MAPE
## Training set -3.352307e-17 0.0720954 0.05027427 -0.001974078 0.30894
##                                     MASE      ACF1
## Training set 0.9070972 0.6436045

##                                     ME      RMSE      MAE      MPE      MAP
E
## Training set -0.004752346 0.06685121 0.05473902 -0.03075597 0.335781
7
##                                     MASE      ACF1
## Training set 0.9876545 0.0006386532

##                                     ME      RMSE      MAE      MPE      MAP
E
## Training set -0.001950813 0.05284739 0.04132444 -0.01304626 0.253586
```

```

4
##           MASE           ACF1
## Training set 0.7456156 -0.008626221

##           ME           RMSE           MAE           MPE           MAPE
## Training set -4.191515e-17 0.2163796 0.1521181 -0.01893062 0.9605385
##           MASE           ACF1
## Training set 1.08279 0.7004715

##           ME           RMSE           MAE           MPE           MAPE
## Training set 7.537302e-17 0.1773106 0.1304212 -0.01274895 0.8216686
##           MASE           ACF1
## Training set 0.9283496 0.6504694

##           ME           RMSE           MAE           MPE           MAPE
## MASE
## Training set -0.01509695 0.1553851 0.1222611 -0.104115 0.7681894 0.8
702653
##           ACF1
## Training set -0.02195118

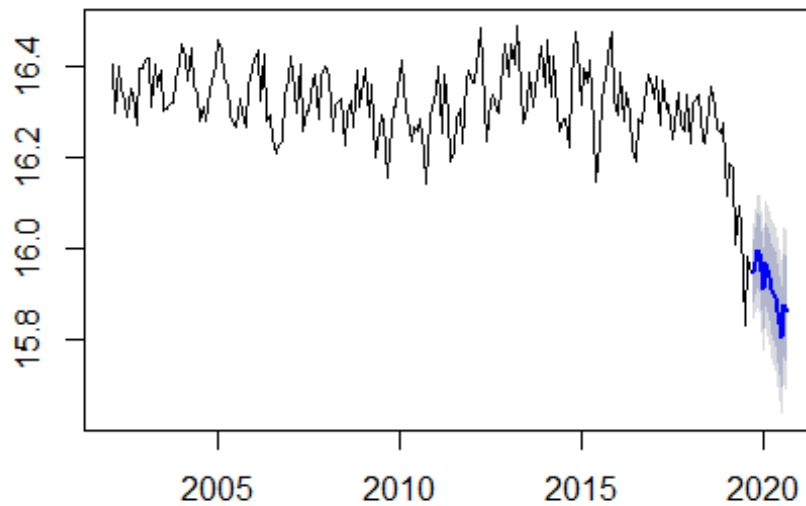
##           ME           RMSE           MAE           MPE           MAPE
## Training set -0.01011434 0.1290303 0.09725765 -0.07090187 0.6111034
##           MASE           ACF1
## Training set 0.6922885 -0.001080694

```

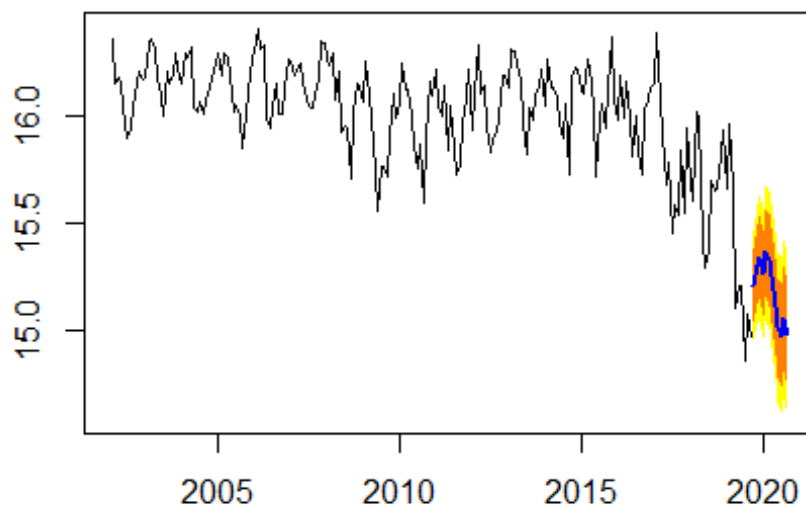
Our selected models has the best BIC and AIC among its siblings.

The error metrics are also demonstrating well-behaved properties of our model. The error is relatively small, regardless of standards utilized.

Forecasting for Crude Lignite



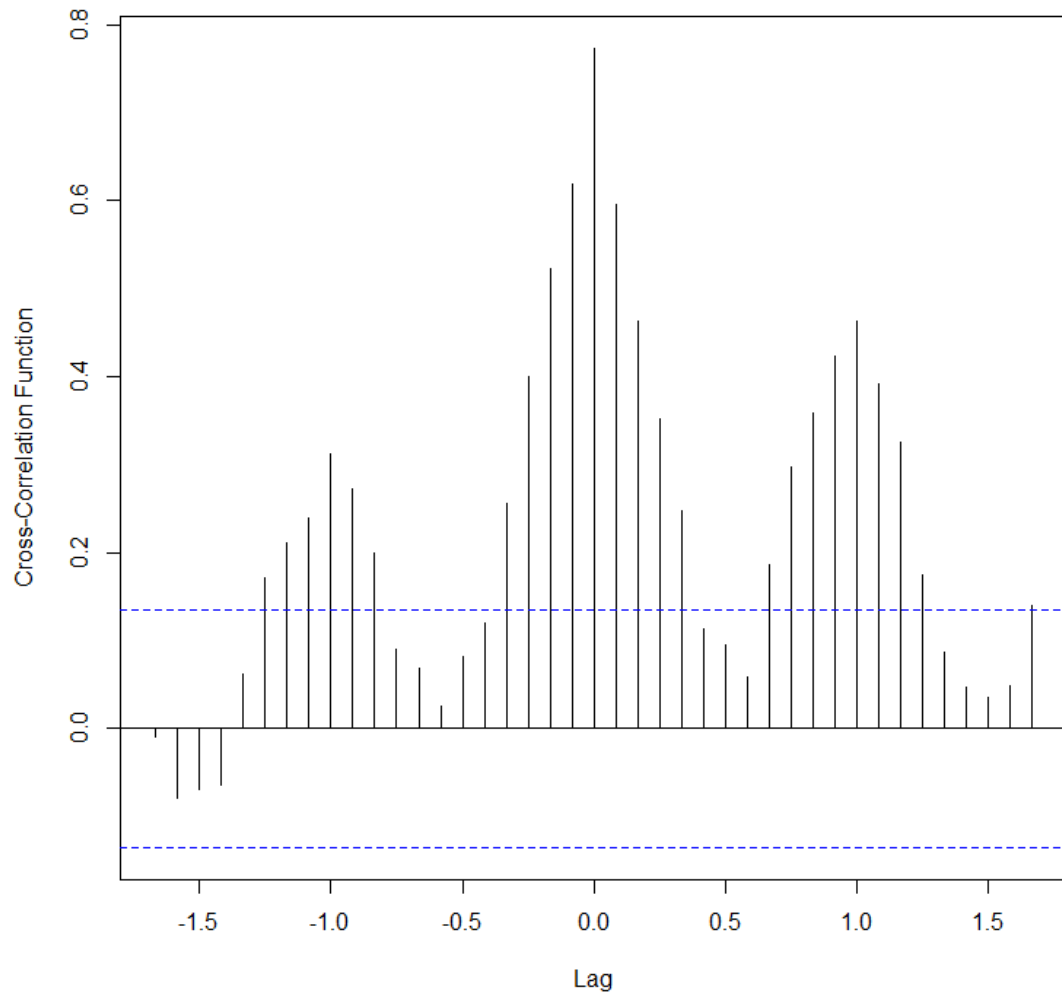
Forecasting for Coal



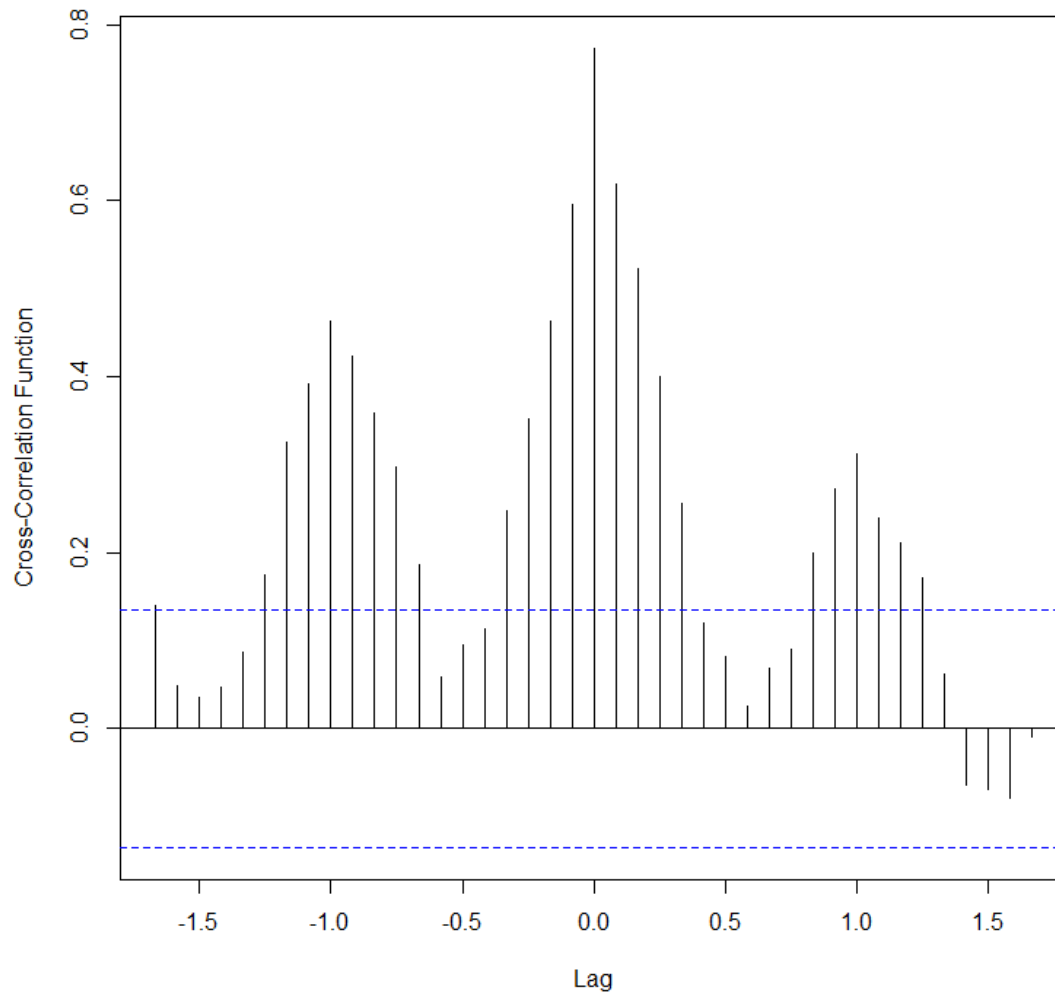
The point forecast from both models, as well as their interval estimates, are reflecting the time-variant nature of both series. The line portrays the point prediction of the

process, the narrower shade suggests the confidence interval and the wider shade indicates predict interval of the proccess, constitute a comprehensive picture of time series forecast.

Crude Lignite and Coal CCF



Coal and Crude Lignite CCF



```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      10      1      1      10
##
## $criteria
##              1              2              3              4
## AIC(n) -9.377897e+00 -9.371342e+00 -9.3383974426 -9.321790e+00
## HQ(n)  -9.338139e+00 -9.305078e+00 -9.2456279607 -9.202515e+00
## SC(n)  -9.279632e+00 -9.207566e+00 -9.1091115626 -9.026993e+00
## FPE(n)  8.457321e-05  8.513082e-05  0.0000879852  8.946424e-05
##              5              6              7              8
## AIC(n) -9.3392056864 -9.357746e+00 -9.380438e+00 -9.541776e+00
## HQ(n)  -9.1934250720 -9.185460e+00 -9.181646e+00 -9.316478e+00
## SC(n)  -8.9788993035 -8.931930e+00 -8.889111e+00 -8.984939e+00
## FPE(n)  0.0000879282  8.632512e-05  8.440451e-05  7.184658e-05
##              9              10
## AIC(n) -9.554560e+00 -9.588284e+00
```



```

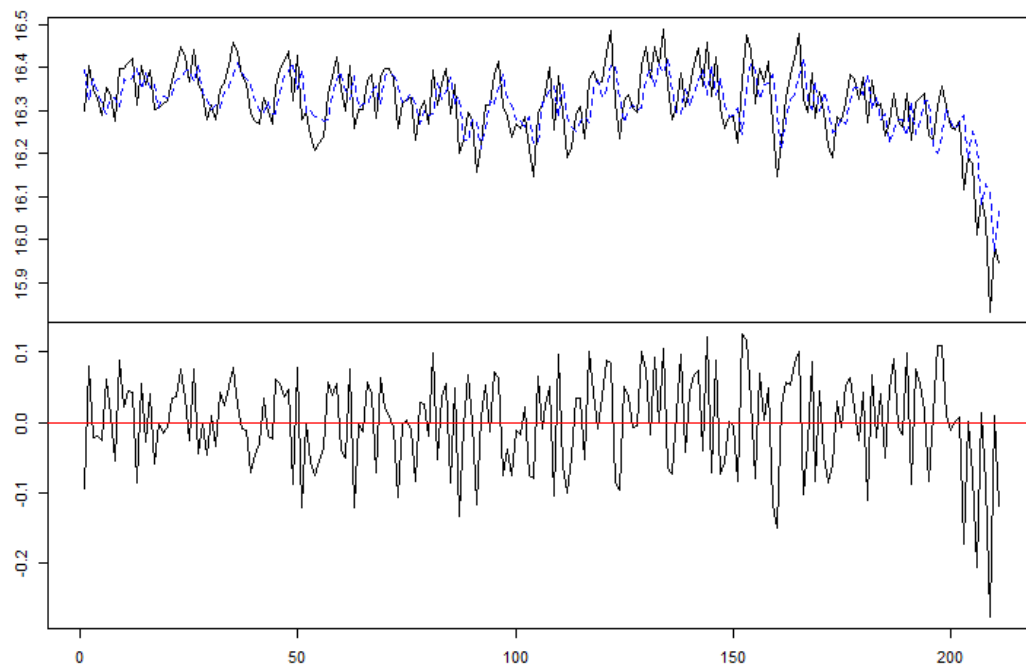
## HQ(n)  -9.302757e+00 -9.309976e+00
## SC(n)  -8.932212e+00 -8.900426e+00
## FPE(n)  7.095644e-05  6.863032e-05

##
## VAR Estimation Results:
## =====
## Endogenous variables: nt, ct
## Deterministic variables: const
## Sample size: 211
## Log Likelihood: 400.517
## Roots of the characteristic polynomial:
## 0.8231 0.355
## Call:
## VAR(y = co, p = 1)
##
##
## Estimation results for equation nt:
## =====
## nt = nt.l1 + ct.l1 + const
##
##      Estimate Std. Error t value Pr(>|t|)
## nt.l1  0.45803    0.07991   5.732 3.46e-08 ***
## ct.l1  0.10355    0.02764   3.746 0.000233 ***
## const  7.18391    1.01294   7.092 2.03e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.06736 on 208 degrees of freedom
## Multiple R-Squared: 0.4689, Adjusted R-squared: 0.4638
## F-statistic: 91.83 on 2 and 208 DF, p-value: < 2.2e-16
##
##
## Estimation results for equation ct:
## =====
## ct = nt.l1 + ct.l1 + const
##
##      Estimate Std. Error t value Pr(>|t|)
## nt.l1  0.36340    0.19290   1.884  0.061 .
## ct.l1  0.72001    0.06673  10.790 <2e-16 ***
## const -1.45195    2.44526  -0.594  0.553
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1626 on 208 degrees of freedom
## Multiple R-Squared: 0.6266, Adjusted R-squared: 0.623
## F-statistic: 174.5 on 2 and 208 DF, p-value: < 2.2e-16
##

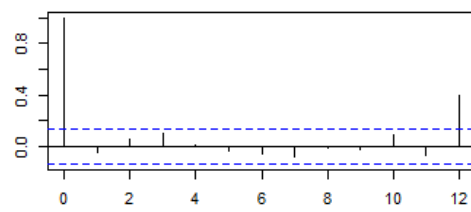
```

```
##  
##  
## Covariance matrix of residuals:  
##          nt          ct  
## nt 0.004537 0.006383  
## ct 0.006383 0.026439  
##  
## Correlation matrix of residuals:  
##          nt          ct  
## nt 1.0000 0.5828  
## ct 0.5828 1.0000
```

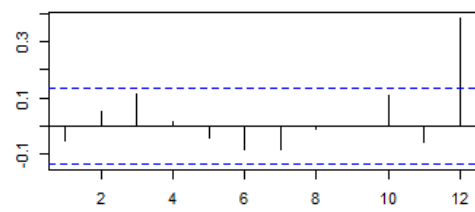
Diagram of fit and residuals for nt

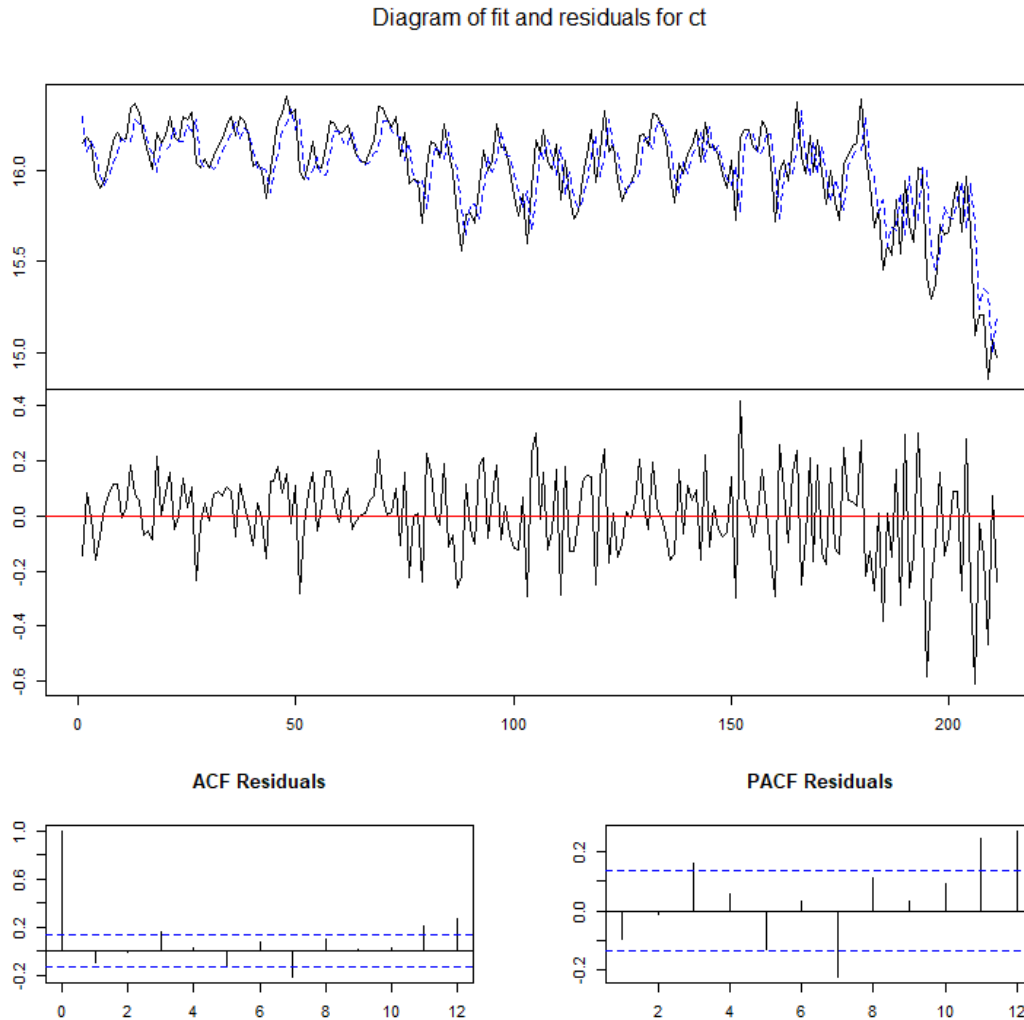


ACF Residuals



PACF Residuals





According to the SARMA model we constructed for Lignite and Hard coal above, they have auto regression process themselves. Then from ccf plot, we can see they are significantly correlated with the other lagged by 0-2 months, which provides a strong support to VAR model.

Then we try to build VAR(1) and VAR(2) model and get better adjust R square and statistical significance. By comparing between AIC, BIC of different models, VAR(1) is best. For Var(1), $nt = nt.l1 + ct.l1 + \text{const}$ the auto regression lag 1 of hard coal, AR(1) crude lignite and constant are all statistically significant, with R-squared 0.4638. For the model, $ct = nt.l1 + ct.l1 + \text{const}$ only the auto regression lag 1 of hard coal itself is statistically significant, with adjust R-square 0.623. The model seems to capture the autoregressive nature of both series, and, above that, have shown some interdependence between the two. The causality is not yet clear, simply judging from CCF plots.

11

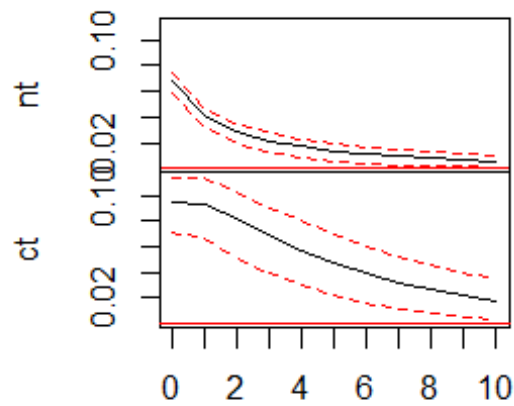
```
##
## Impulse response coefficients
## $nt
##          nt          ct
## [1,] 0.067356377 0.09476722
## [2,] 0.040664086 0.09271008
## [3,] 0.028225158 0.08152908
## [4,] 0.021369989 0.06895842
## [5,] 0.016928468 0.05741632
## [6,] 0.013698977 0.04749189
## [7,] 0.011192138 0.03917265
## [8,] 0.009182507 0.03227176
## [9,] 0.007547476 0.02657278
## [10,] 0.006208476 0.02187532
## [11,] 0.005108770 0.01800652
##
## $ct
##          nt          ct
## [1,] 0.000000000 0.13212808
## [2,] 0.013681262 0.09513317
## [3,] 0.016117062 0.07346828
## [4,] 0.014989435 0.05875456
## [5,] 0.012949407 0.04775080
## [6,] 0.010875620 0.03908668
## [7,] 0.009028629 0.03209484
## [8,] 0.007458675 0.02638948
## [9,] 0.006148822 0.02171107
## [10,] 0.005064439 0.01786658
## [11,] 0.004169678 0.01470446
##
##
## Lower Band, CI= 0.95
## $nt
##          nt          ct
## [1,] 0.060629714 0.071589968
## [2,] 0.029682893 0.066764980
## [3,] 0.017853357 0.051906473
## [4,] 0.011862241 0.039763186
## [5,] 0.008304598 0.030377130
## [6,] 0.006061168 0.023182221
## [7,] 0.004525497 0.017546908
## [8,] 0.003415646 0.013145530
## [9,] 0.002551139 0.009845738
## [10,] 0.001908554 0.007373506
## [11,] 0.001428924 0.005521808
##
## $ct
##          nt          ct
```

```

## [1,] 0.000000000 0.114029918
## [2,] 0.007769651 0.071858482
## [3,] 0.009259827 0.048210658
## [4,] 0.008424227 0.036101622
## [5,] 0.006709278 0.027800940
## [6,] 0.005140552 0.021227789
## [7,] 0.003896753 0.015952829
## [8,] 0.002941983 0.012195735
## [9,] 0.002218472 0.009163338
## [10,] 0.001694616 0.006897660
## [11,] 0.001292220 0.005163503
##
##
## Upper Band, CI= 0.95
## $nt
##          nt          ct
## [1,] 0.07343376 0.11475186
## [2,] 0.04503273 0.10906935
## [3,] 0.03333828 0.10086631
## [4,] 0.02687436 0.08739669
## [5,] 0.02295035 0.07761306
## [6,] 0.01985054 0.06854755
## [7,] 0.01737241 0.06043500
## [8,] 0.01527998 0.05325390
## [9,] 0.01340592 0.04691947
## [10,] 0.01176444 0.04133807
## [11,] 0.01032490 0.03631719
##
## $ct
##          nt          ct
## [1,] 0.000000000 0.14444843
## [2,] 0.019014599 0.10806485
## [3,] 0.020168835 0.08777874
## [4,] 0.019116127 0.07365780
## [5,] 0.017412483 0.06443869
## [6,] 0.015231270 0.05657272
## [7,] 0.013230923 0.04936928
## [8,] 0.011599584 0.04314250
## [9,] 0.010154322 0.03771688
## [10,] 0.008913109 0.03297881
## [11,] 0.007875017 0.02883880

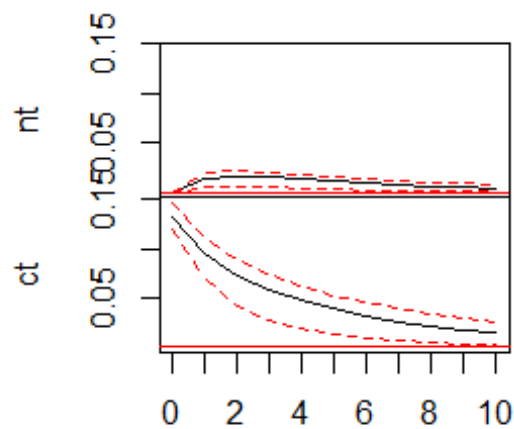
```

Orthogonal Impulse Response from nt



95 % Bootstrap CI, 100 runs

Orthogonal Impulse Response from ct



95 % Bootstrap CI, 100 runs

We can see from the Impulse Response function that both variables are sensitive to their own lagged terms. There is a jump on the value than the influence will decay to zero. As for cross effects, while lagged coal has an influence on crude lignite, the opposite does not hold. If we exert a positive shock on Lignite, coal will jump and

converge to 0 gradually. If we increase value of coal, Lignite will increase a little and then decay.

This asymmetry is pointing to the fact that coal can be used to predict the sequential series of crude lignite.

12

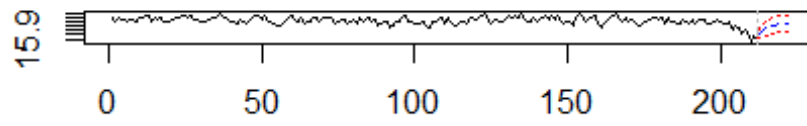
```
## Granger causality test
##
## Model 1: ct ~ Lags(ct, 1:1) + Lags(nt, 1:1)
## Model 2: ct ~ Lags(ct, 1:1)
##   Res.Df Df       F    Pr(>F)
## 1      208
## 2      209 -1 3.5487 0.06099 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Granger causality test
##
## Model 1: nt ~ Lags(nt, 1:1) + Lags(ct, 1:1)
## Model 2: nt ~ Lags(nt, 1:1)
##   Res.Df Df       F    Pr(>F)
## 1      208
## 2      209 -1 14.031 0.0002328 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

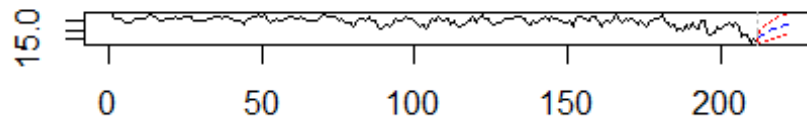
In coherence with the previous IRF results, Granger Causality test result suggests that coal is influencing crude lignite, while the other way around does not hold.

Our intuition is that, crude lignite is a cheaper, but environmentally unfriendly source of energy. Under the pressure from green parties and relevant environment protectionist groups, the usage of crude lignite will face relatively large pressure, making it a supplementary to coal. It is natural, therefore, to suspect that to supplement for shortage of electricity, large amount of coal-generated electricity will be followed by a supplementary amount of crude-lignite-generated electricity. The same happens when there is a drop in coal-generated electricity, which will be followed by a drop in crude-lignite-generated electricity due to a need to reduce social pressure. The Granger-Causality test have verified this hypothesis in some sense.

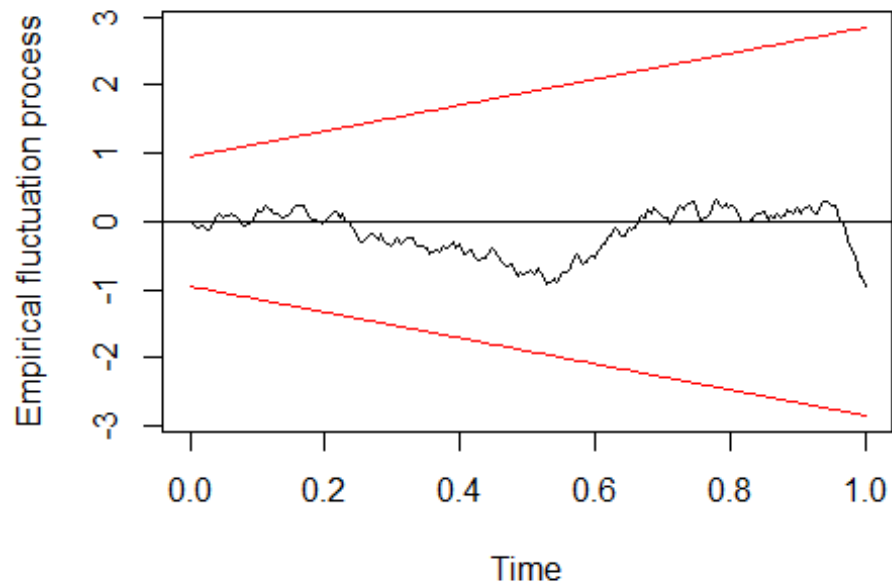
Forecasting for Crude Lignite

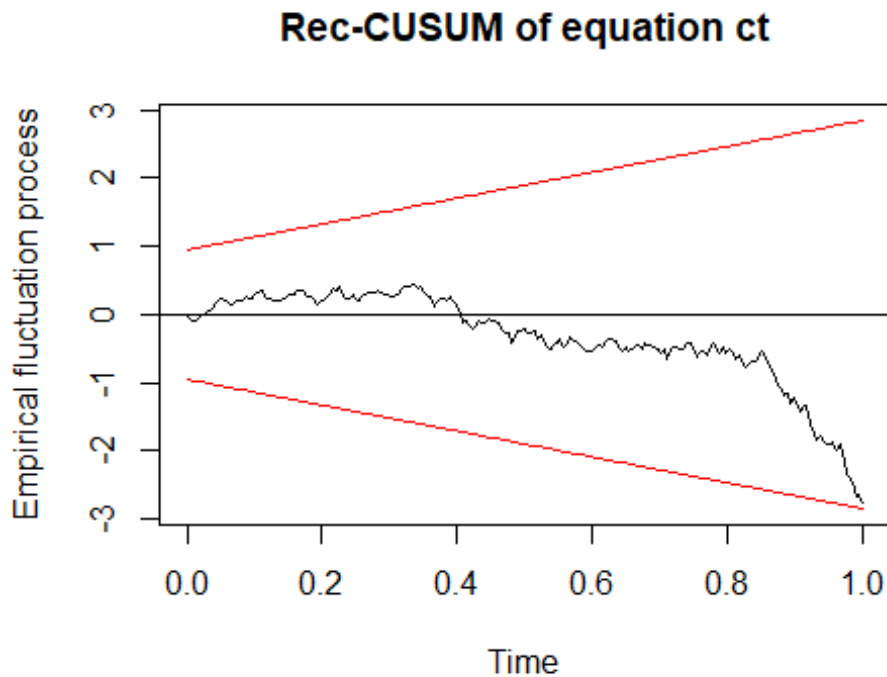


Forecasting for Coal



Rec-CUSUM of equation nt





The CUSUM fluctuates along the zero line, and don't go beyond red boundary, suggesting the model is pretty robust.

Since the VAR model is overly simplified and cannot explain for seasonality or trend, the fit is, not surprisingly, worse compared to ARMA models. The latter, with relatively complicated components, capture the trend comparably more efficiently.

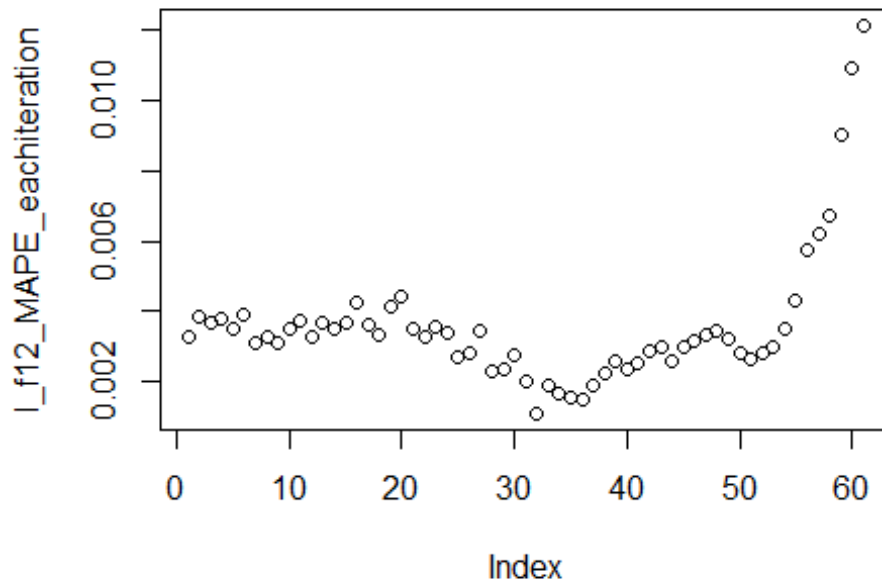
14

(a)

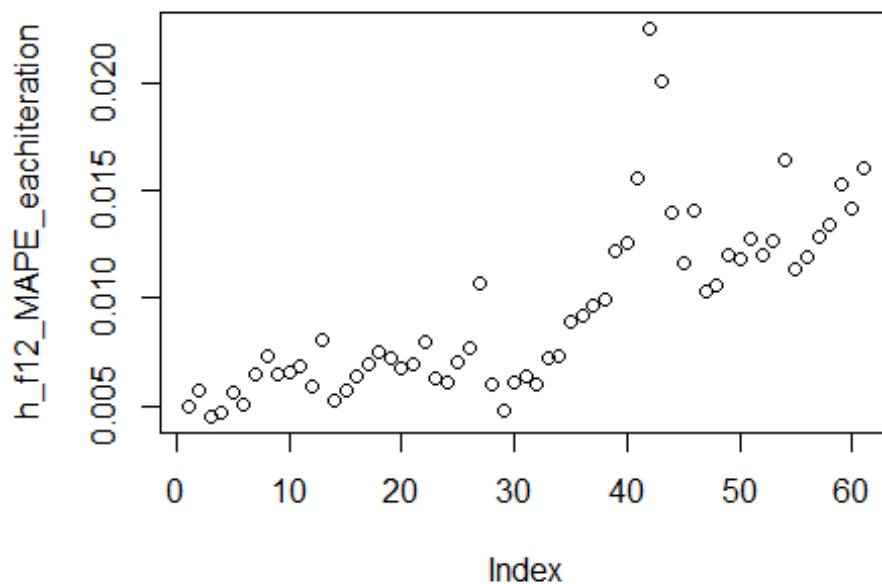
In practice, since we do not know the future yet, we can not evaluate the prediction of our models in a normal way. Most often, we will separate our data into estimation set and prediction set, and used the model fitted in estimation set to predict data in prediction set. Since data in prediction set comes from the past, we usually call this process backtest.

Here in our analysis, we have 212 observations for each variable from Jan 2002 to Aug 2019. We separate them into 140 estimation ones and 72 prediction ones. First, we use a recursive backtesting scheme and forecast 12 steps ahead for each iteration. For evaluating prediction accuracy, we calculate the Mean Absolute Percentage Error at each step and plot the MAPE over each iteration.

Crude Lignite Recursive 12-step MAPE each iteration



Coal Recursive 12-step MAPE each iteration

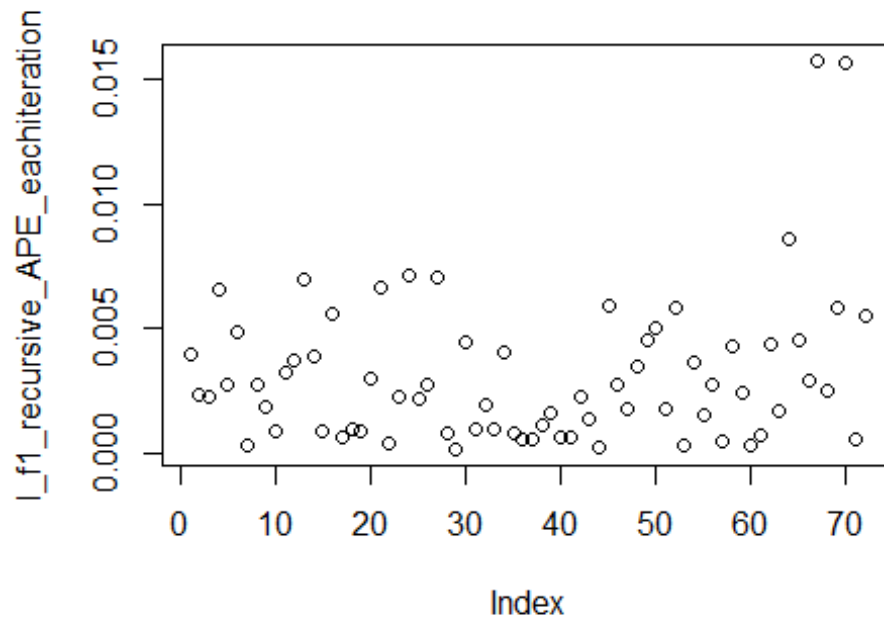


As is shown, the MAPE over each step is increasing, and the MAPE over each iteration is stationarily small, except the rapid rising at end, which may be caused the plummet of our original data.

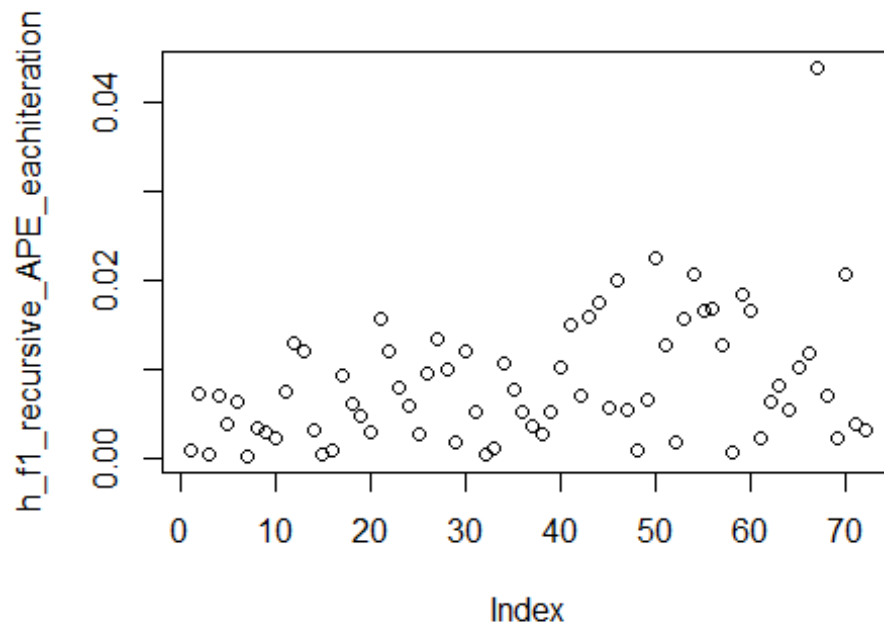
(b)

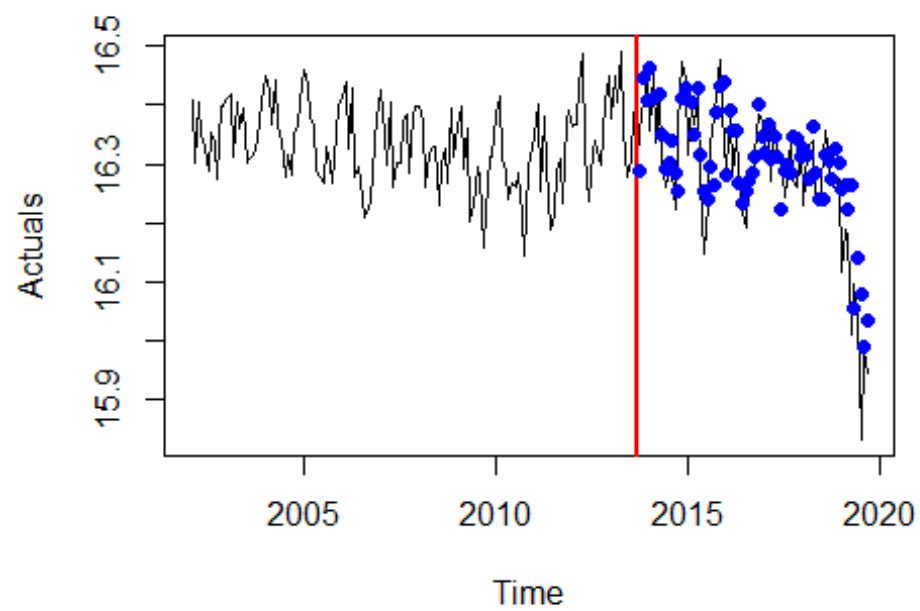
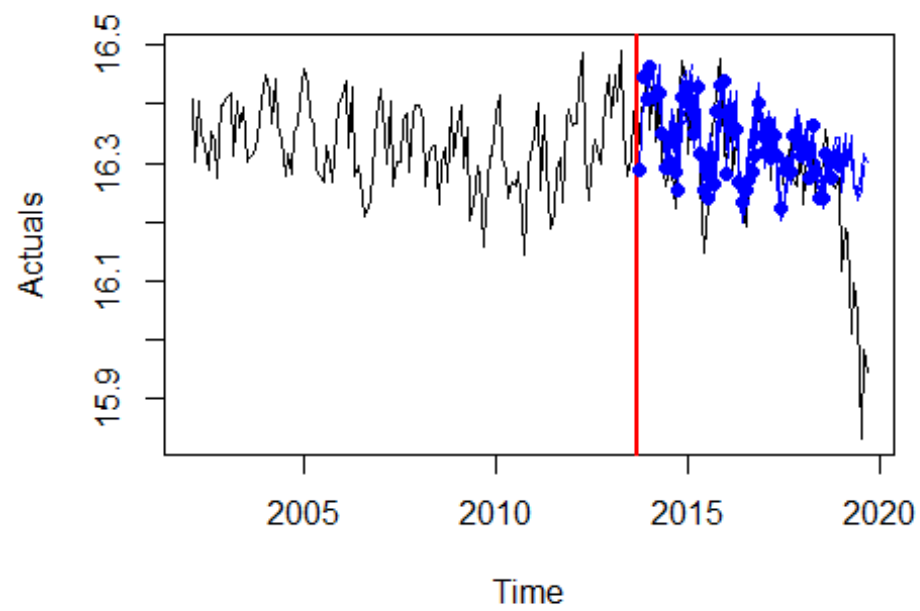
Then, we shorten the forecast horizon to 1 step ahead, and plot Absolute Percentage Error.

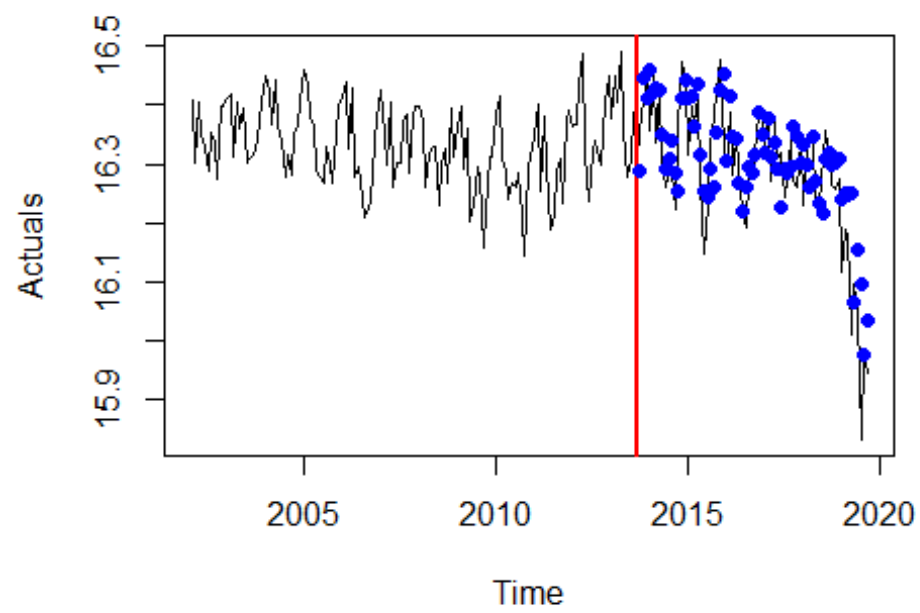
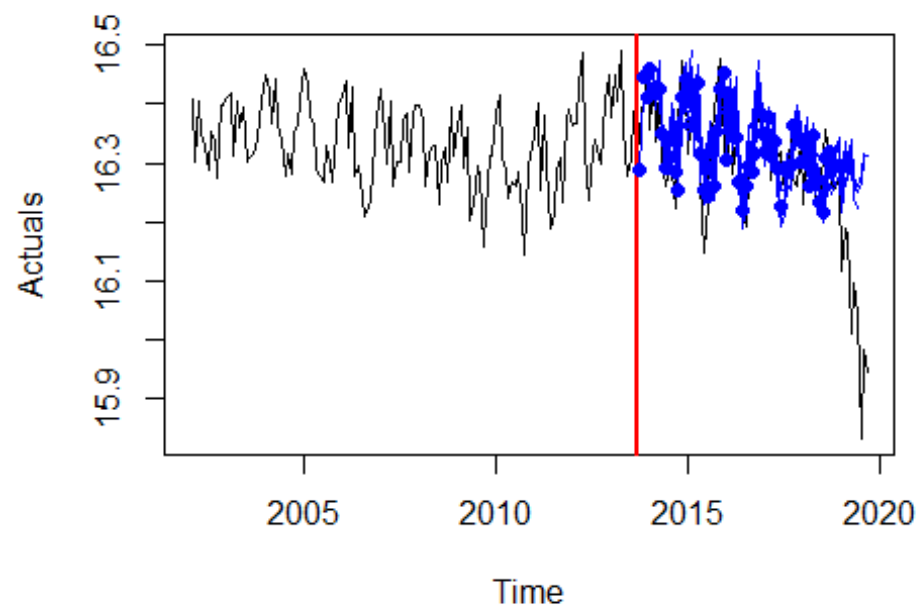
Crude Lignite Recursive 1-step APE each iteration



Coal Recursive 1-step APE each iteration





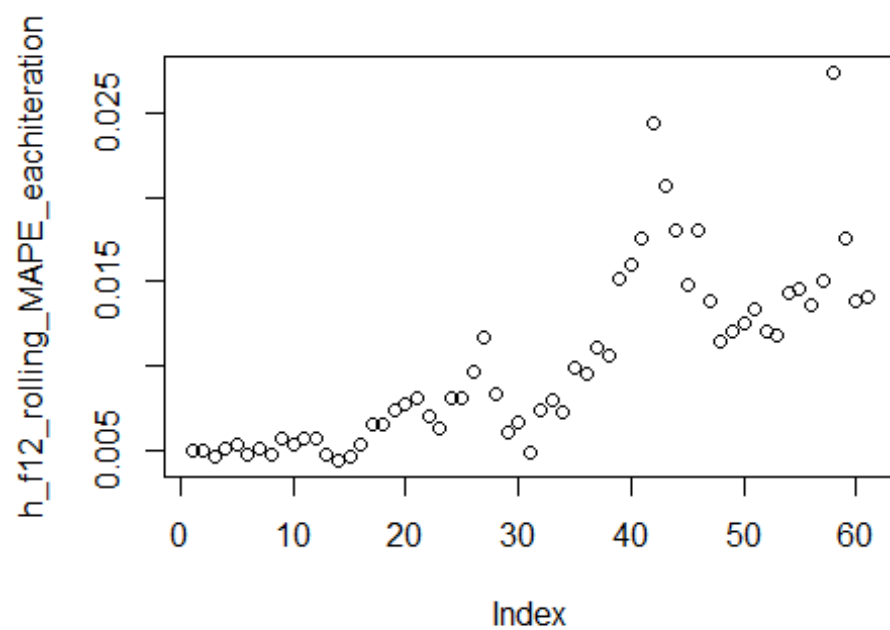
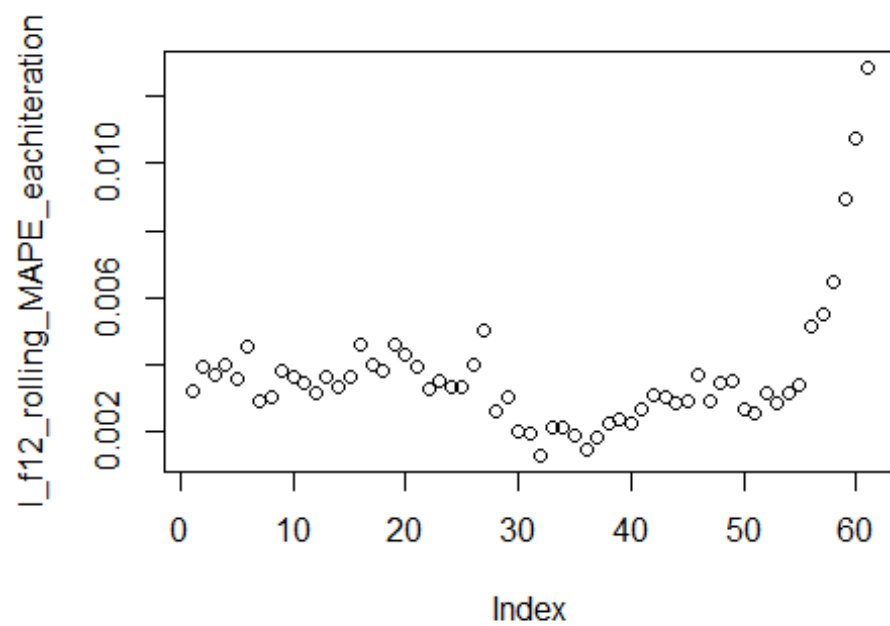


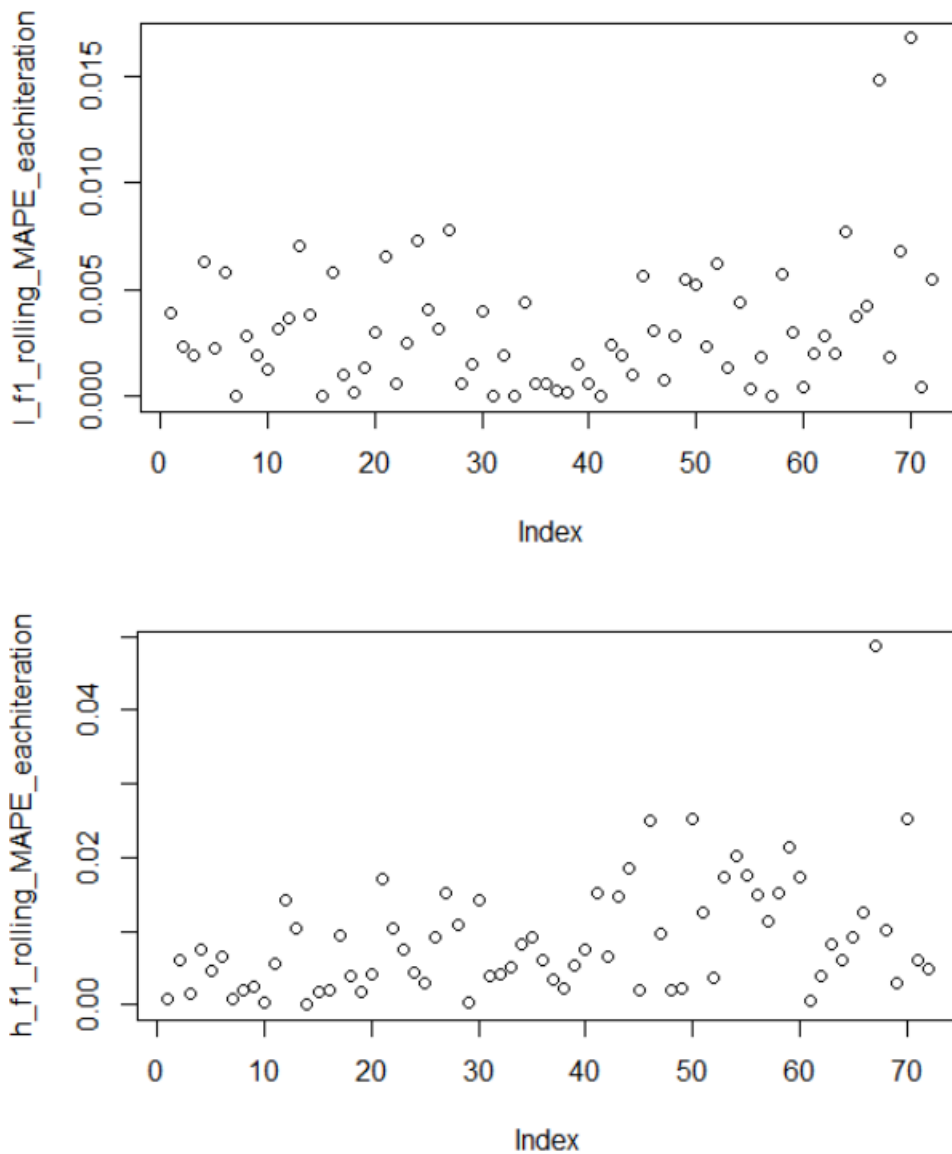
(c)

Based on our findings, for crude lignite the performance of 1 step and 12 steps ahead look similar, almost below 0.005, while for coal the performance of 12 steps ahead is better.

(d)

Next, we switch to the rolling window scheme. Not surprisingly, all the plots show similar patterns as above.





(e)

Comparing the errors found using different schemes, we find that there is no apparent difference and all the errors are very small, which tell us that our data is stationary and our model is quite stable.

III Conclusions and Future Work

In this report, we focus on the development of electricity generation through crude lignite and hard coal in Germany with respect to time. Using time series analysis, we capture the drift, seasonality, and cycles in both log series. We build S-ARIMA models for each series by observing the series, and try linear trend, linear trend plus

seasonality, linear trend plus seasonality and cycle, etc., in a step-by-step fashion. Our full ARIMA models are as follows:

1. Crude Lignite: The log process is stationary after first difference, drift removal, the seasonality of which are modeled combining seasonal autoregression of order 1, and 1-period seasonal moving average with frequency of 12. The full model consists of cycle, composed of 1-period autoregression and 2-period moving average, and drift, seasonality described above. The estimated coefficients are statistically significant, and the residual is white noise, suggesting that we have effectively detected the patterns in the process. The RMSE is 0.0528 and BIC is -594.36, suggesting the model is pretty good, also confirmed by forecast plot and CUSUM plot.

ARIMA(1,1,2)(1,0,1)₁₂ with drift

2. Hard coal: Similarly, the log process is also stationary after first difference, drift removal, and including seasonal autoregression at lag 1 and first order moving average process with frequency of 12. Finally, we get full model by adding cycle composed of first order autoregression and first order moving average. The estimated coefficients are statistically significant, and the residual is white noise, suggesting we have effectively detected the patterns in the process. The RMSE is 0.1290 and BIC is -212, suggesting the model is pretty good, also confirmed by forecast plot and CUSUM plot.

ARIMA(1,1,1)(1,0,1)₁₂ with drift

Via Backtesting our models, we can confirm our conclusion that both models perform well in forecasting, no matter using recursive backtesting or rolling window backtesting. To be more specific, our models are more suitable and reliable for short term prediction than for long term prediction.

On the other hand, we turn to the interaction between two variables and find that there is close cross correlation between them. Granger-Causality test results implies that past electricity generation of hard coal can be used to predict future electricity generation of crude lignite, while the other way around does not work, leading to our VAR(1) model as follows:

$$Lignite_t = 0.4580Lignite_{t-1} + 0.1035Coal_{t-1} + 7.1839$$

$$Coal_t = 0.3634Lignite_{t-1} + 0.7200Coal_{t-1} - 1.4520$$

First lag of hard coal and first lag of lignite itself provide a good model for crude lignite. All the estimated coefficients are statistically significant.

However, since our VAR model does not include the seasonality which influences coal significantly, VAR doesn't provide better prediction power compared to ARIMA model.

As for future work, there are several things we may take into consideration. First, we may try to add seasonality to VAR model to acquire better fit. Next, sudden turning points appeared in the trends of coal series, owing to the government policy to some extent, so it is reasonable to include some policy dummies in the model, and may also consider the use of general Piecewise Regression to better fit the model. We may make comparative study between different countries to cultivate a better understanding of traditional fuel withdrawing process and their interaction with policies.

IV Reference

Svetunkov, I. (2019). Rolling Origin. Retrieved from <https://cran.r-project.org/web/packages/greybox/vignettes/ro.html>.

V Code

```
library(lattice)
library(foreign)
library(MASS)
library(car)
require(stats)
require(stats4)
library(KernSmooth)
library(fastICA)
library(cluster)
library(leaps)
library(mgcv)
library(rpart)
library(pan)
library(mgcv)
library(DAAG)
library("TTR")
library(tis)
require("datasets")
require(graphics)
library("forecast")
#install.packages("astsa")
#require(astsa)
library(xtable)
library(stats)
library(tseries)
library(AER)
library(lmtest)
library(vars)
library(ggplot2)
library(strucchange)
library(cvTools)
```

```
library(boot)
library(rpart)
library(greybox)
```

1

```
# import the data
d = read.csv('t9.csv', header=T)
d = data.frame(d)
colnames(d) = c('time', 'lign', 'coal')
#head(d)

# Transform the data into time series
t = ts(d$time, start=2002.1, freq=12)
nt = ts(d$lign, start=2002.1, freq=12)
ct = ts(d$coal, start=2002.1, freq=12)
lt = log(t)
nt = log(nt)
ct = log(ct)

# Check acf and pacf for Crude Lignite
tsdisplay(nt, lag=48, main='Crude Lignite')

# Check acf and pacf for Coal
tsdisplay(ct, lag=48, main='Coal')

# Plot a rough sketch of the two series
cot = data.frame(t, nt, ct)

g1 = ggplot(cot, aes(x=t)) +
  geom_line(aes(x=t, y=nt), col='#333999') +
  geom_line(aes(x=t, y=ct), col='#ff0066') +
  theme_bw() +
  labs(x='Time', y='Electricity Generated / MWh') +
  theme(axis.text.x = element_text(face="bold", size=10),
        axis.text.y = element_text(face="bold", size=10))
```

g1

2

```
# Build separate arima models for the two series
n1 = auto.arima(nt,xreg=lt)
c1 = auto.arima(ct,xreg=lt)
S(n1)
coeftest(n1)
S(c1)
coeftest(c1)

plot(nt, col='blue1',lwd=1, main='Fit for Crude Lignite', xlab='Time',
      ylab='Electricity generated')
```

```

lines(n1$fitted, col='red1', lwd=1)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('blue1','red1'),bty="n")

plot(ct, col='blue3',lwd=1, main='Fit for Coal', xlab='Time',
      ylab='Electricity generated')
lines(c1$fitted, col='red3', lwd=1)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('blue1','red1'),bty="n")

Box.test(n1$residuals)
Box.test(c1$residuals)

```

3

Try a seasonal + trend model for Crude Lignite

```

n3=tslm(nt~lt+season)

par(mfrow=c(2,1))
plot(nt,ylab="Crude Lignite", xlab="Time", lwd=2, col='skyblue2',main="
seasonal+trend1 model for Lignite")
lines(n3$fitted.values,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(n3$res, ylab="Crude Lignite Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(n3$residuals, lag=36, main="ACF-Crude Lignite Residuals")
pacf(n3$residuals, lag=36, main="PACF-Crude Lignite Residuals")

# Change a trend, the result is not improved
n32=tslm(nt~t+season)

par(mfrow=c(2,1))
plot(nt,ylab="Crude Lignite", xlab="Time", lwd=2, col='skyblue2',main="
seasonal+trend2 model for Lignite")
lines(n32$fitted.values,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(n32$res, ylab="Crude Lignite Residuals",type='l',xlab="Time",lwd=
2)

par(mfrow=c(2,1))
acf(n32$residuals, lag=36, main="ACF-Crude Lignite Residuals")
pacf(n32$residuals, lag=36, main="PACF-Crude Lignite Residuals")

```

```

# Try a trend + cycle model, we see that the fit is not well.
# much of the repeated fluctuations, which can be explained using
# seasonal dummies or seasonal ARIMA are not accounted for here.

n4=Arima(nt, order=c(0,1,3), xreg=lt)

par(mfrow=c(2,1))
plot(nt,ylab="Crude Lignite", xlab="Time", lwd=2, col='skyblue2', main=
"trend + cycle model ")
lines(n4$fitted,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(n4$res, ylab="Crude Lignite Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(n4$residuals, lag=36, main="ACF-Crude Lignite Residuals")
pacf(n4$residuals, lag=36, main="PACF-Crude Lignite Residuals")

# To sum up, after iterated trial and comparison of BIC, we arrive
# at the following model

n2 = Arima(nt, order=c(1,1,2),
           seasonal = list(order=c(1,0,1)), include.drift=T)
par(mfrow=c(2,1))
plot(nt,ylab="Crude Lignite", xlab="Time", lwd=2, col='skyblue2')
lines(n2$fitted,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(n2$res, ylab="Crude Lignite Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(n2$residuals, lag=36, main="ACF-Crude Lignite Residuals")
pacf(n2$residuals, lag=36, main="PACF-Crude Lignite Residuals")

# BIC and AIC suggest the last model is the best
AIC(n3, n32, n4, n2)
BIC(n3, n32, n4, n2)

# The Residuals are White Noise
Box.test(n2$residuals)

# Try a seasonal + trend model for Coal
c3=tslm(ct~lt+season)

par(mfrow=c(2,1))
plot(ct,ylab="Coal", xlab="Time", lwd=2, col='skyblue2')

```

```

lines(c3$fitted.values,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(c3$res, ylab="Coal Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(c3$residuals, lag=36, main="ACF-Coal Residuals")
pacf(c3$residuals, lag=36, main="PACF-Coal Residuals")

# Change a trend, the result is not improved
c32=tslm(ct~t+season)

par(mfrow=c(2,1))
plot(ct,ylab="Coal", xlab="Time", lwd=2, col='skyblue2')
lines(c32$fitted.values,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(c32$res, ylab="Coal Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(c32$residuals, lag=36, main="ACF-Coal Residuals")
pacf(c32$residuals, lag=36, main="PACF-Coal Residuals")

# Try a trend + cycle model

c4=Arima(ct, order=c(1,1,1), xreg=lt)

par(mfrow=c(2,1))
plot(ct,ylab="Coal", xlab="Time", lwd=2, col='skyblue2')
lines(c4$fitted,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(c4$res, ylab="Coal Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(c4$residuals, lag=36, main="ACF-Coal Residuals")
pacf(c4$residuals, lag=36, main="PACF-Coal Residuals")

# To sum up, after iterated trial and comparison of BIC, we arrive
# at the following model

c2 = Arima(ct, order=c(1,1,1),
           seasonal = list(order=c(1,0,1)), include.drift=T)
par(mfrow=c(2,1))
plot(ct,ylab="Coal", xlab="Time", lwd=2, col='skyblue2')

```



```

lines(c2$fitted,col="red2",lwd=2,lty=2)
legend("bottomleft",pch=c(15,15),legend=c("Data","Fit"),
      col=c('skyblue2','red2'),bty="n")

plot(c2$res, ylab="Coal Residuals",type='l',xlab="Time",lwd=2)

par(mfrow=c(2,1))
acf(c2$residuals, lag=36, main="ACF-Coal Residuals")
pacf(c2$residuals, lag=36, main="PACF-Coal Residuals")

# BIC and AIC suggest the Last model is the best
AIC(c3, c32, c4, c2)
BIC(c3, c32, c4, c2)

# The Residuals are White Noise
Box.test(c2$residuals)

# The Final Model for the following parts
lt = log(t)

n2 = Arima(nt, order=c(1,1,2),
          seasonal = list(order=c(1,0,1)), include.drift=T)
S(n2)
coeftest(n2)

c2 = Arima(ct, order=c(1,1,1),
          seasonal = list(order=c(1,0,1)), include.drift=T)
S(c2)
coeftest(c2)

```

4

```

library(ggpubr)
# create the dataframe to put in
pr1 = data.frame(t, fv = n2$fitted, ehat = n2$residuals, dat=nt)
pr2 = data.frame(t, fv = c2$fitted, ehat = c2$residuals, dat=ct)

l = c("Crude Lignite","Coal")

rfv <- function(data, title){
  # Plot for each models' residuals and fitted values
  g1 = ggplot(data, aes(x=t, y=ehat))+
    geom_line(col='#3300cc') +
    labs(x='Time', y='Residuals', title=title) + theme_bw()

  g2 = ggplot(data, aes(x=t, y=fv))+
    geom_line(col='#0000cc', alpha=0.5) +
    geom_line(aes(x=t,y=dat), col='#ff3333') +
    labs(x='Time', y='Fitted Values', title=title) + theme_bw()
  g = ggarrange(g1, g2, ncol=1, nrow=2)
}

```

```

    return(g)
}
rfv(pr1, l[1])

# Paint the graph using two code blocks in r_markdown
# To neatly knit the pictures in the output pdf or doc file

rfv(pr2, l[2])

```

5

```

# acf pacf for crude lignite
par(mfrow=c(2,1))
acf(n2$residuals, main='ACF - Residual from Crude-Lignite-Generated Electricity',
    xlab='Displacement')
pacf(n2$residuals, main='PACF - Residual from Crude-Lignite-Generated Electricity',
    xlab='Displacement')

# acf pacf for coal
par(mfrow=c(2,1))
acf(c2$residuals, main='ACF - Residual from Coal-Generated Electricity',
    xlab='Displacement')
pacf(c2$residuals, main='PACF - Residual from Coal-Generated Electricity',
    xlab='Displacement')

```

6

```

# plot the CUSUM
plot(efp(n2$residuals~1, type='Rec-CUSUM'), main='CUSUM Crude Lignite')

plot(efp(c2$residuals~1, type='Rec-CUSUM'), main='CUSUM Coal')

```

7

```

rn = recresid(n2$residuals~1)
plot(rn, pch=16, col='darkred',
     ylab='Recursive Residuals for Crude Lignite')

rc = recresid(c2$residuals~1)
plot(rc, pch=16, col='skyblue1',
     ylab='Recursive Residuals for Coal')

```

8

```

# AIC and BIC standard for both series of models
AIC(n3, n32, n4, n2)
BIC(n3, n32, n4, n2)
AIC(n3, n32, n4, n2)
BIC(n3, n32, n4, n2)

```

```
# Error Metrics for models
```

```
accuracy(n3)  
accuracy(n32)  
accuracy(n4)  
accuracy(n2)
```

```
accuracy(c3)  
accuracy(c32)  
accuracy(c4)  
accuracy(c2)
```

9

```
# Forecast 12-step ahead
```

```
#new = ts(start=2019.9, end=2020.9, freq=12)  
#new = seq(2019+8/12, by=1/12, length=12)  
#newt = Log(new)  
plot(forecast(n2, h=12), main='Forecasting for Crude Lignite')  
plot(forecast(c2, h=12), main='Forecasting for Coal'  
      ,shadecols="oldstyle")
```

10

```
# Plot
```

```
ccf(nt, ct, ylab="Cross-Correlation Function",  
      main = "Crude Lignite and Coal CCF")
```

```
ccf(ct,nt, ylab="Cross-Correlation Function",  
      main = "Coal and Crude Lignite CCF")
```

```
# Build VAR model
```

```
co = data.frame(nt, ct)
```

```
VARselect(co, lag.max=10)  
v1 = VAR(co, 1)  
summary(v1)
```

```
# plot the model
```

```
plot(v1)
```

11

```
irf(v1)  
plot(irf(v1))
```

12

```
grangertest(nt, ct)  
grangertest(ct, nt)
```

13

Forecast 12-step ahead

```
varpred = predict(v1, h=12)
plot(varpred, main=c('Forecasting for Crude Lignite',
                    'Forecasting for Coal'))
plot(stability(v1, type = "Rec-CUSUM"), plot.type="single")
```

14

Just to make sure the data is correct

```
t = ts(d$time, start=2002.1, freq=12)
nt = ts(d$lign, start=2002.1, freq=12)
ct = ts(d$coal, start=2002.1, freq=12)
lt = log(t)
nt = log(nt)
ct = log(ct)
```

forecast

```
model_l = "forecast(Arima(data, order=c(1,1,2), seasonal = c(1,0,1),
method='CSS', include.drift=T),h = h,level=95)"
model_h = "forecast(Arima(data, order=c(1,1,1), seasonal = c(1,0,1),
method='CSS', include.drift=T),h = h,level=95)"
ourValue = "mean"
```

separation : training 140 test 72, you can make your own

```
n12rec = ro(nt,h=12,origins = 61,call=model_l,value=ourValue,
            ci=F,co=T)
n1rec = ro(nt,h=1,origins = 72,call=model_l,value=ourValue,
           ci=F,co=T)
n12rol = ro(nt,h=12,origins = 61,call=model_l,value=ourValue,
            ci=T,co=T)
n1rol = ro(nt,h=1,origins = 72,call=model_l,value=ourValue,
           ci=T,co=T)
c12rec = ro(ct,h=12,origins = 61,call=model_h ,value=ourValue,
            ci=F,co=T)
c1rec = ro(ct,h=1,origins = 72,call=model_h ,value=ourValue,
           ci=F,co=T)
c12rol = ro(ct,h=12,origins = 61,call=model_h ,value=ourValue,
            ci=T,co=T)
c1rol = ro(ct,h=1,origins = 72,call=model_h ,value=ourValue,
           ci=T,co=T)
```

MAPE over each step for recursive f12

```
l_f12_MAPE_eachstep = apply(abs((n12rec$mean-n12rec$holdout)
                               /n12rec$holdout),1,mean)
l_f12_MAPE_eachiteration = apply(abs((n12rec$mean-n12rec$holdout)
                                     /n12rec$holdout),2,mean)
plot(l_f12_MAPE_eachstep, main='Crude Lignite Recursive 12-step MAPE ea
ch step')
```

```
plot(l_f12_MAPE_eachiteration, main='Crude Lignite Recursive 12-step MA
PE each iteration')
```

```
h_f12_MAPE_eachstep = apply(abs((c12rec$mean-c12rec$holdout)
                               /c12rec$holdout),1,mean)
h_f12_MAPE_eachiteration = apply(abs((c12rec$mean-c12rec$holdout)
                                     /c12rec$holdout),2,mean)
plot(h_f12_MAPE_eachstep, main='Coal Recursive 12-step MAPE each step')
plot(h_f12_MAPE_eachiteration, main='Coal Recursive 12-step MAPE each i
teration')
```

```
# Recursive Backtest 1-step ahead
# APE over each iteration for recursive f1
l_f1_recursive_APE_eachiteration = abs(n1rec$mean-n1rec$holdout)/n1rec
$holdout
l_f1_recursive_APE_eachiteration = t(l_f1_recursive_APE_eachiteration)
plot(l_f1_recursive_APE_eachiteration, main='Crude Lignite Recursive 1-
step APE each iteration')
```

```
h_f1_recursive_APE_eachiteration = abs(c1rec$mean-c1rec$holdout)/c1rec
$holdout
h_f1_recursive_APE_eachiteration = t(h_f1_recursive_APE_eachiteration)
plot(h_f1_recursive_APE_eachiteration, main='Coal Recursive 1-step APE
each iteration')
```

```
plot(n12rec)
plot(n1rec)
plot(n12rol)
plot(n1rol)
```

```
# Rolling Backtest 12-step ahead
```

```
# forecast error over each iteration for rolling f12
l_f12_rolling_FE_eachiteration = n12rol$mean-n12rol$holdout
h_f12_rolling_FE_eachiteration = c12rol$mean-c12rol$holdout
# plot? each step or just 12th step?
```

```
# forecast error over each iteration for rolling f1
l_f1_rolling_FE_eachiteration = n1rol$mean-n1rol$holdout
l_f1_rolling_FE_eachiteration = t(l_f1_rolling_FE_eachiteration)
plot(l_f1_rolling_FE_eachiteration)
```

```
h_f1_rolling_FE_eachiteration = c1rol$mean-c1rol$holdout
h_f1_rolling_FE_eachiteration = t(h_f1_rolling_FE_eachiteration)
plot(h_f1_rolling_FE_eachiteration)
```

```

# forecast error over each iteration for rolling f12
l_f12_rolling_MAPE_eachiteration = apply(abs((n12rol$mean-n12rol$holdout)
                                          /n12rol$holdout),2,mean)
plot(l_f12_rolling_MAPE_eachiteration)

h_f12_rolling_MAPE_eachiteration = apply(abs((c12rol$mean-c12rol$holdout)
                                          /c12rol$holdout),2,mean)
plot(h_f12_rolling_MAPE_eachiteration)


# forecast error over each iteration for rolling f1
l_f1_rolling_APE_eachiteration = abs(n1rol$mean-n1rol$holdout)/n1rol$holdout
l_f1_rolling_APE_eachiteration = t(l_f1_rolling_APE_eachiteration)
plot(l_f1_rolling_APE_eachiteration)

h_f1_rolling_APE_eachiteration = abs(c1rol$mean-c1rol$holdout)/c1rol$holdout
h_f1_rolling_APE_eachiteration = t(h_f1_rolling_APE_eachiteration)
plot(h_f1_rolling_APE_eachiteration)

```