Time Series-2

Mehmet Eyyupoglu - s174448

Github Repo for the structured code: github.com/eyyupoglu/time-series-2

Question 3.1

Plotting

The test data set is shaded with the grey region in the folowwing plots.

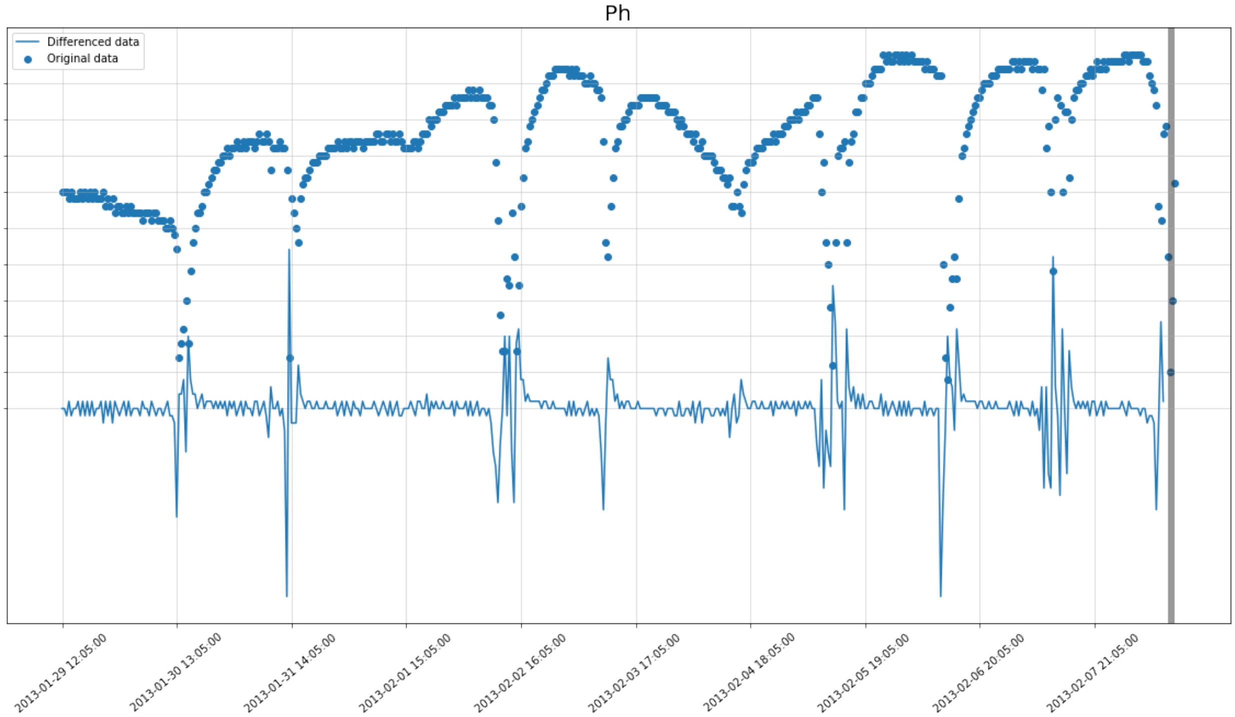


Figure 1: The heating

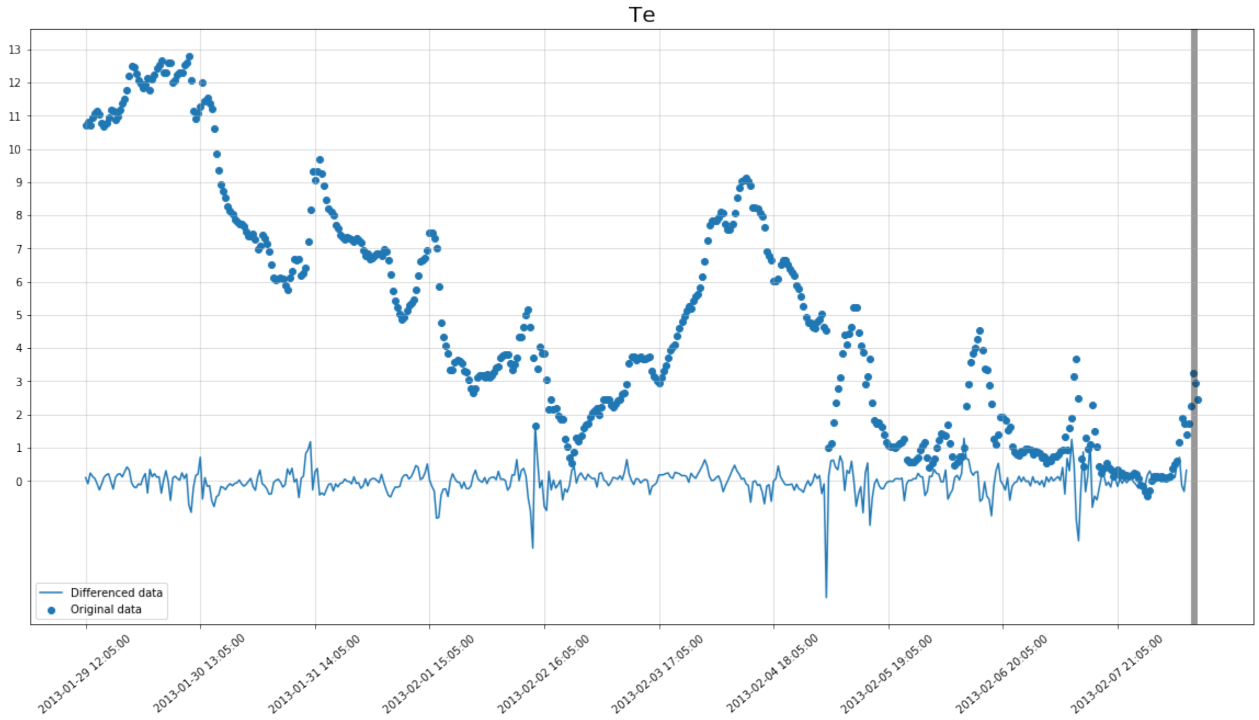


Figure 2: The outdoor temperature

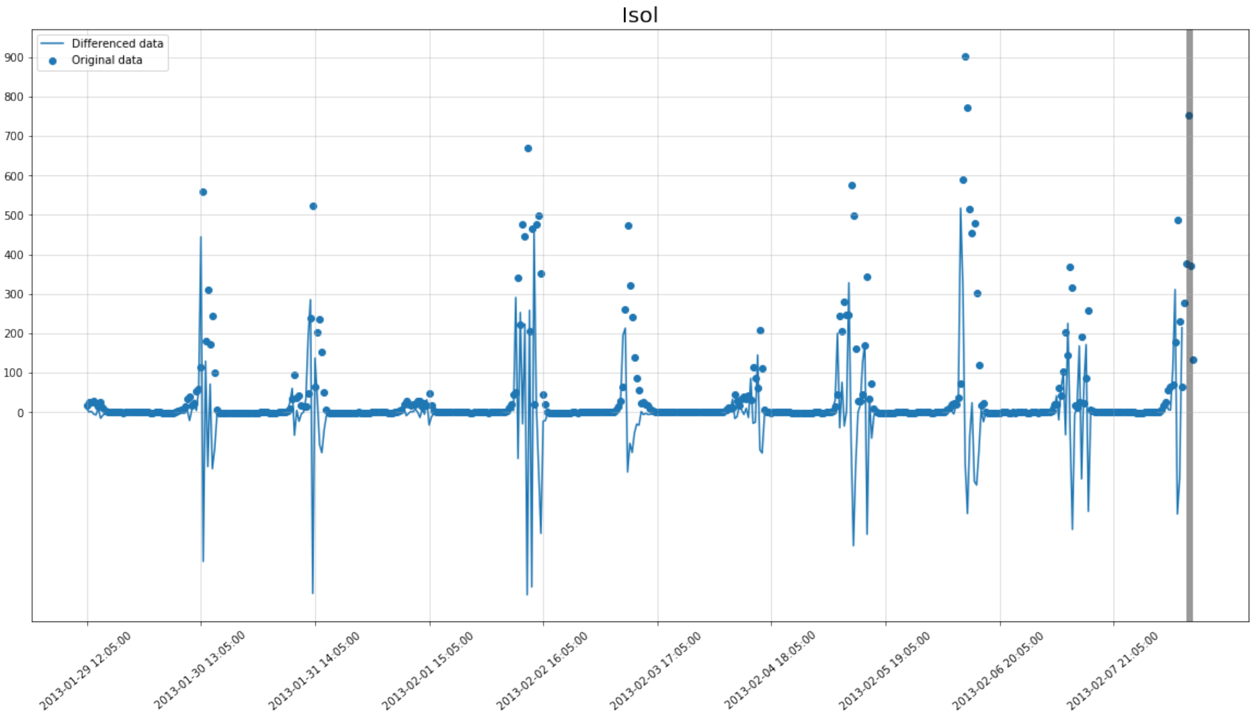
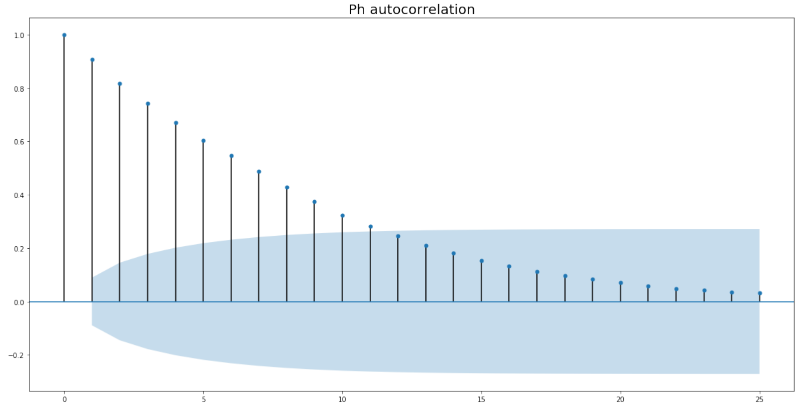


Figure 3: The solar radiation

Question 3.2

Plot the autocorrelation function



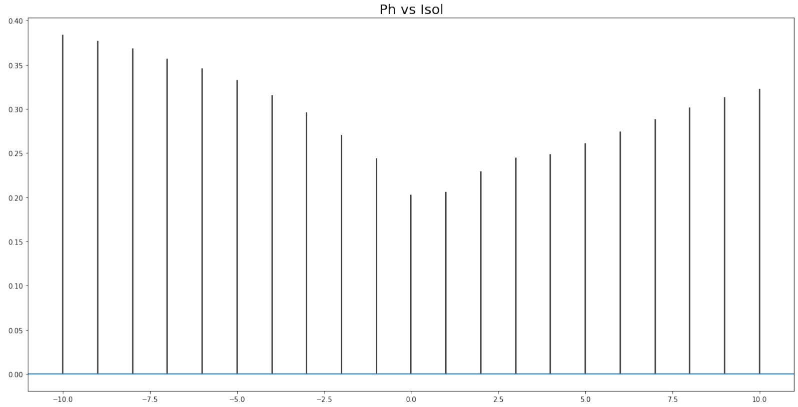
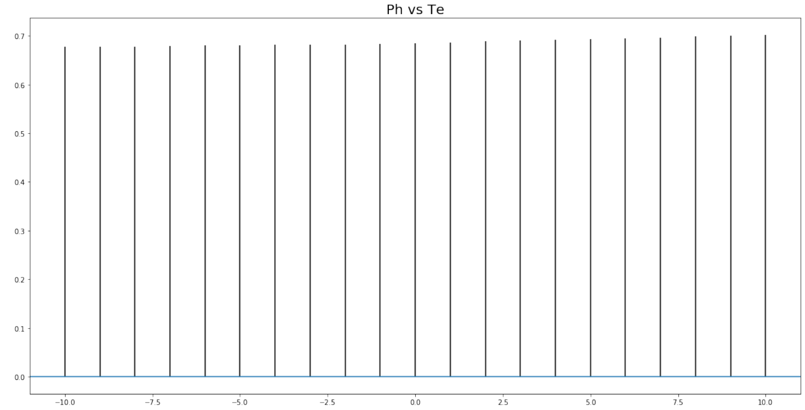


Figure 4: We see that there is fat tails in the cross correlation function between Ph and Isol

In figure 4 the reason why there is fat tails is that the auto correlation function of Ph is not exponentially decaying which means that it is not stationary.



2. What can you say about a model for Ph at this stage?

From the autocorrelation plot of Ph, it is seen that there is a dependency between the historical data points of Ph and current data points. Likewise, the cross correlation between Ph and Isol tells us that there is a dependency.

Question 3.3

1. Try to find a suitable ARIMA model for the heating level Ph.

Box-Jenkins Methodology

1. **Identification** . Using plots of the data, autocorrelations, partial autocorrelations, and other information, a class of simple ARIMA models is selected. This amounts to estimating appropriate values for p, d, and q. First we look at the figure 1 and observe that there is definitely a trend and we suspect that this trend causes the process to be non-stationary. We look at the autocorrelations and p. autocorrelations and realise that large autocorrelations persist even after several lags. Then we may either fit and remove a deterministic trend(prewhitening) or difference the series. This time we choose differencing the series.

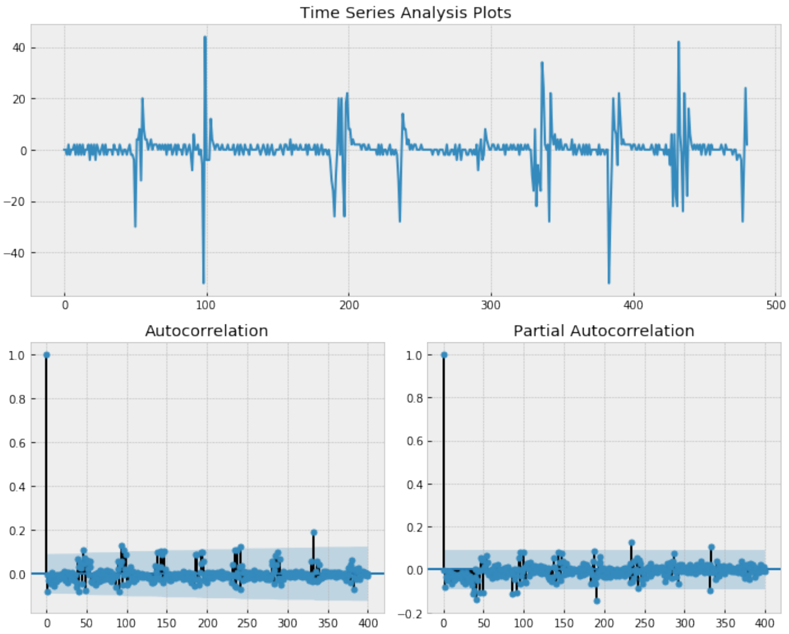
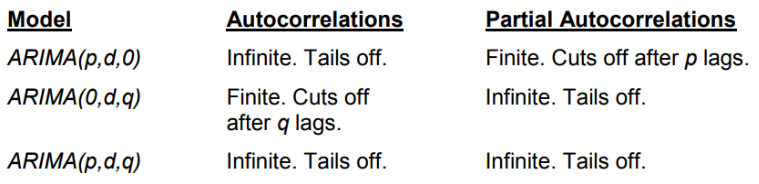


Figure 5: Autocorrelations of the first differencing of Ph

Now we can guess a reasonable model for the data based on the figure 6 also stated in the book(see table 6.1, page 155). After differencing the process, in figure 5 we observe that the autocorrelations go to zero quite fast. They die out quickly, then the appropriate value of d has been found. Therefore, **value of d** is 1.



Model	Autocorrelations	Partial Autocorrelations
ARIMA(p,d,0)	Infinite. Tails off.	Finite. Cuts off after *p* lags.
ARIMA(0,d,q)	Finite. Cuts off after *q* lags.	Infinite. Tails off.
ARIMA(p,d,q)	Infinite. Tails off.	Infinite. Tails off.

Figure 6: Magic table

If the partial autocorrelations cut off after a few lags, the last lag with a large value would be the estimated **value of p**. In figure 5 we see that the last lag with a large value would be 4. So we try p as 4. However, we also see that there are later lags that seems to be significant for example, 46, 92, etc. This gives us the hint that the model should be seasonal. Then we both incorporate a lag-4 and lag -48. The reason we still incorporate the lag-4 because non-seasonal behavior still matter.

$$(1-\varphi B^4)(1-\phi B^{48})(Ph()t)=\epsilon_t \tag{1}$$

2. **Estimation**. The phis and thetas of the selected model are estimated using maximum likelihood techniques, backcasting, etc., as outlined in Box-Jenkins (1976).

3. **Diagnostic Checking**. The fitted model is checked for inadequacies by considering the autocorrelations of the residual series (the series of residual, or error, values). We fit and see the residual structure of $ARIMA(4,1,0)x(1,1,0)_{48}$. In figure we gradually increase the seasonality degree and try to see the relative decrease in AIC-BIC scores.

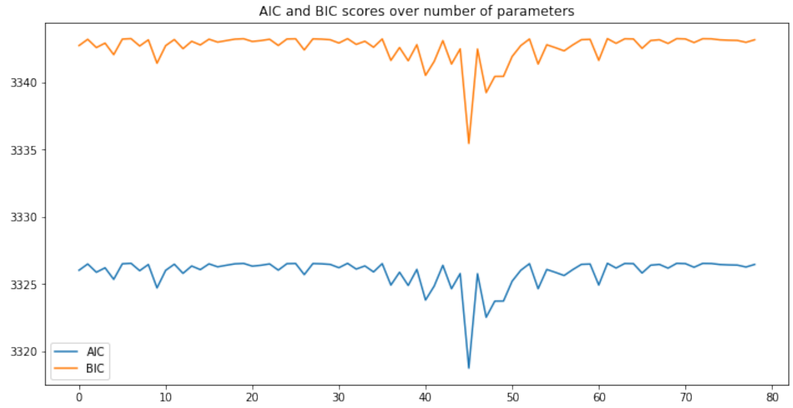


Figure 7: Seasonal p is fitted from 1 to 80 and plotted their AIC and BIC

Around 46-48, we see a "knee", the trend of the decrease changes. Therefore we must consider the minimum AIC scored value as the optimal one. In this case it is 46. When we fit our model we get the plots in figure 14

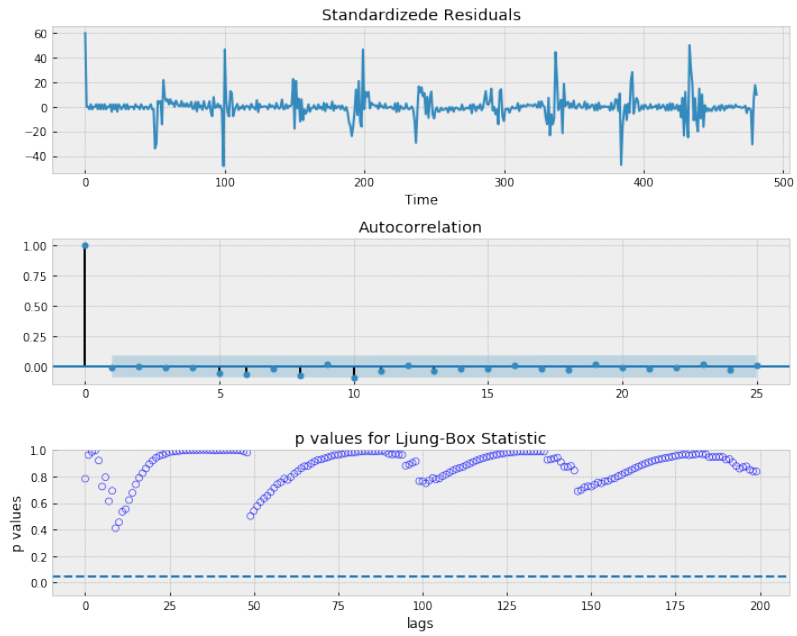


Figure 8: p-values and autocorrelation structure of the fitted model's residuals

These steps are applied iteratively until step three does not produce any improvement in the model.

2. Predict the four observations that were left out.

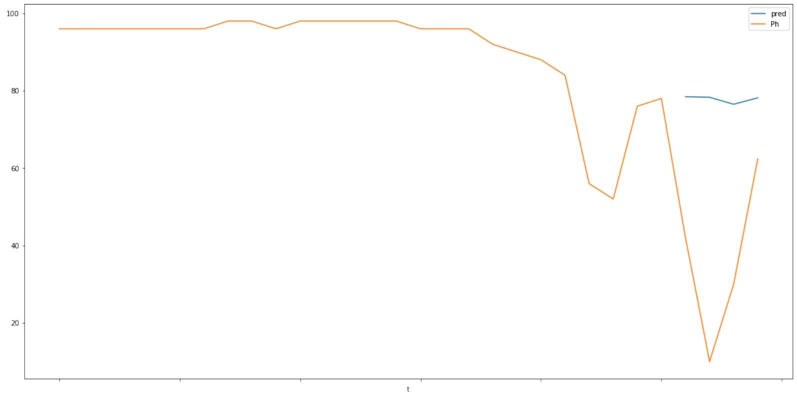


Figure 9: Predictions of the test data and the original data

	1	2	3	4
Predictions	78.45	78.30	76.49	78.15
Original	42	10	30	62.4

Question 3.4

1. Make a model selection by adding/removing the external variables in a reasonable way.

Similarly, since we have the similar initial ACF-PCAF structure, then we again decide to differentiate the data output variable either in th model or before inserting it to the model. I choose to do it inside the model by setting the **d as 1**. However, this time we have an external regressor. Our

$$(1 - \varphi B^4)(1 - \phi B^{48})(Ph()t) = \beta X_t + \epsilon_t \tag{2}$$

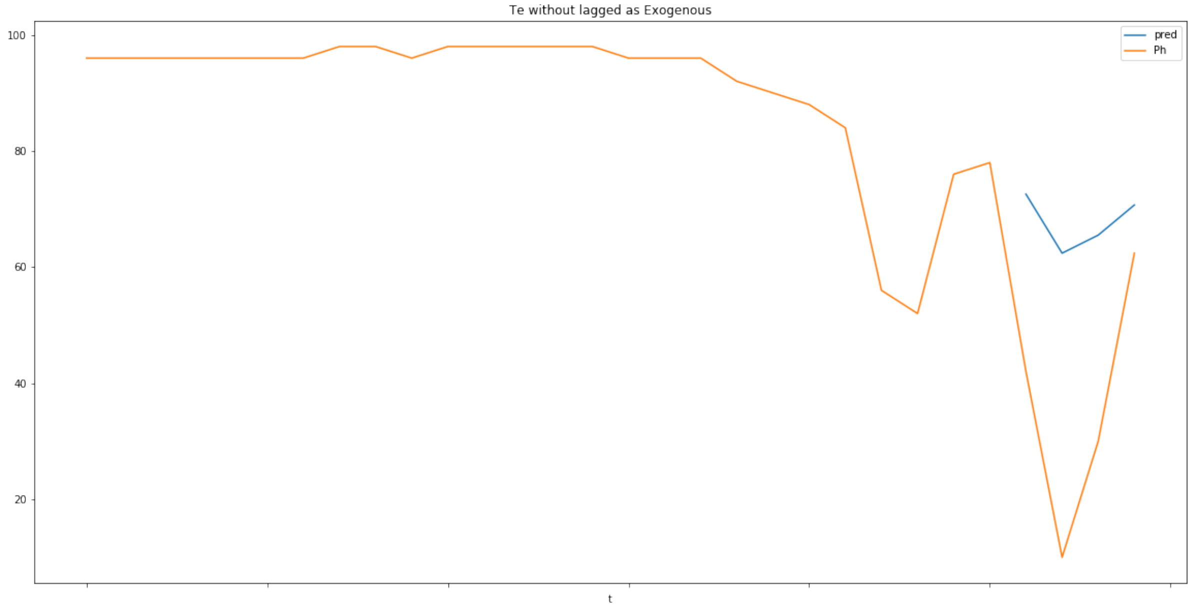


Figure 10: Predictions of the test data and the original data by using Te as exogenous

Next we add Isol as exogenous parameter and see the changes.

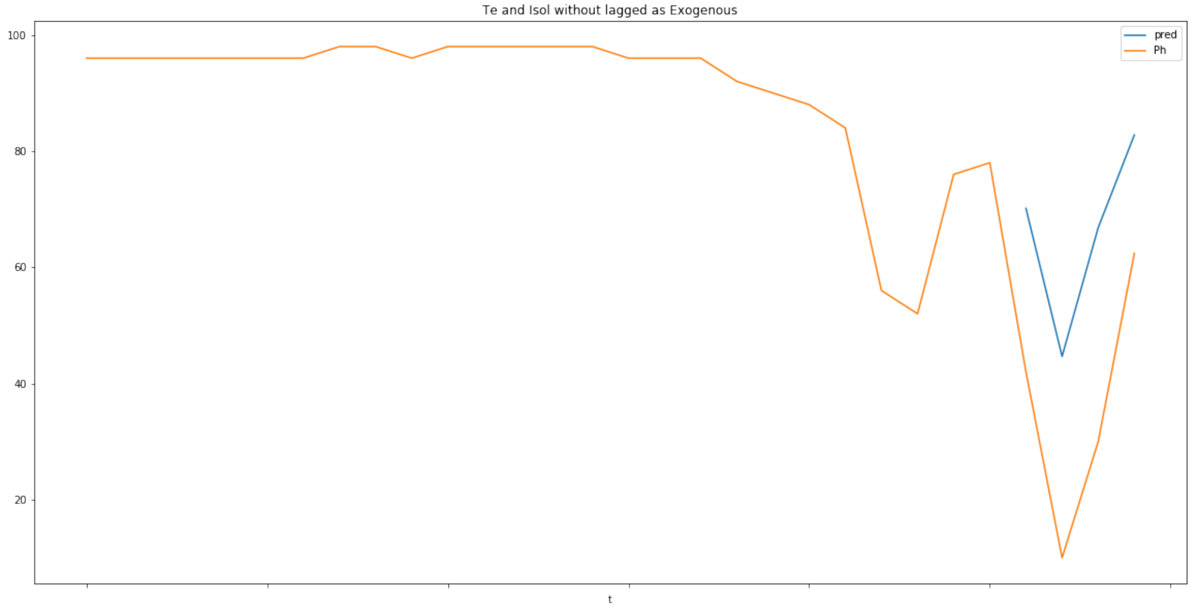


Figure 11: Predictions of the test data and the original data by using Te and Isol as exogenous

Next we add one lagged version of Te, and see the improvement.

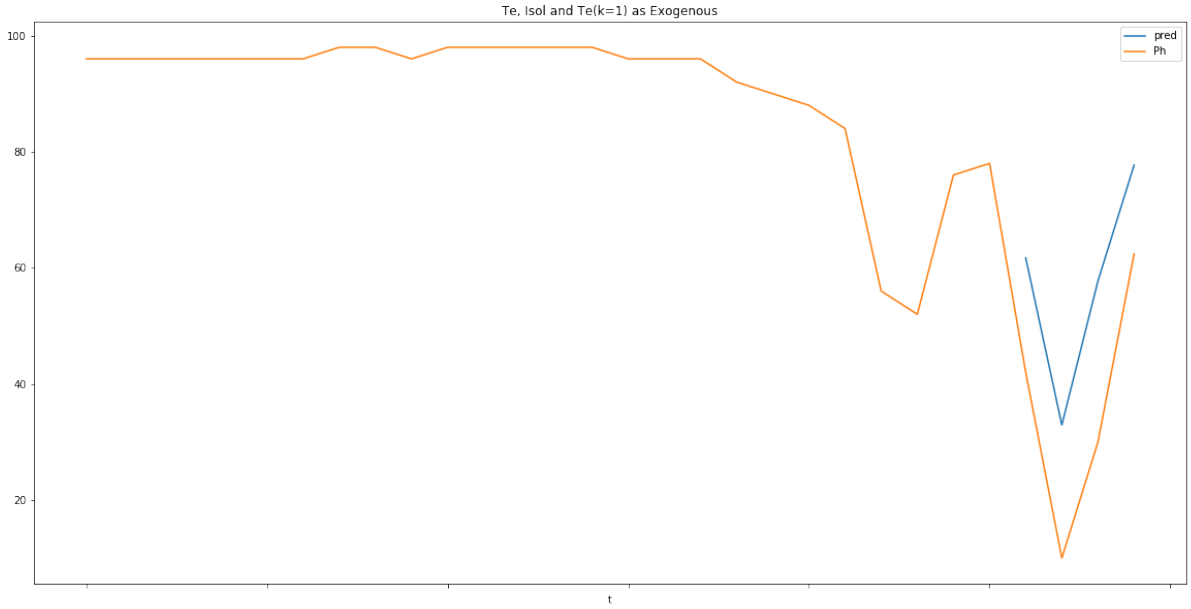


Figure 12: Predictions of the test data and the original data by using Te, Isol and Te(k=1) as exogenous

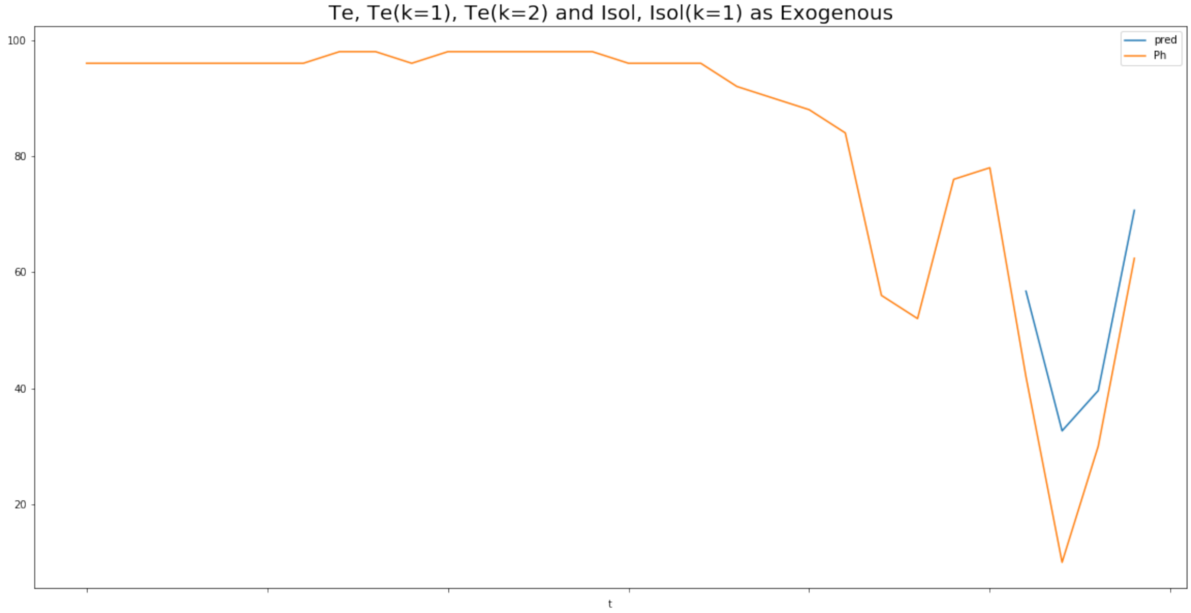


Figure 13: Predictions of the test data and the original data by using Te-Te(1),Te(2) and Isol-Isol(1) as exogenous

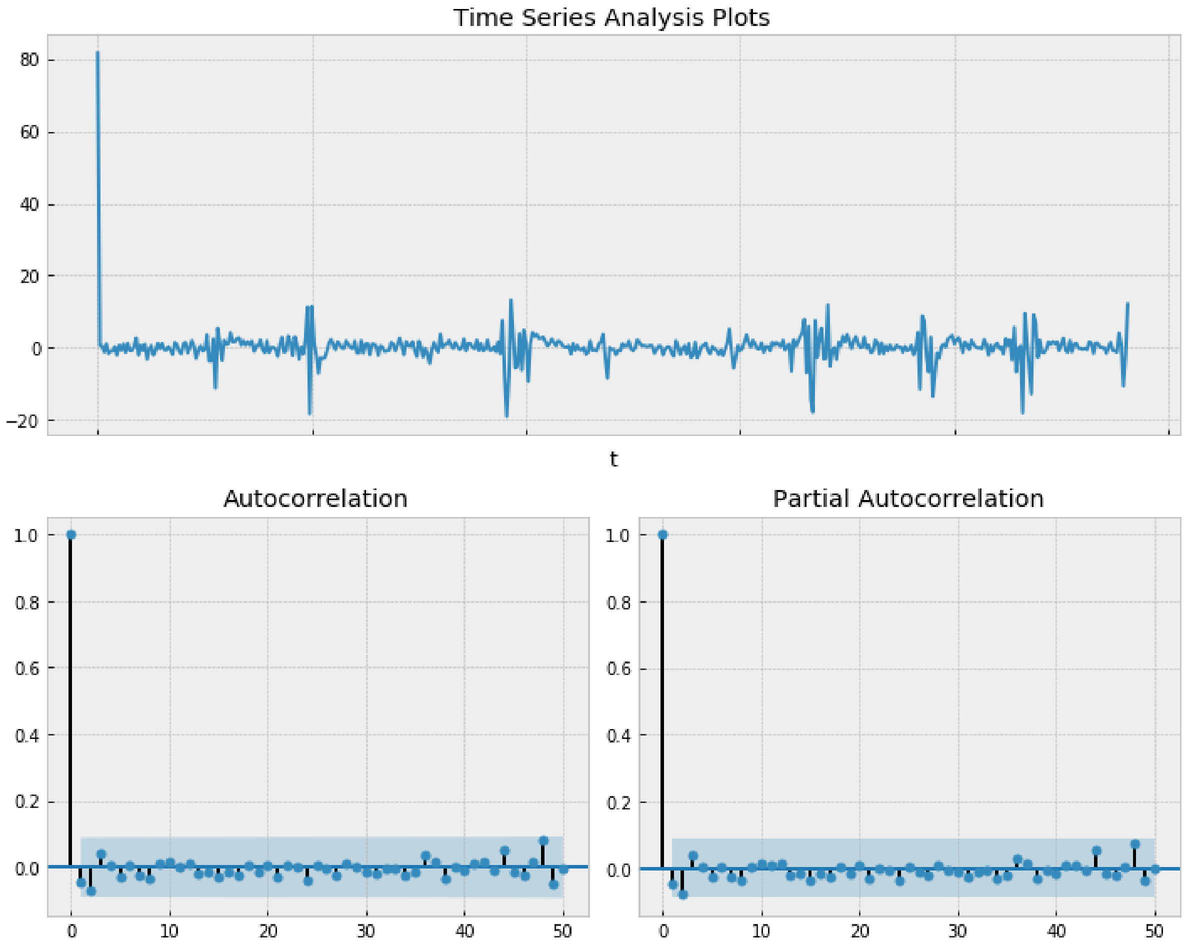


Figure 14: Predictions of the test data and the original data by using Te-Te(1),Te(2) and Isol-Isol(1) as exogenous

Automated way

Here we follow the following backward selection algorithm after we add all the columns until 20 lagged versions of exegenous variables

1. Fit the model,

2. List the p values of the coefficients. Get rid of the maximum one if it is among the exegenous varibles. It means that there is significant evidence that it is not significant if it is greater than 5

3. Check if the AIC score is improving after removing one parameter. It should decrease because it penalise parameters. If it doesnt significantly decreases then keep the exegenous matrix as it is otherwise go back to step 1.

4. Check the residuals if they seem like white noise. If it seems like some is too significant, try to include it as well.

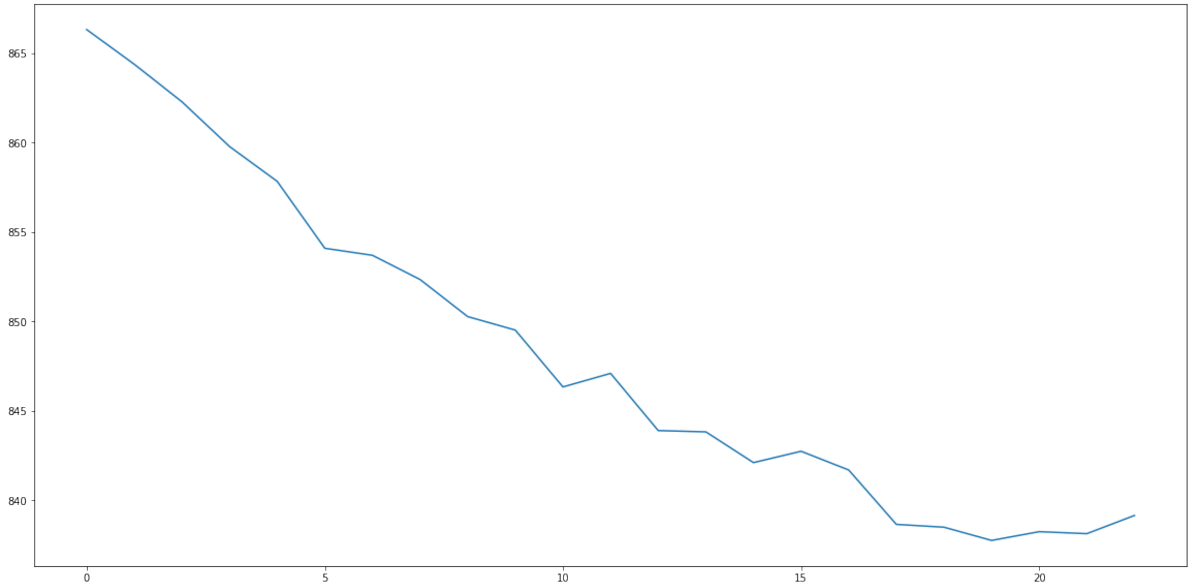
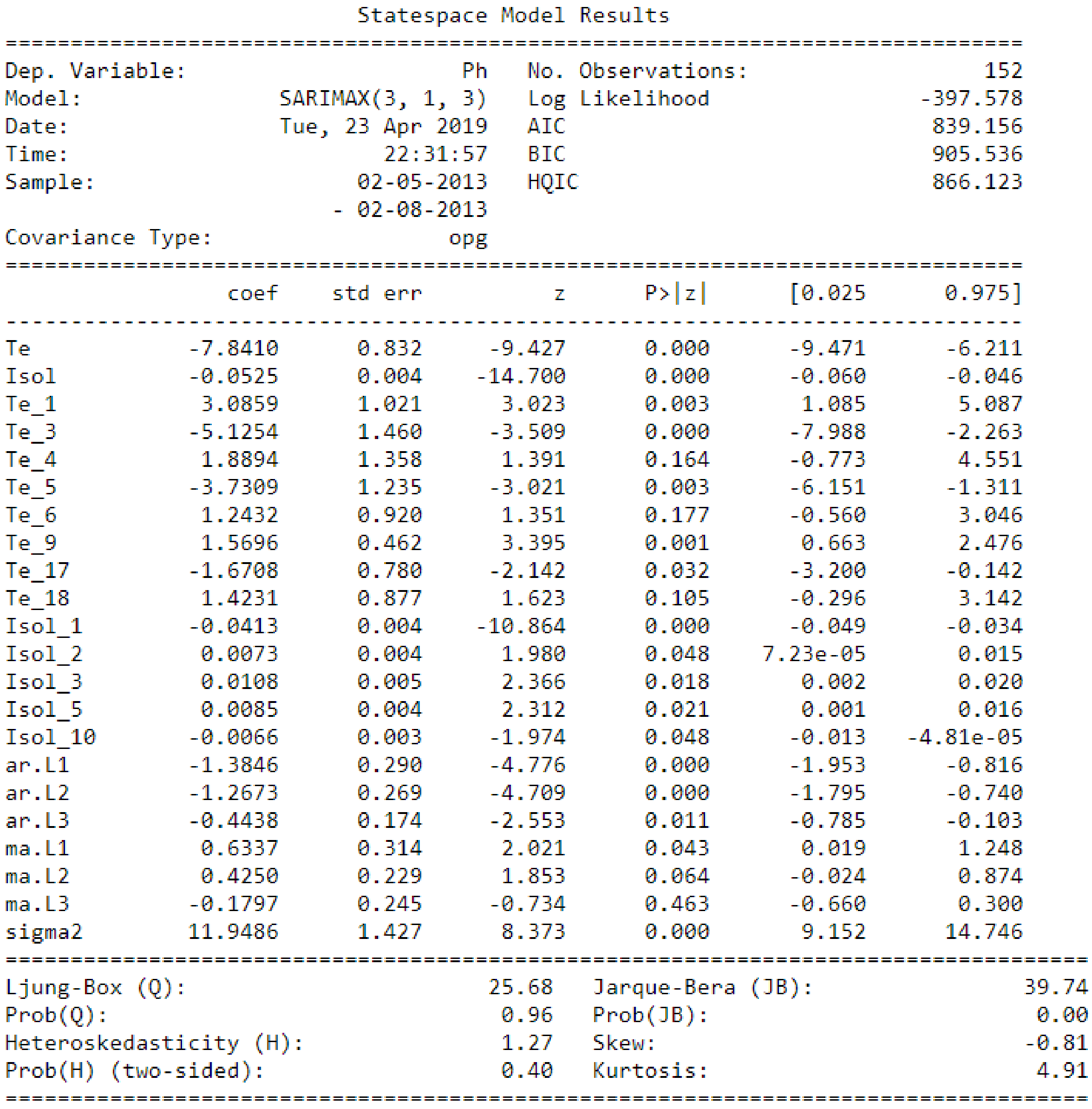


Figure 15: Decrease in the AIC by removing a feature backward selection in every step

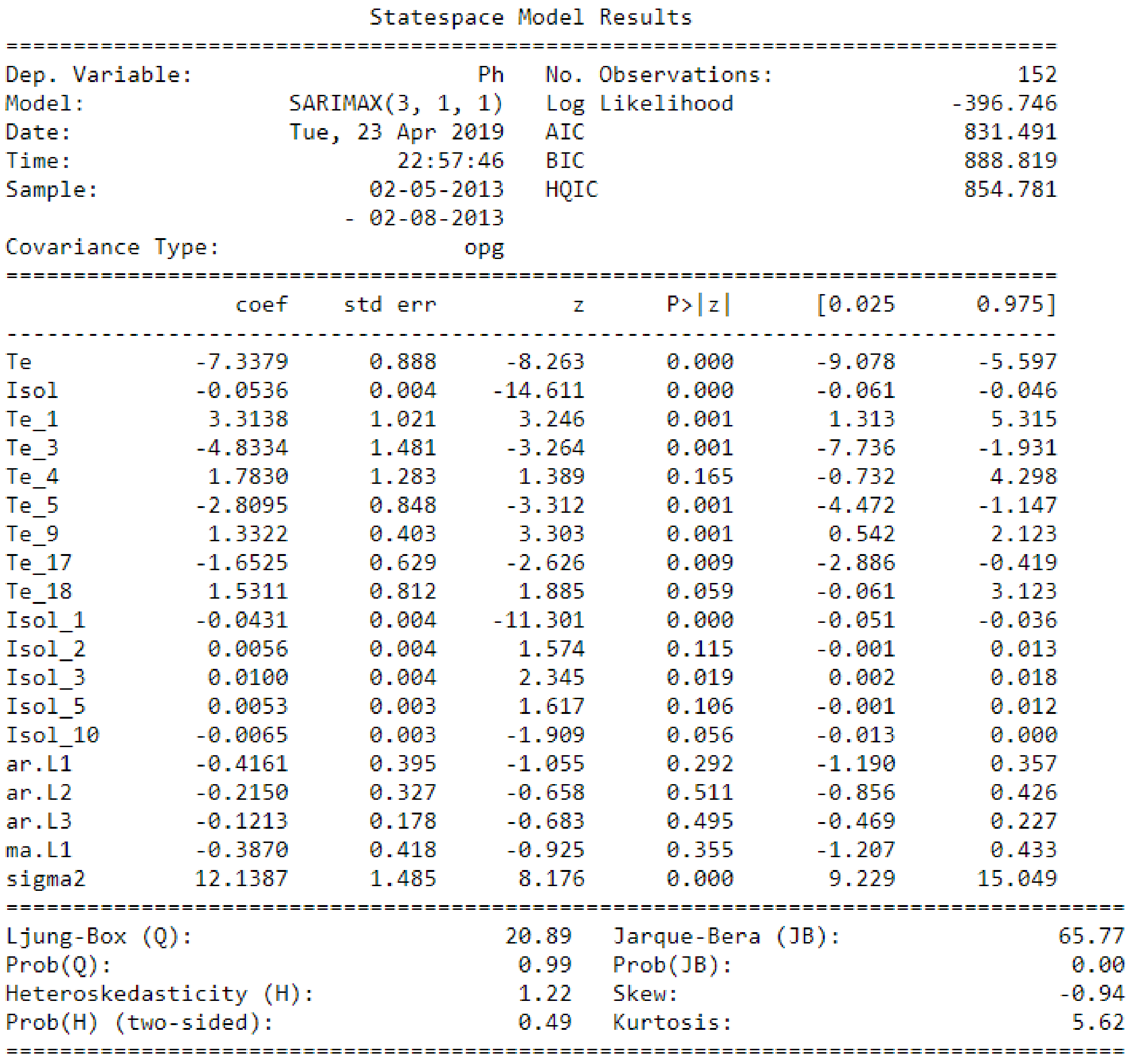
Observe the decrease in AIC score even though we do not select our features according to it. It is only a side effect that we want. After using this algorithm by starting with 20, we end up with the



```
                           Statespace Model Results
==============================================================================
Dep. Variable:                     Ph   No. Observations:                  152
Model:               SARIMAX(3, 1, 3)   Log Likelihood                -397.578
Date:                Tue, 23 Apr 2019   AIC                            839.156
Time:                        22:31:57   BIC                            905.536
Sample:                    02-05-2013   HQIC                           866.123
                         - 02-08-2013
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Te            -7.8410      0.832     -9.427      0.000      -9.471      -6.211
Isol          -0.0525      0.004    -14.700      0.000      -0.060      -0.046
Te_1           3.0859      1.021      3.023      0.003       1.085       5.087
Te_3          -5.1254      1.460     -3.509      0.000      -7.988      -2.263
Te_4           1.8894      1.358      1.391      0.164      -0.773       4.551
Te_5          -3.7309      1.235     -3.021      0.003      -6.151      -1.311
Te_6           1.2432      0.920      1.351      0.177      -0.560       3.046
Te_9           1.5696      0.462      3.395      0.001       0.663       2.476
Te_17         -1.6708      0.780     -2.142      0.032      -3.200      -0.142
Te_18          1.4231      0.877      1.623      0.105      -0.296       3.142
Isol_1        -0.0413      0.004    -10.864      0.000      -0.049      -0.034
Isol_2         0.0073      0.004      1.980      0.048    7.23e-05       0.015
Isol_3         0.0108      0.005      2.366      0.018       0.002       0.020
Isol_5         0.0085      0.004      2.312      0.021       0.001       0.016
Isol_10       -0.0066      0.003     -1.974      0.048      -0.013   -4.81e-05
ar.L1         -1.3846      0.290     -4.776      0.000      -1.953      -0.816
ar.L2         -1.2673      0.269     -4.709      0.000      -1.795      -0.740
ar.L3         -0.4438      0.174     -2.553      0.011      -0.785      -0.103
ma.L1          0.6337      0.314      2.021      0.043       0.019       1.248
ma.L2          0.4250      0.229      1.853      0.064      -0.024       0.874
ma.L3         -0.1797      0.245     -0.734      0.463      -0.660       0.300
sigma2        11.9486      1.427      8.373      0.000       9.152      14.746
===================================================================================
Ljung-Box (Q):                       25.68   Jarque-Bera (JB):                39.74
Prob(Q):                              0.96   Prob(JB):                         0.00
Heteroskedasticity (H):               1.27   Skew:                            -0.81
Prob(H) (two-sided):                  0.40   Kurtosis:                         4.91
===================================================================================
```

Figure 16: Summary table of the final model

We see that in figure 16 among the endegenous variables, there are insignificant ones. Then we can actually exclude the maximum p valued endegenous variables one by one as well.



```
                           Statespace Model Results                           
==============================================================================
Dep. Variable:                     Ph   No. Observations:                  152
Model:               SARIMAX(3, 1, 1)   Log Likelihood                -396.746
Date:                Tue, 23 Apr 2019   AIC                            831.491
Time:                        22:57:46   BIC                            888.819
Sample:                    02-05-2013   HQIC                           854.781
                         - 02-08-2013                                         
Covariance Type:                  opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Te            -7.3379      0.888     -8.263      0.000      -9.078      -5.597
Isol          -0.0536      0.004    -14.611      0.000      -0.061      -0.046
Te_1           3.3138      1.021      3.246      0.001       1.313       5.315
Te_3          -4.8334      1.481     -3.264      0.001      -7.736      -1.931
Te_4           1.7830      1.283      1.389      0.165      -0.732       4.298
Te_5          -2.8095      0.848     -3.312      0.001      -4.472      -1.147
Te_9           1.3322      0.403      3.303      0.001       0.542       2.123
Te_17         -1.6525      0.629     -2.626      0.009      -2.886      -0.419
Te_18          1.5311      0.812      1.885      0.059      -0.061       3.123
Isol_1        -0.0431      0.004    -11.301      0.000      -0.051      -0.036
Isol_2         0.0056      0.004      1.574      0.115      -0.001       0.013
Isol_3         0.0100      0.004      2.345      0.019       0.002       0.018
Isol_5         0.0053      0.003      1.617      0.106      -0.001       0.012
Isol_10       -0.0065      0.003     -1.909      0.056      -0.013       0.000
ar.L1         -0.4161      0.395     -1.055      0.292      -1.190       0.357
ar.L2         -0.2150      0.327     -0.658      0.511      -0.856       0.426
ar.L3         -0.1213      0.178     -0.683      0.495      -0.469       0.227
ma.L1         -0.3870      0.418     -0.925      0.355      -1.207       0.433
sigma2        12.1387      1.485      8.176      0.000       9.229      15.049
===================================================================================
Ljung-Box (Q):                       20.89   Jarque-Bera (JB):                65.77
Prob(Q):                              0.99   Prob(JB):                         0.00
Heteroskedasticity (H):               1.22   Skew:                            -0.94
Prob(H) (two-sided):                  0.49   Kurtosis:                         5.62
===================================================================================
```

Figure 17: Summary table of the final model after removing the insignificant MA parameters

Finally we check the residuals if they seems like a white noise.

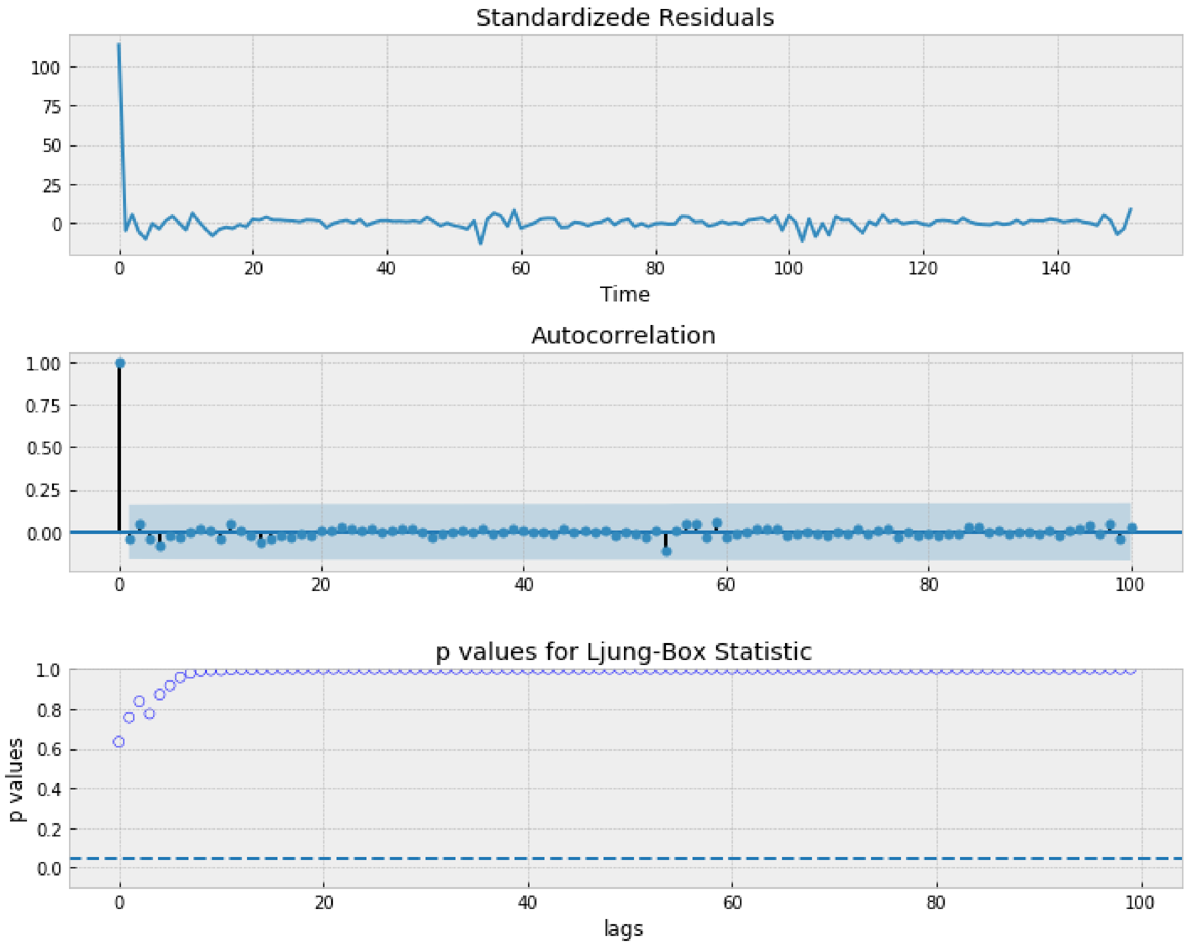


Figure 18: Final residual ACF

Everything seems fine. There is not any significant lagged residuals are autocorrelated.

2. Predict the four observations that were left out.

	1	2	3	4
Predictions	50.41	21.79	31.55	65.30
Original	42	10	30	62.4

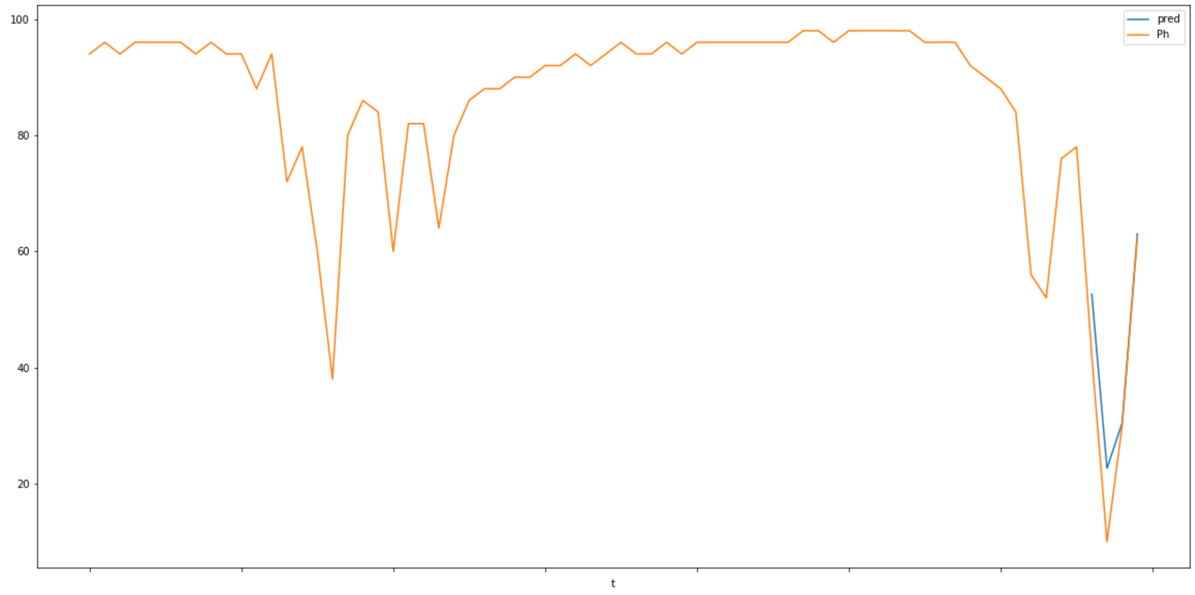


Figure 19: Final plot

Furthermore, we get even lower AIC score after removing the MA2 and MA3.

Question 3.5

When we incorporate the exogenous variables we saw a significant change in the predictions. The difference between figure 9 and figure 19 is obvious.