STAT443 Group 4 Project: Electricity Price in US City

Aaron Abraham, Diana Constantinescu, Lisa Kelly, Minhall Shen

Winter 2024

Contributions:

Minhall Shen: Exploratory Data Analysis + Combination + Conclusion

Diana Constantinescu: Regression + Combination

Lisa Kelly: Smoothing Methods + Combination

Aaron Abraham: Box-Jenkins + Combination

Contents

Motivation

In our project for STAT443, we will be exploring the monthly average electricity usage in kilowatt-hours in a U.S. city from November 1978 to January 2024 and attempting to forecast and predict the monthly average electricity usage over the next 24 months. We will leverage various forecasting techniques that were taught throughout the course to try and create a robust prediction.

This time series dataset about electricity prices stood out to us because it is a real world application that allows us to not only predict numbers and prices, but apply the knowledge we learned in the course to derive applicable insights affecting both the energy industry and homeowners alike. The data clearly suggests that there has been an increase in electricity prices throughout the years and we hope to provide analysis and predictions to help both homeowners and those in the energy industry understand the current and future patterns of electricity prices.

This dataset was obtained from FRED and can be found here: https://fred.stlouisfed.org/series/APU000072610. In the dataset, there are two columns, the date in the format YYYY-MM-DD, where the date is always the first of its respective month, and the cost of electricity per kilowatt-hour in U.S. dollars for that month. The data is not seasonally adjusted.

Exploratory Data Analysis

Let's look at the time series of our dataset

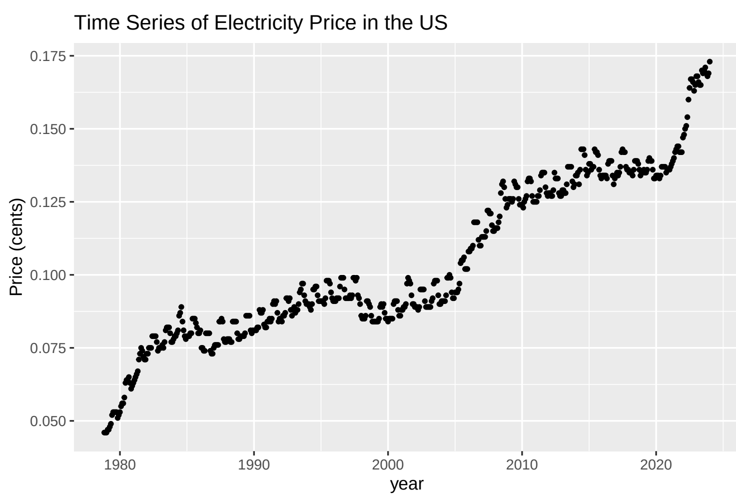

Fig 1: Time series of electricity price per kilowatt hour in the US

There is a missing value in the 83rd month of the dataset, we impute the NA with the mean of its direct neighbours (82nd and 84th months)

Non-constant Variance

From the time series, we see that we may not have constant variance. Let's do Fligner-Killeen tests to make sure. See "Appendix - Exploratory Data Analysis" for exact figures.

After performing a Fligner-Killeen test, wee see that the p-value « 0.05, so there is strong evidence against homogeneity of variances in our time series. We'll try to address this with some transformations.

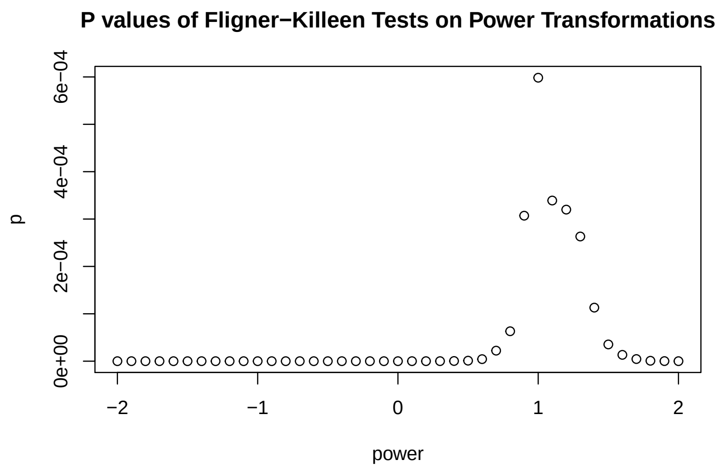

Fig 2: plot of the p−values for Fligner−Killeen Tests for powers from −2 to 2

From this graph, we see that the Fligner-Killeen test still yields very low p-values for all of the powers, so we are unable to get rid of the non-constant variance.

Seasonality and Trend

Next, let's explore seasonality and trend. Looking at the ACF, we see that there is linear decay, indicating trend. Performing a one time differencing, we see that there is a seasonality of 12.

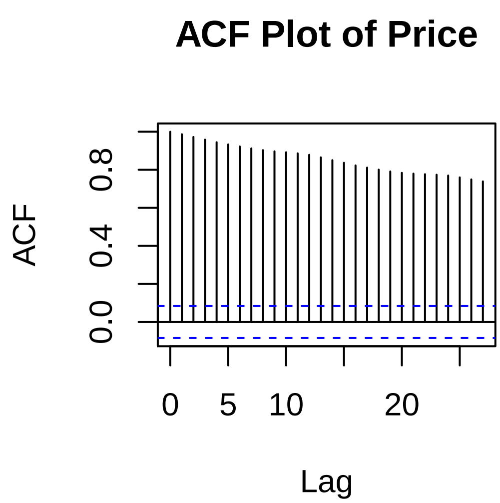

Fig 3: ACF Plot of price data

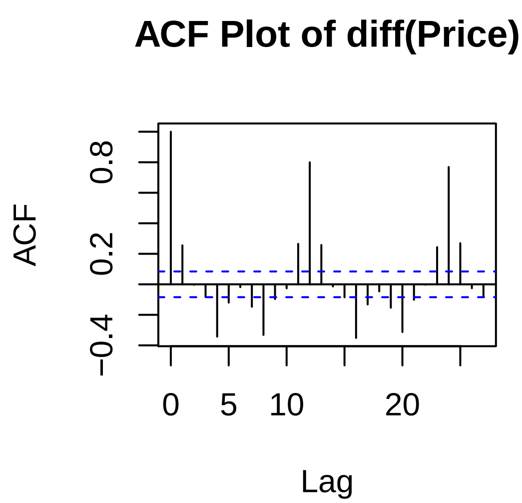

Fig 4: ACF Plot with one time difference

Regression

In this section, we will try regression models up to degree 15 on the data. We decided to use orthogonal polynomials to combat the multicolinearity problem. Details for this choice can be found in the appendix under "Appendix-Regression -> Choice of orthogonal polynomials". Figures 32 and 33 are relevant.

Additionally, from the explanatory analysis we determined that this data has trend and seasonality (with period 12) so we will try modelling using these components from the additive classical decomposition.

Further, we use the non-transformed data since attempts to transform the data did not solve the non-constant variance problem.

Non-regularized regression

First, we try non-regularized regression. We keep the last year of data for the test set and use the rest for training. The resulting APSE values can be seen in Figure 5 below. The dotted vertical red line indicates the degree at which APSE is minimized.

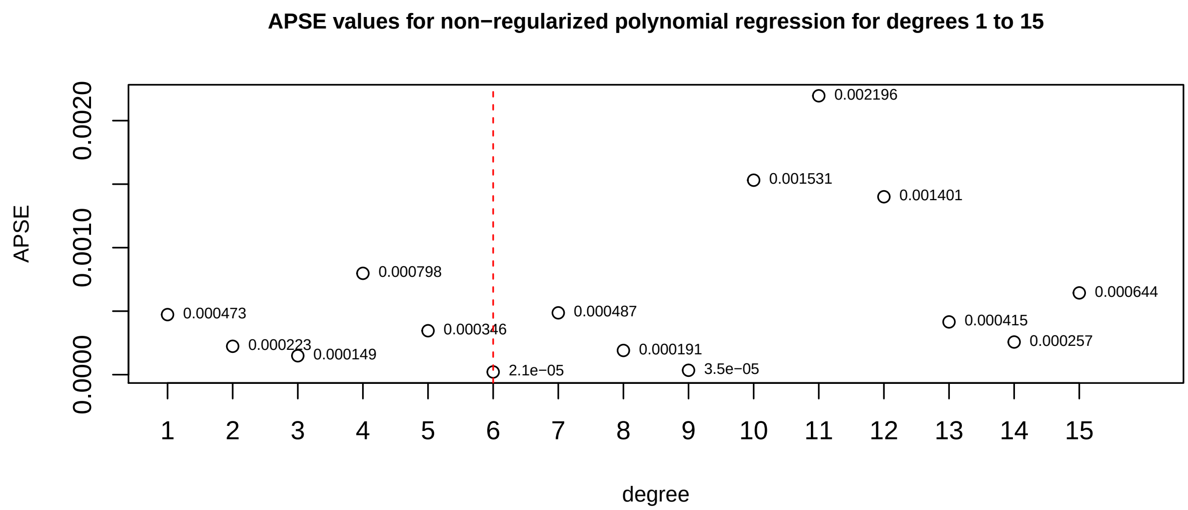

Fig 5: APSE values for non−regularized polynomial regression

The best model for prediction amongst the non-regularized polynomial regression models would be the degree 6 model with an APSE of 2.1281132×10^{-5}. Let's perform model diagnostics for this model to see if the assumptions are satisfied.

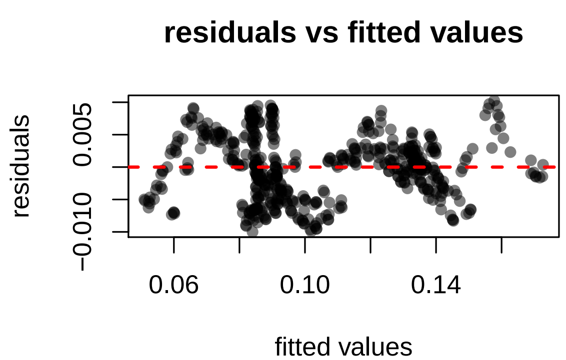

Fig 6: Residuals vs fitted values plot

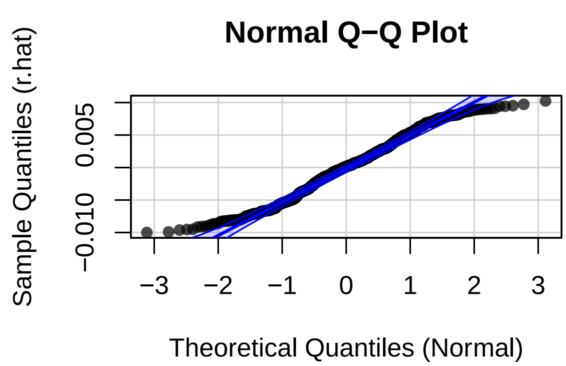

Fig 7: QQplot of residuals

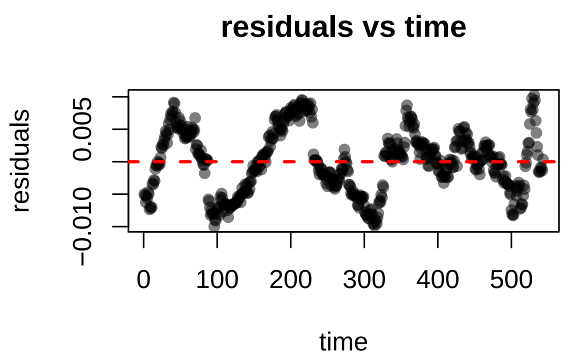

Fig 8: Residuals vs time plot

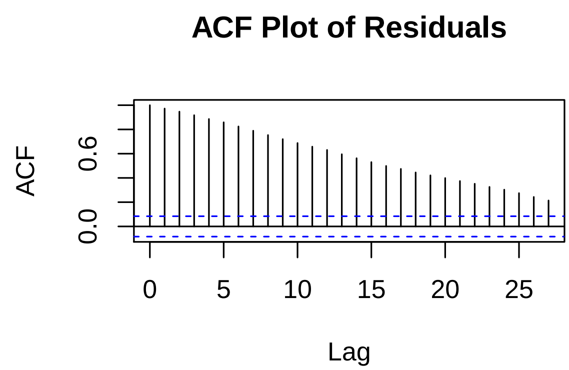

Fig 9: ACF plot of residuals

We notice a trend in Figure 6, which suggests non-constant mean. From Fig 7, the data appears to be non-normal since a lot of points fall outside the confidence bands. From Figure 8, we again see an up and down trend, suggesting non-constant mean. The linear decay in Figure 9 means there is still trend in the residuals. Non-graphical diagnostics were also checked, and details can be found under "Appendix-Regression -> Non-graphical residual diagnostics for all regression models". Figures 47, 48 and 49 are relevant. The conclusions of those tests agree with the ones included here. Thus, the residuals are not stationary, not normal and not random. The assumptions are violated so we cannot use this model for prediction.

Regularized Regression

We try using regularization. We again go up to degree 15 polynomials and we try ridge, elastic net (alpha = 0.25, 0.5, 0.75) and LASSO models. We have calculated CV errors for each of these models and concluded:

The best model for alpha=0 is the degree 13 model, with CV error of 4.5264782×10^{-6}. The best model(s) for alpha=0.25, 0.5, 0.75 and 1 are the degree 15 models, with CV errors of 4.5227208×10^{-6}, 4.5225778×10^{-6}, 4.5222862×10^{-6}, and 4.5225638×10^{-6}, respectively. The CV plots are under "Appendix-Regression -> CV Error plots for regularized models". Refer to figures 34, 35, 36, 37 and 38. We fit these models on the training set with their corresponding optimal lambdas and then predict on the test set to calculate the APSEs.

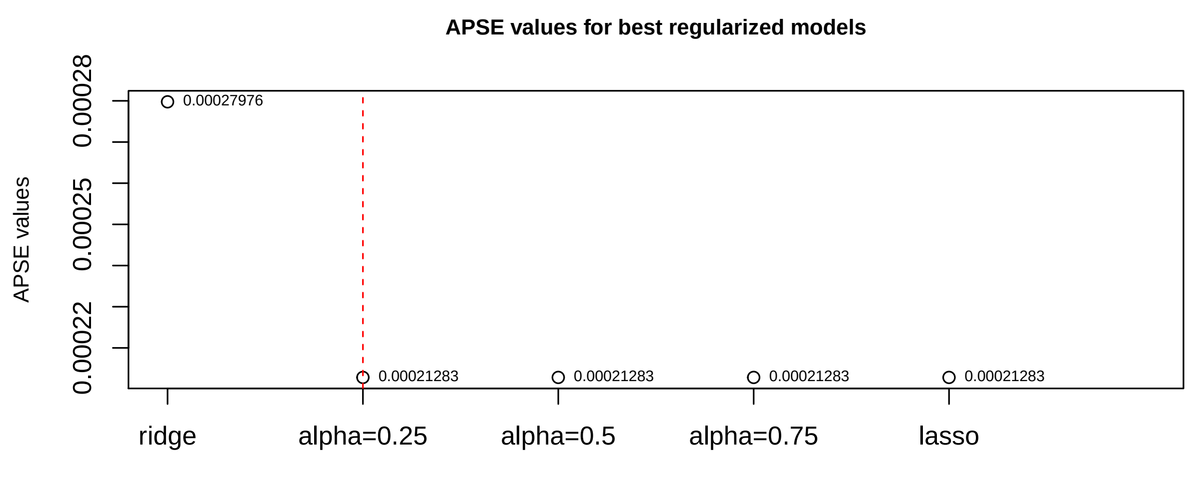

Fig 10: APSE values for best regularized models

According to the APSEs, the best regularized regression model(s) are degree 15 models for alpha=0.25, 0.5, 0.75 and 1.

For brevity, we do not include the details of the model diagnostics for the degree 15 model with alpha=0.25, although they can be found under "Appendix-Regression -> Graphical Residual diagnostics for regularized models" and "Appendix-Regression -> Non-graphical residual diagnostics for all regression models". Figures 39-42 and 50-52 are relevant. We simply state the results of these diagnostics, which is that model assumptions are violated. The conclusion is similar for the degree 15 models with alpha values 0.5, 0.75 and 1 (details in the same appendix section mentioned above). So these models should not be used for prediction.

Although ridge regression model has a higher APSE than these models, we check its residuals as well, hoping that they look better. For brevity, we only state the conclusion that the assumptions are violated for this model as well, but the details are under the same appendix section mentioned above.

Therefore, none of these regularized regression models should be used for prediction.

However, the project asks for us to provide a forecast using the final model, so we compare the APSE values of the best regularized model to the best non-regularized model, and find that the non-regularized polynomial regression model of degree 6 has the smallest APSE (value of 2.1281132×10^{-5}).

We use this model to provide a forecast. However, this forecast should not be used or trusted as assumptions for all these models are violated!

Forecast

Let's forecast the next 24 months. As stated before, this will not be an accurate forecast and should not be trusted as the model assumptions are violated. In particular, since we have non-normal results, the prediction interval is not trustworthy.

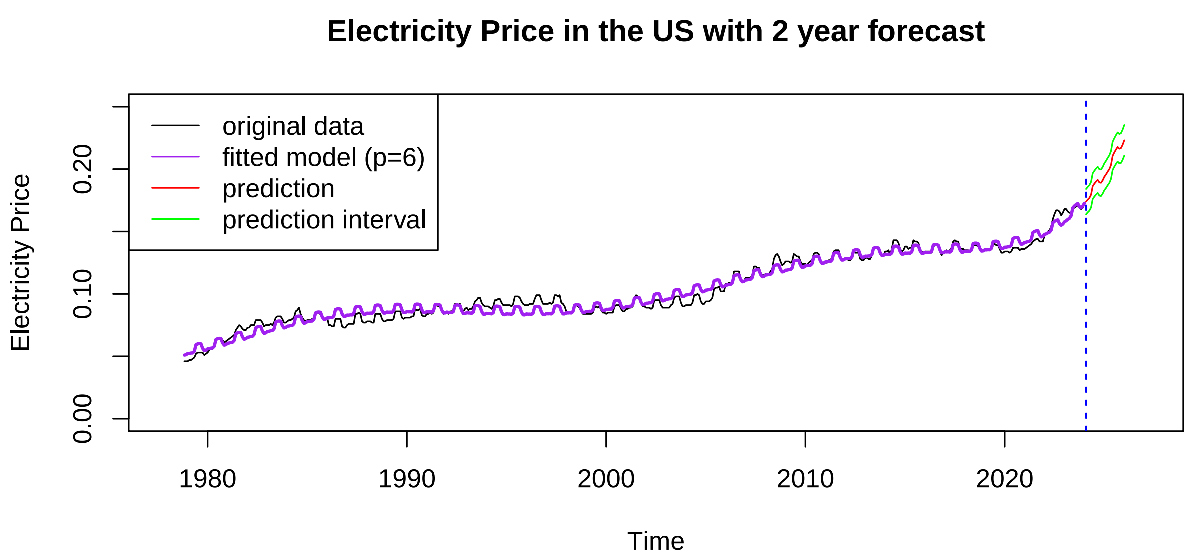

Fig 11: 24−month forecast using best regression model

Overall, polynomial regression models are not great predictive models for this data set. We will now explore other methods/models to try and find a better forecasting model.

Smoothing Methods

We now wish to explore alternative smoothing methods to regression. Since our data has both trend and seasonality, we will investigate Holt-Winters and Differencing as alternative smoothing methods to regression.

Holt-Winters

We will choose the “best” Holt Winters model using the test/train set APSE criterion

Table 1: APSE values of Holt-Winters Models

HWMethod	APSE
1. Simple Exponential Smoothing	5.60e-06
2. Double Exponential Smoothing	4.10e-06
3. Holt-Winters (Additive)	6.09e-05
4. Holt-Winters (Multiplicative)	9.75e-05
5. Season Only (Multiplicative)	5.90e-06
6. Season Only (Additive)	7.50e-06

In reference to Table 1, the smallest APSE value corresponds to the Holt Winters model with trend only (Double Exponential Smoothing). We will examine the plot of the chosen model with its predictions on the test set and also examine the residuals to check for stationarity. Details of other Holt Winters models can be found in the Holt Winters section of the Appendix.

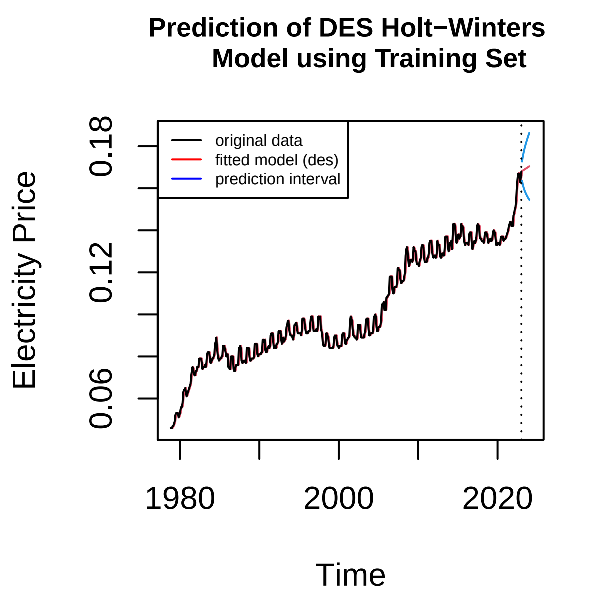

Fig 12: Prediction of DES Holt−Winters Model using Training Set

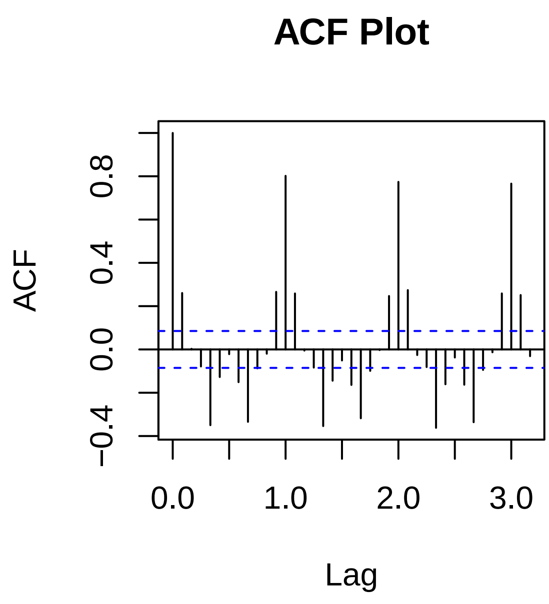

Fig 13: ACF plot of DES Holt Winters Model

Figure 13 shows the existence of slow decay in the lag of season visible in the ACF plot, hence the residuals of our chosen model are non-stationary.

Forecasting for the next 24 months:

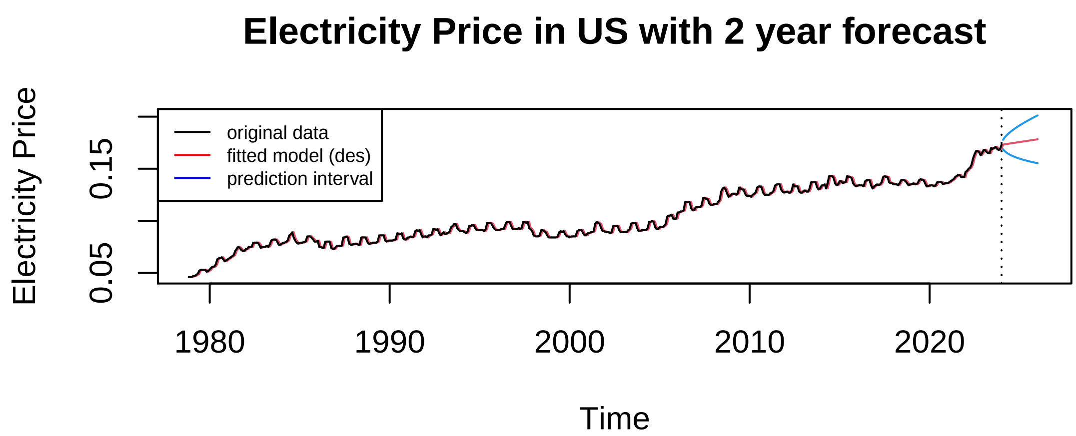

Fig 14: Electricity Price in US with 2 year forecast

The prediction intervals of our plot are quite wide in Figure 14 and our trend only additive model does not take into account the non-stationary seasonal nature of our data, therefore long-term prediction is more than likely unsuitable for this model.

The reason this Holt-Winters model in particular was chosen could be due to the nature of our data and a drastic change point after 2020, which indicates the sudden increase in electricity price, possibly due to the knock on effects of the global pandemic caused by Covid-19.

Differencing

Differencing can be used for trend and/or seasonal elimination. We will try a series of different differencing methods to make our data stationary., such as differencing in lag one (regular differencing), differencing in the lag of season (lag 12, known as seasonal differencing).

One Time Regular Differencing

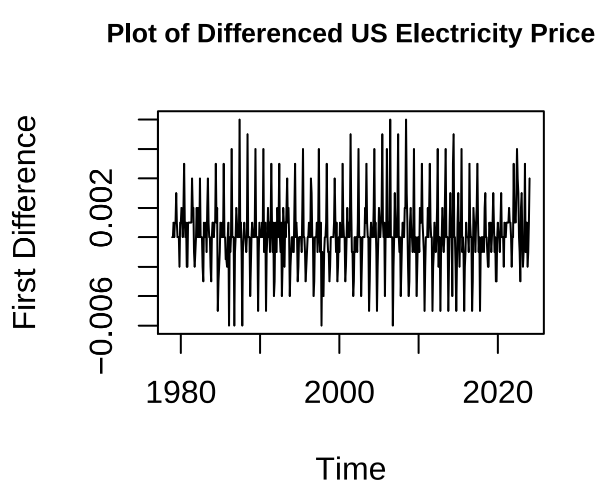

Fig 15: US Electricity Price (One Time Regular Differencing)

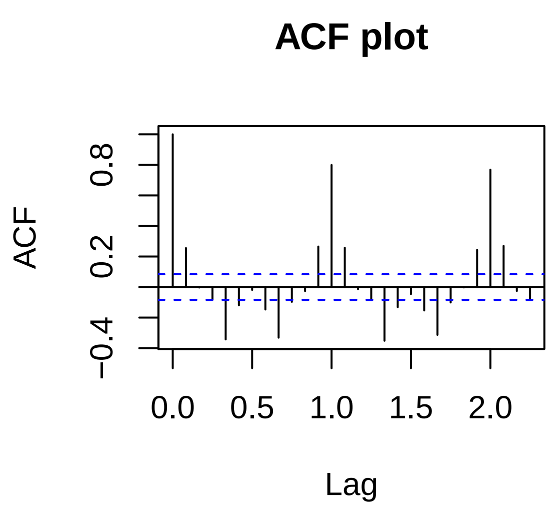

Fig 16: ACF Plot (One Time Regular Differencing)

There is presence of linear decay in the lag of season in the above ACF plot, Figure 16, so the data is non-stationary.

One Time Seasonal Differencing in Lag 12

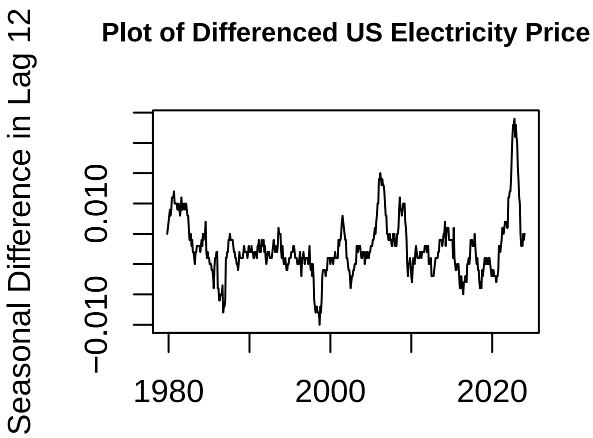

Fig 17: US Electricity Price (One Time Seasonal Differencing)

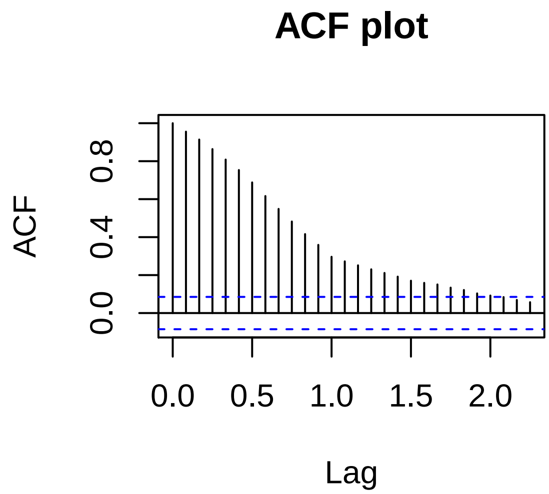

Fig 18: ACF Plot (One Time Seasonal Differencing)

There is exponential decay clearly visible in the ACF plot, Figure 18, so the data is stationary but correlated.

Combination of Regular and Seasonal Differencing

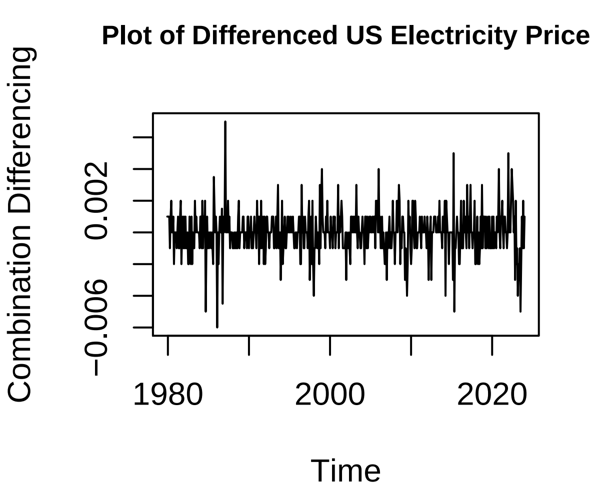

Fig 19: US Electricity Price (Comibination Differencing)

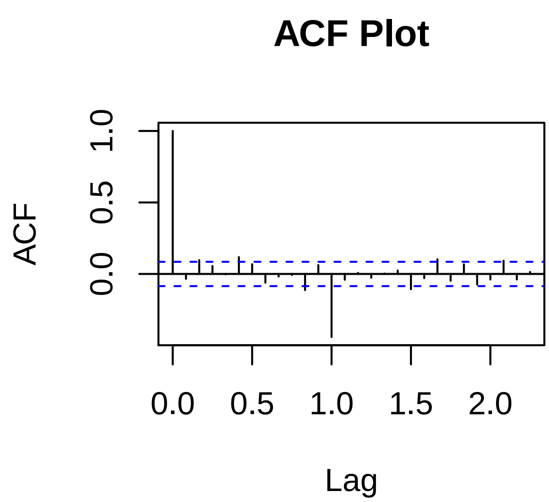

Fig 20: ACF Plot (Combination Differencing)

There is no presence of linear decay or other indicators of non-stationarity in the above ACF Plot, Figure 20, therefore the data is stationary, however we may have overdifferenced here as we achieved stationarity by differencing once in the lag of season i.e. lag 12.

Overdifferencing leads to increased variability which we want to avoid, but it allows for simpler models to be proposed in the Box-Jenkins section below.

Box-Jenkins modelling

From the differencing section of this project, we note that both one-time seasonal differencing and a combined one-time seasonal plus one-time differencing on a lag of 1 can be used to generate SARIMA models. The ACF and PACF plots for both differenced time series is displayed in Figure 21.

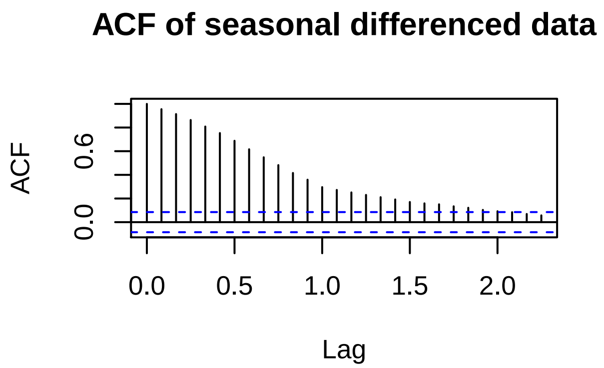

Figure 21A: ACF plot of seasonal differenced data

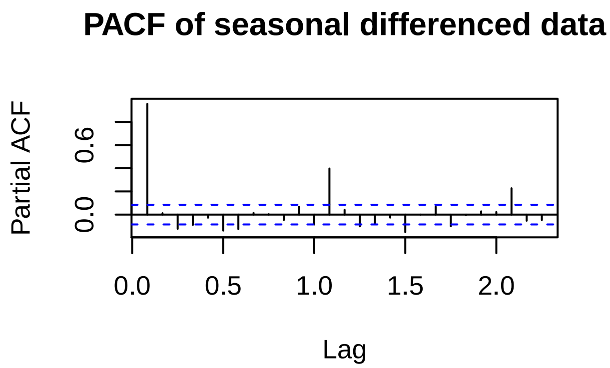

Figure 21B: PACF plot of seasonal differenced data

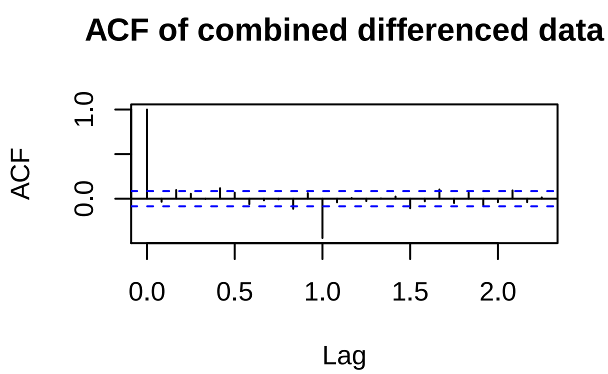

Figure 21C: ACF plot of combined differenced data

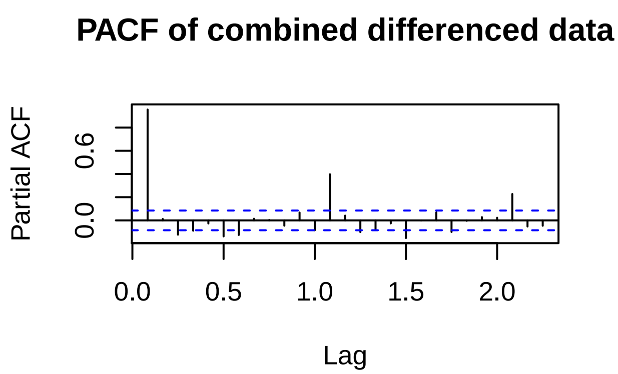

Figure 21D: PACF plot of combined differenced data

We will continue working with the combined differenced data. Modelling on the seasonally differenced data can be found in the appendix section titled "Seasonal differenced SARIMA".

Model proposing

We can see in from Figures 21C and 21D that the ACF cuts off at lag 1 and the partial ACF plot cuts off at around lag 2. Also note that the ACF plot of the data squared does not exhibit any trends, indicating that there is constant variance (can be found in the appendix section titled "Variance analysis of combined difference data")

Since we are working off data that has been one-time differenced on a lag of 1 and a one-time difference on a lag of season (lag = 12), we have d = 1, D = 1 and s = 12. Since we used seasonal and one-time differencing, we will propose a few different SARIMA models (Models 2-10 can be found in the appendix section titled "Additional SARIMA model proposals":

- Proposal 1: $SARIMA(0,1,5) \cdot (0,1,1)_{12}$: If we remove the seasonal lags, we can see that the ACF plot cuts off to 0 after lag 5 (justifying q = 5) and the PACF plot is exponentially decreasing (justifying p = 0). If we only look at the seasonal lags, the ACF plots off after the lag = 12 (justifying Q = 1) and the PACF plot is exponentially decreasing (justifying P = 0).

- Proposal 11: $SARIMA(2,1,1) \cdot (2,1,1)_{12}$: We can interpret both ACF and PACF plots as exponential decay across both regular and seasonal lags, so ARMA modelling can be used. We simply vary p, q, P and Q for this model.
- Proposal 12: $SARIMA(1,1,2) \cdot (2,1,1)_{12}$: Varying the ARMA model parameters for both ARIMA models.

Model fitting

We will fit our models on all data from 1978 to 2023 and keep the 2023-2024 data for testing. We will look at AICc as a measure of fit across all 12 model proposals and select the top 3 models for further investigation

Table 2: Top 3 AICc values and corresponding models

	Model	AICc
1	sarima_model_1	-10.69478
11	sarima_model_11	-10.69379
12	sarima_model_12	-10.69370

The 3 models with the lowest AICc are:

- $SARIMA(0,1,5) \cdot (0,1,1)_{12}$
- $SARIMA(2,1,1) \cdot (2,1,1)_{12}$
- $SARIMA(1,1,2) \cdot (2,1,1)_{12}$

Let us now check the residuals to make sure that model assumptions are not violated

Model residual analysis

For Model 1, we have:

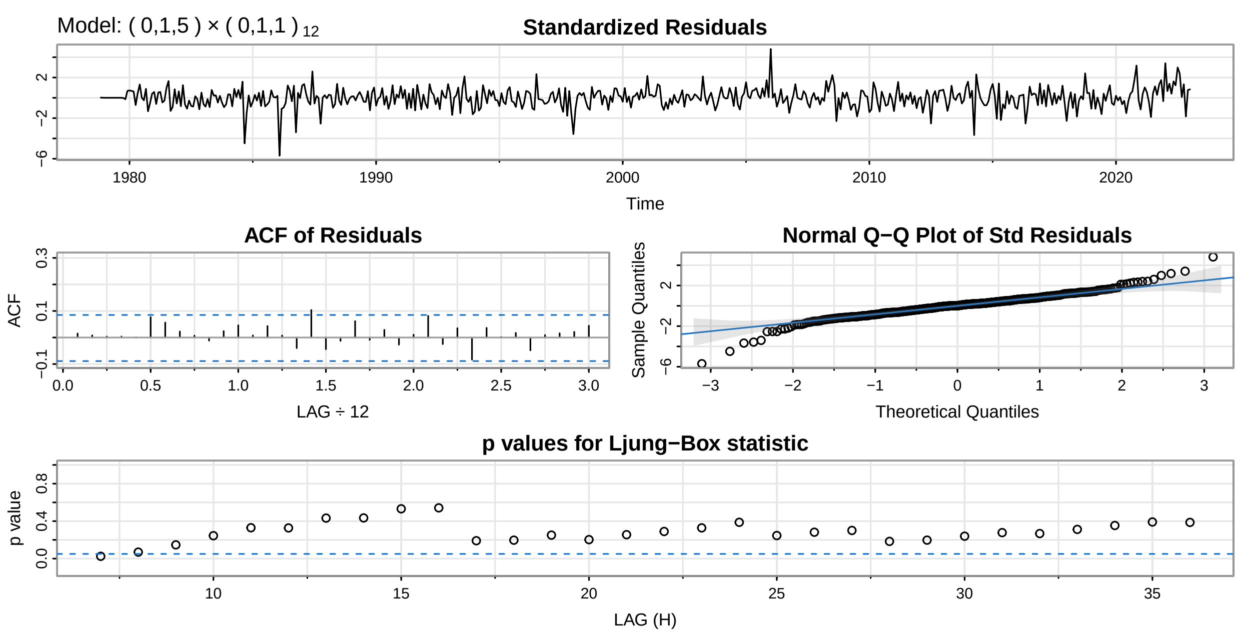

Figure 22: Model 1 residual plot

There are a couple of things to point out from the residual plots in Figure 22:

- The time series of the residuals looks quite random with no significant trend or seasonality present through visual analysis
- There are no major spikes in the ACF across lags, indicating that there is no autocorrelation in the residuals
- For the most part, the residuals do look normally distributed from the Q-Q plot, although there are some heavy tail data points
- Most of the p-values for the Ljung-Box test are above 0.05, indicating that there is not much evidence to suggest that the residuals are not independently distributed

You can see the residual plots for models 11 and 12 in the appendix section titled "Model 11 and 12 model diagnostics".

Model forecasting

Let's forecast each model on the test dataset, which is a 12-step ahead prediction. APSE scores are as follows:

Table 3: APSE scores for SARIMA models

Model	APSE
Model1	2.62e-05
Model11	7.48e-05
Model12	7.23e-05

From the APSE results, we can see that Model 1 has the lowest APSE score on the testing dataset. Let's use this model to forecast for the next two years using the entire dataset.

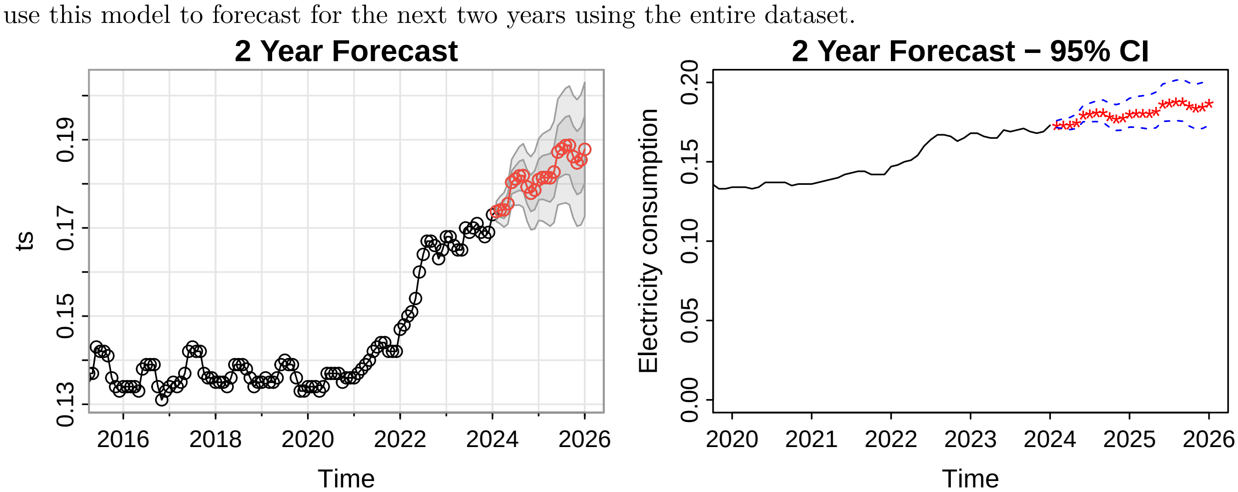

Figure 23: 2 year forecast using SARIMA model 1 and 95% prediction interval

Combination

Looking at combination, we chose to stack Box-Jenkins models on top of the degree 6 regression model. We used both the seasonal differencing and one-time differencing of the residuals for the model as the ACF plot showed indication that they were non-stationary (see Fig 76 and Fig 77 in Appendix-Combination). The two models were selected using the lowest AICc scores when fit on the test set, and they were:

- SARIMA(6, 0, 0)x(0, 1, 1)_12
- SARIMA(5, 1, 0)x(2, 0, 0)_12

For the first model, d=0, D=1 and s=12 were chosen from differencing. We choose p=6 as PACF cuts off after lag 6 and ACF has exponential decay. We choose Q=1 as the ACF cuts off after the first season and PACF is exponentially decaying.

For the second model, d=1, D=0 and s=12 were chosen from differencing. We choose p=5 as PACF cuts off after lag 5 and ACF has exponential decay. We choose P=2 because one can argue that the PACF cuts off after the second season.

Let's take a look at the forecasts:

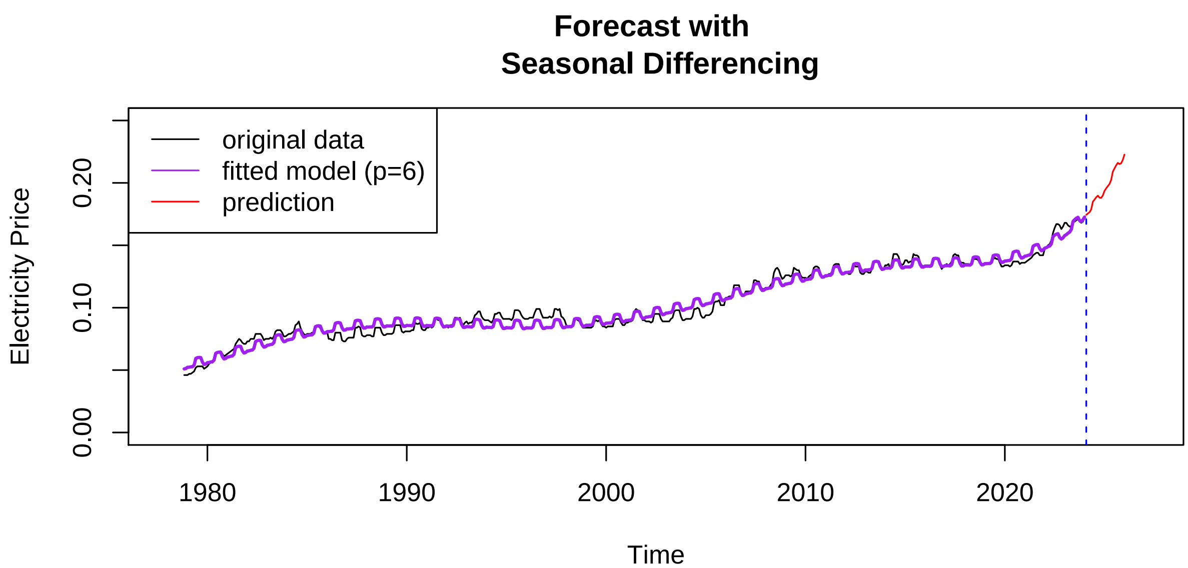

Fig 24: SARIMA Model stacked on seasonal differencing of regression model

The model based on seasonal differencing had an APSE of 6.77572e-05 and the one based on one-time differencing had an APSE of 1.12479e-04, which are not as good as model 1 yielded from Box-Jenkins.

Conclusion

The model with the best forecast based on the lowest APSE score was the Double Exponential Smoothing model with an APSE of 4.137981e-06. However, since the residuals were found to be non-stationary, the SARIMA(0,1,5)x(0,1,1)12 model that was discovered in the Box-Jenkins section could be an alternative with a higher APSE but stationary residuals.

Referring to all of the forecasts that were done throughout the report (Fig 11, 14, 23 and 24), we can see that electricity prices in the next two years are predicted to increase along with some seasonality. These patterns are important information crucial for homeowners to assist in their financial planning, make energy efficient investments in their appliances and policy makers make informed decisions to implement energy efficient programs. Rising electricity prices may ultimately incentivize businesses and homeowners to reduce their carbon footprint by using less electricity or investing in renewable energy sources.

Appendix

Exploratory Data Analysis

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  data$PRICE and seg
## Fligner-Killeen:med chi-squared = 19.602, df = 4, p-value = 0.0005983
```

Fig 25: Fligner-Killeen test of electricity price data

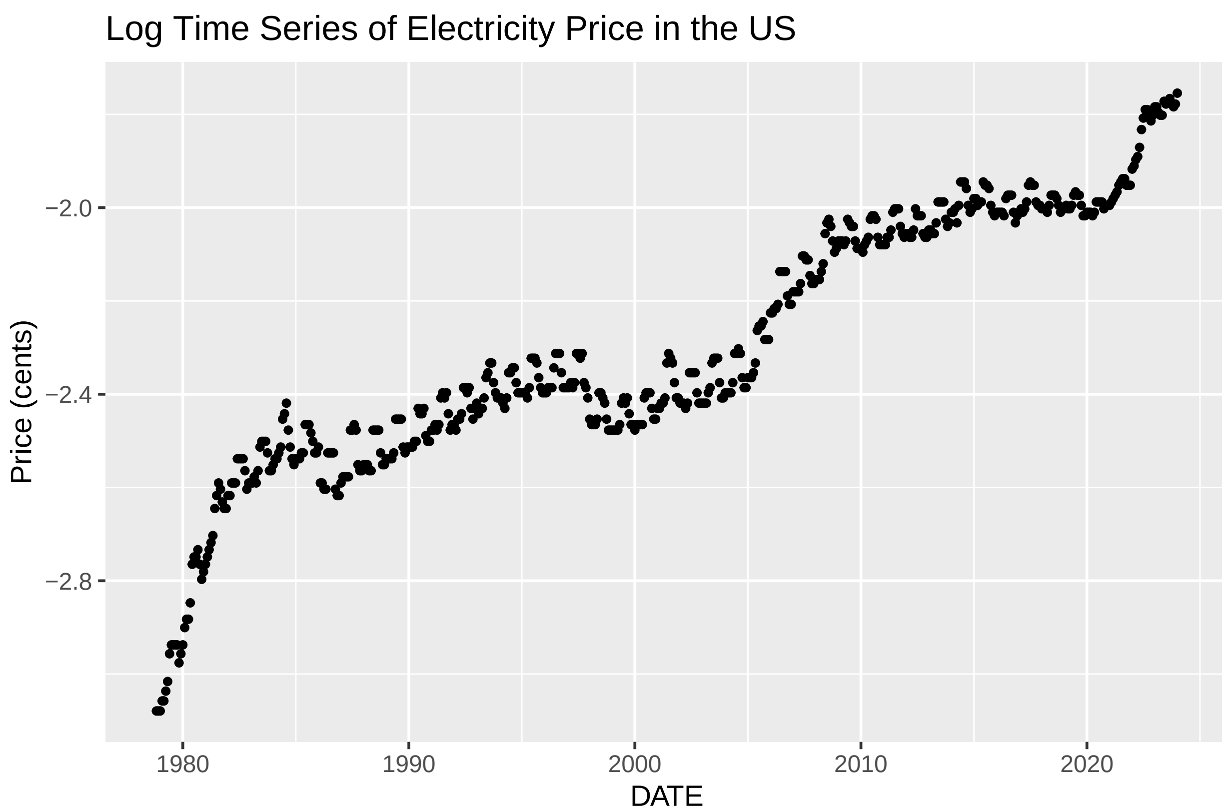

Fig 26: Log transformed time series of electricity price per kilowatt hour in the US

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  log(data$PRICE) and seg
## Fligner-Killeen:med chi-squared = 46.408, df = 4, p-value = 2.025e-09
```

Fig 27: Fligner-Killeen test of log transformed electricity price data

```
##    Month     Price
## 1      1 0.1027174
## 2      2 0.1011556
## 3      3 0.1014222
## 4      4 0.1014000
```

```
## 5        5 0.1026000
## 6        6 0.1079111
## 7        7 0.1084444
## 8        8 0.1085556
## 9        9 0.1083000
## 10      10 0.1045778
## 11      11 0.1014783
## 12      12 0.1016739
```

Fig 28: Table of average price per month

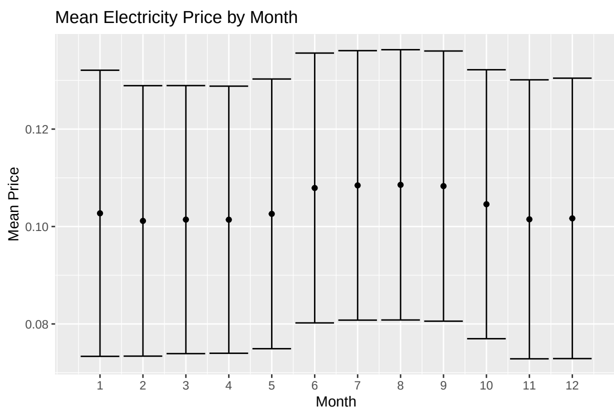

Fig 29: Average electricity price for all the months with a 95% confidence interval

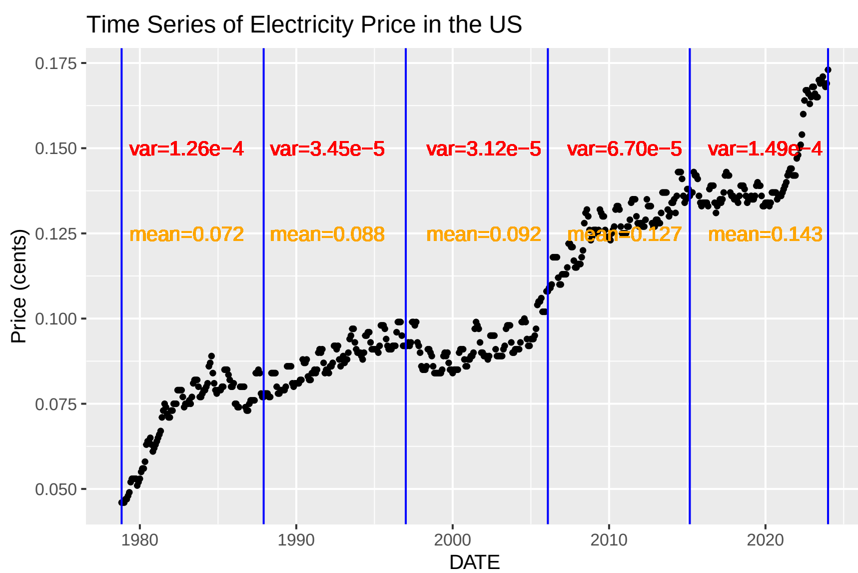

Fig 30: Dividing the time series into 5 equal sections and checking each one's variance and mean

```
## No summary function supplied, defaulting to ‘mean_se()‘
```

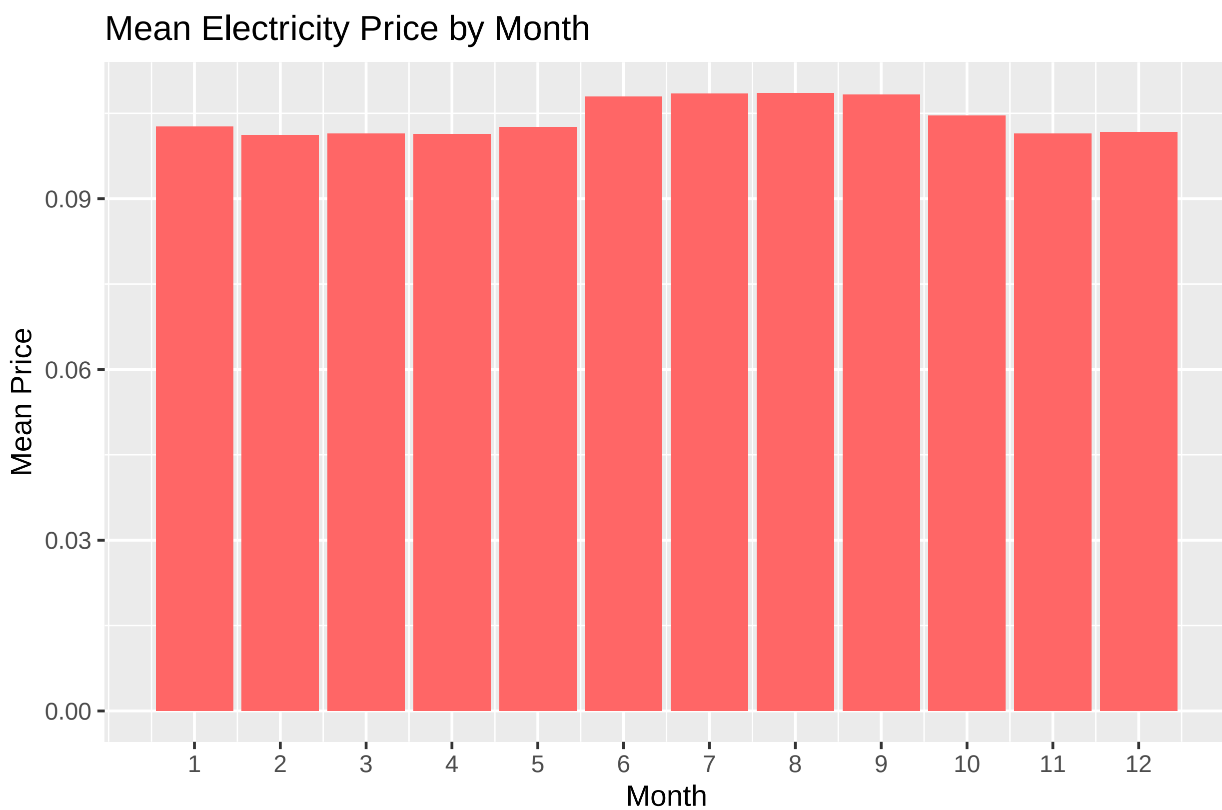

Fig 31: Average electricity price for all the months

Regression

Choice of orthogonal polynomials

We check correlation of regular polynomials:

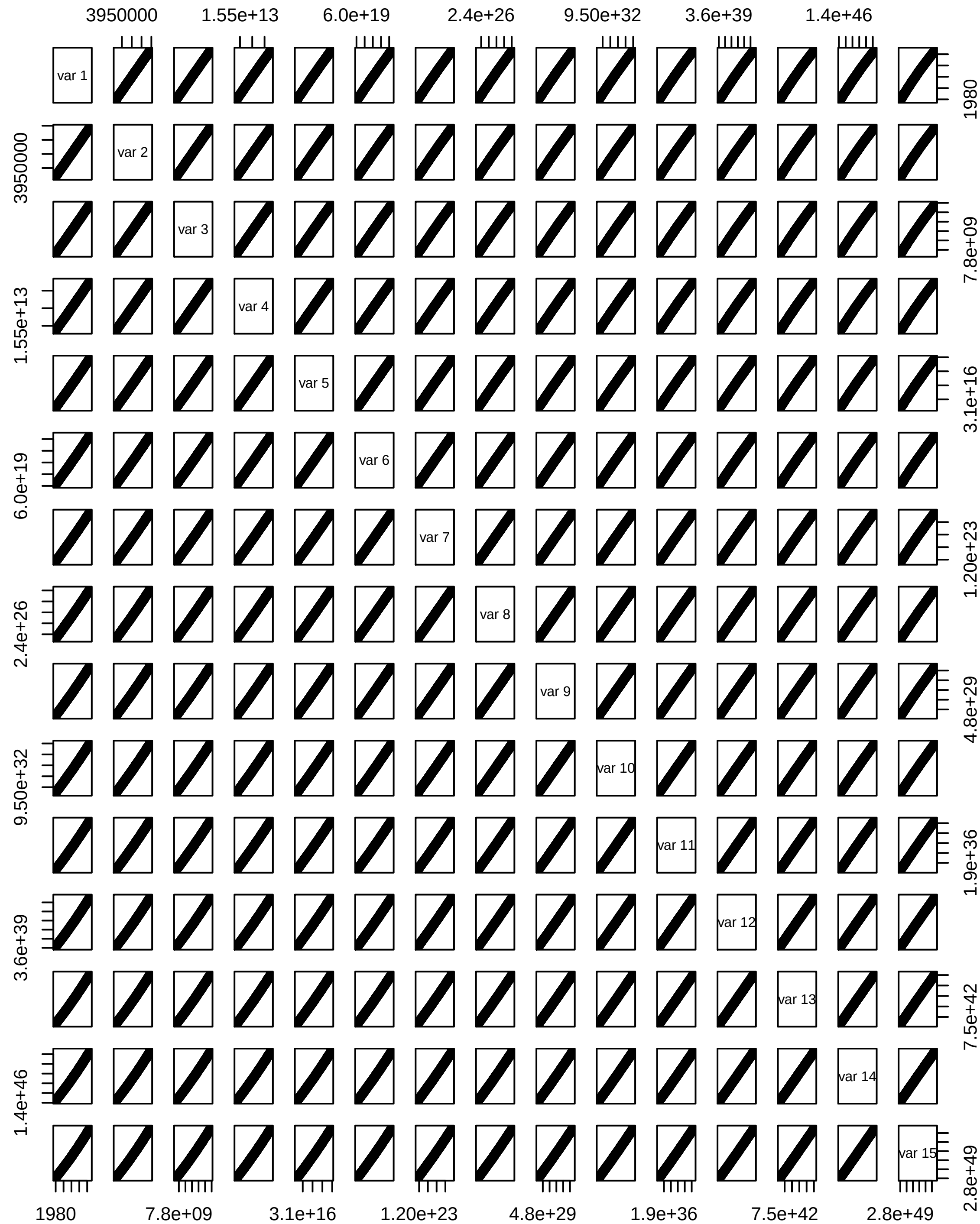

Fig 32: Correlation plot of regular polynomials

From the correlation plot, we can see that there is high linear correlation between the columns of the design matrix when using regular polynomials, so let's try orthogonal polynomials and check the correlation of their corresponding design matrix columns.

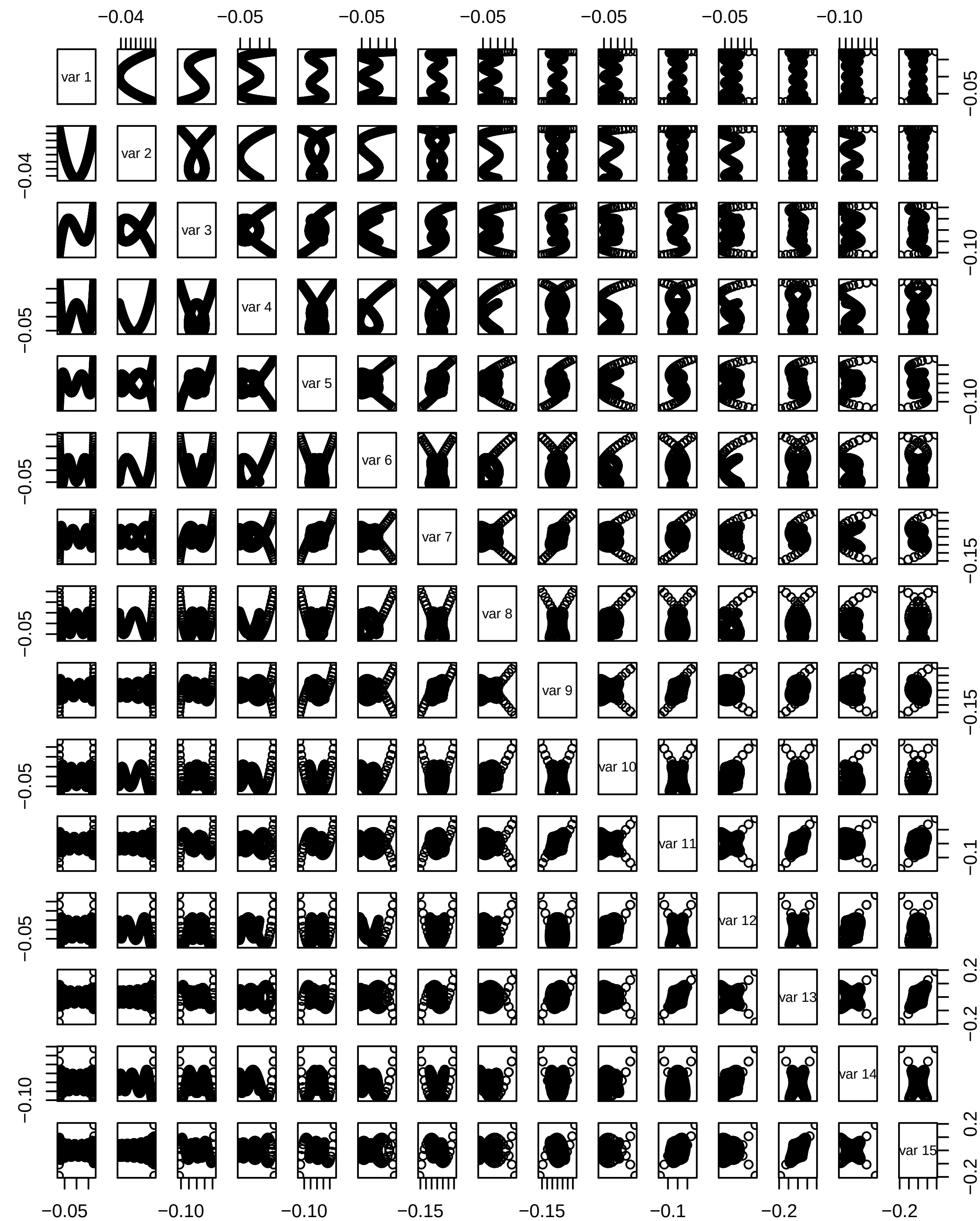

Fig 33: Correlation plot of orthogonal polynomials

From the correlation plot, there is no evidence of linear relationship between the columns of the design matrix when using the orthogonal polynomials.

Therefore, orthogonal polynomials were used throughout the regression portion of this analysis.

CV Error plots for regularized models

We plot the CV errors vs the degree for each alpha.

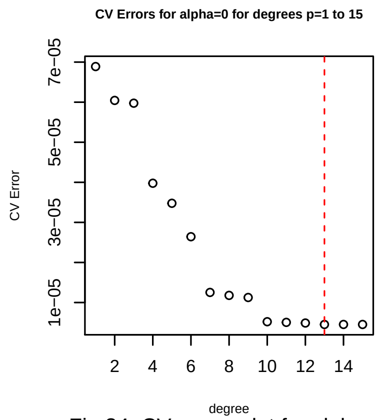

Fig 34: CV errors plot for alpha=0

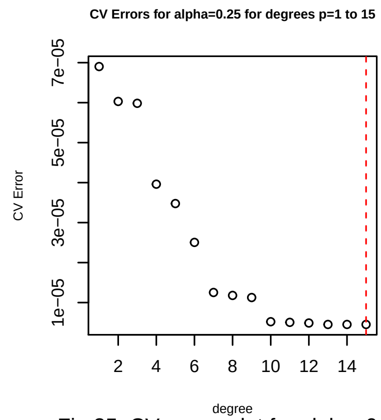

Fig 35: CV errors plot for alpha=0.25

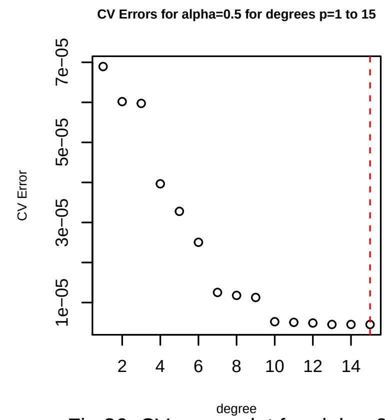

Fig 36: CV errors plot for alpha=0.5

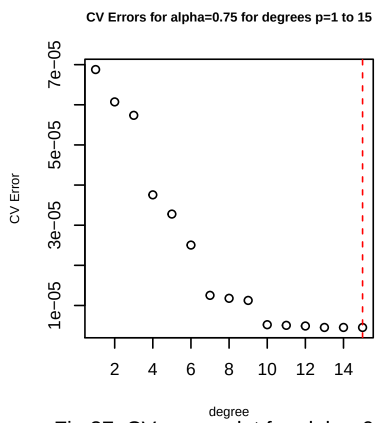

Fig 37: CV errors plot for alpha=0.75

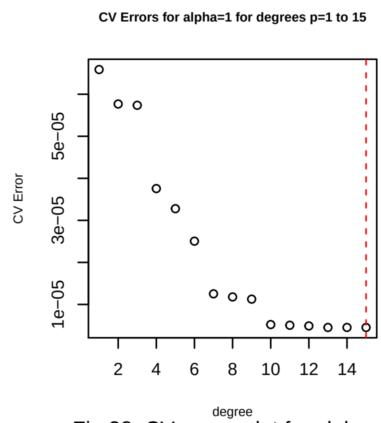

Fig 38: CV errors plot for alpha=1

Graphical Residual diagnostics for regularized models

Here are the graphical residual diagnostics for the degree 15 model with alpha=0.25:

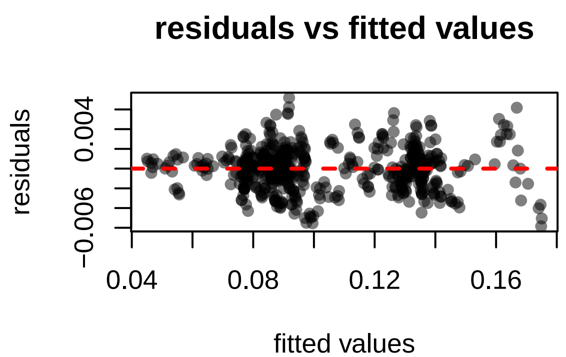

Fig 39: Residuals vs fitted values plot

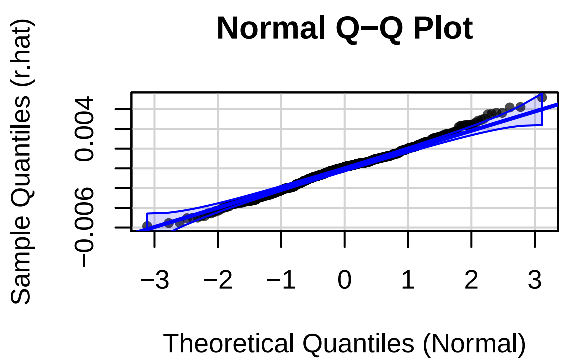

Fig 40: QQplot of residuals

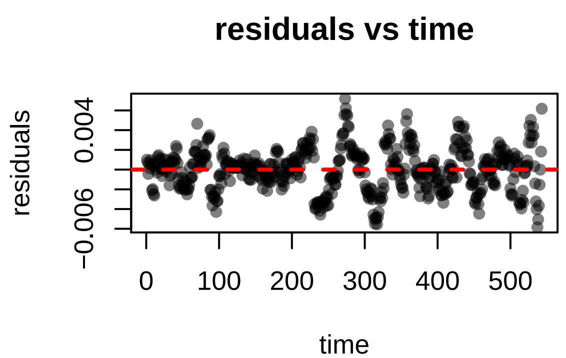

Fig 41: Residuals vs time plot

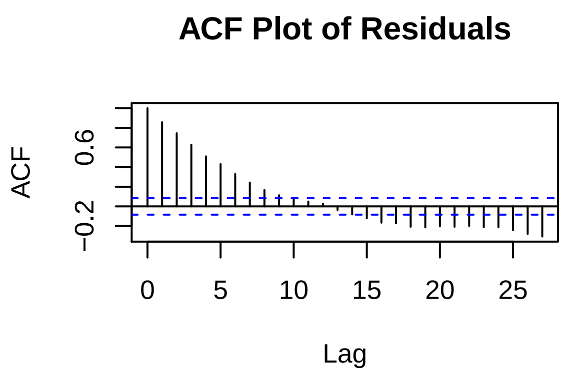

Fig 42: ACF plot of residuals

I notice a fanning out shape on the residuals vs fitted values plot, suggesting non-constant variance of the residuals.

There are a lot of data points outside of the confidence band in the QQplot, suggesting the residuals may not be normal.

There is some fanning of the residuals in the residuals vs time plot, suggesting non-constant variance.

We see what appears to be some linear decay in the ACF plot, meaning there is some trend left and the residuals are not stationary.

We note that in the alpha=0.25, alpha=0.5, alpha=0.75 and alpha=1 cases, the APSE was minimized at degree 15, with optimal lambda=0. This implies that the chosen alpha value is meaningless and the model is the same in all of these cases. In particular, this means that the residuals are the exact same across these models and since the diagnostics were already performed, they are not repeated. We confirm that the residuals of all these models are identical:

```
which(res.025-res.05 != 0) # all 0
```

```
## integer(0)
```

```
which(res.025-res.075 != 0) # all 0
```

```
## integer(0)
```

```r
which(res.025-res.lasso != 0) # all 0
```

```
## integer(0)
```

However, the diagnostics for the ridge regression model were skipped in the body of the report to save space. The details of those diagnostics are included here for completeness.

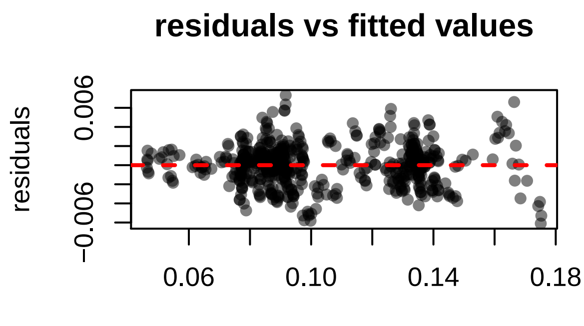

Fig 43: Residuals vs fitted values plot

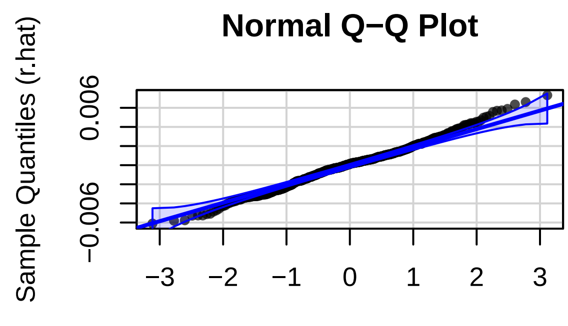

Fig 44: QQplot of residuals

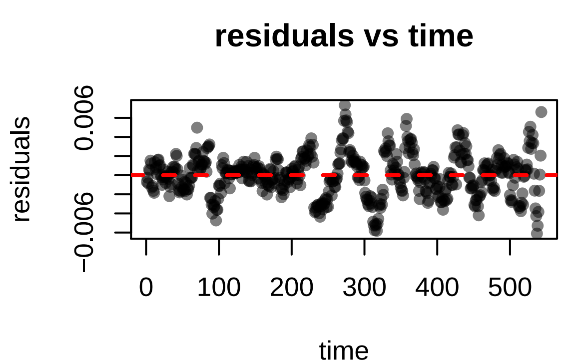

Fig 45: Residuals vs time plot

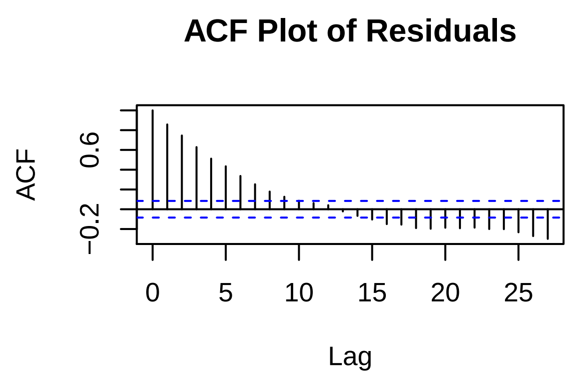

Fig 46: ACF plot of residuals

There seems to be a fanning out shape in the residuals vs fitted values plot, suggesting non-constant variance.

A lot of points fall outside the confidence band of the QQplot, suggesting the data is not normal.

It seems like there is non-constant variance in the residuals vs time plot.

There is linear decay in the ACF plot, indicating there is still some trend and the residuals are not stationary.

From these diagnostics, we concluded that the assumptions for this model were violated, which was stated in the main report.

Non-graphical residual diagnostics for all regression models

Non-regularized degree 6 model Firstly, for the non-regularized degree 6 model:

```
## 
##  Shapiro-Wilk normality test
## 
```

```
## data:  residuals.model6
## W = 0.98035, p-value = 1.087e-06
```

The small p-value indicates that there is strong evidence against the null hypothesis that the residuals are normal.

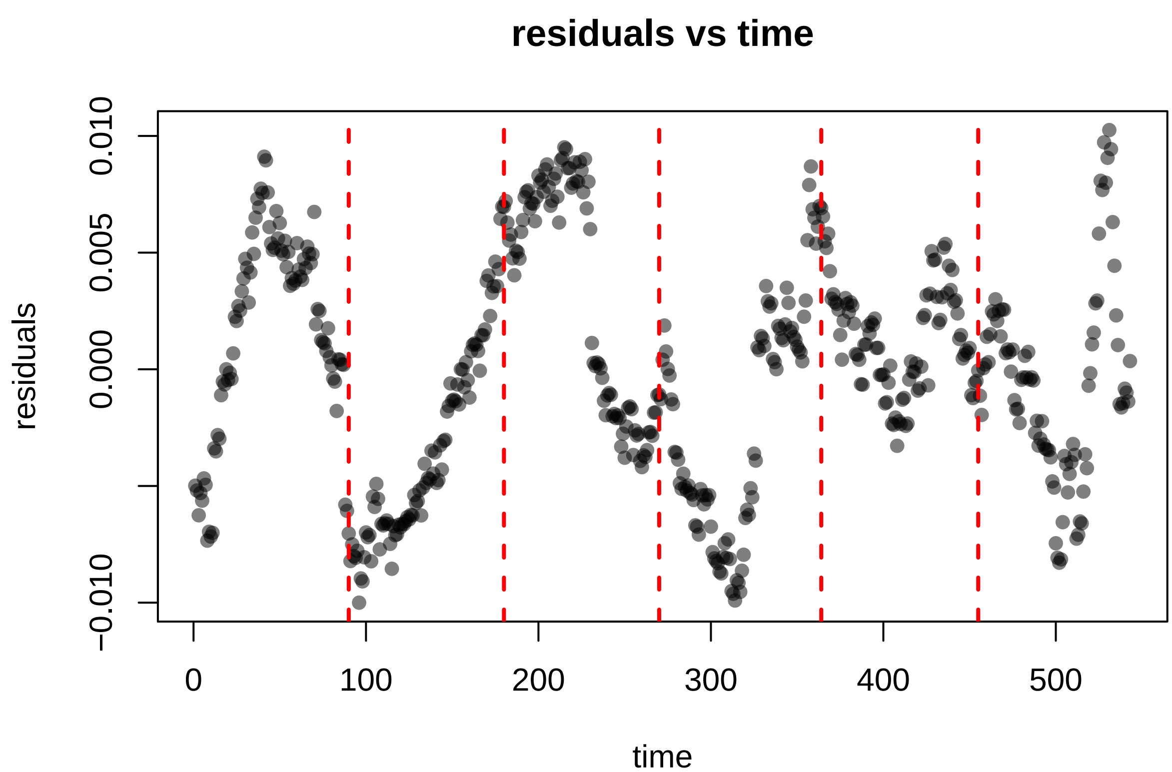

Fig 47: residuals vs time plot for Filgner−Killeen test with 6 groups

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  residuals.model6 and segments
## Fligner-Killeen:med chi-squared = 51.127, df = 5, p-value = 8.145e-10
```

The small p-value indicates that there is strong evidence against the null hypothesis of constant variance.

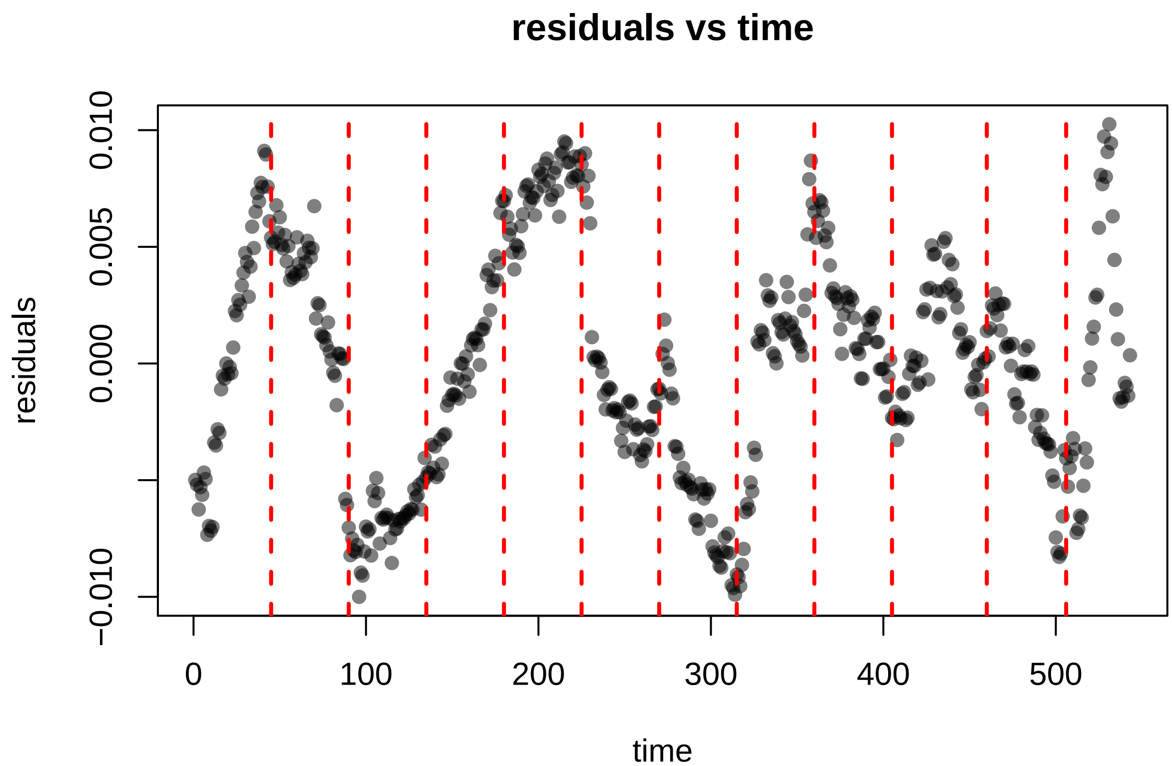

Fig 48: Residuals vs time plot for Fligner−Killeen test with 12 groups

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  residuals.model6 and segments
## Fligner-Killeen:med chi-squared = 111.66, df = 11, p-value < 2.2e-16
```

The small p-value indicates that there is strong evidence against the null hypothesis of constant variance.

```
##
##  Difference Sign Test
##
## data:  residuals.model6
## statistic = -2.8219, n = 543, p-value = 0.004774
## alternative hypothesis: nonrandomness
```

The small p-value indicates that there is strong evidence against the null hypothesis that the residuals are random.

```
##
##  Runs Test
##
## data:  residuals.model6
## statistic = -19.949, runs = 40, n1 = 271, n2 = 271, n = 542, p-value <
## 2.2e-16
## alternative hypothesis: nonrandomness
```

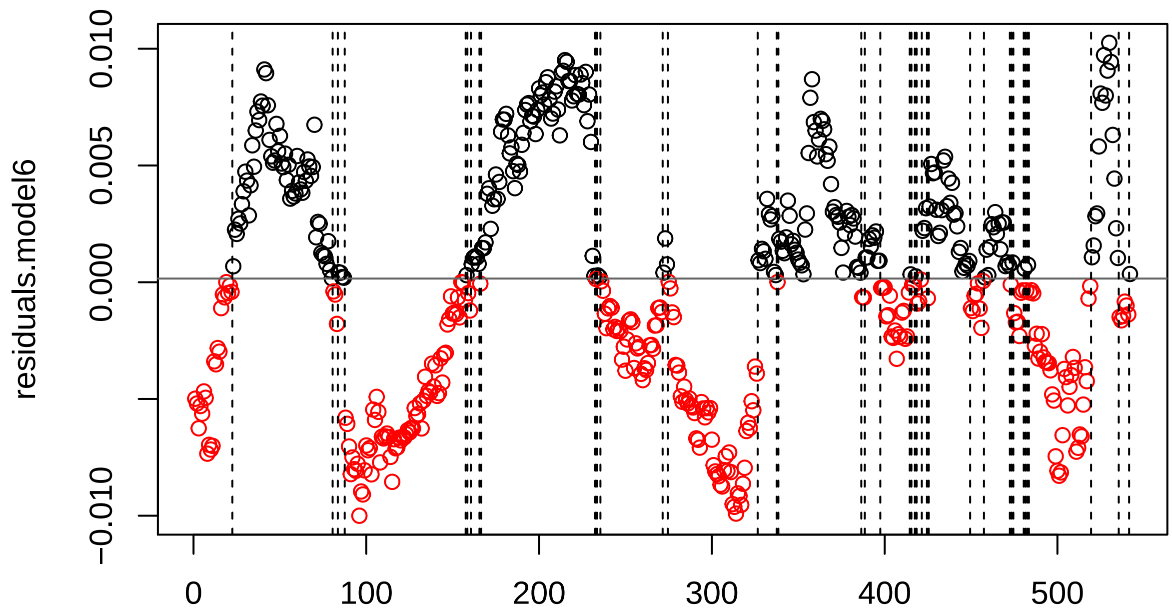

Fig 49: Runs test plot

The small p-value indicates that there is strong evidence against the null hypothesis that the residuals are random.

Regularized degree 15 model Next, for the regularized degree 15 model (alpha=0.25, 0.5, 0.75 and 1 have the same residuals, so included here only once):

```
##
##  Shapiro-Wilk normality test
##
## data:  res.025
## W = 0.99455, p-value = 0.05013
```

Since the p-value is greater than 0.05, but really close to 0.05, there is weak evidence against the null hypothesis that the residuals are normal.

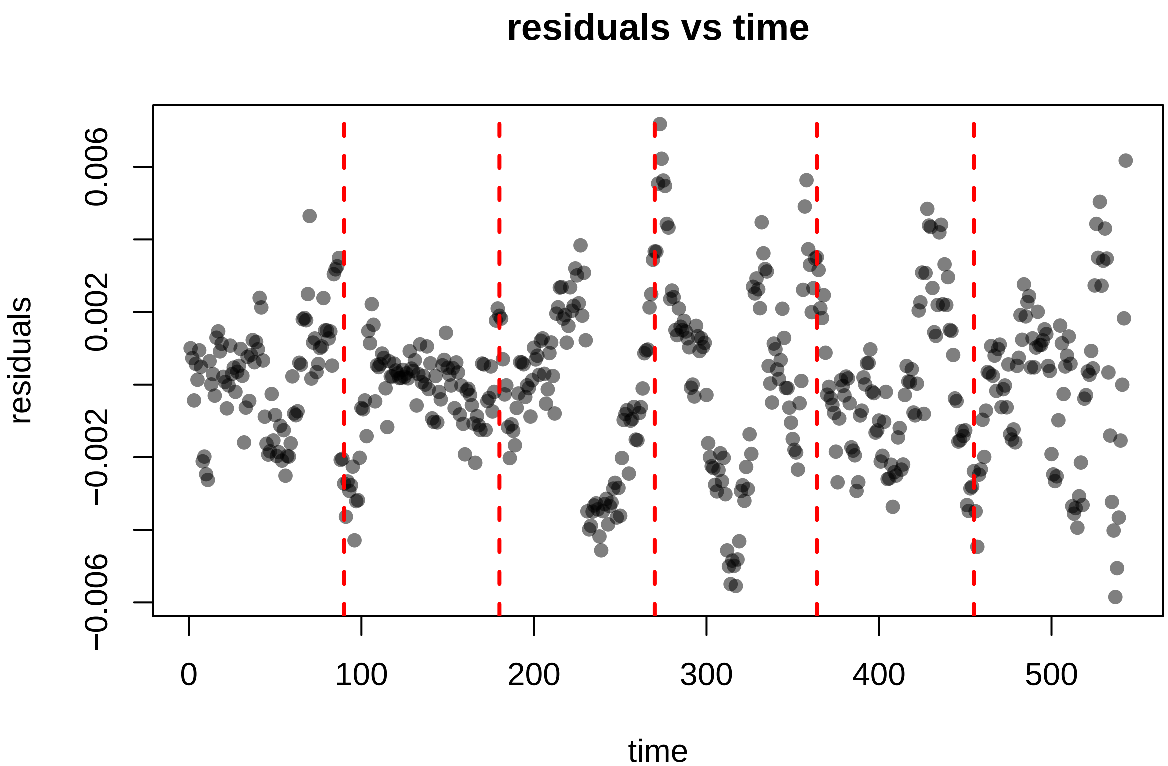

Fig 50: Residuals vs time plot for Fligner−Killen test with 6 groups

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  res.025 and segments
## Fligner-Killeen:med chi-squared = 63.654, df = 5, p-value = 2.131e-12
```

Since the p-value is small, there is strong evidence against the null hypothesis that the residuals have constant variance.

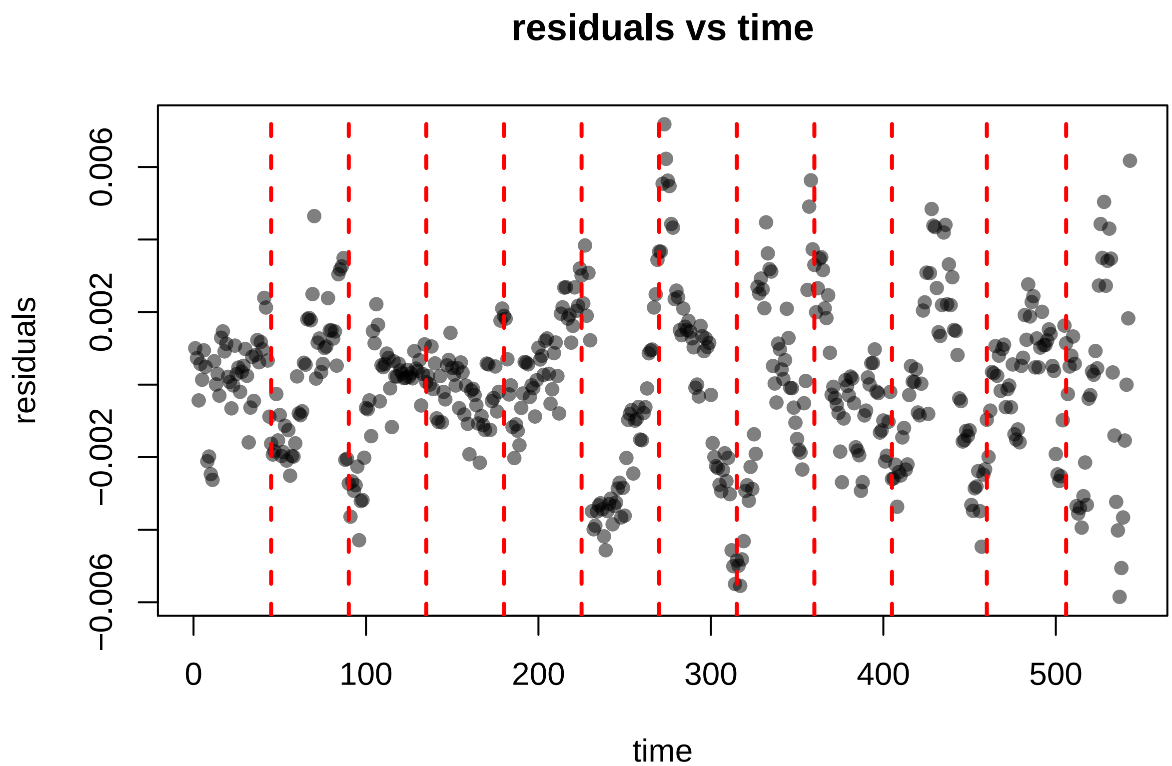

Fig 51: Residuals vs time plot for Fligner−Killeen test with 12 groups

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  res.025 and segments
## Fligner-Killeen:med chi-squared = 94.618, df = 11, p-value = 2.064e-15
```

Since the p-value is small, there is strong evidence against the null hypothesis of constant variance.

```
##
##  Difference Sign Test
##
## data:  res.025
## statistic = -0.29704, n = 543, p-value = 0.7664
## alternative hypothesis: nonrandomness
```

Since the p-value is large, there is not enough evidence to reject the null hypothesis that the residuals are random. Recall, however, that the sign difference test does not consider the order/placement of the points, just their sign.

```
##
##  Runs Test
##
## data:  res.025
## statistic = -14.876, runs = 99, n1 = 271, n2 = 271, n = 542, p-value <
## 2.2e-16
## alternative hypothesis: nonrandomness
```

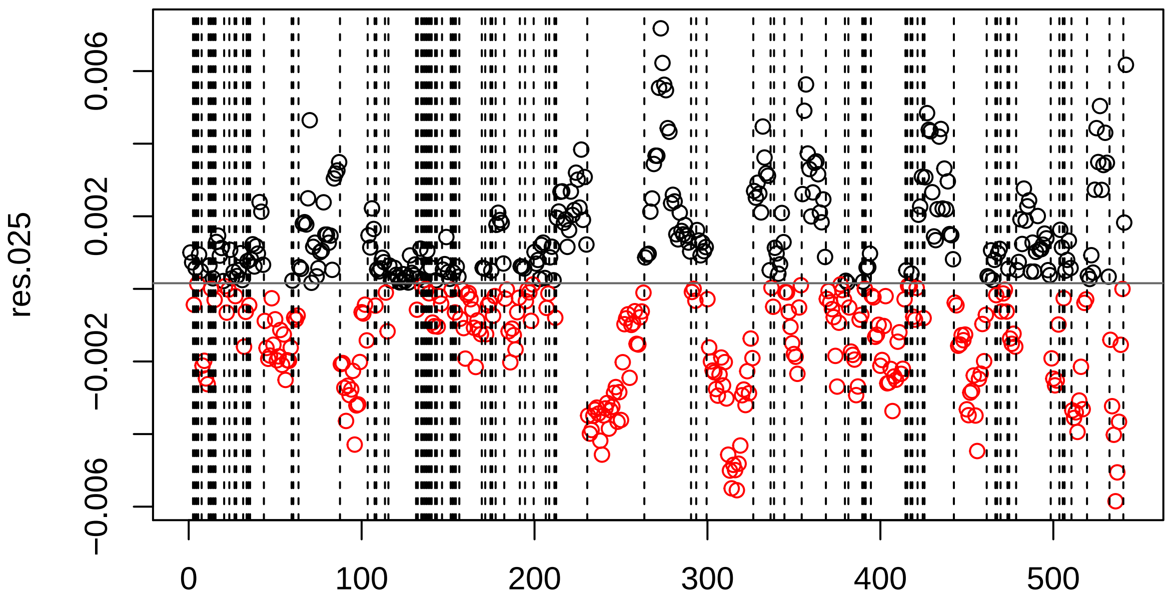

Fig 52: Runs test plot

Since the p-value is small, there is strong evidence against the null hypothesis that the residuals are random.

Regularized degree 13 model (alpha=0) Finally, for the regularized degree 13 model (alpha=0):

```
##
##  Shapiro-Wilk normality test
##
## data:  res.ridge
## W = 0.99349, p-value = 0.01927
```

Since the p-value is small, there is evidence against the null hypothesis that the residuals are normal.

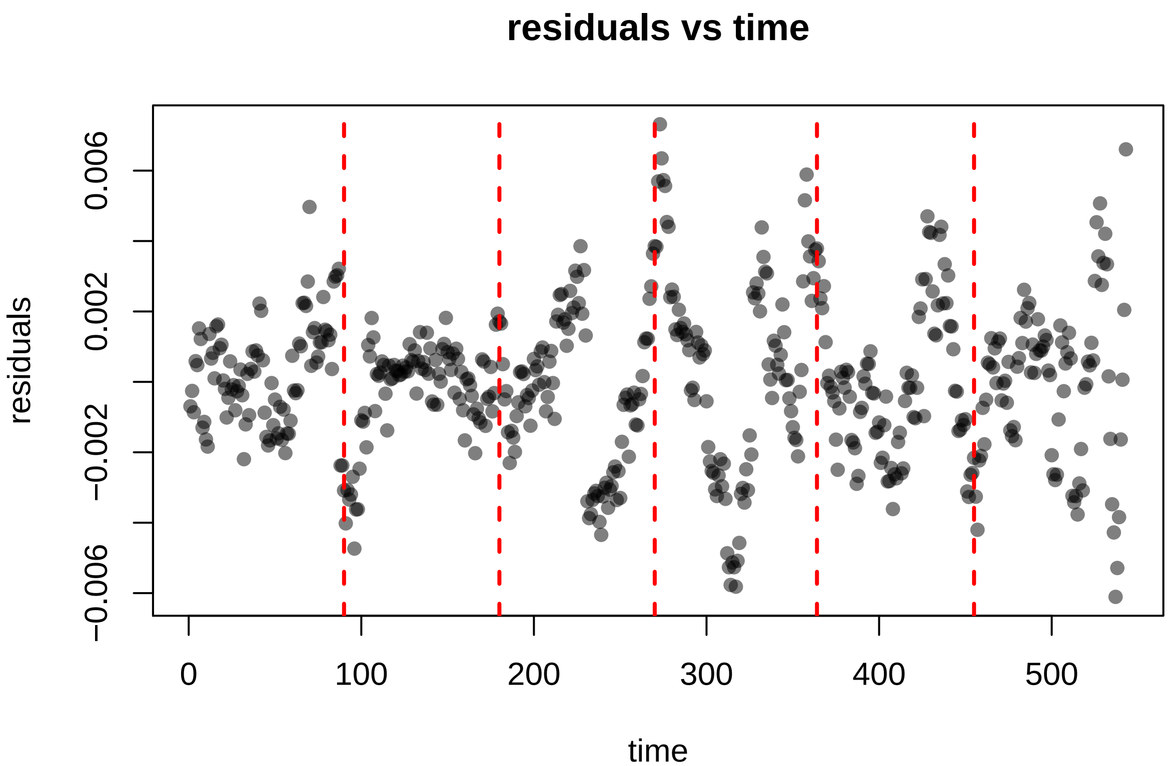

Fig 53: Residuals vs time plot for Fligner−Killeen test with 6 groups

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  res.ridge and segments
## Fligner-Killeen:med chi-squared = 59.785, df = 5, p-value = 1.346e-11
```

Since the p-value is small, there is strong evidence against the null hypothesis of constant variance

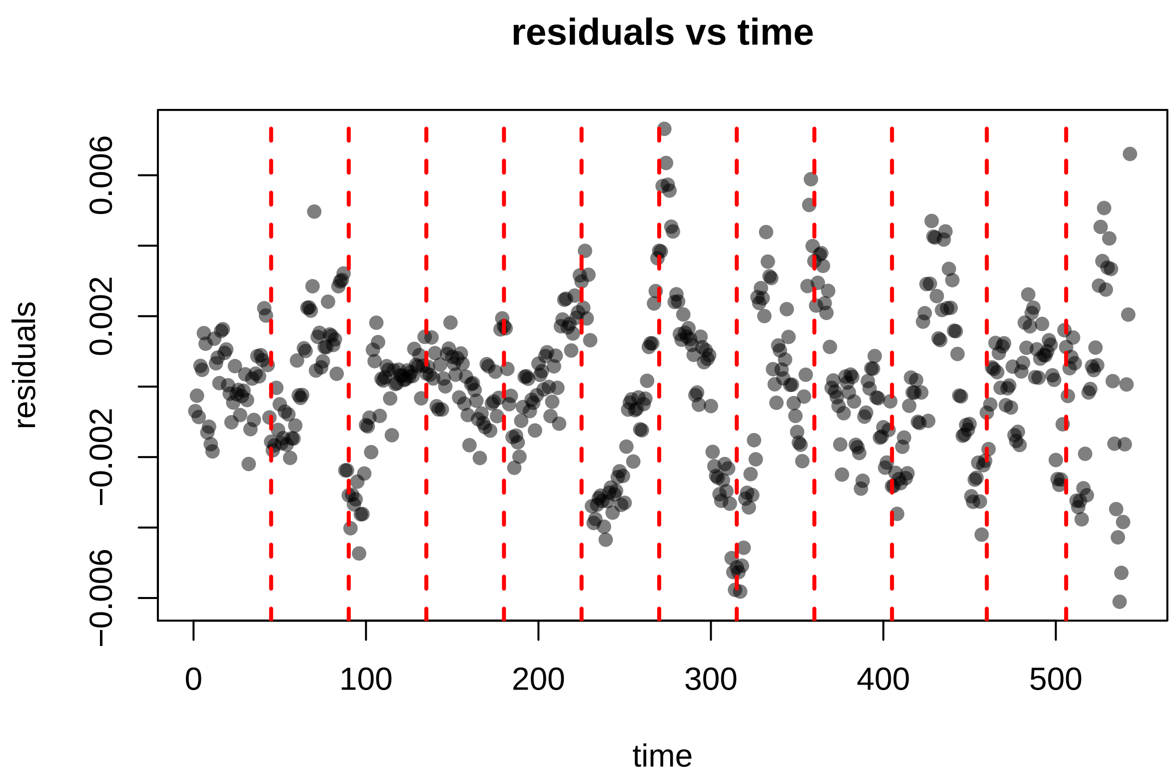

Fig 54: Residuals vs time plot for Fligner−Killeen test with 12 groups

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  res.ridge and segments
## Fligner-Killeen:med chi-squared = 90.607, df = 11, p-value = 1.268e-14
```

Since the p-value is small, there is strong evidence against the null hypothesis of constant variance.

```
##
##  Difference Sign Test
##
## data:  res.ridge
## statistic = -0.29704, n = 543, p-value = 0.7664
## alternative hypothesis: nonrandomness
```

Since the p-value is large, there is not enough evidence to reject the null hypothesis that the residuals are random. Recall, however, that the sign difference test does not consider the order/placement of the points, just their sign.

```
##
##  Runs Test
##
## data:  res.ridge
## statistic = -14.962, runs = 98, n1 = 271, n2 = 271, n = 542, p-value <
## 2.2e-16
## alternative hypothesis: nonrandomness
```

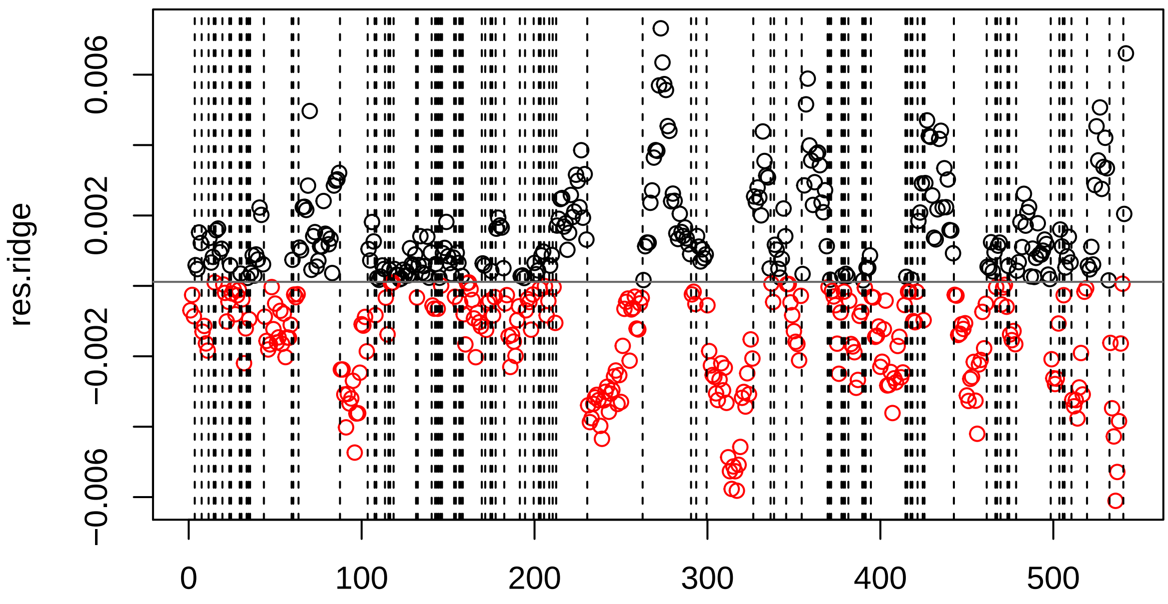

Fig 55: Runs test plot

Since the p-value is small, there is strong evidence against the null hypothesis that the residuals are random.

Smoothing Methods

1. Additive Holt Winters Method, no trend, no seasonality.

Training Prediction Plot of the Additive Holt Winters Method, no trend, no seasonality.

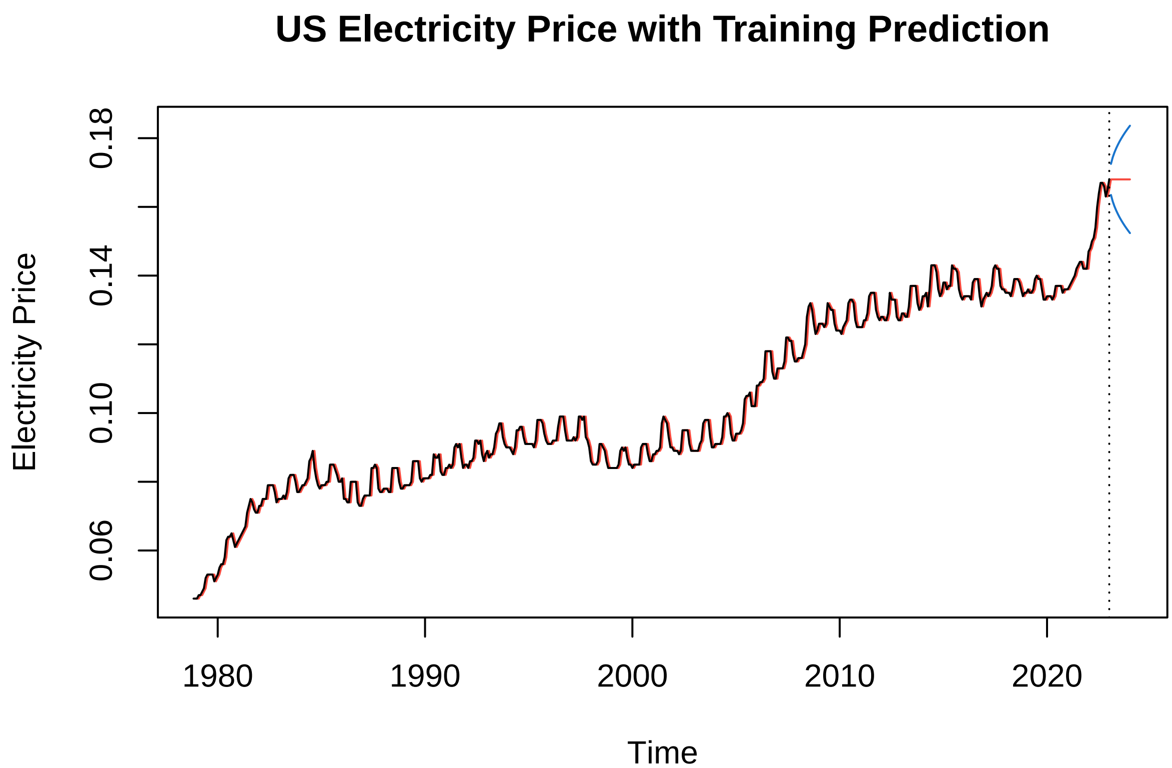

Fig 56: US Electricity Price with Training Prediction

Residuals of the Additive Holt Winters Method, no trend, no seasonality

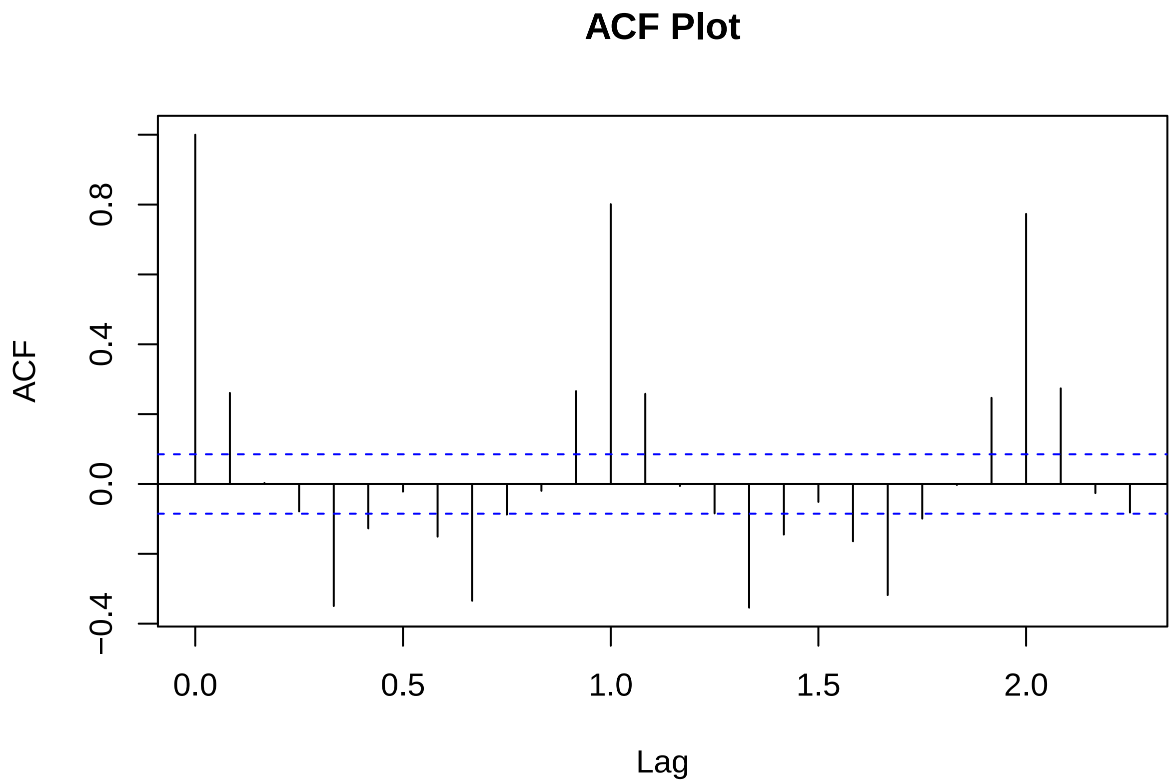

Fig 57: ACF Plot of the Holt Winters Residuals

There appears to be slow decay in the lag of season of the ACF Plot for Figure 54, hence the residuals are non-stationary.

2. Additive Holt Winters Method with both trend and seasonality.

Training Prediction Plot of the Additive Holt Winters Method with both trend and seasonality.

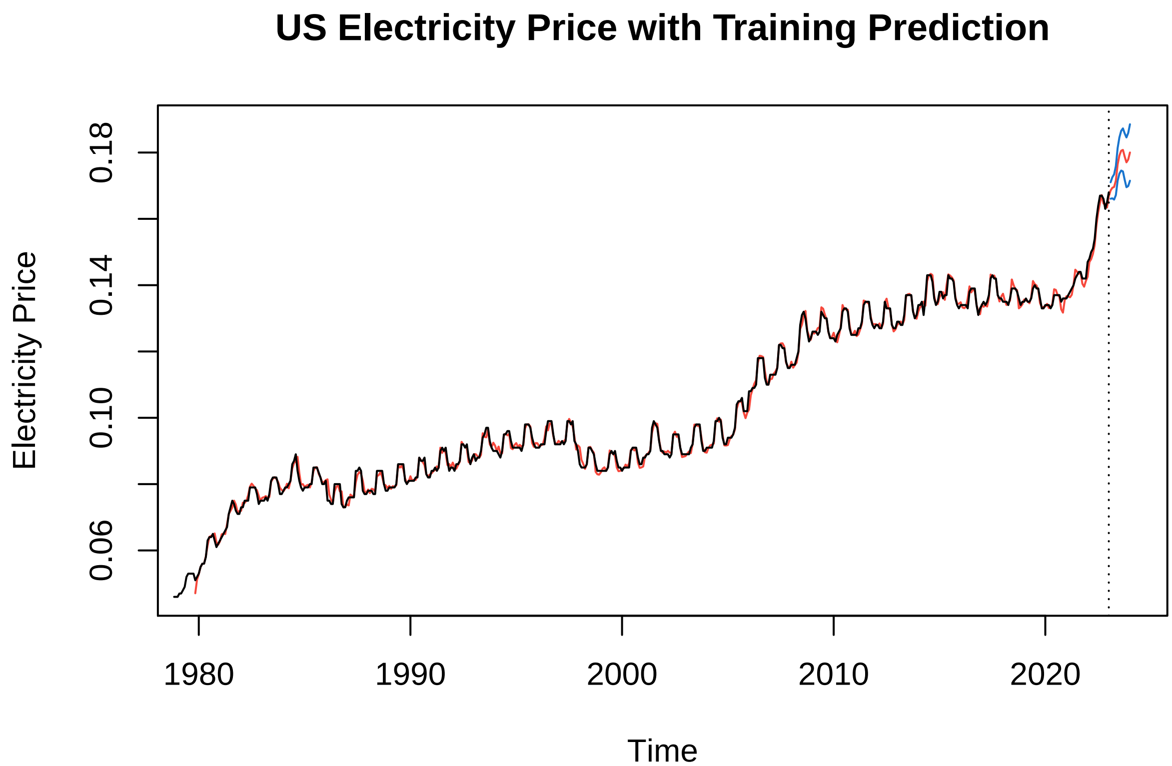

Fig 58: US Electricity Price with Training Prediction

Residuals of the additive Holt Winters model with both trend and seasonality.

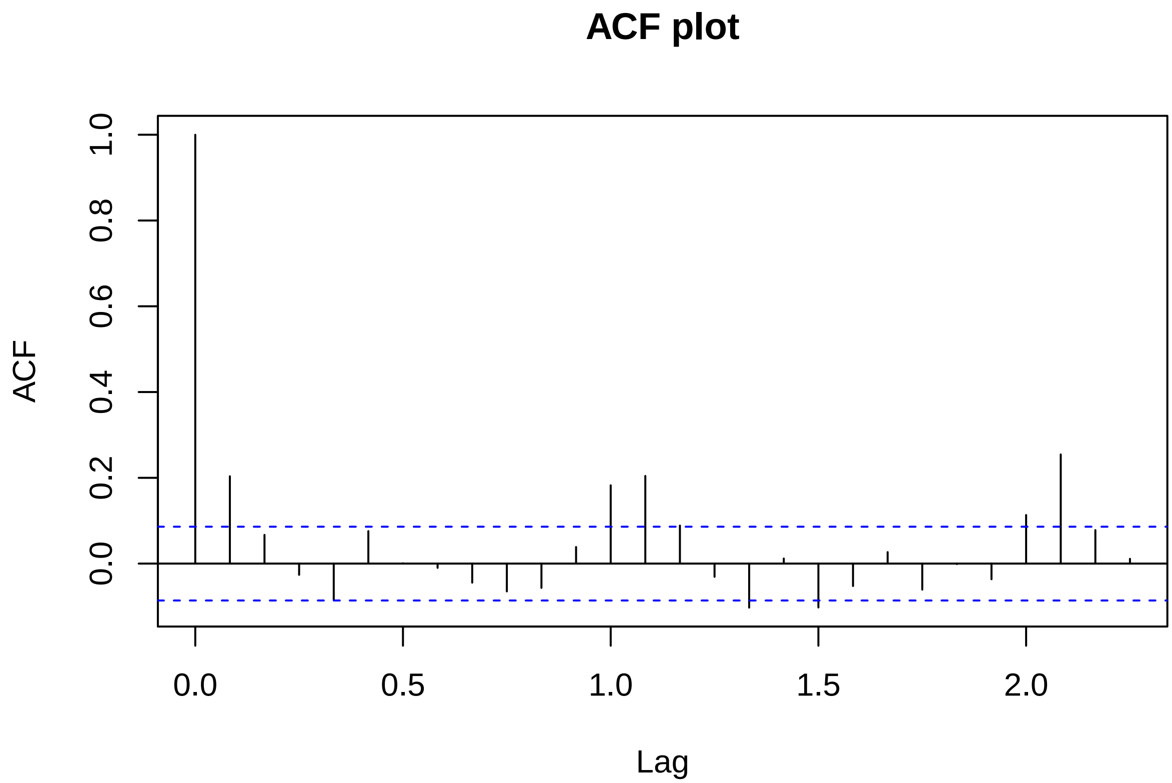

Fig 59: ACF Plot of the Holt Winters Residuals

There is no sign of linear decay, periodic pattern or decay in the lag of season of the ACF Plot in Figure 59, so we conclude that the residuals are stationary.

3. Multiplicative Holt Winters Method with both trend and seasonality.

Training Prediction Plot of the Multiplicative Holt Winters Method with both trend and seasonality.

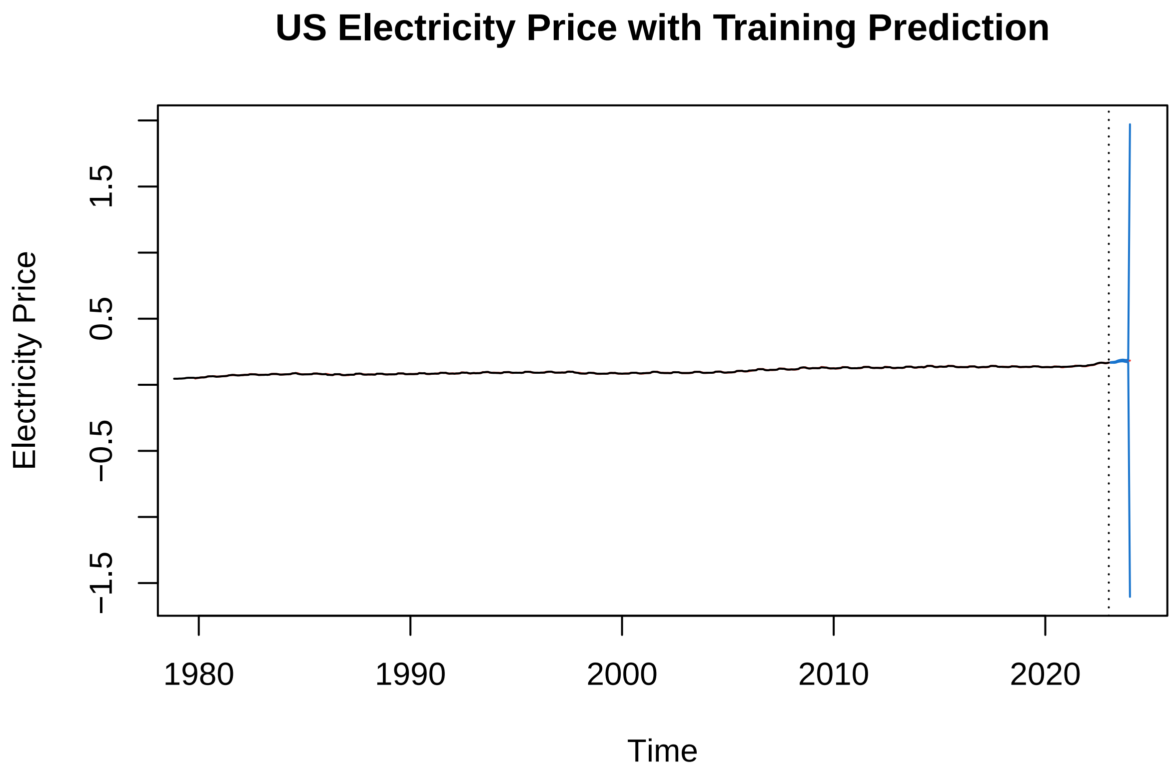

Fig 60: US Electricity Price with Training Prediction

Restricting the y-axis of Figure 60 so we can see the fit better.

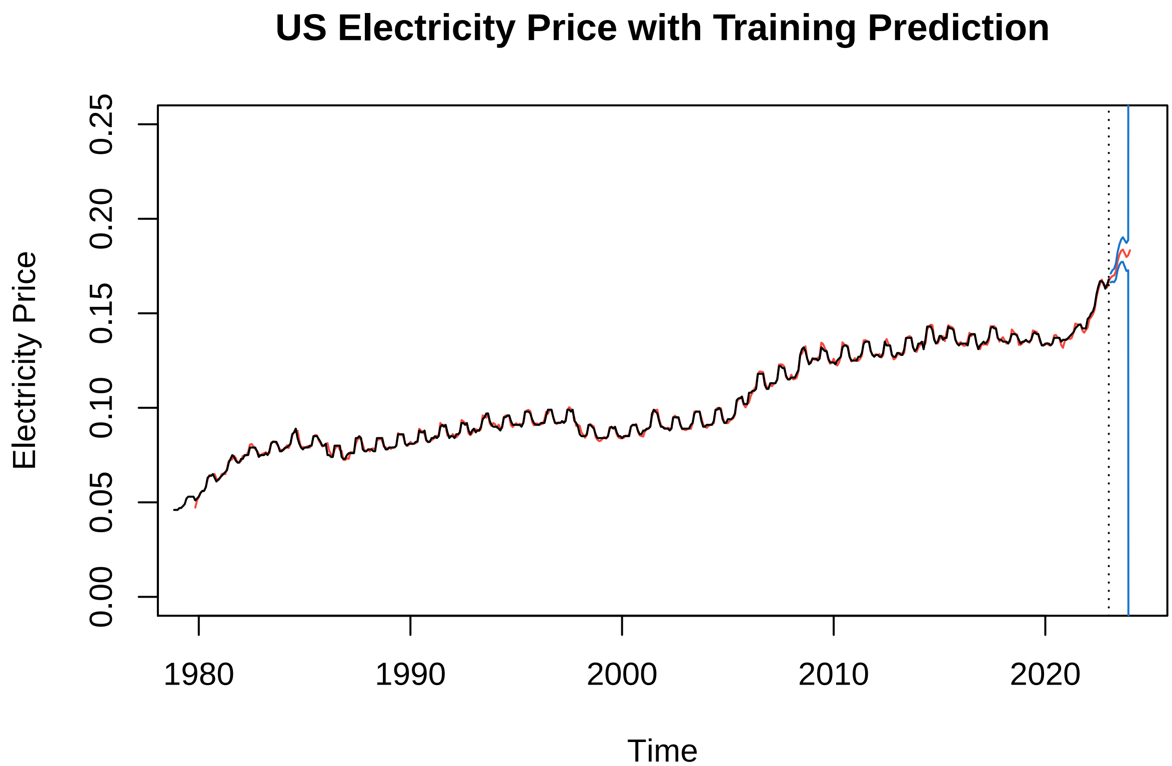

Fig 61: US Electricity Price with Training Prediction

Residuals of the Multiplicative Holt Winters Method with both trend and seasonality.

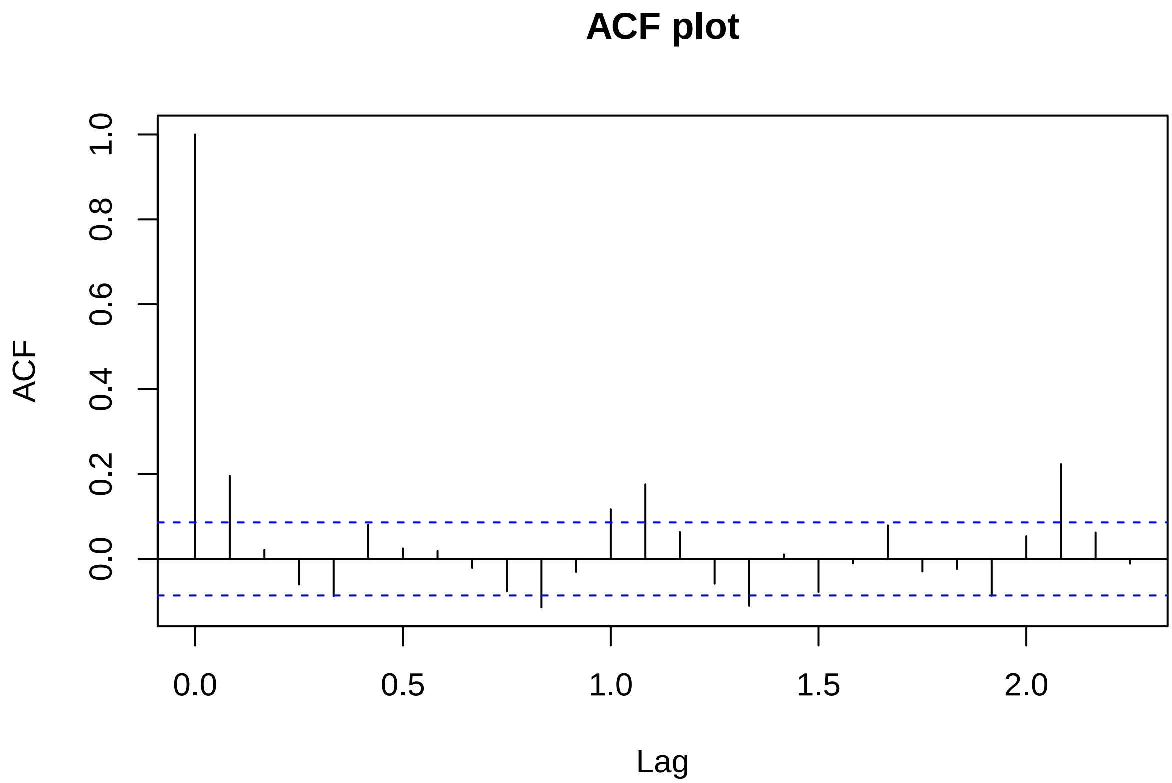

Fig 62: ACF Plot of the Holt Winters Residuals

There is no sign of linear decay, periodic pattern or decay in the lag of season of Figure 62 so we conclude that the residuals are stationary.

4. Multiplicative Holt Winters Method with seasonality but not trend.

Training Prediction Plot of the Multiplicative Holt Winters Method with seasonality but not trend.

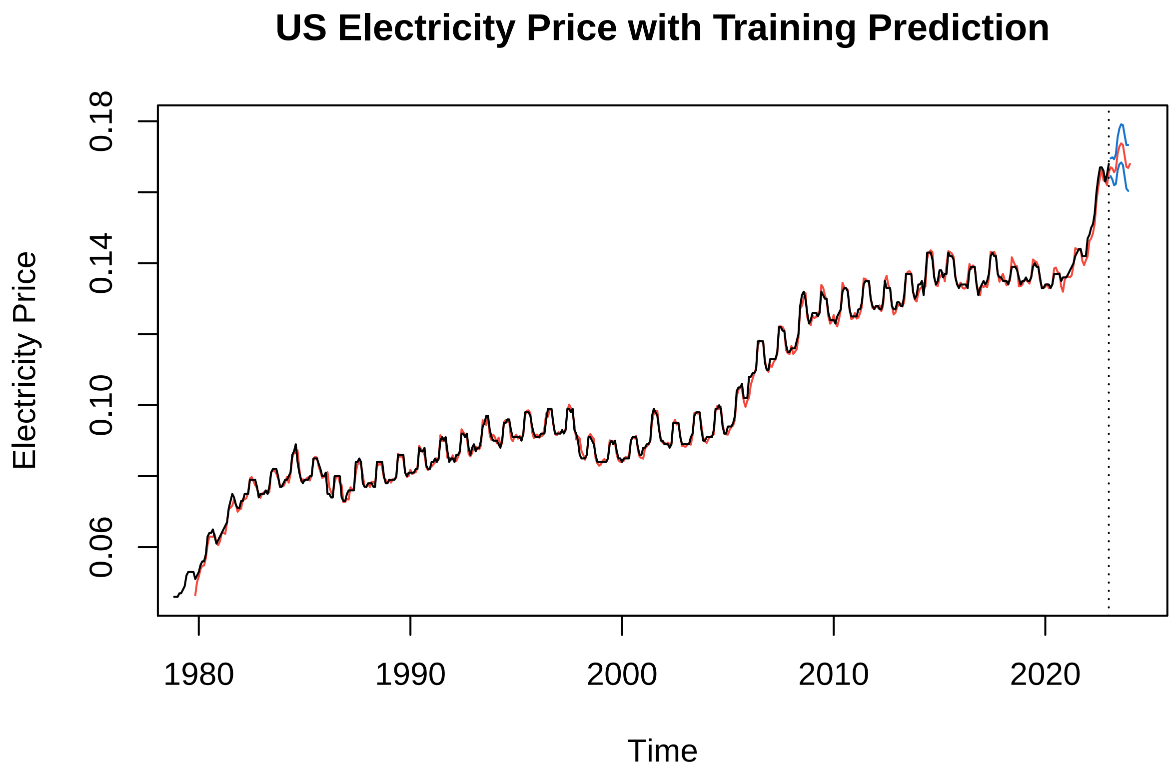

Fig 63: US Electricity Price with Training Prediction

Residuals of the Multiplicative Holt Winters Method with seasonality but not trend.

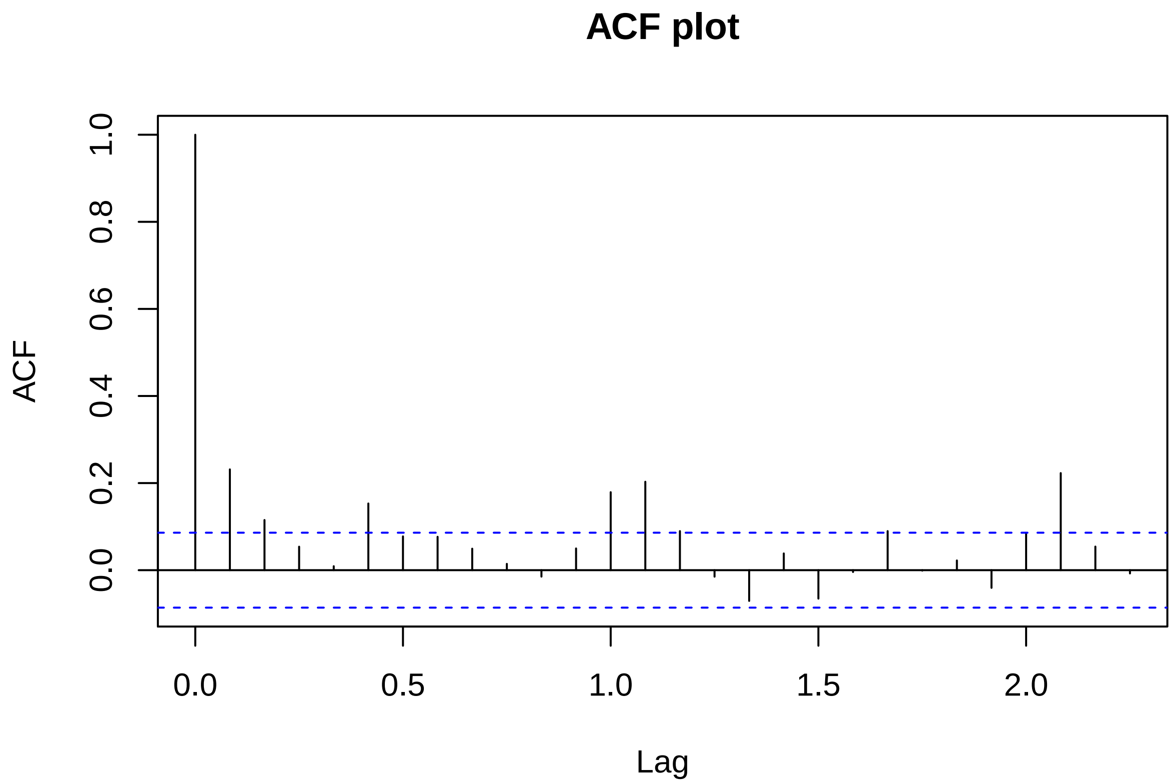

Fig 64: ACF Plot of the Holt Winters Residuals

There is no sign of linear decay, periodic pattern or decay in the lag of season in the ACF Plot in Figure 64, so we conclude that the residuals are stationary.

5. Additive Holt Winters Method with seasonality but not trend.

Training Prediction Plot of the Additive Holt Winters Method with seasonality but not trend.

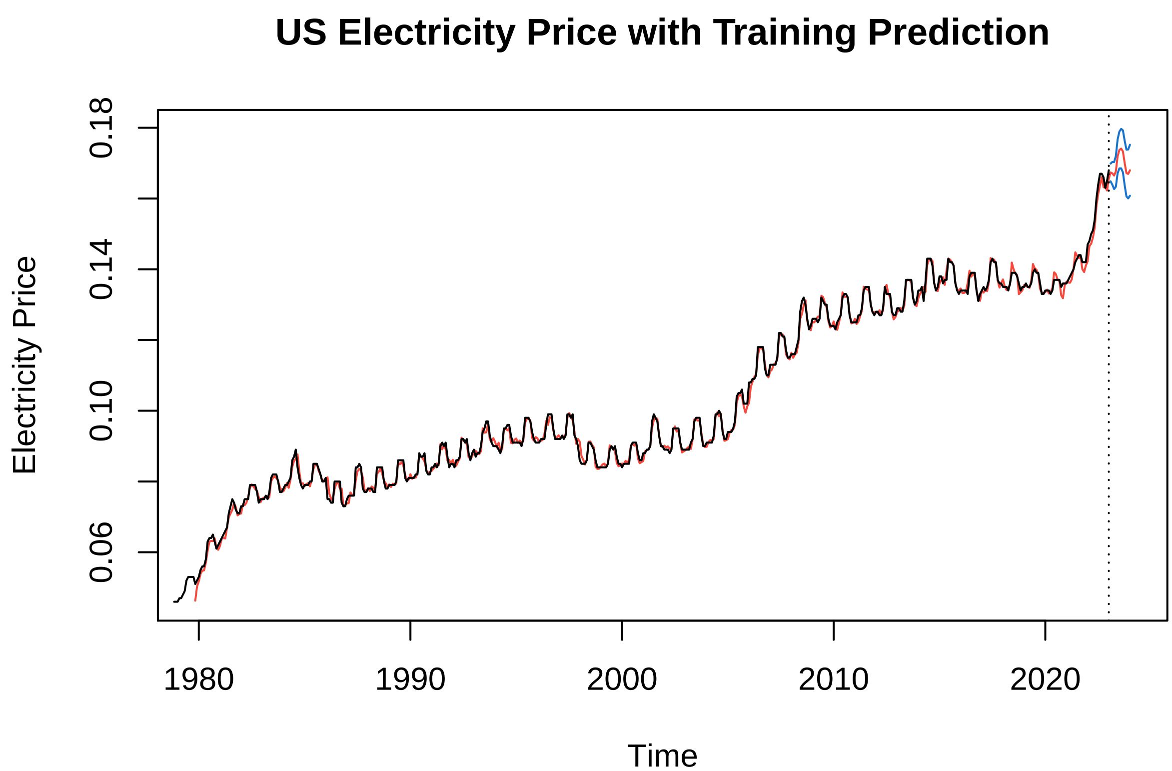

Fig 65: US Electricity Price with Training Prediction

Residuals of the Additive Holt Winters Method with seasonality but not trend.

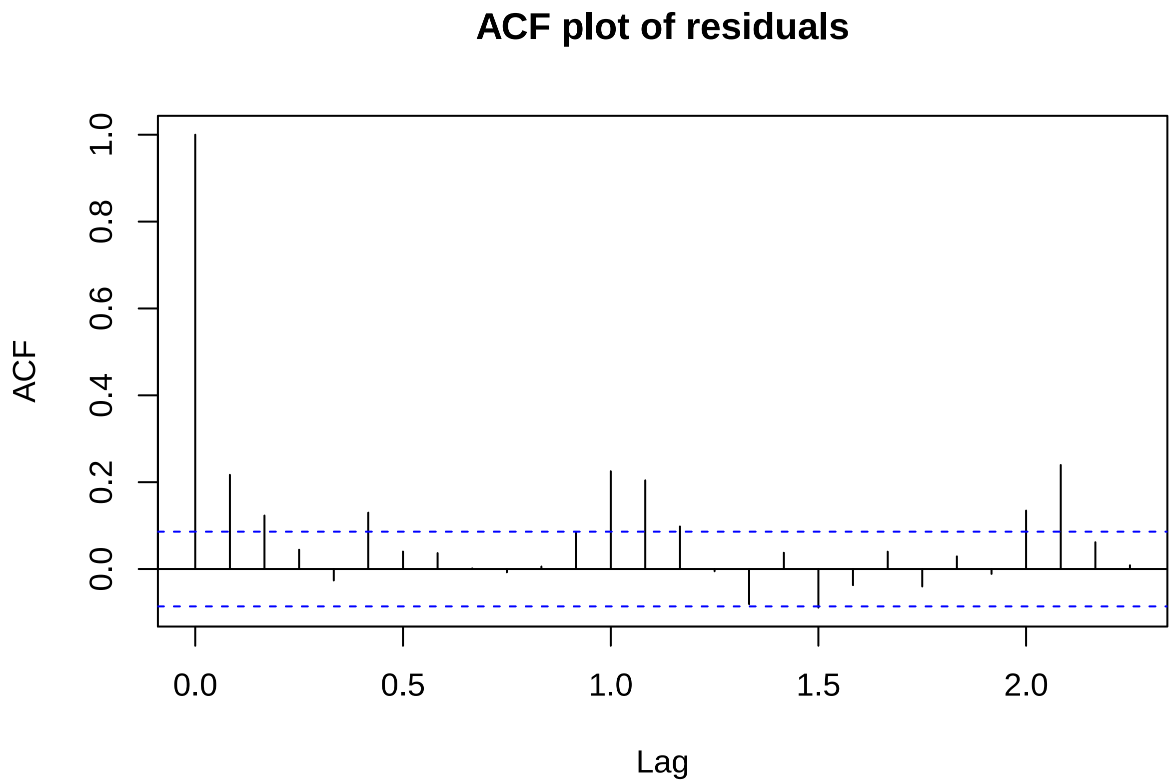

Fig 66: ACF Plot of the Holt Winters Residuals

There is no sign of linear decay, periodic pattern or decay in the lag of season in the ACF Plot in Figure 66, so we conclude that the residuals are stationary.

Box-Jenkins

Variance analysis of combined difference data

We will look at the plot of the ACF of the squared combined difference data:

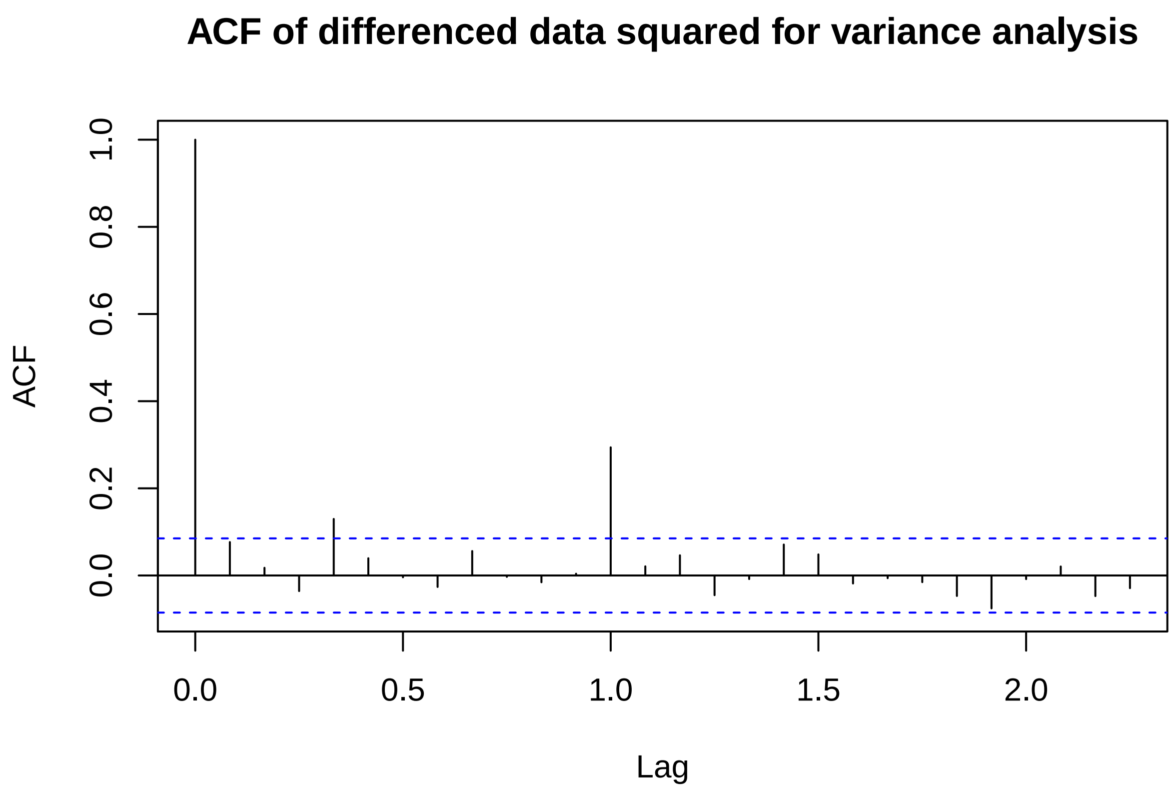

Fig 67: ACF of differenced data squared for variance analysis

Note that the ACF plot of the data squared does not exhibit any trends, indicating that there is constant variance.

Additional SARIMA model proposals

- Proposal 2: $SARIMA(0, 1, 5) \cdot (2, 1, 0)_{12}$: We use the same justification for the regular differencing ARIMA model. If we only look at the seasonal lags, the ACF plots can be considered exponentially decreasing the the PACF plot cuts off after the second seasonal lag, justifying P = 2 and Q = 0
- Proposal 3: $SARIMA(1, 1, 1) \cdot (0, 1, 1)_{12}$: If we only look at the ACF and the PACF of the regularly differenced data, we can claim that both plots exhibit exponential decay or sinusoidal damping patterns. Thus, we can propose an ARMA(1, 1) model for the regular differenced data. We use the same justification as Proposal 1 for the seasonal ARIMA model.
- Proposal 4: $SARIMA(1, 1, 1) \cdot (2, 1, 0)_{12}$: We can propose an ARIMA(1, 1) model for the regular differenced data based off the same justification as Proposal 3. We use the same justification as Proposal 2 for the seasonal ARIMA model.
- Proposal 5: $SARIMA(1, 1, 2) \cdot (0, 1, 1)_{12}$: Varying the ARMA model parameters for the regular differenced ARMA model.
- Proposal 6: $SARIMA(2, 1, 1) \cdot (0, 1, 1)_{12}$: Varying the ARIMA model parameters for the regular differenced ARIMA model.
- Proposal 7: $SARIMA(1, 1, 2) \cdot (2, 1, 0)_{12}$: Varying the ARIMA model parameters for the regular differenced ARIMA model. We'll use the same ARMA model parameters for the seasonally

differenced data as proposal 4.

- Proposal 8: $SARIMA(1,1,1) \cdot (1,1,1)_{12}$: Varying the ARIMA model parameters for the regular differenced ARIMA model. We can also propose an ARMA model for the seasonal differenced data as the ACF and PACF plots both look like they are exponentially decreasing or sinusoidally damped.
- Proposal 9: $SARIMA(1,1,1) \cdot (1,1,2)_{12}$: Varying the ARMA model parameters for both ARIMA models.
- Proposal 10: $SARIMA(1,1,1) \cdot (2,1,1)_{12}$: Varying the ARMA model parameters for both ARIMA models.

Model 11 and 12 model diagnostics

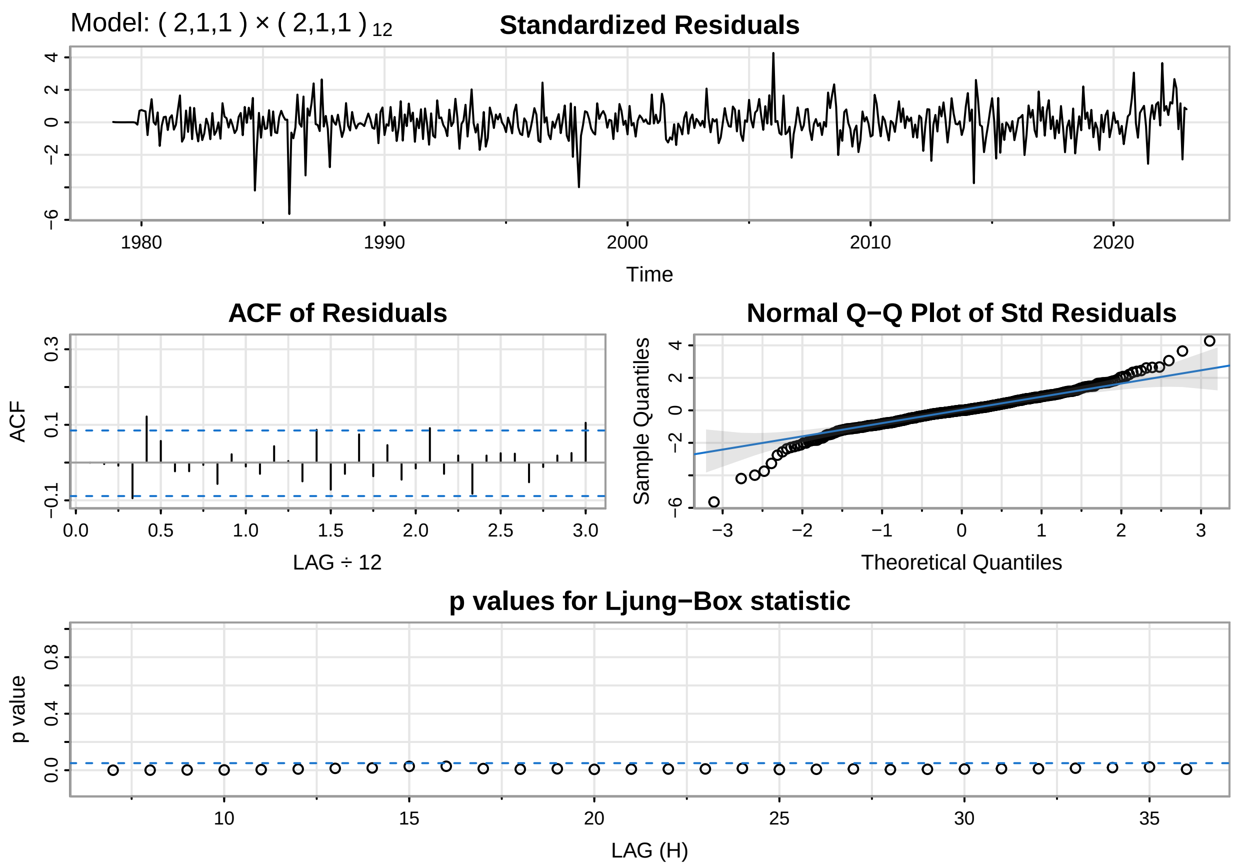

Figure 68: Model 11 model diagnostics

We immediately note that there are some spikes in the ACF graph of the residuals and the p-values of the Ljung-Box test are all significant. This indicates that there is still some correlation that the model does not sufficiently model. Also note that the residuals are not closely following the line in the Q-Q plot, indicating that the residuals are not normally distributed. Model 11 violates its assumptions. Model 12 follows a similar conclusion, as can be seen here:

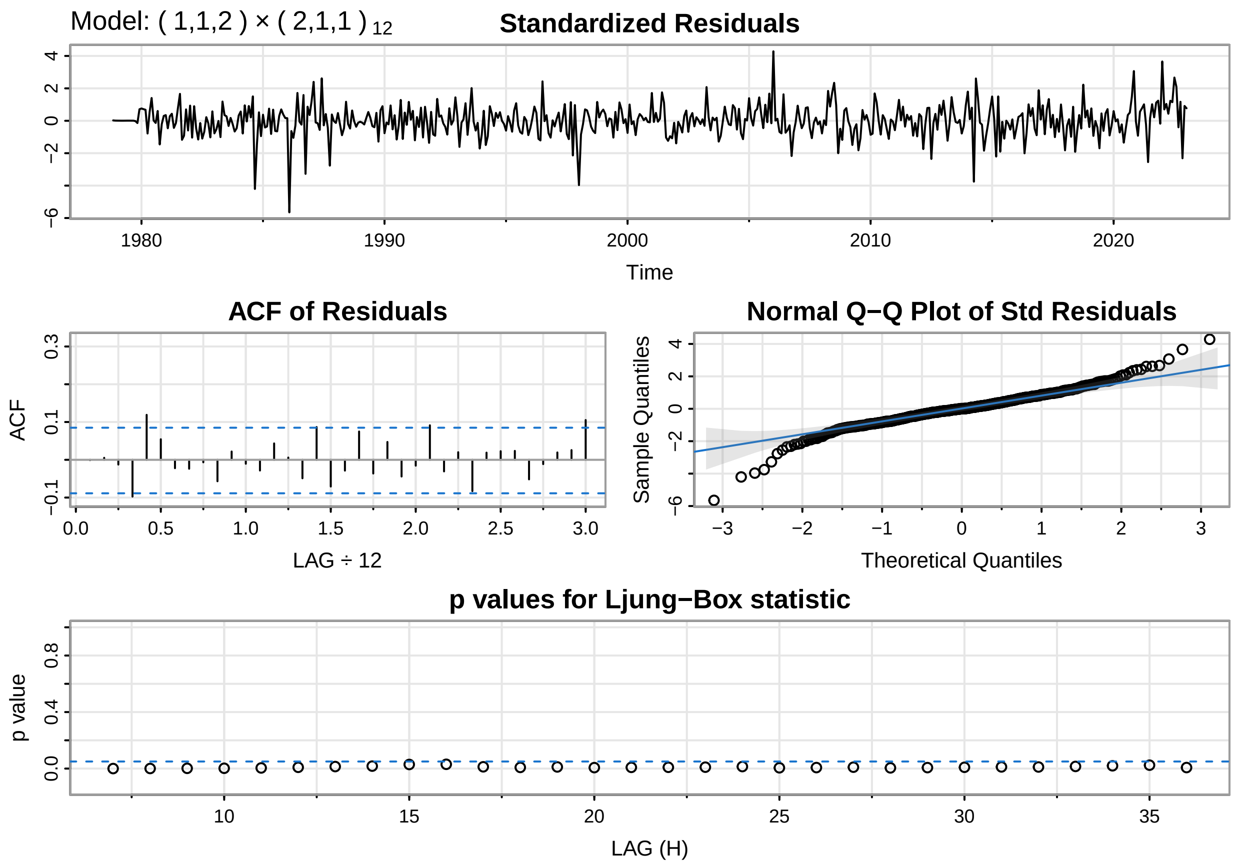

Figure 69: Model 12 model diagnostics

Seasonal differenced SARIMA

We will redo the SARIMA modelling with just one-time seasonal differencing. Since we have a exponential decay in our ACF plot and the PACF plot cuts off after a certain lag (as evidenced in section 3), we can propose the following models:

- p = 7, q = 0, P = 2, Q = 0, d = 0, D = 1, S = 12
- p = 1, q = 1, P = 2, Q = 0, d = 0, D = 1, S = 12
- p = 2, q = 1, P = 2, Q = 0, d = 0, D = 1, S = 12
- p = 1, q = 2, P = 2, Q = 0, d = 0, D = 1, S = 12
- p = 1, q = 1, P = 1, Q = 1, d = 0, D = 1, S = 12

After fitting all of these models, we compare the models based off AICc:

Table 4: AICc for top 3 SARIMA models

	Model	AICc
5	sarima_model_5	-10.66512
1	sarima_model_1	-10.62902
4	sarima_model_4	-10.59000

From the AICc results, we see that Model 5, 1 and 4 are the top performing models. We will now perform model diagnostics on each to see if residual diagnostics pass across all models. For Model 5, we have the following diagnostics

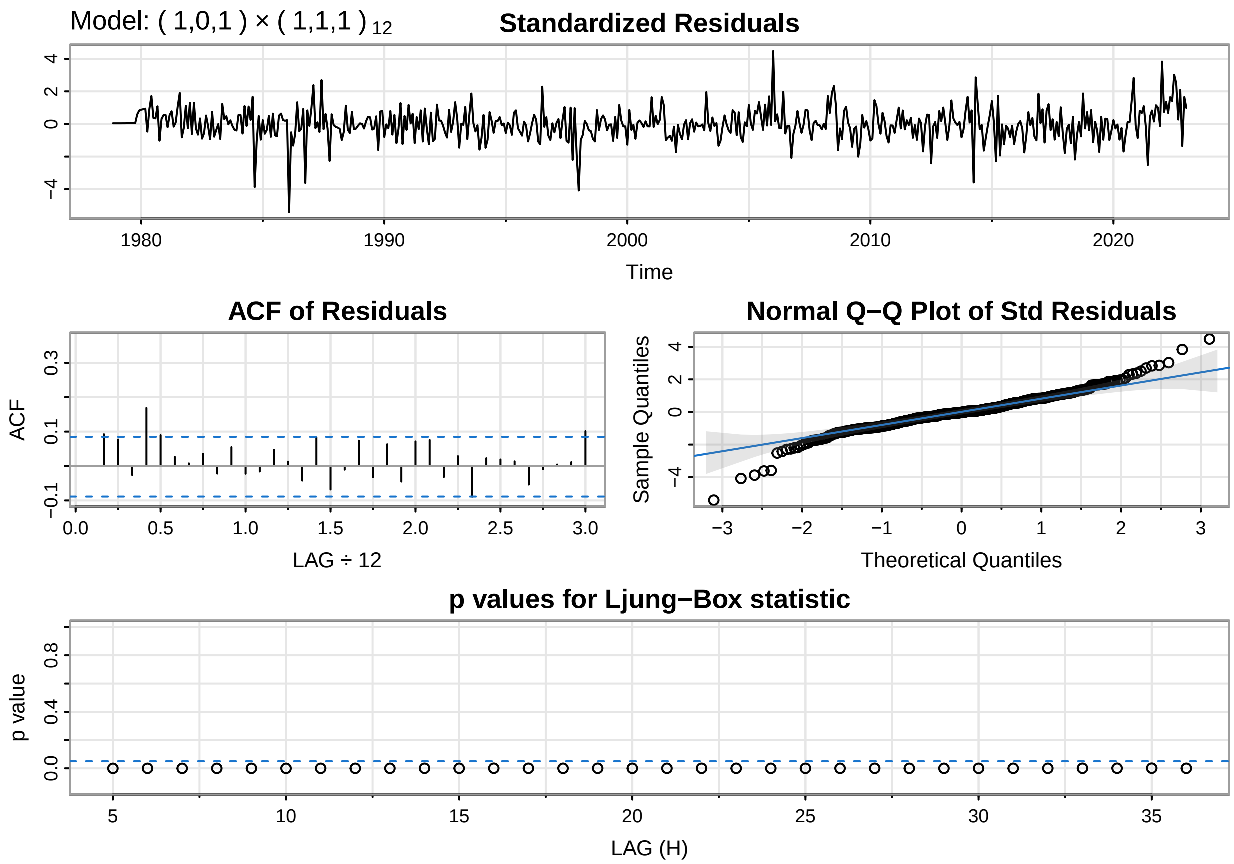

Figure 70: Model 5 residual plots

For Model 1, we have:

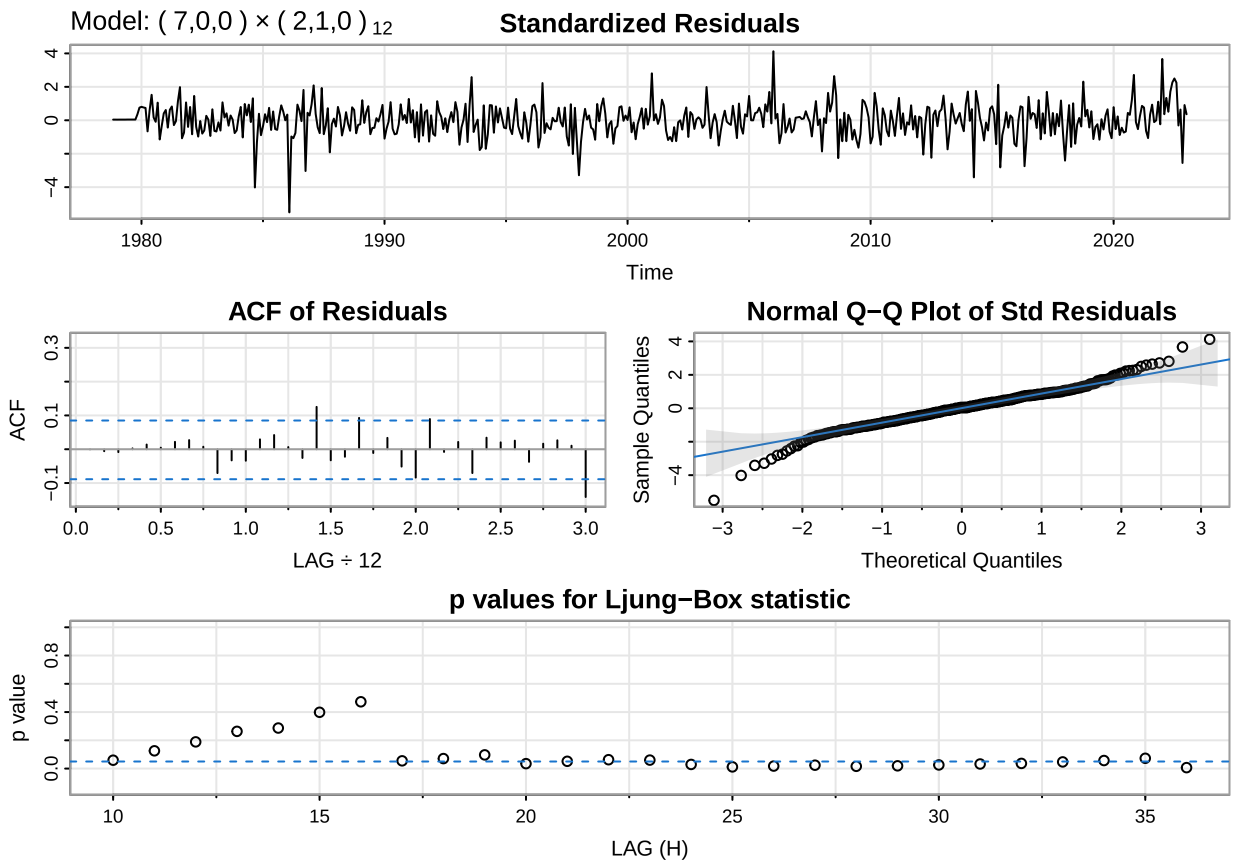

Figure 71: Model 1 residual plots

For Model 4, we have:

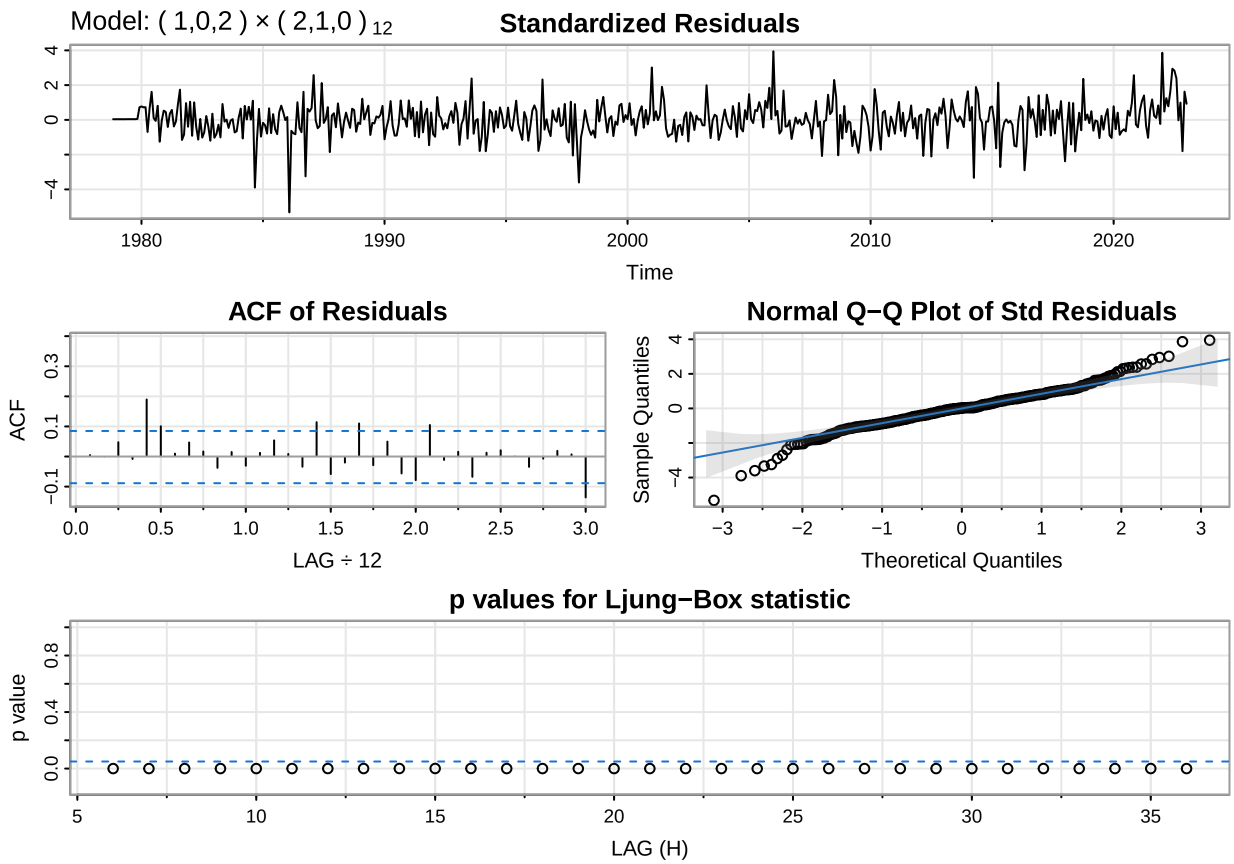

Figure 72: Model 4 residual plots

Across all 3 models, we notice that the p-values for the Ljung-Box test are significant, indicating that the proposed models do not capture all autocorrelations effectively. Hence, we will stop using one-timed seasonal differencing for SARIMA modelling.

Let's now analyze model 11:

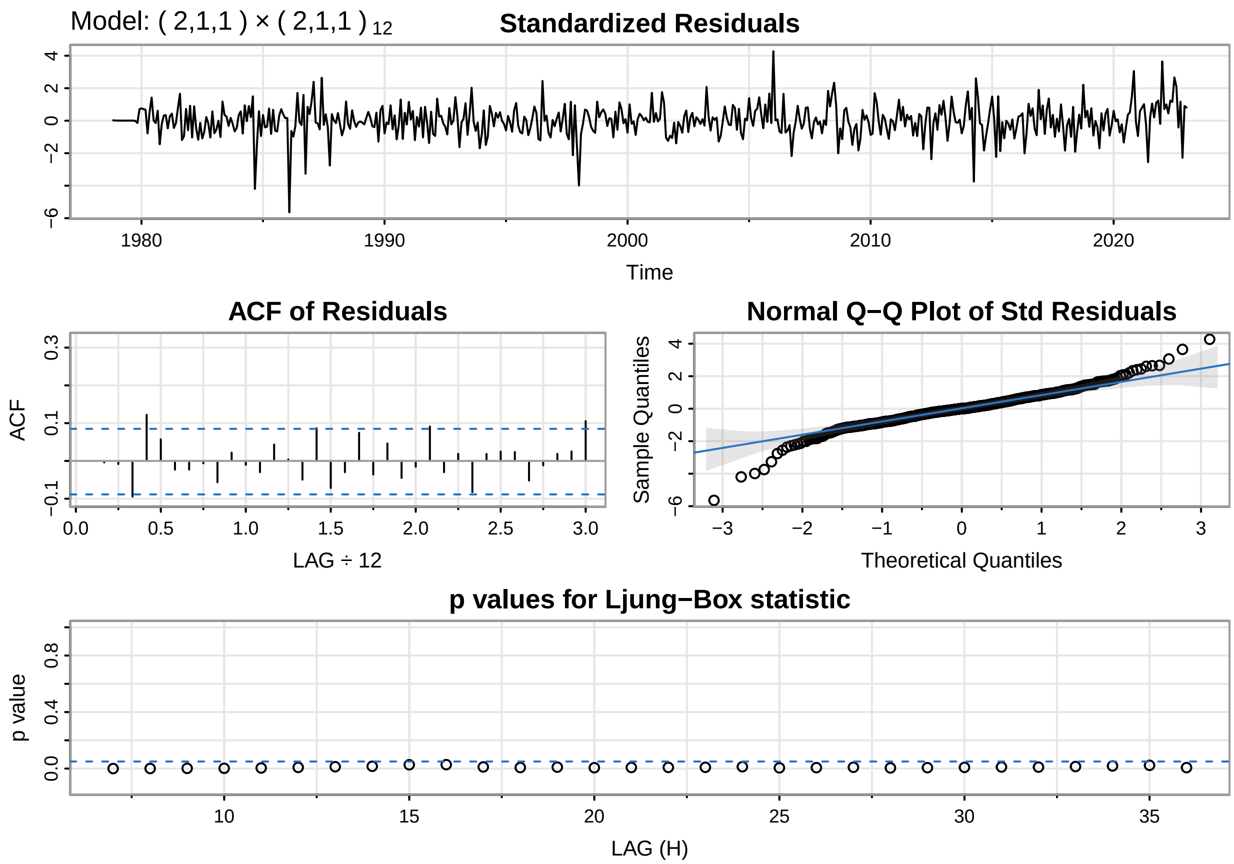

Figure 73: Model 11 residual plot

We immediately note that there are some spikes in the ACF graph of the residuals and the p-values of the Ljung-Box test are all significant. This indicates that there is still some correlation that the model does not sufficiently model. Also note that the residuals are not closely following the line in the Q-Q plot, indicating that the residuals are not normally distributed. Model 11 violates its assumptions. Similar results are found in model 12 (can be found in the appendix section titled "Model 12 model diagnostics").

Even though Model 11 and 12 have violated assumptions, we are still going to evaluate APSE on all three chosen models. We will only use Model 1 for forecasting the next 2 years of data.

Let's forecast each model on the test dataset, which is a 12-step ahead prediction.

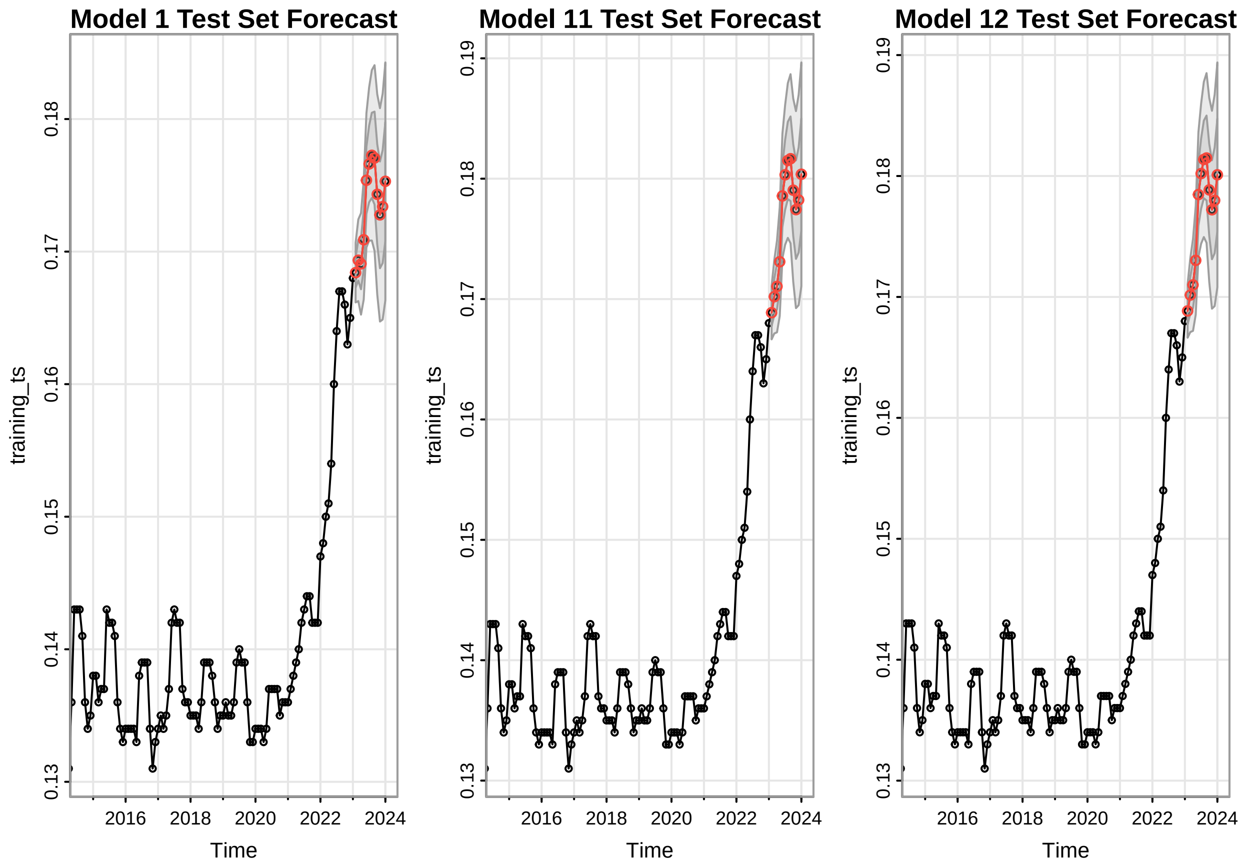

Figure 74: Model 1, 11, 12 forecasts on test set

Combination

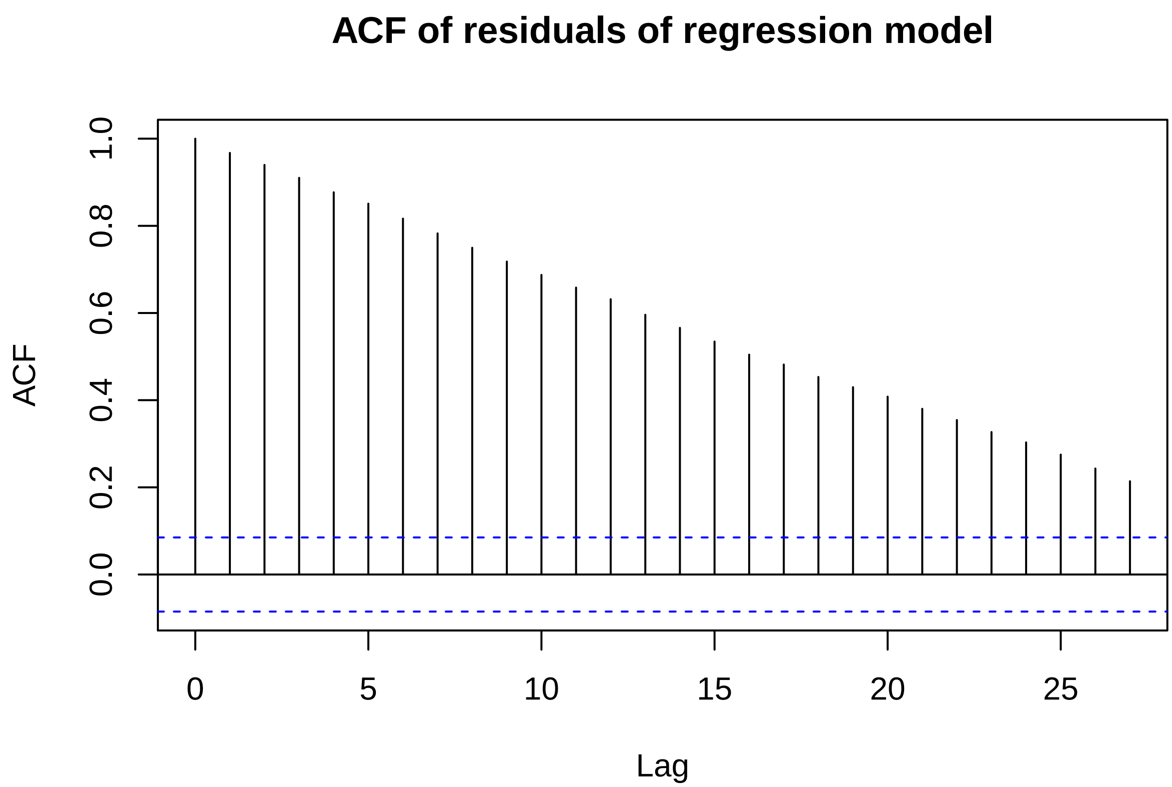

Fig 75: ACF of residuals of regression model

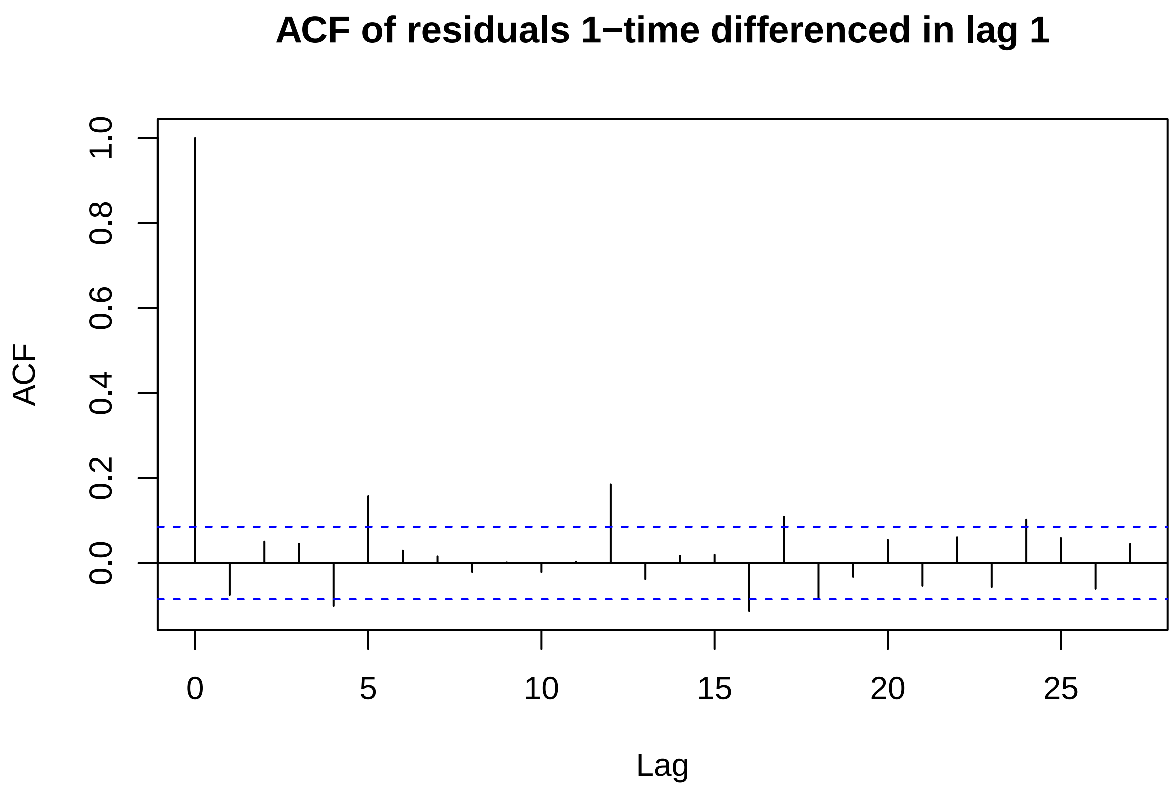

Fig 76: ACF of 1−time differenced residuals of regression model

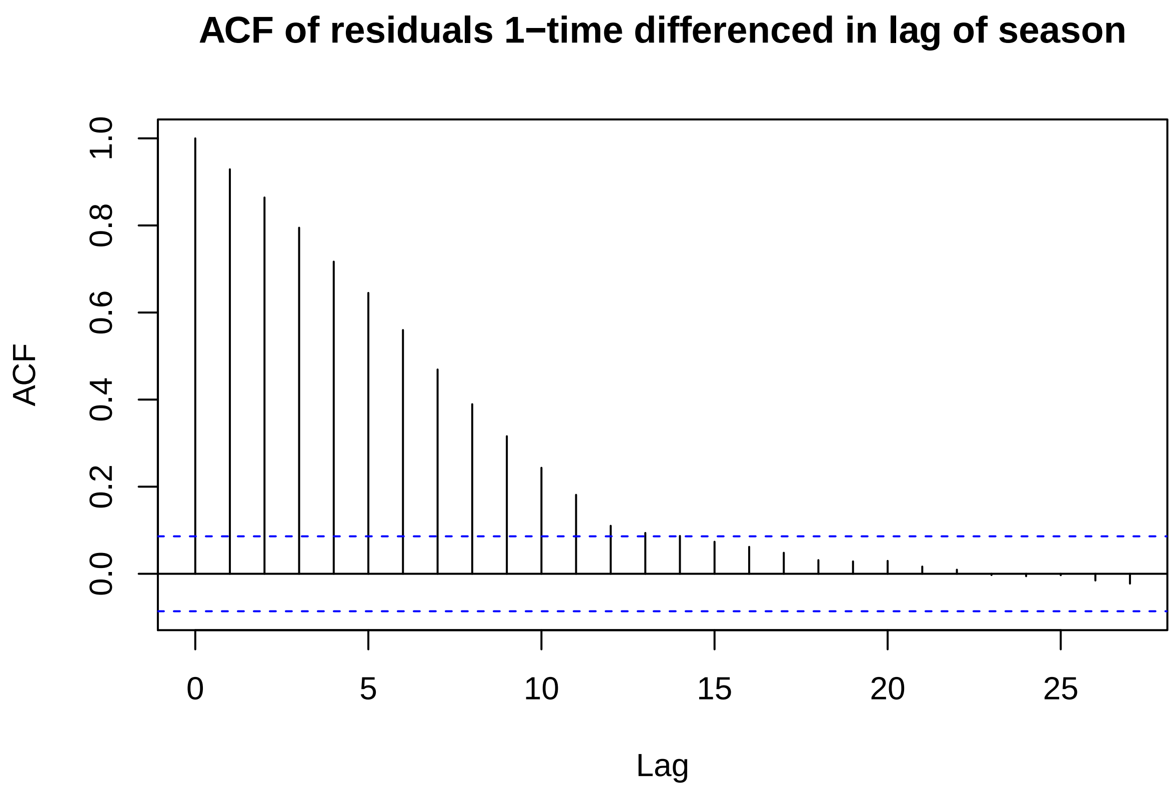

Fig 77: ACF of 1−time differencing in lag of season

We see exponential decay, so we have reached stationary. Let's get the PACF.

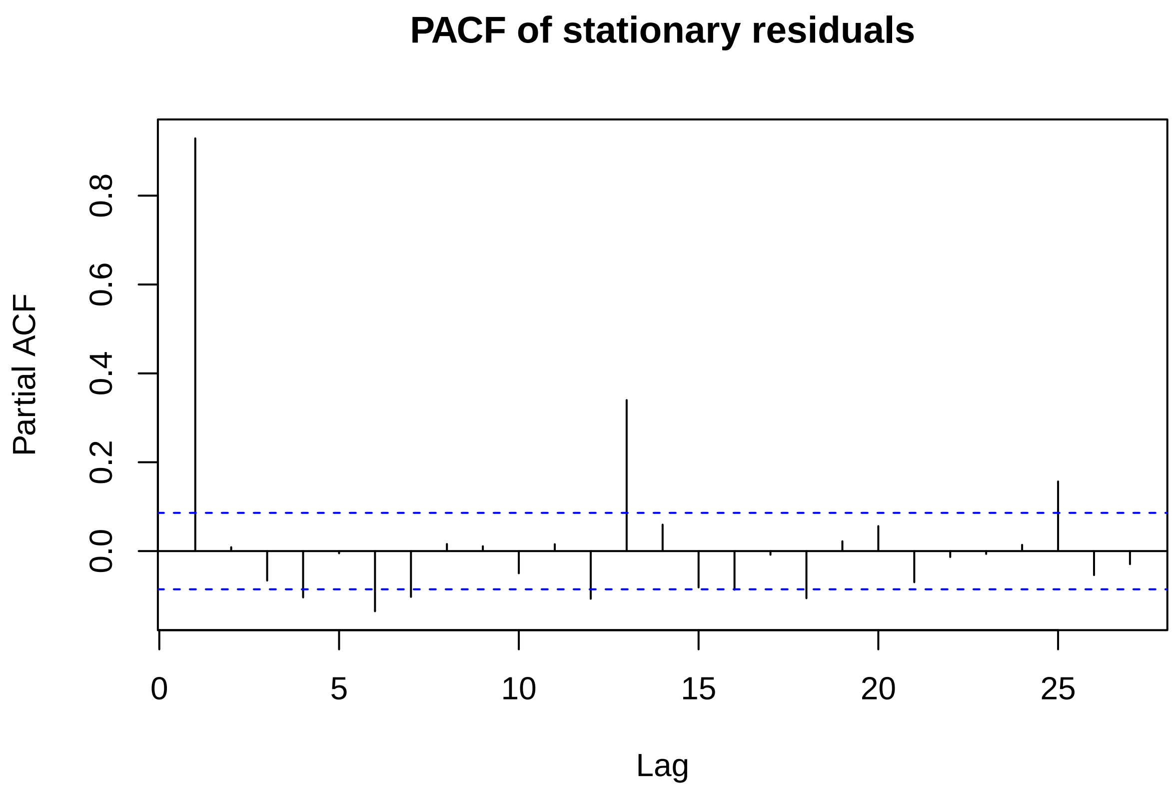

Fig 78: PACF of 1−time differencing in lag of season

Ignoring seasonal lags:

We see exponential decay in the ACF, so consider q=0. The PACF appears to cut off after lag 6, so consider p=6.

Alternatively, both ACF and PACF exponentially decay so consider:

- p=1, q=1
- p=1, q=2

Ignoring non-seasonal lags:

Alternatively, ACF cuts off after season 1 and PACF decays exponentially, so P=0 and Q=1.

Alternatively, ACF decays exponentially and PACF cuts off after season 1, so P=1 and Q=0.

Alternatively, both ACF and PACF decay exponentially, so consider:

- P=1, Q=1
- P=1, Q=2

This gives 3*4=12 models, but we propose a subset of them here.

So we propose the following SARIMA models:

- Model 1: p = 6, q = 0, P = 1, Q = 0, d = 0, D = 1, S = 12
- Model 2: p = 6, q = 0, P = 0, Q = 1, d = 0, D = 1, S = 12
- Model 3: p = 1, q = 1, P = 1, Q = 0, d = 0, D = 1, S = 12

- Model 4: p = 1, q = 1, P = 0, Q = 1, d = 0, D = 1, S = 12
- Model 5: p = 1, q = 1, P = 1, Q = 1, d = 0, D = 1, S = 12
- Model 6: p = 1, q = 2, P = 1, Q = 2, d = 0, D = 1, S = 12

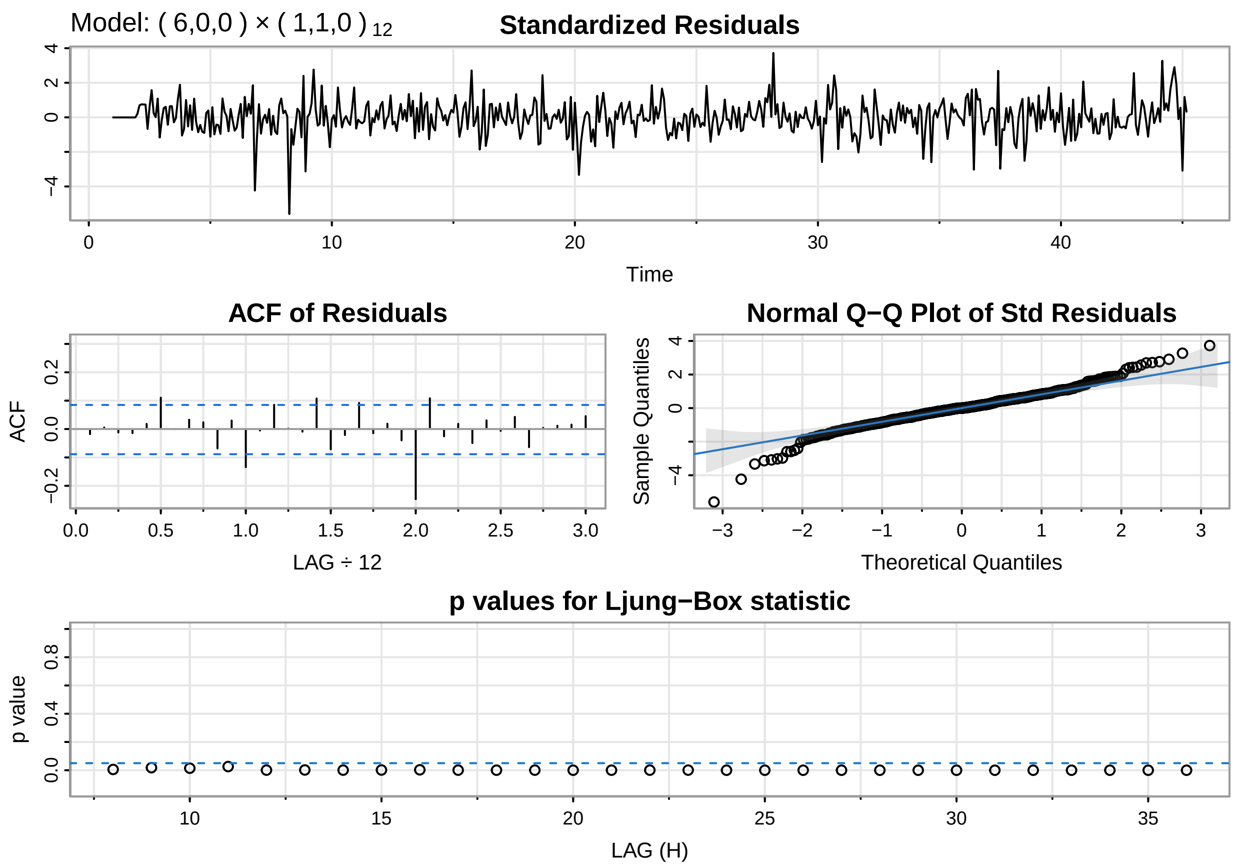

Fig 79: SARIMA model 1 diagnostics

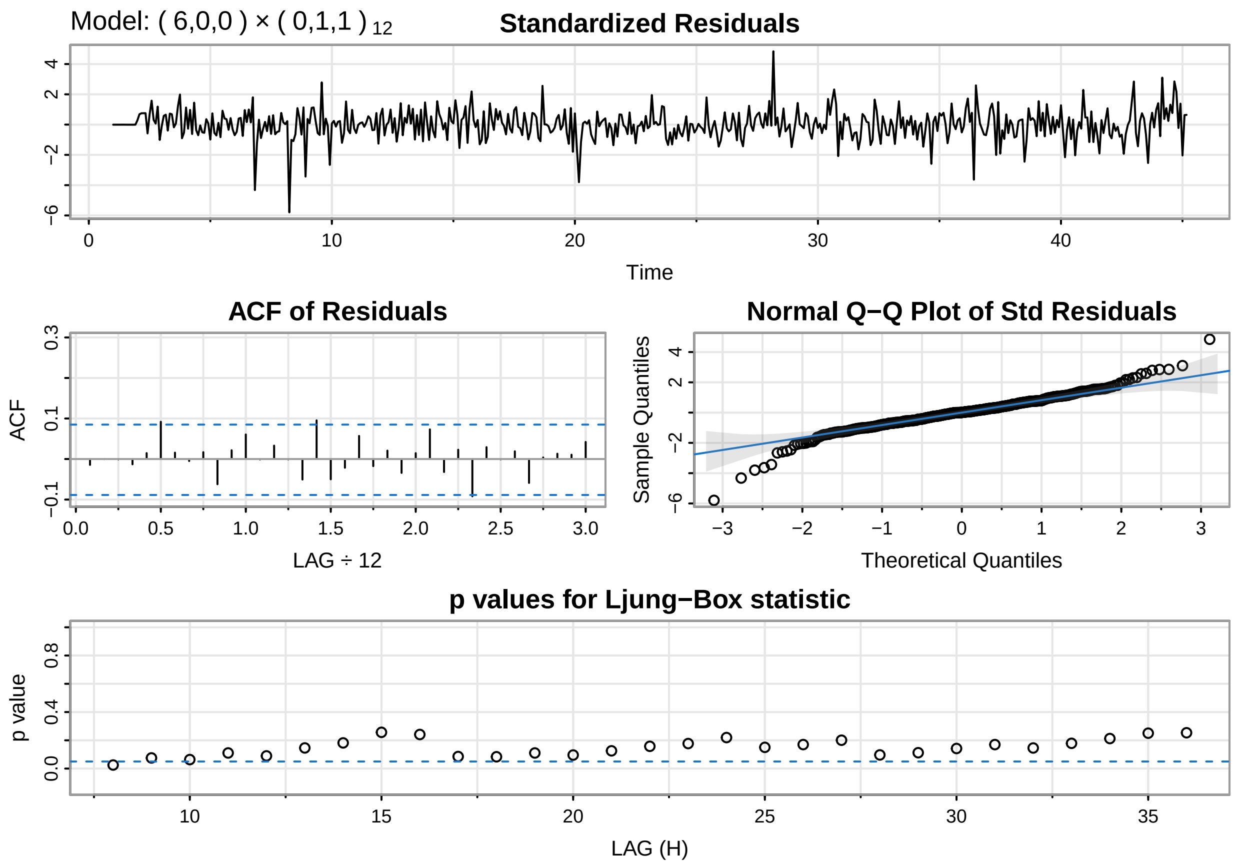

Fig 80: SARIMA model 2 diagnostics

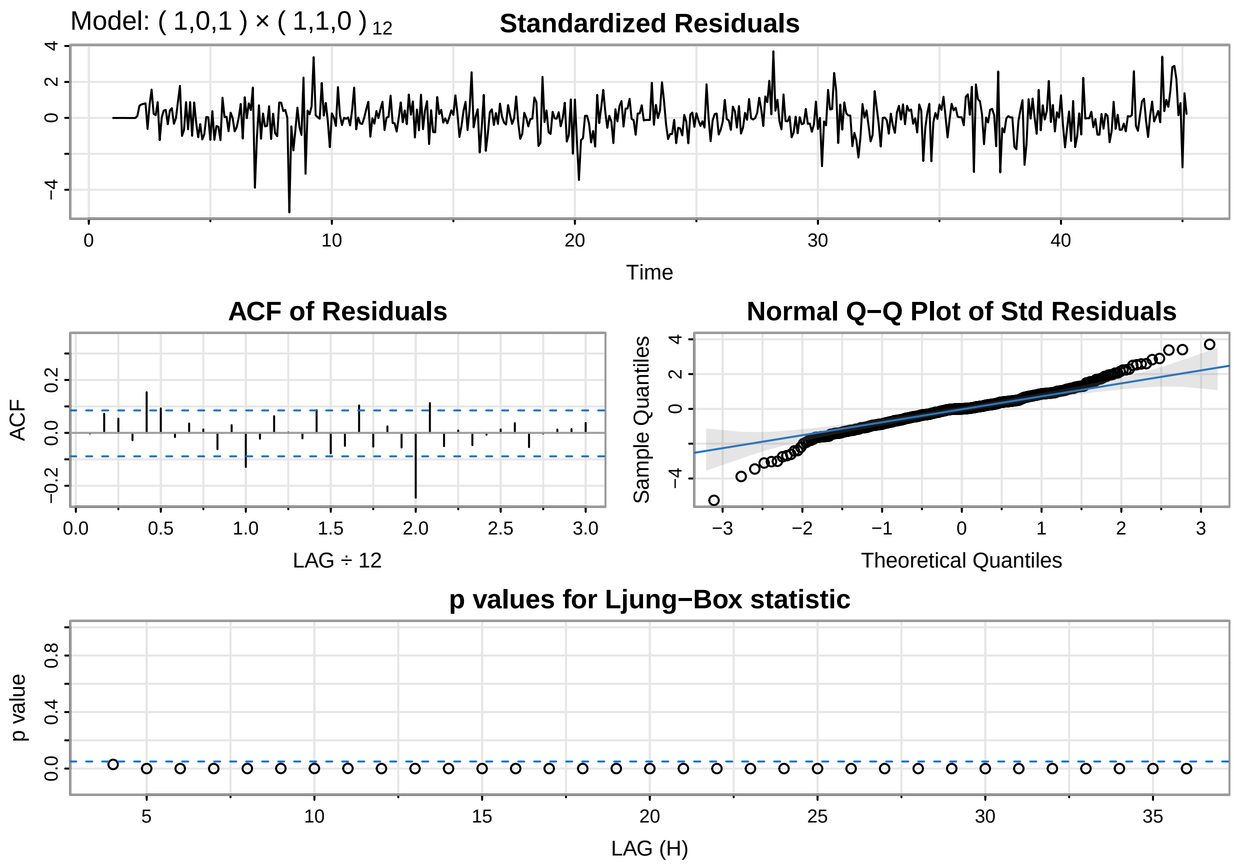

Fig 81: SARIMA model 3 diagnostics

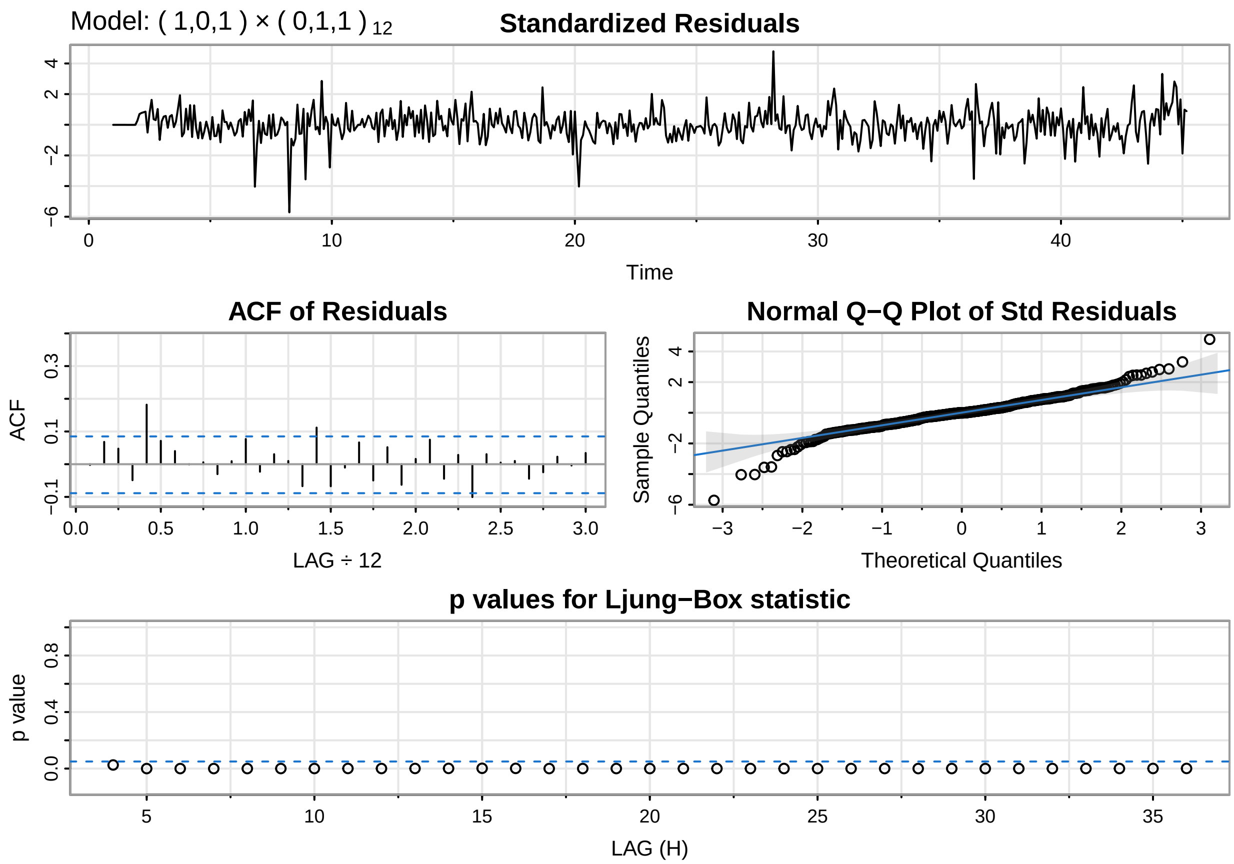

Fig 82: SARIMA model 4 diagnostics

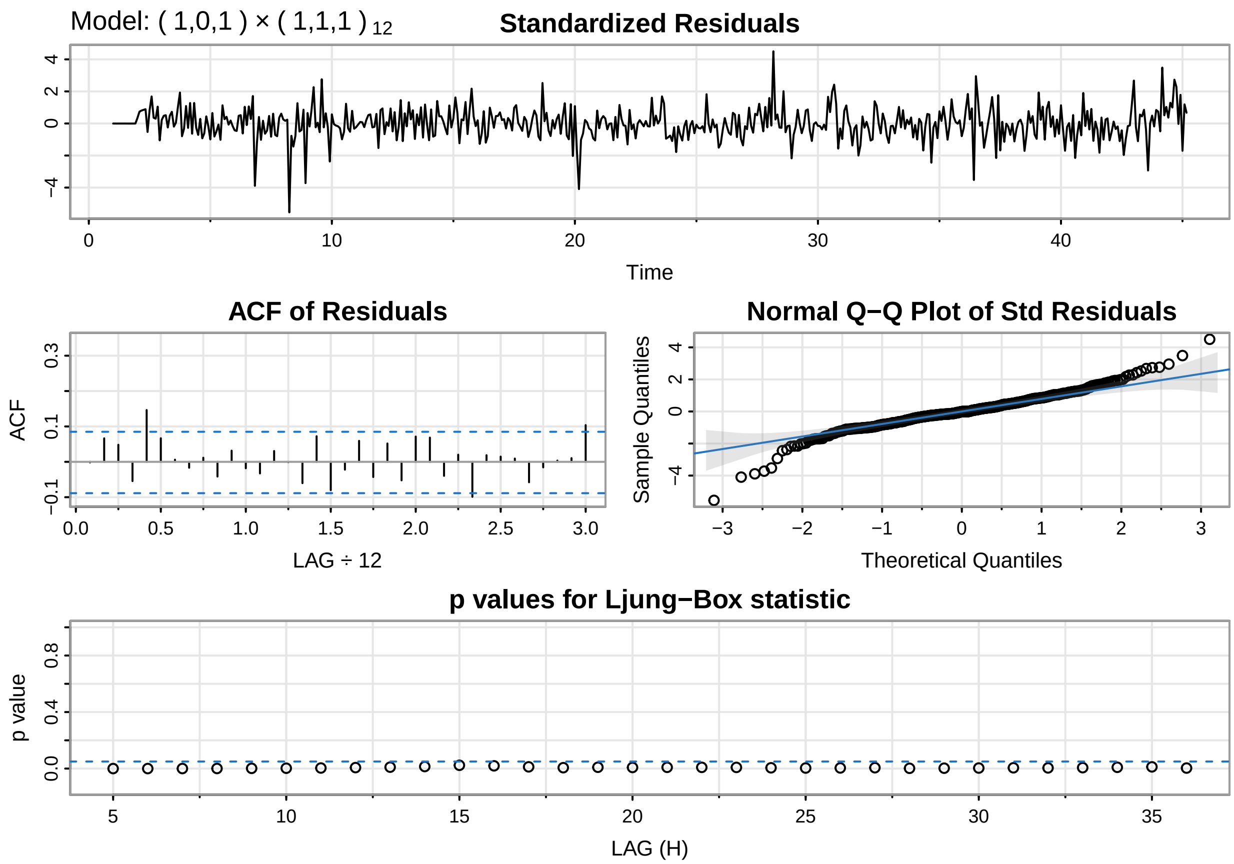

Fig 83: SARIMA model 5 diagnostics

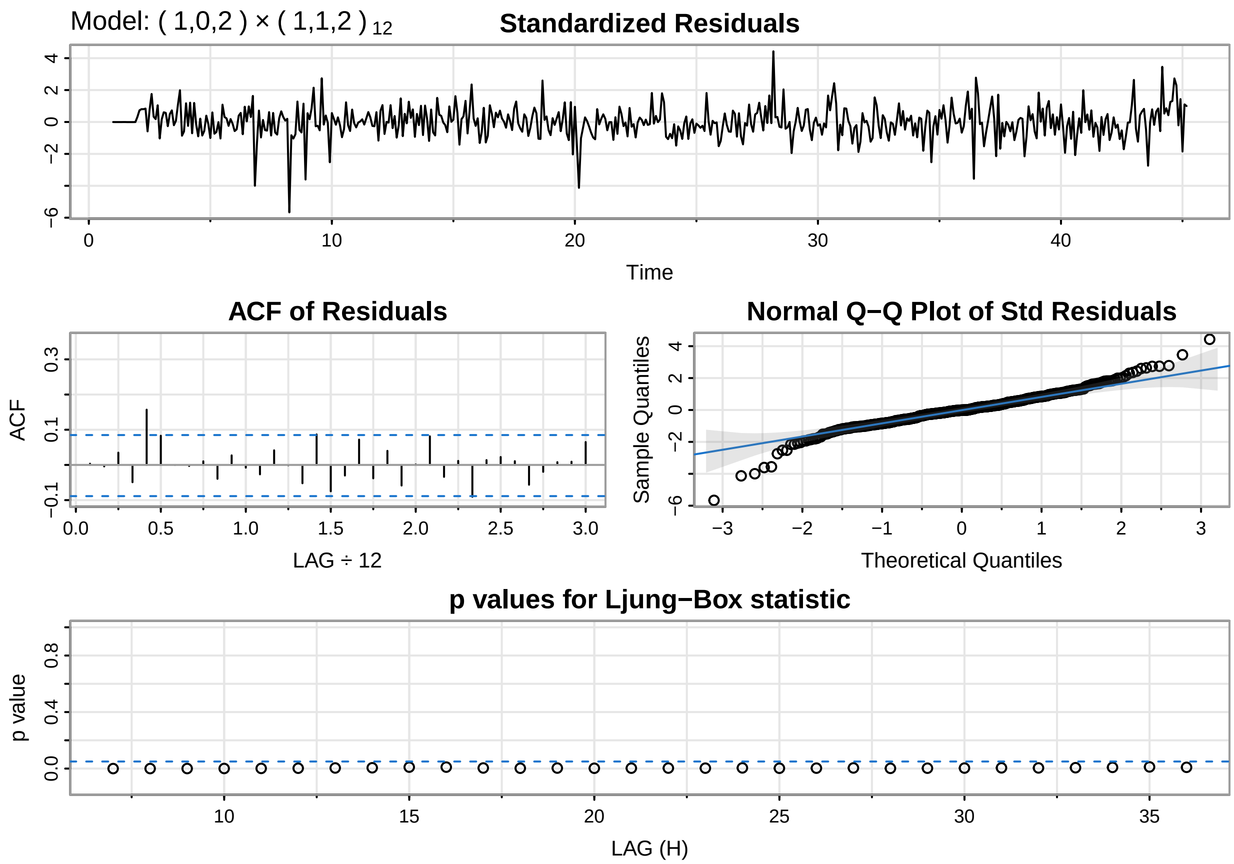

Fig 84: SARIMA model 6 diagnostics

Only model 2 passes the diagnostics test. Let's look at APSE

```
## [1] 6.77572e-05
```

Fig 85: APSE of model 2

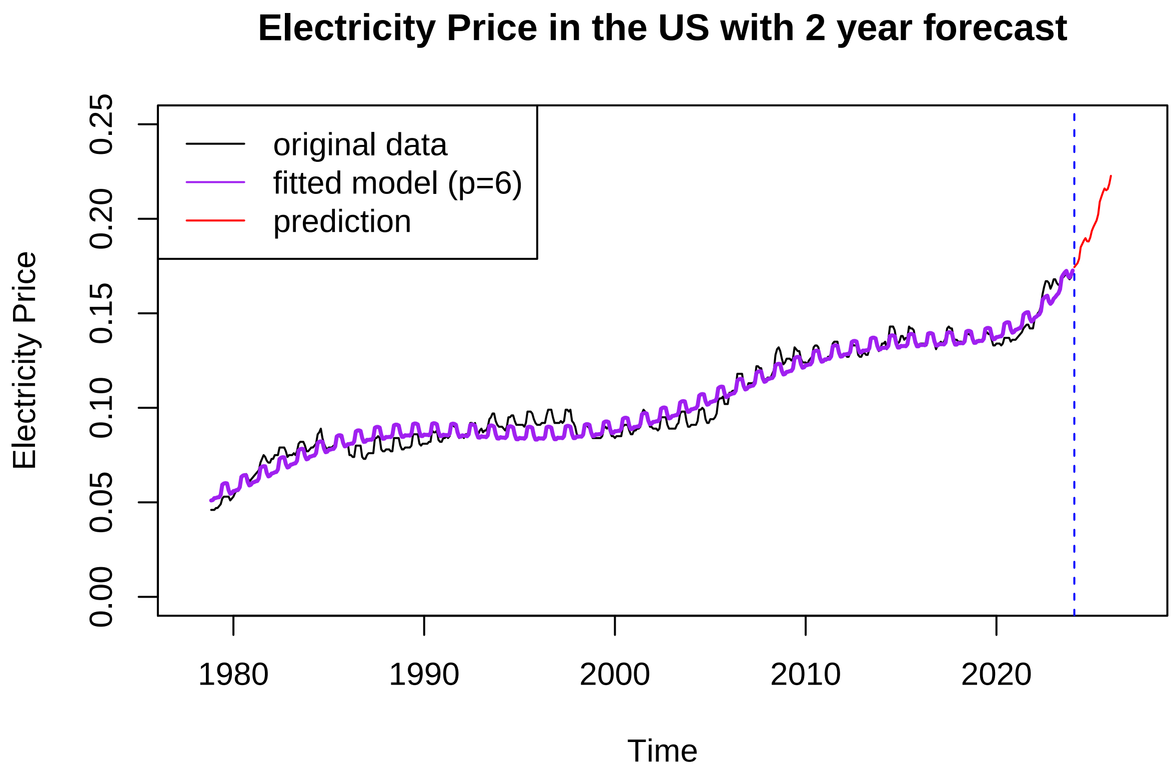

Fig 86: 2−year forecast using model 2

Now, we will try fitting SARIMA models on the residuals differenced once on lag 1.

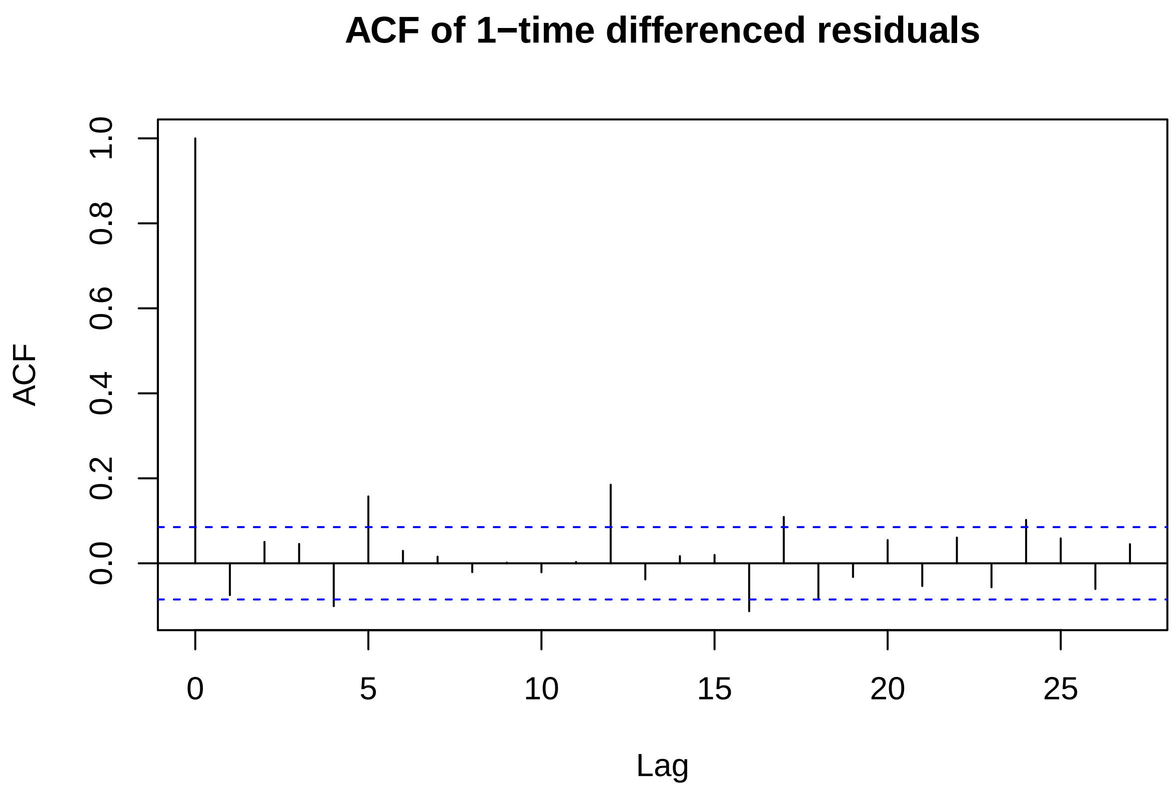

Fig 87: ACF of 1−time differenced regression residuals

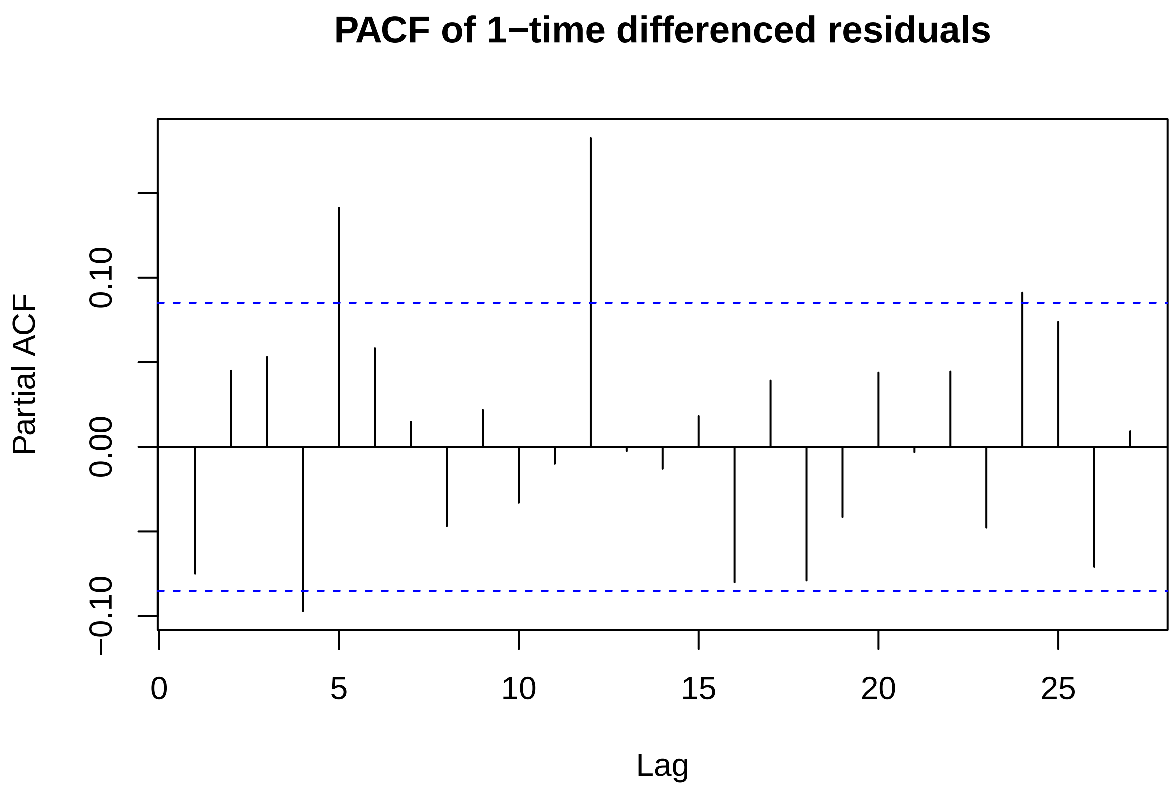

Fig 88: PACF of 1−time differenced regression residuals

Here are the values we chose for SARIMA:

- p=5, q=0 because the pacf cuts off at lag 5 and acf is exponentially decreasing
- P=1 because one could argue that the pacf cuts off at season 1
- P=2 because one could argue that the pacf cuts off at season 2
- Q=1 because one could argue that acf cuts off at season 1
- Q=2 because one could argue that acf cuts off at season 2

So we propose the models with the following values:

- p: 5
- q: 0
- d: 1 (one time differencing)
- P: 1 2 0 0
- Q: 0 0 1 2
- D: 0
- s = 12

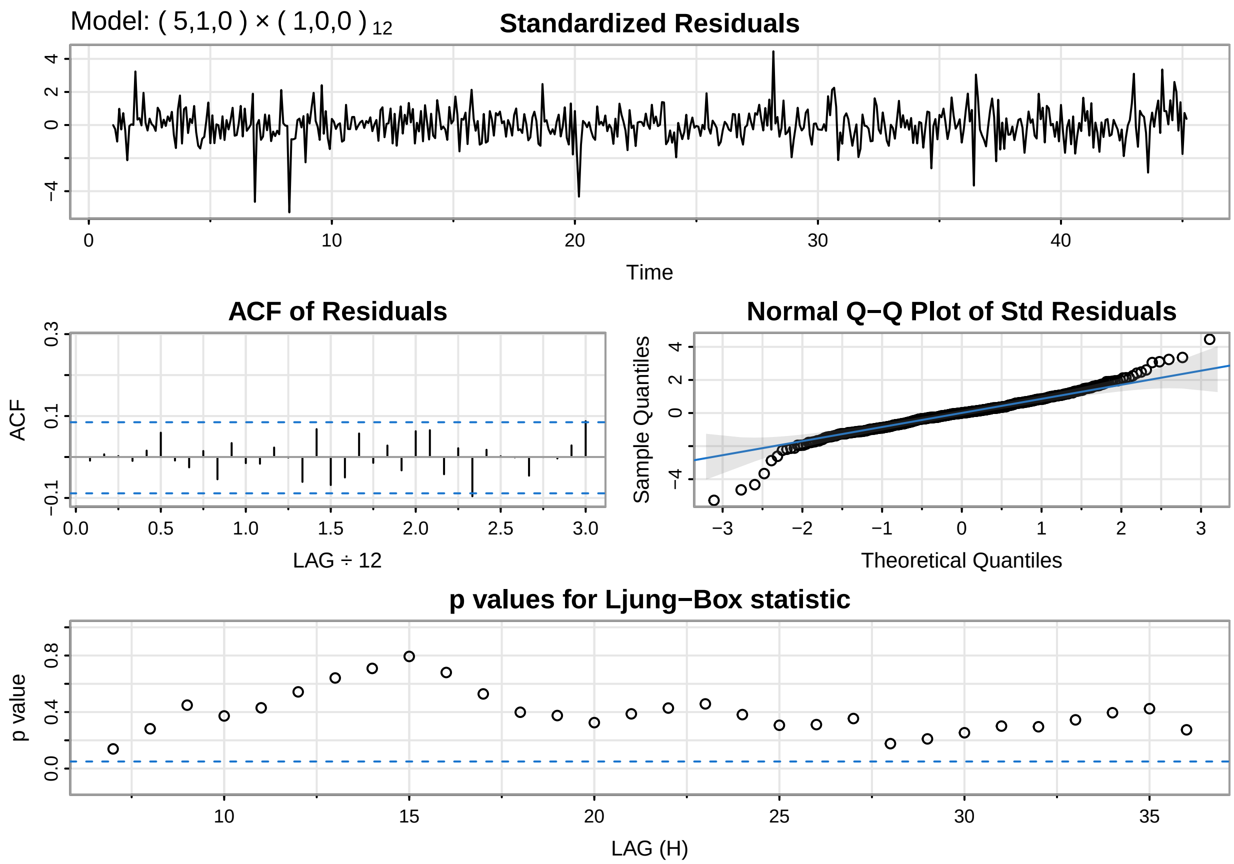

Fig 89: SARIMA model residual diagnostics

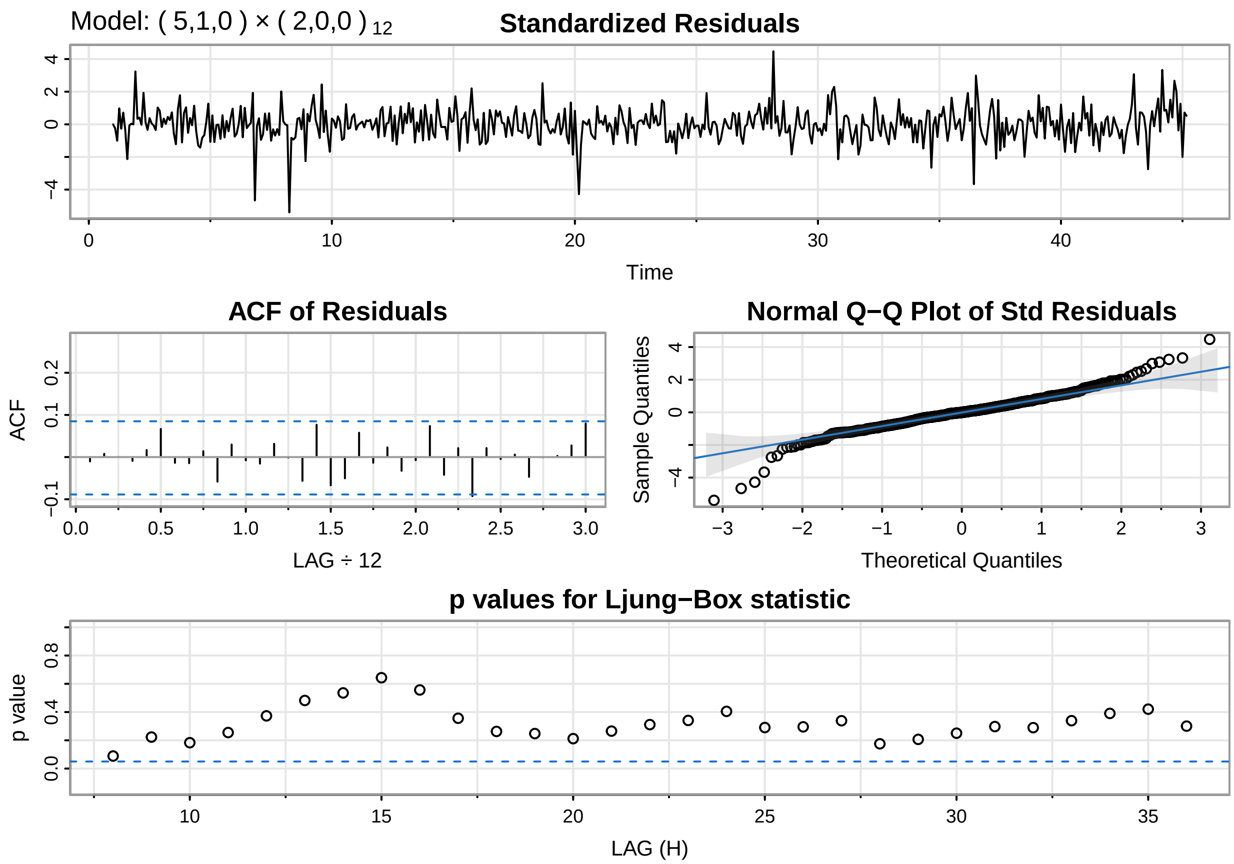

Fig 90: SARIMA model residual diagnostics

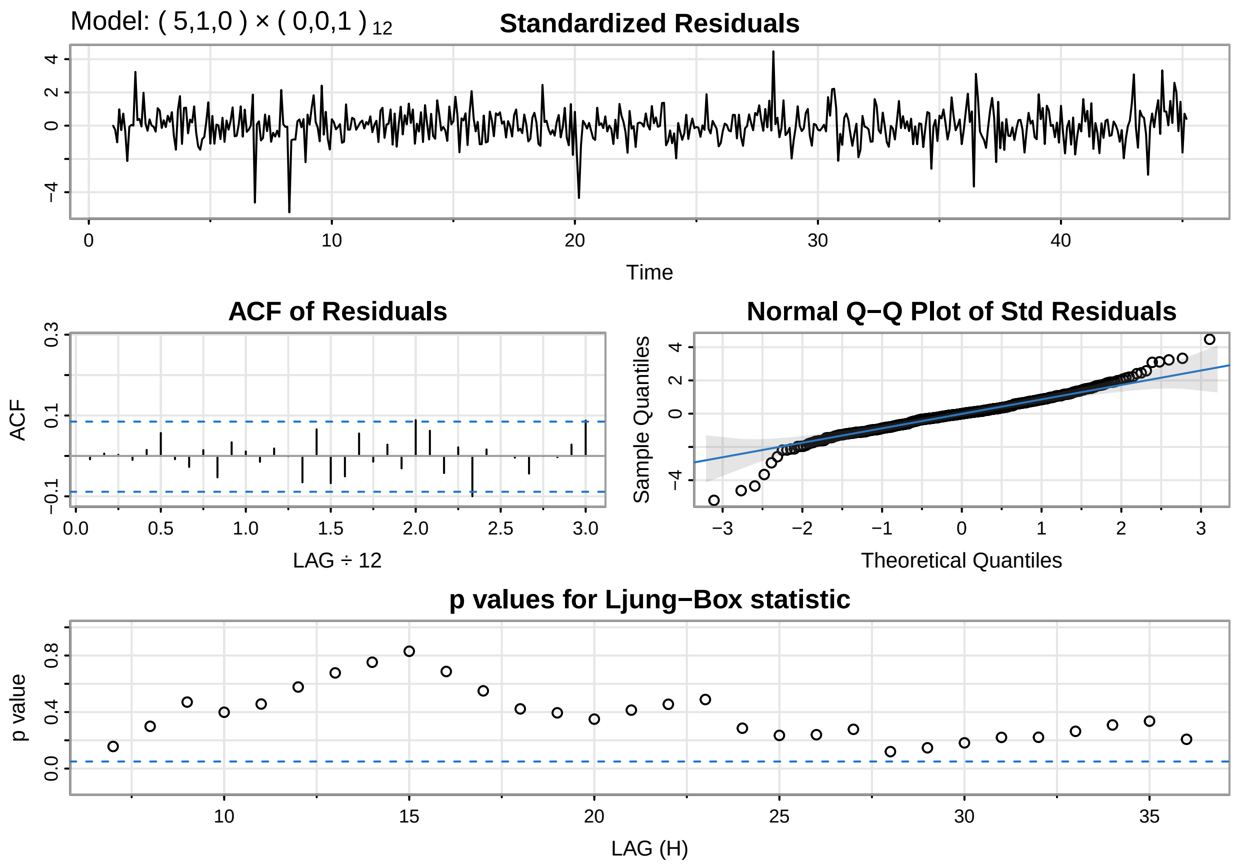

Fig 91: SARIMA model residual diagnostics

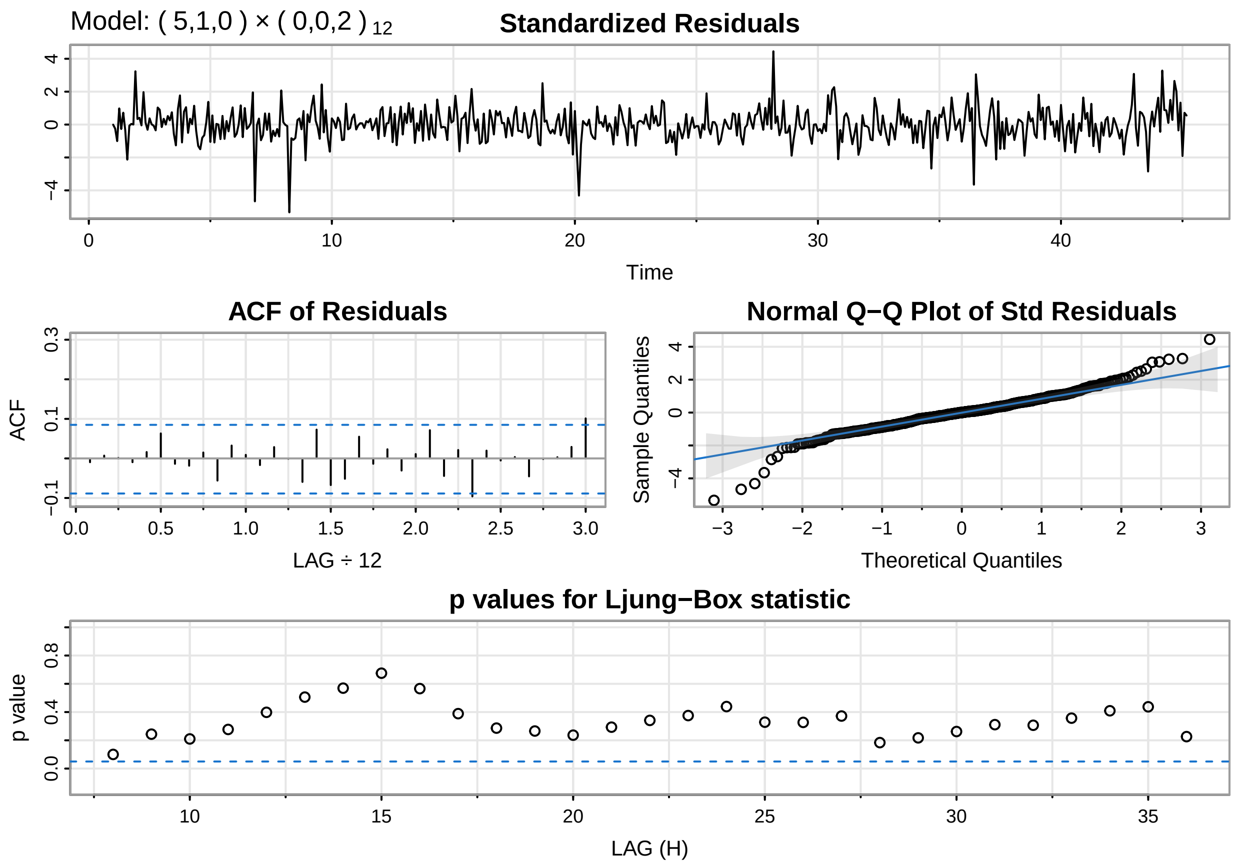

Fig 92: SARIMA model residual diagnostics

Let's check APSE for this last model, since it passed the diagnostics.

```
## [1] 0.000112479
```

Fig 93: APSE for SARIMA model

We fit this model and forecast with it.

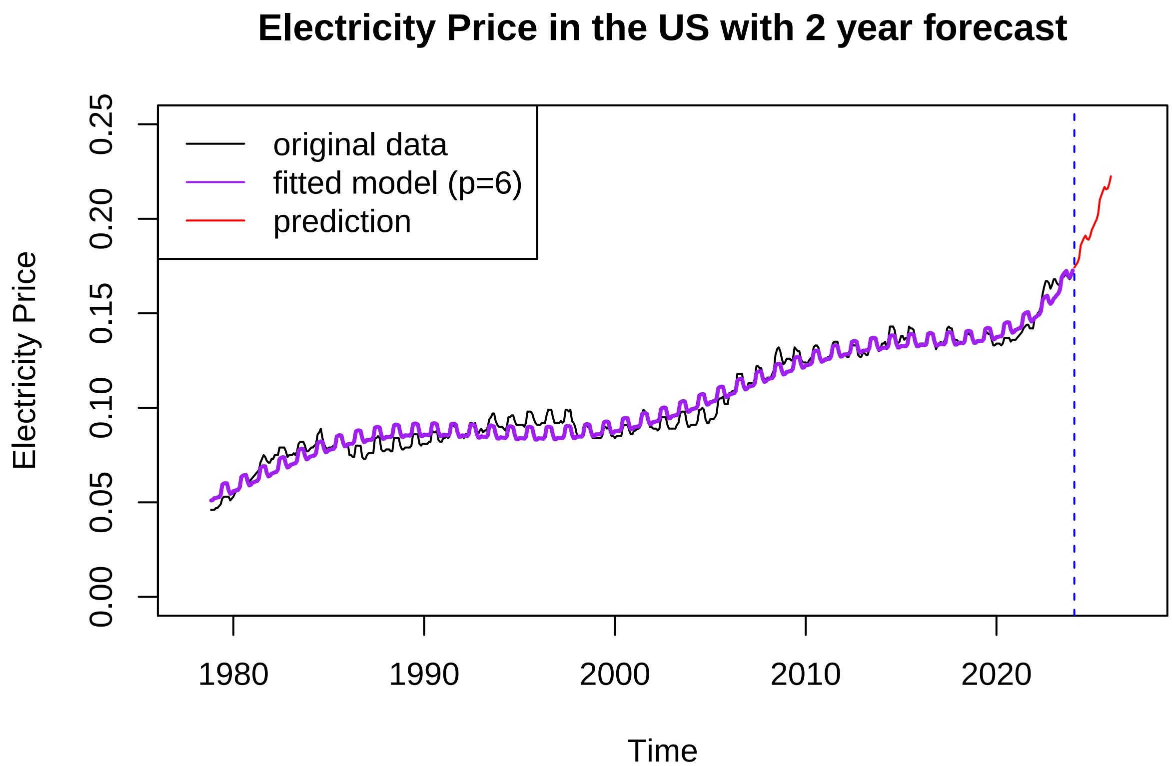

Fig 94: 2 year forecast with seasonal differencing

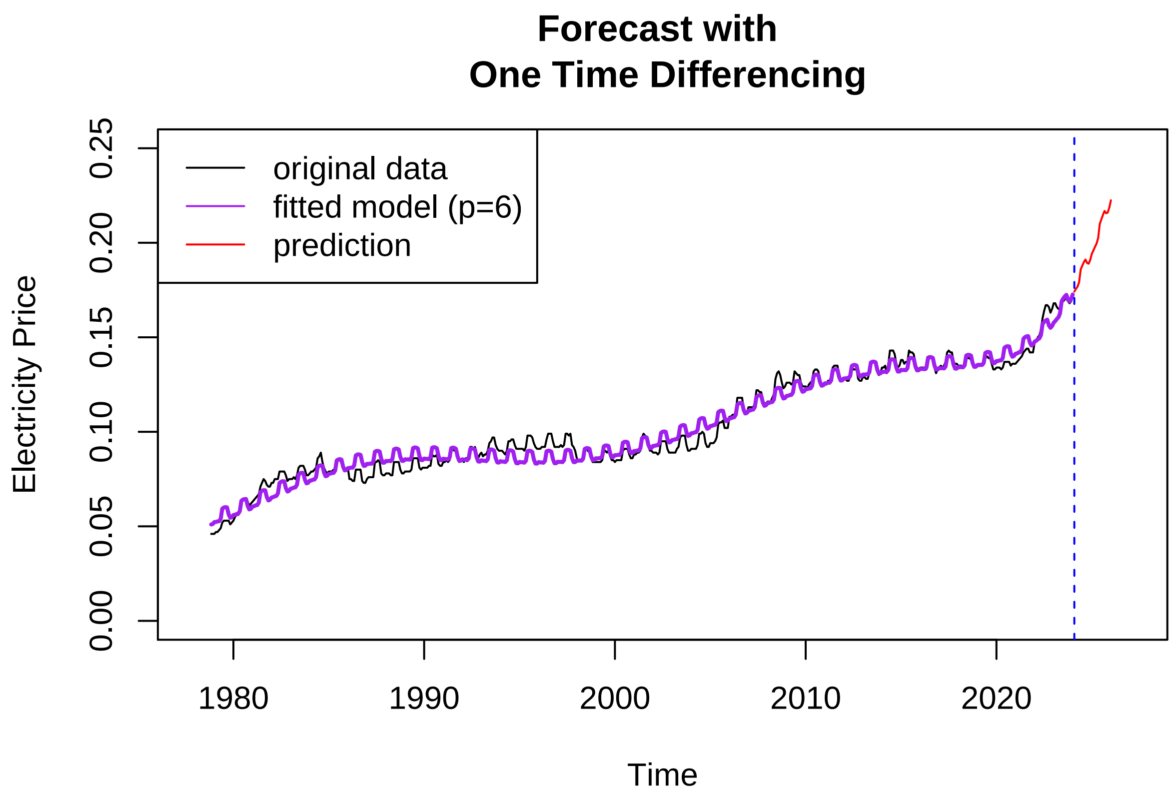

Fig 95: 2 year forecast with regular differencing

Combination Method using Additive Double Exponential Smoothing Holt-Winters Model

Making the Residuals of the DES HW Model Stationary

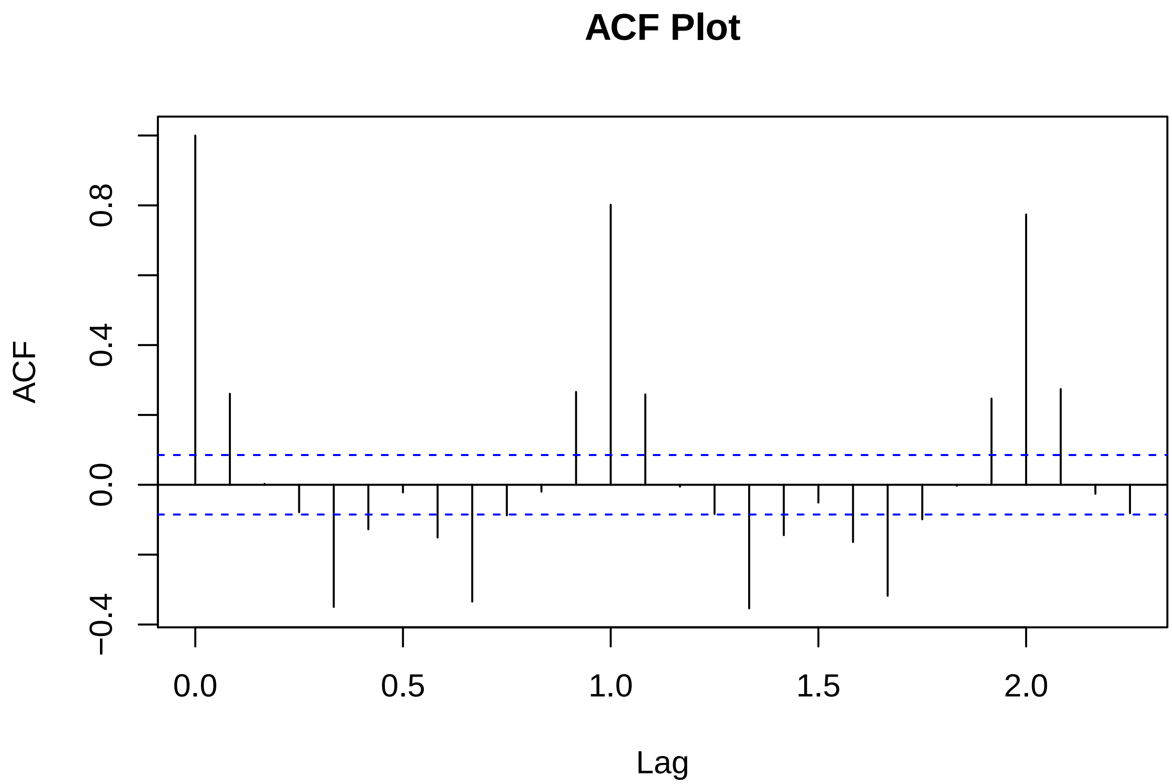

Fig 96: ACF Plot of the Holt Winters Residuals

There is linear decay in the lag of season in the ACF Plot in Figure 96, hence the residuals are non-stationary.

One time Regular Differencing

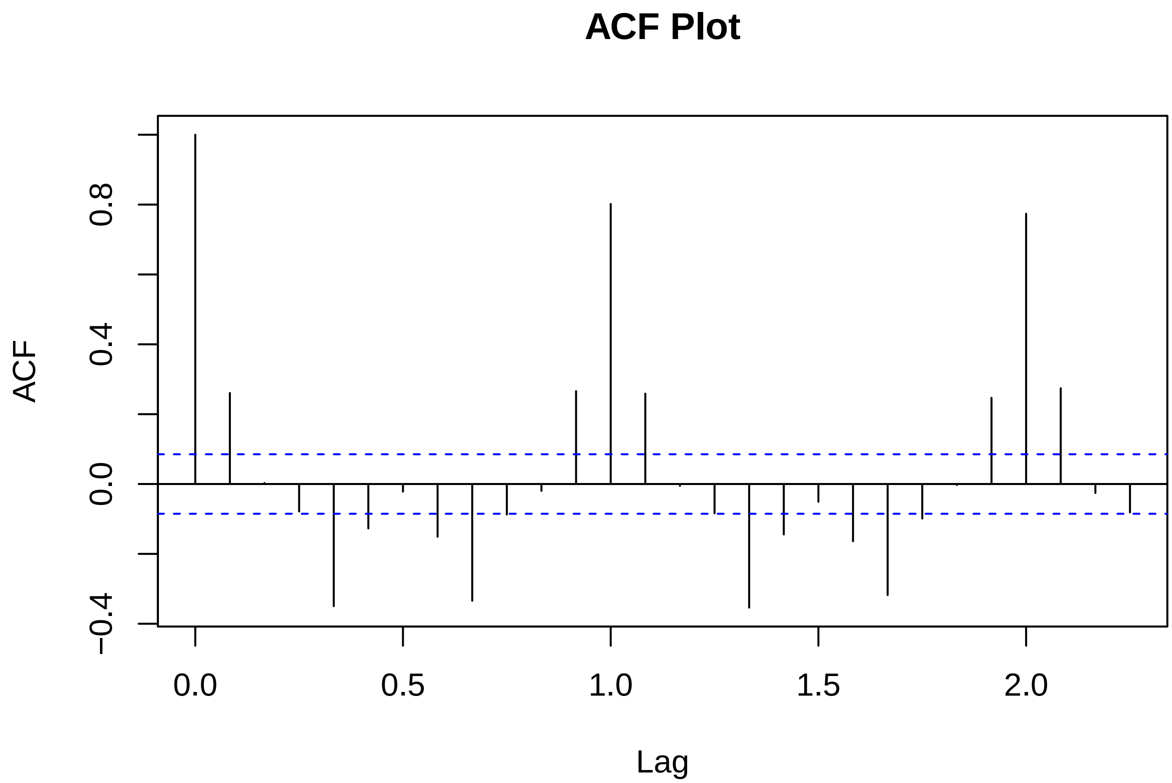

Fig 97: ACF Plot of the Holt Winters Residuals

There is linear decay in the lag of season in the ACF Plot in Figure 97, hence the residuals are non-stationary.

One Time Seasonal Differencing

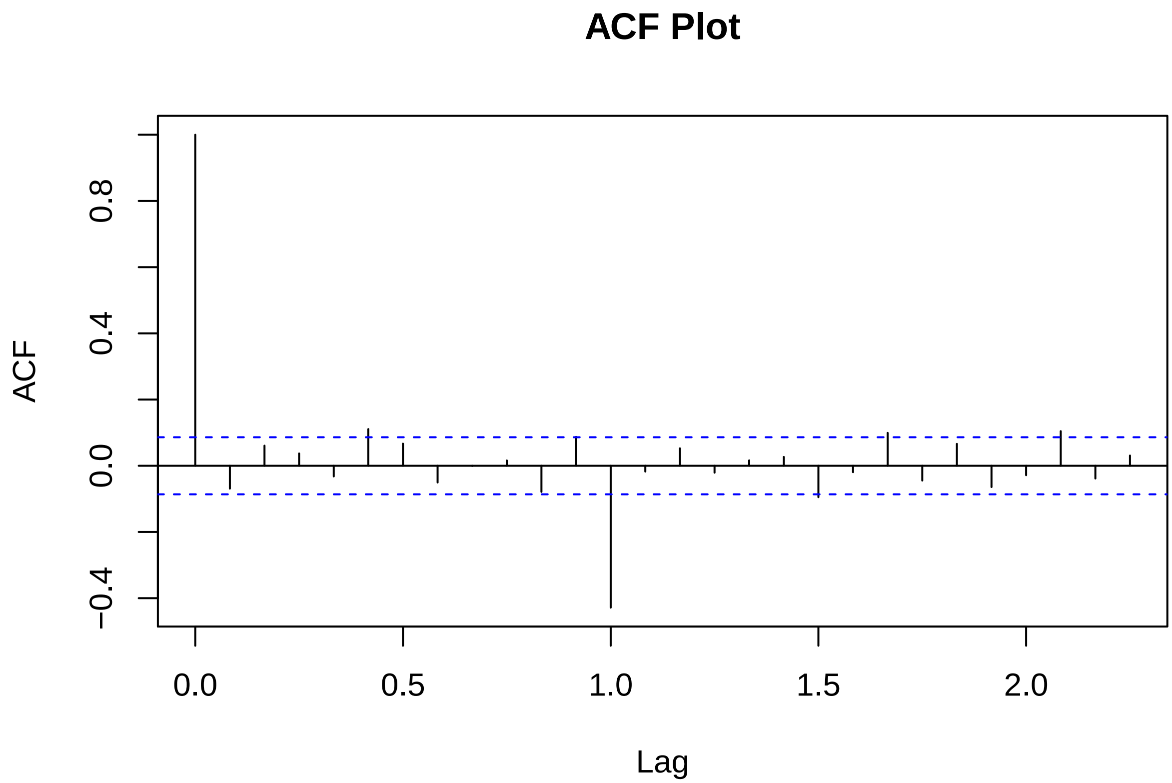

Fig 98: ACF Plot of the Holt Winters Residuals

There are no signs of non-stationarity in the above ACF plot in Figure 98, hence our residuals are now stationary.

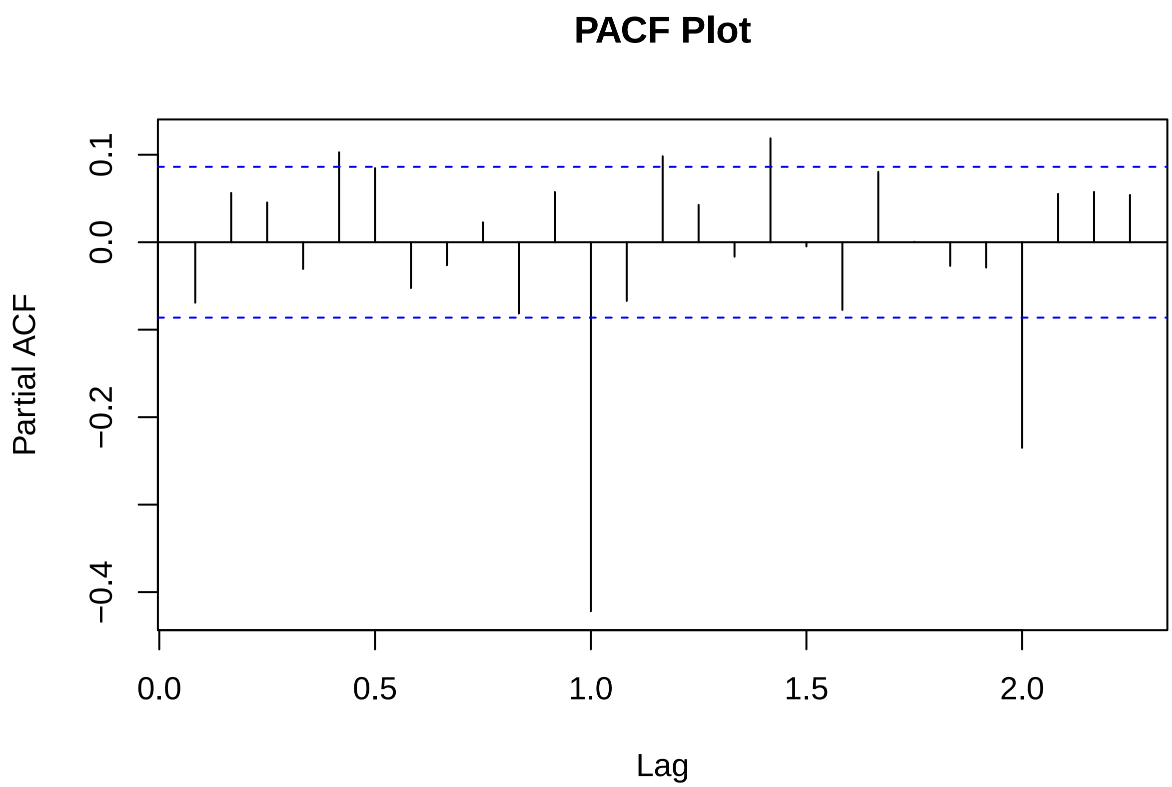

Fig 99: PACF Plot of the Holt Winters Residuals

Model Proposal Since we seasonally differenced our residuals, immediately we know that D=1, s=12 and d=0.

Looking at the above ACF and PACF plots, we can propose the following values for p,q,P and Q.

For p and q, we ignore seasonal lags:

The following are in reference to the ACF Plot in Figure 98 and the PACF plot in Figure 99.

One can argue that the ACF plot cuts off after lag 5 and there is immediate exponential decay in the PACF plot, therefore I will propose $p = 0$ and $q = 5$.

One can argue that there is exponential decay in the ACF plot and the PACF plot cuts off after lag 5, therefore I will propose $p = 5$ and $q = 0$.

One can argue that there is exponential decay in both the ACF plot and the PACF plot, therefore I will propose $p = 1$ and $q = 1$

For P and Q we focus on the seasonal lags:

One can argue that the ACF plot cuts off after lag 5 and there is immediate exponential decay in the PACF plot, therefore I will propose $p = 0$ and $q = 1$.

One can argue that there is exponential decay in the ACF plot and the PACF plot cuts off after lag 2, therefore I will propose $p = 2$ and $q = 0$.

One can argue that there is exponential decay in the lag of season in both the ACF plot and the PACF plot, therefore I will propose $P = 1$ and $Q = 1$

Therefore I propose to fit a total of $3 \times 3 = 9$ models, those being:

Model 1: $SARIMA(0,0,5) \times (0,1,1)_{12}$

Model 2: $SARIMA(0,0,5) \times (2,1,0)_{12}$

Model 3: $SARIMA(0,0,5) \times (1,1,1)_{12}$

Model 4: $SARIMA(5,0,0) \times (0,1,1)_{12}$

Model 5: $SARIMA(5,0,0) \times (2,1,0)_{12}$

Model 6: $SARIMA(5,0,0) \times (1,1,1)_{12}$

Model 7: $SARIMA(1,0,1) \times (0,1,1)_{12}$

Model 8: $SARIMA(1,0,1) \times (2,1,0)_{12}$

Model 9: $SARIMA(1,0,1) \times (1,1,1)_{12}$

Fitting the Combination HW SARIMA Models

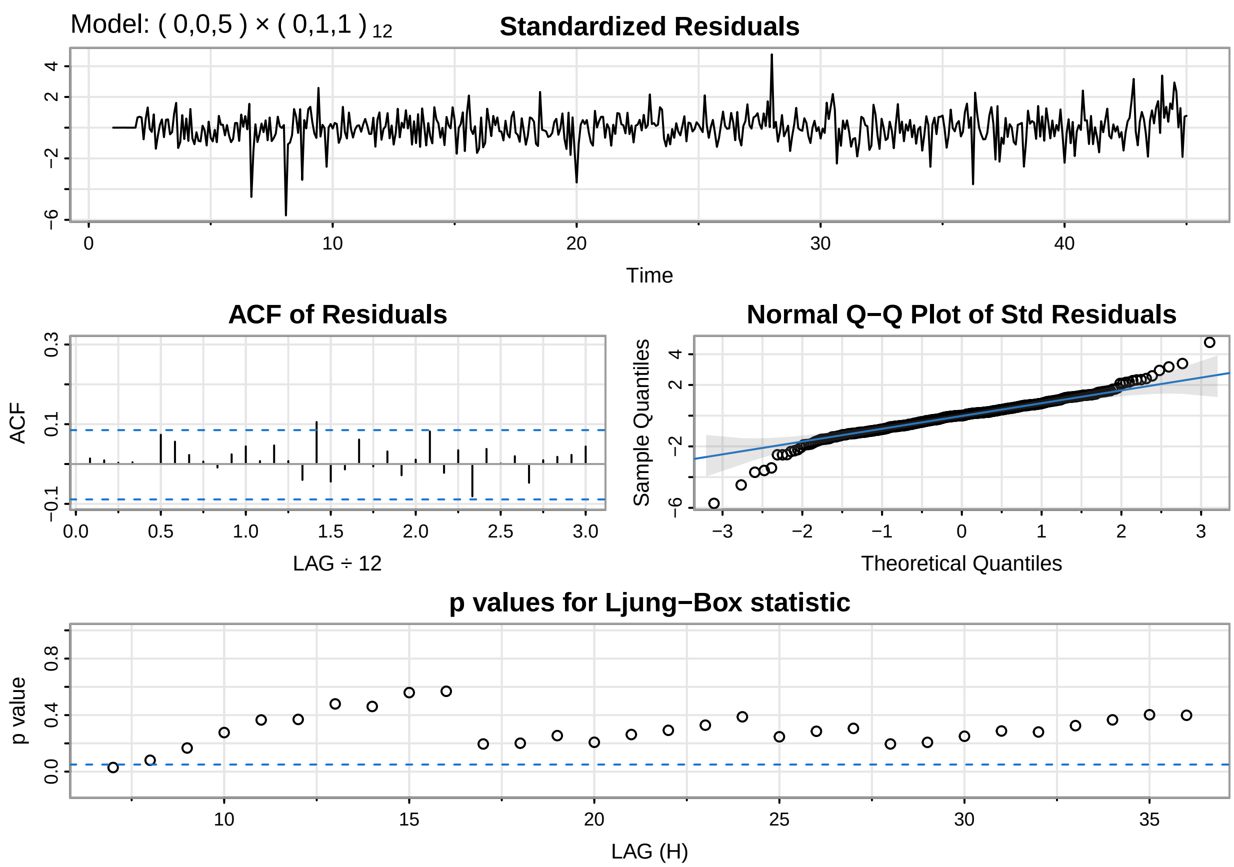

Fig 100: Residuals Diagnostics for SARIMA

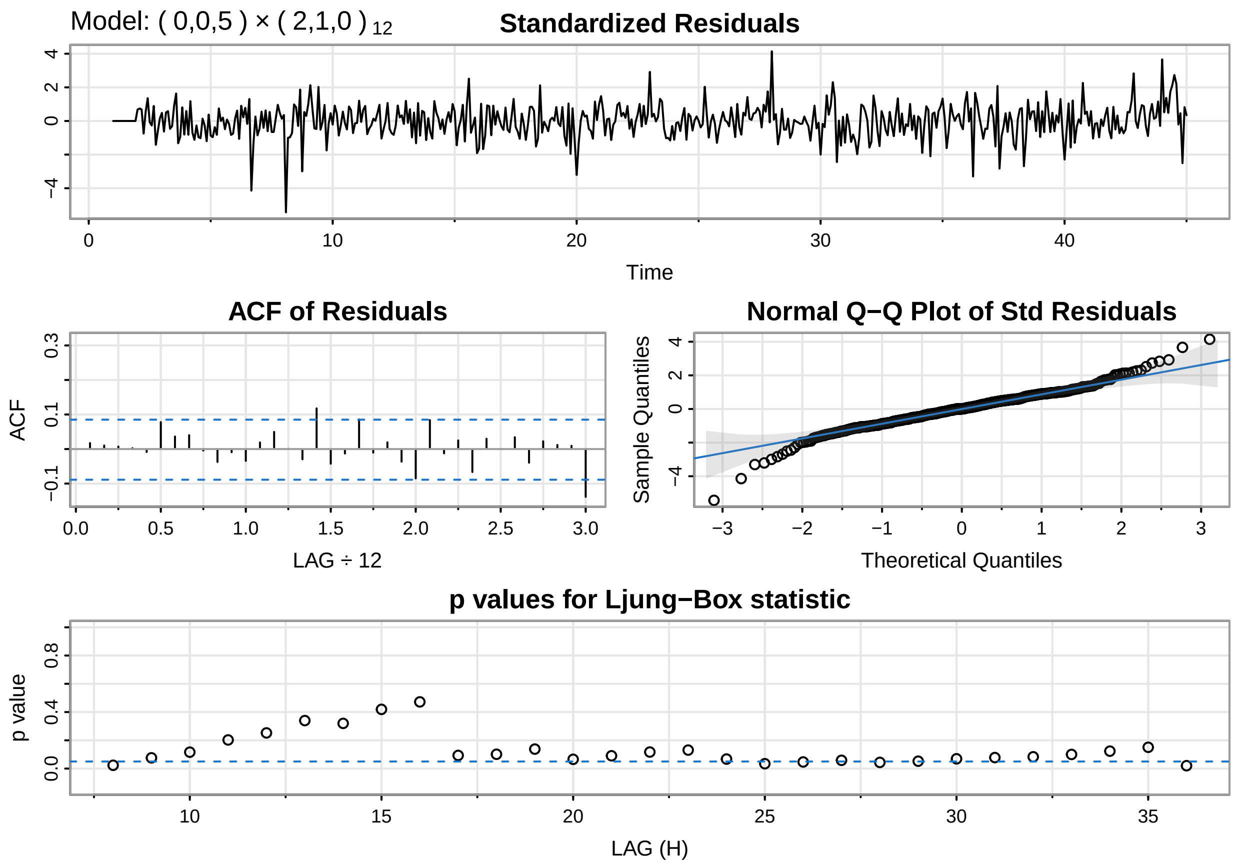

Fig 101: Residual Diagnostics for SARIMA

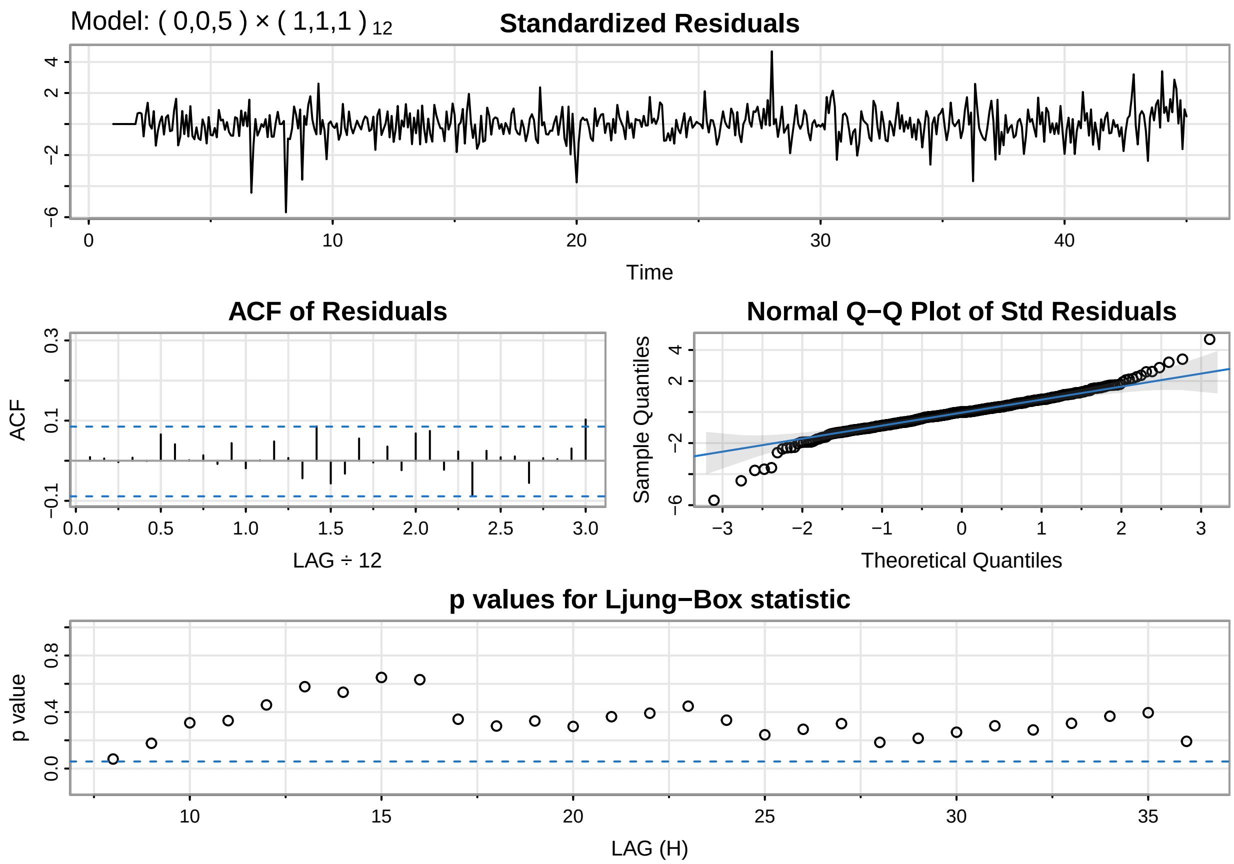

Fig 102: Residual Diagnostics for SARIMA

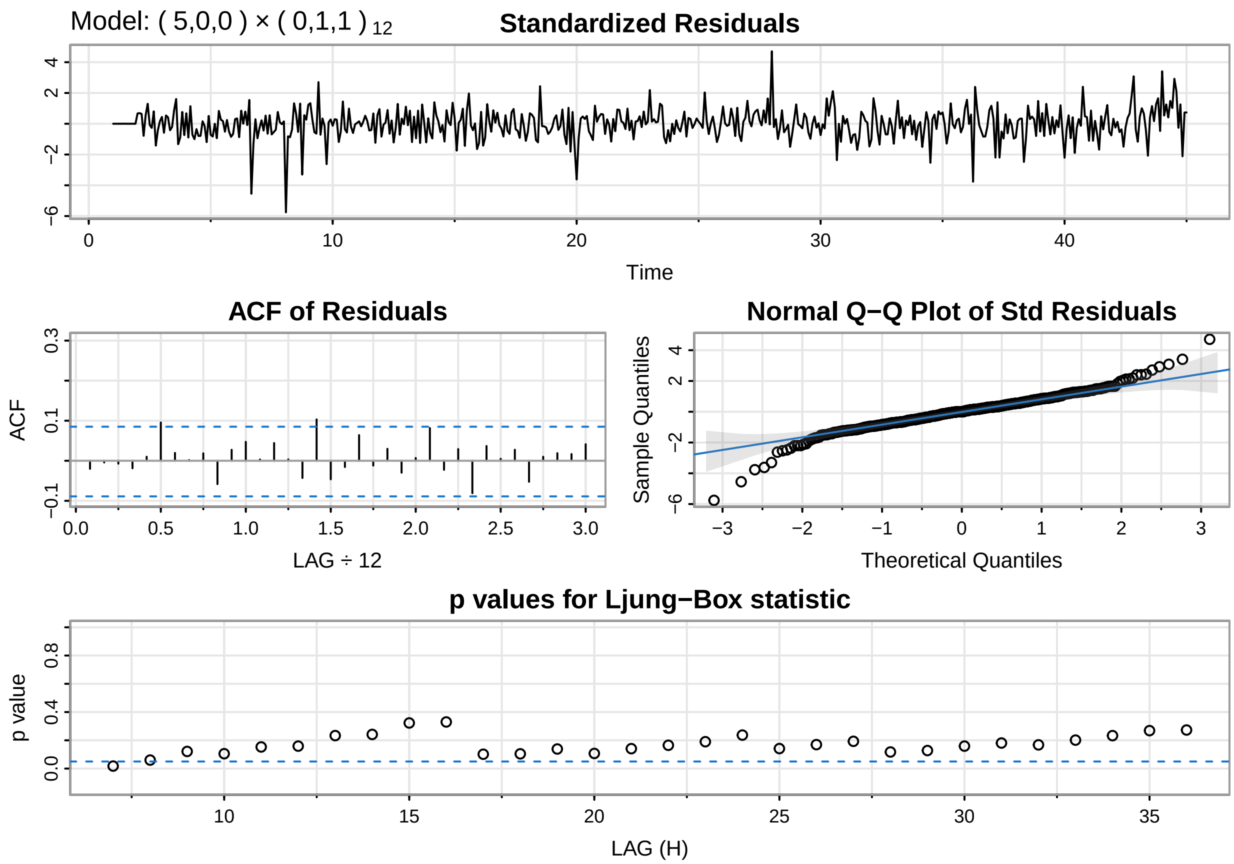

Fig 103: Residual Diagnostics for SARIMA

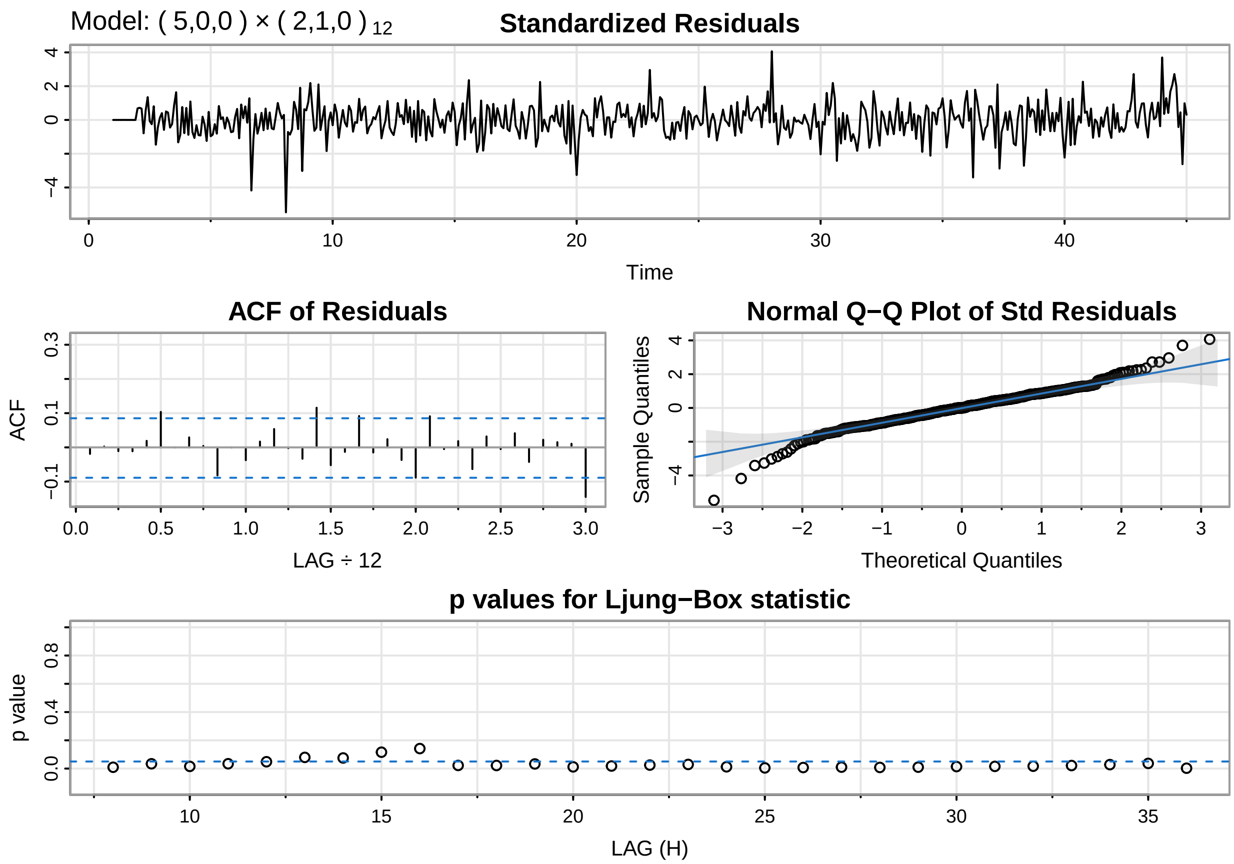

Fig 104: Residual Diagnostics for SARIMA

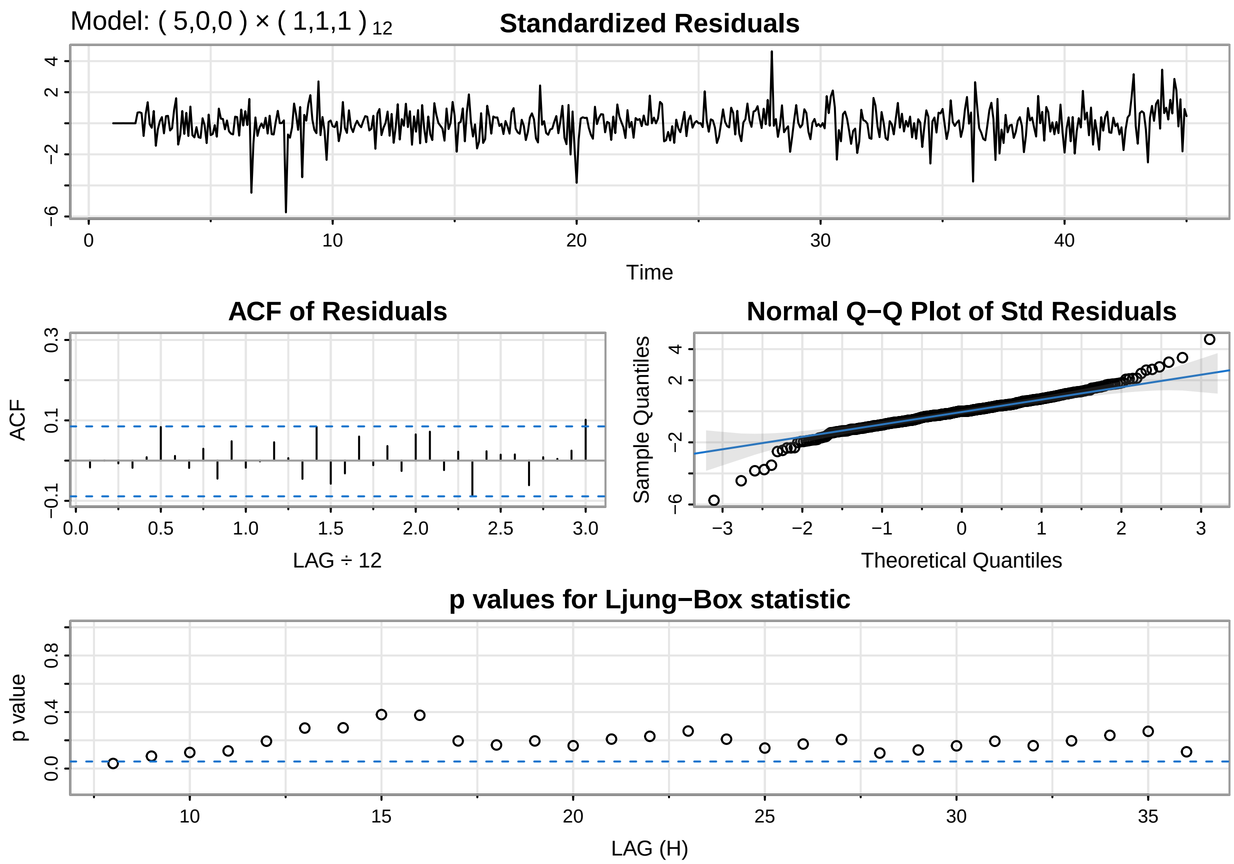

Fig 105: Residual Diagnostics for SARIMA

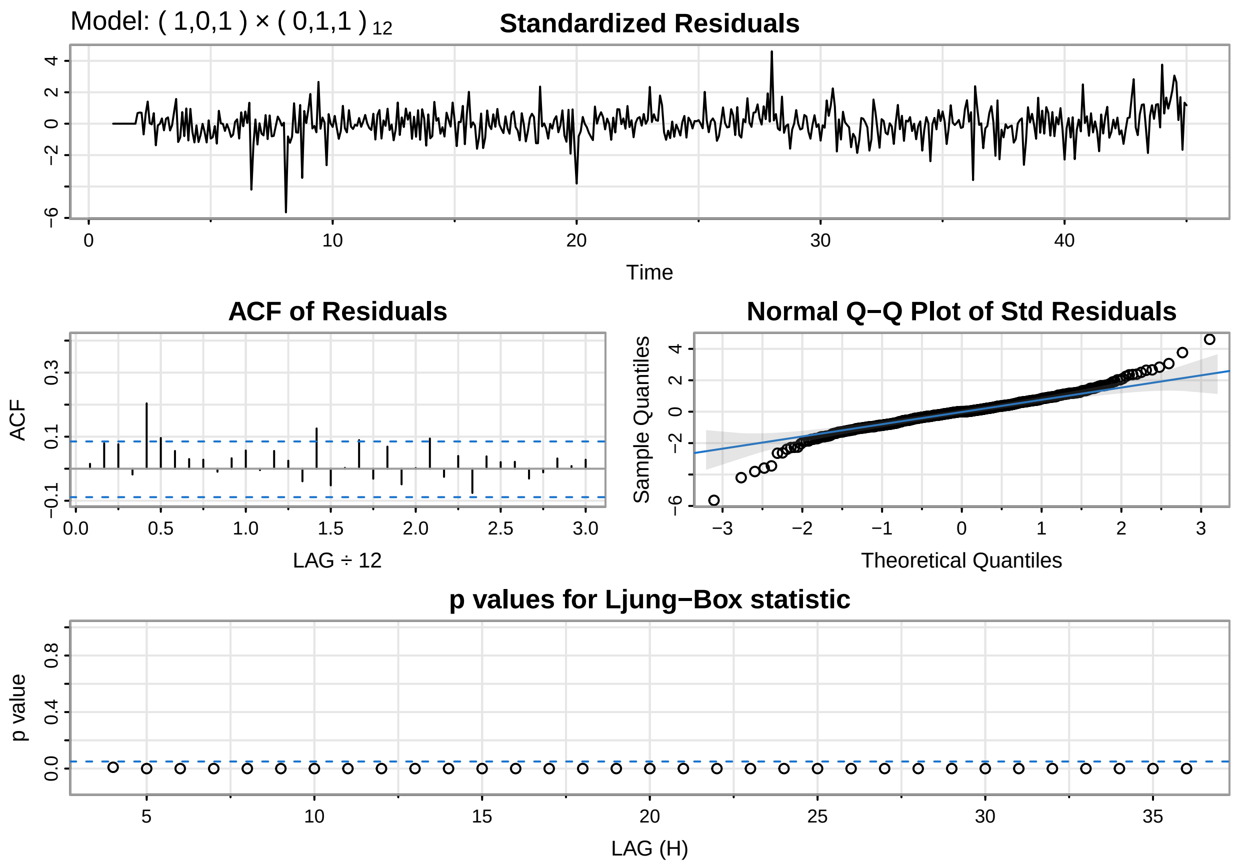

Fig 106: Residual Diagnostics for SARIMA

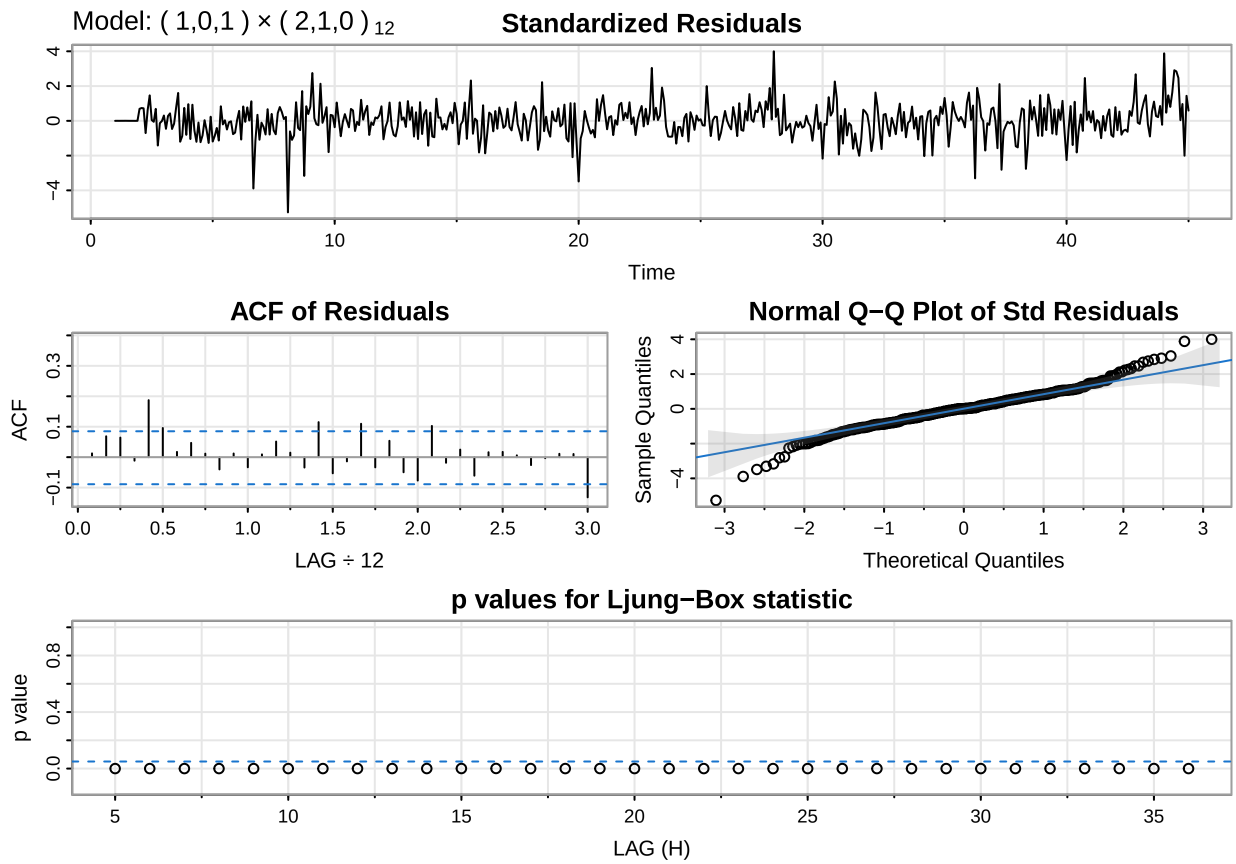

Fig 107: Residual Diagnostics for SARIMA

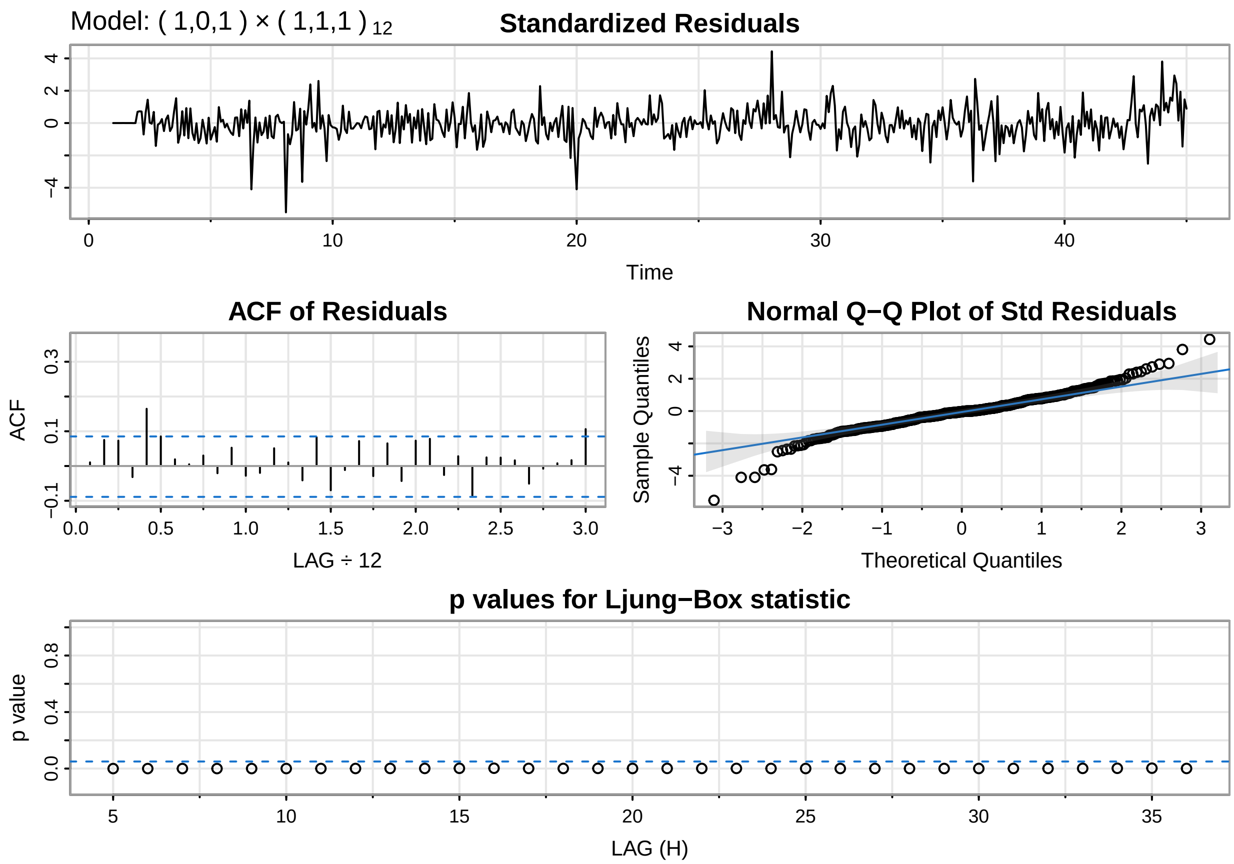

Fig 108: Residual Diagnostics for SARIMA

From the p-values for the Ljung-Box statistic in Figures 104,106,107 and 108, most p-values are below the 0.05 dotted line in the above plots so we will not further investigate models 5,7,8 and 9. We will continue investigation in models 1,2,3,4, and 6 as the majority of their p-values sit above the 0.05 line as can be seen in Figures 100,101,102,103 and 105.

APSE Calculations for Proposed Models

Calculating the APSE value for Model 1 using the train/test APSE criterion method.

```
## [1] 9.688002e-05
```

Fig 109: APSE values for model 1

Calculating the APSE value for Model 2 using the train/test APSE criterion method.

```
## [1] 9.696472e-05
```

Fig 110: APSE values for model 2

Calculating the APSE value for Model 3 using the train/test APSE criterion method.

```
## [1] 9.805467e-05
```

Fig 111: APSE values for model 3

Calculating the APSE value for Model 4 using the train/test APSE criterion method.

```
## [1] 9.727467e-05
```

Fig 112: APSE values for model 4

Calculating the APSE value for Model 6 using the train/test APSE criterion method.

```
## [1] 9.832425e-05
```

Fig 113: APSE values for model 6

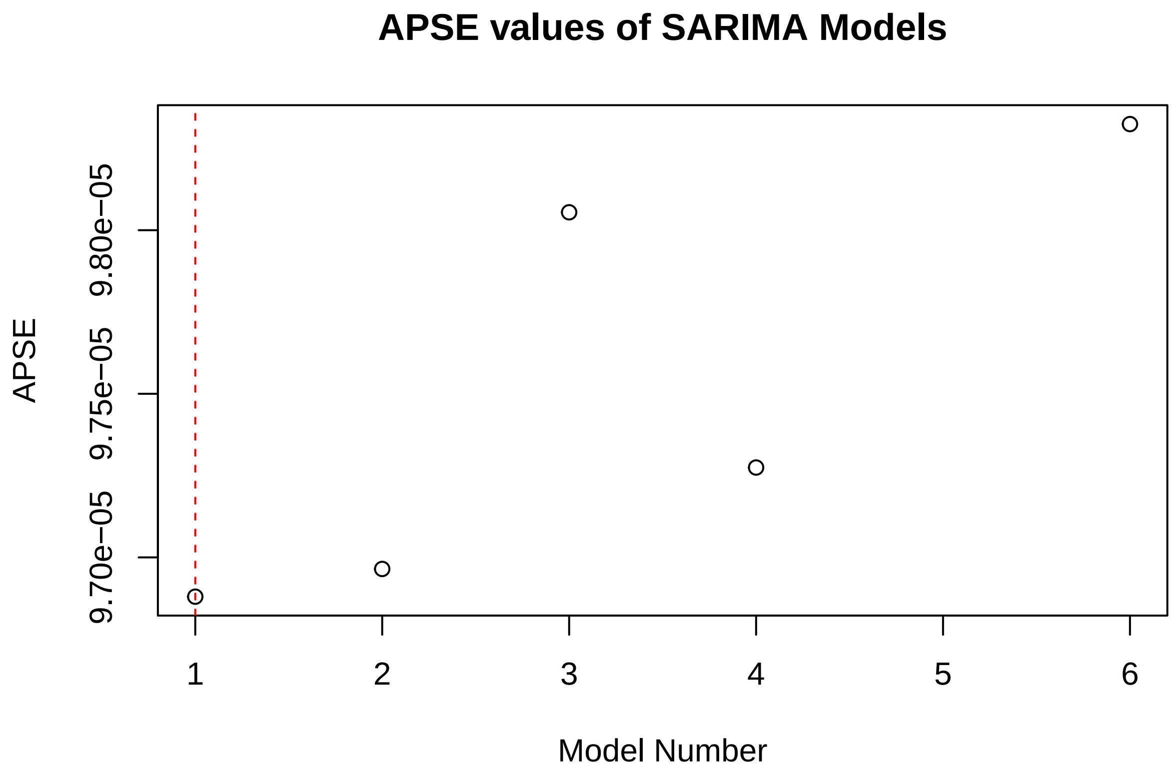

Fig 114: APSE values of Combination HW SARIMA models

Model 1 gives us the smallest value for APSE out of the proposed SARIMA models as can be seen in Figure 114, therefore for using SARIMA on differenced Holt-Winters residuals, I would recommend the $SARIMA(0,0,5) \times (0,1,1)_{12}$ model. However, the APSE for this SARIMA model does not outperform the SARIMA model used for combination methods using the residuals of regression, so we will not use it for forecasting purposes.