



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat d'Informàtica de Barcelona



*Data Analysis*

# Energy consumption in United States of America

Matija Jakovac

Barcelona

April 28, 2025

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Project objective . . . . .	2
1.2	Expected benefits . . . . .	2
<b>2</b>	<b>Basic description of the dataset</b>	<b>3</b>
2.1	Context . . . . .	3
2.2	Variables . . . . .	3
2.3	Sample size and coverage . . . . .	3
2.4	Data collection method . . . . .	3
<b>3</b>	<b>Summary of Methodology</b>	<b>4</b>
<b>4</b>	<b>Results and interpretation</b>	<b>5</b>
4.1	Identification . . . . .	5
4.2	Estimation . . . . .	11
4.2.1	Model 1 . . . . .	11
4.2.2	Model 2 . . . . .	11
4.3	Validation . . . . .	12
4.3.1	Model 1 . . . . .	12
4.3.1.1	Causality and invertibility . . . . .	12
4.3.1.2	Homoscedasticity . . . . .	12
4.3.1.3	Normality . . . . .	14
4.3.1.4	Independence . . . . .	16
4.3.1.5	Forecasting . . . . .	18
4.3.2	Model 2 . . . . .	20
4.3.2.1	Causality and invertibility . . . . .	20
4.3.2.2	Homoscedasticity . . . . .	20
4.3.2.3	Normality . . . . .	22
4.3.2.4	Independence . . . . .	24
4.3.2.5	Forecasting . . . . .	26
4.3.3	Comparison . . . . .	28
4.4	Forecasting . . . . .	29
4.5	Outlier Treatment . . . . .	29
4.5.1	Automatic outlier detection . . . . .	29
4.5.2	Re-estimation . . . . .	32
4.5.3	Re-validation . . . . .	32
4.5.4	Re-forecasting . . . . .	38
4.6	Comparison of three models . . . . .	39
<b>5</b>	<b>Conclusions and Discussion</b>	<b>41</b>
5.1	Strengths and weaknesses at a glance . . . . .	41
5.2	Final conclusion . . . . .	41

# 1 Introduction

Accurately measuring the trajectory of energy demand is critical for **government policies, investors, and climate activities**.

The **EnerUSA** series provides a long run monthly record of **primary energy consumption in the United States (billions of British Thermal Units - BTU)** which is an indicator that combines fossil, nuclear, and renewable sources.

A validated forecast of this series can

- support government in setting targets and policies;
- determine utilities in mix fuel decisions;
- inform private sector investment in infrastructure
- serve as an input to carbon emissions scenarios.

## 1.1 Project objective

Corresponding to the guidelines, objective is to **construct and validate an ARIMA-based forecasting with automatic outlier detection for the EnerUSA series**. Specifically, following objectives will be done:

1. **Identify** the transformations required for stationarity, justified by plots and formal tests.
2. **Estimate** two ARIMA specifications and compare them.
3. **Validate** model assumptions via full residual analysis (independence, normality, homoscedasticity), check stability, and determine accuracy by reserving the final 12 observations for validation.
4. **Forecast** U.S. primary energy consumption 12 months ahead, providing point forecasts and confidence bands.
5. **Detect and interpret outliers**, re-estimate the chosen model on the adjusted series with outlier treatment, and compare the new forecasts with the originals.

## 1.2 Expected benefits

- **Investment timing** – investors gain a great knowledge for making financing decisions.
- **Policy evaluation** – government can track the efficiency standards or incentives by comparing consumption against forecast.
- **Climate modelling** – enabling more credible long-term climate-impact decisions.

By combining the Box–Jenkins methodology with outlier-adjustment techniques, goal is to produce **high-confidence forecasts** that remain robust even in the presence of unusual activity.

## 2 Basic description of the dataset

### 2.1 Context

The dataset represents fossil-fuel, nuclear, and renewable inputs into a single line that people track as the measure of U.S. energy demand.

### 2.2 Variables

Variable	Description	Unit	Notes
date	Calendar month	—	January 1990 – December 2019
energy	Total primary energy consumption	<b>Bilions Btu</b>	Sum across all fuels and sectors

### 2.3 Sample size and coverage

- **Frequency:** Monthly
- **Period:** Jan 1990 – Dec 2019
- **Observations:** 348 (29 years  $\times$  12 months)

### 2.4 Data collection method

The **EnerUSA** series represents the *total primary energy consumed in the United States*, reported every month by the **U.S. Energy Information Administration (EIA)** in its *Monthly Energy Review*. Data is given through the EIA’s **Total Energy Data Browser**, which offers CSV, Excel, and API downloads.

### 3 Summary of Methodology

The project applies a **five-step Box–Jenkins methodology**—together with automatic outlier handling to the monthly *EnerUSA* series:

#### 1. Identification

- Transform (log) and difference (regular + seasonal) the series to secure stationarity, justify transformations with both numerical (comparing variance) and graphical (plotting series after differencing) results;
- Analyze lags of ACF/PACF graphs to identify two possible models, in other words getting all needed parameters (p,q,P,Q) for ARIMA model

#### 2. Estimation

- Fit both models with maximum-likelihood (`arima()`)
- Performing t-test for significance of model coefficients using t-test
- Compare AIC/BIC of the models

#### 3. Validation

- Validating if residuals are casual and invertible calculating modulus of characteristic polynomial roots
- Validating if residuals are homoscedastic plotting the residuals and square root of absolute residuals with smoothing
- Validating if residuals follow normal distribution plotting Q-Q plot and histogram with superimposed normal curve
- Validating if residuals are independent computing Shapiro-Walk, Anderson-Darling and Jarque Bera tests and plotting Ljung-Box test
- Assessing forecasting accuracy of the models by reserving last 12 observation as test dataset for model trained on the remaining dataset
- Computing model fit metrics, RMSE and MAE, RMSPE and MAPE and mean of prediction confidence interval
- Selecting the better model for forecasting depending on the performed validation and model metrics

#### 4. Forecasting

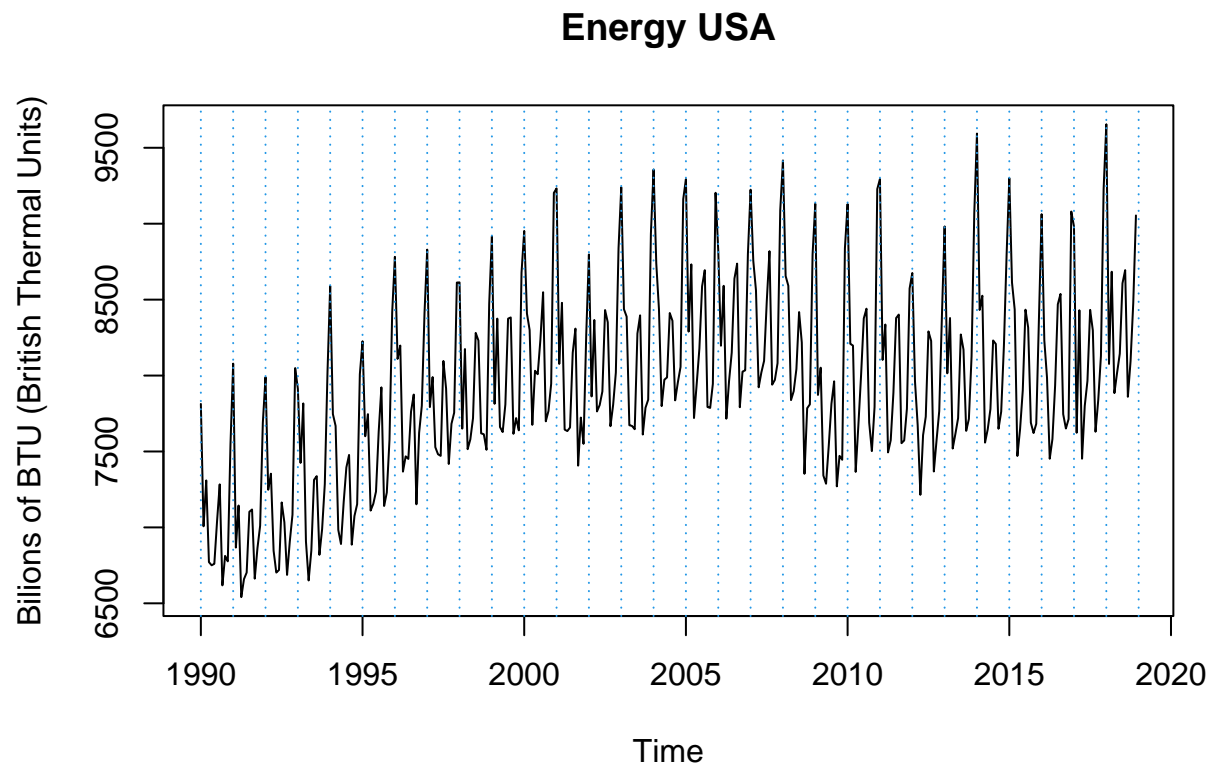
- Generate 12-month future forecasts and confidence intervals from the chosen model
- Visualizing them against the observed data

#### 5. Outlier Treatment

- Apply automatic outlier detection for selected model and interpret detected anomalies
- Adjust the series for outliers
- Perform re-estimation and validation steps again
- Compare the results with the original series
- Generate forecast for the original with adjusted model and compare the forecast to the previous ones

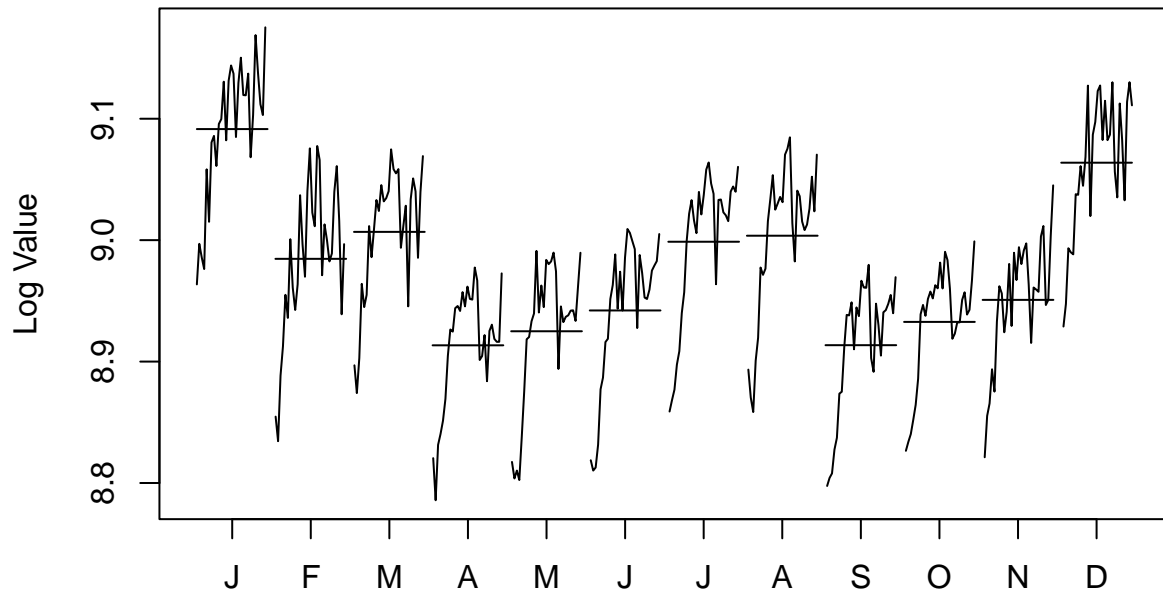
## 4 Results and interpretation

### 4.1 Identification



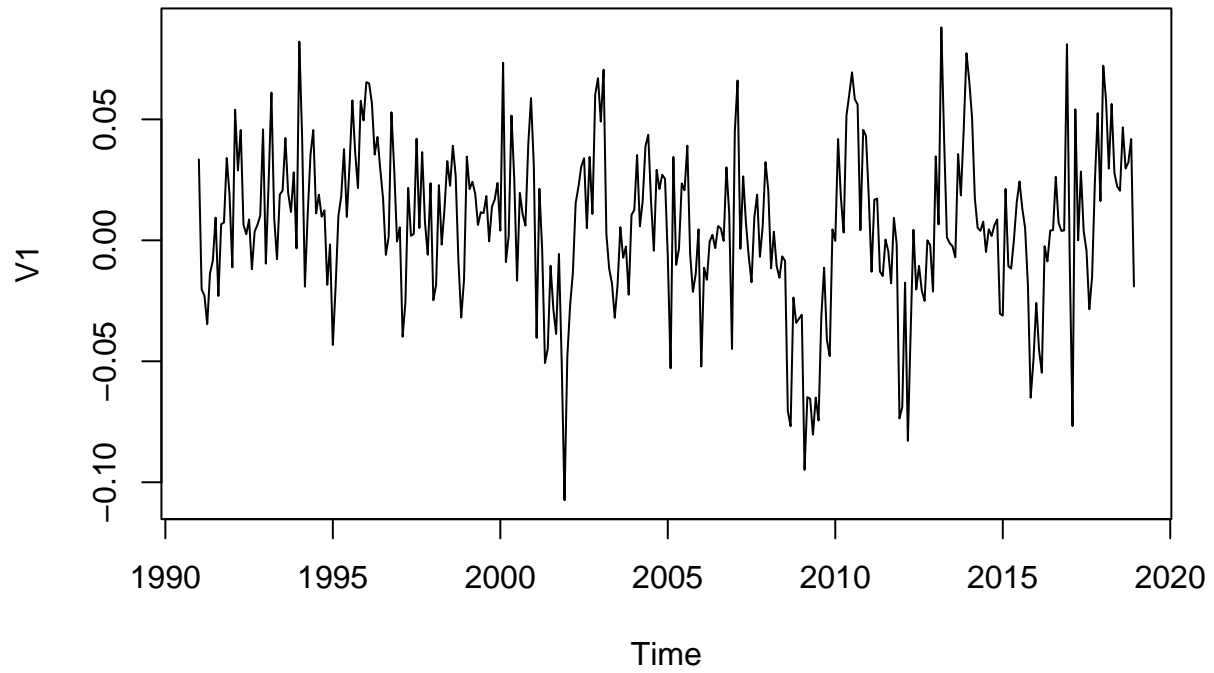
This plot represents original obtained time series. Logarithmic operation will be performed on the time series for easier observation.

## Monthplot of Energy USA



The month-plot of log values shows seasonal pattern with winter peaks and spring minimal. This pattern represents more energy spending in winter where energy is used for heating in buildings and for electricity during the night which is longer during winter. Minimal values are present in spring because it is the time heating is no longer needed and days are becoming longer. Summer comes with high temperature which means spending more energy on air conditioning. Autumn is similar to spring as it has lower temperatures. Overall, month-plot confirms stable and winter high-peaking seasonality in energy consumption.

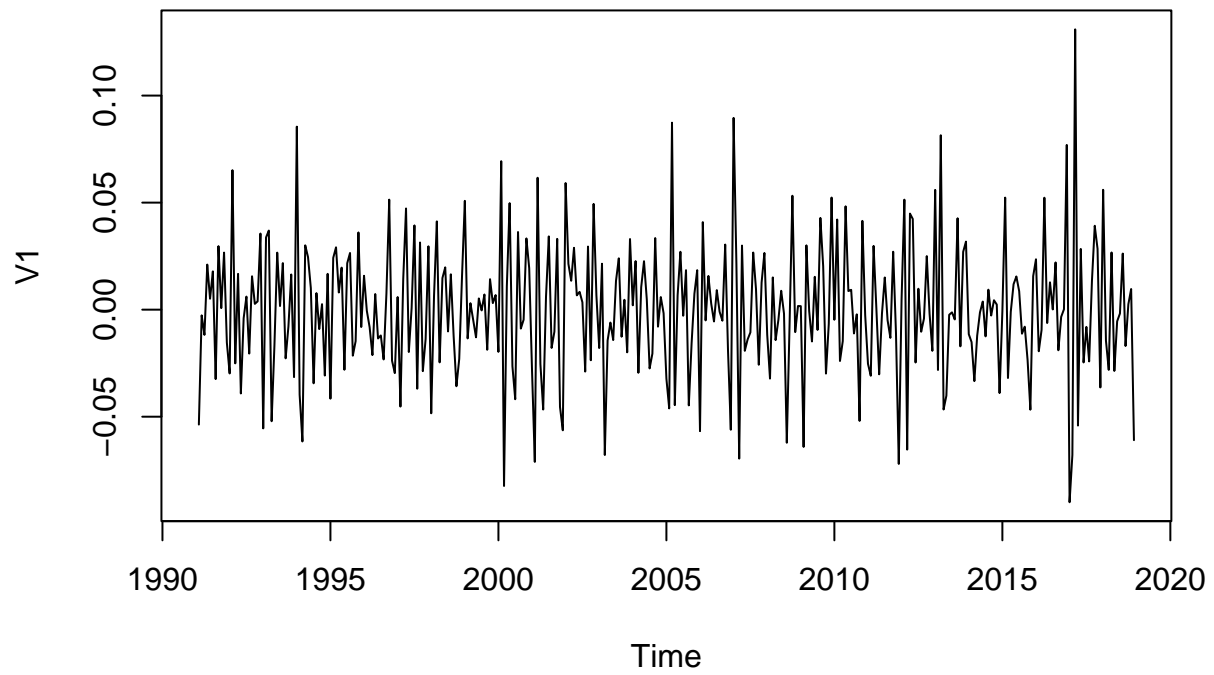
### Series after 1 seasonal differencing (d=0, D=1)



After performing seasonal difference with  $s=12$  as that represent year seasonal pattern, we can clearly see that the series is still not stationary having non-constant variance. We must perform regular difference.

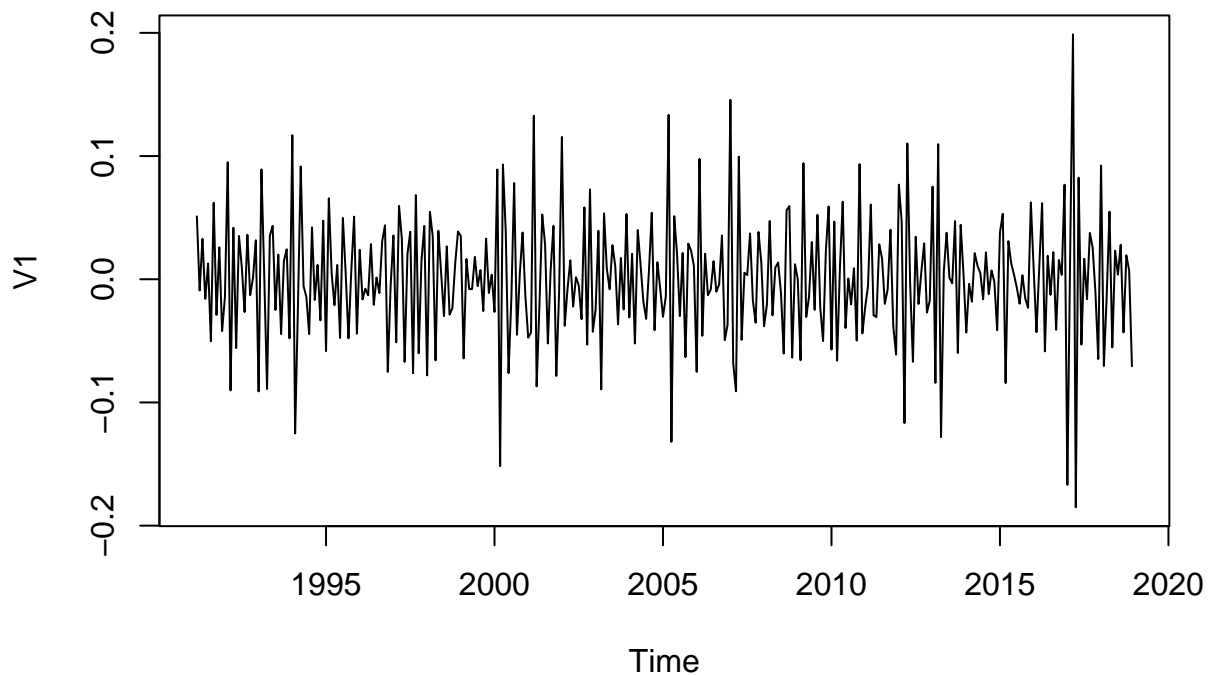


### Series after 1 regular and seasonal differencing (d=1, D=1)



After performing **one** regular and seasonal difference we can see that the series is **stationary** and we can use this transformed series having constant variance.

### Series after 2 regular and 1 seasonal differencing (d=2, D=1)



Maybe someone could say the series is still not stationary but after performing another regular difference we can assume that would be over-differencing.

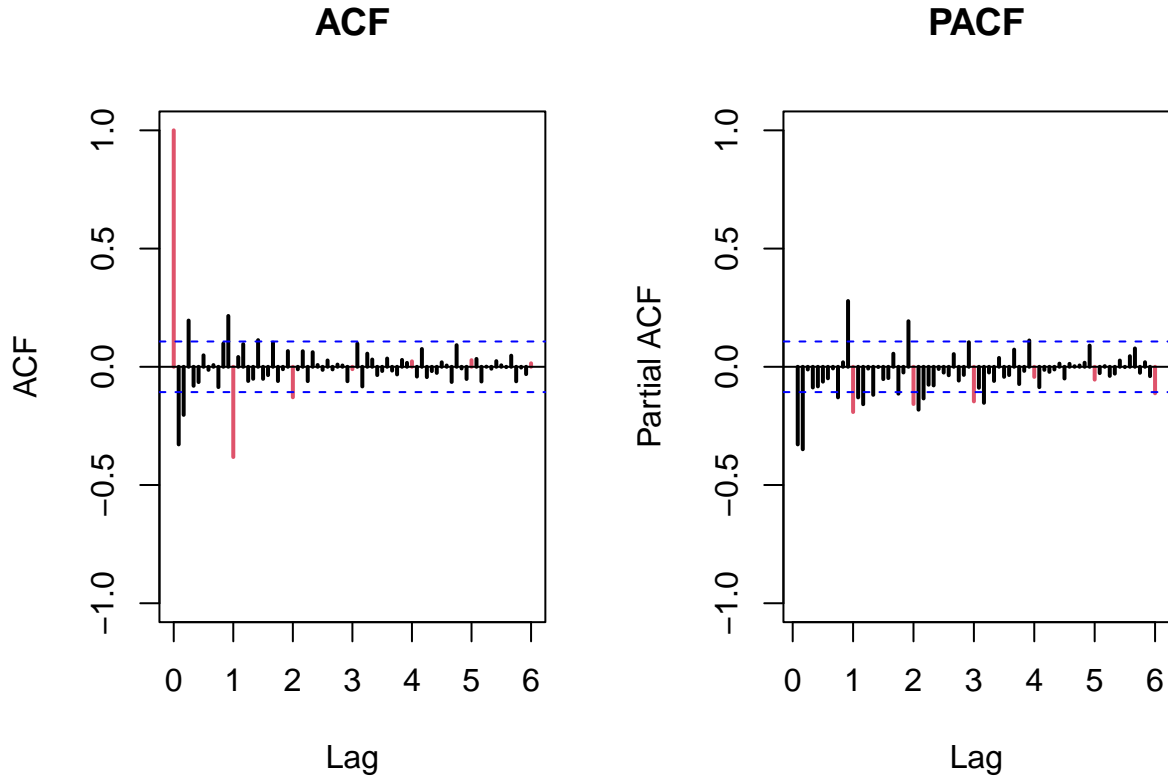
Variance of log-transformed original (d=0,D=0) series: 0.006160523

Variance of after one seasonal differencing (d=0,D=1) at s=12: 0.001068652

Variance of after one regular and seasonal differencing (d=1,D=1) differencing: 0.0009817458

Variance of after two regular and one seasonal differencing (d=2,D=1) differencing: 0.002596128

We prove that by computing variance of all series we perform differencing and look for the series having smallest variance. It is series with **one** regular and seasonal differencing so that first parameters of model are **d=1** and **D=1**.



**SEASONAL COMPONENT  $ARMA(P,Q)_{12}$ :** We will focus only on the red-colored lags and apply the identification criteria used for standard lags.

ACF: We can see that last red-colored lag that crosses blue line which represents confidence band is at lag=2 so I can propose  $MA(2)$ .

PACF: I can propose  $AR(3)$  as the third red line(lag=3) is over the confidence band (blue line) by a little bit. The sixth red line, last one, could be taken in consideration but as we usually want less parameters in models, I will not take it into consideration.

**REGULAR COMPONENT  $ARMA(p,q)$ :** We will focus only on the first 5-6 lags given the seasonality is 12. We must keep in mind that near the multiples of the seasonality (lags of order 12, 24, 36, etc.), there may be significant satellite lags that should not be considered for identification.

ACF: I can propose  $MA(3)$  because third black lag is over the confidence band and it is the last one do cross it.

PACF: I can propose  $AR(2)$  because the last lag that is over the band is second one.

With that I can propose models with  $d=1$ ,  $D=1$  and regular comp. of  **$AR(2)$** ,  **$MA(3)$**  and seasonal comp. of  **$AR(3)$** ,  **$MA(2)$**  of which I will take two models.

Model 1)  $ARIMA(2,1,0)(0,1,2)_{12}$  where I took  **$AR(2)$**  for regular comp. and  **$MA(2)$**  for seasonal comp.

Model 2)  $ARIMA(0,1,3)(3,1,0)_{12}$  where I took  **$MA(3)$**  for regular comp. and  **$AR(3)$**  for seasonal comp.

## 4.2 Estimation

Firstly, we specify the transformed stationary series ( $W_t$ ) to obtain mean estimation. After that, we check if the mean is significant. Last step is to check if coefficients are significant. Those checks are performed using t-test.

For t-test we can use following hypothesis:

$H_0: u_{wt} = 0$

$H_1: u_{wt} \neq 0$

$t = u/S_u$

$\text{abs}(t) > 2 \Rightarrow \text{reject } H_0, \text{abs}(t) \leq 2 \Rightarrow H_0$

### 4.2.1 Model 1

The model we are going to estimate is  $ARIMA(2, 1, 0)(0, 1, 2)_{12}$ .

Call:

```
arma(x = did12lnserie, order = c(2, 0, 0), seasonal = list(order = c(0, 0, 2), period = 12))
```

Coefficients:

	ar1	ar2	sma1	sma2	intercept
	-0.4391	-0.3111	-0.7464	-0.1365	0e+00
s.e.	0.0526	0.0527	0.0631	0.0641	1e-04

sigma<sup>2</sup> estimated as 0.0004674: log likelihood = 800.48, aic = -1588.95

By looking at intercept we can see that  $u=0.0$  while  $S_u=0.0001$ . In this case  $\text{abs}(t)=0$  which means we keep  $H_0$  therefore mean is not significant and we can re-estimate model with  $\log(X_t)$

Call:

```
arma(x = lnserie, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 2), period = 12))
```

Coefficients:

	ar1	ar2	sma1	sma2
	-0.4392	-0.3111	-0.7460	-0.1362
s.e.	0.0526	0.0527	0.0631	0.0641

sigma<sup>2</sup> estimated as 0.0004676: log likelihood = 800.43, aic = -1590.87

On this model we can also perform t-test for every coefficient to see if they are significant or not. We can see that every coefficient is significant as their s.e. is at least two times smaller than absolute value of coefficient.

### 4.2.2 Model 2

Model 2)  $ARIMA(0, 1, 3)(3, 1, 0)_{12}$

Call:

```
arma(x = did12lnserie, order = c(0, 0, 3), seasonal = list(order = c(3, 0, 0), period = 12))
```

Coefficients:

ma1	ma2	ma3	sar1	sar2	sar3	intercept
-----	-----	-----	------	------	------	-----------

	-0.4859	-0.2362	0.0692	-0.6448	-0.5695	-0.3396	0e+00
s.e.	0.0551	0.0675	0.0543	0.0537	0.0580	0.0553	2e-04

sigma^2 estimated as 0.000495: log likelihood = 794.12, aic = -1572.23

Same as in model 1 the mean here is not significant and we can continue with log(X\_t).

Call:

```
arima(x = lnserie, order = c(0, 1, 3), seasonal = list(order = c(3, 1, 0), period = 12))
```

Coefficients:

	ma1	ma2	ma3	sar1	sar2	sar3
	-0.4859	-0.2361	0.0692	-0.6448	-0.5695	-0.3396
s.e.	0.0551	0.0675	0.0543	0.0537	0.0580	0.0553

sigma^2 estimated as 0.000495: log likelihood = 794.11, aic = -1574.22

Performing t-test on every coefficient we can exclude coefficient ma3 as insignificant because absolute t-value of that coefficient is close to 1 then rejecting H\_0 in t-test.

Call:

```
arima(x = lnserie, order = c(0, 1, 2), seasonal = list(order = c(3, 1, 0), period = 12))
```

Coefficients:

	ma1	ma2	sar1	sar2	sar3
	-0.4937	-0.2037	-0.6535	-0.5755	-0.3347
s.e.	0.0585	0.0680	0.0531	0.0580	0.0553

sigma^2 estimated as 0.0004972: log likelihood = 793.32, aic = -1574.63

After removing that coefficient, we can see AIC metric got smaller which is good indication that model is fitting better to the data. Also, model has fewer parameters which improves computation speed. Removing it we obtain model  $ARIMA(0, 1, 2)(3, 1, 0)_{12}$

## 4.3 Validation

### 4.3.1 Model 1

The model we are going to validate is  $ARIMA(2, 1, 0)(0, 1, 2)_{12}$ .

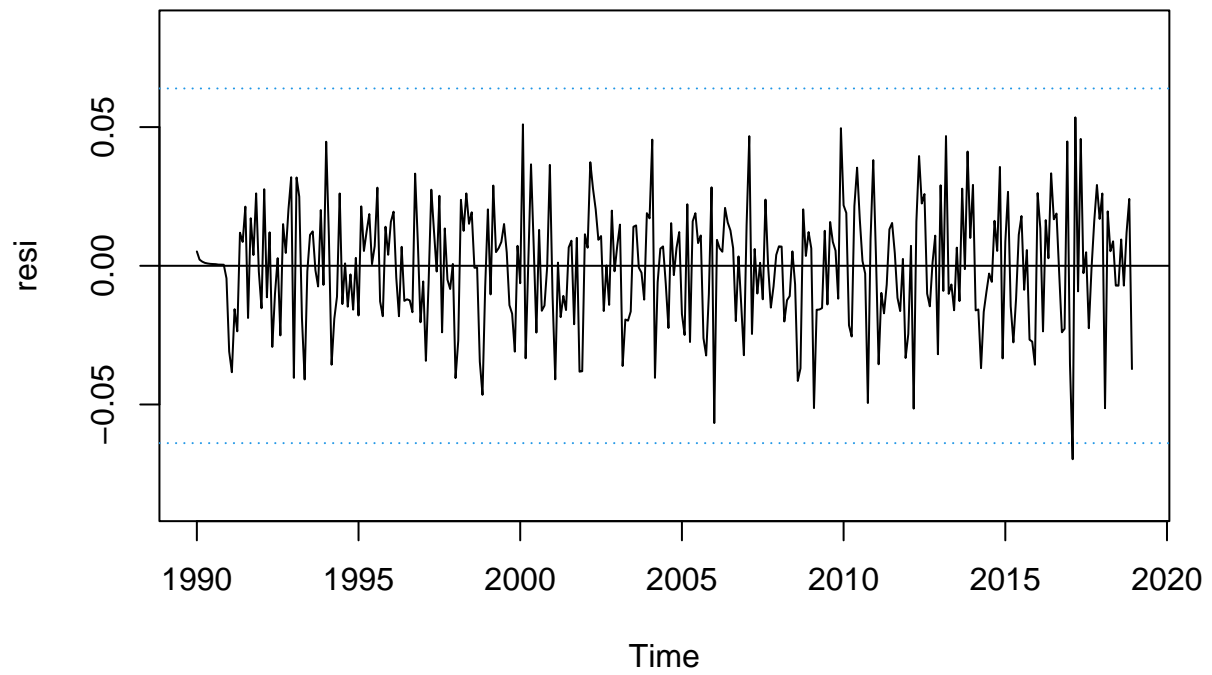
**4.3.1.1 Causality and invertibility** To check causality AR part of model is observed. To check invertibility MA part of model is observed. For both parts, the roots of characteristic equation must be computed and need to lie outside the unit circle (have modulus > 1).

Modulus of AR roots are 1.792904 1.792904

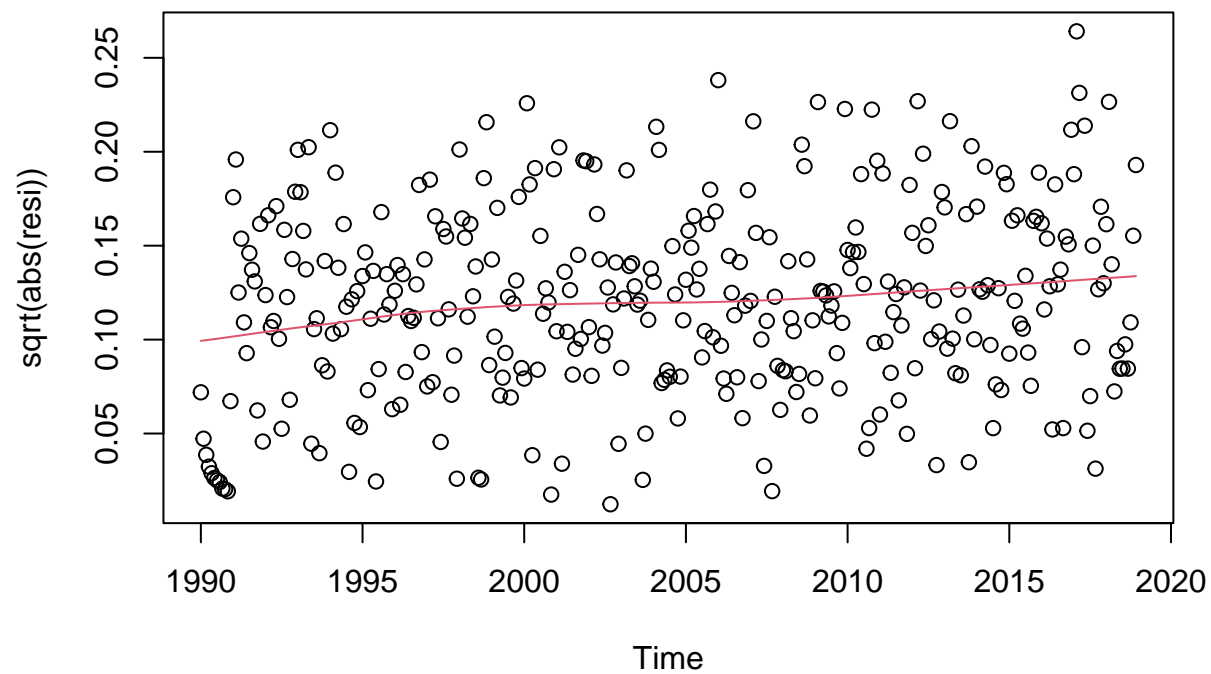
Unique modulus of MA roots are 1.009 1.1702

All modulus of AR and MA part have modulus above 1 meaning that this model is **casual** and **invertible**.

### 4.3.1.2 Homoscedasticity

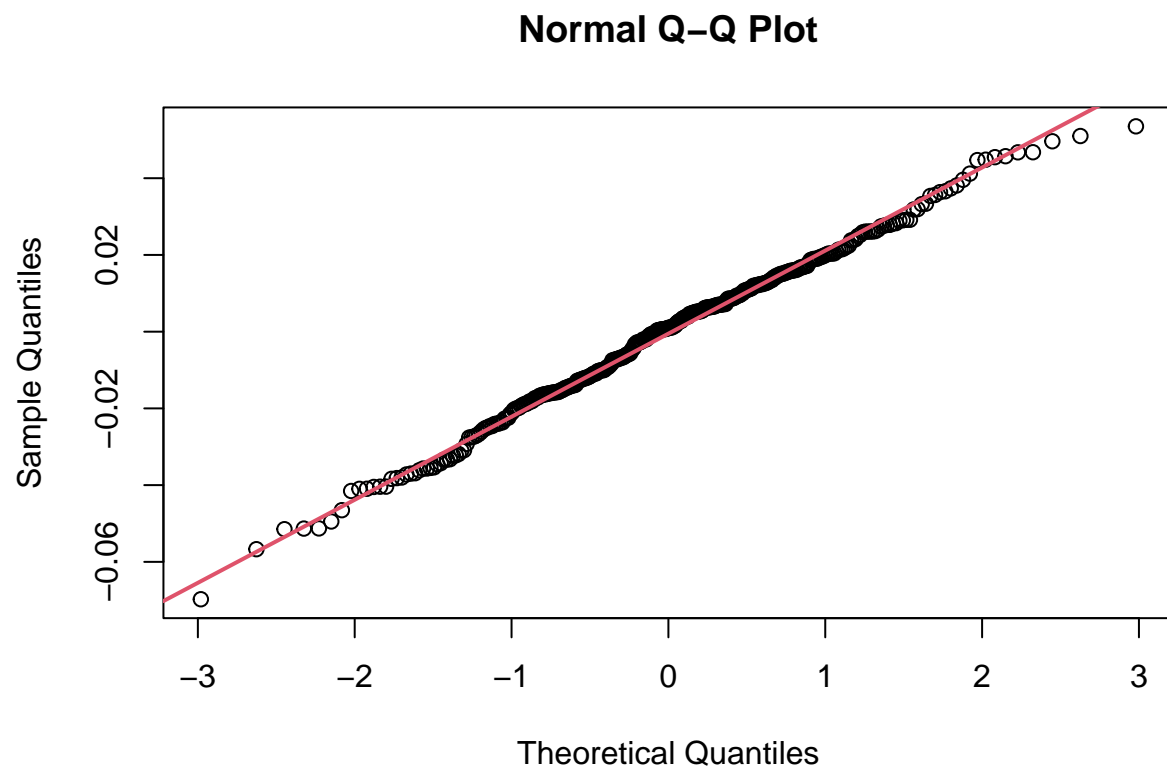


Observing residuals plot we can see only one is crossing the confidence band around 2017. However, having only one “crossing” we **can assume** that residuals are homoscedastic.



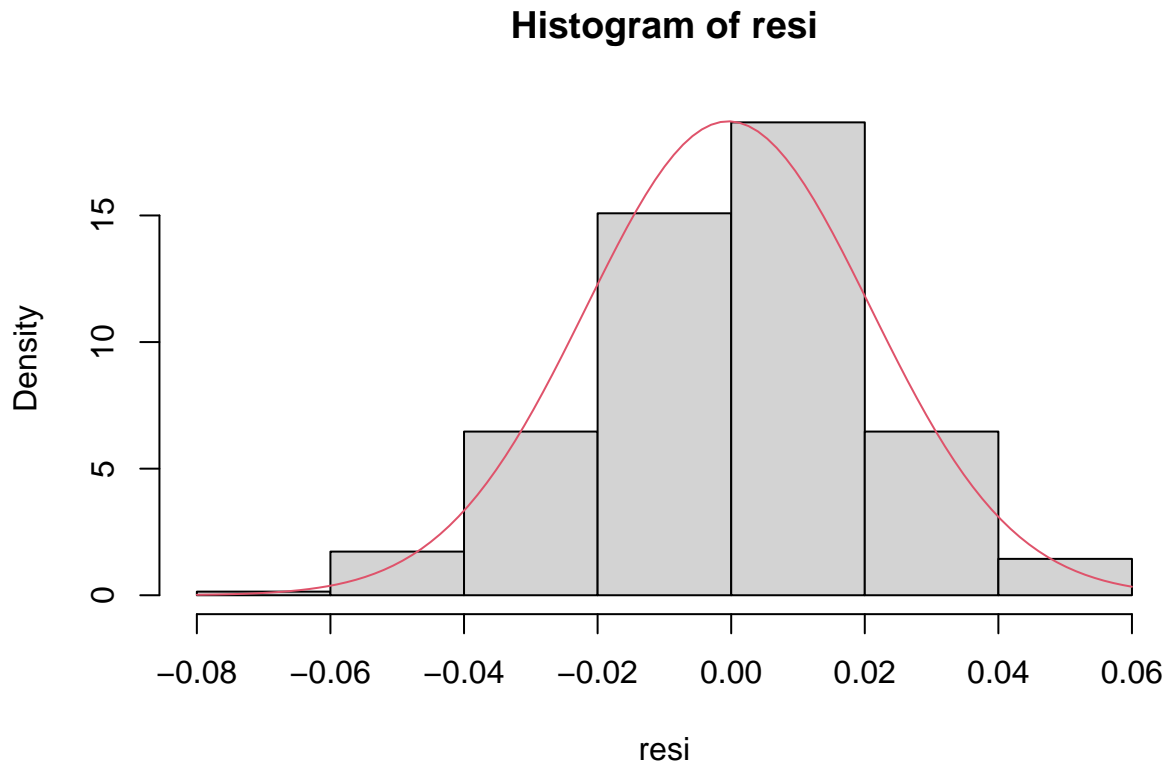
This square root of absolute residuals plot also proves that variance **is constant** as in theory the red line should be flat and we can only see slight increase during years.

#### 4.3.1.3 Normality



The residuals follow theoretical Q-Q line. We can say only one residual probably does not follow the line but vast majority of them do.





Histogram also shows that residuals follow superimposed normal curve but this is not the most reliable way to test normality as it depends on how many breaks does the histogram have.

#### Shapiro-Wilk normality test

```
data: resi
W = 0.99541, p-value = 0.3977
```

#### Anderson-Darling normality test

```
data: resi
A = 0.5056, p-value = 0.2009
```

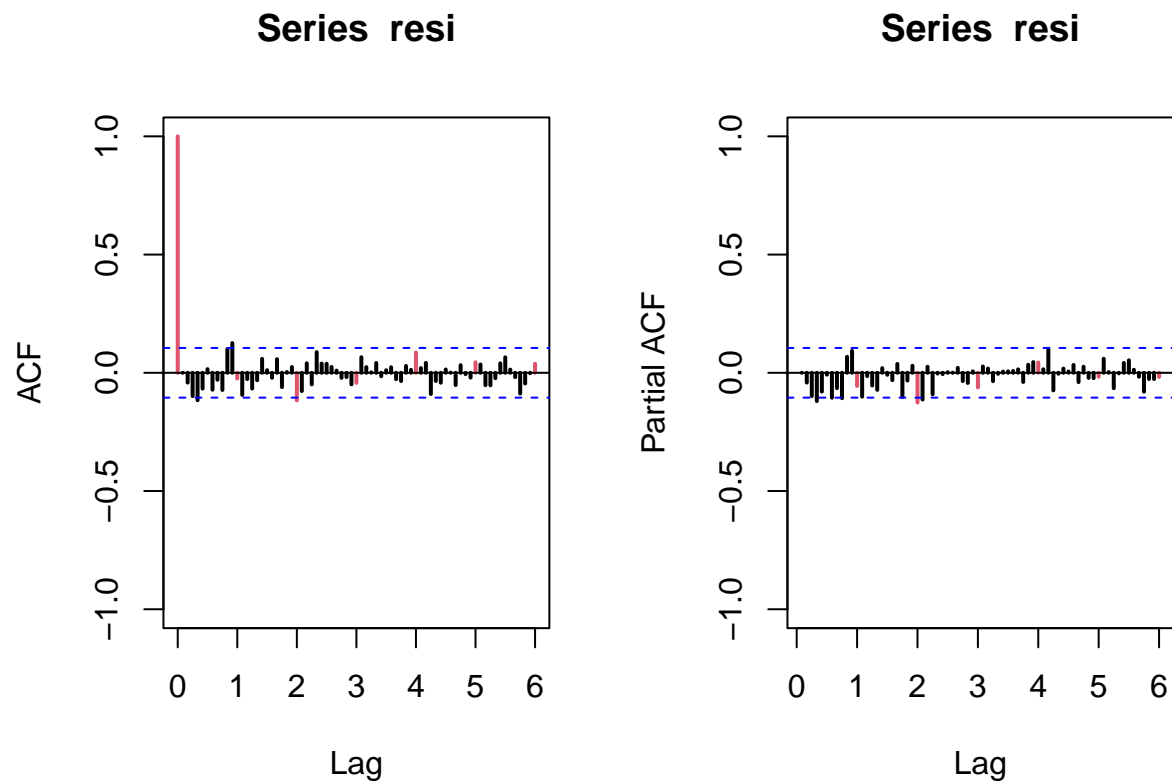
#### Jarque Bera Test

```
data: resi
X-squared = 1.4845, df = 2, p-value = 0.476
```

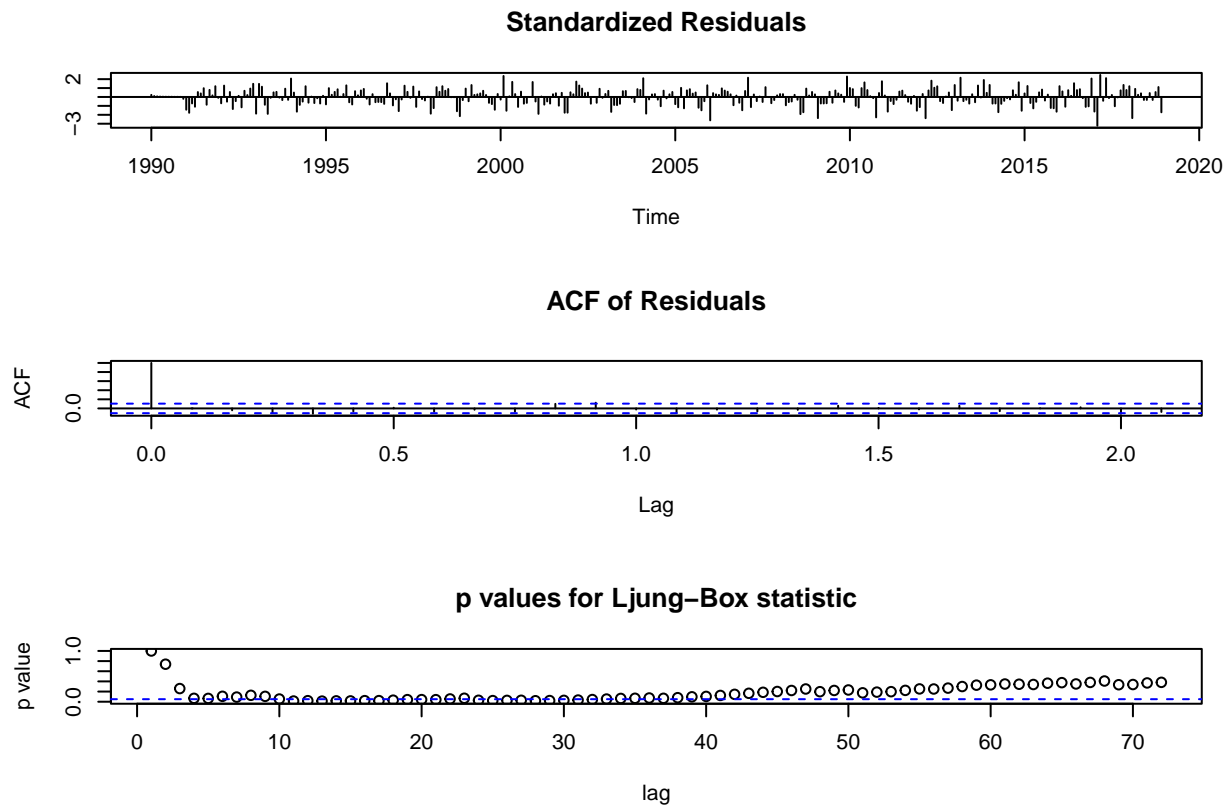
All the normality tests have p-value larger than 0.05 which means we keep  $H_0$  which in those tests states that the residuals follow normal distribution.

Combining all the ways to test normality, we can conclude that the residuals **follow normal distribution**.

#### 4.3.1.4 Independence



From the ACF/PACF of residuals we can see that some lags are crossing the confidence band representing autocorrelation between residuals and their lagged values.



More insight in independence of residuals we can get from Ljung-Box test. Observing p-values of lags we can see a lot of them have p-value below 0.05 then rejecting  $H_0$  which says lags are independent. With those results, we can assume that residuals **are not independent**.

Because of that, this model cannot be validated. Problem could probably be solved with removing outliers, adding more differencing or adding higher order AR or MA terms.

#### 4.3.1.5 Forecasting

AIC of model1 is -1590.866

BIC of model1 is -1571.795

Original model 1

Call:

```
arima(x = lnserie, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 2), period = 12))
```

Coefficients:

	ar1	ar2	sma1	sma2
	-0.4392	-0.3111	-0.7460	-0.1362
s.e.	0.0526	0.0527	0.0631	0.0641

sigma<sup>2</sup> estimated as 0.0004676: log likelihood = 800.43, aic = -1590.87

Model 1 without last 12 observations

Call:

```
arima(x = lnserie2, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 2),  
  period = 12))
```

Coefficients:

	ar1	ar2	sma1	sma2
	-0.4223	-0.3015	-0.7895	-0.0906
s.e.	0.0536	0.0537	0.0678	0.0694

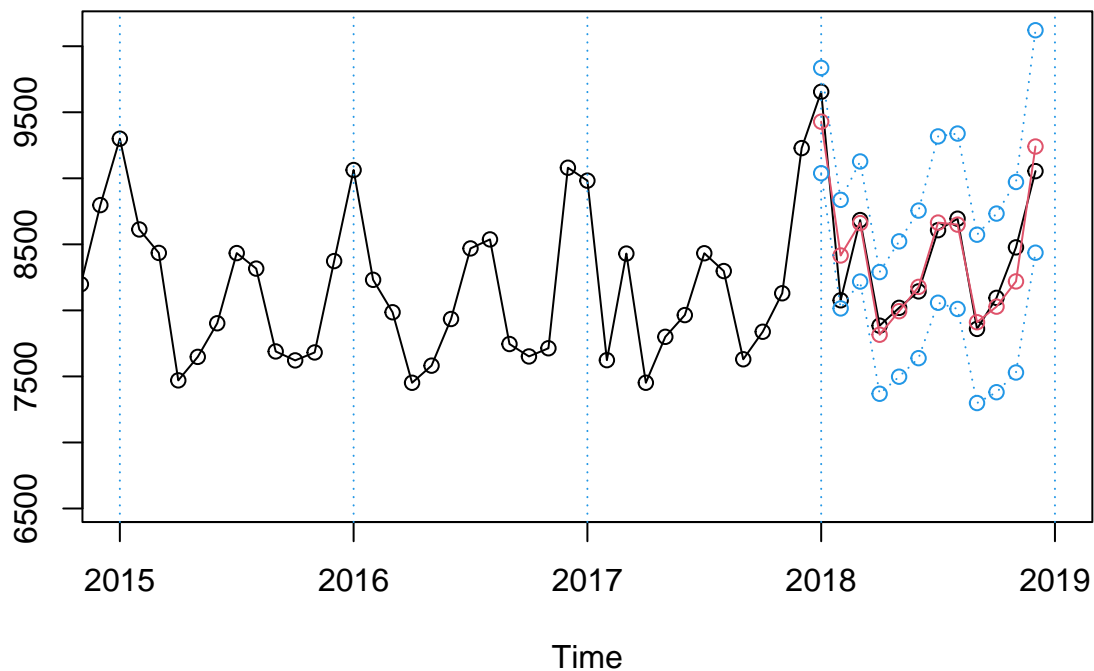
sigma<sup>2</sup> estimated as 0.000465: log likelihood = 772.31, aic = -1534.63

Here we can see difference between original model and the model without last 12 observations.

Modulus of AR roots of mod1b are 1.821218 1.821218

Unique modulus of MA roots of mod1b are 1.0097 1.2099

Observing modulus of AR and MA roots as well as sign of coefficients and absolute t-values for them we can see that the model without last 12 observations is still **stable**.



Red line represents predicted last 12 values based on the model without them. The original values are represented by black line. Two blue lines, upper limit and lower limit, represent 95% confidence interval. We can see that predicted values are similar to the original ones proving that the model trained without last 12 observations is generalized and can predict really accurate values.

The predictions values are

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2018	9428.821	8415.811	8661.133	7815.928	7993.845	8177.987	8664.763	8649.488
	Sep	Oct	Nov	Dec				
2018	7910.234	8028.411	8219.336	9240.690				

The original values are

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2018	9655.100	8076.522	8683.815	7884.303	8019.308	8143.949	8607.324	8694.376
	Sep	Oct	Nov	Dec				
2018	7859.887	8094.910	8477.193	9054.374				

RMSE of the model1 is 154.59

MAE of the model1 is 114.96

RMSPE of the model1 is 0.0181

MAPE of the model1 is 0.0135

Width of prediction confidence interval of model1 is 1161.06

Really low values of RMSPE and MAPE prove that model has a good fit and with this model we could predict further values as well.

### 4.3.2 Model 2

The model we are going to validate is  $ARIMA(0, 1, 2)(3, 1, 0)_{12}$ .

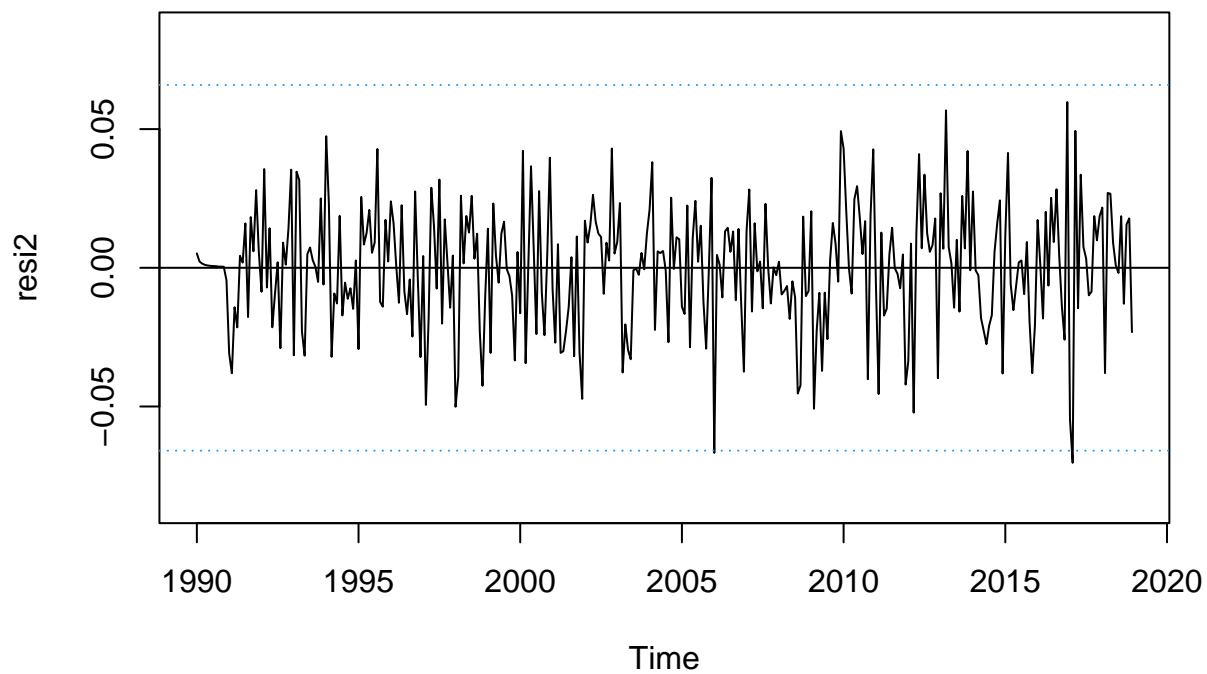
#### 4.3.2.1 Causality and invertibility

Unique modulus of AR roots are 1.0253 1.0421

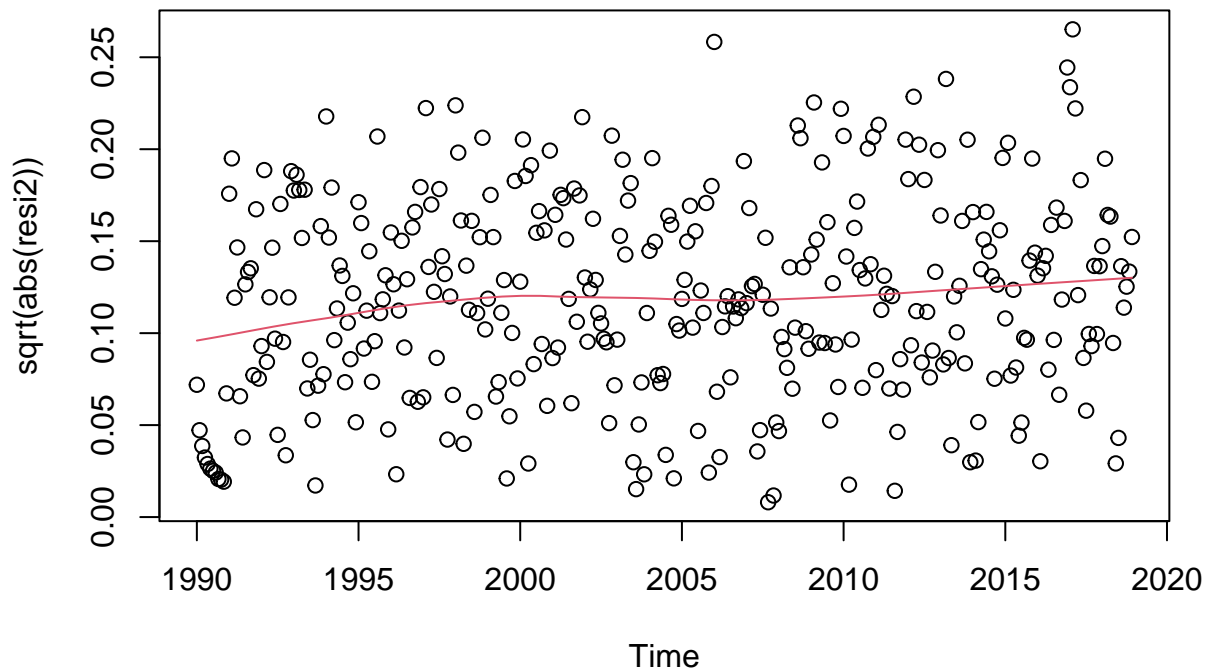
Modulus of MA roots are 1.313569 3.737627

As well as in first model, the roots of characteristic polynomials are greater than 1 showing that the model is **casual** and **invertible**.

#### 4.3.2.2 Homoscedasticity

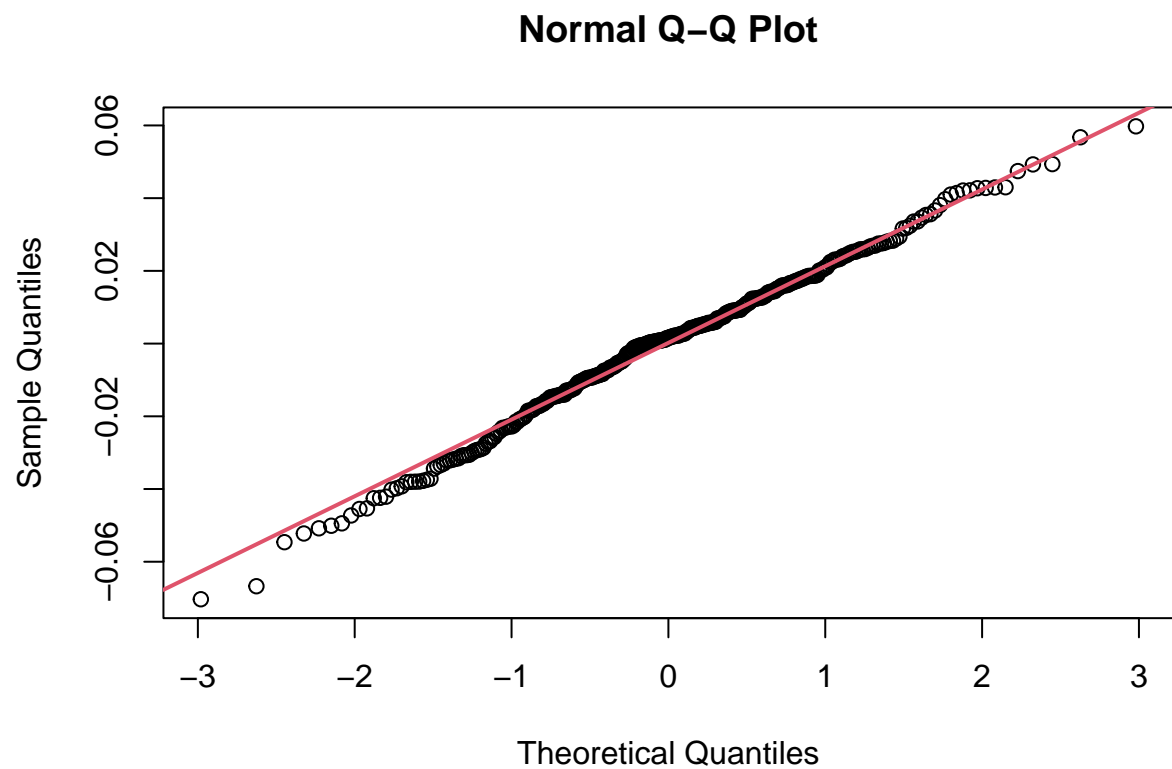


Observing the residuals plot variance seems to be constant but in this case the second model has one more residuals that is crossing the confidence band around 2006. Both models seems to have larger variance coming to the end of observation period.



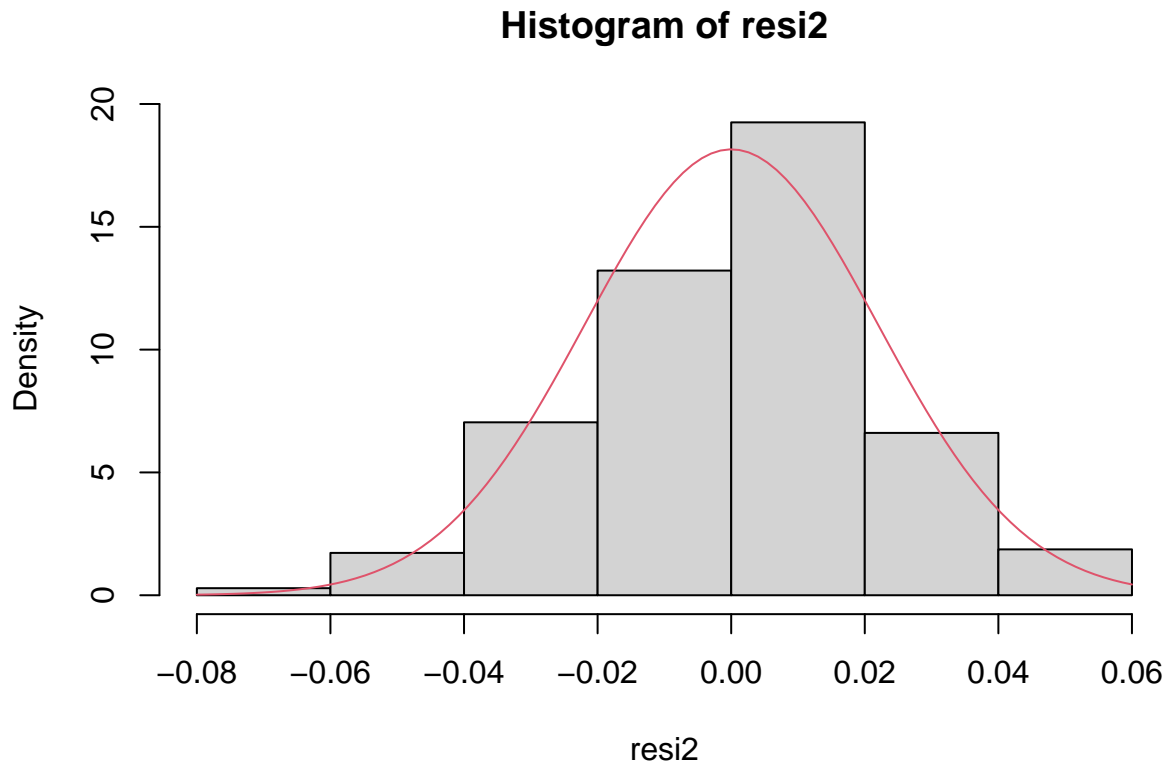
That is also proved for second model using the smoothing line having increase and larger values at the end. Even with that in mind, those increases are small and do not have large influence. Second model is also **homoscedastic** having constant variance.

#### 4.3.2.3 Normality



Residuals of second model follow the theoretical Q-Q line despite having larger number of “unfollowers”. We could assume that those residuals do not follow normal distribution as well as the residuals of first model.





Histogram of those residuals show normality in the residuals distribution but we can not rely on this method.

#### Shapiro-Wilk normality test

```
data: resi2
W = 0.99383, p-value = 0.1686
```

#### Anderson-Darling normality test

```
data: resi2
A = 0.80897, p-value = 0.03606
```

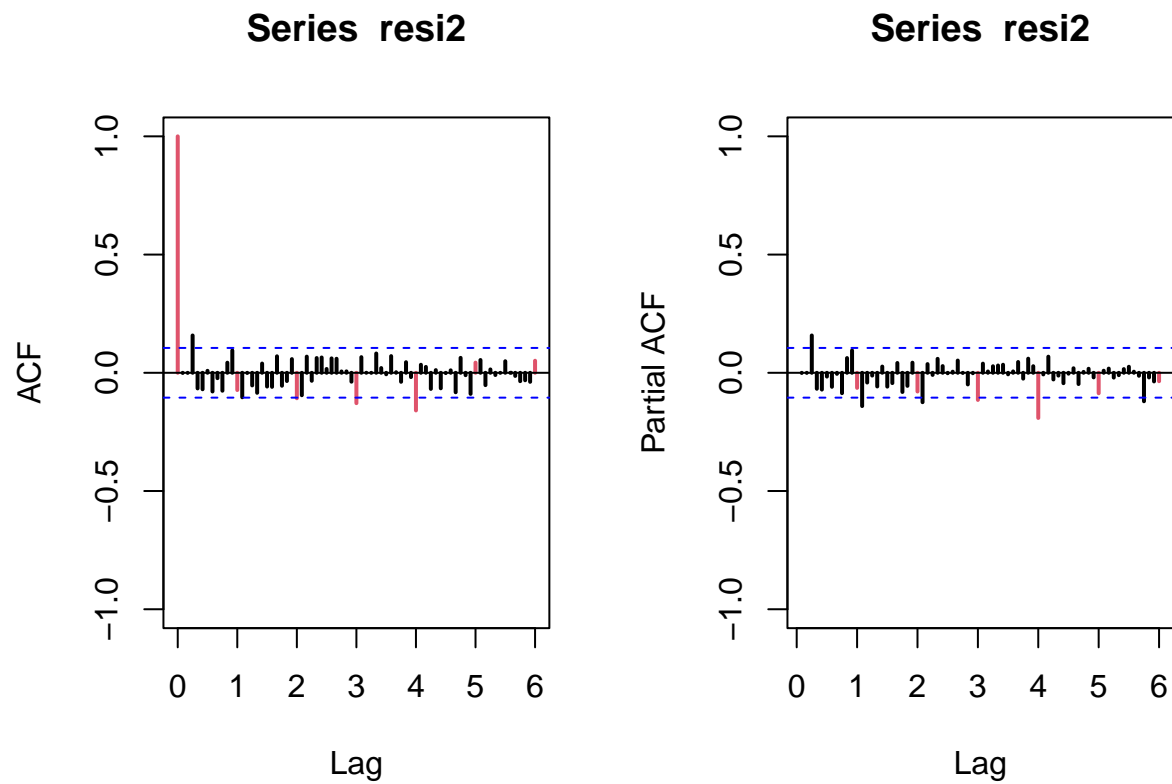
#### Jarque Bera Test

```
data: resi2
X-squared = 3.3146, df = 2, p-value = 0.1907
```

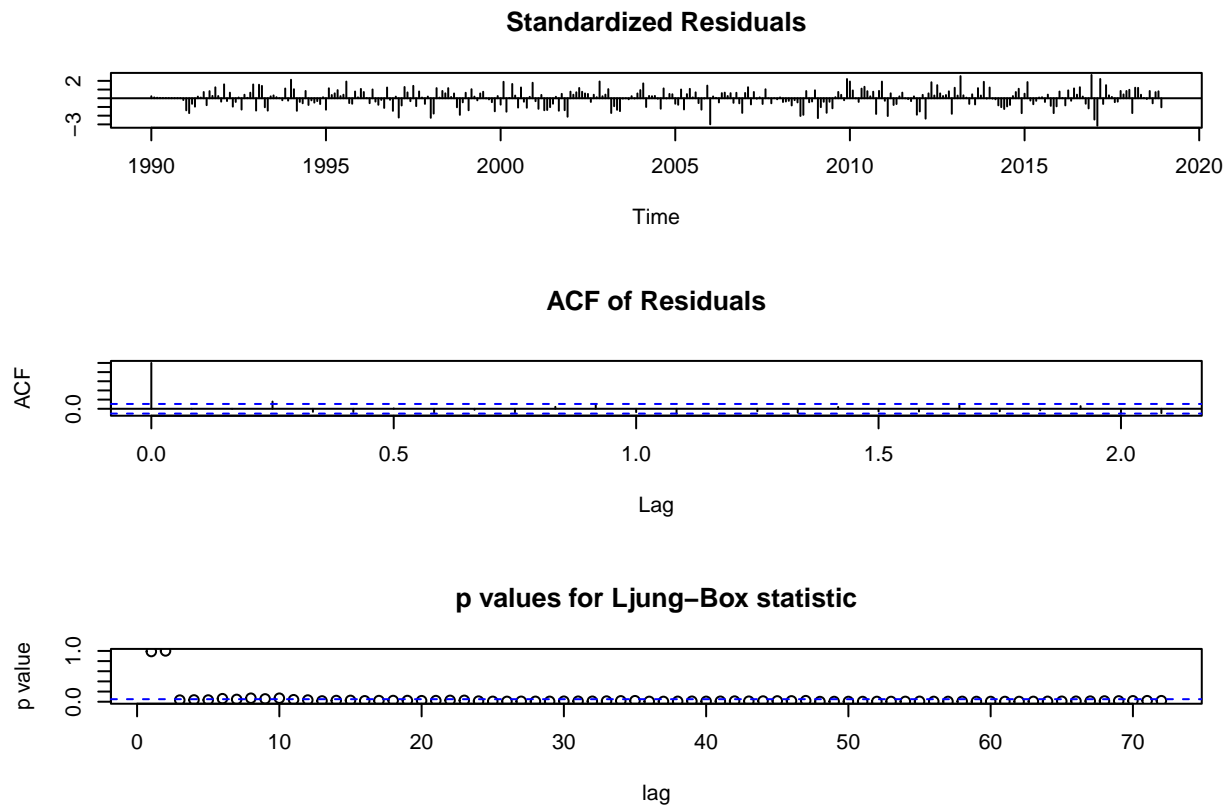
Residuals fail one of three test for normality having p-value of Anderson-Darling normality test less than 0.05 rejecting the normality hypothesis. Other two test show the opposite as they do not reject that hypothesis.

With all those results, we can say the residuals **do follow normal** distribution but not as well as the residuals of first model.

#### 4.3.2.4 Independence



From the ACF/PACF of residuals we can see some lags that are crossing the confidence band showing there is some autocorrelation.



Further proof of that is p-value of Ljung-Box test that are showing for almost all lags p-value lower than 0.05. This results shows us the residuals **are not independent** and **can not** be validated.

We should further check the model's parameters or treat outliers and after that perform test for independence again.

#### 4.3.2.5 Forecasting

AIC of model2 is -1574.634

BIC of model2 is -1551.749

Original model 2

Call:

```
arima(x = lnserie, order = c(0, 1, 2), seasonal = list(order = c(3, 1, 0), period = 12))
```

Coefficients:

	ma1	ma2	sar1	sar2	sar3
	-0.4937	-0.2037	-0.6535	-0.5755	-0.3347
s.e.	0.0585	0.0680	0.0531	0.0580	0.0553

sigma<sup>2</sup> estimated as 0.0004972: log likelihood = 793.32, aic = -1574.63

Model 2 without last 12 observations

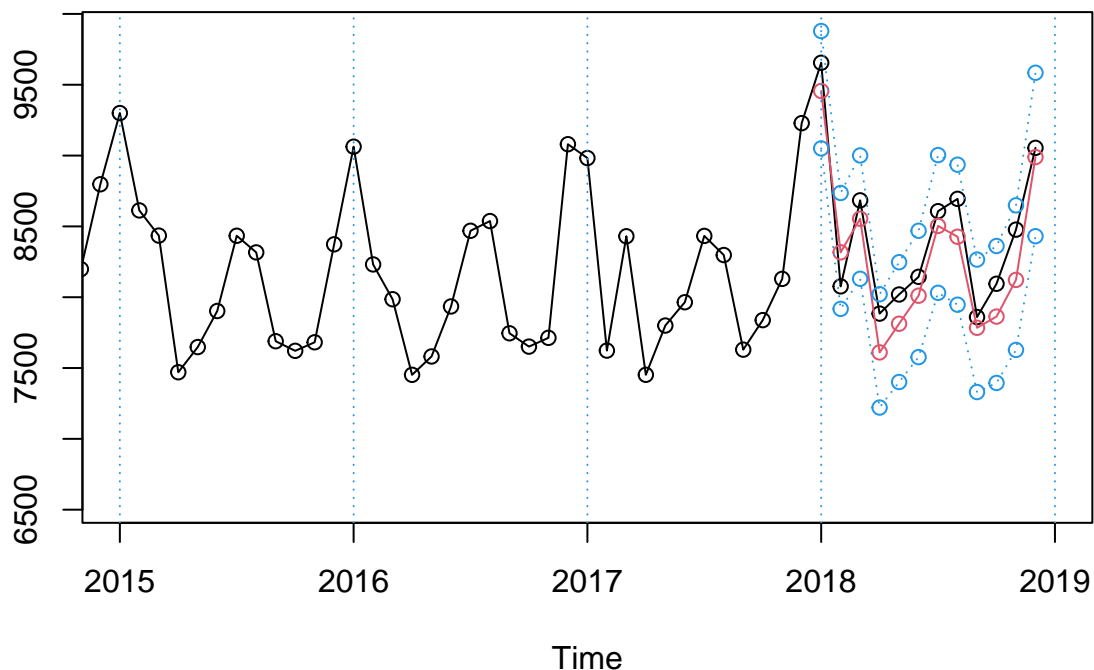
```
Call:
arima(x = lnserie2, order = c(0, 1, 2), seasonal = list(order = c(3, 1, 0),
  period = 12))
```

Coefficients:

	ma1	ma2	sar1	sar2	sar3
	-0.4901	-0.2126	-0.6749	-0.5826	-0.3255
s.e.	0.0588	0.0686	0.0559	0.0604	0.0565

sigma<sup>2</sup> estimated as 0.0004995: log likelihood = 763.92, aic = -1515.84

For the forecasting of last 12 observations we will use model trained on all observation except the last 12. The parameters and absolute t-values of this model are very similar to the original model showing that the model is **stable**.



The predictions of trained model represented by the red line are very similar to the original observations represented by the black line as we can see that on the graph.

The predictions values are

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2018	9455.869	8316.686	8554.255	7610.431	7812.778	8010.666	8503.360	8426.965
	Sep	Oct	Nov	Dec				
2018	7784.088	7863.283	8122.224	8989.255				

The original values are

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
-----	-----	-----	-----	-----	-----	-----	-----

2018	9655.100	8076.522	8683.815	7884.303	8019.308	8143.949	8607.324	8694.376
	Sep	Oct	Nov	Dec				
2018	7859.887	8094.910	8477.193	9054.374				

RMSE of the model2 is 208.47

MAE of the model2 is 190.13

RMSPE of the model2 is 0.025

MAPE of the model2 is 0.0227

Width of prediction confidence interval of model2 is 924.38

The model metrics of the second model are also acceptable for further forecasting but they are bit worse then the first model.

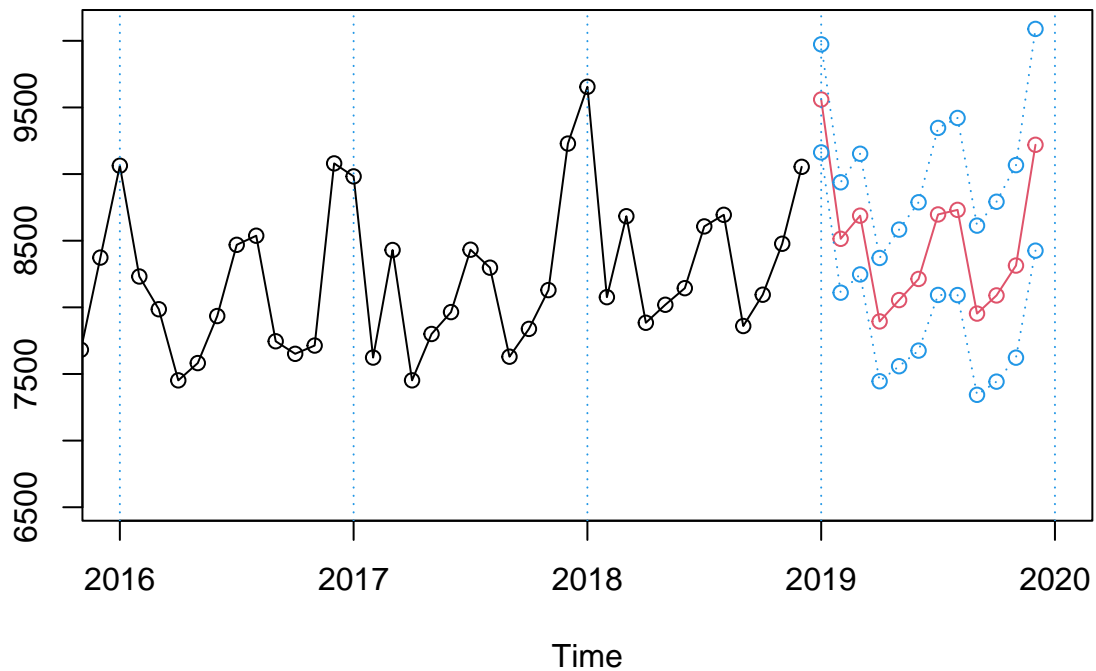
### 4.3.3 Comparison

	ARIMA(2,1,0)(0,1,2) <sub>12</sub>	ARIMA(0,1,2)(3,1,0) <sub>12</sub>
parameters	4.0000000000	5.0000000000
sigma2	0.0004676107	0.0004971541
AIC	-1590.8659819749	-1574.6335272536
BIC	-1571.7953293158	-1551.7487440626
RMSE	154.5936313018	208.4745614018
MAE	114.9559148796	190.1274771157
RMSPE	1.8147510653	2.4965134053
MAPE	1.3461352842	2.2692119134
CIml	1161.0585008747	924.3823564173

Residuals of both models are **casual, invertible, homoscedastic**. For the first model all normality test have p-value greater than 0.05, but for second model one test have value less than 0.05 showing that **first model** follows normal distribution better. Residuals of both models **are not independent**. First model has **less parameters, lower AIC, RMSPE and MAPE** but **larger** length of prediction interval. Both models perform good in forecasting but they is slight difference between them.

Having all of that in mind, it is better to choose **first model**,  $ARIMA(2, 1, 0)(0, 1, 2)_{12}$ , for further forecasting and outlier detection as it has slightly better performance in predictions and less parameters in original model.

## 4.4 Forecasting



Long-term forecast (red line) for year 2019 shows similar seasonal pattern as the for the previous years. Blue lines represent 95% confidence interval for those predictions.

	l13	pr3	ul3
Jan 2019	9162.843	9559.630	9973.599
Feb 2019	8110.337	8514.267	8938.314
Mar 2019	8246.872	8687.920	9152.556
Apr 2019	7445.302	7894.380	8370.545
May 2019	7558.239	8054.589	8583.534
Jun 2019	7675.518	8213.095	8788.324
Jul 2019	8092.981	8697.081	9346.274
Aug 2019	8092.218	8731.379	9421.022
Sep 2019	7343.724	7953.231	8613.325
Oct 2019	7442.584	8089.377	8792.379
Nov 2019	7622.343	8313.487	9067.299
Dec 2019	8425.417	9219.836	10089.159

Mean confidence interval for next year(2020) is 1159.829

It can be said the model 1 is **robust** enough for the predictions in the future.

## 4.5 Outlier Treatment

### 4.5.1 Automatic outlier detection

Obs	type_detected	W_coeff	ABS_L_Ratio	Fecha	PercVar
-----	---------------	---------	-------------	-------	---------

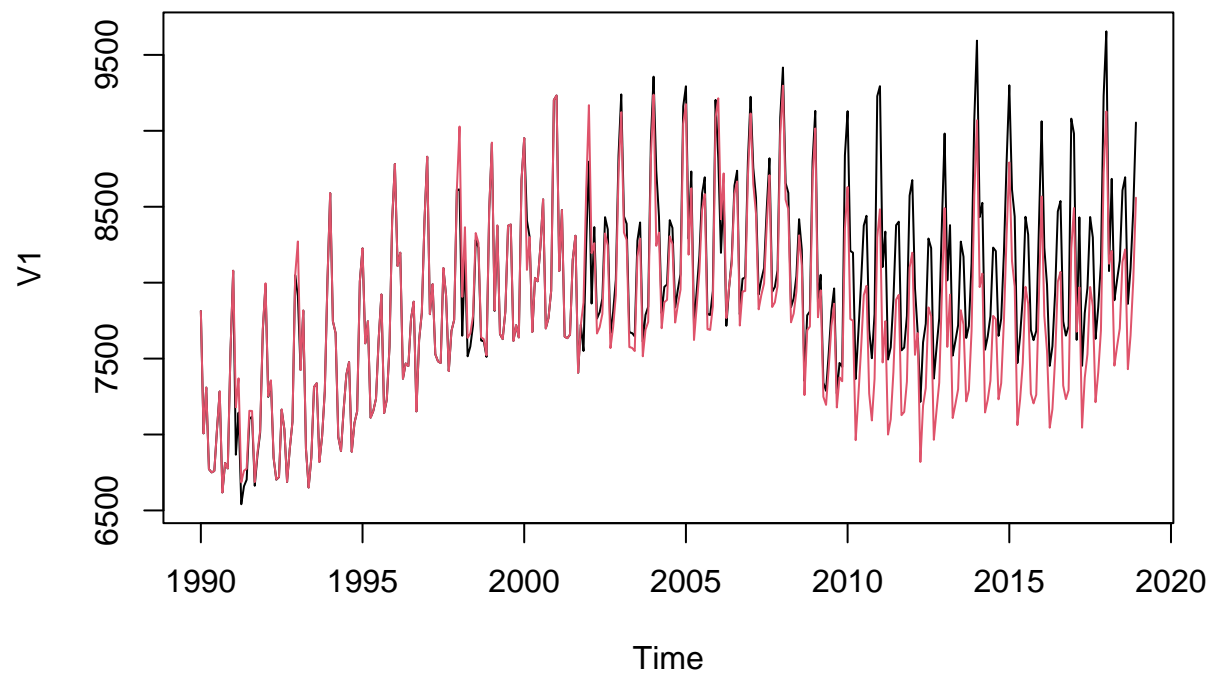
9	14	TC	-0.04451019	2.783713	Feb 1991	95.64659
8	37	AO	-0.04437019	2.942928	Jan 1993	95.65998
7	97	TC	-0.04705586	2.874609	Jan 1998	95.40341
13	122	AO	0.03939983	2.773680	Feb 2000	104.01863
14	143	LS	-0.04138535	2.799755	Nov 2001	95.94593
5	147	LS	0.05415069	3.289649	Mar 2002	105.56437
6	170	AO	0.04580807	2.963807	Feb 2004	104.68735
4	193	TC	-0.05639576	3.308117	Jan 2006	94.51650
10	240	LS	0.04344414	2.792411	Dec 2009	104.44017
11	252	TC	0.05004848	3.200825	Dec 2010	105.13221
3	267	AO	-0.05581504	3.444280	Mar 2012	94.57140
12	324	AO	0.04285197	2.947975	Dec 2016	104.37834
1	326	AO	-0.06323336	3.735123	Feb 2017	93.87244
2	338	AO	-0.06211550	3.622581	Feb 2018	93.97743

In this table we can see all **14 outliers** that are present in the dataset based on the first model. We must be taking in the context that criterion for selecting is setted at 2.7. If we set that criterion to lower value we would have more outliers. That criterion is usually set from 2.6 to 2.9.

There are **three** types of outlier, AO - additive outlier, LS - level shift and TC - temporary change. Our main focus will be on LS outliers as they changed the mean of the series to a new one. TC also changes the mean but after some time it goes back to the original one and AO is just slight outlier happening in one month and not longer.

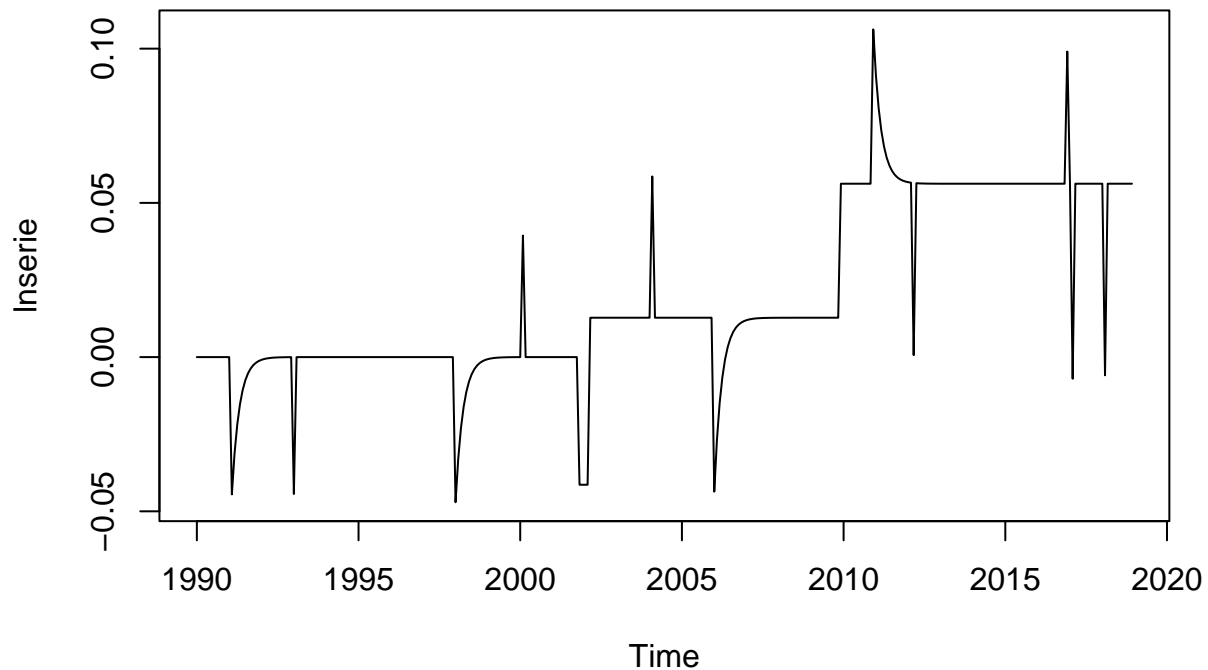
There are **three** level shifts in presented outliers at **Nov 2001, Mar 2002 and Dec 2009**. The first two happened because of the same incident which was **Twin Towers incident in 2001** which forced air travel ban and industrial cutbacks. By the next year around March U.S. economy was rebuilding and reconstructing.

The third and last one happened in the winter year after **great depression in 2008**. Because of the financial crisis the industrial activity was reduced but after initial reduction, it increased in the year after which happens to be last level shift.



Red line represent what would the series look like without computed outliers in comparison to original observations. We can see slight difference between two series as because of the level shifts outliers having impact on the observations.





This plot represents impact of outliers on the original series. Three LS are clearly seen, 2001 and 2009, as well other outliers which are not that impactful as level shifts. Those shift changed the mean of the series which can be seen on the plot by observing straight lines in time. Short spikes represent AO outliers and similar to those spikes but taking more time are the TC outliers.

#### 4.5.2 Re-estimation

Call:

```
arima(x = lnserie.lin, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 2),
  period = 12))
```

Coefficients:

	ar1	ar2	sma1	sma2
	-0.3981	-0.2872	-0.7051	-0.1543
s.e.	0.0530	0.0537	0.0579	0.0564

sigma<sup>2</sup> estimated as 0.0003164: log likelihood = 866.89, aic = -1723.78

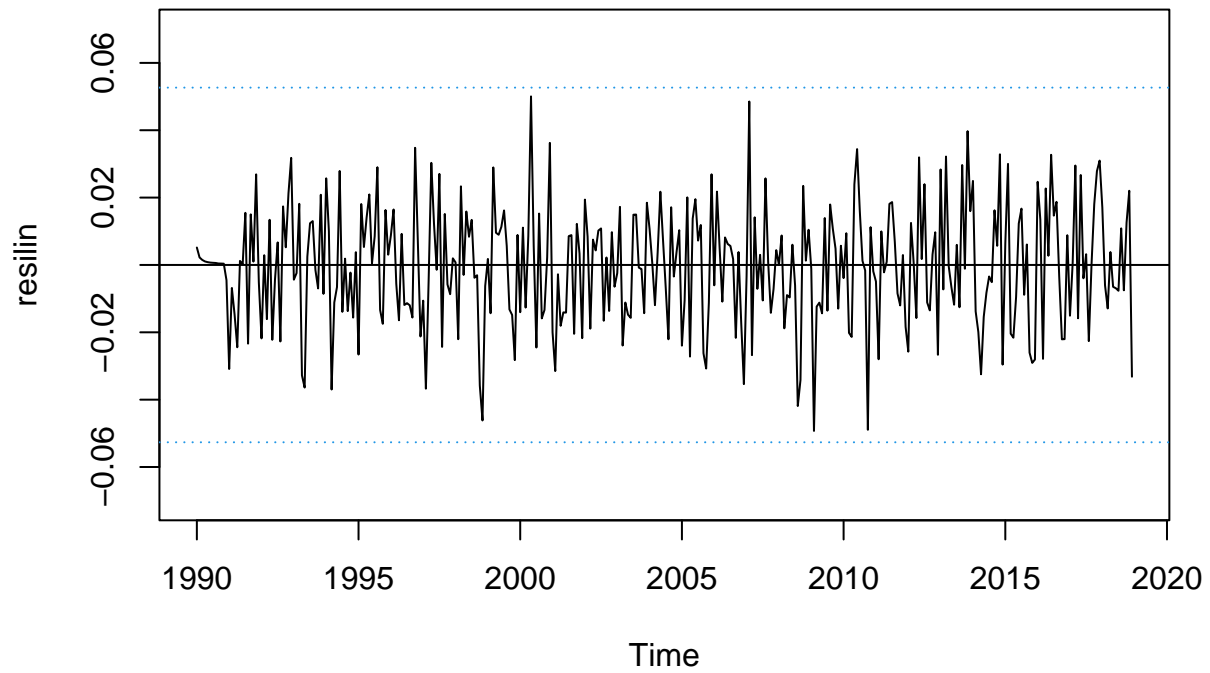
All the coefficients of the new model adjusted to the outliers are **significant**.

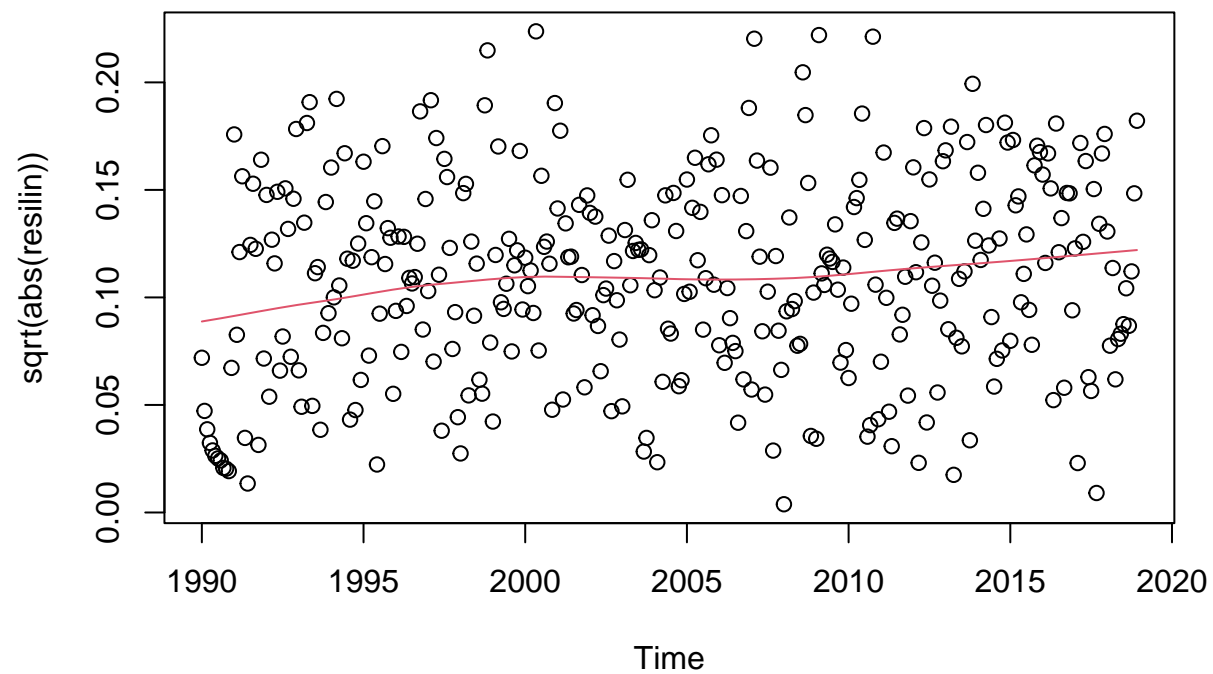
#### 4.5.3 Re-validation

Modulus of AR roots are 1.865869 1.865869

Unique modulus of MA roots are 1.1562 1.0107

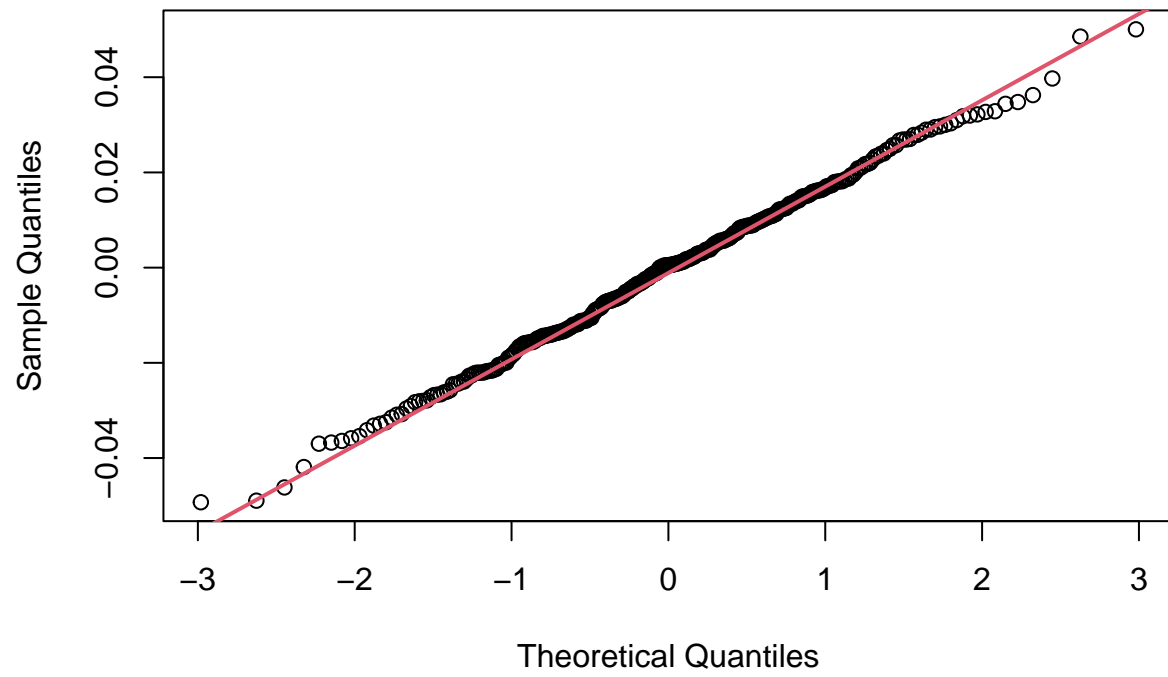
The roots of AR and MA part are greater than 1 making model **casual** and **invertible**.

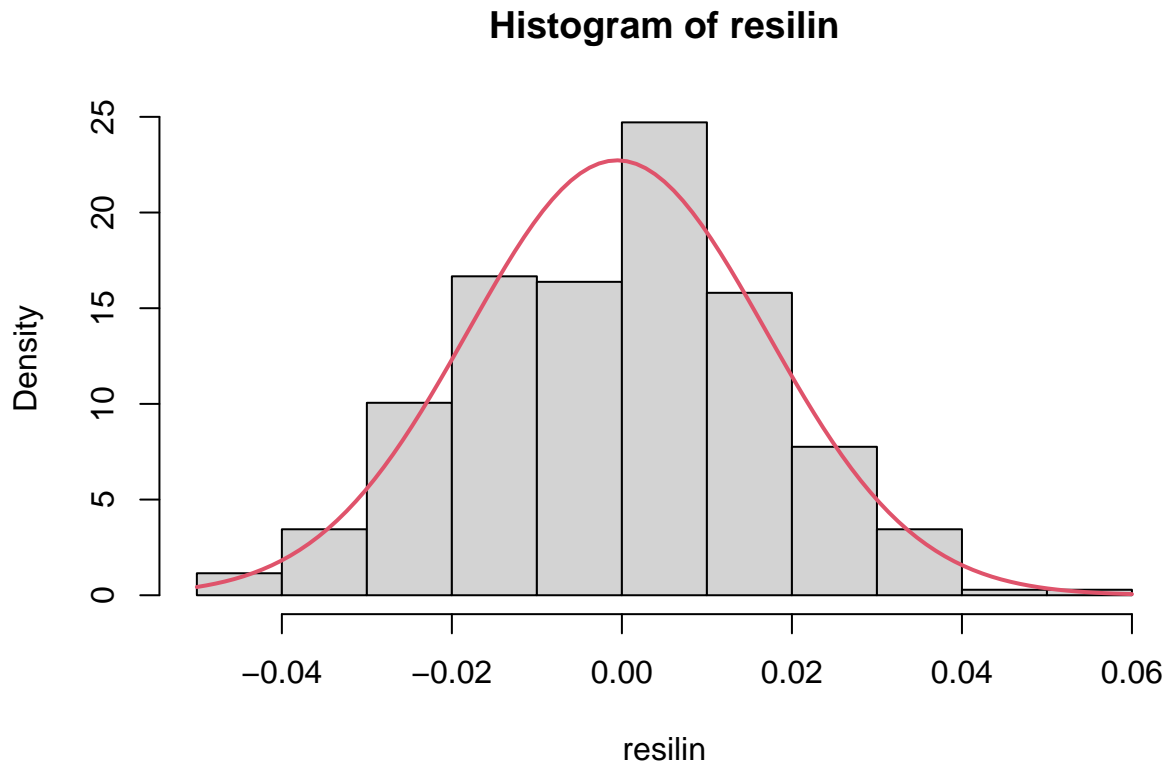




Adjusted model performs has **more constant variance** as not one residuals crosses confidence band but we can also see a slight increase in variance over time.

Normal Q-Q Plot





Shapiro-Wilk normality test

```
data:  resilin
W = 0.99757, p-value = 0.8933
```

Anderson-Darling normality test

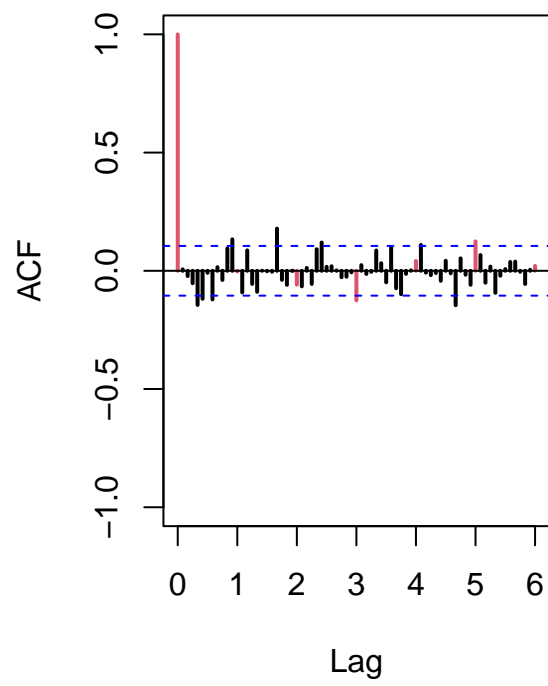
```
data:  resilin
A = 0.21688, p-value = 0.8424
```

Jarque Bera Test

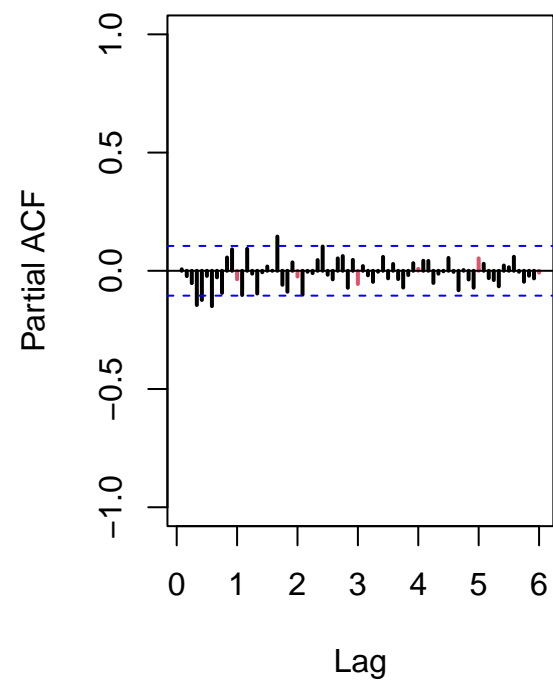
```
data:  resilin
X-squared = 0.5568, df = 2, p-value = 0.757
```

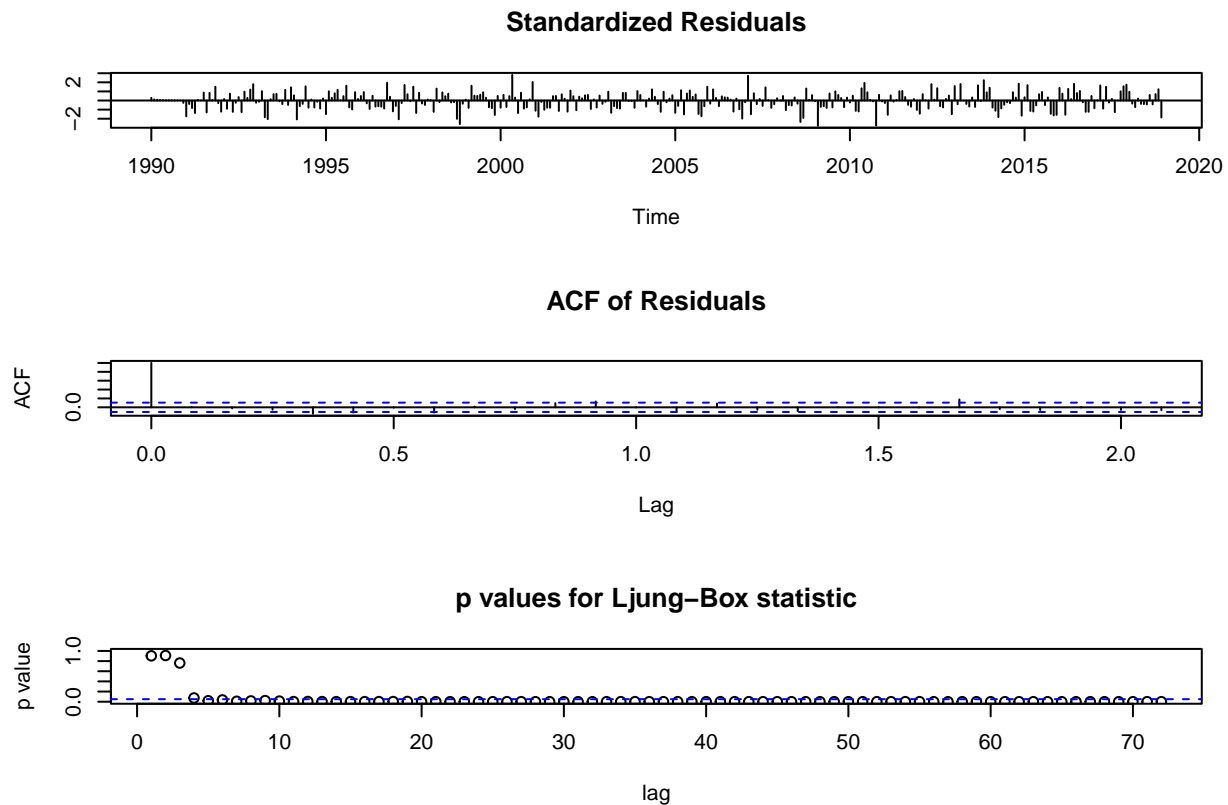
All the plots and test show the residuals **follow** normal distribution. Observing the Q-Q plot it can be said that the residuals of adjusted model follow the normal distribution better than the pre-adjusted models.

**Series resilin**



**Series resilin**





As well as models before even with adjusting the model with outliers residuals **are not independent** as the p-values of Ljung-Box are below 0.05 rejecting normality hypothesis. Same conclusion could be concluded from ACF/PACF plots as there are multiple lags crossing the confidence band.

AIC of adjusted model is -1695.776

BIC of adjusted model is -1622.775

#### 4.5.4 Re-forecasting

Original adjusted model 1

Call:

```
arima(x = lnserie.lin, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 2),
  period = 12))
```

Coefficients:

	ar1	ar2	sma1	sma2
	-0.3981	-0.2872	-0.7051	-0.1543
s.e.	0.0530	0.0537	0.0579	0.0564

sigma<sup>2</sup> estimated as 0.0003164: log likelihood = 866.89, aic = -1723.78

Adjusted model 1 without 12 last observations

Call:

```

arima(x = lnserie.lin2, order = c(2, 1, 0), seasonal = list(order = c(0, 1,
2), period = 12))

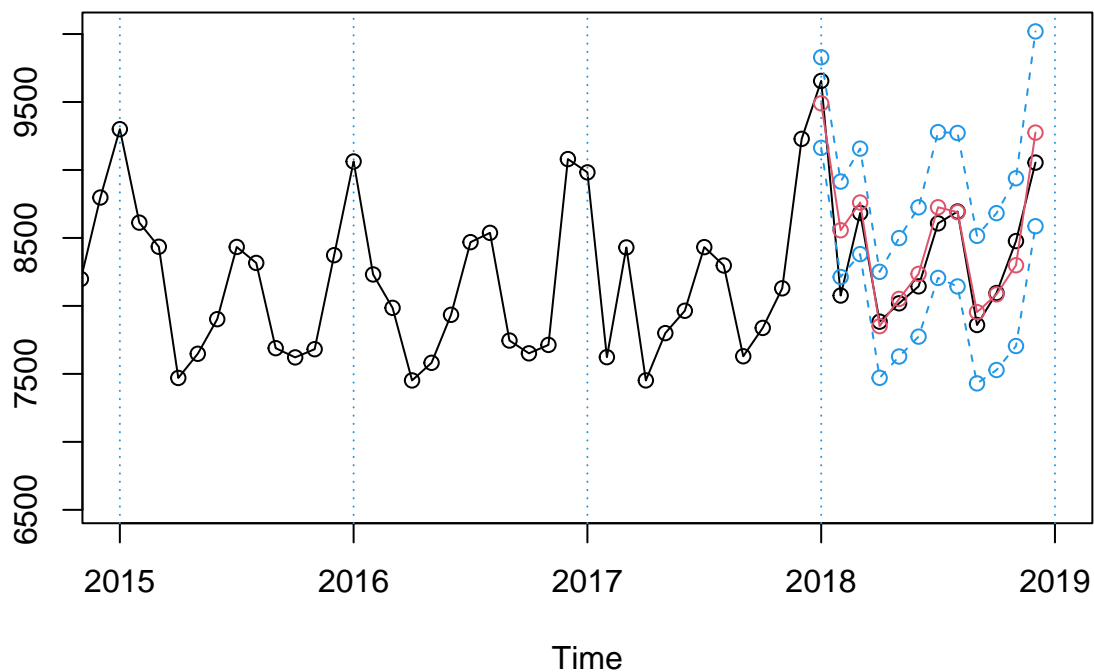
```

Coefficients:

	ar1	ar2	sma1	sma2
	-0.3953	-0.2848	-0.6926	-0.1656
s.e.	0.0538	0.0546	0.0593	0.0582

sigma<sup>2</sup> estimated as 0.0003203: log likelihood = 833.68, aic = -1657.36

Adjusted model without last 12 observations is also stable as the coefficients and absolute t-values are similar to the original adjusted model so we can use it for forecasting.



After comparing predictions using the adjusted model with previous ones we can see that the predictions are **more similar** to the original ones than with two models before. Also the confidence interval seems narrower which shows greater precision, less uncertainty around the forecast and stronger information.

#### 4.6 Comparison of three models

	par	sigma2	AIC	BIC	RMSE	MAE
ARIMA (model1)	4	0.0004676107	-1590.866	-1571.795	154.5936	114.9559
ARIMA (model2)	5	0.0004971541	-1574.634	-1551.749	208.4746	190.1275
ARIMA+OutTreat (model1)	18	0.0003164344	-1695.776	-1622.775	177.4293	125.4964
	RMSPE	MAPE	CIml			
ARIMA (model1)	0.01814751	0.01346135	1161.0585			
ARIMA (model2)	0.02496513	0.02269212	924.3824			
ARIMA+OutTreat (model1)	0.02128622	0.01482239	987.8202			



Simple model 1 is the **simplest** model, only **4** parameters, and had **lowest** forecast errors, RMSE/MAE/RMSPE/MAPE so it is the best on 1-year prediction.

The outlier adjusted model 1 has a **lot lower AIC/BIC** and residual variance ( $\sigma^2$ ) which is proof of accounting for level shifts and other outliers as it gives much better likelihood fit. The **drawback** is the it has 18 parameters (4 parameters of model 1 + 14 parameters for 14 outliers).

Model 2 has **narrowest** average confidence interval, but its variance errors are larger.

## 5 Conclusions and Discussion

The goal of this data analysis was to generate forecast based on last observation using the Box-Jenkins methodology through different ARIMA models. It also includes outlier treatment of one selected model. After generating forecast seasonal-ARIMA models were compared against each other in order to figure out which one is the best to predict U.S. monthly primary-energy demand one year ahead.

Three models were compared:

- **Model 1** –  $ARIMA(2, 1, 0)(0, 1, 2)_{12}$  (4 coefficients)  
Best pure forecast: lowest RMSPE = 1.8% and MAPE = 1.3 %. AIC = -1591.
- **Model 2** –  $ARIMA(0, 1, 2)(3, 1, 0)_{12}$  (5 coefficients)  
Narrowest prediction interval but higher error (RMSPE = 2.5%, MAPE = 2.3 %). AIC = -1575.
- **Model 1 + outlier treatment**

Same structure as Model 1 plus 14 LS/AO/TC outlier treatment (18 coefficients)

Gives the best in-sample likelihood (AIC = -1696) and lowest variance, but its one year forecast error lands between the first two models

All three pass causality and invertibility checks (roots outside the unit circle) but none completely remove residual autocorrelation (Ljung-Box  $p < 0.05$ ).

### 5.1 Strengths and weaknesses at a glance

aspect	Model 1	Model 2	Model 1 + OT
Forecast accuracy	Best	Worst	Medium
Prediction interval width	Wide	Narrowest	Medium
Simplicity (coefficients)	4	5	18
Homoscedasticity	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
Normality	<b>Yes</b>	<b>Yes (2/3)</b> (one test fails)	<b>Yes</b>
Residual independence	<b>No</b>	<b>No</b>	<b>No</b>
Causal / Invertible	<b>Yes / Yes</b>	<b>Yes / Yes</b>	<b>Yes / Yes</b>

### 5.2 Final conclusion

The objective of project **was met** as three ARIMA models were able to generate reliable forecasts that are similar to the original values. All models had really small RMSPE and MAPE which proves that point.

For further forecasting the plain  $ARIMA(2, 1, 0)(0, 1, 2)_{12}$  is **recommended** as it is accurate and has smallest number of parameters. Outlier adjusted model improves the model fit but at the cost of 14 extra parameters without fixing the residual dependence issue. Model 2 offers the tightest confidence bands yet makes larger point-errors.

Future work should focus on **fixing the remaining autocorrelation** by adding extra AR/MA lags which could make model fit worse or including calendar effects (holidays, working days, etc.) into a consideration.