Time Series Final Project

Analysis of the monthly users of the Barcelona metro

Victor Monge, Ximena Moure

January 2024

Contents

1 Introduction

The objective of this project is to examine the trends and patterns in the usage of the Barcelona Metro over the past two decades. By employing the Box-Jenkins ARIMA methodology on the provided data set, our goal is to conduct a comprehensive time series analysis and to generate future forecasts. The initial phase will involve an exploratory data analysis to assess the key characteristics of the time series data. Subsequently, we aim to identify and estimate a range of suitable models, selecting the most effective one for predictive purposes. Finally, we will evaluate the time series for potential anomalies that could influence the accuracy of our predictions.

1.1 Dataset

The dataset comprises data detailing the volume of passengers conveyed by the metro system within the Barcelona metropolitan area.

We will start by plotting the time series.

From it we can observe the following (see Figure 1):

- There appears to be an overall increasing trend in the usage of the metro over time.
- There are regular fluctuations within each year, which indicates seasonality. This might be due to seasonal variations in tourism, local events, or weather conditions affecting how often people use the metro. For instance, we know that during summer the number of metro passengers drops.
- The variability in the number of passengers seems to change over time. In certain periods, the swings between the high and the low passenger counts within a year are more pronounced.

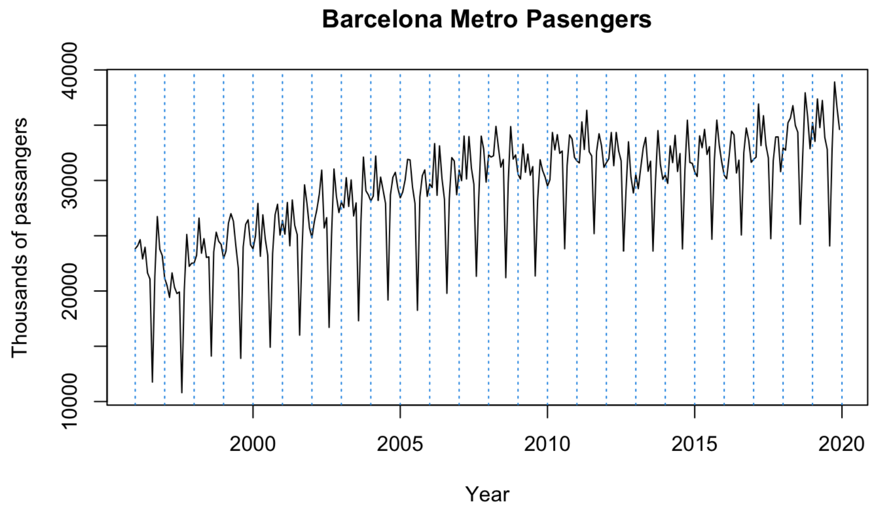

Figure 1: Barcelona monthly metro passengers

2 Identification

2.1 Stationarity

In the following analysis, we will conduct a test to determine the stationarity of the time series. Should the series prove to be non-stationary, we will apply the necessary transformations to stabilize its variance and mean. Achieving stationarity is essential for the reliability of parameter estimates, which in turn is vital for drawing meaningful and statistically significant conclusions.

2.1.1 Constant variance

Ensuring constant variance, also known as homoscedasticity, is a fundamental requirement for a time series to be considered stationary. The assessment of this property can be effectively conducted through the use of Box plots and Mean-Variance plots, which provide a visual representation of variance consistency across the time series.

The mean-variance plot (Figure. 2 does not exhibit a discernible trend with regard to the variance of the dataset, implying that the time series could be stationary. Additionally, the boxplot visualization (Figure. 3 generally indicates that the variance remains relatively consistent over the observed period. Consequently, it can be inferred that the variance is sufficiently constant, obviating the need for data transformation to stabilize variance before further analysis.

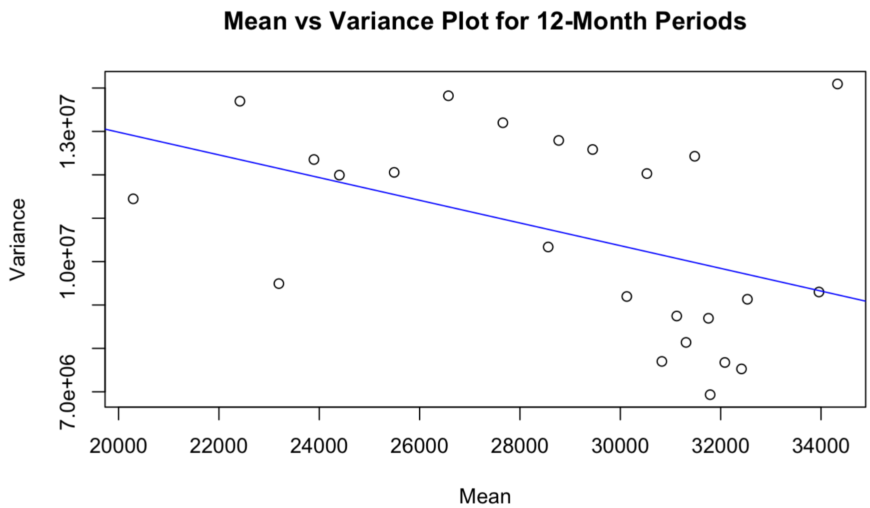

Figure 2: Mean vs Variance plot

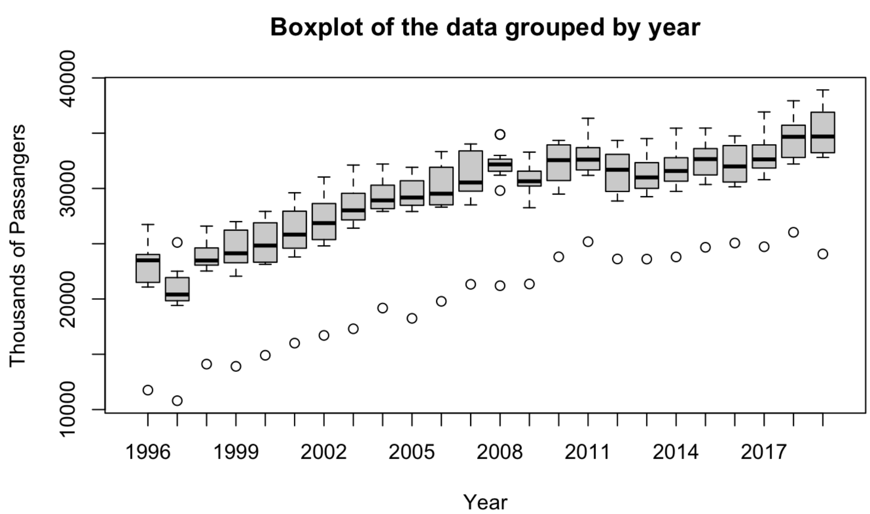

Figure 3: Boxplot of the data grouped by year

It should be noted that we explored the possibility of a logarithmic transformation to assess its impact on the data's variance. Upon application, we observed that the transformation led to worst results. Consequently, we decided to proceed with our project without applying the logarithmic transformation.

2.1.2 Seasonality

As previously noted, the data exhibits a discernible seasonal pattern, which aligns with expectations since the usage of the metro system is influenced by seasonal variations. This cyclical behavior is typical for transportation data, reflecting changes in passenger numbers due to factors such as weather, holidays, and school schedules.

From the plots (Figure. 4, we can observe that there is a clear seasonal pattern. In August, the number of passengers using the metro decreases. The reason behind this, can be the fact that during the holidays in August people tend to travel abroad more and consequently they do not use the metro. There is a consistent trough around the middle of the year, suggesting a regular annual low point in the data. There are peaks that occur in certain months, and these peaks appear to be consistent across multiple years.

Since we found that seasonality is present we need to eliminate it by applying a seasonal difference. $(1 - B^{12})X_t$.

2.2 Constant mean

Next step is to check whether the mean is constant.

From Figure 5. we can observe that it appears that the mean of the series is relatively close to zero, which would suggest that the seasonal differencing has made the mean of the series relatively constant over time.

We decided to test applying one regular difference. After applying it we can see (Figure 6) that the the mean appears to hover around zero, which is an indication that the mean is constant over time — a key characteristic of a stationary series. Additionally, the variability (or variance) doesn't seem to be changing dramatically over time, which is another property we'd like to see in a stationary series.

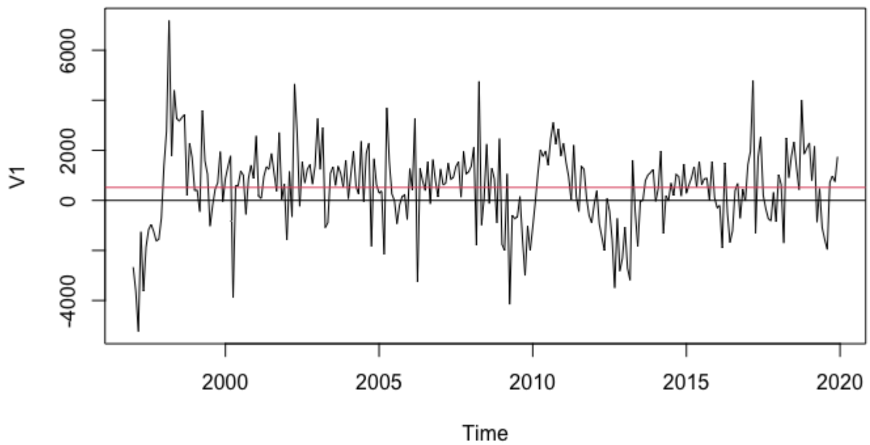

Figure 5: Time series without seasonal pattern

We also tried applying one more regular difference, but it became evident that this was not necessary.

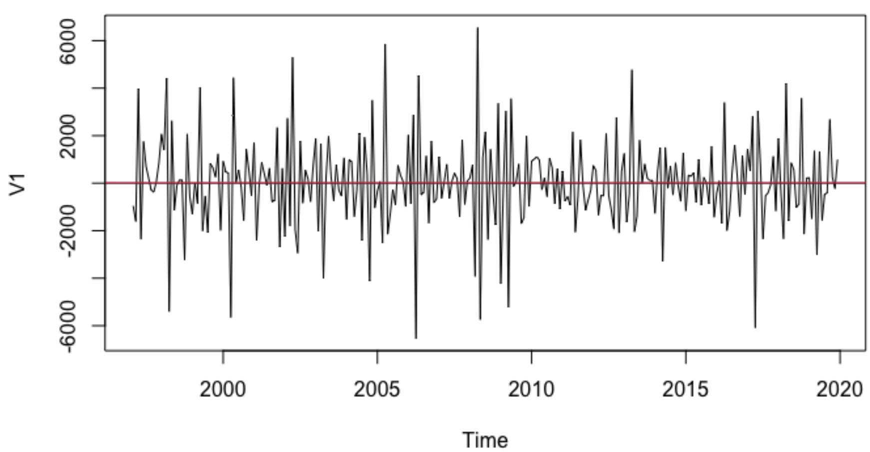

Figure 6: Time series without seasonal pattern and one regular difference

Table 1 depicts the variance of the time series with the different regular differences applied.

	d12serie	d1d12serie	d1d1d12serie
Variance	2710242	3628631	11357707

Table 1: Variance of the different transformed series

As a last visualization tool to check whether we should apply or not a regular difference, we plotted the ACF plots (Figure 9 for the time series with and without the regular difference. From the first plot (Figure 7) we can see that the ACF starts with a high value at lag 1 and then decreases but does not decay rapidly towards zero, as the subsequent lags hover above or below the significance bounds. The second plot (Figure 8)shows an ACF where the initial lags quickly approach the zero line and stay within the bounds of significance, indicating a rapid decay towards zero. In summary, it seems that the second plot depicts an ACF that decays more rapidly towards zero, which is more indicative of a stationary time series.

Despite the variance rising after applying a single regular differencing, the mean stabilizes notably, converging towards zero. This stabilization of the mean justifies our decision to proceed with the time series that has been subjected to one regular differencing.

Consequently, we suggest the following transformation to stabilize the time series:

$$W_t = (1 - B)(1 - B^{12})X_t \tag{1}$$

This equation represents the series d1d12metro, which has undergone both seasonal differencing (with a period s=12) and regular differencing. The transformed series is deemed to be stationary, exhibiting a mean that is approximately zero. Consequently, our model parameters are set with d=1 and D=1, indicating one non-seasonal and one seasonal difference, respectively, with s=12 reflecting the annual seasonality cycle.

2.3 Identification of plausible models

Moving forward, we will explore various configurations of the seasonal ARIMA model, denoted as (p,d,q)(P,D,Q)s. Given that we have established d and D at 1, and the seasonal period s at 12, we will utilize the ACF and PACF plots to discern the likely values for p, q, P, and Q. This approach will guide us in pinpointing the optimal parameters for the ARIMA model to capture the underlying patterns in the time series effectively.

From Figure 12 we can observe the following:

- Looking at the lags of the PACF, the model suggests non-seasonal AR(4), AR(6), depending on if we consider the lag after lag 4 close enough to the confidence band to be worth considering. For the seasonal part only the first seasonal lag being significant, so we would have a seasonal AR(1), or AR(4).
  Arguably, from the PACF plot, if we consider the 4th and 6th lag to be random happenstance we could also consider an AR(2).
- Looking at the ACF, the model suggests non-seasonal MA(5) or MA(8), depending on if the lag 7 is a random happenstance or if it really is significant. For the seasonal part, due to how all the seasonal lags seem significant, we can test seasonal MA(1) and MA(6), but it seems likely that the seasonal AR component is better.

Thus, the following candidate models are decided:

- mod1: $ARIMA(6,0,0)(1,0,0)[12]$.
- mod2: $ARIMA(6,0,0)(4,0,0)[12]$.
- mod3: $ARIMA(6,0,0)(1,0,0)[12]$.
- mod4: $ARIMA(0,0,8)(1,0,0)[12]$.
- mod5: $ARIMA(0,0,5)(1,0,0)[12]$.
- mod6: $ARIMA(0,0,7)(1,0,0)[12]$.
- mod7: $ARIMA(0,0,6)(0,0,6)[12]$.
- mod8: $ARIMA(0,0,3)(1,0,0)[12]$.
- mod9: $ARIMA(0,0,3)(4,0,0)[12]$.
- mod10: $ARIMA(0,0,7)(4,0,0)[12]$.
- mod11: $ARIMA(0,0,3)(0,0,1)[12]$.
- mod12: $ARIMA(0,0,7)(0,0,6)[12]$.
- mod13: $ARIMA(0,0,3)(0,0,6)[12]$.
- mod14: $ARIMA(6,0,0)(0,0,6)[12]$.

3 Estimation

In this section we estimate the models proposed in the previous section on the transformed data series ‘d1d12metro‘ and check the T-ratios to see if the coefficients are significant.

From Table 5 we can observe that model 9 (mod9) and model 10 (mod10) are the ones with the lowest AIC. Thus, they are selected.

Model	**AIC**
mod1	4752.63
mod2	4678.93
mod3	4752.63
mod4	4751.84
mod5	4756.3
mod6	4750.71
mod7	4682.79
mod8	4755.18
mod9	4674.05
mod10	4672.8
mod11	4695.46
mod12	4684.26
mod13	4687.17
mod14	4692.97

Table 2: AIC for each model

From Figure 15 it can be seen that the intercepts in both models exhibit t-ratios below the treshold of 2, indicating a lack of statistical significance. Additionally, several other parameters also appear to be insignificant. It may be beneficial to explore whether setting these particular coefficients to zero could improve the AIC.

After removing some of the not significant variables, the result is two new models with a slightly reduced AIC.

Now, these two models can be fitted to the original data series ‘metro‘ as: $SARIMA(0,1,2)(4,1,0)_{12}$ (from now on model 1) and $SARIMA(0,1,6)(4,0,0)_{12}$ (from now on model 2).

```
Call:
arima(x = d1d12metro, order = c(0, 0, 3), seasonal = list(order = c(4, 0, 0),
    period = 12))

Coefficients:
          ma1     ma2      ma3     sar1     sar2     sar3     sar4  intercept
      -0.7008  0.1547  -0.0940  -0.6240  -0.6525  -0.3546  -0.4807     2.9902
s.e.   0.0634  0.0791   0.0676   0.0575   0.0701   0.0698   0.0609     8.4252

sigma^2 estimated as 1228986:  log likelihood = -2328.03,  aic = 4674.05

T-ratios: -11.05 1.96 -1.39 -10.86 -9.31 -5.08 -7.89 0.35
```

Figure 13: Output of the Seasonal ARIMA Model Fit for the Differenced Series d1d12metro with Non-Seasonal MA(3) and Seasonal AR(4) Components (model 9)

```
Call:
arima(x = d1d12metro, order = c(0, 0, 7), seasonal = list(order = c(4, 0, 0),
    period = 12))

Coefficients:
          ma1     ma2      ma3      ma4     ma5      ma6      ma7     sar1     sar2     sar3     sar4  intercept
      -0.7028  0.1544  -0.0230  -0.0946  0.0467  -0.1408  -0.0552  -0.6256  -0.6609  -0.3418  -0.4713     0.9074
s.e.   0.0638  0.0752   0.0764   0.0707  0.0749   0.0902   0.0760   0.0578   0.0709   0.0720   0.0628     4.5856

sigma^2 estimated as 1185245:  log likelihood = -2323.4,  aic = 4672.8

T-ratios: -11.02 2.05 -0.3 -1.34 0.62 -1.56 -0.73 -10.83 -9.33 -4.75 -7.51 0.2
```

Figure 14: Output of the Seasonal ARIMA Model Fit for the Differenced Series d1d12metro with Non-Seasonal MA(7) and Seasonal AR(4) Components (model 10)

Figure 15: Model 9 and 10 outputs

Upon fitting the two models, we observe that Model 1 contains a single insignificant parameter; however, nullifying this parameter results in a higher AIC, indicating a less optimal fit. Additionally, Model 1 estimates a white noise variance of 1240050 and has an AIC value of 4672.11. Similarly, Model 2 has one parameter that lacks significance, and setting this parameter to zero leads to a deterioration in the AIC value, suggesting a decrease in model quality. Model 2 is associated with an estimated white noise variance of 1190489 and an AIC of 4666.34.

The statistical expressions of our two final models are:

- $$(1 - B)^1 X_t = (1 + \theta_1 B + \theta_2 B^2)(1 - B^{12})^1 \Phi(B^{12}) Z_t \tag{2}$$

  where Z_t is white noise with $Z_t \sim \mathcal{N}(0, \sigma^2)$.

- $$\begin{aligned}(1-B)^1(1-B^{12})^1X_t &= \theta_1 B + \theta_2 B^2 + \theta_4 B^4 + \theta_5 B^5 + \theta_6 B^6 \\ &\quad - \Phi_1 B^{12} - \Phi_3 B^{36} - \Phi_4 B^{48} + Z_t \\ Z_t &\sim \mathcal{N}(0,\sigma^2)\end{aligned} \tag{3}$$

  where θ_3 and Φ_2 are fixed at zero.

4 Validation

4.1 Residual analysis

To validate our models, we employed the evaluation function provided by the professor. This approach was used to assess the performance of both models and conduct a thorough analysis of their residuals.

Lets first address model 1 (see Figure 16).

Upon examining the residual plot, we observe that the residuals predominantly oscillate around the zero line, suggesting a satisfactory model fit without obvious patterns or trends. Although there is a hint of variance fluctuation at the beginning of the series, this may potentially indicate the impact of outliers.

The plot of the square root of absolute residuals indicates relatively stable variance throughout most of the series, aside from the initial segment, where outliers might influence the variability.

The Normal Q-Q Plot reveals that the residuals align well with the theoretical normal distribution, except for several points in the tails, which suggests the potential presence of outliers or heavy tails in the distribution of residuals. Similarly, the histogram of residuals approximates a normal distribution, although with some slight discrepancies, especially in the tails, implying that the residuals may not be perfectly normally distributed.

The ACF and PACF show some lags outside the confidence band, too many to not consider that there might be some autocorrelation present in the data.

The results from the Shapiro-Wilk, Anderson-Darling, and Jarque-Bera tests lead to the rejection of their null hypotheses, indicating that the residuals do not follow a normal distribution. This deviation from normality could be attributed to potential outliers in the data. Conversely, the Studentized Breusch-Pagan test does not reject its null hypothesis ($p = 0.608$), suggesting that the assumption of constant variance in the residuals is reasonable (we can assume homoscedasticity). The Durbin-Watson test does not reject the null hypothesis of no autocorrelation (with $p = 0.5579$).

Ljung Box test results show that from lag 11 the points show some autocorrelation, confirming what was observed in the ACF/PACF plots.

The sample ACF and PACF are quite similar for the origin lags, losing some of the similarities the further from the origin the lags find themselves.

After all this analysis, we can conclude that the residuals do not follow a normal distribution, show homoscedasticity and are not independent, probing the model is not correct.

Moving on to Model 2 (see Figure 13), we observe results akin to those of Model 1, with a notable improvement in the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. In Model 2, the number of lags falling outside the confidence bands is within acceptable limits. Furthermore, the Ljung-Box statistic indicates an absence of autocorrelation, suggesting that the residuals, while not normally distributed, are both homoscedastic and independent. With effective outlier treatment, the primary concern of Model 2 should be addressed, making it a more suitable choice for our predictive analysis compared to Model 1.

Thus, model 2 is the one that we select to make the predictions.

4.2 Causality and Invertibility

For a model to qualify as causal, the roots of the characteristic polynomial for the Autoregressive (AR) component, denoted by $\Phi(B)$, must reside outside the unit circle. This implies that the reciprocals of these roots should be located within the unit circle. In a similar vein, a model is considered invertible if the roots of the characteristic polynomial for the Moving Average (MA) component, represented by $\Theta(B)$, are positioned outside the unit circle, or equivalently, if their inverses fall inside the unit circle.

4.2.1 Model 1

From Figure 18 we can see that for the AR roots, all of them are slightly greater than 1 (all values are more than 1.012), which suggests that the AR part of the model is causal. Additionally, the MA roots have moduli of approximately 2.087 and 5.041, which are both greater than 1, indicating that the MA part of the model is invertible. This is visually evident on Figure 19.

```
Modul of AR Characteristic polynomial Roots:  1.012843 1.018335 1.018335 1.018335 1.012843 1.018335 1.012843 1.012843 1.018335 1.018335 1.018335 1.012843 1.012843 1.018335
843 1.018335 1.012843 1.012843 1.018335 1.018335 1.018335 1.018335 1.018335 1.012843 1.012843 1.012843 1.012843 1.012843 1.012843 1.018335 1.018335 1.018335 1.018335 1.012
012843 1.012843

Modul of MA Characteristic polynomial Roots:  2.086977 5.041168

Psi-weights (MA(inf))

--------------------
      psi 1       psi 2       psi 3       psi 4       psi 5       psi 6       psi 7       psi 8       psi 9      psi 10      psi 11      psi 12      psi 13
-0.67752862  0.09504978  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000 -0.61543046  0.41697175
     psi 14      psi 15      psi 16      psi 17      psi 18      psi 19      psi 20
-0.05849653  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000

Pi-weights (AR(inf))

--------------------
         pi 1          pi 2          pi 3          pi 4          pi 5          pi 6          pi 7          pi 8          pi 9         pi 10         pi 11
-0.6775286208 -0.3639952506 -0.1822182528 -0.0888604125 -0.0428856676 -0.0206101044 -0.0098876623 -0.0047401883 -0.0022717931 -0.0010886510 -0.0005216588
        pi 12         pi 13         pi 14         pi 15         pi 16         pi 17         pi 18         pi 19         pi 20
-0.6156804224 -0.4170915239 -0.2240711553 -0.1121701626 -0.0547005812 -0.0263994599 -0.0126871114 -0.0060866182 -0.0029179509
```

Figure 18: Root Moduli and Polynomial Weights for ARIMA Model 1.

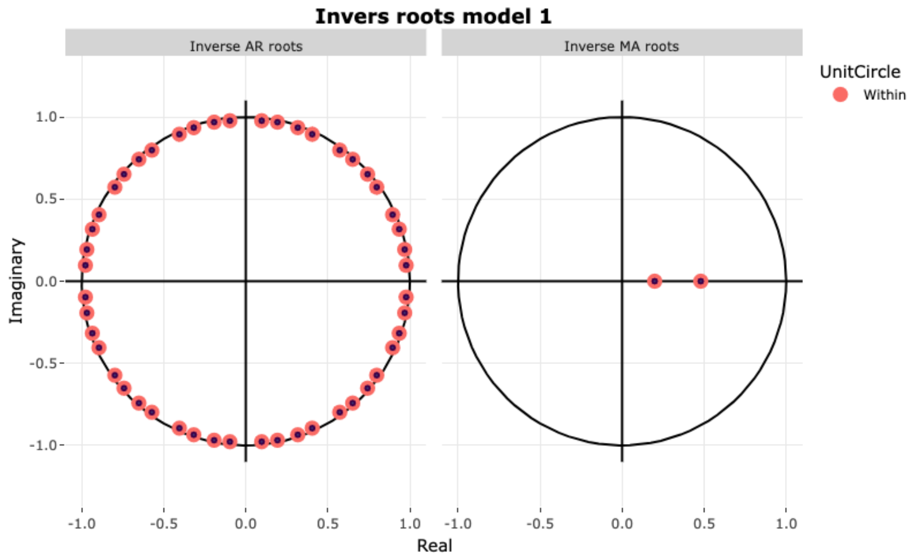

Figure 19: Inverse roots model 1.

4.2.2 Model 2

From Figure 20 we can see that the moduli of the AR characteristic polynomial roots are all greater than 1. This fact implies that the AR part of the model is causal. Also, the moduli of the MA characteristic polynomial roots are also all greater than 1, which indicates that the MA part of the model is invertible. This is visually evident on Figure 21.

```
Modul of AR Characteristic polynomial Roots:  1.01265 1.018358 1.018358 1.018358 1.01265 1.018358 1.01265 1.01265 1.018358 1.01265 1.018358 1.01265 1.01265 1.01265 1.01265 1.01265 
018358 1.018358 1.018358 1.01265 1.018358 1.01265 1.01265 1.018358 1.01265 1.01265 1.018358 1.01265 1.01265 1.018358 1.018358 1.018358 1.01265 1.018358 1.018358 1.01265 1.018358 1.

Modul of MA Characteristic polynomial Roots:  1.281479 1.527861 1.281479 1.10157 1.532022 1.527861

Psi-weights (MA(inf))

--------------------
      psi 1       psi 2       psi 3       psi 4       psi 5       psi 6       psi 7       psi 8       psi 9      psi 10      psi 11      psi 12      psi 13
-0.71054620  0.14154316  0.00000000 -0.07664134  0.00000000 -0.15457266  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000 -0.62359485  0.44309295
     psi 14      psi 15      psi 16      psi 17      psi 18      psi 19      psi 20
-0.08826559  0.00000000  0.04779315  0.00000000  0.09639071  0.00000000  0.00000000

Pi-weights (AR(inf))

--------------------
      pi 1       pi 2       pi 3       pi 4       pi 5       pi 6       pi 7       pi 8       pi 9      pi 10      pi 11      pi 12      pi 13      pi 14
-0.7105462 -0.3633327 -0.1575917 -0.1371903 -0.1296312 -0.2551096 -0.2848278 -0.2329500 -0.1595008 -0.1211180 -0.1053507 -0.7385946 -0.5661448 -0.3430194
     pi 15      pi 16      pi 17      pi 18      pi 19      pi 20
-0.1963259 -0.1662750 -0.1500319 -0.2235255 -0.2401464 -0.2047616
```

Figure 20: Root Moduli and Polynomial Weights for ARIMA Model 2.

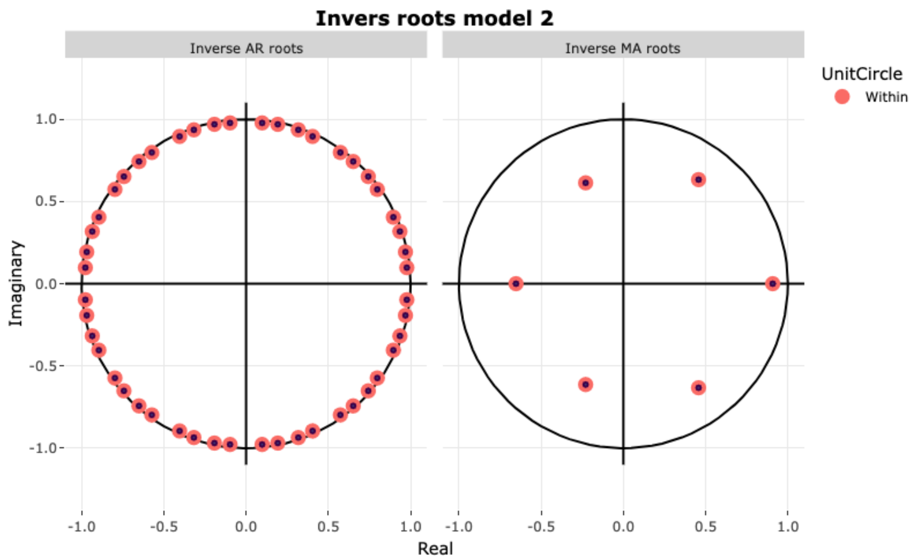

Figure 21: Inverse roots model 2.

4.3 Stability

To determine the most suitable model between the two, we reserve the last 12 data observations and retrain one of the models without them. Following this, we use the retrained model to make predictions for these 12 observations, providing a confidence interval for each prediction. Finally, we compare these predicted values to the actual values, which were initially set aside for this purpose.

4.3.1 Model 1

As depicted in Figure 22, the models can be considered stable as the signs and magnitudes of the coefficients are similar, and the significance levels have not changed drastically.

```
Call:
arima(x = serie1, order = pdq, seasonal = list(order = PDQ, period = 12))

Coefficients:
          ma1     ma2     sar1     sar2     sar3    sar4
      -0.6775  0.0950  -0.6154  -0.6458  -0.3438  -0.476
s.e.   0.0595  0.0674   0.0575   0.0702   0.0697   0.061

sigma^2 estimated as 1240050:  log likelihood = -2329.06,  aic = 4672.11

Call:
arima(x = serie2, order = pdq, seasonal = list(order = PDQ, period = 12))

Coefficients:
          ma1     ma2     sar1     sar2     sar3     sar4
      -0.7051  0.1234  -0.6216  -0.6916  -0.3445  -0.4801
s.e.   0.0609  0.0670   0.0575   0.0713   0.0694   0.0612

sigma^2 estimated as 1186384:  log likelihood = -2222.53,  aic = 4459.06
```

Figure 22: Stability check for model 1

4.3.2 Model 2

As depicted in Figure 22, the models can be considered stable as the signs and magnitudes of the coefficients are similar, and the significance levels have not changed drastically.

```
Call:
arima(x = serie1, order = pdq, seasonal = list(order = PDQ, period = 12), fixed = c(NA,
    NA, 0, NA, 0, NA, NA, NA, NA, NA))

Coefficients:
          ma1     ma2  ma3      ma4  ma5      ma6     sar1     sar2     sar3     sar4
      -0.7105  0.1415    0  -0.0766    0  -0.1546  -0.6236  -0.6665  -0.3477  -0.4779
s.e.   0.0612  0.0611    0   0.0507    0   0.0589   0.0576   0.0699   0.0701   0.0617

sigma^2 estimated as 1190489:  log likelihood = -2324.17,  aic = 4666.34

Call:
arima(x = serie2, order = pdq, seasonal = list(order = PDQ, period = 12), fixed = c(NA,
    NA, 0, NA, 0, NA, NA, NA, NA, NA))

Coefficients:
          ma1     ma2  ma3      ma4  ma5      ma6     sar1     sar2     sar3     sar4
      -0.7323  0.1609    0  -0.0671    0  -0.1237  -0.6375  -0.7137  -0.3532  -0.4823
s.e.   0.0630  0.0627    0   0.0516    0   0.0612   0.0579   0.0711   0.0697   0.0618

sigma^2 estimated as 1150807:  log likelihood = -2219.14,  aic = 4456.27
```

Figure 23: Stability check for model 1

4.4 Predictive power

Now, we perform out of sample predictions using the models obtained, to check if they correctly predict the data of the last year.

Tables 3 and 4 display the forecasting accuracy metrics for model 1 and model 2, allowing us to draw the following conclusions:

- RMSE (Root Mean Squared Error): Model 2 has a lower RMSE than Model 1 (1384.55 compared to 1541.137), indicating that Model 2 has a better fit in terms of the magnitude of errors.
- RMSPE (Root Mean Squared Percentage Error): Model 2 also has a lower RMSPE than Model 1 (0.0474867 compared to 0.0536584), which means that Model 2 has smaller error percentages relative to the actual values.
- MAE (Mean Absolute Error): Again, Model 2 performs better with a lower MAE (1101.47 compared to 1148.807), suggesting that, on average, the forecasts of Model 2 are closer to the actual values.
- MAPE (Mean Absolute Percentage Error): Model 2 has a lower MAPE (0.0347239 compared to 0.0368677), indicating that Model 2's predictions are, in terms of percentage, closer to the actual values.
- Mean CI (Mean Confidence Interval): The mean confidence interval for Model 2 is narrower (5317.951 compared to 5780.784), which suggests that Model 2 is more certain about its forecasts.

Additionally, from Figures 24 and 24 we can observe that Model 1 appears to show a slightly wider confidence interval than Model 2, which might suggest that Model 1 is less certain about its forecasts or that it predicts greater volatility in passenger numbers.

Table 3: Forecasting Accuracy Metrics for Model 1

Metric	Value
RMSE	1541.137
RMSPE	0.0536584
MAE	1148.807
MAPE	0.0368677
Mean CI	5780.784

Table 4: Forecasting Accuracy Metrics for Model 2

Metric	Value
RMSE	1384.55
RMSPE	0.0474867
MAE	1101.47
MAPE	0.0347239
Mean CI	5317.951

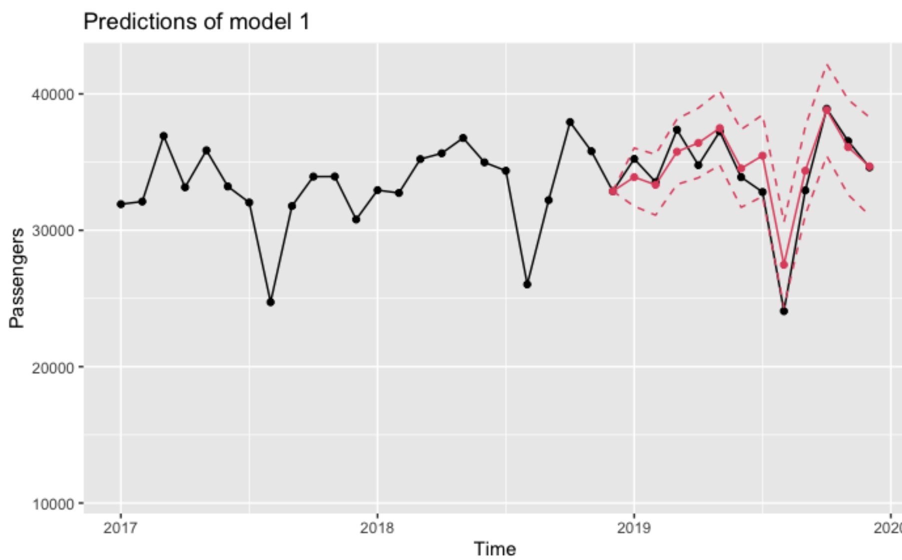

Figure 24: Predictions of model 1

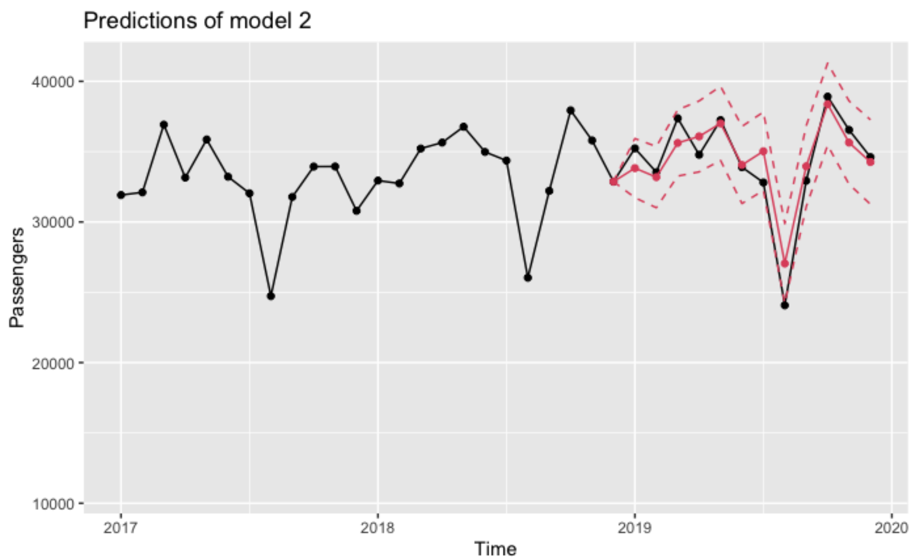

Figure 25: Predictions of model 2

4.5 Best model

Considering the comprehensive analysis and the metrics previously discussed, Model 2 emerges as the better choice. It demonstrates lower error rates in both absolute figures and in proportion to the data's magnitude, as evidenced by the percentage errors. Additionally, it offers a forecast range with greater precision, enhancing the reliability of its predictions.

5 Predictions

Figure 26 depicts the predictions for the next twelve months.

The forecasted values align with prior expectations, exhibiting no apparent anomalies. Furthermore, the confidence intervals appear to be well-calibrated, lending credibility to the forecast's reliability.

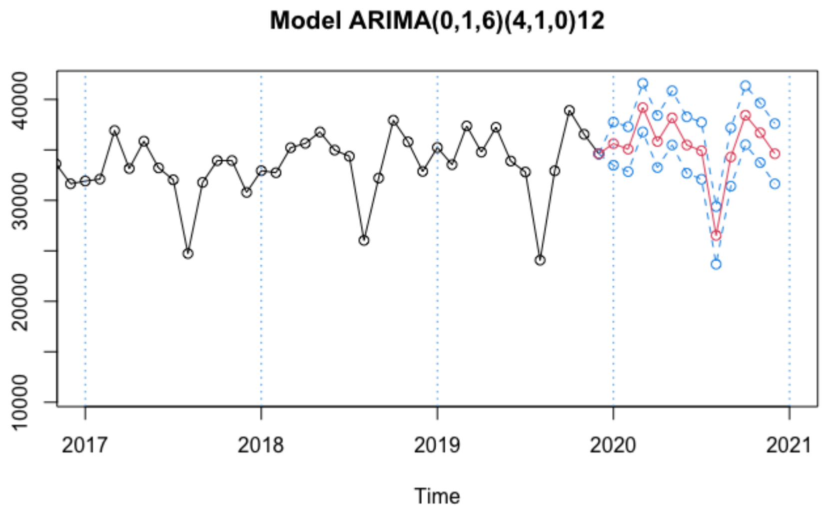

Figure 26: Predictions

6 Outlier detection and calendar effect

Following the generation of forecasts, it's critical to evaluate potential calendar anomalies or outliers that might refine our predictive accuracy.

Throughout this section we made use of the functions provided to us by the professor in order to analyze the calendar effect and the outliers.

6.1 Calendar Effect

We initiate this process by examining the influence of trading days and the Easter holiday. Trading days refer to the dates within a month when the stock market operates. Additionally, the occurrence of Easter, which alternates between March and April, may introduce seasonal fluctuations that warrant adjustment in the data.

We start by fitting the model to the series but taking into account the trading days and the Easter effect. Table 5 shows the AIC for each model. It can be observed that there is an improvement in the AIC when taking into account both the Easter week and the trading days.

Comparing the results obtained when validating the base model (Figure 17) and the model taking into account Easter and trading days (Figure 27) we can see the following improvements:

- the square root of absolute residuals show a flatter line, indicating constant variance for the residuals.

- the ACF/PACF plots show less lags outside the confidence bands, showing an improvement over the lack of autocorrelation of the points.
- The new model represents a marked advancement in terms of normality tests, successfully retaining the null hypotheses and thereby affirming that the residuals are normally distributed—a significant enhancement over the original model, which failed to demonstrate this attribute.

In summary, we can confirm that the model has the residuals normally distributed, with constant variance and that they are independent.

Model	AIC
Base Model	4666.34
Model with trading days	4608.75
Model with Easter	4574.64
Model with trading days and Easter	4502.83

Table 5: Comparative Analysis of AIC Values for ARIMA Models Adjusted for Trading Days and Easter Effects

6.2 Calendar Effect + Outliers

Next, we look at the outliers. To do so, we continue with the new model described in the previous section (the one that takes into account Easter and trading days).

Figure 28 shows all the different types of outliers detected. Note that AO refers to an additive outlier, while LS denotes a level shift and TC denotes a transitory change. Additive outliers impact only a single data point. Instead of observing the expected value in the linearized series, we observe a different data point. In the case of a level shift, there is a disruption in the series that remains uncorrected, resulting in a permanent effect. Transitory change affects the series for multiple time points but with decreasing intensity over time.

Additive outliers were identified at observations 73 in January 2002, 77 in May 2002, 116 in August 2005, and 149 in May 2008. While the specific catalysts for these anomalies remain largely undetermined, May 2008 coincides with record-breaking torrential rainfall, which may have contributed to atypical transit patterns. In January 2002, the introduction of the Euro could have played a role, as transitions to new currencies can disrupt economic activities and alter consumer behaviors. Such changes, alongside the potential logistical complexities during the currency switch, might have led to a temporary shift in metro usage patterns, reflecting in the observed passenger count deviations. In October 2017, a negative Additive Outlier (AO) was observed, which might be attributed to the aftermath of the terrorist attack that occurred in Las Ramblas, Barcelona, in mid-August 2017. The incident's lingering impact could have led to heightened apprehension among the public, resulting in fewer outings and, consequently, a reduction in metro usage during that period.

No information about what might have happened for the levels shifts was found. Observation 25 on January 1998 presents a positive level shift that affect all the future values. Another positive level shift happens for observation 59 on November 2000 and a negative shift for observation 157 on January 2009. Also, a positive level shift occurs in March 2012.

There are negative transitory changes for observations 34 October 1998, 40 April 1999 and 165 September 2009. No information was found about the possible causes of these changes.

After identifying the outliers, the subsequent step involves a comparison between the observed series and the linearized series (from which the outliers have been removed). From Figure 29 we can observe that the plot reveals that level shifts significantly influence the time series data. The original data predominantly resides above the adjusted series, with the level shift in January 1998 marking the onset of this divergence. This discrepancy is further exacerbated by an upward level shift in November 2000. Post the downward level shift in January 2009, both the original and adjusted series exhibit parallel patterns, with the original consistently registering slightly higher than the outlier-adjusted series

6.3 Identification and Estimation with outlier treatment

Next step is to repeat the process done before but now using the linearlized data series with the outliers treated.

We look at the ACF and PACF plots (Figure 30). From them a few new models can be identified looking at the lags near the origin outside the confidence bands of the ACF/PACF plots, one of them being an ARIMA model (0,0,7), another an ARMA(3,0,0) and an ARMA(5,0,0), with the possible seasonal components being (0,0,6) and (3,0,0). Seasonal component (4,0,0) and ARMA model (0,0,6) are also considered due to the original model. Table 6 shows the different models identified. We looked at the significant values and visualize that the intercept is not significant for any of the models and models 1 (mod.lin1_1) and 2 (mod.lin1_2) will end up as the original one. Thus, we continue with model mod.lin1_1 and model mod.lin4_1 which are the two models with the lowest AIC value.

For model mod.lin1_1 we remove some of the non significant values and we take into account the Easter week and the trading days. Once we do that, we analyze if it is invertible and causal by looking at the roots. Since not all the roots are larger than 1, the model is not invertible. However, it is causal because for MA all roots for phi are smaller than 1. We also take into account the Easter week and the trading days for model mod.lin4_1 and we observe that the AIC is better. We analyze if it is invertible and causal. Due to the absolute value of all the roots being larger than 1, model is causal and it is also invertible because for AR all roots for theta are smaller than 1.

The subsequent phase involves confirming the validity of the models. Given that both models failed the Ljung-Box test, we have opted to proceed with the model designated as mod.lin1_1, retaining all coefficients (we keep the coefficients that we removed before), including those that are not statistically significant. Following this adjustment, there is a slight deterioration in the Akaike Information Criterion (AIC) value; however, the model now successfully meets the criteria of the Ljung-Box test.

Model	**(p, d, q)**	**(P, D, Q)**	**Seasonal Period**	**AIC**
mod.lin1_1	(0, 0, 6)	(4, 0, 0)	12	4573.51
mod.lin1_2	(0, 0, 6)	(3, 0, 0)	12	4637.54
mod.lin1_3	(0, 0, 6)	(0, 0, 6)	12	4602.76
mod.lin2_1	(0, 0, 7)	(4, 0, 0)	12	4575.22
mod.lin2_2	(0, 0, 7)	(3, 0, 0)	12	4639.41
mod.lin2_3	(0, 0, 7)	(0, 0, 6)	12	4604.63
mod.lin3_3	(3, 0, 0)	(0, 0, 6)	12	4612.5
mod.lin4_1	(5, 0, 0)	(4, 0, 0)	12	4585.37
mod.lin4_2	(5, 0, 0)	(3, 0, 0)	12	4641.79
mod.lin4_3	(5, 0, 0)	(0, 0, 6)	12	4613.82

Table 6: ARIMA Models for d1d12metro.lin

6.4 Stability

We use the same procedure as we did before.

Results (see Figure 31) show that the model is stable, coefficients for both models are similar in significance, sign and magnitude, meaning the correlation structure has not changed and the complete series can be used for predictions reliably.

6.5 Predictive power

We applied the same procedure as before. Figure 32 shows that the predicted data seems to follow the same pattern as the observed data, with the observed data being inside the confidence intervals. Examining Table 7, it is evident that both the Root Mean Square Percentage Error (RMSPE) and the Mean Absolute Percentage Error (MAPE) fall below the 5% threshold, indicating that the predictions are notably accurate. Additionally, the average width of the confidence intervals stands at 3340.301. When juxtaposed with the initial model, which was the best one derived prior to adjusting for outliers and calendar effects, this model demonstrates a marked enhancement in performance.

6.6 Predictions

Figure 33 shows the predictions for the next twelve months. The predicted values conform to previous anticipations and display no noticeable irregularities.

Table 7: Forecasting Accuracy Metrics

Metric	Value
RMSE	1185.268
RMSPE	0.03918465
MAE	903.4818
MAPE	0.02820977
Mean CI	3340.301

Additionally, the confidence intervals are appropriately determined, enhancing the trustworthiness of the forecast's accuracy.

6.7 Comparison with the previous model

As mentioned before, if we look at Tables 4 and 7 we can see that the model with the best metrics is the one in which the outliers were treated. Additionally the model with the outlier treatment has a better AIC than the previous model. Finally, looking at the predictions of the model we can see that the model with outlier treatment has narrower confidence intervals. Consequently, we conclude that the model with the outlier treatment is a better fit for the data.

7 Conclusions

During the course of the project, we were able to apply the Box-Jenkins ARIMA methodology, along with its extensions for addressing calendar effects and outliers, to the analysis and forecasting of a real-world time series of our selection (Barcelona metro passengers).

Our final best model is an $ARIMA(0,1,6)(4,1,0)_{12}$ with treatment for trading days, Easter week and outliers. The accuracy of the predictions for 2020 is compromised by the unforeseen impact of the Covid-19 pandemic, which was not incorporated into the model. It's prudent to anticipate discrepancies in the forecasts for that year. Over time, as the ramifications of the pandemic become clearer, it may be possible to categorize it as an anomalous event within the model. However, due to the recency of the event and the ongoing nature of its effects, it is currently challenging to make adjustments to account for it.

References

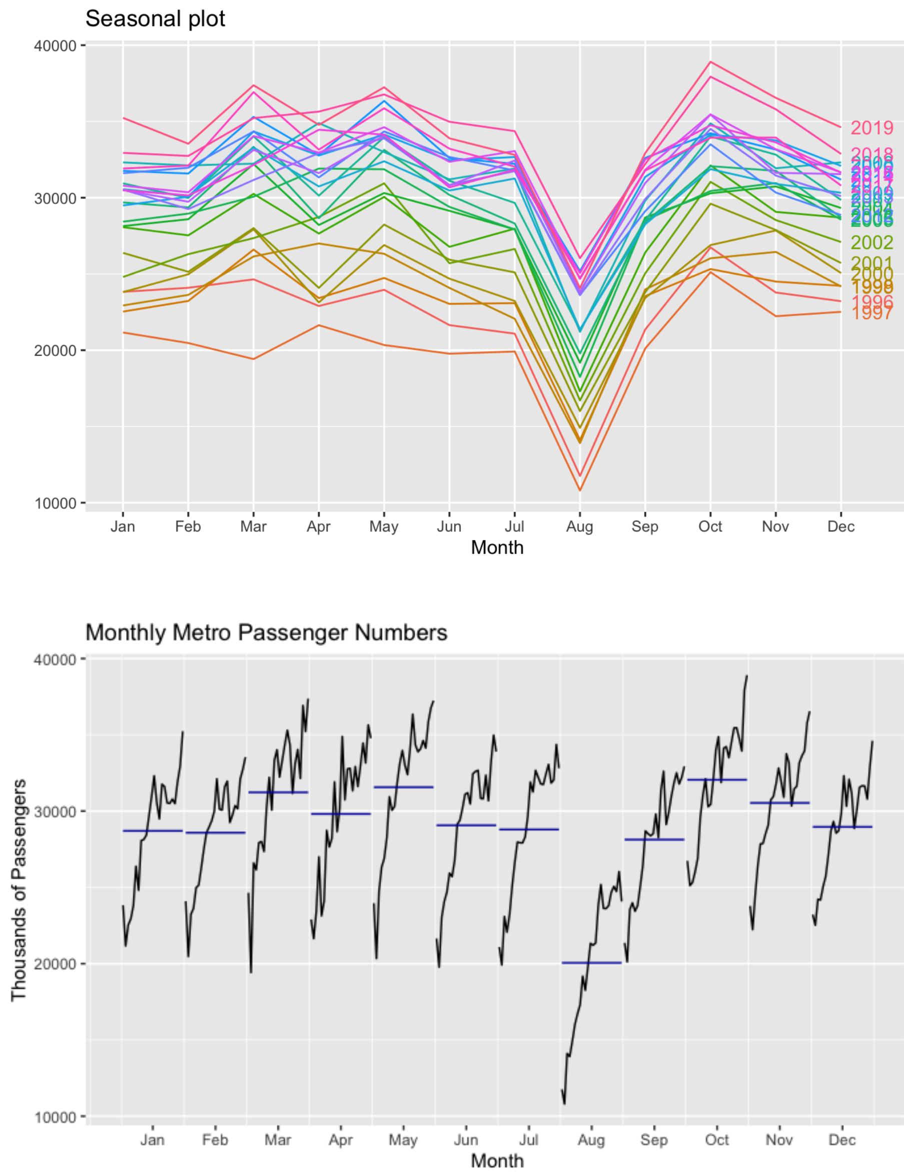

Figure 4: Seasonal plots.

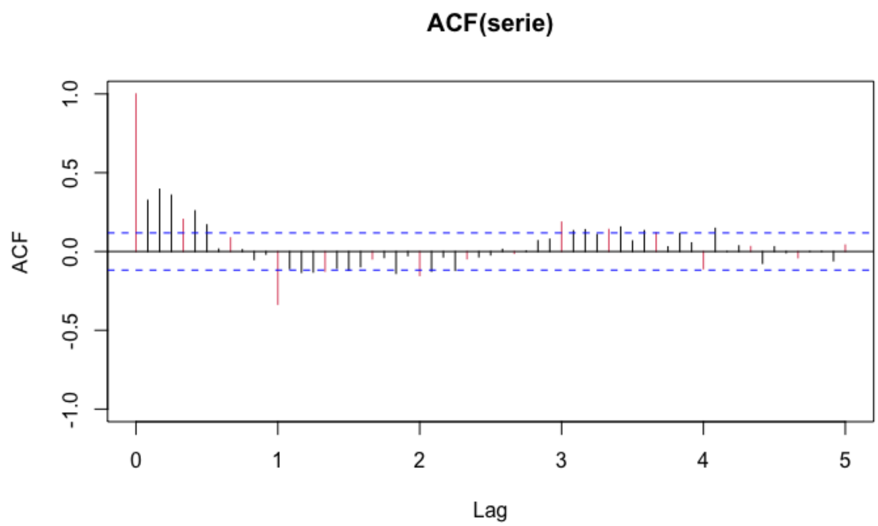

Figure 7: ACF plot time series d12metro

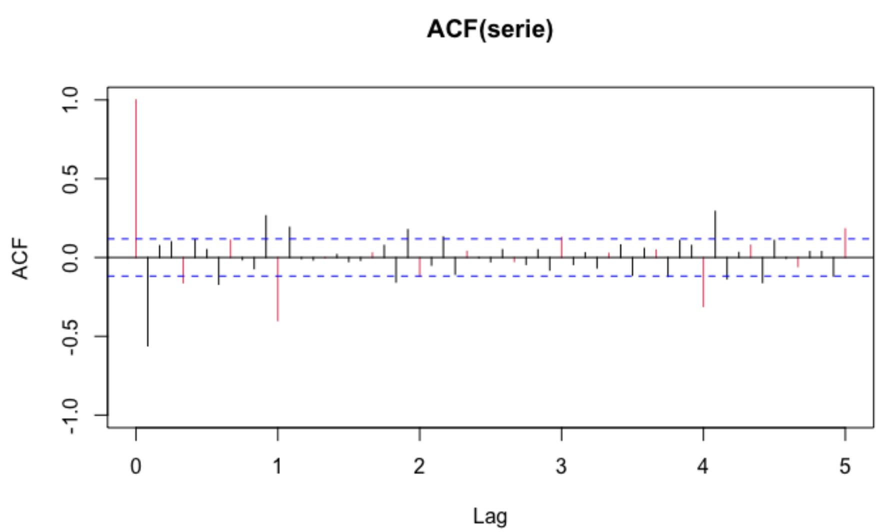

Figure 8: ACF plot time series d1d12metro

Figure 9: ACF plots

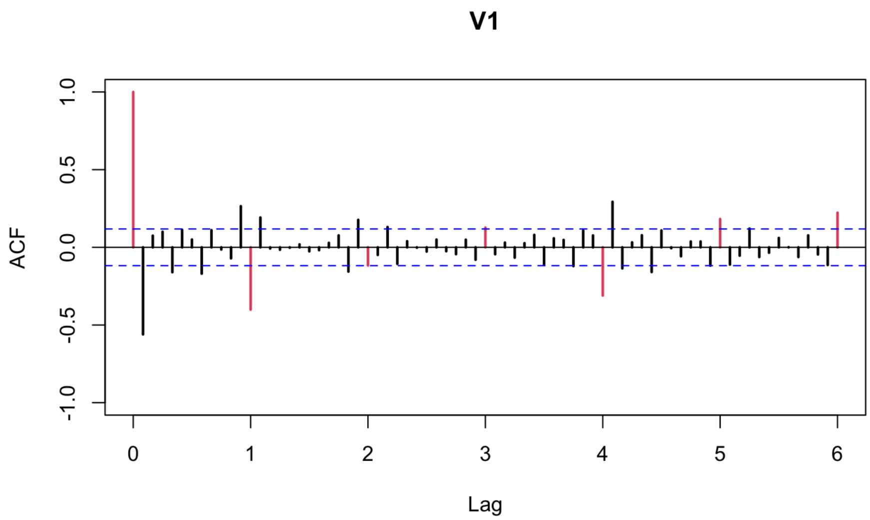

Figure 10: ACF plot

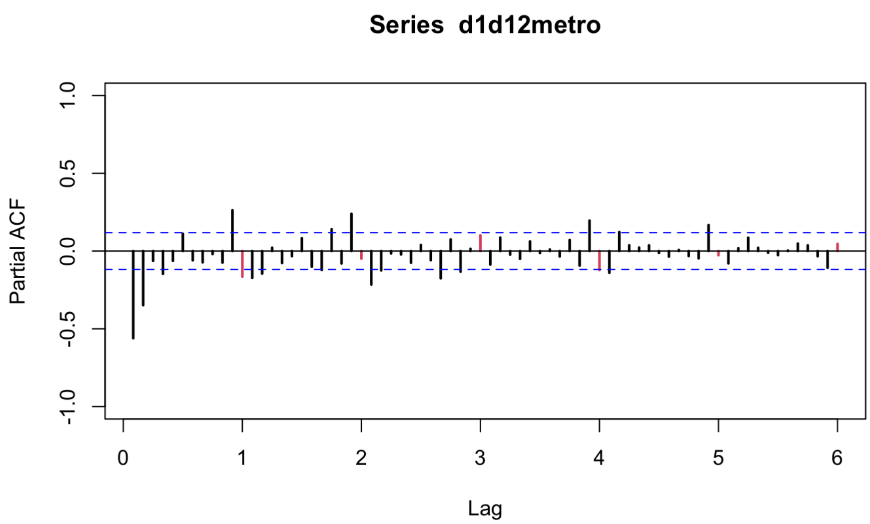

Figure 11: PACF plot

Figure 12: ACF and PACF plots

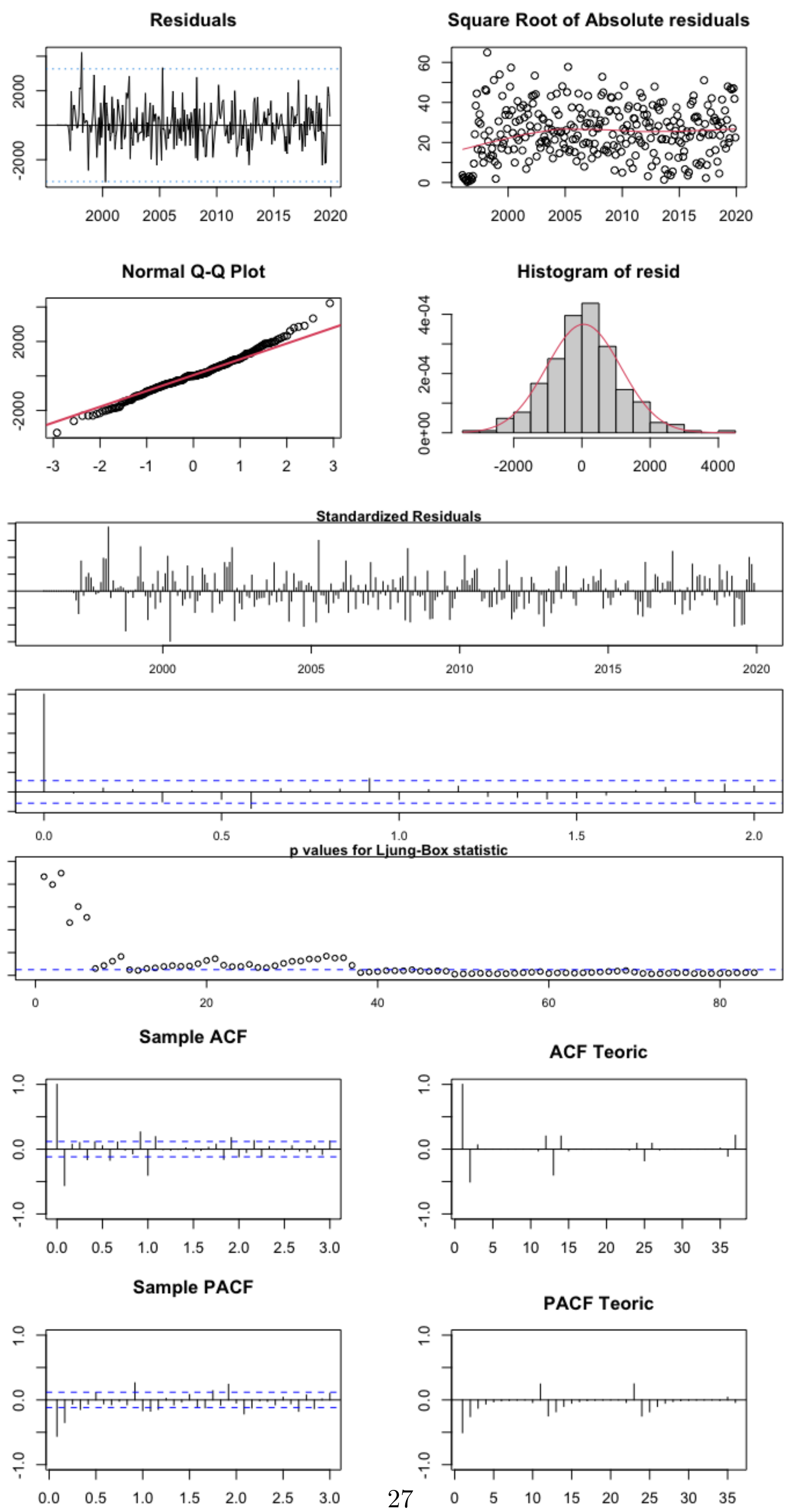

Figure 16: Model 1 validation

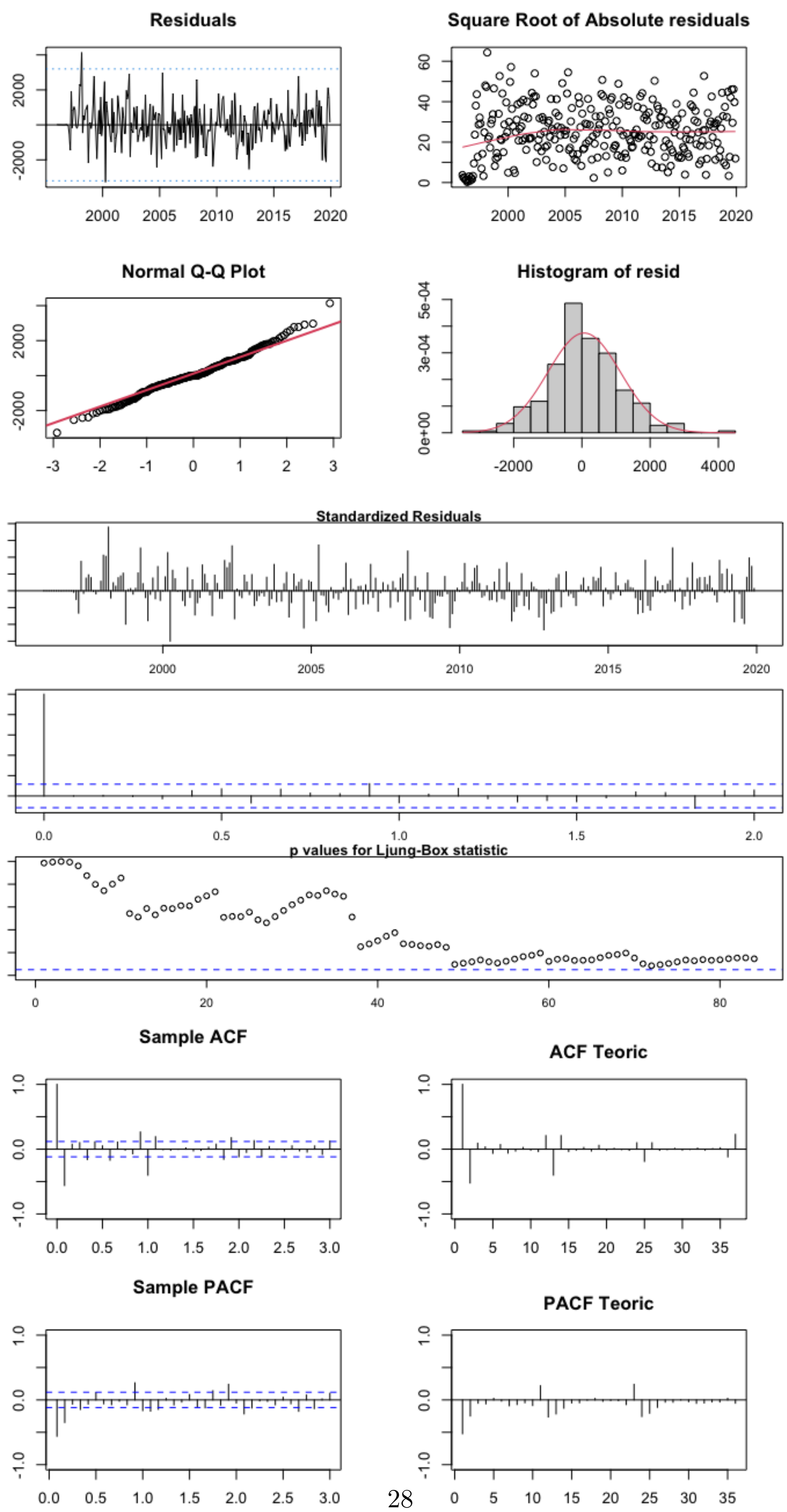

Figure 17: Model 2 validation

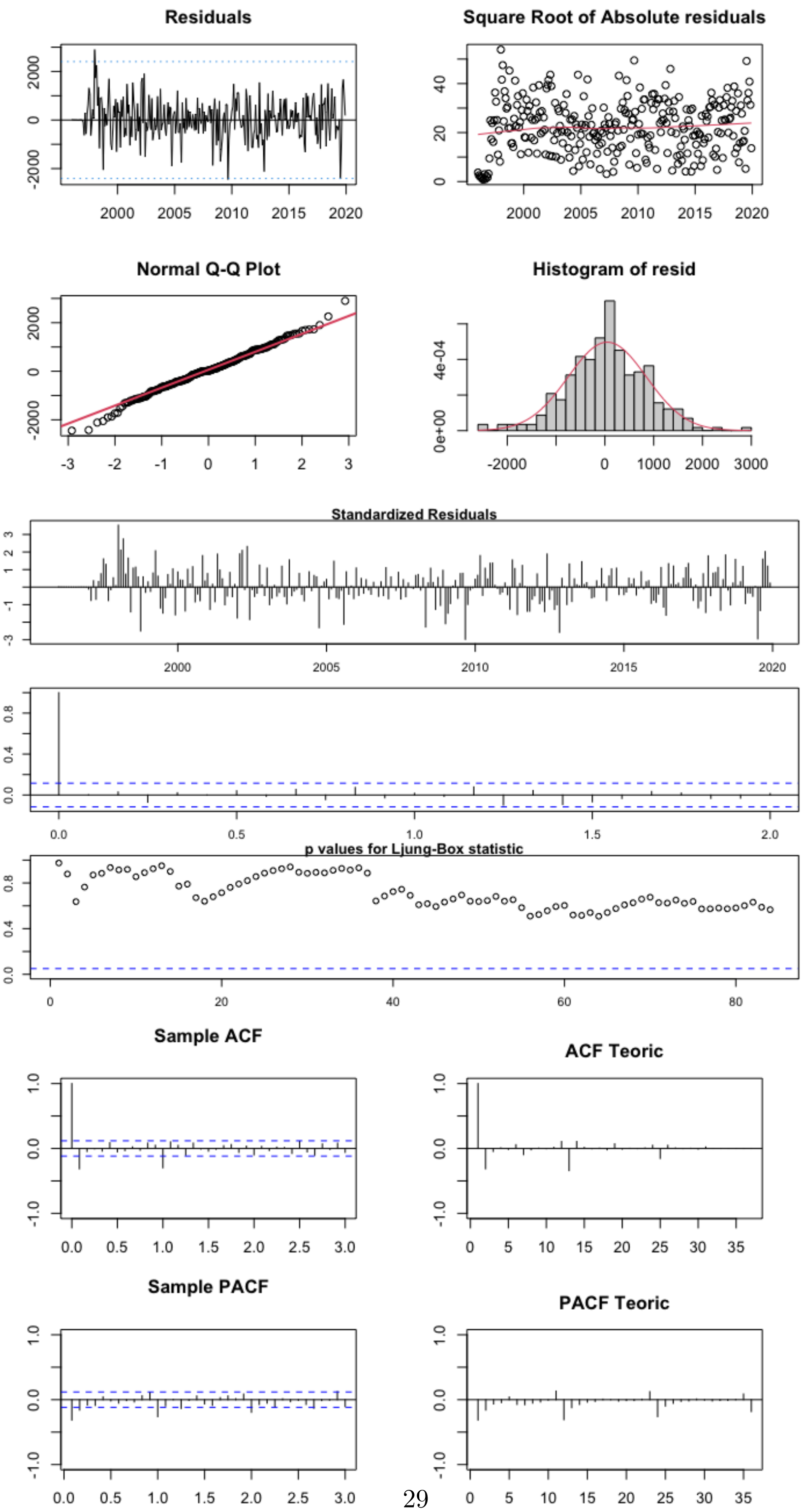

Figure 27: Model with Easter and trading days validation

	Obs	type_detected	W_coeff	ABS_L_Ratio	Fecha	perc.Obs
	<int>	<chr>	<dbl>	<dbl>	<chr>	<dbl>
1	25	LS	2555.822	4.193320	Ene 1998	Inf
6	34	TC	-1882.549	3.238132	Oct 1998	0
7	40	TC	1990.315	3.495359	Abr 1999	Inf
15	59	LS	1439.304	3.071668	Nov 2000	Inf
3	73	AO	-2095.782	3.838070	Ene 2002	0
5	77	AO	1682.518	3.189023	May 2002	Inf
9	116	AO	-1597.516	3.263475	Ago 2005	0
14	149	AO	-1387.872	3.089536	May 2008	0
12	157	LS	-1510.713	3.070984	Ene 2009	0
4	165	TC	-1902.247	3.158742	Sep 2009	0
11	195	LS	-1662.264	3.325053	Mar 2012	0
10	198	TC	1773.021	3.288814	Jun 2012	Inf
13	246	AO	-1447.500	3.081941	Jun 2016	0
8	262	AO	-1689.335	3.259113	Oct 2017	0
2	283	TC	-2872.740	3.972744	Jul 2019	0

Figure 28: Outliers detected

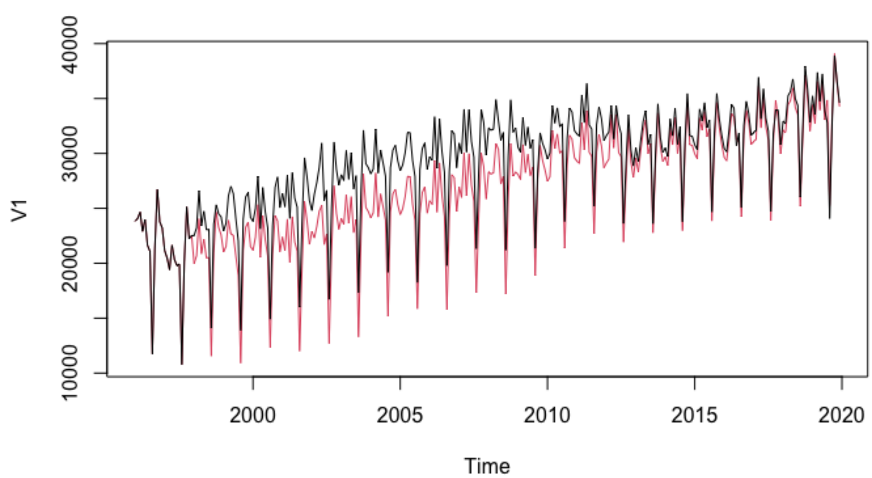

Figure 29: Comparison between the observed series and the linearized series

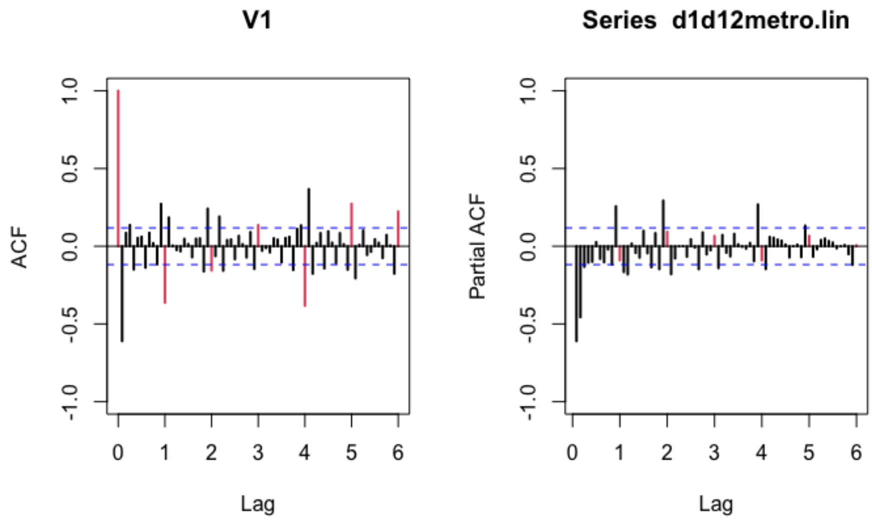

Figure 30: ACF and PACF plots for the linearlized model with treated outliers

```
Call:
arima(x = train, order = pdq, seasonal = list(order = PDQ, period = 12), xreg = data.frame(TradDays2,
    Eastw2))

Coefficients:
          ma1      ma2     ma3     ma4      ma5      ma6     sar1     sar2     sar3     sar4  TradDays2     Eastw2
      -0.4329  -0.0085  0.0035  0.0056  -0.0854  -0.1743  -0.4588  -0.4442  -0.3372  -0.2526   154.6070  -2314.3470
s.e.   0.0616   0.0694  0.0723  0.0620   0.0619   0.0674   0.0673   0.0753   0.0738   0.0710    12.0268    123.4758

sigma^2 estimated as 378081:  log likelihood = -2066.51,  aic = 4159.03

Call:
arima(x = complete, order = pdq, seasonal = list(order = PDQ, period = 12),
    xreg = data.frame(TradDays, Eastw))

Coefficients:
          ma1      ma2     ma3     ma4      ma5      ma6     sar1     sar2     sar3     sar4  TradDays      Eastw
      -0.4302  -0.0144  0.0094  0.0005  -0.1038  -0.1641  -0.4512  -0.4416  -0.3167  -0.2458  158.0842  -2299.5680
s.e.   0.0601   0.0679  0.0702  0.0614   0.0603   0.0662   0.0652   0.0722   0.0725   0.0704   11.6040    120.2086

sigma^2 estimated as 379975:  log likelihood = -2161.11,  aic = 4348.21
```

Figure 31: Stability check

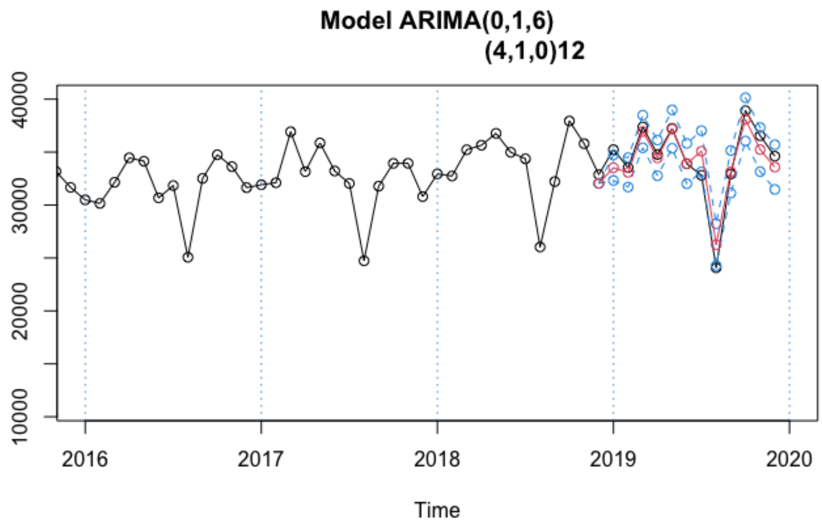

Figure 32: Predictive power

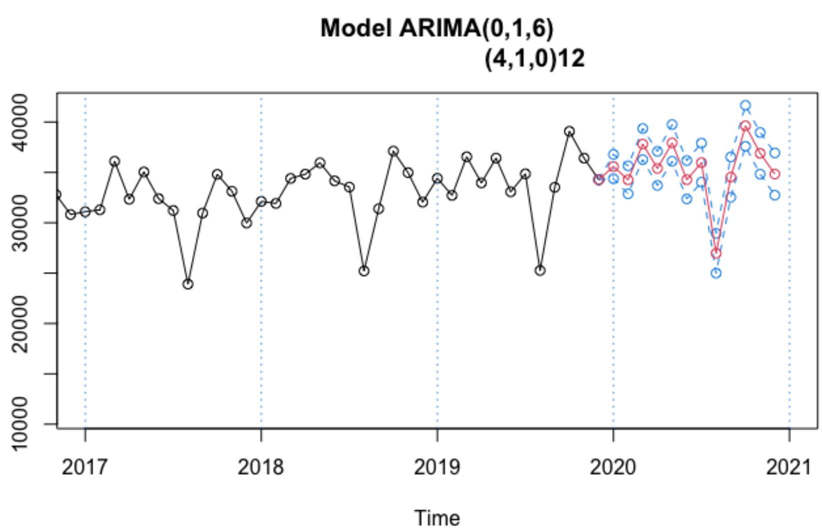

Figure 33: Predictions model with outlier treatment